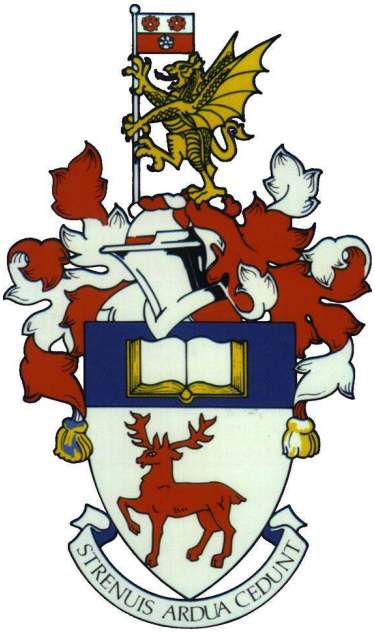
UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Clinical and Experimental Sciences



A genetic study of the childhood eye disorder: Congenital Nystagmus

by

Chelsea Sarah Norman

Thesis for Degree of Doctor of Philosophy

November 2019

University of Southampton

Abstract

Faculty of Medicine

Clinical and Experimental Sciences

Doctor of Philosophy

A Genetic Study of the Childhood Eye Disorder: Congenital Nystagmus

By Chelsea Sarah Norman

Nystagmus is an ocular disorder characterised by uncontrolled, repetitive eye movements. Nystagmus may occur alongside a broad spectrum of other diseases, or may occur in isolation. The variety of causal clinical conditions can produce varying, overlapping, hypomorphic phenotypes making early nystagmus presentations hard to diagnose. A multitude of genes are associated with nystagmus, but a molecular diagnosis is not always possible. Here, a large cohort of patients with varying phenotypes were studied to determine the genetic cause of nystagmus. Next Generation Sequencing techniques (NGS) were combined with detailed phenotyping information to inform on the best genetic workflow for efficient diagnosis of infantile nystagmus syndrome (INS).

One common cause of INS is albinism, which describes a broad group of pigmentary disorders that can be sub-grouped by genetic diagnosis and are usually considered phenotypically distinguishable. However, this work describes a hypomorphic albinism subtype that may be hidden behind a diagnosis of idiopathic INS. Cases of hypomorphic albinism are particularly susceptible to missing heritability which is prevalent through albinism and INS in general. A novel pattern of inheritance was identified in patients with hypomorphic albinism, whereby common population variants play a role in the disease. These SNPs are investigated in multiple family pedigrees as well as in a robust functional assay, providing proof of their pathogenicity.

Furthermore, a gene panel was analysed for its effectiveness in diagnosing a cohort of 81 patients with various presentations of INS. The panel was able to provide a diagnosis for 43% of the cohort, which is a high yield for a young INS cohort. The identified genotypes were grouped based on detailed clinical phenotypes to assess the correlation between the two. Overall, the two phenotype groups ‘idiopathic nystagmus’ and ‘albinism’ correlate with the identified causal genotypes but without complete fidelity.

Overall, this work has improved the diagnosis of nystagmus by improving the efficiency of genetic diagnosis and by providing evidence of a novel albinism genotype The commonality of the novel genotype provides an explanation for a large portion of the missing heritability in this cohort. It also sheds light on the complex inheritance of albinism disorders and how the current bioinformatics guidelines may need to be adjusted when analysing pigment gene variants.

Table of Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, Chelsea Sarah Norman

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

A genetic study of the childhood eye disorder: Congenital Nystagmus

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Norman, C. S. *et al.* (2017). Identification of a functionally significant tri-allelic genotype in the Tyrosinase gene (*TYR*) causing hypomorphic oculocutaneous albinism (OCA1B). *Scientific Reports*, *7*, 4415.

Arshad, M. W. *et al.* (2018). Mutations in TYR and OCA2 associated with oculocutaneous albinism in Pakistani families. *Meta Gene*, 17, 48-55.

O'Gorman, L., et al. (2019). A small gene sequencing panel realises a high diagnostic rate in patients with congenital nystagmus following basic phenotyping. *Scientific Reports* 9(1): 1-8.

Signed:

Date:

Acknowledgements

I would like to express my deepest gratitude to my supervisor Mr. Jay Self, whom has always provided guidance with seemingly boundless enthusiasm. His patience, kindness, motivation and knowledge have made him a brilliant PhD mentor. Further thanks go to the rest of my supervisory team, Dr. Arjuna Ratnayaka and Prof. Diana Baralle for their valuable ideas and guidance throughout the entire process. It has also been a pleasure to work with Luke O'Gorman whom has carried out the bioinformatics for this project. Having input from Luke and his supervisors has really helped focus the project and produce quality publications.

I must thank Vision Group in its entirety for their willingness to help, sense of humour and provision of moral support. You have all been a pleasure to work with. In particular Rhiannon Page and Tutte Newall who both provided a great deal of support and carried out many DNA extractions. I would also like to thank Dr. John Butler for his training and humour when setting up the techniques in Chapter 4. A large thank you goes to Chris, my fiancé, for his continued support and confidence in me. His patience has amazed me.

Finally, I would like to thank Gift of Sight, not only for their funding of this research project, but also for the encouragement provided by Ailsa and for the outreach opportunities I was able to experience.

Abbreviations

ACMG = American College of Medical Genetics and Genomics

ApE = A Plasmid Editor, produce by M. Wayne Davis, University of Utah

AROA = Autosomal Recessive Ocular Albinism

BSA = Bovine Serum Albumin

BCA = Bicinchoninic Acid

CEMAS = Classification of Eye Movement Abnormalities and Strabismus

CHS = Chèdiak-Higashi Syndrome

CMBS = Cross-McKusick-Breen Syndrome

CSNB = Congenital Stationary Night Blindness

DMSO = Dimethyl Sulfoxide

ELM = External Limiting Membrane

ERG = Electroretinogram

ES = Elejalde Syndrome

FHONDA = Foveal Hypoplasia, Optic Nerve Decussation defects and Anteria segment dysgenesis

GATK = Genome Analysis Tool Kit

GCL = Ganglion Cell Layer

GERP = Genomic Evolutionary Rate Profiling

GS = Griscelli Syndrome

HFEA = Human Fertilisation and Embryology Authority

HGMD = Human Gene Mutation Database

HGNC = HUGO (HUman Genome Organisation) Genome Nomeclature Committee

HPS = Hermansky Pudlak Syndrome

Chapter 1

HRP = Horseradish Peroxidase

INS = Infantile Nystagmus Syndrome

INL = Inner Nuclear Layer

IPL = Inner Plexiform Layer

IS/OS = Inner Segment/Outer Segment

IVF = *In vitro* Fertilsation

LB = Luria Broth

LCA = Leber Congenital Amaurosis

L-DOPA = l-3,4-dihydroxyphenylalanine

MBTH = 3-methyl-2-benzothiazolone hydrazine

MLPA = Multiple Ligation-dependent Probe Amplification

NGS = Next Generation Sequencing

OA = Ocular Albinisim

OCA = Oculocutaneous Albinism

OCT = Ocular Coherence Tomography

OKN = Optokinetic Nystagmus

OKR = Optokinetic Reflex

OMIM = Online Mendelian Inheritance in Man

ONL = Outer Nuclear Layer

PCR = Polymerase Chain Reaction

PMSF = Phenylmethylsulfonyl Fluoride

RNFL = Retinal Nerve Fibre Layer

RPE = Retinal Pigment Epithelium

SCA = Spinocerebellar Ataxia

SIFT = Sorting Intolerant From Tolerant

SOC = Super Optimal broth with Catabolite repression

TSO = TruSight One (targeted enrichment sequencing panel)

UKGTN = UK Genetics Testing Network

VCF = Variant Call Format

VEP = Visual Evoked Potential

VOR = Vistibulo-ocular Reflex

WES = Whole Exome Sequencing

WGA = Whole Genome Amplification

WGS = Whole Genome Sequencing

Chapter 1: Introduction

Eye movement disorders are detrimental to the day-to-day lives of those affected, causing partial sightedness, special learning requirements and social stigma[1]. Fortunately for diagnostic purposes, eye movements are readily available for clinical evaluation without the need for invasive investigation. The accessible nature of eye movements means that systematic examination is possible and they have been used as tools to investigate a broad range of neurological disorders[2]. Often, eye-movement abnormalities can point to a particular pathophysiology or pharmacological ailment, however, this is not always the case with nystagmus[1, 3].

1.1 Infantile Nystagmus Syndrome (INS)

Nystagmus is disorder of eye movement characterised by a repetitive to-and-fro oscillation of the eyes initiated by a slow drift away from the visual target[2]. Nystagmus can be a result of damage to the parts of the brain involved in oculomotor functions or visual pathways (acquired) or it can be inherited. There is variation over the nomenclature used to describe inherited nystagmus. The terms congenital, sensory/motor, and infantile are often used interchangeably to mean early-onset nystagmus. A national eye institute workshop determined the term 'infantile nystagmus syndrome' (INS) was most appropriate and has broadly been used since (Classification of Eye Movement Abnormalities and Strabismus (CEMAS) Workshop report, 2001). However, the CEMAS definition is purely a description of nystagmus onset/waveform and clinical features rather than referencing cause, thus 'congenital nystagmus' still occurs in the literature. Throughout this report 'INS' will be used to refer to the genetic disorder or early-onset, inherited nystagmus.

Nystagmus results in a functional impairment of daily activity and has been compared to that of patients with age-related macular degeneration[4]. INS has a prevalence of 1 in 1000 and most children with nystagmus are registered as sight impaired[5, 6]. Due to the continued maturation of oculomotor control systems postnatally, the genetic disorder generally presents at a young age, typically between two and five months, rather than at birth[1, 7]. Hence, there is a postnatal developmental window that may be exploited for treatment (after specific genetic diagnosis) using therapeutics to target the damaged visual control systems before they mature. However, current treatments only aim to manage the consequences of the disorder once they have manifest.

There are many ocular disorders where visual deficits and abnormal visual feedback during development are considered the likely cause of nystagmus. One prevalent syndromic link to nystagmus is albinism, where defects in the melanin biosynthesis pathway result in ocular abnormalities, such as hypopigmentation in ocular tissues, nystagmus, optic nerve misrouting and foveal hypoplasia. However, it is not clear which intermediate components in this pathway are responsible, or indeed how these abnormal features relate to the presentation of nystagmus[8].

The multitude of nystagmus manifestations, theories behind the mechanism, and controversy in the field leads to a common question of whether there is one underlying mechanism or whether there are multiple factors that cause disease. The possibility of many different underlying mechanisms is further endorsed by the multitude of genes that have been implicated in the disorder. Within the Online Mendelian Inheritance in Man (www.OMIM.org, Accessed: January 2019) database there are nearly 800 entries relating to nystagmus.

Currently, even after detailed ocular phenotyping, clear diagnosis for a child or adult with nystagmus is difficult for most cases. Targeted Sanger sequencing is occasionally used (usually in older children with the most unambiguous phenotypes) to provide a final diagnosis which can be crucial for treatment and genetic counselling. However, there is a lack of available gene panels, as well as, a shortage of knowledge about the genetics of nystagmus and how genetic mutations within a single gene can result in such varied phenotypes (allelic heterogeneity) and mutations in disparate genes can result in very similar phenotypes (locus heterogeneity).

1.2 The Eye

The eye is composed of several important structures that work together to focus light and send information to the visual cortex. The large structural components of the eye include the cornea, iris, lens, retina and optic nerve (*Figure 1*). The cornea is the initial transparent protective cover at the front of the eye, behind which lies the iris with a central pupil that contracts and relaxes dependent upon the amount of light. The crystalline lens is positioned behind the pupil where it can focus the incoming light onto the small central area of the retina known as the macula. The retina is a complex multi-cell layer at the back of the eye and containing photoreceptors, the retinal pigment epithelium (RPE), blood vessels and neuronal cells that transduce the image to the visual cortex. The macula is a region of the retina with few blood vessels and a high density of photoreceptors for high acuity vision. Within this area, there is a specialised cone-rich depression known as the fovea[9]. During retinal development, the outer and inner retinal layers regress to expose the photoreceptors and allow light to directly fall upon

them. Importantly, foveal development is thought to continue between 3 and 5 months postnatally, although morphological changes have been shown to occur up to the age of 12 years[10].

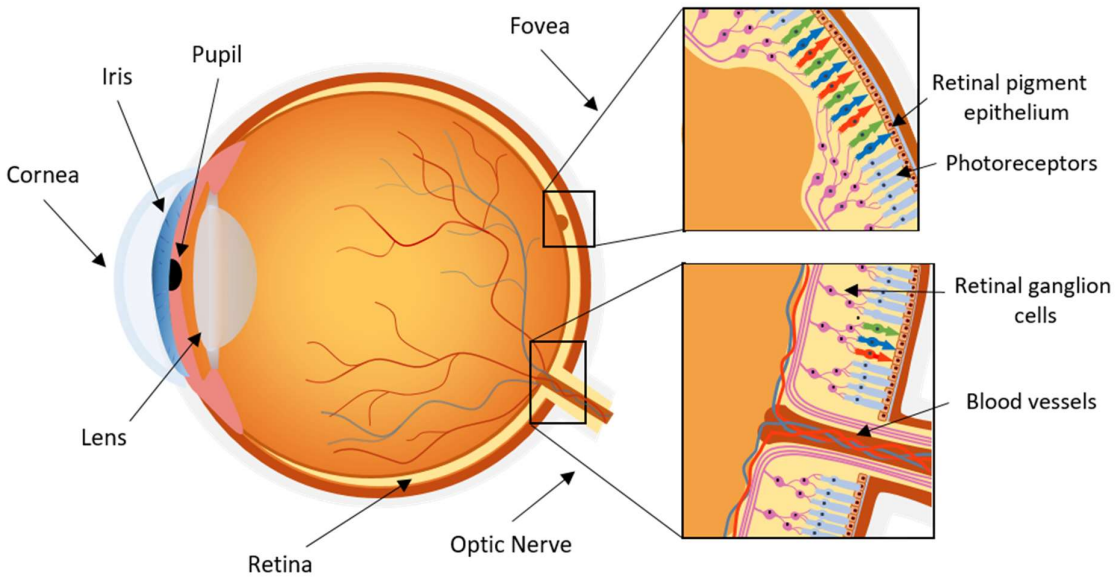


Figure 1. The structure of the eye, optic nerve and fovea. The main structural components of the front of the eye include the cornea, the iris, the pupil, and the crystalline lens which is held in place by intraocular muscles. Behind the lens is the vitreous humour which is transparent gelatinous mass in contact with the retina. The neural retina consists of several layers of interconnected neurons transporting signals from the photoreceptors, down the optic nerve, to the visual cortex. The photoreceptors and nerve cells are supported by a layer of pigmented epithelial cells. The fovea is a pitted region within the retina where the inner retinal layer has thinned to reveal densely packed cone-cells for high acuity vision[9].

1.3 Gaze stability

Gaze stability is a crucial component of normal vision function, enabling stabilisation of images on the retina. Two mechanisms of eye movement have evolved to achieve it. The first comprises the Vestibulo-Ocular Reflexes (VOR), which counteract movement sensed by the labyrinthine mechanoreceptors during head perturbations. The second consists of visually mediated reflexes (Optokinetic and Smooth-pursuit tracking), which react to the retinal-slip as an image 'drifts' across the retina. Together, these reflexes stabilize the angle of gaze, allowing eyes to remain directed towards the visual target regardless of head movement[2].

Detailed vision is a result of the fovea (photoreceptor-rich depression at the centre of the retina). Hence, it is necessary to change the line of sight to centre and hold an object of regard

on the fovea (foveation). Saccades are fast foveating eye movements that bring the object onto the fovea. The gaze-holding system then prevents the gaze from drifting off target[2]. The optokinetic reflex (OKR) and vestibulo-ocular reflex each provide physiological nystagmus in health. Optokinetic nystagmus (OKN) consists of an eye roll to stabilise the image of a moving object on the retina in healthy vision. OKN is composed of a slow, compensatory movement in the direction of moving object followed by a quick reset back to the start position. A vestibular nystagmus can be induced by self-rotation or by irrigating the ears with warm or cold water [11]. Gaze stability may be disrupted by unwanted nystagmus and abnormal saccades which pose a common diagnostic challenge for clinicians. Nystagmus is characterised by slow drift movement followed by a corrective movement in the opposite direction and is distinguished from saccadic intrusions as these are characterised by fast movements in both directions [2, 11]. INS differs to physiological nystagmus as it hinders vision and in most cases, does not cause oscillopsia (illusory motion of the visual world). Frequently, either horizontal OKN is not present in patients with nystagmus or an 'inverted' OKN is reported although the reasons for this debated observation are not clear[12-14]. Some individuals also exhibit end-point nystagmus, a jerky physiological nystagmus only seen when a healthy individual is made to fixate in extreme eccentric gaze[11, 14]. It is important to distinguish end-point nystagmus from its pathological variant; gaze evoked nystagmus, which is seen in less extreme eccentricities of gaze[15].

Acquired nystagmus can also be frequently attributed to disturbances of the mechanisms controlling gaze stability, most often abnormalities of vestibular input. However, acquired nystagmus is generally a result of diseases that cause progressive loss of vision, lesions in the cerebellum and brain injuries[11]. As it tends to occur after development of visual control systems, acquired nystagmus is usually accompanied by oscillopsia[11].

INS is most commonly horizontal but the oscillation can be vertical or torsional or a combination of all three (*Figure 2)*[2, 6]. The nystagmus develops at birth or a few months afterwards and persists throughout life. Usually, INS consists of slow-phase smooth pursuit movements away from an object of fixation followed by fast-phase saccades in the opposite direction. This produces a slow movement followed by a corrective movement which is either fast, making jerk nystagmus, or a second slow movement which creates pendular nystagmus. Nystagmus can also exhibit other, more atypical waveforms, such as see-saw nystagmus[3]. Frequently a patient can find a null zone, this is a gaze angle (orientation of eyes in orbit) in which the nystagmus oscillation is reduced and best vision is found[16]. There is an on-going debate over whether visual acuity is actually increased, or indeed whether the measured improvement in visual acuity is a consequence of the increased length of foveation[17, 18]. Increased periods of foveation allow for faster recognition of objects, faces, and letters on a

clinical chart. Therefore, there may be a temporal component in visual acuity measurements[18, 19]. The increased visual performance experienced in the null zone leads patient to use this gaze angle and an eccentric null point, which can lead to an abnormal head posture. Eye muscle surgery can be used to adjust eye position within the orbit, centring the null zone and correcting head posture[20]. There are many theories for the mechanistic cause of INS, implicating different ocular control systems including smooth pursuit, fixation, and optokinetic systems[8, 21-23]. Recent models have investigated retinal mis-wiring as the cause of nystagmus, though it is unknown whether this is a consequence of foetal development or due to abnormal afferent input during early visual development[8].

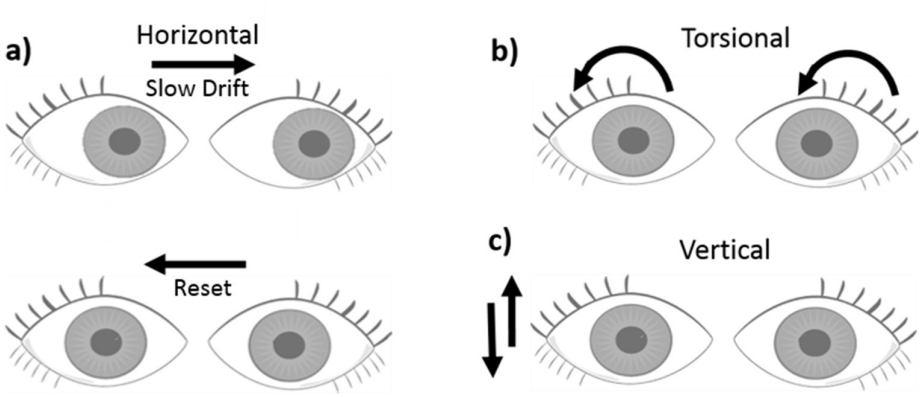


Figure 2. Example eye movements during nystagmus. a) Represents the horizontal jerk nystagmus that is most common b) Torsional nystagmus c) Vertical nystagmus, can be downbeat or upbeat but commonly represents acquired nystagmus rather than inherited.

1.3.1 Nystagmus waveforms

The INS waveforms, as measured through eye tracking, were originally classified as four different types, but the classification was expanded to an exhaustive list of 12 different waveforms by Dell'Osso and Daroff (1975)[24] with a further waveform described by Dell'Osso's group in 2007 [25]. However the most common waveforms are pendular, pendular with foveating saccades, pseudocycloid, jerk, and jerk with extended periods of foveation (represented in *Figure 3)*[26]. Furthermore, jerk nystagmus can also be described in terms of whether there is an accelerating, decelerating or linear slow phase.

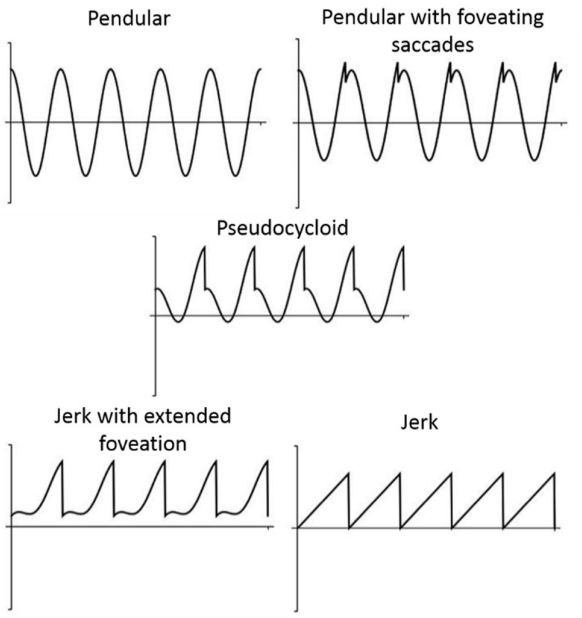


Figure 3. Common nystagmus waveforms plotted as eye position (vertical axis) against time (horizontal axis). Adapted from Theodorou and Clement (2016)[3].

Examination of nystagmus waveforms can sometimes be used to determine between INS and acquired nystagmus[15]. Attempts have been made to correlate the different waveforms of INS with corresponding ocular abnormalities[27, 28]. However, they remain limited in most cases as waveforms can vary widely across family members with the same genotype. The variability of INS means that although some clinical characteristics can suggest a given cause, there does not appear to be a clear deterministic relationship between the underlying mechanism and the waveform it produces, thus generating complex theories on its aetiology[29, 30],[31].

1.4 Idiopathic Infantile Nystagmus Syndrome

Idiopathic INS refers to isolated nystagmus, or often INS with no known ocular disorder as the cause. One current clinical challenge is the frequent mis-diagnosis of potentially subtle disorders such as ocular albinism. Detailed ocular phenotyping can sometimes reveal that cases with a diagnosis of idiopathic INS actually have underlying ocular abnormalities. The classic features of the idiopathic INS phenotype are nystagmus with accelerating slow phases, a loss of horizontal OKN, frequent null zones and convergence dampening and no apparent ocular or systemic cause for the nystagmus [32].

Idiopathic INS appears to be genetically heterogeneous and some idiopathic INS gene loci have been identified: *NYS1* (Xq26.2 MIM 310700), *NYS2* (6p12 MIM 164100) and *NYS3* (7p11.2 MIM 608345). *NYS1* has now been identified as the FERM domain containing 7 gene (*FRMD7*), resulting in X-linked nystagmus with variable penetrance in females [33]. *NYS2* and *NYS3* are both

associated with autosomal dominant idiopathic INS[34], but autosomal recessive inheritance has also been reported[35]. The role of X-inactivation has been investigated to explain apparent dominant inheritance of females affected by a single heterozygous *FRMD7* mutation[36]. The *FRMD7* gene was first discovered by Tarpey *et al*. (2006)[33], a significant discovery as it opens up the possibility of understanding the mechanism of idiopathic INS. The *FRMD7* gene is thought to hold a role in neurite outgrowth and mutations have been shown to cause a loss of horizontal direction selectivity in retinal ganglion cells in mice, which may explain the loss of OKN in both mice and humans[33, 37]. *FRMD7* is now known as 'the nystagmus gene' yet only accounts for approximately 50% of idiopathic nystagmus cases coming from large X-linked pedigrees, and less than 5% of cases with no family history[38]. The low number of cases accounted for by *FRMD7* mutations suggests that other causative genes exist for apparent IIN phenotypes[39].

1.5 INS associated disorders

Syndromic congenital nystagmus, generally refers to disorders of vision causing visual deficits with nystagmus as an accompanying condition. There are a wide range of syndromes that can result in nystagmus, some of which are degenerative or severely affect the quality of life, such as spinocerebellar ataxias that shorten life span. The results of a 2009 nystagmus survey by Sarvananthan *et al*. showed that nystagmus most commonly presents in association with disorders causing low vision. These disorders include aniridia, congenital cataracts, foveal hypoplasia and optic nerve hypoplasia. The survey revealed that albinism is one of the most common causes of INS, second to only that of idiopathic cases[40]. This high number of albinism cases may even be an underestimation due to reportsof subtle albinism initially being diagnosed as idiopathic IN[41].

In many cases, INS is caused by heterozygous mutations to the paired box protein Pax-6, encoded by the *PAX6* gene. *PAX6* codes for a transcription factor known to have a central role in ocular development and dominant loss of function mutations are known to result in aniridia, an ocular disorder characterised by iris malformation[42]. This is not always the case and mutations in *PAX6* are also associated with visual defects such as foveal hypoplasia, and presenile cataracts, corneal diseases and major ocular structural anomalies[43].

There are numerous retinal diseases associated with nystagmus, such as congenital stationary night blindness and achromatopsia. Achromatopsia is an inherited disorder characterised by poor visual acuity, pendular nystagmus and photophobia. Individuals with achromatopsia have impaired colour discrimination due to cone dysfunction. Diagnosis is established with a detailed ocular exam, colour vision assessment, and an electroretinogram

(ERG) which can be used to test cone function[44] (described in section 1.7.2.3). The UK Genetic Testing Network (UKGTN) approved a retinal gene sequencing panel (referred to as the Retinal Degeneration 105 Gene Panel) available on the NHS, that can be used to genetically diagnose retinal phenotypes. X-linked congenital stationary night blindness (CSNB) is retinal disorder caused by mutations to the gene *CACNA1F* and is characterised clinically by poor visual acuity, nystagmus, myopia and strabismus (abnormal alignment of the eyes/squint). Complete CSNB can be detected through obvious abnormalities on an ERG due to the absence of rod-dark adaptation. Incomplete CSNB, however, does not always cause detectable abnormalities on an ERG and has a heterogeneous phenotype, which may present as isolated nystagmus in young children[45].

Another significant cause of INS is in degenerative neurologic disease, one example of this is spinocerebellar ataxia (SCA). SCA is typically characterised by progressive loss of coordination, particularly of gait and fine motor skills, followed soon after by visual problems [46]. The *CACNA1A* gene, responsible for SCA type 6, has been reported to present as isolated infantile nystagmus, as well as nystagmus with late onset ataxia[47]. As nystagmus may occur without ataxic symptoms, *CACNA1A* is another important gene to consider when sequencing patients with apparent idiopathic INS. The significant overlap between some of these disorders and albinism/idiopathic INS have resulted in a confused picture regarding diagnosis and genetic testing [42]. It is apparent that detailed ocular phenotyping is crucial but also that genetic testing is often necessary to give an accurate diagnosis[48].

1.5.1 Foveal hypoplasia

Foveal hypoplasia is caused by a disruption of normal retinal development and can be responsible for a lack of high acuity vision. Normal foveal development begins in foetal week 25; progressing from an area with a thick layer of ganglion cells, to a pit exposing cone photoreceptors. Development of visual systems and structures has been shown to continue for a crucial period after birth. Formation of a normal fovea is thought to be fully completed between 15-45 months postnatally, though developments have been recorded in volunteers up to the age of 12 years old[10, 49]. Foveal hypoplasia is a characteristic morphological abnormality in various ocular disorders including albinism, *PAX6* mutant phenotypes, and achromatopsia, and can also occur in isolation. Foveal hypoplasia, optic-nerve-decussation defects, and anterior segment dysgenesis (FHONDA) syndrome has been recently described and associated with the gene *SLC38A8*[50]. The presence of a normal or hypoplastic fovea can be determined with optical coherence tomography (OCT) (described in section 1.7.2.4.) (*Figure 4*).

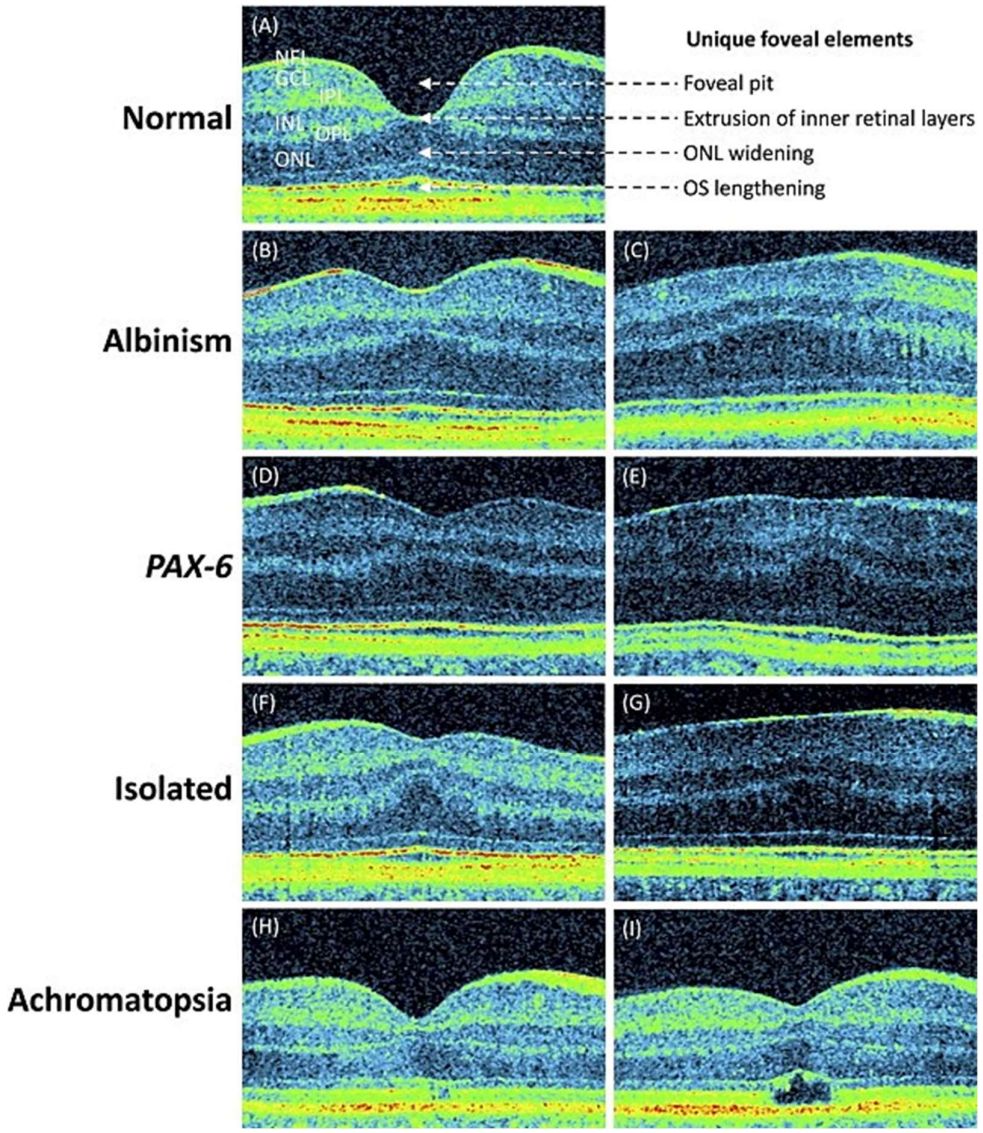


Figure 4. OCT scans of normal and hypoplastic foveas, by Thomas *et al.* (2011). (A) Normal fovea (B, C) Albinism (D, E) *PAX6* mutations (F, G) Isolated cases (H, I) Achromatopsia - A hyporeflective zone is also seen in (I) suggests photoreceptor degeneration. NFL = nerve fibre layer; GCL = ganglion cell layer; IPL = inner plexiform layer; INL = inner nuclear layer; OPL = outer plexiform layer; ONL = outer nuclear layer; OS = photoreceptor outer segment.

1.6 Albinism

1.6.1 Oculocutaneous albinism

Oculocutaneous albinism (OCA) refers to a group of inherited disorder of the melanin biosynthesis pathway, with a worldwide prevalence of approximately 1/17,000[51]. OCA causes varied levels of hypopigmentation in skin, hair, and eyes, as well as ocular abnormalities which can leave patients with severe ocular deficits and increased sensitivity to light (photophobia). Morphological changes of the retina and the optic nerve accompany reduced pigmentation of

the iris, retina and choroid. Characteristic ophthalmic features include reduced visual acuity, nystagmus, strabismus and photophobia[51]. Clinical examination using OCT may reveal foveal hypoplasia. A slit-lamp biomicroscope may reveal iris transillumination, and it is possible to determine excessive nerve decussation at the optic chiasm by measuring visual evoked potentials across the scalp[51].

There are seven reported OCA subtypes, each with varying phenotypes and a different associated gene (*Table 1*). OCA types 1-4 are widely recognised and well phenotyped, whereas, OCA types 5-7 are newly reported in the literature and OCA5 is currently attributed to a chromosomal locus.

Table 1. Genes and phenotypes associated with each albinism subtype.

Albinism subtype	HGNC symbol	HGNC name	Inheritance	Phenotype	Population Prevalence
OCA1A	*TYR*	Tyrosinase	Autosomal recessive	Complete OCA - White hair and skin	Most common OCA in Caucasians and 50% of all OCA worldwide.[52]
OCA1B				Partial OCA - Some pigment, wide range	
OCA2	*OCA2* *(P gene)*	OCA2 melanosomal transmembrane protein	Autosomal recessive	Partial OCA - Some pigment, wide range	Africa ~1 in 10,000 to 1 in 1000 and 30% of cases worldwide.[52]
OCA3	*TYRP1*	Tyrosinase related protein 1	Autosomal recessive	Red hair and reddish skin pigment	Southern Africa ~1 in 8,500 and 3% of cases worldwide.[52]
OCA4	*SLC45A2* *(MATP)*	Solute carrier family 45 member 2 (Membrane-associated transporter protein)	Autosomal recessive	Ranges from partial to near complete OCA	Prevalent in Japan, rare among Caucasians and Africans, 17% of cases worldwide. [52]
OCA5	-	Chromosomal location 4q24	Autosomal recessive	Yellow hair, white skin	1 case reported Pakistan.[53]
OCA6	*SLC24A5*	Solute carrier family 24 member 5	Autosomal recessive	White to brown hair, varied skin pigment	Reported in Caucasians, Africans, and Arabs.[54, 55]
OCA7	*LRMDA* *(C10orf11)*	Leucine rich melanocyte differentiation associated	Autosomal recessive	Partial OCA - Some pigment	1 case reported in Denmark. [56]

OCA1 is the most common OCA subtype in Caucasian populations and accounts for nearly half of reported OCA cases worldwide[51, 57, 58]. This has led to OCA1 being the best characterised subtype. Many studies have focused on the OCA1 *TYR* gene which encodes the enzyme tyrosinase, an important enzyme catalysing the initial rate-limiting step in the melanin synthesis pathway[59-63]. A common finding for OCA1 is a high level of missing heritability, wherein a patient presents with the phenotype of OCA1 but is missing a mutation in *TYR* (most often the

other OCA genes are also sequenced with no variants found). The incidence of missing heritability in OCA1 may be as high as 30%[57, 64]. Hence, the genetic cause is not yet fully understood for even some of the obvious clear-cut cases of OCA1.

1.6.2 X-linked ocular albinism

X-linked ocular albinism (OA), like OCA, is a disorder affecting the melanin biosynthesis pathway, particularly affecting melanosome biogenesis. Unlike OCA, the hypopigmentation exhibited in OA is localised to the eyes, causing hypopigmentation of the iris pigment epithelium and ocular fundus. Skin can also be mildly hypopigmented when compared with unaffected family members. Currently there is a single gene known to carry causal mutations for X-linked ocular albinism, *GPR143*, identified by Bassi *et al.* in 1995[65]. *GPR143* is known as the gene responsible for OA type 1 (OA1), however further OA subtypes are yet to be defined.

Infantile nystagmus is usually the first clinical sign that leads to investigation of an underlying disorder. Ocular albinism was first recognised as distinct from INS by Vogt in 1925[66]. However, presentations vary widely and initially OA1 can be hard to distinguish from other congenital eye diseases such as achromatopsia, Leber's congenital amaurosis, or idiopathic INS[67, 68]. Iris and fundus hypopigmentation can be particularly difficult to identify in Asian and Caucasian patients, and previous cases of OA1 have been incorrectly diagnosed as idiopathic INS[69, 70]. A study into a large Chinese family with OA1, reported nystagmus as the only consistent finding between all affected family members[71]. Severe foveal hypoplasia has sometimes been described as a prominent clinical feature and can be used to establish a diagnosis for OA1[68]. However, diagnosis of OA1 often requires a comprehensive ophthalmic examination including retinal function and assessment of retinal structure. Molecular screening is also required as other overlapping phenotypes exist such as *PAX6* gene related disorders.

OA1 is inherited in an X-linked trait. Females who inherit a single pathogenic variant will usually be unaffected. However, there are reported cases of supposedly dominant OA which are possibly due to skewed X-inactivation[66]. Skewed X-inactivation of the non-mutant chromosome can result in expression of a disease phenotype that is normally X-linked recessive It has been shown that female carriers exhibit some fundus hypopigmentation and iris transillumination[72]. A recent publication has characterised asymptomatic female carriers exhibiting streaks of hypopigmentation throughout the fundus (*Figure 5*)[68].

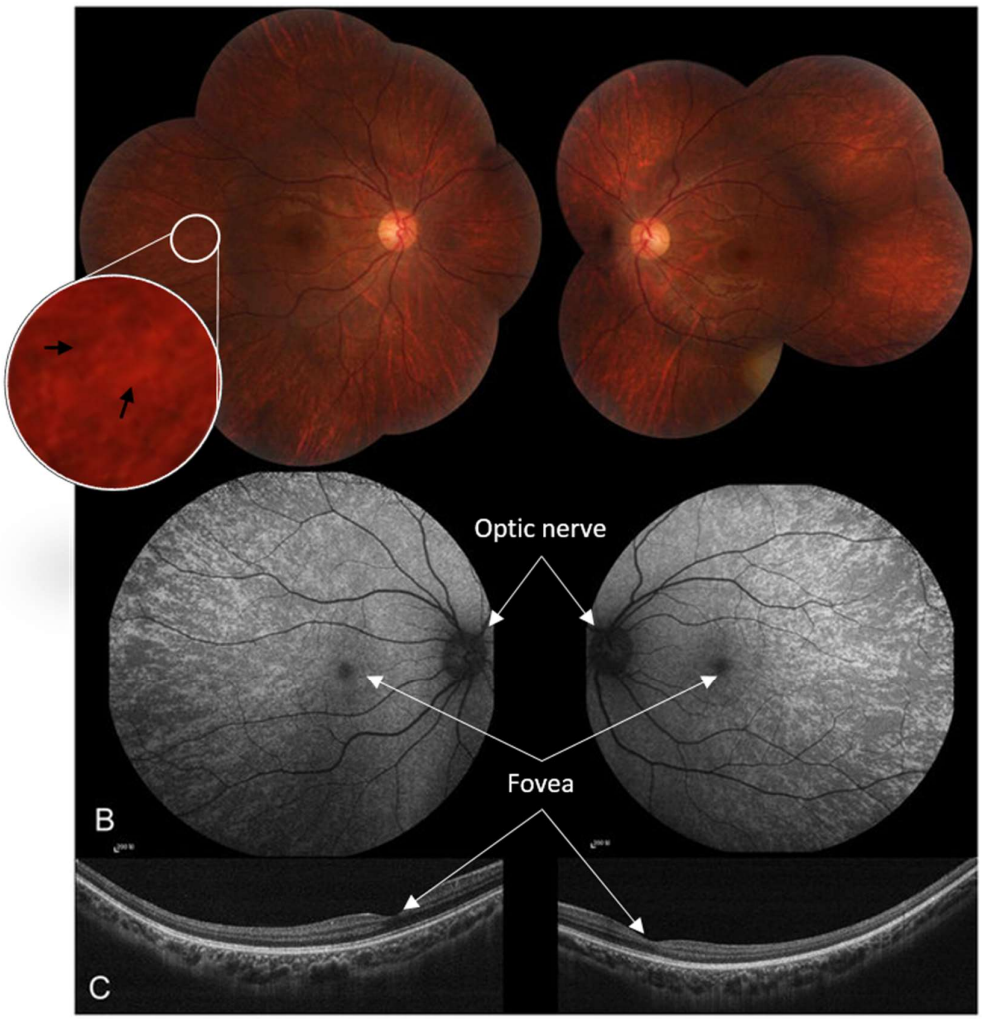


Figure 5. Example fundus images for a female carrier of a *GPR143* (OA1) loss of function mutation. Adapted from Zou *et al.* (2017)[68]. A) Presumed pigmentary mosaicism in the retinal pigment epithelium examined using OCT. Arrows point to pigmented and hypopigmented regions B) and examined with fundus autofluorescence. C) The fovea is normal on OCT.

1.6.3 Autosomal recessive ocular albinism

Autosomal recessive ocular albinism (AROA) was first described in 1978 as a new form of ocular albinism that affects females as severely as males[73]. It was important to recognise that an autosomal form of ocular albinism exists, particularly as it affects the genetic counselling received by families. However, since its first description, most of the AROA cases reported have been associated with mutations in *OCA2* and *TYR*. It can therefore be suggested that AROA is an unnecessary distinction as it represents clinically mild forms OCA2 and OCA1[74],[75]. Evidence is accumulating for the idea that OCA can produce phenotypes covering a whole continuum of pigment, from white hair, skin and translucent irides to near normal pigment with the

characteristic albinoid ocular abnormalities[68]. It has been suggested that the term AROA has been made redundant[76].

1.6.4 Syndromic oculocutaneous albinism

There are several albinism-associated syndromes that can result in multisystem pathology. These include Hermansky-Pudlak Syndrome (HPS), Chèdiak-Higashi Syndrome (CHS), Griscelli Syndrome (GS), Elejalde Syndrome (ES), Cross-McKusick-Breen Syndrome (CMBS) and Waardenburg Syndrome (WS). Each syndrome differs in phenotype and genetic origin but most of the syndromes have a common aetiology of genetic defects causing malformation of secretory vesicles and lysosomes[77].

HPS is a rare autosomal recessive disorder affecting 1 in 500,000 to 1 in 1,000,000 individuals. HPS has variable manifestation resembling tyrosinase-positive OCA in terms of pigmentation and ocular deficits. However, there are other systemic complications such as bleeding diathesis and eventually pulmonary fibrosis, inflammatory bowel disease, and kidney disease due to the lysosomal storage defects[78]. The different syndromic causes of albinism differs greatly according to the phenotype, but often manifest early as a severe multisystem disease (*Table 2*).

Table 2. Syndromic albinism genes and distinct clinical manifestations.

Disease	Subtype	Gene	Common Clinical Manifestations
HPS	HPS1	*HPS1*	OCA, ceroid deposition, occasional pulmonary fibrosis.
	HPS2	*ADTB3A*	Variable, occasional pulmonary fibrosis.
	HPS3	*HPS3*	Mild symptoms.
	HPS4	*HPS4*	Mild symptoms, iris transillumination, variable hair/skin pigmentation, absent platelet dense bodies, occasional pulmonary fibrosis and granulomatous colitis.
	HPS5	*HPS5*	Mild oculocutaneous albinism and easy bruising.
	HPS6	*HPS6*	Oculocutaneous albinism and nosebleeds, no pulmonary or gastrointestinal symptoms, normal platelet count.
	HPS7	*DTNBP1*	Little is known.
CHS		*LYST*	Accelerated phase lympho-proliferation.
GS	GS1/ **ES**	*MYO5A*	Neurological disease, seizures and silver hair.
	GS2	*RAB27A*	Hemophagocytic syndrome.
	GS3	*MLPH*	Hypopigmentation, no immune defects.
CMBS		Unknown	Severe mental retardation with spastic tetraplegia.
WS		*MITF*	Small eyes, hypopigmentation, deafness and neurological defects.

1.6.5 Melanogenesis

Melanogenesis is the multistep process of converting L-tyrosine into either pheomelanin, eumelanin-black or eumelanin-brown via various intermediates. The production of melanin is restricted to melanocytes and the RPE. RPE melanogenesis is thought to differ slightly from that of melanocytes, but the precise mechanisms are still poorly understood. Melanogenesis within melanocytes has been broadly studied and will be briefly summarised here. The synthesis of melanin requires the enzyme tyrosinase, a glycoprotein localised to the membrane of distinct endosomal structures termed melanosomes. Tyrosinase catalyses the critical rate-limiting step of the hydroxylation of L-tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA) and then catalyses the oxidation of L-DOPA to Dopaquinone, (*Figure 6*).

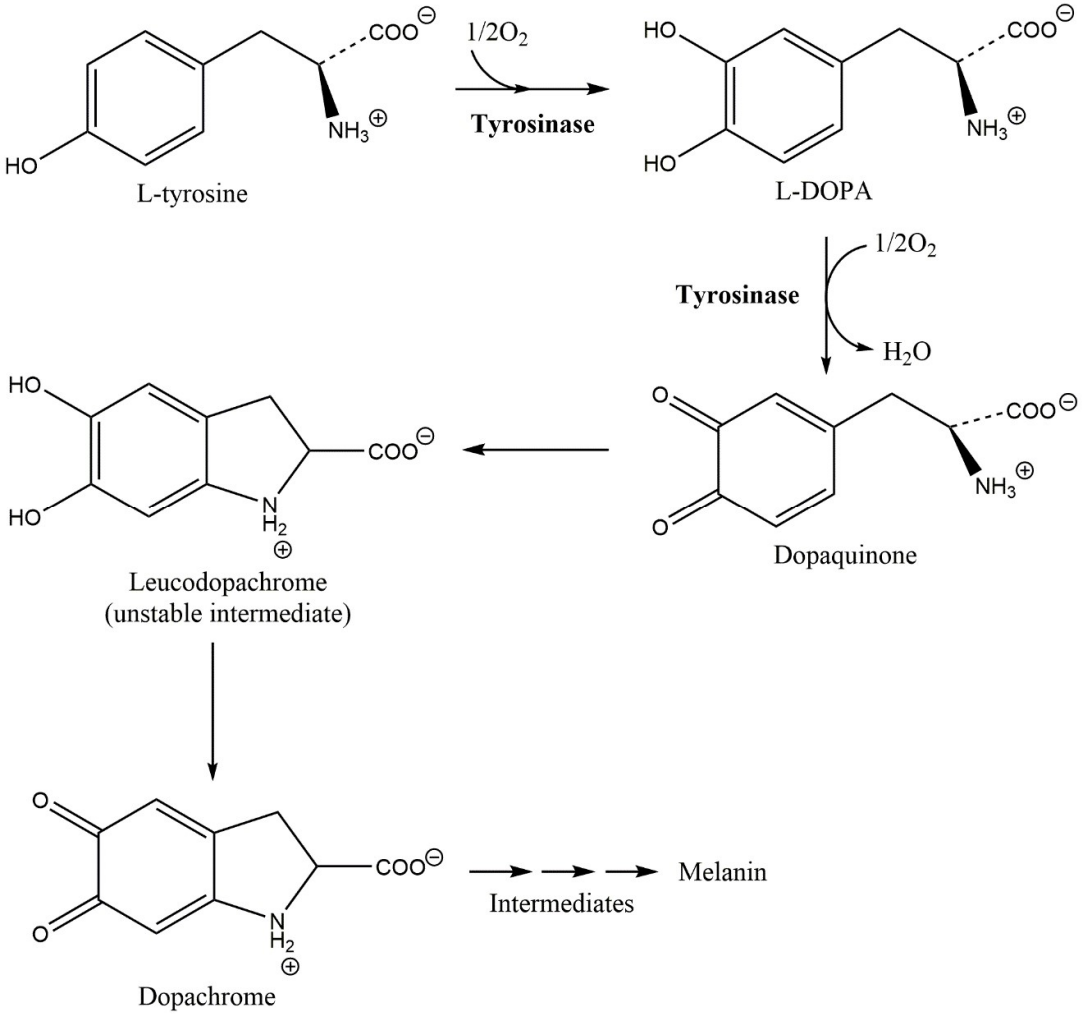


Figure 6. The role of tyrosinase in the early melanogenesis chemical pathway. Tyrosinase catalyses the oxidation of L-tyrosine to L-DOPA, and then catalyses the hydrolysis of L-DOPA to Dopaquinone.

During development, the RPE is the first tissue in the body to become pigmented[79]. However, the role of melanin in the eye in unclear. It has been suggested that melanin can be beneficial by reducing intraocular scatter of light, as well as binding toxins and stabilising free radicals. Nevertheless, blond fundi do not seem to suffer from excess scatter and it has been suggested that melanin binding toxins can create be harmful byproducts[80]. The pigmentary pathway is thought to differ between RPE and melanocytes due to a study showing that the RPE cell line ARPE-19 can produce melanin when transduced with a vector expressing tyrosinase[81]. There was no need for the strict melanosome structure and maturation pathways seen in melanocytes, though the overall production of melanin remained the same[81].

Melanogenesis is stimulated by multiple pathways. These include stem cell-factor (SCF) stimulation of c-kit, α-Melanocyte-stimulating hormone (α-MSH), adrenocorticotropic hormone (ACTH) stimulation of melanocortin 1 receptor (MC1R) and, in melanocytes, α1 and β2

adrenergic receptors can be stimulated be catecholamines produced by keratinocytes from L-DOPA. Furthermore, DNA damage caused by exposure to ultraviolet radiation (UVR) is thought to stimulate melanogenesis as a protection mechanism. The downstream signalling pathways lead to upregulation of the melanogenesis transcription factor MITF and the activation of PKC-β. MITF increases transcription of *TYR* and other melanogenic genes, whilst PKC-β phosphorylates and activates tyrosinase when it is anchored in the melanosome membrane. The role of OCA2 is uncertain and has been suggested to both enable the correct folding and trafficking of tyrosinase to the melanosome, possibly though the regulation of melanosome pH[82, 83]. TYRP1 shows an association with tyrosinase which inhibits the enzymes function but improves folding and trafficking of the protein[84]. The membrane-associated transporter protein (MATP/SLC45A2) has also been shown to influence tyrosinase activity through regulating pH of the melanosome[85]. As the melanosomes mature in the melanocyte they move across the membrane into keratinocytes to increase skin pigmentation (*Figure 7*).

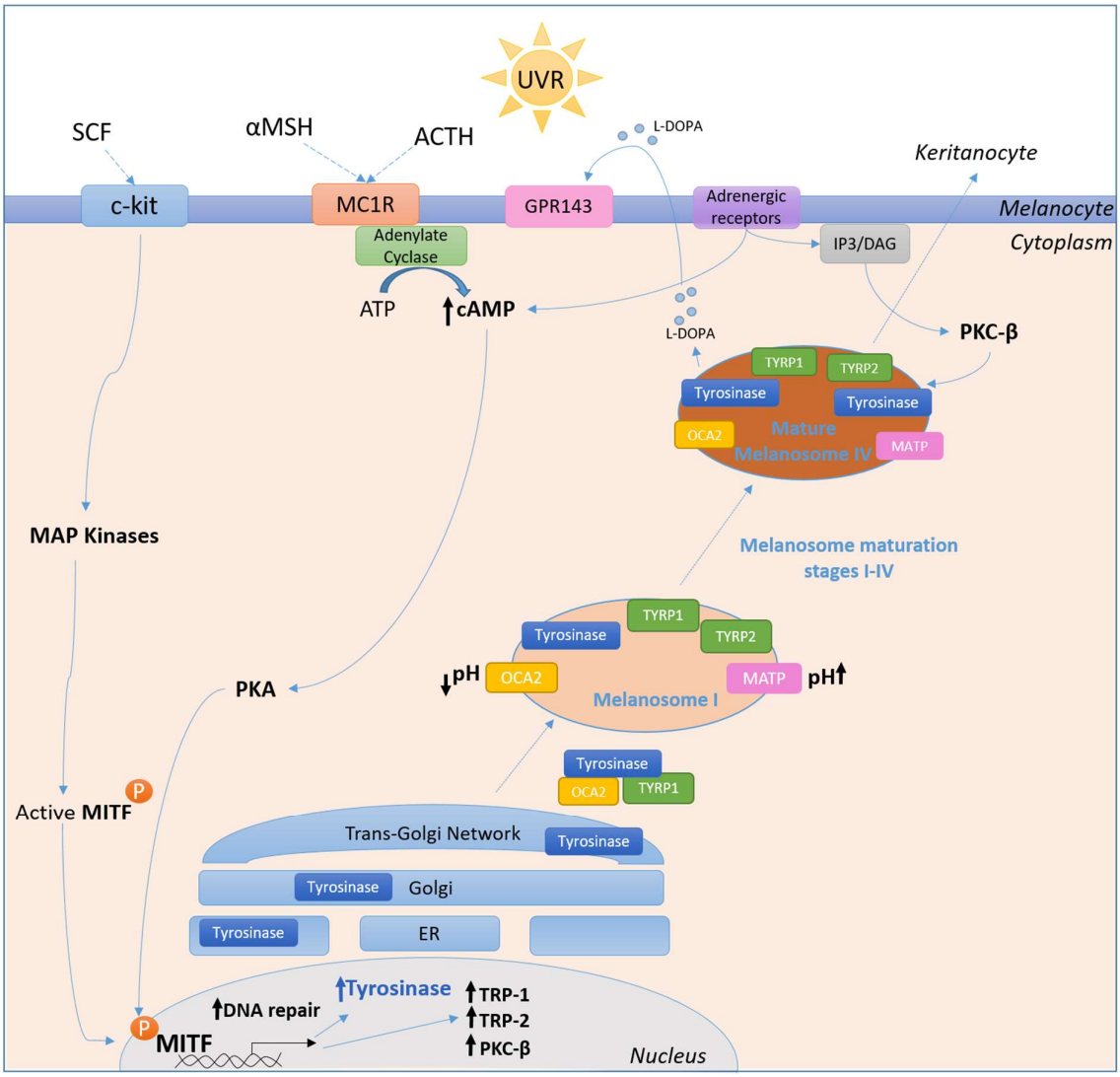


Figure 7. Melanosome formation within melanocytes. The melanocyte production of melanin occurs within specific lysosome-like structures known as melanosomes. The maturation of melanosomes is dependent upon numerous protein components, in which mutations can result in albinism.

Roffler-Tarlov *et al.* wrote about 'L-DOPA and the albino riddle'. They described how a large increase in L-DOPA production in mice 30 days post birth causes conformational changes and suggest that this mechanism is responsible for changes in early infant visual pathways[86]. In 2008, Lopex *et al.* provided evidence that L-DOPA is an endogenous ligand for the GPR143 receptor (mutations in *GPR143* gene are responsible for OA1). A review by McKay discussed the interaction of L-DOPA and the receptor GPR143, suggesting there is an autocrine loop between ligand and the cell surface receptor[87] Mckay also suggested L-DOPA has a role in paracrine signalling between the RPE cells and the neuroretina. A recent paper has shown that overexpression of *GPR143* inhibits neurite outgrowth in pheochromocytoma cells (a neuronal tumour cell line). The effect was mitigated by treatment with the L-DOPA antagonist L-DOPA

cyclohexylester. The findings show that mutant GPR143 protein, as in OA1, is unable to act as an inhibitor potentially causing overgrowth of neurites[88]. However, the results also suggest that a lack of L-DOPA production, as in OCA1, may cause inhibition to neurite growth. This may allude to an explanation for the foveal hypoplasia and chaismal misrouting exhibited by patients with albinism, though the complete mechanism is not clear.

1.7 Diagnosis of infantile nystagmus syndrome

1.7.1 Importance of diagnosis

Diagnosis of infantile nystagmus can be very difficult due to the subtlety of the ocular disorders. Some patients with INS may remain undiagnosed and miss out on helpful management plans. For example, the hypopigmentation in albinism may be understated and missed in early infants, but then become apparent with age and after years of damaging sun exposure. Furthermore, nystagmus can be a sign of neurological damage that requires immediate treatment, or a sign of severe hereditary conditions requiring medical intervention[14].

Most treatments available for nystagmus are used to treat specific aetiologies. For example, the nystagmus resulting from familial episodic ataxia and spinocerebellar ataxia type 6 can be reduced with acetazolamide, but 4-aminopyridine can be more effective in some patients episodic ataxia type 2[89]. Neither of these drugs can be used to help idiopathic INS[89]. Accurate diagnosis of the various underlying causes can enable more effective targeted treatment and may pave the way to revisiting potential therapies and identifying new ones.

In addition, genetic counselling is an important service providing support to patients and relatives at risk of an inherited disorder. In the case of nystagmus it is important to advise families of the nature of the disorder as the consequences of different genetic mutations vary in severity. Genetic counselling also discusses the probability of transmission to offspring and the options available in family planning. When there is a recurrence risk of offspring being affected it is possible to carry out prenatal diagnostic testing. This can be carried out by chorion villus sampling at 10–12 weeks gestation, or by culturing amniocytes post 16 weeks, and then sequencing extracted DNA[51]. Preimplantation genetic diagnosis has been used for multiple monogenic disorders wherein *in vitro* fertilisation (IVF) is used as a means to select embryos without pathogenic mutations. The Human Fertilisation and Embryology Authority (HFEA) has approved OCA1 and OCA2 for testing, but in order to interpret results correctly we must first define which mutations are causal.

1.7.2 Clinical techniques

Nystagmus can be clinically investigated using a number of non-invasive tests. The techniques discussed here are available in the Southampton Eye Unit and are frequently used to phenotype probands in the clinical aspect of this project.

1.7.2.1 Nystagmus eye tracking

Electrographic eye movement recordings are useful in determining the presence of nystagmus in very subtle cases, as well as characterising different aspects such as waveform and fixation. The patient is told to fixate on a point and the horizontal and vertical uni-oculor movements are measured, though vertical and pendular movements can also be recorded. The eye tracker characterises the nystagmus waveform which can be used to guide diagnosis, for example, if the characteristic slow phase is accelerating then IIN is a more likely diagnosis than neurological nystagmus[90]. The Southampton Eye Unit has access to an EYElink10000+ (SR research) eye tracker (*Figure 8*).

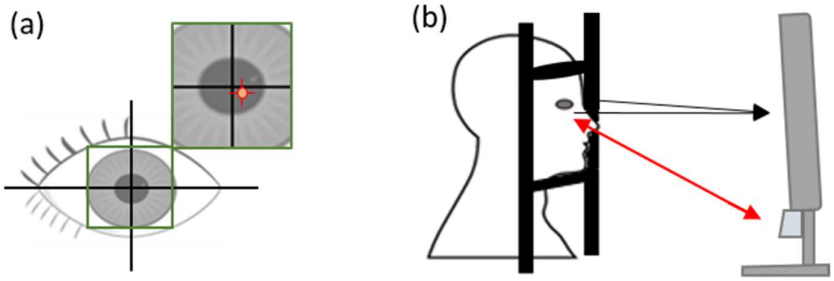


Figure 8. EYElink10000+ eye tracker (SR research). a) An infra-red light is reflected of the eye and detected by a camera (b) mounted below the computer in the light grey box. The head is held stationary by the chin rest (b) and eye movement is tracked.

1.7.2.2 Optokinetic nystagmus

Optokinetic nystagmus (OKN) can be induced in health using an optokinetic drum (*Figure 9*). As the drum is spun it presents a moving pattern of stripes which can test smooth pursuit tracking (slow phases) and automatic saccades (quick phases). Clinician's also look for any asymmetry which may suggest neurological damage[15].

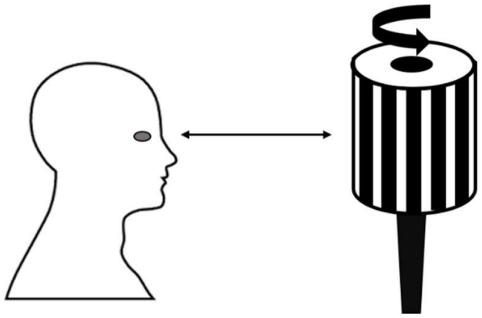


Figure 9. Optokinetic drum used to test for a normal OKR. The drum is spun either vertically or horizontally while the patient is told to focus on the black lines and a clinician or optometrist examines the ocular movements for abnormalities.

1.7.2.3 **Electroretinogram**

An ERG, also known as electroretinography, is a clinical test used to measure the electrical response of photoreceptors to light stimulation. The basic ERG method stimulates the retina with a bright light source such as a LED flash and then measures the response action potential produced (a composite of electrical activity from photoreceptors, Muller cells and Retinal Pigment Epithelium) using electrodes touching the cornea or skin around the eye (*Figure 10*(a)). The potential changes elicited produce a biphasic waveform, composed of two main waves. The a-wave can be used to assess the health photoreceptor outer segments, whereas the b-wave reflects the health of the inner layers of the retina (*Figure 10*(b))[91]. ERG is used to test the function of the macula and is frequently used to identify retinal conditions which may be the underlying cause for nystagmus. Congenital stationary night blindness (CSNB) describes a group of rod dystrophies resulting in night blindness, a reduction of central vision and, infrequently, nystagmus. CSNB cases typically give an ERG reading with a reduced or absent b-wave and a normal a-wave. Leber congenital amaurosis (LCA) is an inherited retinal degenerative disease resulting in low vision from birth. LCA can be detected through very low retinal activity, presented as a very small or non-existent ERG waveform[2].

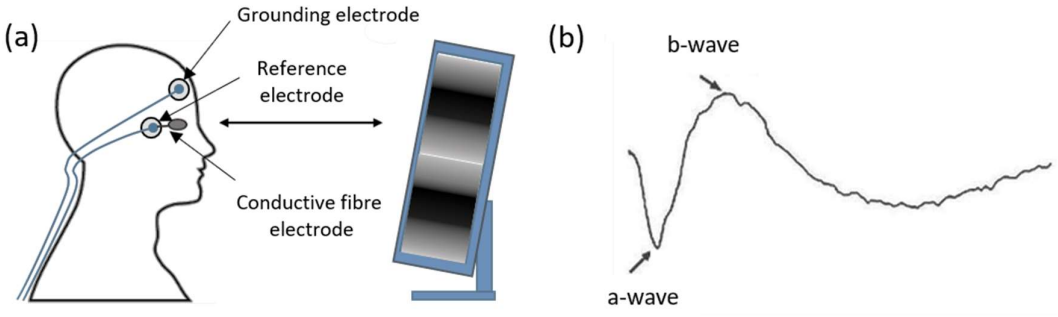


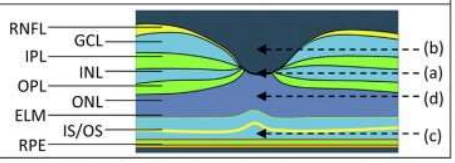
Figure 10. Electroretinography. a) An electroretinogram being performed on a patient with a grounding skin surface electrode in the centre of the forehead, a reference electrode to the right of the eye and a conductive fibre electrode leading to the corner of the eye b) The biphasic ERG waveform seen in a normal patient[91].

1.7.2.4 **Retinal Optical Coherence Tomography (OCT)**

OCT is a high-speed imaging technique providing cross-sectional images of *in vivo* retinal morphology that can be used for both research and diagnostic purposes. These high-resolution images are created from measurements of the echo time delay and magnitude of light reflected/backscattered from the retina[92]. OCT is frequently used to determine the presence of a normal or hypoplastic fovea as foveal hypoplasia is a common feature of albinism and *PAX6* mutation phenotypes. The scan can also reveal gross retinal dystrophies. For example, Stargardt disease is characterised by a loss of the junction between photoreceptor inner and outer segments which can be observed by OCT.

Thomas *et al*. used ultra-high-resolution spectral-domain OCT to document the varying degrees of foveal hypoplasia and determine a 1-4 grading system based on the stage at which foveal development was arrested (*Figure 11*). The grading system has been used to demonstrate that the degree of foveal maldevelopment correlates with decreasing visual acuity[93], which may be the cause of a proportion of sensory-deficit nystagmus.

(A)

Normal foveal structural features detectable using optical coherence tomography	Illustration
(a) Extrusion of plexiform layers (b) Foveal pit (c) OS lengthening (d) ONL widening	

(B)

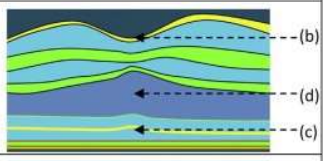
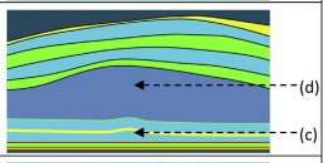
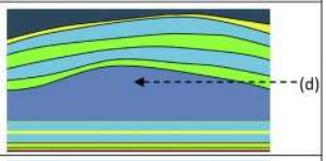
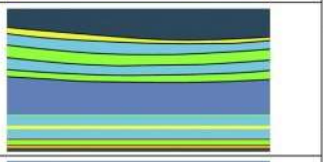
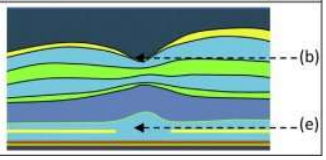
Grade of foveal hypoplasia	Structural features detected on optical coherence tomography	Present or absent	Illustration
1	**(a) Extrusion of plexiform layers** (b) Foveal pit – Shallow (c) OS lengthening (d) ONL widening	**(a) Absent** (b) Present (c) Present (d) Present	
2	**(a) Extrusion of plexiform layers** **(b) Foveal pit** (c) OS lengthening (d) ONL widening	**(a) Absent** **(b) Absent** (c) Present (d) Present	
3	**(a) Extrusion of plexiform layers** **(b) Foveal pit** **(c) OS lengthening** (d) ONL widening	**(a) Absent** **(b) Absent** **(c) Absent** (d) Present	
4	**(a) Extrusion of plexiform layers** **(b) Foveal pit** **(c) OS lengthening** **(d) ONL widening**	**(a) Absent** **(b) Absent** **(c) Absent** **(d) Absent**	
Atypical	**(a) Extrusion of plexiform layers** (b) Foveal pit – Shallow (e) IS/OS disruption	**(a) Absent** (b) Present (e) Present	

Figure 11. Foveal hypoplasia grading figure adapted from Thomas *et al.* (2011). A) Illustration showing features of a normal fovea detectable on OCT B) Illustration of typical grades of foveal hypoplasia. Grade 1 being the closest to a fully formed fovea to grade 4 for which most foveal features are absent. ELM = external limiting membrane; GCL = ganglion cell layer; INL = inner nuclear layer; IPL = inner plexiform layer; OPL = outer plexiform layer; RNFL = retinal nerve fibre layer; RPE = retinal pigment epithelium.

1.7.2.5 **Visual Evoked Potential (VEP)**

Asymmetry of visual evoked potential (VEP) responses suggests excessive decussation at the optic chiasm, a feature that is commonly used to distinguish albinism from other disorders with similar phenotypes[94-96]. However, misrouting of nerves at the optic chiasm has also been shown to be present with some *PAX6* mutations[97].

VEP recordings are taken for each eye in response to full-field pattern-onset stimulation (light stimulation generally using a checkboard pattern). Electrical signals are measured through

electrodes placed on the scalp across the occipital cortex (at the back of the head). The readings are plotted as amplitude over time and the measurements for each eye compared (*Figure 12*).

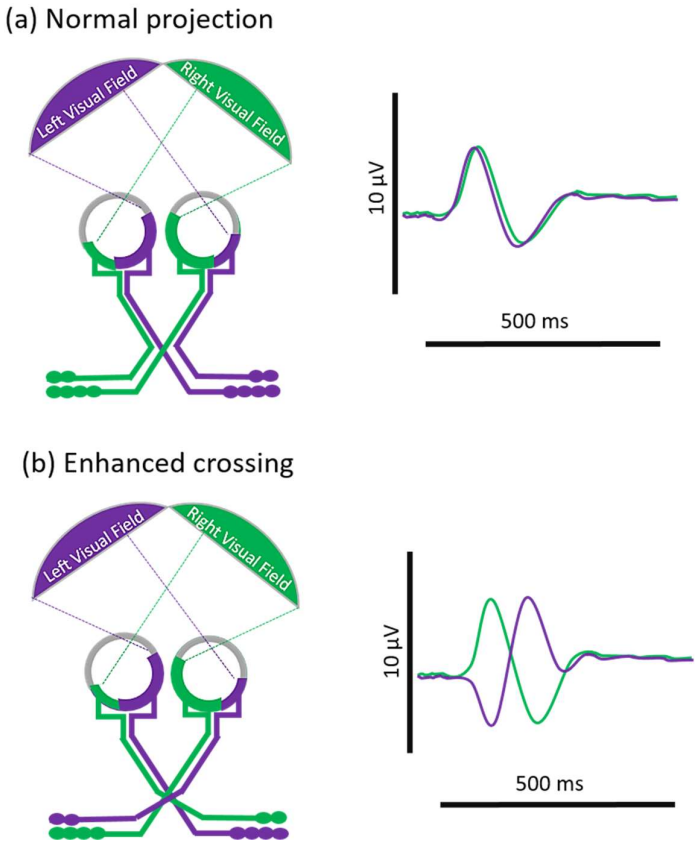


Figure 12. Enhanced crossing of nerves at the optic chiasm with representative VEP recordings. A) Normal projection on the right with representative VEP readings on the left B) Enhanced crossing associated with albinism on the right with representative VEP readings on the left adapted from Hoffman *et al.* 2015[98].

1.7.2.6 **Slit lamp biomicroscopy**

Slit lamp biomicroscopy uses a high-intensity light source focused to shine a thin sheet of light into the eye and then both the posterior and anterior segments of the eye can be viewed through a biomicroscope (*Figure 13*). Slit lamp biomicroscopy provides a magnified detailed view of the eye structures enabling identification of any abnormalities in the eyelid, sclera, conjunctiva, iris, lens, or cornea, enabling diagnoses for a variety of eye conditions. In the case of nystagmus, the slit lamp exam is specifically important for determining the presence of iris transillumination. Iris transillumination can be difficult to determine when a patient has blue eyes, but it is considered present if the light from the microscope can be seen reflecting back out through the hypopigmented iris. Iris transillumination is often a sign of albinism or *PAX6* disease phenotypes[51].

Figure 13. Slit lamp biomicroscope. The patient's head is held steady by the chin rest so the eyes can be examined by the clinician. Adjustments can be made to magnification and light intensity.

1.7.3 Current genetic testing

Genetic testing is often considered beneficial from a patient point of view, providing a sense of relief from uncertainty and enabling patients to make informed decisions about treatment and the management of their care. Genetic results can determine the likelihood of passing on a disorder to offspring and help individuals make decisions when preparing to have children. Research genetic testing can also benefit patients through finding new genes, linking genes and mutations to specific medical conditions, thus determining the effect on individuals and feeding back into clinical diagnoses and management.

At the beginning of this project, there were a few UKGTN single gene sequencing and gene panels available for some of the genes associated with nystagmus including, retinal degeneration (£897), the X-linked idiopathic INS gene *FRMD7* (£400), the *PAX6* gene (£605), and OCA1 and OCA3 genes within a large developmental disorder panel (£1300). In some cases, a patient would need to have multiple genetic tests due to a mixed phenotype, taking over 100 days per test which was not always economically viable for the NHS. Consequently, the vast majority of patients did not undergo genetic testing of any sort.

1.8 Treatment and management

As already described, INS can be caused by numerous disorders, each of which may require specific treatment. There are few treatments for the nystagmus itself but no cure. If a refractive error is detected, optical correction can be attempted with glasses or even contact lenses as these move with the eye movement. Surgery known as four-muscle tenotomy is used to move the null point to less eccentric position, correcting any head turn that has developed. The procedure can result in improved visual acuity, broadening of the null zone, and improved target acquisition[99]. There are several drug treatments with variable levels of success across the various forms of nystagmus. There have been studies suggesting that some forms of nystagmus, such as the pendular nystagmus that can occur in neurodegenerative disease, may be a result of instability in the gaze-holding mechanism. The neurotransmitters important in gaze-holding are GABA and glutamate which have been targeted by gabapentin, baclofen (GABA target)[100] and memantine (glutamate target)[101]. Gabapentin and memantine both had positive effects when tested in cases of acquired nystagmus due to multiple sclerosis[100, 101]. However, both drugs have shown little effect in cases with idiopathic INS, possibly due to an alternative mechanism[102]. Familial episodic vertigo and ataxia type 2 are both inherited disorders of calcium channels that can result in nystagmus. These disorders can be treated with calcium channel blockers and acetazolamide[103]. There has been research into the use of gene therapy to treat OCA1 and HPS1. So far, these disorders have been corrected in animal models using systems such as CRISPR/Cas9 and lentiviral vectors[104],[105, 106]. However, there is yet to be any considerable development towards clinical use. Improving knowledge of the mechanisms behind nystagmus will allow investigation of the pharmacology, leading to further drug treatments.

1.9 Molecular diagnosis through genetic sequencing

1.9.1 Next generation sequencing

The rise of high-throughput sequencing has completely transformed the field of genomics into an accessible and cost-effective range of techniques that can rapidly provide a wealth of information. The cost of whole-genome sequencing was immediately reduced by 50,000-fold on the introduction of the first high-throughput sequencing platform[107]. The ease with which we can now sequence the human genome means that genomics has the potential to revolutionise patient diagnosis and provide tailored treatments. Having said this, only a small proportion of the human genome is understood and only then to a relatively small degree. There will always

be a place for conventional clinical judgements and diagnosis, but the combination of genetic and clinical data gives the best possibility of diagnosis and patient-tailored management.

In 2005 the dbSNP database contained over 9 million human SNPs, by 2008 the number was 12.8 million and still rising[108]. The large amount of variation in the human genome combined with the mass of data available leaves a problem for determining which variants are pathogenic and which are benign. Furthermore, misclassification of variants can lead to an inflated number of rare variants being reported as pathogenic[109]. However, a shift towards correcting classification of variants has been observed as there are increasing contributions to databases such as ClinVar[109].

Data intensive, high-throughput sequencing techniques such as whole genome, whole exome, or targeted panels are frequently used to investigate the genetic cause of undiagnosed, suspected genetic conditions. As the cost of sequencing drops and the demand for personalised medicine increases and these techniques become more appealing. However, there is a gap between acquiring the data and interpreting it for clinical diagnosis. There needs to be a clear method of data interpretation in order to provide a useful output for clinicians and patients.

The correct assignment of pathogenicity to sequence variants is an important and difficult hurdle that needs to be tackled. Once the sequencing reads have been aligned to the reference genome thousands of variants are identified which need to be assessed and filtered following a structured workflow. Initially, the filtering of variants is generally based on the gene(s) of interest, type of variant, population frequencies, and computational pathogenicity scores. The variants passing the filtering thresholds can then be further investigated using data from previous literature reports which may provide segregation and functional evidence of pathogenicity. If available, additional family members can be sequenced to see whether the identified variant segregates with the disease phenotype. The American College of Medical Genetics and Genomics (ACMG) standards and guidelines can guide how much weight is given to each piece of evidence. Functional studies provide the most convincing evidence for variants of questionable pathogenicity. However, set up of a functional study may not be possible and also may require a great deal of time and resources.

1.9.2 ACMG standards and guidelines

In 2013 the ACMG and the Association for Molecular Pathology revised standards and guidelines used for interpreting sequence variants. They produced a document that advises specific terminology for the description of variants found in Mendelian disorders. The ACMG standards and guidelines describe a detailed classification system based on different criteria

such as population data, computational predictive data, segregation data, and functional studies. The strength of each piece of evidence for a variants pathogenicity is taken into consideration. The overall recommendation of the ACMG guidelines is the use of a specific five-tier terminology system with the terms 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign'[110]. Here, the different types of evidence for pathogenicity are discussed in more detail, and the criteria for pathogenicity classification has been summarised (*Table 3*).

Table 3. A brief summary of the ACMG 2015 standards and guidelines, recommended terminology, and example evidence criteria for description of variants in Mendelian disorders. There are a total of 28 criterion falling into two sets: pathogenic/likely pathogenic (P/LP) and benign/likely benign (B/LB). Due to the extent and complexity of the guidelines only example criteria are given here.

Classification of variant	**Pathogenic/Likely Pathogenic (P/PL)**	**Benign/Likely Benign (B/LB)**
Very strong criteria (PVS1) or stand-alone (BA1)	• A null variant (nonsense, frameshift, canonical splice site, initiation codon, exon deletion).	• Allele frequency >5% in population data e.g. 1000 Genomes Project.
Strong criteria (PS1-4 or BS1-4)	• Same amino acid as a previously established LOF variant. • Functional studies show damaging effect. • Prevalence of variant in affected cohort is greater than in controls.	• Allele frequency is greater than expected for the disorder. • Functional studies show no damaging effect. • Does no segregate with affected family members. • Observed in a healthy adult.
Moderate Criteria (PM1-6)	• Novel missense change in the same location as a previously confirmed pathogenic mutation. • For a recessive disorder, there is another pathogenic variant in *trans*. • Absent in population databases. • Within a well-studied functional domain. • *De novo* variant (unconfirmed maternally and paternally)	-
Supporting criteria (PP1-5 or BP1-7)	• Segregation with multiple affected family members (strength of evidence increases with number of family members). • Multiple types of computational evidence. • Missense variant in gene with low rate of benign missense variation.	• Missense variant when only truncating variants are known to cause disease. • In a dominant disorder, there is a known pathogenic variant in *trans*. • In-frame indel in a repetitive region with no known function. • Multiple types of computational evidence.

1.9.2.1 **Population databases**

There are a number a databases containing ever growing population variant data from NGS projects, such as Exome Variant Server (http://evs.gs.washington.edu/EVS/) and the 1000 Genomes Project (accessed through ensembl: http://grch37.ensembl.org/index.html). Population variant data can also be submitted to databases, such as dbSNP (https://www.ncbi.nlm.nih.gov/SNP/) and dbVar (https://www.ncbi.nlm.nih.gov/dbvar), which combine data from smaller studies. The database dbSNP contains short genetic variations (typically <50 bp) which can be submitted from small studies as well as national/international sequencing projects. However, the details of the original studies may be lacking and variants may be pathogenic rather than typical SNPs. Population databases can be used to determine the frequencies of variants in large populations. It is important to note that these databases likely contain data for both healthy and diseased individuals, and the population data will be skewed for different ethnicities. On the other hand, disease databases mostly contain variants found in patients with disease though some variants may be incorrectly classified.

1.9.2.2 **Previously reported mutations**

Previously reported mutations can sometimes be found in disorder-specific databases. These tend to be established by smaller organisations e.g. The Albinism Database (http://www.ifpcs.org/albinism/) which was set up by the University of Minnesota and edited by William Oetting. This database was last updated in September 2009, so it can be a helpful resource but is out of date. Some disease causing variants can be found in specific disease databases such as ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), and the Human Gene Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/ac/). These databases mostly report variants that have been annotated as disease causing in the literature with links to the original research so the evidence for pathogenicity can be assessed by the viewer.

1.9.2.3 **Computational predictive programs**

The ACMG standards and guidelines recommend the use of multiple software programs when interpreting sequence data due to the varying intrinsic strengths and weaknesses. However, they also state that a combination of multiple predictive tools can still only be considered a single piece of evidence and is not enough to make a clinically relevant assertions[110]. The Southampton pipeline currently uses the software tool ANNOVAR to functionally annotate genetic variants[111]. The output annotates variants for their frequencies in population databases as well as pathogenicity scores when run through numerous predictive

programs. The pathogenicity scores include SIFT, PolyPhen, GERP, CADDPhred and MaxEntScan, each of which uses a different algorithm to predict a variant's pathogenicity.

Sorting Intolerant From Tolerant (SIFT) is an algorithm that aims to predict the pathogenicity of nonsynonymous variants using homology data to decide whether evolutionary conservation would tolerate the substitution. An amino acid substitution is predicted to be damaging to the protein if the SIFT score is ≤0.05 and predicted to be tolerated if the score is >0.05 (http://sift.jcvi.org/www/SIFT_enst_submit.html)[112].

Similar to SIFT, Genomic Evolutionary Rate Profiling (GERP) predicts conservation but at the nucleotide level. Nucleotide conservation is quantified with a 'rejected substitutions' score, defined as the number of substitutions to be expected neutrally minus the number of substitutions 'observed' at the position. The GERP tool was first created by Cooper *et al.* in 2005, however, a more rigorous GERP++ was implemented by Davydov *et al*. in 2010 using algorithms to calculate site-specific RS scores and to discover evolutionarily constrained elements[113]. Positive scores represent fewer substitutions than the average neutral site and indicate that a site may be under evolutionary constraint, whereas negative scores indicate that a site is probably evolving neutrally. High scores (max 6.18) suggest evolutionary constraint.

PolyPhen2 is a commonly used predictive tools (alongside SIFT). PolyPhen2 uses protein structure/function and evolutionary conservation data to predict the pathogenicity of missense mutations. PolyPhen2 Human Variant (HumVar) describes variants using the terms: benign, possibly damaging, and probably damaging (http://genetics.bwh.harvard.edu/pph2/)[114].

Combined Annotation-Dependent Depletion (CADD) is a tool for scoring the deleteriousness of single nucleotide variants and indels in the human genome (https://cadd.gs.washington.edu/). It differs to the other scores described as it aims to integrate the sometimes conflicting results of many different tools into one comprehensive annotation, thus removing algorithm specific bias[115]. The creators of CADD, Kircher *et al.,* later improved the scores interpretability by ranking each score in against all potential single nucleotide variants in the human genome. The improved score was termed CADDPhred with a higher score pertaining to a greater predicted deleteriousness[116].

Splicing variants are notoriously difficult to predict when they are not directly within the 5' acceptor or 3' donor splicing sites. Annovar annotates variants as 'splicing' if they are within 2 base pairs on the intronic side of the intron/exon boundary, though predictive tools do exist and can be used in addition. MaxEntScan is the name for a tool based on 'maximum entropy modelling' of short sequence motifs involved in splicing. The tool scores variants in 5' splice sites

or in any associated sequence motifs for their potential disruption to normal splicing. Similar to CADDPhred, MaxEntScan amalgamates previous methods of analysis to come up with a single comprehensive predictive score[117].

1.9.2.4 **Functional studies**

Functional studies are often considered the most powerful tool when providing evidence for the pathogenicity of a variant. When assessing proteins the differences between the assay conditions and the biological environment mean they may not accurately predict the impact on a protein's physiological function. Assaying a protein expressed *in vitro* is not considered as valid as protein taken directly from biopsied tissue. Likewise, an assay examining a single component of enzyme function is less valid than an assaying biological function in its entirety. However, well established assays with reproducible and robust findings can provide trustworthy evidence for loss of protein function[110]. Chapter 5 describes the future functional studies that will be used in this project to assess variants in OCA1. Besides protein, there are assays that assess the impact of variants on messenger RNA. The levels, stability, and processing of mRNA can be assessed[110].

1.9.2.5 **Segregation analysis**

Segregation analysis within a family can be used as evidence for linkage of the locus to the disorder rather than for the variant pathogenicity. For heterozygous disorders, segregation can be used to determine whether two variants are coinherited in *trans* by determining separation on parent alleles. The ACMG guideline consider segregation data only supporting to moderate evidence dependent upon family size. However, segregation can be combined with other evidence of pathogenicity to make a compelling case.

Chapter 2: Materials and Methods

2.1 Patient recruitment

2.1.1 Ethics

Patients were recruited following the tenets of the declaration of Helsinki. Informed consent was obtained and the research was approved by the Southampton and South West Hampshire Research Ethics Committee (LREC number: 028/O4/t). This work also contributes to a study into albinism in Pakistani populations (appendix B.2). Work carried out by Arshad *et al.* was approved by the ethical approval committee institutional review board of the International Islamic University, Islamabad, Pakistan[118]. The method of sequencing used and number of probands and family members sequenced in each chapter is detailed in *Table 4*.

Table 4. Number of samples and method of sequencing per results chapter. *This sequencing was performed prior to this study.

Results Chapter	Sequencing method	Number of samples
3 **(total = 59)**	WES*	2
	HaloPlex Custom Panel*	3
	TruSight One Panel	29
	Sanger Sequencing Only	25
4 (N/A)	N/A	N/A
5 **(total = 83)**	TruSight One Panel	81
	Sanger Sequencing Only	2

2.1.2 Phenotyping

The majority of probands were recruited from a regional paediatric nystagmus clinic located in Southampton General Hospital Eye Unit. In Southampton, clinical workup and saliva sample collection were carried out by Mr Jay Self and his orthoptic team. All Southampton patients underwent detailed phenotyping comprising of, orthoptic examination, electrodiagnostic tests, OCT of the macular, and eye movement recording. Furthermore, probands were phenotyped for skin and hair tone in context of family pigmentation. Abnormal

ERG results excluded a patient from the study as the NHS-available UKGTN retinal degeneration panel would be used for genetic diagnosis instead. Phenotyping was carried out for as many relatives as were available and willing to participate at the time of clinic appointment.

Direct ophthalmoscopy, using a slit-lamp biomicroscope, was used to examine anterior and posterior segments for structural abnormalities. Electrodiagnostic tests included an ERG and to assess retinal function and VEPs to detect the visual pathway from the retina to the visual cortex. The Great Ormond Street (GOSH) paediatric protocol was used in children (up to the age of 18)[119], and the International Society for Clinical Electrophysiology of Vision (ISCEV) standard protocols were used for adults[120, 121]. At birth, the retina and neural visual pathways are functional but are rapidly changing, particularly in the first 5 months of life. Therefore, noting the age at which recordings were taken was important for correct interpretation of the phenotype. Cross sectional OCT images of the macular were taken to examine retinal layers. In young/uncooperative children a hand-held Leica OCT system was used. However, a Spectralis OCT was used in adults as it provided the highest resolution (Heidelberg Engineering). Eye movement recordings were made on an EYElink10000+ (SR research) eye tracker to determine the presence and waveform of nystagmus.

For probands and family members recruited in Pakistan, a detailed history was taken before samples were genotyped. Those identified as carrying pathogenic mutations in *TYR* and *OCA2* were revisited for phenotyping following confirmation of pathogenic variants. Skin pigmentation and presence of nystagmus was recorded, along with visual acuity (using Snellen chart) and findings from direct ophthalmoscopy of the fundus.

2.1.3 DNA collection

Saliva samples were collected from probands and family members using Oragene-DNA kit: OG-575 for infants and OG-500 for adults (DNA Genotek, Canada). The OG-575 contains sponges for collecting saliva from young children who require assistance. Mostly, the collection of DNA occurred in the eye unit after recruitment of families to the Nystagmus study. However, on some occasions, the collection kits were posted with instructions and consent forms. For Pakistani samples, DNA was extracted from blood by colleagues in Exeter.

2.1.4 Data storage

Patient data was pseudonymised with an 'NG' ID number, e.g. NG123, and a corresponding family ID e.g. #123. Data was originally entered into two different databases. The first, a password protected excel spreadsheet containing detailed information on proband's

phenotype and genotyping status, and the second, a password protected spreadsheet with simplified phenotyping information about proband and family members. These databases were combined to streamline data entry and the terms of phenotyping were refined into a few selectable options to avoid unclear language and allow for an accurate bioinformatics analysis (*Figure 14*).

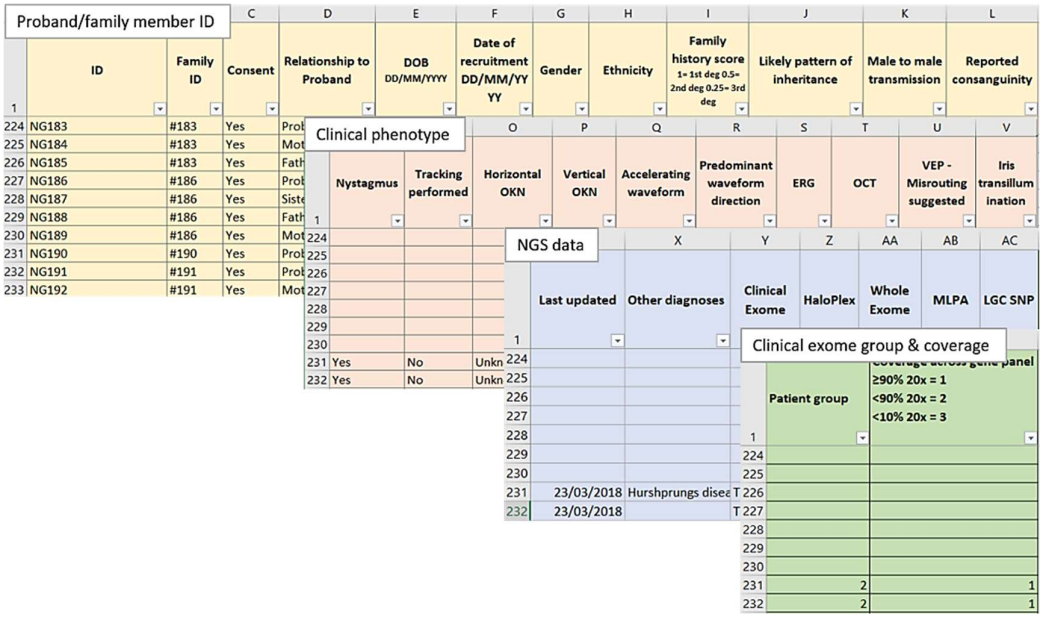


Figure 14. Database containing the entire cohort of probands and family members. Proband/ family member ID and clinical phenotyping information is being continually updated by Research Orthoptists in the eye unit. The databased was also updated with NGS data as samples were run on the clinical exome and quality control results became available.

2.2 DNA extraction

The extraction of DNA from saliva samples was carried out by me or other members of the research team. DNA was extracted using the prepIT•L2P (DNA Genotek) kit, following the protocol for manual purification of genomic DNA.

Briefly, saliva samples were incubated and agitated at 50°C in a shaking incubator for 1-2 hours. Samples were transferred to 15 ml centrifuge tubes and, for the purpose of improved yield, any undissolved precipate were removed with a sterile pipette. The prepIT L2P purifier was added to the tube at a volume of (estimated saliva volume in µl)/25 and the tube was placed on ice and gently shaken for 10 minutes. The sample was then centrifuged at room temperature for 20 minutes at 330 *xg* to pellet cell membrane and unwanted material. After centrifugation, the supernatant was transferred into a new 15 ml tube and the pellet discarded. An equal volume of 100% ethanol was added to the sample and the tube inverted 10 times before being left for 10 minutes. The centrifugation step was repeated in order to pellet genomic DNA. The

supernatant was gently removed afterwards and discarded, followed by addition of 1 ml of 70% ethanol wash step. After 1 minute, the ethanol was removed and the tube was left to air dry. Depending on the size of the pellet, 50 µl-200 µl of Tris-EDTA buffer was added to rehydrate the DNA. The tubes were lightly vortexed and incubated for two days at room temperature or for 1 hour at 50°C.

2.2.1 DNA quality assessment

A NanoDrop ND-1000 (Thermo Fisher Scientific, UK) spectrophotometer was used to measure the concentration and purity of the extracted DNA samples. The ratio of absorbance at wavelengths at 260 nm and 280 nm was used to assess purity with a value of 1.8 or above being accepted as 'pure'. As the DNA samples were extracted from saliva, the 260/280 ratio was often appreciably lower, indicating contaminants such as proteins or phenols with a spectral absorbance in proximity to 280 nm. This sometimes negatively affected PCR reactions so a genomic clean and concentration kit was used to improve the purity of DNA (section 2.2.2).

A Qubit® 2.0 fluorometer was used with dsDNA broad range assay kit (Thermo Fisher Scientific, UK) for quantitation of DNA concentration. The Qubit reading was considered more accurate than the measurement by NanoDrop as the reagent used in the Qubit contains a fluorescent dye which directly binds DNA, thus removing potentially misleading background signals due to aforementioned contaminants.

2.2.2 Cleaning and concentrating DNA

DNA samples of poor quality were cleaned using a Genomic DNA Clean & Concentrator™ kit (Zymo Research, CA, USA). The ChiP DNA binding buffer was added at a ratio of 2:1 to the DNA samples and the mix was transferred to a spin column and then centrifuged at 13,000 x*g* for 30 seconds to bind the DNA to the column matrix. Columns were washed with 200 µl of DNA wash buffer and centrifuged for 1 minute. This step was repeated and the column was transferred to a sterile 1.5 ml microcentrifuge tube. Following this step, >10 µl of elution buffer was added to the column matrix, incubated at room temperature for 1 minute and then centrifuged 13,000 x*g* for 30 seconds. The concentration and purity of eluted DNA was measured using a NanoDrop spectrophotometer ND-1000.

2.2.3 Whole genome amplification

Whole genome amplification (WGA) is a method for the amplification of an entire genome, starting with very small (nanogram) quantities of gDNA and making microgram

quantities of the amplified DNA product. WGA was used to amplify samples of very low volume. We typically used this approach to obtain large DNA quantities from samples obtained from young children when there was minimal DNA after extraction from a small saliva sample. When using WGA it is important to consider any possible bias that may be introduced. Particularly, copy number bias which would be important when assessing factors such as SNP frequencies between cells and qPCR technologies [122]. WGA should not alter the sequence itself and therefore is unlikely to cause a problem for Sanger sequencing.

A REPLI-g Mini Kit (Qiagen, Germany) was used to achieve WGA. First, the buffers D1 (9 µl of reconstituted buffer DLB with 32 µl of nuclease-free water) and buffer N1 (12 µl stop solution and 68 µl nuclease-free water) were prepared. To denature DNA, buffer D1 was added to 2.5 µl of template DNA (roughly 200 ng), vortexed and incubated for 3 minutes, then 5 µl of buffer N1 was added to stop denaturation. A PCR master mix was prepared on ice by combining 10 µl of nuclease free water with 29 µl REPLI-g Mini reaction buffer and 1 µl REPLI-g Mini DNA polymerase. The denatured DNA was added to the PCR master mix and incubated at 30˚C for 16 hours, followed by heating at 65˚C for 3 minutes to inactivate the polymerase. The concentration of the PCR product was checked using the Qubit dsDNA broad range protocol as described in section 2.2.1.

2.3 Next generation DNA sequencing

Next generation sequencing (NGS) is the name given to high-throughput sequencing technologies. The technology used in this project is predominantly illumina DNA sequencing which involves the clonal amplification or short DNA strands to create clusters, followed by the nucleotide pairing of reversible dye-terminators. The clusters of the same sequence are dense enough that a camera can detect the signal emitted by the reversible dye-terminators. A vest number of short reads (100-150 bps) are sequenced at the same time and can be pieced together to build up a longer sequence.

Advances in NGS have enabled the use of sequencing as a clinical tool[107]. Whole genome sequencing provides us with a vast amount of data, but it is currently difficult to interpret any variants that lie outside the protein-coding regions of the genome. Diagnostic and research laboratories, predominantly search for coding variants, making cheaper targeted sequencing panels more viable options. Here, a clinical exome was carried out using a commercially available targeted sequencing panel (TruSight One v1.0, Illumina) to provide exonic sequencing data on 4813 genes associated with human disease.

2.3.1 TruSight One library preparation

A library of DNA fragments from all coding exons of disease-causing 4813 genes, as determined by the Online Mendelian Inheritance in Man (OMIM) database (www.omim.org), were prepared with a TruSight One (TSO) target enrichment kit v1.0 (Illumina, SD, USA), following the manufacturer's instructions. Briefly, 50 ng of genomic DNA from each sample, diluted with nuclease-free water to a final concentration of 5 ng/µl was fragmented and tagged using a single transposome step. Fragmented gDNA tagged with adapter sequences was cleaned with purification beads and then amplified in a PCR step adding unique index adapters to each sample. The DNA was purified with beads and then quantified using the Qubit broad range fluorometric method, as described in section 2.2.2., to calculate concentrations for equimolar pooling. Samples were pooled into three groups of 12, containing 500 ng of DNA from each sample. The exonic regions of the genes of interest were targeted in the pooled libraries by the hybridisation of biotinylated probes. The target regions were captured using streptavidin magnetic beads. The hybridisation and capture was repeated and cleaned with wash buffer before a second PCR step to amplify the captured library. Finally, the enriched library was purified and quantified using two fluorometric methods; a Qubit 3.0 with a high sensitivity assay kit (same protocol as the Qubit 2.0 broad range in section 2.2.2.) and an Agilent 2100 Bioanalyser (Agilent, CA, USA). The three libraries were equimolar pooled then requantified using the same flourometric methods. NanoMolar concentration was calculated using the formula:

$$\frac{\text{concentration ng/µl}}{(660\ \text{g mol}^{-1} \times \text{mean library size}^{*})} \times 10^{6}$$

*Average library size is calculated using the Bioanalyser readings to be 500 base pairs. This value was used for all Clinical Exome preparations as the Bioanalyser step was deemed costly and unnecessary.

Pooled libraries were diluted to 2 nM with resuspension buffer and then normalized for loading into the sequencer.

2.3.2 HaloPlex custom panel preparation – previous work

The Haloplex target enrichment panel (Agilent Technologies, Santa Clara, CA, USA) was custom designed to capture the exomes of 55 genes associated with nystagmus, other related ocular disorders, oculocutaneous albinism, and population pigmentation genes according to OMIM (Appendix table A.1). The HaloPlex panel work was carried out by other members of the research team in 2014. There were 33 samples with varying ocular phenotypes including, albinism and CIN, were sequenced. Briefly, 225 ng at 5 ng/µl of each genomic DNA sample was digested in eight different reactions, each containing 2 restriction enzymes. A library of

biotinylated HaloPlex probes was hybridized to the newly created DNA restriction fragments and unique index sequences were incorporated. The target fragments were bound to magnetic streptavidin beads and purified. DNA ligase was added to the bead-bound samples to close nicks in the circularized HaloPlex probe-target DNA hybrids and 50 mM NaOH was then used to elute the captured DNA libraries. PCR amplification was carried out on 20 µl of supernatant from each tube. The target libraries were purified, and enrichment was validated by gel electrophoresis before loading the enriched library onto the NextSeq 550 sequencer.

2.3.3 Sequencing using the Illumina NextSeq 550 sequencer

The prepared 2 nM TSO libraries were denatured and diluted for sequencing on the Illumina NextSeq system. First, 0.2 M NaOH in water was added to denature the DNA fragments, after 5 minutes 200 mM Tris-HCL was added to stop the process. Then, 970 µl of chilled buffer HT1 was added to the library and vortexed, before further diluting the library to a final concentration 1.8 pM (117 µl library solution in 1183 µl HT1).

The NextSeq sequencer interface was synced with an online BaseSpace (Illumina) account containing information prepared for the run. Then a new flow cell was loaded in the sequencer, followed by a waste cartridge and buffer cartridge containing the library loaded in the allocated well. The NextSeq run was initiated and the read data was automatically transferred across to a secure folder location to be accessed for bioinformatics processing. The NextSeq 550 sequencer provides run metrics available both on the machine and on BaseSpace. Metrics such as cluster density and Q score can indicate the quality of the run and sequencing data achieved.

2.3.4 Whole exome – previous work

The agilent SureSelect v5.0 was used to capture the whole-exome for a total of 8 samples, 6 of which were nystagmus cases. Whole-exome sequencing was performed on the Illumina HiSeq 2000 at Wellcome Trust Centre for Human Genetics, Oxford. This work was carried out beforehand in cooperation with colleagues in Oxford and the data was primarily analysed by the bioinformatician Luke O'Gorman (PhD student, University of Southampton).

2.4 Bioinformatic pipeline

FASTQ is the file format used by the NextSeq to store read sequence data, containing information on the run, lane, barcode, quality (Phred) score, and paired-end sequencing read. FASTQ files were processed and quality checked by the bioinformatician Luke O'Gorman. The

data was originally processed using a Southampton pipeline (2016), but when this was updated in 2017, where all TruSight One data was reprocessed.

The 2016 pipeline was first used to process data and the results in chapter 3 use this method of data processing. This briefly consisted of aligning FASTQ files to the human reference genome GRCh37 (hg19) using NovoAlign v2.08.02, an alignment tool specifically designed for short read mapping. This was followed by variant calling performed using SAMtools v0.1.19 which is a set of utilities that combine alignment and read depth information called SNVs and indels. BCFtools was then used to merge NovoAlign and SAMtools information into Variant Call Format (VCF) files. VCF files were further annotated using ANNOVAR against RefSeq transcripts[123],[111]. ANNOVAR is a tool that can utilise different population databases such as 1000 Genomes (1000g), Exome Aggregation Consortium (ExAC), information databases such as UCSC Genome browser, and tools to predict the functional consequences of variants such as SIFT and PolyPhen. Additional annotation was applied using the Human Gene Mutation Database (HGMD)[124].

The Southampton pipeline was updated in 2017 and was used to reprocess and analyse all data for chapter 5, the overall workflow is shown in *Figure 15*. Briefly, FASTQ files were demultiplexed and then aligned to the updated human reference genome GRCh38 (hg38) with the algorithm BWA-MEM (Burrows-Wheeler Aligner software). BWA-MEM is an alignment tool similar to NovoAlign in accuracy but with the benefit of being faster [125]. The Genome Analysis Tool Kit (GATK) v3.7 was used to find SNPs and small indels within the VCF files, followed by annotation of variants by ANNOVAR. The mutation databases InterVar (2018), DECIPHER v9.23 and HGMD (2016) were additionally used to annotate variants. Any low quality candidate causal variants were examined in Integrative Genome Viewer, which provided a visual analysis of the reads covering the area of interest. To assess quality, FastQC v0.11.3 software was used to look at the raw sequencing data and SAMtools v1.3.1 and BEDtools v2.17.0 were used to determine coverage across the genes. Possible contamination was assessed using VerifyBamID v1.0 software, which looks at excessive heterozygosity at a number of reference sites. Further to this, the known inter-sample relationships and ethnicities were checked within the NGS data by looking at the variants shared between all samples.

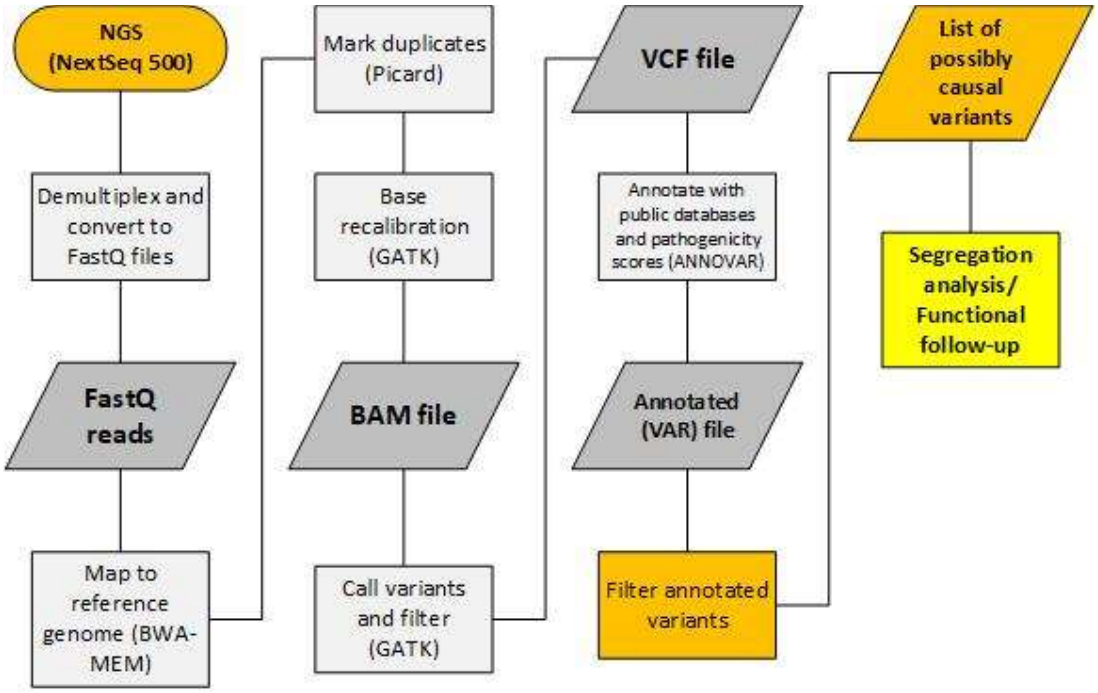


Figure 15. Data pipeline from NextSeq to functional follow-up. Areas in grey are bioinformatic steps carried out by Luke O'Gorman. The bioinformatic team also had input in the filtering of variants to produce the list of candidate causal variants. This was followed up by segregation of variants with affected family members and functional study.

2.4.1 Validating NGS data with SNP fingerprinting

SNP fingerprinting is a quality check step that can determine the correct NGS data is matched to the corresponding sample ID. The simple, yet powerful SNP fingerprinting step was designed and introduced by Pengelly *et al*. to determine NGS data provenance in WES studies where mismatches are possible due to multiple processing steps both *in vitro* and *in silico* [126]. The method involves sending replicates of the original samples to be sequenced on an optimised SNP profiling panel (LGC genomics). The panel contains 24 SNPs (Appendix table A.2) captured by commonly used enrichment kits (captured by the TSO sequencing panel) allowing a comparison between the 'SNP fingerprint' and the NGS data for accurate validation of each data-sample pairing. This quality control step was analysed by Luke O'Gorman.

2.4.2 Multiple Ligation-dependent Probe Amplification

Multiple ligation-dependent probe amplification (MLPA) is a NGS tool that can be used in conjunction with normal sequencing methods to find large deletions that may be missed by the limitations of NGS. MLPA was carried out by Simon Thomas at Salisbury Hospital. The SALSA MLPA P325 *OCA2* probe mix was used (MRC-Holland, the Netherlands) according to manufacturer's instructions. The probe mix targets both *TYR* and *OCA2* in a multiplex PCR

method. MLPA detects abnormal copy numbers by comparing a DNA sample's peak patterns to the peak pattern of reference sequences. Subsequent data were analysed using the MLPA analysis function of the GeneMarker (version 1.85) software (SoftGenetics, USA).[127] MLPA was carried out in a batch fashion, testing numerous proband DNA samples have from the large cohort to determine if this is a useful tool in genetic diagnosis.

2.4.3 Analysis of NGS variants

The TruSight One sequencing panel contains 4813 genes, each of which can contain multiple variants. A single sample of genomic DNA sequenced using this panel can have over 10,000 variants. Nearly all of this genetic variation is benign but the correct filtering is important to determine which variants are pathogenic. The first step for filtering through the large number of variants was to initially focus on genes known to cause the phenotype. Gene prioritisation is discussed in more detail in chapter 5 as it was crucial for determining which variants could provide a diagnosis for the maximum number of patients and still be time and cost effective for use in the NHS. The variants within these genes were then filtered using population frequencies and pathogenicity scores.

2.4.3.1 Prioritisation of variants

The initial filtering constraints used to prioritise variants were: (1) 1000 Genomes Project all populations (1000g_all) Minor Allele Frequency (MAF) of less than or equal to 5%, (2) the pathogenicity prediction tool SIFT with a threshold of less than 0.05, (3) PolyPhen v2 HumVar (possibly damaging and probably damaging) and (4) a GERP++ score of greater than 3.

The tools SIFT and PolyPhen are commonly used and referred to when considering *in silico* predictive scores for nonsynonymous SNVs. SIFT (Sorting Intolerant From Tolerant) is an algorithm designed to predict pathogenicity of missense mutations based on sequence homology, whereas PolyPhen predicts pathogenicity based on conservation of protein structure in relation to function [114]. GERP++ was used as a complimentary tool as it is designed to score the SNV location for evolutionary constraint [128]. These thresholds were used to define pathogenic/likely pathogenic variants in chapter 3.

The method of filtering was edited over time to account for advances in the field of pathogenicity algorithms[129]. The filtering constraints used in chapter 4, differed between two categories of 'pathogenic' and 'likely pathogenic'. A variant was defined as 'pathogenic' if it carried a 'pathogenic' annotation in ClinVar or InterVar, or 'disease causing mutation' (DM) annotation in HGMD. A 'likely pathogenic' variant was defined as a not synonymous, with a MAF

of less than or equal to 5% in the three databases; 1000 Genomes Project all populations (1000g_all), Exome Sequencing Project 6500 all populations (esp6500_all) and Exome Aggregation Consortium all populations (ExAC_all), and the variant had either a CADD Phred score greater than or equal to 15 or an absolute MaxEntScan score greater than or equal to 3.

Variants were only considered causal if the zygosity of the variant(s) were compatible with the known inheritance pattern i.e. two compound heterozygous variants in *TYR* were considered causal for OCA1. If there was a single variant determined as pathogenic/likely pathogenic for an autosomal recessive disease the genotype was considered partially resolved with a level of missing heritability.

2.5 Sanger sequencing

Sanger sequencing is a DNA sequencing technique that is based upon selective incorporation of chain-terminating dideoxynucleotides (ddNTPs) during *in vitro* DNA replication. The method was developed by Frederick Sanger and colleagues in 1977 and is still frequently used to confirm variants found with NGS methods [130]. The Sanger method was further developed by Hood *et al*. to incorporate fluorescently tagged ddNTPs (dATP, dGTP, dCTP and dTTP), each of which emit light at different wavelengths that can be distinguished with capillary electrophoresis [131]. This permits sequencing in a single reaction and has led to the development of the high-throughput sequencing systems.

The reliability of Sanger sequencing is due to the separation of new strand synthesis and the actual reading of the sequence, whereas Illumina sequencing detects each nucleotide incorporation at once, potentially resulting in less reliable signals. However, it must be noted that quality of the read is not constant throughout Sanger sequencing as the beginning and the end of the reads tend to have noisy and low quality signals, thus actually only the middle ~500bp segment of the whole fragment can be read reliably. Furthermore, NGS sequencing has the advantage of immense high-throughput due to clonal amplification, meaning each original fragment is sequenced multiple times and errors in a single read will most likely be overwritten, resulting in high accuracy. Nevertheless, Sanger sequencing remains the gold standard and has been used here to confirm variants found in NGS data and to determine presence of known variants in family members.

2.5.1 Polymerase chain reaction

The polymerase chain reaction (PCR) was used to amplify small regions of genomic DNA for Sanger sequencing. A high-fidelity Phusion DNA polymerase (Thermo Fisher Scientific, UK)

was used, which possesses 5′→ 3′ polymerase activity, 3′→ 5′ exonuclease activity. The Phusion polymerase is suited to a higher annealing temperature (+3˚C) which was factored in during primer design and the temperature cycling steps. The standard 50 µl reaction mix used when carrying out PCR for sequencing is listed below (*Table 5*) along with the temperature cycling steps (*Table 6*).

Table 5. Standard Phusion® polymerase PCR reaction mix. The HF buffer is used for standard reactions; the GC buffer is for GC rich amplicons (New England BioLabs). DMSO can also be added at a 3% final concentration.

Component	**50 µl reaction mix**	**Final concentration**
Phusion Pfu Polymerase	0.5 µl	1.0 units
5X Phusion HF or GC Buffer	10 µl	1X
10 µM Forward Primer	2.5 µl	0.5 µM
10 µM Reverse Primer	2.5 µl	0.5 µM
10 mM dNTPs	1 µl	200 µM
Template DNA	variable	< 250 ng
Nuclease-free water	to 50 µl	

Table 6. Temperature cycling steps for a standard PCR reaction using the Phusion polymerase. Annealing and extension time are variable due to length of primer and amplicon respectively.

Cycle step	**No. of cycles**	**Temperature**	**Time**
Initial denaturation	1	98°C	30 seconds
Denaturation **Annealing** **Extension**	30	98°C 45–72°C 72°C	10 seconds 10-30 seconds 15-30 seconds/kb
Final extension	1	72°C	5–10 minutes
Hold	1	4°C	∞

Gel electrophoresis was used to size-dependently separate PCR products. The PCR products were run on a 1.5% agarose gel alongside a 100 bp ladder. The gel was made with GelRed nucleic acid gel stain at a concentration of 1:30000 (1:10000 is recommended, however this was optimised to save on reagent use) and PCR product was loaded with a loading buffer containing glycerol and bromophenol blue. The gel was imaged using a UV transilluminator to determine band size and thus correct sequence amplification (example image is given in *Figure 16*).

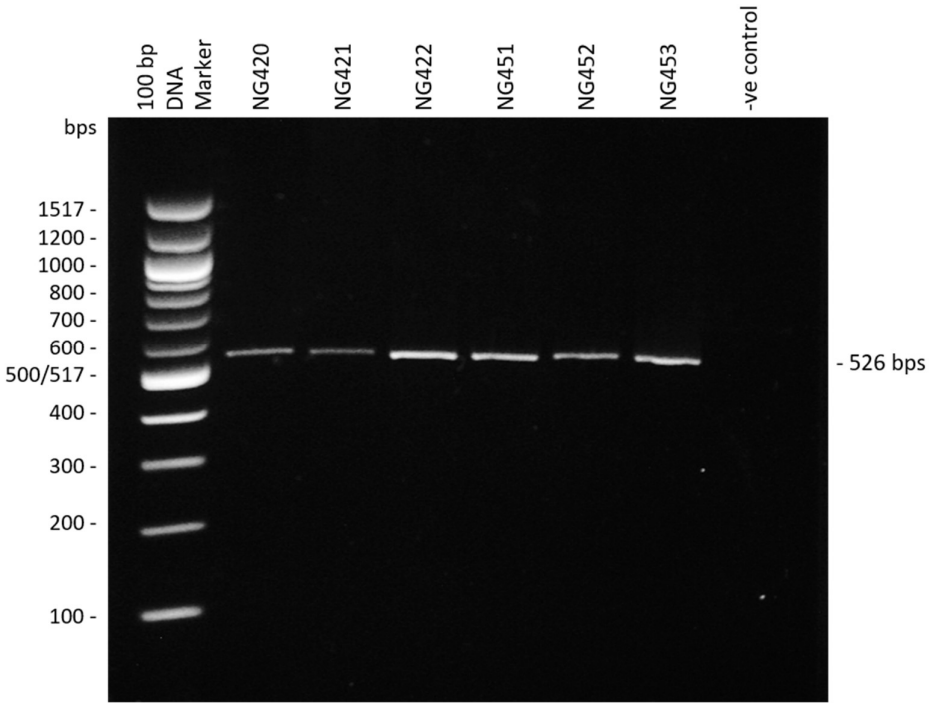


Figure 16. Example electrophoresis gel (1.5% agarose) of *TYR* exon 1 part 2 amplicon with a 100 bp ladder (New England Biolabs). The gDNA of each sample was used as a template for PCR with the primers TYR-1-part2 to produce a 526 base pair amplicon covering the SNP S192Y. The PCR products were used to sequence this region to determine the presence/absence of S192Y.

2.5.2 Designing primers

Where possible primers were found using a literature search of previous publications sequencing the genes of interest, those that were not already available were designed using multiple different tools. Firstly, the genomic and cDNA sequencing were downloaded from Ensembl in FASTA format. The sequence was manipulated and labelled in the 'A Plasmid Editor' (ApE, Univeristy of Utah, M. Wayne Davis) file, a program that allows easy design of primers and prediction of amplicon size. Primers were designed to capture regions of interest, generally producing an amplicon greater than 200 bps to make it suited for Sanger sequencing, but smaller than 800bps to improve the efficiency of the PCR amplification step. The primer sequences were designed to be between 15-25 bps, to contain roughly equal ratio of GC to AT, to be devoid of long repeats, and end in a G/C-clamp where possible. The lower primer sequence was reverse complemented to the template gDNA using an online tool (http://www.bioinformatics.org/sms/rev_comp.html), after which the forward and reverse primers were analysed using the PremierBiosoft NetPrimer online tool (http://www.premierbiosoft.com/NetPrimer/AnalyzePrimer.jsp) to predict any unwanted dimerization or hairpin structures. In the case of difficult sequences, Primer3

(http://primer3.ut.ee/) was used to choose primers by inputting a sequence of interest and defining specific parameters. The suggested primers were then analysed on NetPrimer and confirmed on ApE. A nucleotide BLAST search of primers against the human genome was carried out to check for primer specificity (https://blast.ncbi.nlm.nih.gov/). All primer oligonucleotides were ordered from Sigma-Aldrich and then optimised over a range of annealing temperatures before use. It was important to use specific primers for exon 4 of *TYR* due to the 98.55% sequence similarity between the C-term region of this gene and that of TYR-like gene (*TYRL*)[132].

Table 7. Primers used for polymerase chain reaction and Sanger sequencing.. NCBI accession numbers: *TYR* NM_000372, *PAX6* NM_001258464. *Primers were taken from Kalahroudi, V. G., et al. (2014)[133]

Gene	Exon	Forward Primer (5'-3')	Reverse Primer (5'-3')
TYR	1a	TTCAGAGGATGAAAGCTTAAGATAAA *	CGTCTCTCTGTGCAGTTTGG *
TYR	1b	CATCTTCGATTTGAGTGCCC *	CCCTGCCTGAAGAAGTGATT *
TYR	3	AGTTATAAATCAAATGGGATAATCA *	ACATTTGATAGGCACCCTCT *
TYR	4	CTGTTTCCAATTTAGTTTTATAC *	CTGTTTCCAATTTAGTTTTATAC *
PAX6	13	GACTCATTTCCCCTGGTGTG	TGTGTCCCCATAGTCACTGAC
OCA2	13	GCTCCCTGTTCTTAAAGTCACT	CTGGAATGCAGTGAGCTGTG

2.5.3 PCR purification for sequencing

Following the PCR reaction, the PCR product needed to be purified for Sanger sequencing. To achieve this a QIAquick PCR purification kit (Qiagen) was used. Binding buffer (buffer PB) was combined with the PCR product at a ratio of 1:5 and this was applied to a column inside a 2ml collection tube. The column was then centrifuged for 1 minute at 17,900 x*g* to bind the DNA. The wash buffer (buffer PE) was applied and the column was centrifuged again, followed by a final centrifugation to ensure no wash buffer remained, before the column was transferred to a clean 1.5 ml microcentrifuge tube. Elution buffer (buffer EB) (~30 µl for an initial PCR product volume of 40 µl to 60 µl) was placed on the centre of the column matrix, incubated for 1 minute and the column centrifuged to elute the purified DNA.

The concentration of the purified DNA product was measured on the NanoDrop ND-1000 so the DNA could be diluted to 10 ng/µl for sequencing. Sanger sequencing was outsourced to Nottingham Source Bioscience and results were returned as both a chromatogram and a sequence file. Both forward and reverse reads were sequencing for each sample.

2.5.4 Analysis of Sanger sequencing

The Sanger sequencing results are represented as a chromatogram, each peak denoting a nucleotide. Sequencing quality can be determined by analysing the height and spread of the peaks. Frequently, the peaks at the start and end of the sequencing read are mixed and with lower signal intensities, the sequence is therefore more reliable when the chromatogram is clearer. Two peaks at one point suggest the sample is heterozygous in that poisition (*Figure 17*). The sequencing files were opened in ApE and the data was aligned with a reference sequence.

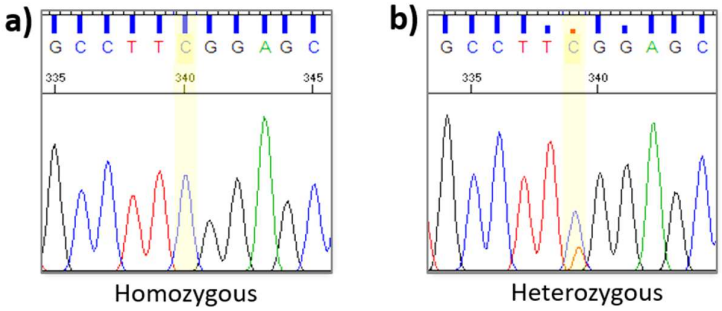


Figure 17. Sanger sequencing chromatograms.. a) Homozygous for a 'C' at the location highlighted b) Heterozygous for 'C/T' at the same location shown by overlapping peaks.

2.6 Tyrosinase Functional Work

2.6.1 Tyrosinase plasmid vector

The tyrosinase plasmid vector was purchased from Addgene (Massachusetts, USA) and was initially deposited by Ruth Halaban (*Figure 18*)[134]. It was decided that the C-terminus 3XFLAG-tag would be helpful for protein purification as well as being a good recognition site for antibodies in western blotting. The vector was suited to both bacterial cloning for site-directed mutagenesis and to transient transfection of mammalian cells for expression the protein.

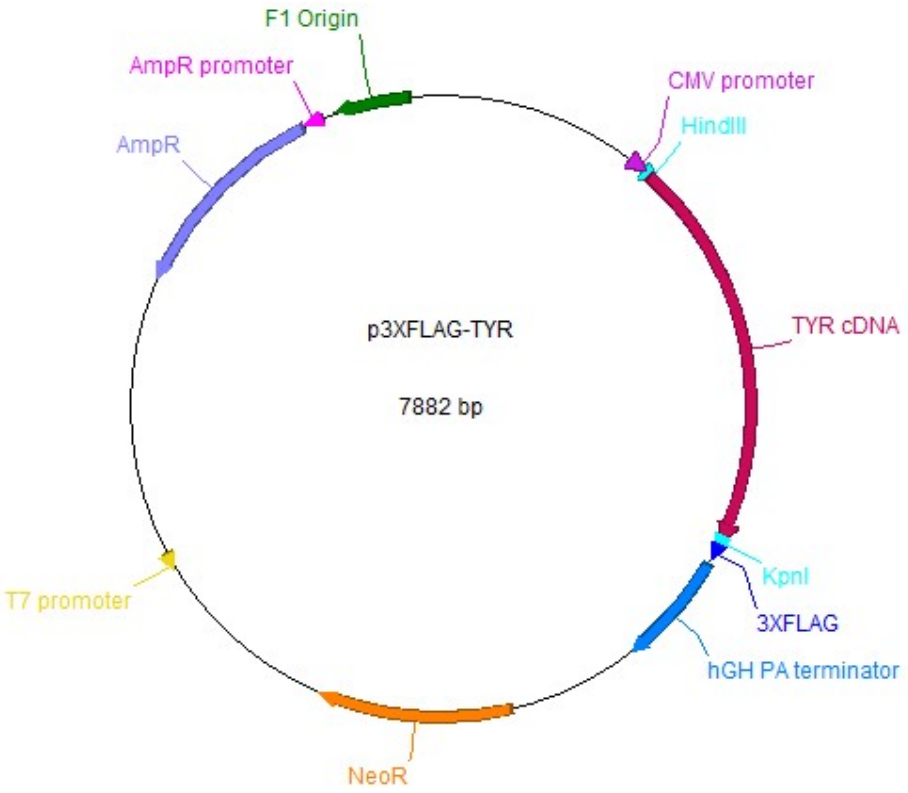


Figure 18. The plasmid vector p3XFLAG-CMV-14 containing *TYR* cDNA. The vector contains and F1 origin for high copy number in bacteria and a CMV promoter for eukaryotic protein expression. The ampicillin resistance gene (AmpR) was used for bacterial selection. When expressing protein, the plasmid produces a recombinant tyrosinase protein with a 3XFLAG-tag at the C-terminus.

2.6.2 Site-directed mutagenesis

On arrival of the p3XFLAG-TYR vector, the *TYR* cDNA was Sanger sequenced to confirm its wildtype sequence. However, sequencing revealed the S192Y (c.C575A) common population variant to be present, therefore this needed to be removed to create the wildtype sequence (c.575C, p.192Ser). The primers used for each variant inserted through site-directed mutagenesis are listed in *Table 8*.

Table 8. The primers used to introduce each sequence mutation within the p3XFLAG-TYR vector.

Nucleotide Mutation	Amino Acid Change	Upper Primer (5'- 3')	Lower Primer (5'- 3')
c.575A>C	Tyr192Ser	ctgcttgggggatctgagatctggagagacattgatttt	aatgtctctccagatctcagatcccccaagcagtgcatcc
c.1205G>A	Arg402Gln	tttgaacagtggctccaaaggcaccgtcctcttcaagaag	tgaagaggacggtgcctttggagccactgttcaaaaatac

Site-directed mutagenesis was carried out using the non-strand displacing activity of *Pfu* DNA polymerase to incorporate and extend the mutagenic primers. The PCR reaction mix and thermocycler program were the same as previously described in section 2.5 (*Table 5* and *Table 6* respectively). The PCR product was treated with the enzyme DpnI to digest the methylated parental DNA. The product was then used to transform competent DH5α cells.

2.6.3 Cloning the plasmid vector

2.6.3.1 Transformation of competent DH5α cells

Competent cells (New England BioLabs, USA, 5-alpha Competent E. coli (High Efficiency)) were transformed via heat shock method. Briefly, 50 µl of thawed cells were kept on ice and combined with approximately 100 ng of plasmid DNA and incubated for 30 minutes. The cell-DNA mixture was heat shocked at 42˚C for 30 seconds and then placed on ice for five minutes. Cells were given SOC medium and incubated for an hour in a shaking incubator before being plated on ampicillin selection (100 µg/ml) LB agar plates. After overnight incubation at 37˚C, single ampicillin resistant colonies were picked and grown in LB broth for approximately 16 hours, at which point the cells were pelleted by centrifugation and the DNA extracted. When the stocks were diminished, competent cells were produced through treatment with $CaCl_2$ and subsequently transformed using the heat shock method described above.

2.6.3.2 DNA purification from bacterial cells

A QIAprep Spin Miniprep Kit (Qiagen) was used to extract the DNA from the transformed DH5α cells. This involved pelleting the cells at 8,000 x*g*, resuspension in a buffer containing RNase, followed by cell lysis and then pelleting undesired membrane and proteins at 13,000 x*g*. The supernatant was applied to a spin column where it was washed with ethanol containing wash buffer and then eluted in Tris-HCl pH 8.5. The concentration of purified samples was determined via NanoDrop. An aliquot of the plasmid DNA was diluted to a concentration of 100 ng/µl and sent for Sanger sequencing as described in section 2.5. The commercial pCMVF primer was used as well as the *TYR*_1a and *TYR*_1b primers shown in *Table 7*.

2.6.4 Protein production in mammalian cells

2.6.4.1 Culturing HEK293F cells

The Freestyle 293 Expression System, containing Human Embryonic Kidney 293 Freestyle (HEK293F) cells, was purchased from Invitrogen (California, USA). Cells were cultured in Freestyle culture medium at 37˚C in a shaking incubator at 125 rpm with 8% CO_2. HEK293F cells

were counted on a haemocytometer in the presence of trypan blue dye to determine their viability. Cells were passaged when they exceeded a density of 3 $x10^6$ cells/ml and vortexed for up to 20 seconds to break up clumps and create a single cell suspension. The cultures were not treated with antibiotics as this slowed cell growth and reduced viability. Stocks were created by freezing down 1 $x10^6$ early passage cells in 1 ml of Freestyle media with 10% DMSO. The cryovials containing the cells were cooled at 1˚C per minute using a Nalgene Mr. Frosty freezing container (Thermo Fisher Scientific, UK) kept at -80˚C overnight. The vials were transferred to liquid nitrogen afterwards for long term storage.

2.6.4.2 **Transfection of HEK293F cells**

To prepare for transfection, cells were vortexed and diluted to a density of 5 $x10^5$ cells/ml the day before. This allowed cells to achieve >90% viability, with a single cell suspension at a density of 1 $x10^6$ cells/ml at time of transfection. Opti-MEM transfection medium was combined with 30 µg of plasmid DNA up to the volume of 1 ml and sterilised with a 0.22 µm filter. The lipid-based reagent, 293fectin (60 µl), was diluted in Opti-MEM and incubated at room temperature for 5 minutes. DNA and 293fectin was combined, gently mixed and incubated at room temperature for 30 minutes before adding to cells. Cultures were then lysed after 72 hours of growth at 37˚C in a shaking incubator.

2.6.4.3 **Cell lysis**

Cells were pelleted by centrifugation at 100 *xg* for 5 minutes and then washed twice with ice-cold PBS, re-pelleting each time. The washed cells were lysed on ice with NP-40 lysis buffer containing 1 mM phenylmethylsulfonyl fluoride (PMSF) (in DMSO with a final concentration of 1%) and 1X protease and phosphatase inhibitor (Halt™ Phosphatase Inhibitor Cocktail, Thermo Fisher Scientific, UK). The lysis reaction was incubated for 30 minutes on ice and vortexed every 10 minutes. Once lysed, cells were centrifuged at 10,000 *xg* for 10 minutes at 4˚C and the supernatant was aliquoted into 0.2 ml tubes and stored at -80˚C.

2.6.4.4 **Determining protein concentration**

A bicinchoninic acid (BCA) assay (Pierce™ BCA Protein Assay Kit, Thermo Fisher Scientific, UK) was used to determine the total protein concentration of each whole cell lysate. A plate reader was used to measure the optical density at a wavelength of 570 nm. A protein standard curve was generated using bovine serum albumin (BSA) as the standard and all samples and standards were in triplicate. When creating the standard curve of known BSA concentrations an R^2 value of > 0.95 was considered acceptable, example standard curve in *Figure 19*.

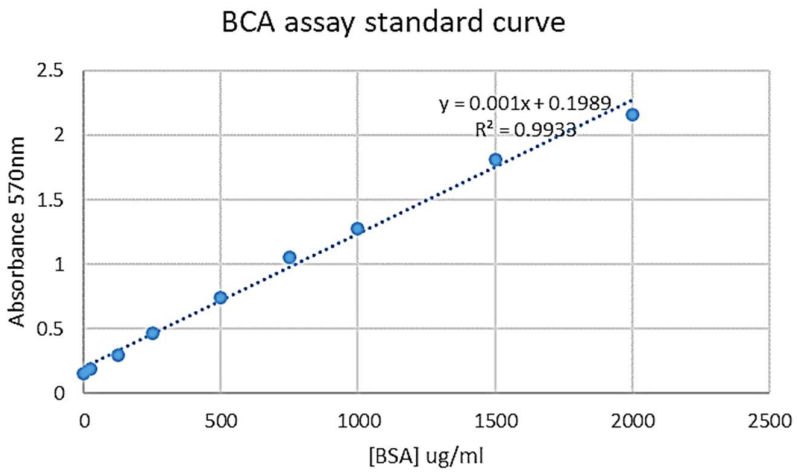


Figure 19. Example BCA assay standard curve with BSA as the known protein. Here the equation for the line of best fit is y=0.001x+0.1989 which can be rearranged to calculate 'x' for unknown protein samples. The R^2 value is greater than 0.95 so the graph is considered reliable.

2.6.5 Endo H digest

Endoglycosidase H (endo H) cleaves mannose rich oligosaccharides that are present on glycosylated yet still immature protein. As the protein is processed through ER and the compartments of the Golgi it is still susceptible to digestion, that is until Golgi alpha-mannose II removes the mannose sugars and the protein is no longer recognised by endo H[135]. For the endo H digestion (New England Biolabs), 20 µg of total protein was denatured in 1X denaturing buffer at 100˚C for 10 minutes. This was followed addition of 3 µl of endoglycosidase-H and 1X Glycobuffer 3. The reaction was incubated at 37˚C for 1 hour and then a western blot was used to determine the quantity of each tyrosinase fragment according to known fragment size.

2.6.6 Western blot of endo H digested tyrosinase

Protein was prepared for SDS-PAGE through further denaturation by the addition of the reducing agent β-mercaptoethanol in a solution containing 4 µg of digested protein, 1X NuPage LDS sample buffer, and 1% β-mercaptoethanol. As the protein has already undergone denaturation during endo H digest, the solution was incubated at 37˚C for 10 minutes, furthermore, the lower temperature avoided any possibly aggregation of the protein as a result of its transmembrane domain. The prepared protein was separated in a 10% SDS-PAGE gel run at 125 V for an hour and then transferred at 35 V for 1 hour to a nitrocellulose membrane. The membrane was blocked for 90 minutes in a buffer of tris-buffered saline containing 5% bovine serum albumin and 0.1% tween (1X TBS-T), at pH 7.6. The membrane was then incubated overnight at 4˚C with the mouse anti-DDDDK tag HRP (Horseradish Peroxidase) conjugated

antibody (equivalent to anti-FLAG) (ab49763, Abcam, UK) at a concentration of 0.1 µg/ml in 5% BSA. Membranes were washed three times in 1X TBS-T and an enhanced chemiluminescent solution (Bio-Rad, UK) was applied to develop the membrane. The blot was imaged using a ChemiDoc Gel imaging system (Bio-Rad, UK).

2.6.7 DOPA-oxidase assays

A DOPA-oxidase assay was used to determine the enzymatic activity of the recombinant tyrosinase mutants in comparison with wild-type. The assay relies on the colour change produced by the oxidation of L-DOPA to Dopaquinone and subsequent spontaneous conversion of dopaquinone to dopachrome[79]. The reaction mix contained 200 ng of total protein in 1 mM phosphate buffer containing 3.3 mM L-DOPA at pH 7.4. On addition of protein, an immediate reading was taken of optical density at UV max of 492 nm. The reaction was incubated at 37 ˚C and measured every 5 minutes for the first 15 minutes and then every 30 minutes over a 3 hour time-course.

A similar DOPA-Oxidase assay was used with the incorporation of 3-methyl-2-benzothiazolone hydrazine (MBTH). The reaction mix contained 120 ng of total protein in 50 mM phosphate buffer containing 1.1 mM L-DOPA and 6 mM MBTH at a pH of 7.0. The addition of MBTH was to sequester dopaquinone as this is the direct product of tyrosinase's DOPA-Oxidase activity. As MBTH bound the quinone it produced a pink colour which darkened as more dopaquinone was produced[136]. The reaction was incubated at 37 ˚C and the absorbance measured at 492nm every 5 minutes for the first 15 minutes, and then every 30 minutes over a 3 hour time-course.

2.6.8 Transmission Electron Microscopy

HEK293F cells transfected with p3XFLAG-*TYR* were centrifuged at 100 x*g* for 5 minutes to loosely pellet cells. The supernatant was discarded and the cells fixed in 3% glutaraldehyde and 4% formaldehyde in 0.1 M PIPES buffer at pH 7.2 for 1 hour. After fixation, the cells were processed and imaged. Processing was carried out by the University's Biomedical Imaging Unit. Briefly, cells were pelleted and rinsed twice for 10 minutes in 0.1 M PIPES at pH 7.2. They were stained with 2% aqueous uranyl acetate for 20 minutes and washed with increasing concentrations of ethanol in 10 minute intervals (30%, 50%, 70%, 95%, 100%). The sample was further dehydrated with application of acetonitrile for 10 minutes followed by overnight incubation in a 1:1 ratio of acetonitrile and Spurr resin. The fixed and dehydrated cells were embedded in resin and the resin beads cut in ultra-thin slices in a microtome (Leica Biosystems,

Germany). The slices were mounted on mesh copper grids, stained with Reynolds lead stain and imaged using a FEI Tecnai T12 Transmission Electron Microscope.

Chapter 3: Missing heritability in albinism

3.1 Introduction

Albinism is a term used to describe a group of genetic disorders affecting melanin biosynthesis and resulting in hypopigmentation of skin, hair, and ocular tissues[51]. Furthermore, it has significant effects on visual development and results in profound visual disability. The most striking clinical features in patients with albinism include reduced visual acuity, nystagmus, strabismus, and photophobia. A detailed ophthalmic examination may reveal foveal hypoplasia (lack of foveal pit at the back of the retina), asymmetry of visual evoked potentials (VEP) which suggest excessive decussation of the optic chiasm, and iris transillumination[51],[95]. Current treatment of albinism is to manage symptoms and comprises: correction of any refractive errors, management of head postures/strabismus and effective sun protection. Another important factor in the management albinism is genetic counselling, which makes genetic diagnosis critical for patients.

Six genes involved in melanin biosynthesis pathway are known to cause forms of OCA and OA: *TYR, OCA2*, *TYRP1, SLC45A2*, *SLC24A5*, and *LRMDA* (previously *C10orf11*) accounting for OCA subtypes 1-4 and 6-7 respectively, and *GPR143* accounting for OA1 (*Table 9*). Apart from X-linked OA1, all of the OCA subtypes are understood to be inherited as autosomal recessive disorders but the subtypes are heterogeneous in pigmentary phenotype[51, 137, 138]. OCA may also be caused by syndromic conditions such as Chèdiak-Higashi syndrome and Hermansky-Pudlack syndrome. The syndromic conditions damage more than just melanogenesis with multisystem consequences such as immune defects and bleeding problems[77].

Partial (hypomorphic) phenotypes have been described widely in the literature where those with possible albinism exhibit some ocular or pigmentation features but lack others (e.g. nystagmus or foveal hypoplasia). However, phenotyping methods have varied significantly and the detailed phenotype has rarely been described in detail [139] [140] [141]. OCA1 has a mixed phenotype and is further split into OCA1A and OCA1B. OCA1A describes complete loss of tyrosinase activity (previously described as 'tyrosine negative' albinism) and is characterised by an apparent total lack of pigment. Some tyrosinase function is retained in OCA1B, allowing pigment to accumulate and generate a phenotype of minimal to near normal skin pigmentation, as is also the case for the other described OCA and OA phenotypes [51, 138]. Phenotypes of partial OCA also overlap with those seen in patients with dominant mutations in the *PAX6* gene, which is involved in ocular development, where a variety of phenotypes have been described including

foveal hypoplasia, iris transillumination, and nystagmus [97]. Furthermore, syndromic causes of OCA may be incorrectly diagnosed as isolated albinism if detailed phenotyping is not performed.

Table 9. The HGNC approved gene names associated with the subtypes of OCA, OA, and *PAX6* phenotypes. OCA5 has been attributed to a chromosomal location but does not yet have an associated gene[53].

HGNC symbol	HGNC name	Associated disorder	Mode of inheritance
TYR	Tyrosinase	OCA1A	Autosomal recessive
		OCA1B	
OCA2 (P gene)	OCA2 melanosomal transmembrane protein	OCA2	Autosomal recessive
TYRP1	Tyrosinase related protein 1	OCA3	Autosomal recessive
SLC45A2	Solute carrier family 45 member 2	OCA4	Autosomal recessive
-	Chromosomal location 4q24	OCA5	Autosomal recessive
SLC24A5	Solute carrier family 24 member 5	OCA6	Autosomal recessive
LRMDA (C10orf11)	Leucine-rich melanocyte differentiation-associated protein	OCA7	Autosomal recessive
GPR143	G protein-coupled receptor 143	OA1	X-linked recessive
PAX6	Paired-box protein 6	Aniridia, nystagmus, foveal hypoplasia	Autosomal dominant

As the most severe form of OCA, OCA1A is often recognised in early infancy. King *et al*. proposed that white hair from birth can be used to predict OCA1[138] and showed that 85% of patients identified in this way tested positive for pathogenic *TYR* mutations. However, 15% of supposed OCA1 cases identified didn't appear to carry an accountable genetic mutation, and 29% of those confirmed as OCA1 had only a single heterozygous *TYR* mutation [138]. It is widely recognised that the OCA genes do not account for all non-syndromic cases, with as many as 30% of OCA1A occurrences having an unknown genetic origin [57, 64]. Indeed, this percentage may be even higher for cases of partial albinism [52]. It is also important to note that a variety of techniques have been employed to screen for tyrosinase gene mutations in previous studies and

no method has 100% sensitivity, often overlooking the possible effect of intronic and regulatory regions.

An individual's pigmentary phenotype depends on polymorphisms in many genes, including polymorphisms in the OCA genes [142],[60, 143]. It is possible that pigmentary background may play a large role in a person's susceptibility to the albinism phenotype, with 'not so damaging' variants having a greater effect on an already less active pigmentary pathway [62, 75]. In a letter to the American Journal of Medical Genetics, it was suggested that inheritance of OCA2 is not purely recessive, with the example of haploinsufficiency noticeably reducing skin pigmentation in a Hispanic family, arguably due to their already fair skin tone [142]. It has also been suggested that a synergistic interaction between genes throughout the pigment pathway may exist in albinism phenotypes, evidenced by one family exhibiting an OCA2 phenotype that is modified by a mutation in the OCA3 gene and a correlation between *OCA2* and *MC1R* variants in a small albinism cohort[143], [67]. The quantitative effect of pigmentation also has relevance to OCA1B, particularly the notion of autosomal recessive ocular albinism (AROA), an arbitrary characterisation that has been used previously to describe cases with clinically mild OCA1B [74, 144].

AROA sparked a debate over the possible pathogenicity of two *TYR* polymorphisms, rs1126809 (p.R402Q) and rs1042602 (p.S192Y), common in Caucasian populations with allele frequencies ~28%-36% (1000g_EUR)[145]. Functional studies have shown the R402Q polymorphism produces a thermolabile enzyme, retained by the cells endoplasmic reticulum, with a 75% reduction in catalytic activity compared to the wild-type [60, 146, 147]; and S192Y results in a 40% reduction of tyrosinase enzymatic activity [59]. Multiple OCA1 studies have shown the R402Q allele is strongly associated with albinism patients who have only one other identifiable pathogenic variant [52, 74, 75].

R402Q has been proposed as a causal variant when inherited on the *trans* allele to a null activity *TYR* mutation [74, 144]. However this was disputed with evidence of no OCA phenotype in the parents of affected probands even when they carried a combination of null mutation and R402Q[148]. This has led to the suggestion that an additional variant (in addition to a null mutation and R402Q) is necessary for manifestation of the ocular phenotype. The combination of two common variants may produce a reduction in tyrosinase activity that, when co-inherited with a deleterious *TYR* mutation, provide a sufficient loss of activity to cause an albino phenotype[60, 62].

In this chapter, the role of common population variants in albinism has been investigated using detailed phenotyping and segregation studies.

3.2 Hypothesis

The controversy over the pathogenicity of common SNPs led to the hypothesis: 'The common polymorphisms R402Q and S192Y are pathogenic when co-inherited with a loss of function mutation on the *trans TYR* allele'.

3.3 Next generation sequencing preliminary work

The background work, carried out in sections 2.3.2 and 2.3.4, involved NGS data that was collected previous to this body of work. Firstly, 33 samples with infantile nystagmus were sequenced using Agilent's HaloPlex Select custom panel, targeting the coding regions for 55 genes associated with nystagmus, albinism and related ocular disorders. A further six nystagmus samples were sequenced with the Agilent's SureSelect v5.0 system which targets the entire exome. The data was then analysed and used as preliminary work for this project.

3.3.1 Methods and results for preliminary work

Library preparation using the custom HaloPlex panel was carried out by a colleague (methods section 2.3.2) and run on a NextSeq 500 sequencer (methods section 2.3.3). The raw data was processed through the 2016 Southampton bioinformatics pipeline (methods section 2.4) and prioritisation of variants was carried out following guidelines described earlier (methods section 2.4.3.1). Whole-exome sequencing with Agilent's SureSelect v5.0 system (methods section 2.3.4) was carried out by colleagues in Oxford and the data processed by Luke O'Gorman (PhD student, bioinformatician).

Notably, no HaloPlex samples with a putative *TYR* mutation (five samples) had a second *TYR* variant with a deleterious prediction, providing a high level of missing heritability. The common variants S192Y and R402Q were investigated, and found in combination with the deleterious mutations in three out of five probands. There were a further two cases with the same *TYR* plus the R402Q:S192Y genotype identified in the whole-exome cohort (*Table 10*).

Table 10. Phenotype and genotype of probands. The presence of variants as determined from NGS data. *TYR* transcript accession number: NM_000372.

	Proband	**Sex**	**Phenotype**	***TYR* variant 1**	***TYR* R402Q**	***TYR* S192Y**
Custom HaloPlex Panel	NG066	Male	Partial albinism	T373K	Het	Hom
	NG083	Female	Albinism	R217W	Het	Hom
	NG154	Female	Partial albinism	G254V	Het	Hom
Whole-Exome	NG222	Female	Albinism (Also has Noonan syndrome)	H19R	Het	Hom
	NG241	Male	Mild albinism	P81L	Het	Hom

Sanger sequencing was used to further confirm the variants in the proband and to investigate segregation in family members (*Figure 20*). As the project continued, the collection of DNA and phenotyping of family members was improved so the amount of phenotype data available for these early samples is comparatively limited to those collected in later years.

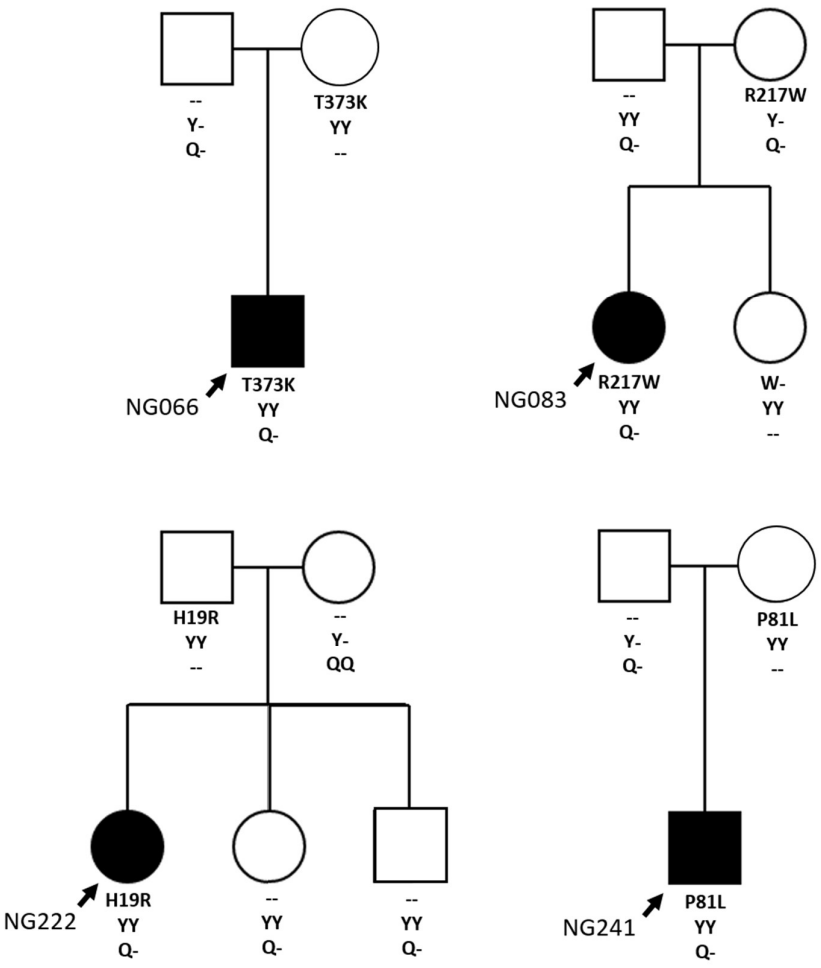


Figure 20. Pedigree trees demonstrating the inheritance and segregation of variants in each family. Wild-type at each locus is indicated by '-', putative variant amino acid change is for *TYR* NM_000372, presence of the variant S192Y is indicated by 'Y' when heterozygous or 'YY' when homozygous, and presence of the R402Q variant is indicated by a 'Q' in the same format. NG154 does not have any family data available so has not been included.

3.3.2 Discussion of preliminary data

The preliminary data suggests that a combination of S192Y and R149Q mutant alleles ranging from heterozygous to homozygous could modify the phenotypic effects of other *TYR* mutations. Both the research exome and HaloPlex panel provided good data for the albinism samples, however they are on either end of the broad-to-targeted NGS spectrum. It was decided that next step was to use a targeted approach that was already commercially available, thus reducing cost, with broader application in terms of nystagmus and the NHS in general. The

TruSight One targeted exome panel was chosen, dubbed a 'clinical exome' by those already using it. Analysis of clinical exome data for more recent samples (with their greater detail of phenotyping information) was chosen as the most appropriate methodology to examine the common polymorphism hypothesis.

3.4 Methods

The known albinism genes were investigated in eighteen patients with possible hypomorphic albinism phenotypes, identified through detailed ocular phenotyping in a tertiary eye clinic. Hypomorphic albinism was defined as at least two phenotypic features of albinism i.e. skin and hair pigmentation deemed to be low within the family context, nystagmus, foveal hypoplasia, VEP crossing or iris transillumination (probands with full features were excluded to select for hypomorphic phenotypes).

3.4.1 NGS data

Patient recruitment and DNA collection was carried out as described in methods section 2.12.1. Proband DNA samples were processed for NGS sequencing following the protocols described in methods section 2.22.3 for the Illumina TruSight One v1.0 targeted sequencing panel and the Illumina NextSeq 550 sequencer. Following sequencing, the raw data output underwent processing and quality checks by the bioinformatician Luke O'Gorman (following the protocol in methods section 2.4.3). Multiple ligation-dependent probe amplification (MLPA) was carried out for the *TYR* and *OCA2* genes for samples (following protocol in methods section 2.4.2).

Variants within the albinism genes; OCA genes 1-4 and 6, the OA1 gene, syndromic albinism genes and *PAX6* were filtered to provide only pathogenic and likely pathogenic variants using population databases and pathogenicity scores. Firstly, synonymous variants were excluded, followed by exclusion of very common variants, and then the predictive scores from SIFT, PolyPhen2, and GERP++ were used to provide *in silico* evidence for the pathogenicity of variants that remained. A threshold for each score was set; SIFT greater than or equal to 0.05, PolyPhen2 equal to possibly damaging and probably damaging, and GERP++ greater than or equal to 3 (methods section 2.4.3.1).

3.4.2 Sanger sequencing data

Probands left as partially resolved with only a single heterozygous *TYR* mutation were further investigated for the common SNPs S192Y and R402Q. Sanger sequencing was used to confirm and segregate each *TYR* variant in probands and family members (primers used are listed in *Table 7*). The workflow and outputs at each stage are shown in *Figure 21*.

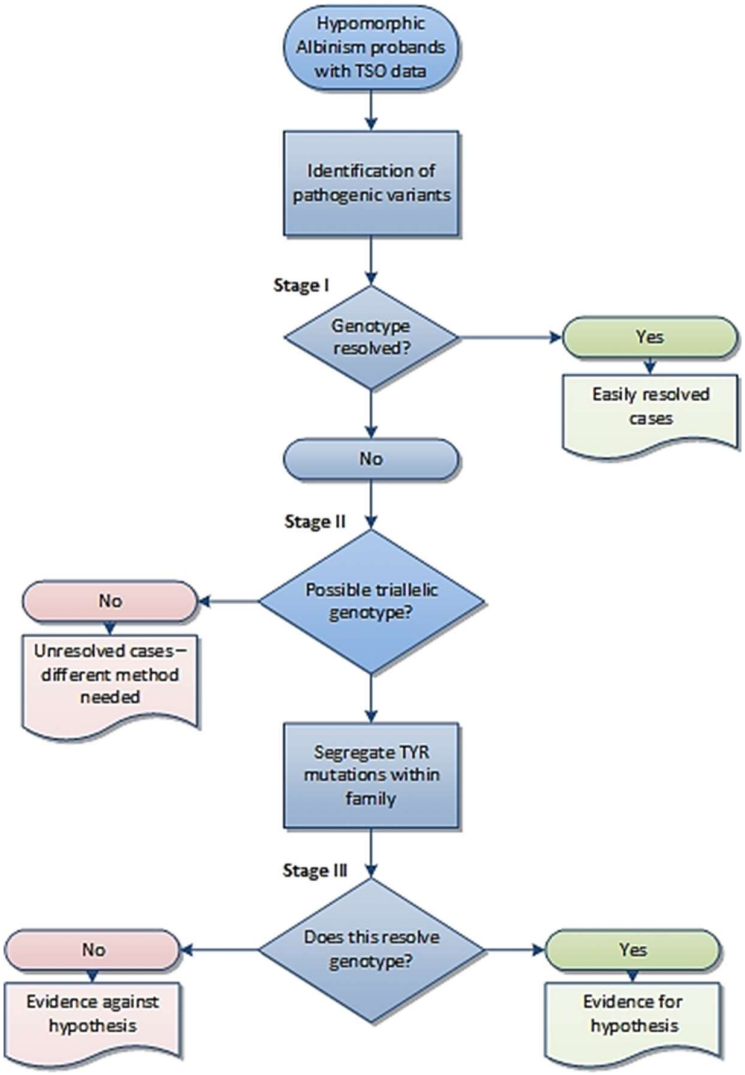


Figure 21. Workflow for investigating the genotype of probands with hypomorphic albinism, followed by segregation of the tri-allelic genotype within families. Probands without an easily resolvable phenotype as stage I, but that do carry a *TYR* variant, are taken to stage II and inspected for common variants. Those carrying the common variants are taken stage III where family member genotypes are investigated with Sanger sequencing to map inheritance.

3.5 Results

3.5.1 Phenotyping results

Presentation of hypomorphic albinism varied in both ocular phenotype and cutaneous pigment level between probands and between affected family members. For example, NG257 and his mother both have a phenotype consistent with partial albinism. However, NG257 appeared to have a lower cutaneous pigment level in comparison with family but does not have iris transillumination, whereas his mother has normal cutaneous pigment levels in context of family but iris transillumination defects. The level of foveal hypoplasia also varied between patients and within families. Example OCT images from the cohort demonstrate the broad range of foveal developmental anomalies (*Figure 22*).

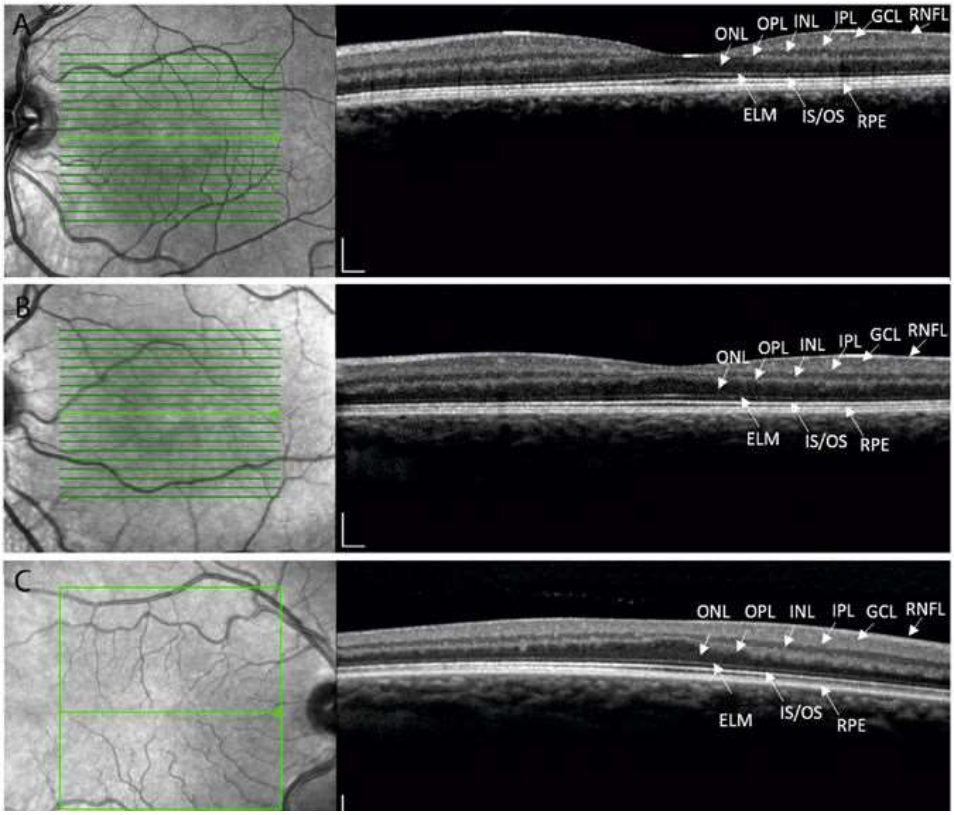


Figure 22. OCT images using the Heidelberg Spectralis Diagnostic imaging platform. A) Normal fovea (Unaffected mother of NG322) B) Foveal hypoplasia grade 1 (Affected brother of proband NG322) C) Foveal hypoplasia grade 3 (Affected mother of proband NG356). Foveal grading according to the Thomas *et al*. grading system. Outer nuclear layers (ONL), outer plexiform layers (OPL), inner nuclear layers (INL), inner plexiform layers (IPL), ganglion cell layers (GCL) and retinal nerve fibre layers (RNFL) are labelled.

3.5.2 Variants after prioritisation and filtering

Four of the probands can be categorised as genetically resolved. NG270 has a likely pathogenic mutation in the *PAX6* and NG272 (male proband) has a deletion resulting in a frameshift mutation in the X-linked gene, *GPR143*. NG213 and NG340 each have two compound heterozygous mutations in the *OCA2*.

There are a remaining 16 probands with no/partial genetic resolution. NG265 has a single mutation in *OCA2* and a second mutation in *TYRP1.* As these genes cause different OCA subtypes, this genotype requires further investigation before concluding a combined pathogenicity. Two probands, NG322 and NG344, each have a single heterozygous mutation in the *OCA2* gene with no second mutation identified. Furthermore, six probands each had a single heterozygous mutation in the *TYR* gene with no further variants passing the filtering threshold (*Table 11*).

Table 11. Predicted causal variants from clinical exome data, in eighteen probands with phenotypes matching hypomorphic albinism. Pathogenicity was determined by filtering all variants in the genes; *TYR*, *OCA2*, *TYRP1*, *SLC45A2*, *SLC24A5*, *C10orf11* and *PAX6*, with the parameters MAF <0.05, SIFT <0.05, PolyPhen2 = possibly damaging or probably damaging, and GERP++ >2. The prediction scores for non-synonymous variants are included, for some mutations a prediction score was not available at the time of analysis. NCBI transcript accessions numbers: *TYR* NM_000372, *OCA2* NM_000275, *PAX6* NM_001258465, T*YRP1* NM_000550, and *GPR143* NM_000273. *indicates novel mutations

Proband	Variant 1	Variant 2	Variant 3
NG167	-	-	-
NG178	-	-	-
NG213	***TYR* c.G529T p.V177F** (SIFT= . PolyPhen=D GERP=5.16)	*OCA2* c.G822C p.W274C* (SIFT=0 PolyPhen=D GERP=4.66)	*OCA2* c.C2020G p.L674V (SIFT=0.03 PolyPhen=D GERP=5.75)
NG250	***TYR* c.1467dup p.T489fs**	-	-
NG251	***TYR* c.505_507del p.D169del**	-	-
NG257	***TYR* c.732_733del p.C244X**	-	-
NG263	***TYR* c.C1204T p.R402X**	-	-
NG265	*OCA2* c.A1465G p.N489D (SIFT=0.01 PolyPhen=D GERP=5.33)	*TYRP1* c.C1037G p.P346R (SIFT=0 PolyPhen=D GERP=5.73)	-

NG270	*TYR* c.C1217T p.P406L (SIFT= . PolyPhen=D GERP=4.68)	*PAX6* c.C1264A p.Q422K (SIFT=0 PolyPhen=D GERP=6.16)	-
NG272	*GPR143* c.485delG p.W162fs*	-	-
NG296	-	-	-
NG309	***TYR* c.C1217T p.P406L** (SIFT= . PolyPhen=D GERP=4.68)	-	-
NG322	*OCA2* c.C1606T p.R536C* (SIFT=0.01 PolyPhen=D GERP=5.8)	-	-
NG327	-	-	-
NG333	-	-	-
NG340	*OCA2* c.G1327A p.V443I (SIFT=0.02 PolyPhen=D GERP=5.2)	*OCA2* c.A1025G p.Y342C (SIFT=0 PolyPhen=D GERP=5.55)	-
NG344	*OCA2* c.G1327A p.V443I (SIFT=0.02 PolyPhen=D GERP=5.2)	-	-
NG356	*TYR* c.C1264T p.R422W (SIFT= . PolyPhen=D GERP=2.69)	-	-

MLPA of *TYR* and *OCA2* was carried out in NG167, NG178, NG213, NG250, NG251, NG257, NG265, NG270, NG272, NG296, and NG309 to search for large deletions that would be missed in the clinical exome NGS data. MLPA results revealed no abnormal copy numbers, ruling out whole gene/exon deletions.

3.5.3 Pakistani families with mutations in TYR and OCA2

Sequencing identified three novel variants in the *OCA2* gene (p.S820P, p.R588W, p.R137I across three families) and a single novel variant in the *TYR* gene (p.W80C in two families). Three

further known *TYR* mutations were identified in three families and three OCA2 mutations were identified in four of the families (*Table 12*). Co-segregation was consistent with causality throughout, with each affected individual harbouring a causal genotype according to a recessive pattern of inheritance. The data is published in a paper under the title 'Mutations in TYR and OCA2 associated with oculocutaneous albinism in Pakistani families' (Appendix B.1)

Table 12. Mutations in TYR and OCA2 associated with oculocutaneous albinism in Pakistani families. The NCBI accession numbers: *TYR* NM_000372 and *OCA2* NM_000275. *indicates novel mutations.

Family	Gene	Nucleotide variant	Protein variant	Allelic status
Family 1	*TYR*	c.G240GC	p.W80C*	Homozygous
Family 2	*TYR*	c.G240GC	p.W80C*	Homozygous
Family 3	*TYR*	c.C649T	p.R217W	Homozygous
Family 4	*TYR*	c.G1255A	p.G419R	Homozygous
Family 5	*TYR*	c.C832T	p.R278X	Homozygous
Family 6	*TYR*	c.T132A	p.S44R	Homozygous
Family 7	*OCA2*	c.1045-15T>G	Splicing mutation	Homozygous
Family 8	*OCA2*	c.1045-15T>G	Splicing mutation	Homozygous, Heterozygous in one affected individual
Family 9	*OCA2*	c.T2458C	p.S820P*	Homozygous
Family 10	*OCA2*	c.C2020G	p.L674V	Heterozygous
	OCA2	c.408_409delTT	p.R137fs*	Heterozygous
Family 11	*OCA2*	c.G1327A	p.V443I	Heterozygous
	OCA2	c.C1762T	p.R588W*	Heterozygous

3.5.4 Segregation of the OCA1B tri-allelic genotype

Six probands were found to have a single heterozygous *TYR* variant. A *TYR* variant was found in two further samples, however NG270 also has a variant in the *PAX6* gene predicted to be pathogenic by computational scores, and NG213 has two compound heterozygous mutations within OCA2. Additionally, neither of these samples were found to have both S192Y and R402Q in the NGS data. The single *TYR* variants were Sanger sequenced in both probands and family members (families #250, #251, #257, #263, #309, and #356 from *Table 11*). In total, twenty probands and family members were phenotyped and genotyped (*Table 13*). The phenotyping results of these six families suggests a total nine cases of partial albinism (six probands and three affected family members). Sanger sequencing confirmed the predicted causal variants in probands and revealed variants segregated with affected family members in every case, with three unaffected family members as carriers.

To explore the apparent missing heritability in these cases the potential pathogenicity of common variants R402Q and S192Y was investigated. The NGS data was examined in probands with *TYR* mutations. All six probands were found to have both variants. The common variants were confirmed in probands with Sanger sequencing and variant segregation was determined across available members of the six pedigrees (*Figure 24*). The combined presence of both common polymorphisms and a putative *TYR* mutation segregates with affected family members. This is true for all cases, except the mother of NG356, who does not have the same R422W mutation as the son (*Figure 23*). It is possible that either the mother has a different causal variant not inherited by her son, or that we are missing the true causal variant.

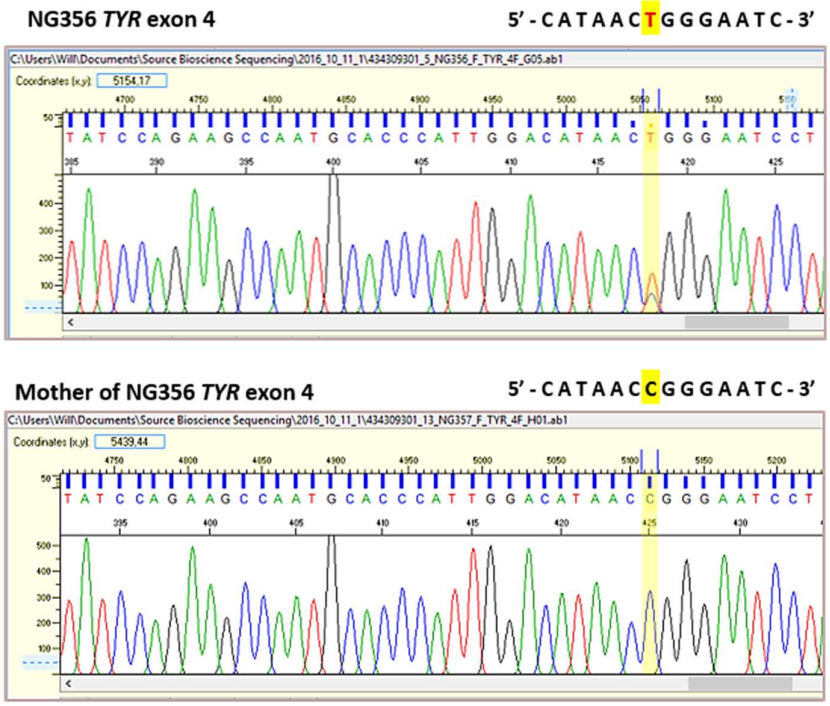


Figure 23. Forward read of *TYR* exon 4 variant p.R422W Sanger sequencing in proband NG356 (top) and mother of NG356 (bottom). The variant R422W is a result of the non-synonymous base substitution c.C1264T highlighted in yellow, carried by NG356 but not by mother.

Table 13. Phenotype-genotype table of families with Sanger-confirmed *TYR* variants. Phenotype information (from left to right): cutaneous and hair pigmentation in context of family background, presence of nystagmus, foveal *hypoplasia* (FH), iris trans-illumination, and VEP asymmetry indicating (over) crossing of the optic nerve. NP = not possible, too young. *TYR* accession number NM_000372.

ID	Relation to proband	Abnormal pigment	Nystagmus	OCT	Trans-illumination	VEP	Genotype		
							Variant 1	R402Q	S192Y
#250	Proband	Yes - OCA1A	No	FH	No	Crossed	*TYR*:c.1467dupT:p.T489fs	Het	Het
	Father	No	No	Normal	No	-	*TYR*:c.1467dupT:p.T489fs	WT	WT
	Mother	No	No	Normal	No	-	WT	Het	Hom
	Sister	No	No	Normal	No	-	WT	WT	Het
#251	Proband	Yes	No	FH	Yes	Abnormal	*TYR*:c.505_507del:p.169_169del	Het	Het
	Mother	No	No	-	-	-	WT	Het	Het
	Father	No	No	-	-	-	TYR:c.505_507del:p.169_169del	WT	Het
#257	Proband	Yes - OCA1A	No	FH	No	Normal	*TYR*:c.732_733del:p.244_245del	Het	Het
	Mother	No	No	FH	Yes	-	*TYR*:c.732_733del:p.244_245del	Het	Het
	Father	No	No	Normal	No	-	WT	Het	Het
	Sister	No	No	-	-	-	WT	Het	Het

ID	Relation to proband	Abnormal pigment	Nystagmus	OCT	Trans-illumination	VEP	Genotype		
							Variant 1	R402Q	S192Y
#263	Proband	Yes	Yes	-	Yes	-	*TYR*:c.C1204T:p.R402X	Het	Het
	Sister	Yes	Yes	-	Yes	-	*TYR*:c.C1204T:p.R402X	Het	Het
	Mother	No	No	-	-	-	*TYR*:c.C1204T:p.R402X	WT	Het
	Father	No	No	-	-	-	WT	Het	Het
#309	Proband	No	Yes	FH	No	Crossed	*TYR*:c.C1217T:p.P406L	Het	Het
	Mother	No	No	-	-	-	*TYR*:c.C1217T:p.P406L	WT	Het
	Grandmother	No	No	-	-	-	WT	Het	WT
#356	Proband	Yes	Yes	FH	Mild	Inconclusive	*TYR*:c.C1264T:p.R422W	Het	Het
	Mother	No	Yes	FH	No	-	WT	Het	Het

Some probands had the typical physical appearance of OCA1A but their ocular phenotyping revealed they are missing some of ocular features you would expect in 'full albinism' (*Figure 24*). The mother of NG257 has no obvious albinoid features, i.e. hypopigmentation of skin and hair or nystagmus, but does have two ocular features that indicate albinism which may have been missed without detailed phenotyping. Both NG356 and mother have nystagmus and foveal hypoplasia, however NG356 differs by also exhibiting some hypopigmentation. Furthermore, the mother does not have the same genotype as her son as she is not carrying the p.R422W mutation in *TYR*. All probands and family members with foveal hypoplasia have the tri-allelic genotype, further confirming OCT is a sensitive method for detecting albinism[149].

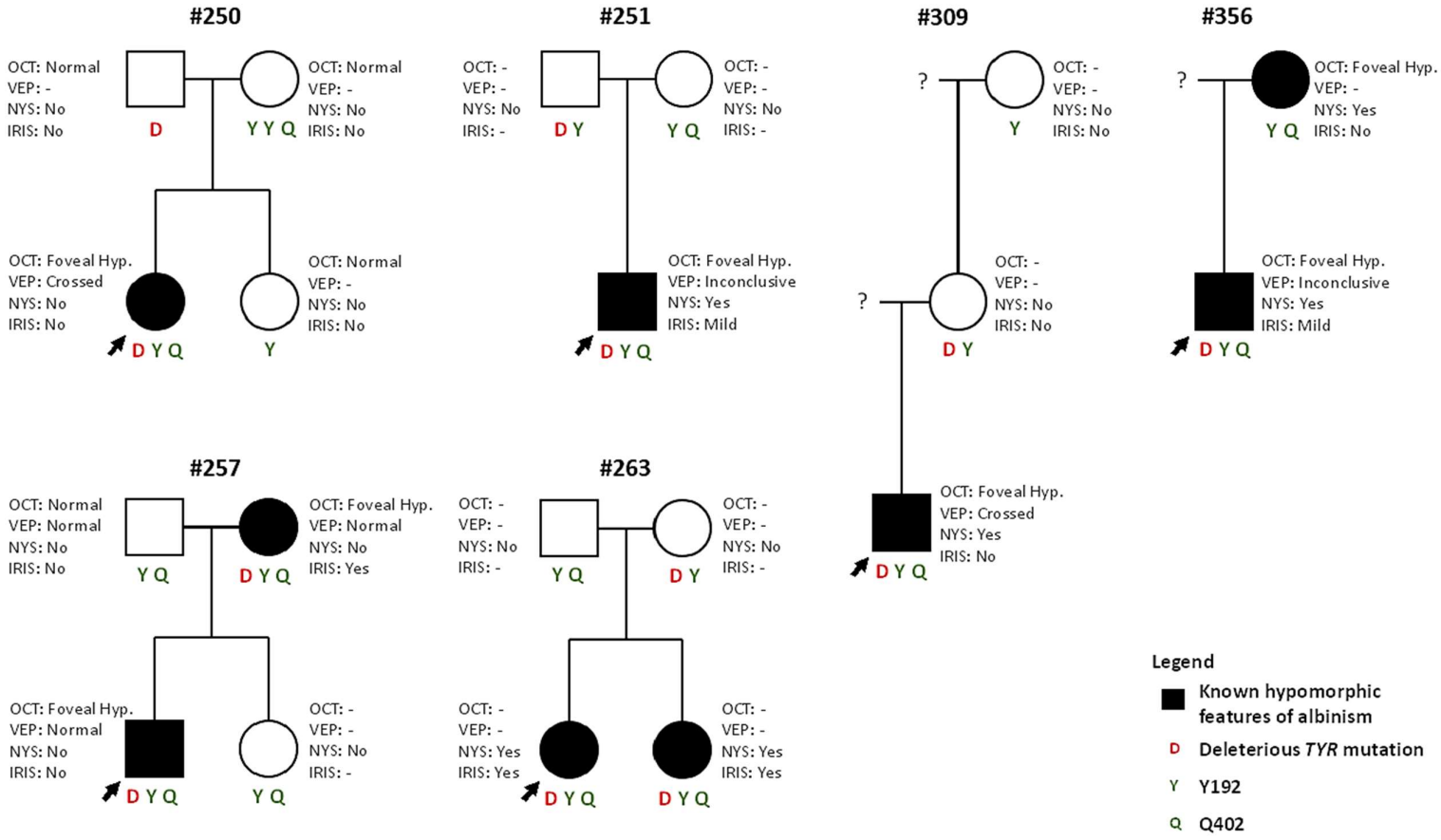


Figure 24. Pedigree diagrams for six families with a single *TYR* pathogenic mutation and common polymorphism phenotyping. TYR variants are listed beneath each family. Sanger sequencing was performed on family members as opposed to the full exonic region sequenced in probands. Family number corresponds with proband number.

3.6 Discussion

High resolution phenotyping, a targeted NGS panel, MLPA, and segregation analysis have all been utilised to study a cohort of nystagmus patients fitting the phenotype of 'partial albinism' for the first time. This in depth study provides the opportunity for a detailed assessment of genotype and phenotype which has not yet been carried out in this subgroup of patients. One important point to note is that some probands with a genotype proving albinism, may have been excluded if the misnomer 'normal optic nerve routing excludes albinism' was strictly followed [150].

The TruSight One (TSO) clinical exome has identified a novel variant in the *PAX6* gene, a novel frameshift variant in the *GPR143* gene, two novel variants in the *OCA2* gene, five previously reported variants in *OCA2*, seven previously reported variants in the *TYR* gene, and one previously reported variant in TYRP1 in eighteen probands. When combined, these variants provide a convincing genetic diagnosis for only 22% of the original hypomorphic albinism cohort when excluding those with only a single heterozygous variant for a recessive condition (i.e. OCA1). In comparison, the Pakistani families have complete causal genotypes which may be due to the greater consanguinity of the population, as well as, the complete OCA phenotype exhibited by each affected member.

The novel *PAX6* mutation p.Q422K is in the same location as a previously reported mutation p.Q422R [151]. This *PAX6* mutant was shown to cause Aniridia but the authors concluded that this C-terminus mutation may only have a subtle effect on the protein as the resulting phenotype was mild [151]. Proband NG270 does not have any iris abnormality but it is clear that background genetics can influence the ocular phenotype, which in this instance has resulted in nystagmus and chiasmal misrouting. Interestingly, NG270 was also diagnosed with autism, a disorder with which *PAX6* has strong links due to its crucial role in brain development [152].

The novel variant in *GPR143*, c.485delG, causes a frameshift mutation likely causing the ocular albinism (OA1) phenotype in NG272. Of the six different mutations found in *OCA2*; p.N465D [138], p.V443I [153], p.Y342C [154] and p.L674V [155] are all reported as disease causing in the HGMD database. The variant p.V443I is a mutation seen quite frequently in cases of OCA2 yet is considered a conservative mutation due to the high similarity between valine and isoleucine suggested to affect its penetrance [153, 156, 157].

The *OCA2* variants p.R536C and p.W274C have not been reported before but both variants predicted to be deleterious by SIFT, PolyPhen2 and GERP++ (*Table 11*). Probands NG213 and NG340 each carry two variants in *OCA2* so their molecular diagnosis can be referred back to the clinician. Probands NG322 and NG344 both have a single *OCA2* variant so require more investigation. A brief investigation through lowering of the filtering parameters (removing SIFT) allows a second *OCA2* variant, p.A339E (Sift = 0.07), to pass the threshold in NG322 which could account for the initial missing heritability in this proband. NG265 is the final proband to have a single *OCA2* variant, but this sample also carries a single putative mutation in *TYRP1*. The synergistic relationship between melanogenic genes may lead to mutations across different genes having an additive effect. Chiang *et al*. previously suggested a synergistic relationship between OCA2 and OCA3, where a single heterozygous mutation in each of the responsible genes altered the overall albinism phenotype. The report was only within a single family and still requires greater proof before the concept can be widely accepted [143].

Within the Pakistani cohort there is a much higher level of homozygosity due to the higher consanguinity sometimes seen within this population. There are four novel mutations identified, one in the *TYR* and three in the *OCA2* gene. The *TYR* variant p.W80C has not been reported before but it is in the same location of two previously reported mutations (W80R and W80X[158, 159]) and has been predicted as damaging by pathogenicity scores. The variants p.S820P, p.R137fs, p.R588W are novel, though each variant is predicted to be damaging by pathogenicity scores and segregation data within the large family pedigrees suggests each variant is pathogenic (Appendix B.1).

3.6.1 The causal role of common population variants

Seven different mutations were found within the *TYR* gene: p.V177F, c.1467dup, c.505_507del, p.C244X, p.R422W, p.R402X and p.P406L. The variant V177F has been previously reported as a pathogenic mutation in an albinism cohort[160]. *TYR* c.1467dup results in a frameshift and has been reported as a causal mutation multiple times [57, 74, 138, 160]. R402X has been reported previously and creates a premature stop codon, considered highly deleterious[74, 160, 161]. The mutation P406L has also been reported many times before in association with albinism [138, 160], and it has been shown to reduce enzyme activity to 35% [63]. R422W has been reported as disease causing [138], however functional studies of this mutation have conflicting results. Mondal et al. assayed the tyrosine hydroxylase and DOPA oxidase activity of the R422W

mutant and found that the enzyme retained no activity[62], whereas, Dolinska et al. assessed only DOPA oxidase activity and found that the R422W mutant retained 95% of wild-type enzyme activity. Dolinska et al. also state that R422W is temperature sensitive and the immature glycoprotein is degraded more quickly at 37°C[63], potentially accounting for the difference between assays. Reported literature ascribes many variants as disease causing throughout the coding regions of both *OCA2* and *TYR*, however recent functional studies have questioned the deleterious effect of some of these variants, particularly in the *TYR* gene [62, 63]. There is currently no functional evidence of the deleterious effect of the mutations *TYR* c.505_507del and TYR C244X though the deletions have previously been reported as causal mutations, and the introduction of a premature stop codon is considered highly deleterious[162, 163]. The dispute between pathogenicity of variants has highlighted the necessity of further functional analyses to produce a curated list of mutations for accurate genetic diagnosis[156].

Six probands were found to have single *TYR* variant, but no variant in another known gene. As there is no functional evidence for the variants in NG309 and NG356 there remains the possibility of another causal gene mutation. It has been suggested that this high level of missing heritability could be due to mutations in the *TYR* promoter or an interacting distal gene enhancer [164]. Notably, all six had also inherited R402Q and S192Y common *TYR* SNPs, producing a tri-allelic genotype. A similar tri-allelic hypothesis has been demonstrated in Bardet-Biedl syndrome (BBS) which was originally thought to be a traditional autosomal recessive disorder. However, the tri-allelic genotype described in BBS is across two genes with one carrying two mutations that are not causal without the third mutation in a second gene[165]. Such a complex genotype is yet to be demonstrated in albinism. It is not possible to determine whether R402Q and S192Y are in *cis* (i.e. on the same *TYR* allele) or in *trans* (i.e. separate *TYR* alleles) for all of these pedigrees. However, R402Q:S192Y in *cis* may provide the best explanation for the apparent additive effect, providing a pathogenic haplotype.

The SNP R402Q is located in exon 4, near to the CuB catalytic site, and there is evidence that the SNP results in a thermolabile enzyme, but it has been argued that the reduction of tyrosinase activity is not enough to produce a phenotype[16,146]. The controversy over the R402Q variant stems from a paper by Oetting *et al.* (2009) which argues that segregation of R402Q with a known pathogenic variant on the homologous allele does not confer albinism [148].

The SNP S192Y is located in the CuA catalytic site of tyrosinase and has been shown to lower enzymatic activity in an independent manner to R402Q [60]. Previous studies have had stringent criteria for an OCA1 phenotype (white hair and skin and translucent irides from birth)

[148], whereas here, we have considered hypomorphic presentations that do not appear as severe but result in ocular deficits nonetheless. Here we suggest that a combination of a pathogenic mutation inherited with both variants in a tri-allelic genotype may cause a large enough reduction in tyrosinase activity for a partial OCA1 phenotype. The background work also suggests the tri-allelic phenotype is not always restricted to partial (OCA1B) e.g. NG309, who appears to have normal skin pigment and no iris transillumination, and may extent to full (OCA1A) when in the context of background mutations e.g. NG222, who has a full OCA1 phenotype and Noonan's syndrome.

AROA, as a suggested subtype within OCA1, is not an appropriate diagnosis for probands in this cohort as cutaneous and hair pigment is noticeably decreased in most probands and many family members, and there is a great variation in ocular phenotype. It is likely that the background level of pigmentation may determine the severity of the mutations as lower pigment levels will be affected more severely by the same dosage loss of tyrosinase. Therefore, these results support the theory of a causal tri-allelic genotype and may go some way to account for a proportion of the of OCA1 cases with apparent missing heritability. Functional studies would assist in confirming pathogenicity, thus allowing the tri-allelic genotype to be considered for both future and retrospective genetic diagnosis of OCA1.

If the proposed tri-allelic genotype hypothesis is correct, this would increase the diagnostic yield of genetic testing from 22% as described earlier, to 56% in the cohort. Given that hypomorphic albinism is a difficult cohort to diagnose clinically and genetically, evidenced here by NG265 carrying mutations for two separate OCA subtypes, further exome-sequencing is a suitable method for the genetic diagnosis. A sequencing technique with broad capture allows for the pickup of genetic variants, which may be causing abnormal ocular phenotype.

Since publication of these findings under the title 'Identification of a functionally significant tri-allelic genotype in the Tyrosinase gene (*TYR*) causing hypomorphic oculocutaneous albinism (OCA1B)' [166], there have been two further studies taking the common SNPs into consideration. Lasseaux *et al.* screened known albinism genes in 990 patients and considered the phenotype of patients with an R402Q–OCA1 genotype [167]. They describe a milder phenotype with less severe foveal hypoplasia and ocular/skin hypopigmentation. The SNP S192Y was found in *cis* with R402Q in 21.9% of the alleles, a number much higher than the estimated frequency in the general population. Lasseaux *et al.* conclude it is likely that there is significant involvement of the 192Y-402Q haplotype in OCA1.

Monfermé and colleagues investigated the link between R402Q and mild albinism phenotypes (equivalent to the partial albinism phenotype described in this chapter). They conclude that the subtly of features and ability to accumulate pigment in these cases provides clear risk of underdiagnosis or belated diagnosis. It may be that the supposed negative proof provided by Oetting *et al.* is simply due to a lack of detailed phenotyping. They suggest foveal hypoplasia is the most sensitive feature to diagnose by, which appears to agree with the results in this cohort. Monfermé et al. also consider the additive effect of S192Y and found the variant to be present in the majority of their R402Q-OCA1 cases, though not in all, and of these just over 50% are in a *cis* p.[R402Q:S192Y] haplotype. They do not suggest to prove or disprove the tri-allelic genotype theory and have not considered the phenotypic cross-over with mutations in the *PAX6* gene and have therefore not sequenced this gene. However, the Monfermé group suggest that the single R402Q SNP in *trans* with a *TYR* null mutation can cause partial OCA1B.

There is no current treatment for the underlying molecular anomaly in albinism and present treatments are supportive. Therapeutics are under development but an effective treatment for any of the underlying molecular defects has not yet reached clinical practice. This work and that of others appears to suggest that small variations in melanin biosynthesis between related family members dictate the extent of the phenotype in OCA pedigrees. Furthermore, the net loss of TYR function (caused by cumulative effects of multiple variants, each of which reduce TYR function by differing amounts), appear to result in a continuum of clinical features. This work supports the assertion that small modulations in components of the melanin biosynthesis pathways, through therapeutic means, may be sufficient to rescue some of the visual disability seen in patients with albinism phenotypes.

3.6.2 Conclusion

The TSO clinical exome suitably covers exonic regions of the OCA genes of interest, as well as splicing sites. However, it has been suggested that variants in the upstream promoter region could alter the expression of *TYR* so the TSO panel could be improved to cover this region too. *SLC38A8,* a gene not captured by TSO, has been reported as the cause of some cases of isolated foveal hypoplasia and nystagmus, thus an improved clinical exome would cover this gene. A so called 'spike-in' has recently been used to improve gene coverage in custom designed clinical NGS [168]. Farooqi *et al.* tested a spike-in for a SeqCap EZ panel (Roche Nimblegen) and demonstrated improvements to coverage by addition of oligo baits during the capture process. It may be possible to utilise this technique to increase coverage of the whole *TYR* gene (promoter included) as well as add in genes that are currently missing.

Currently, a single variant is considered benign if the MAF is greater than 5% [110]. These findings suggest standards and guidelines could be revised to consider the combined impact of variants, particularly for more complex disorders such as albinism. So far, *TYR* has been investigated but this work could be extended to *OCA2* as the second most common cause of albinism in European populations with a similar level of missing heritability.

Furthermore, the diagnosis of albinism currently focusses on compound mutations in single genes without considering the potential for synergistic relationships between functionally related genes such as that previously suggested for OCA2 and OCA3 genes (*OCA2* and *TYRP1*) [143] and for which there is potentially one example within this hypomorphic cohort. These findings suggest that the inheritance pattern of OCA may not be as simplistic as 'autosomal recessive'.

Chapter 4: A functional study of common population variants in the Tyrosinase gene (*TYR*)

4.1 Introduction

4.1.1 Structure and function of tyrosinase

Tyrosinase is a copper-dependent monooxygenase enzyme that catalyses multiple steps within the melanin biosynthesis pathway. Tyrosinase has both monophenolase and diphenol oxidase activities and is crucial for catalysing the initial rate limiting steps: hydroxylation of L-tyrosine to L-DOPA (L-3,4-dihydroxyphenylalanine) and the oxidation of L-DOPA to dopaquinone (*Figure 25*). Damaging mutations within the tyrosinase gene, *TYR,* result in Oculocutaneous albinism type 1 (OCA1) which is split into subtypes OCA1A and OCA1B depending upon whether or not any tyrosinase function remains. OCA1A is the most severe form of non-syndromic albinism with a complete loss of pigmentation and severe ocular defects, suggesting the tyrosinase enzyme is one of the most important proteins in melanin biosynthesis. Mutations within the *TYR* gene contribute to population differences in pigment and some mutations are considered risk factors for skin cancer [169, 170]

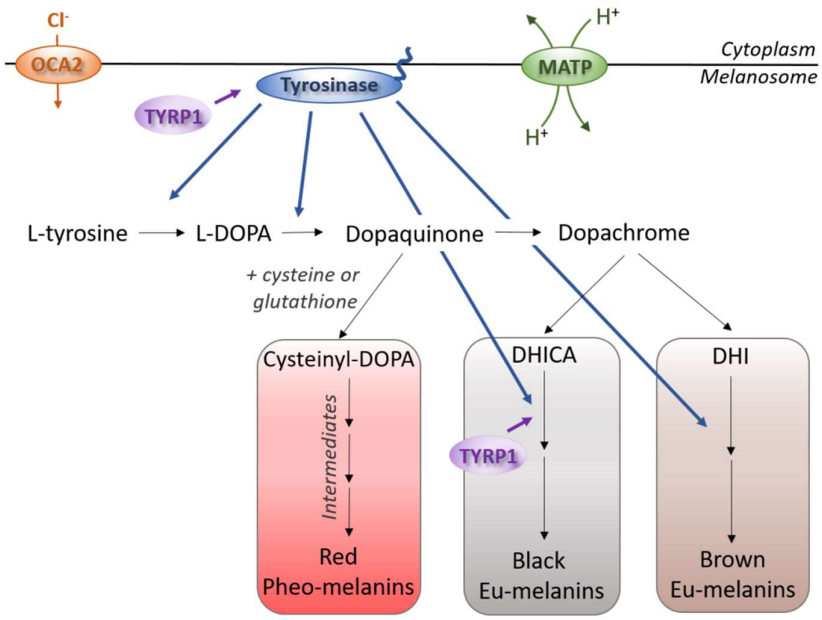


Figure 25. Melanogenesis within the melanosome. The proteins associated with OCA1-4 (tyrosinase, OCA2/P-protein, TRP1/TYRP1 and MATP/SLC45A2 respectively) are included. Tyrosinase (blue) is important throughout the pathway, particularly for the first two rate limiting steps: hydroxylation of L-tyrosine and oxidation of L-DOPA. OCA2 is thought to transport anions such as Cl^- across the membrane to maintain electrogenic neutrality and support the acidic melanosome environment [171]. TYRP1 catalyses the oxidation of 5,6-dihydroxyindole-2-carboxylic acid (DHICA) to a carboxylated indolequinone in the black eumelanin pathway, as well as, directly associating with tyrosinase and possibility stabilising its structure[84, 172]. MATP is a membrane transport protein that regulates melanosomal pH and influences tyrosinase activity [85].

The tyrosinase enzyme is anchored in the melanosome membrane, with a large internal domain, a transmembrane domain, and a small cytoplasmic domain. There is no crystal structure of human tyrosinase but the homology of bacterial tyrosinase has been used to computationally model human tyrosinase (*Figure 26*) [173]. Tyrosinase contains two copper binding sites, CuA and CuB. Each of the two sites each contain three histidine residues surrounding the bound copper and are known to be important for the catalytic activity of the enzyme [174]. The protein also has a C-terminus transmembrane domain followed by small chain extending into the cytoplasm.

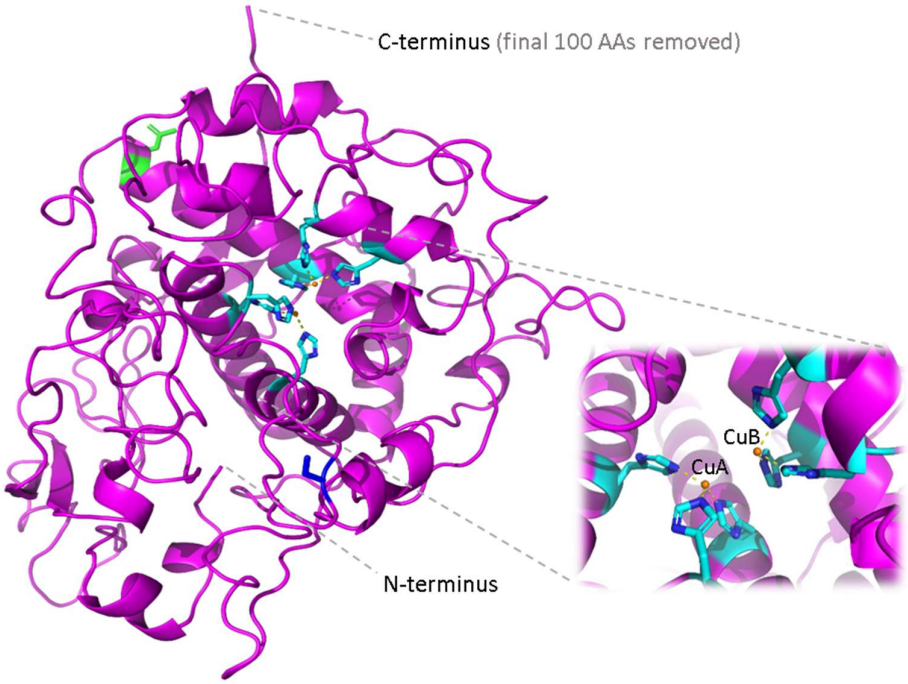


Figure 26. A homology model of the tyrosinase protein showing the CuA and CuB binding sites thought to be important for activity. The tyrosinase protein was modelled from the closely related TYRP1 protein (PDB: 5M8L)[175] using the Modeller program[176] and figure was produced in open source PyMOL.

The expression and production of tyrosinase is crucial for melanogenesis. A number of factors including age, ethnicity, and UVR exposure each determine the amount of melanin produced. The initial stimulation, e.g. by UVR exposure in skin, is followed by signalling systems and transcription factors within the melanocyte, including MC1R, MITF, the tyrosine kinase receptor KIT, KIT's ligand SCF. The activation of these signalling pathways upregulates the expression of melanogenic genes, such as *TYR*[177]. Once translated, the immature protein makes its way through the ER to the trans-golgi network, undergoing signal cleavage, folding and glycosylation. Tyrosinase is a highly processed protein with seven known N-linked glycosylation sites important for enzymatic function[178]. Glucosidase enzymes trim the oligosaccharide sugars and release tyrosinase from chaperone proteins such as calnexin and calreticulin[179]. If the protein is not released from the ER, it undergoes ER-associated protein degradation (ERAD), a process known to occur with many OCA1 mutants[62, 147, 180]. Tyrosinase is then trafficked to the melanosome where it is anchored into the melanosomal membrane by its C-terminus transmembrane domain. It has been suggested that the trafficking of tyrosinase to the melanosome is further enabled by association of TYRP1[172]. The protein is thought to be

activated by PKC-β phosphorylation of two serine residues within the cytoplasmic tail[181](*Figure 27*).

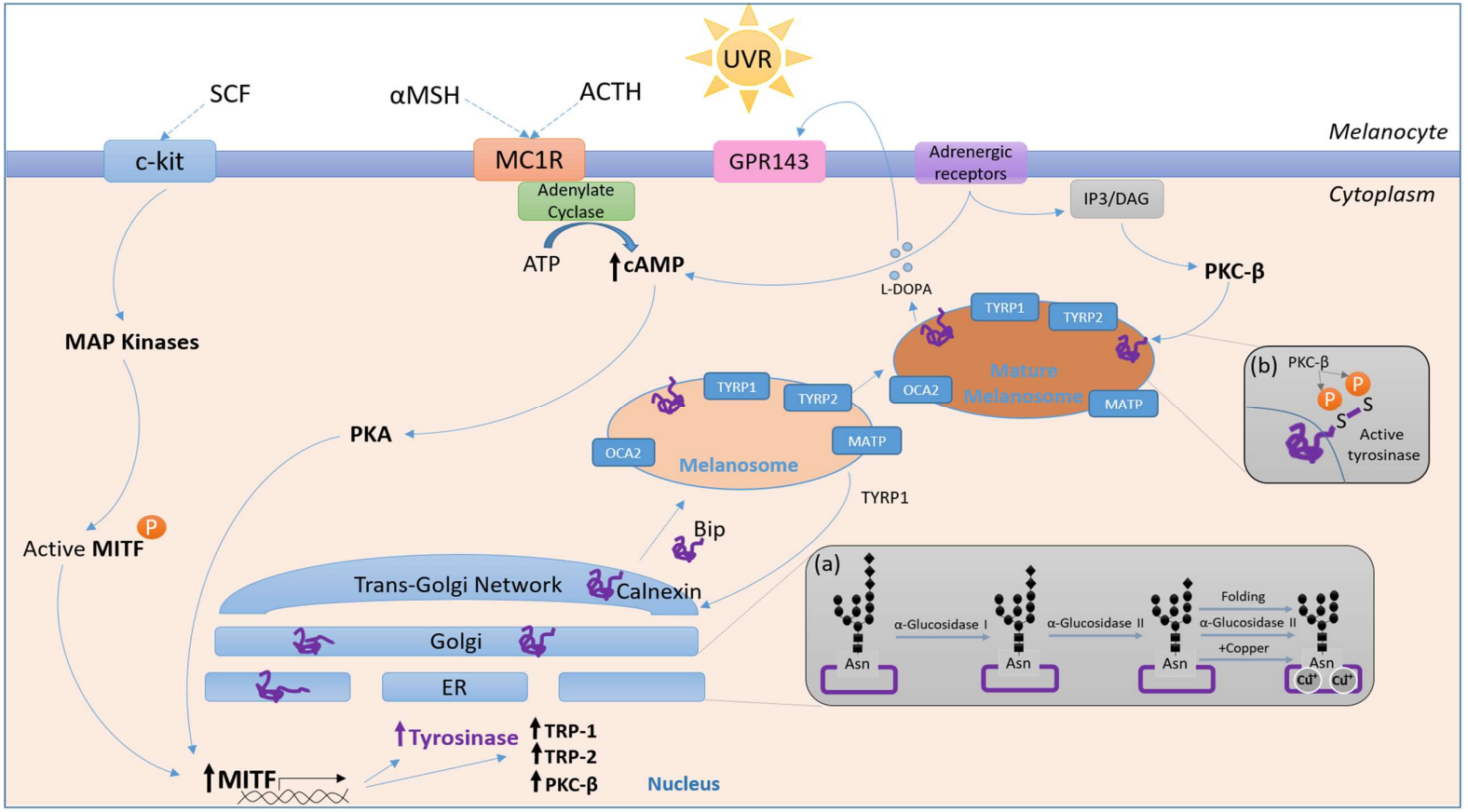


Figure 27. Tyrosinase production and processing within the melanocyte. UVR exposure activates melanocyte membrane proteins, stimulating downstream signalling pathways to upregulate the melanogenesis transcription factor MITF. MITF increases transcription of *TYR* and other melanogenic genes. Once translated, the immature tyrosinase moves through the ER and the Golgi network undergoing posttranslational modifications to produce a mature protein. Tyrosinase is trafficked to the early melanosome where its C-terminus anchors it in the melanosome membrane. The melanosome matures and tyrosinase is further activated by phosphorylation of two serine residues within the C-terminus cytoplasmic tail.

How defective melanogenesis effects the development of the eye has not yet been fully explored, however, there appears to be a connection between tyrosinase and the development of the neural retina. The expression of tyrosinase during neuroblast divisions has been shown to affect pathfinding by retinal ganglion cells in mice and expression of wild-type tyrosinase in albino mice was able to rescue the abnormal chiasmatic projections [182],[183]. Is it thought that the melanogenesis intermediate DOPA plays a role in cell to cell signalling during neural development. L-DOPA is known to be an agonist for the GPR143 protein which causes OA1 when mutated. Binding of L-DOPA causes an immediate increase in intracellular calcium by activating this G protein-coupled receptor [87, 184]. The *GPR143* gene is expressed within both melanocytes and RPE but its ligand L-DOPA has been suggested to have a paracrine effect on the retina [185].

4.1.2 Genetic variation within the Tyrosinase gene (*TYR)*

Tyrosinase is encoded by the *TYR* gene on chromosome 11. There are 321 known pathogenic mutations within tyrosinase according the HGMD [HGMD-PUBLIC, Dec 2017]. These mutations are uniformly spread across the gene with no specific areas seemingly worse affected; it is only the C-terminus transmembrane domain that does not carry many damaging mutations (*Figure 28*). Whether a mutation in the *TYR* gene is defined as an OCA1A or OCA1B mutation is dependent upon the phenotype of the patient. However, this categorisation can be quite arbitrary due to the heterogeneity of phenotype between patients. Functional data provides strong evidence of a variant's pathogenicity; however, some previous functional studies of the same variants provide contradictory results [62, 186].

The *TYR* R402Q and S192Y variants are common polymorphisms that were initially considered benign but were seemingly enriched in OCA cohorts. For example, Kalahroudi *et al.* sequenced the *TYR* gene in Iranian patients with OCA1 and found R402Q and S192Y in 17.5% and 35% of the patients respectively, (the expected prevalence according to 1000 genomes is between 0-5% and 0-6% in Asian populations) [133]. The S192Y SNP is located adjacent to the CuA catalytic site and the R402Q SNP is located close to the CuB catalytic site (*Figure 28*). Evidence for a tri-allelic genotype containing both SNPs in combination with a rare TYR variant suggests these mutations could have an additive effect [166].

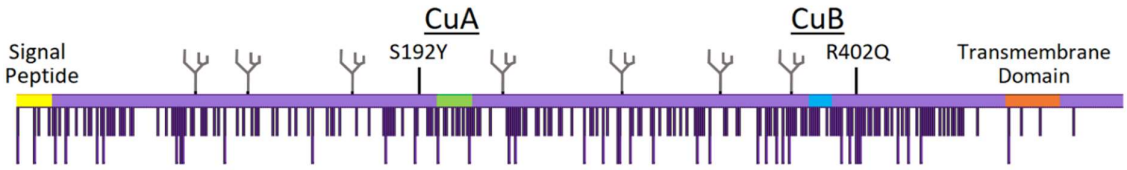


Figure 28. Map of OCA1 mutations throughout the tyrosinase. The protein length is 529 amino acids with signal peptide up residue 18, CuA site from residue 202 to 219, CuB site from residue 383 to 394, and transmembrane domain over residues 474 to 500 (UniProt, L8B082_HUMAN, Nov 2018). The numbering of tyrosinase starts at the signal peptide Met 1 even though this section is cleaved once the protein is translocated to the ER [179]. The mutations are shown as purple lines (HGMD-PUBLIC, Dec 2017). Lines of double height represent two different mutations at the same site. Tyrosinase is known to have seven glycosylation spread sites across its intramelanosomal domain[178].

4.1.3 Previous functional studies of tyrosinase variants

Various research groups have investigated the functional effect of variants on the *TYR* encoded protein tyrosinase. Different approaches in both expressing the protein and characterising its enzymatic activity have been taken. The findings from these studies can be summarised as follows.

Mondal *et al.* investigated 34 missense *TYR* variants from the albinism database established by Oetting 2009 (www.ifpcs.org/albinism/). They transfected HEK293F cells with the variants in a pEGFP-TYR-WT vector, and used site-directed mutagenesis to insert base changes into the *TYR* cDNA. Endoplasmic reticulum retention was studied using an endo H assay. Hydrolysis of tyrosine was studied using a scintillation counter to measure the release of tritiated water (^{3}H2O) from [^{3}H]-tyrosine. DOPA oxidase activity was determined through spectrophotometer measurements of absorbance change as L-DOPA was oxidised to DOPA-quinone. These assays showed that five out of eight OCA1A variants completely eliminated enzymatic activity, two variants had residual activity, whilst one retained enzymatic function equivalent to that of the wild-type tyrosinase. For OCA1B, ten of 14 variants had no activity, two variants had residual activity, and two had the same activity as wild-type protein. The tyrosinase common variants investigated were found to partially reduce the function of tyrosinase, retaining 30-60% activity. Three were completely null and one remained the same as wild-type. All the variants studied were mapped to positions within the protein and there was no clear relationship between structural changes and decrease in specific function.

Dolinska *et al.* (2014) investigated two *TYR* variants, R422Q and R422W, expressing the protein in insect cells (*Trichoplusia ni* Larvae) with the final 80 residues of the trans-membrane C-terminus removed[61]. A colourimetric assay with L-DOPA as the substrate was used to

determine the conditions for maximum protein activity and followed by the use of a spectrophotometer to determine DOPA-chrome formation on addition of either L-tyrosine or L-DOPA. R422Q and R422W both had lowered enzymatic activity when grown and expressed at 37˚C. This activity was recovered when the variant protein was synthesised at 31˚C. In 2017, Dolinska *et al.* went on to study four more mutants, as well as revisiting R422Q and R422W[186]. The same expression and enzyme activity methods were used, with the addition of circular dichroism and measurements of fluorescence emission to determine changes to secondary structure and efficiency of folding respectively. Their findings revealed that OCA1A-related mutants T373K and R77Q were unstable and had no enzymatic activity, whilst OCA1B-related mutants R402Q, R422W, R422Q, and P406L were biochemically similar to the wild-type, yet, showed decreased enzymatic activity and protein stability.

Chaki *et al.* functionally validated six *TYR* mutations identified in an Indian cohort with OCA1. All the mutants were found to be endo H sensitive, further confirming OCA1 as largely a disorder of ER retention. The group went on to investigate the S192Y SNP which has a biased distribution in European populations but a very low occurrence in the Indian population. Sequencing 1,600 healthy volunteers from diverse Indian populations, they found Y192 was over represented among the indo-europeans, suggesting an association with reduced pigmentation. Functional assessment of the tyrosine hydroxylase and DOPA-oxidase activities of the S192Y-mutant revealed a 40% reduction in efficiency for both activities [59].

Jagirdar *et al.* employed microscopy to compare the melanogenic activities of cultured melanocytes carrying each of the common *TYR* SNPs, R402Q and S192Y as either homozygous or heterozygous variants[60]. They also assayed tyrosinase DOPA-oxidise activity, glycosylation and temperature sensitivity. Homozygous 402Q/Q melanocytes yielded significantly less protein, showed altered glycosylation, and had reduced DOPA oxidase activity. However, activity could be recovered at a lower growth temperature of 31˚C to near wild-type *TYR* levels. They also described a decrease in expression of protein when homozygous for 192Y/Y, though it was not statistically significant. The 192Y/Y DOPA oxidise activity was significantly reduced but surprisingly not as much as that of the heterozygous 192S/Y genotype. Jagirdar *et al*.'s findings suggest that both common SNPs are likely important modifiers of general population pigmentation, but also their high population frequency suggests they should be investigated in combination with each other with a potential role in albinism.

These previous functional studies have not included a full assessment of the tri-allelic theory described in Chapter 3 due to the assumption that S192Y is a benign SNP commonly occurring in the Caucasian population. The studies described have used a range of methods to

produce tyrosinase and quantify its activity. The contradictory results question which method best reflects the true loss of tyrosinase function and how this translates to human phenotypes. A reliable and robust method is needed for the assessment of tyrosinase variants so that validated mutations can be used to confidently provide patients with a genetic diagnosis.

4.2 Aim

The aim of this study was to develop a robust assay for the functional assessment of recombinant tyrosinase mutants. The assay was designed to be simple and accurate to resolve the data already published, as well as provide functional data for the double R402Q:S192Y tyrosinase mutant. The ultimate aim was to create a method of tyrosinase production and analysis that could be manipulated in future studies to assess the efficacy of therapeutics.

4.3 Hypothesis

The hypothesis was that the introduction of the common *TYR* SNPs R402Q and S192Y would have a damaging effect on the tyrosinase protein, effecting both maturity and activity of the enzyme. It was proposed that the most common SNP S192Y would be damaging to a lesser degree with a combination of the two variants on the same allele having an additive effect, thus resulting in an even greater loss of available functional tyrosinase.

4.4 Methods

The methods were developed to create a robust assay for the comparison of tyrosinase mutants to wild-type tyrosinase activity. Previous studies have produced recombinant tyrosinase in cultured melanocytes, insect cells, HeLa cells, and HEK293 cells [59-62]. Human tyrosinase has seven known N-linked glycosylation sites thought to affect enzyme activity and therefore was produced within a human cell line that does not already undergo pigment production[178].

4.4.1 Tyrosinase protein production

The HEK293F expression system was chosen for protein production as it is a non-melanogenic cell line reported to have a high transfection efficiency. HEK293F cells were cultured as described in section 2.6.4.1. A p3XFLAG-TYR plasmid vector was purchased from Addgene. The vector contained the 1.79 Kb *TYR* cDNA encoding a recombinant tyrosinase protein with a 3XFLAG tag at the C-terminus. The vector also encoded genes allowing for growth

and selection in both bacteria and mammalian cells. Site-directed mutagenesis was used to mutate the plasmid vector to produce wild-type cDNA (due to initial presence of S192Y), as well as, the variants of interest (listed in *Table 14*).

Table 14. TYR variants to undergo functional investigation. Whether or not there is previous functional evidence of each mutation and reason for study. *TYR* accession number NM_ 000372

***TYR* variants (exon:cDNA:protein)**	**Functional data**	**Reason for study**
exon1: c.C575A: p.S192Y*	Yes – not in combination with R402Q [60]	Considered a benign SNP, but incidence of S192Y is higher in patients with OCA1. Possibly forms part of tri-allelic inheritance of OCA1B.
exon4:c.G1205A:p.R402Q*	Yes – not in combination with S192Y [60, 186]	Generally considered a benign SNP, but functional evidence suggests the mutation has a negative effect on activity.

*These common variants are potentially only pathogenic when in combination and will be also be studied together.

Once created, the p3XFLAG-TYR wild-type and mutant plasmid vectors were used to transiently transfect HEK293F cells in a 30 ml reaction. Cells were cultured for 72 hours post transfection to allow for adequate protein expression (methods section 2.6.4).

4.4.2 Assay of protein activity

A DOPA-oxidase assay was used to determine the enzymatic activity of the recombinant tyrosinase. This is simple assay relies on the diphenol oxidase activity of tyrosinase, wherein the substrate L-DOPA is oxidised to dopaquinone. There are two methods of measuring this activity, both of which require spectrophotometry to measure a colour change, but they differ in the product they are measuring. The most frequently used assay is to measure accumulation of the downstream product dopachrome as this turns the solution brown [59-61, 63] (methods section 2.6.7).

The second option is to sequester the direct product using 3-methyl-2-benzothiazolinone hydrazine (MBTH), a compound known to bind quinones (methods section 2.6.7). This removes the lag period for spontaneous dopachrome formation and an outside interaction that could affect the unstable intermediates (*Figure 29*). Furthermore, once trapped as a dopaquinone-MBTH adduct, the compound is stable unlike the unstable intermediate dopachrome.

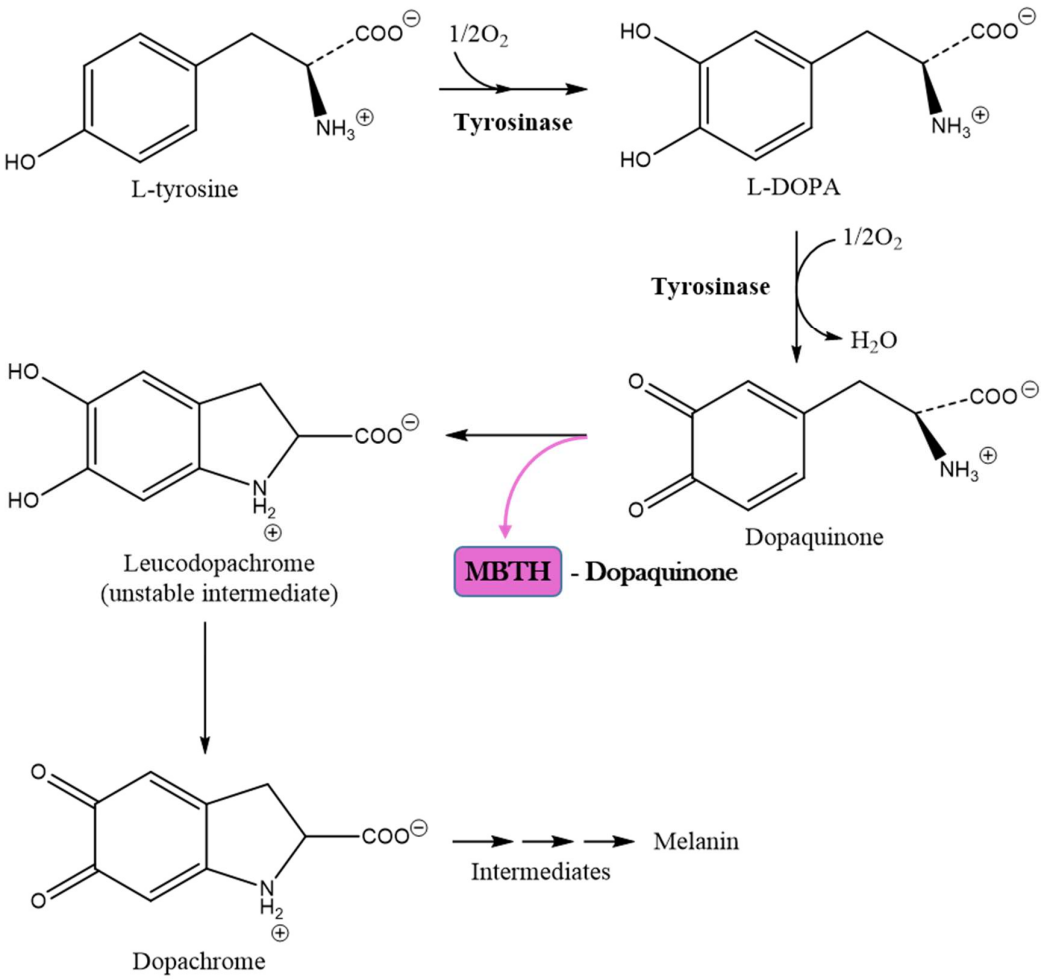


Figure 29. Chemical Mechanism for the initial stage of melanin synthesis. The two rate limiting steps, hydroxylation of L-tyrosine and oxidation of L-DOPA, are catalysed by tyrosinase. The downstream intermediates, leucodopachrome and dopachrome, form spontaneously though, physiologically, some dopaquinone becomes cysteinyl-dopa and is used in the production of pheomelanin (red melanin).

The assays of enzyme activity were carried out at 37˚C and as close to a neutral pH as possible in order to replicate the *in vivo* physiological environment. Even though melanosomes are often described as acidic, the optimal pH for tyrosinase activity is higher as the pH increases to physiological levels during melanosome maturation[179]. The buffer for the MBTH-coupled assay was only taken to a pH of 6.9 as any higher began to push particles out of solution. Both assay formats were tested to determine which would be optimal (*Figure 30*). MBTH causes the reaction to reach maximum saturation very quickly as it is more sensitive to tyrosinase activity, therefore it is important for revealing small amounts of remaining activity. It was thought that the DOPA-oxidase assay without MBTH would be more useful in elucidating any change in activity that was time-dependent. Therefore, both the original DOPA-oxidase assay MBTH-coupled assay were used to investigate the wild-type and mutant tyrosinase. The MBTH-coupled

assay reached maximum saturation very quickly, therefore the initial time points were shortened to five minute intervals and the final protein concentration was reduced.

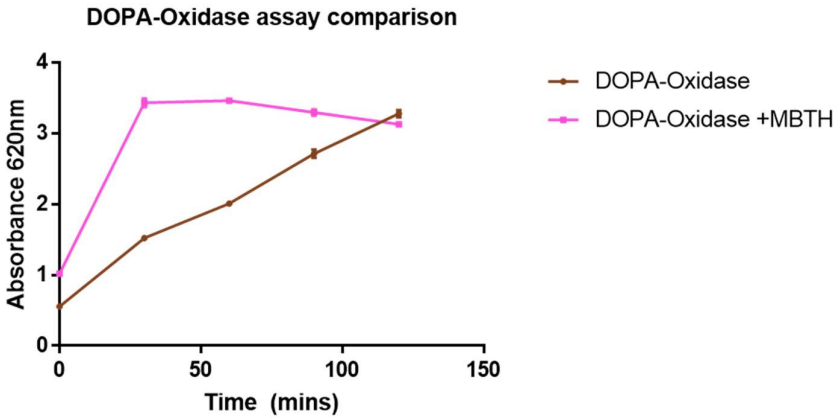


Figure 30. Comparison of DOPA-oxidase assay format, with and without addition of MBTH, to measure tyrosinase enzymatic activity. Whole cell lysate for wild-type tyrosinase transfected HEK293F cells was used. The final protein concentration was 1 µg/ml for both assays. For each point there are two repeats and SEM bars are shown.

To determine the optimal time course for the DOPA-oxidase assays a plate containing wild-type tyrosinase and blank was measured every 30 minutes for 4 hours and for a final time point at 16 hours (*Figure 31*). The DOPA-Oxidase (DOPA-only) assay showed a gradual increase of absorbance over time, never reaching saturation even at the 16 hour time point. However, blank absorbance also increased over time and when wild-type is blank-normalised the graph plateaus by the 180 minute time point. The DOPA-Oxidase +MBTH (MBTH-coupled) assay showed a fast increase in absorbance reaching saturation after one hour of incubation. The blank is more stable and only shows a slight increase overtime. Therefore, the DOPA-only assay should be run for a minimum of 3 hours and the MBTH-coupled assay should be run for at least an hour but can be continued for longer as the buffer is quite stable overtime, however the product appears to break down after 16 hours of incubation.

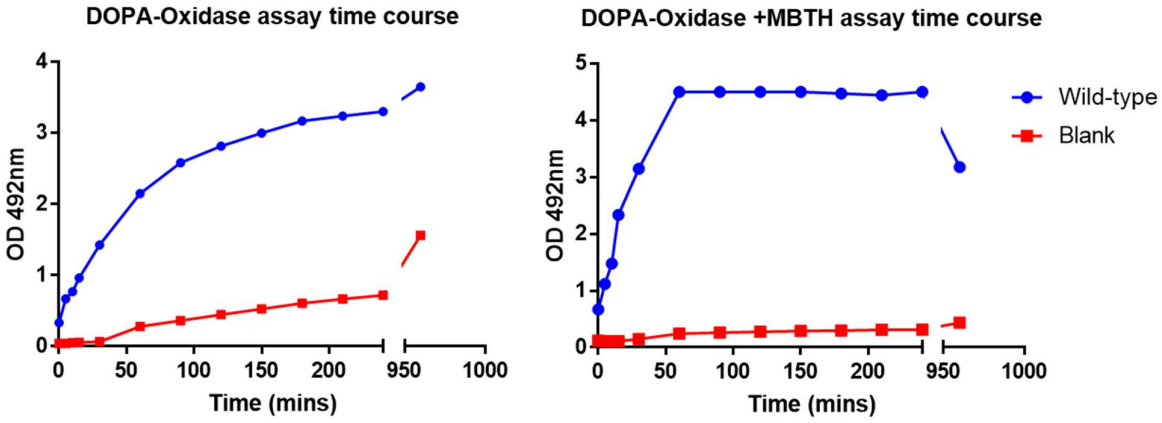


Figure 31. Extended time course for both DOPA-only DOPA-Oxidase (left) and MBTH-coupled DOPA-Oxidase (right) assays of tyrosinase enzymatic activity. Cell lysate containing wild-type tyrosinase activity was run alongside a buffer-only blank. The assay plate was incubated at 37˚C between measurements.

4.4.3 Determining protein maturity/retention in endoplasmic reticulum

Retention of *TYR* mutants in the endoplasmic reticulum (ER) is believed to be a common defect of the melanin biosynthesis pathway causing OCA1[147, 187, 188]. Berson *et al*. used an endo H assay to show the R402Q variant is likely retained in the ER, counting towards the loss of tyrosinase function seen in HeLa cells[180]. However, they also demonstrated wild-type human tyrosinase expressed in non-melanogenic exits the ER inefficiently, just to a lesser degree.

Protein digestion with the enzyme endo H was used to investigate the maturity and potential ER retention of the tyrosinase mutants. The assay involves digestion of equal concentrations of whole protein from whole cell lysate followed by SDS-PAGE to separate protein fragments by size and western blot to detect the tyrosinase fragments (using monoclonal mouse anti-DDDK antibody). Endo H cleaves the bond in the core of the oligosaccharide between two N-acetylglucosamine (GlcNAc) subunits directly next to the asparagine residue, truncating the sugar molecule to a single N-acetylglucosamine residue. The enzyme recognises the early mannose-rich oligosaccharides which are present on the immature protein throughout the ER and the compartments of the Golgi until Golgi alpha-mannose II removes the mannose sugars (*Figure 32*) [189]. As a more mature protein has fewer sugars, it is less susceptible to digestion and thus, lack of digestion suggests the protein has not been retained in its immature form in the ER.

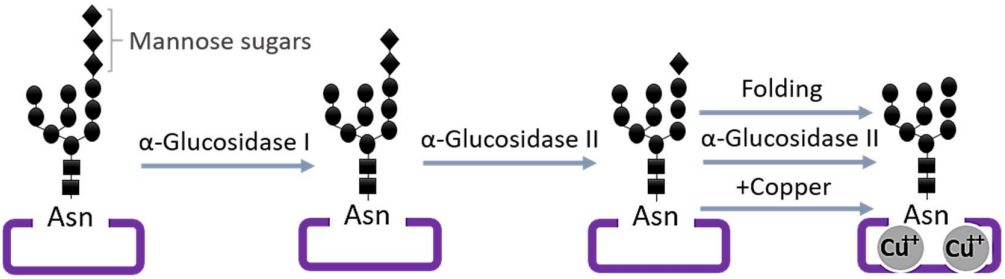


Figure 32. Endo H recognises mannose rich oligosaccharides and will deglycosylate the immature tyrosinase glycoprotein. As the protein enters the ER an oligosaccharide molecule containing 14 sugar subunits is selectively linked to Asparagine (Asn/N) residues. As the protein matures, moving through the trans-Golgi network, the large oligosaccharide is edited and trimmed by α-Glucosidase enzymes. Endo H can cleave the sugars up until the point at which two mannose sugars are removed [189].

4.4.4 Visualisation of intracellular melanin

Transmission electron microscopy was used to investigate the apparent melanin production within transfected cells. Cells were pelleted 96 hours post transfection and placed in fixative. The Southampton Biomedical Imaging Unit then processed and imaged ultrathin sections of the fixed cell pellet (methods section 2.6.8). Melanin appears as a darker contrast within the image due the greater absorption of electrons and is easily distinguishable [190].

4.5 Results

4.5.1 Decreased enzymatic activity of tyrosinase mutants

4.5.1.1 DOPA-oxidase assay

The DOPA-Oxidase assay provides a steadily increasing curve for both wild-type tyrosinase and S192Y-mutant tyrosinase. The S192Y-mutant is not as active as wild-type and the two activities appear to show greater divergence over time. R402Q and S192Y:R402Q do not show any greater absorbance than the buffer-only blank, even at the final time point (*Figure 33*).

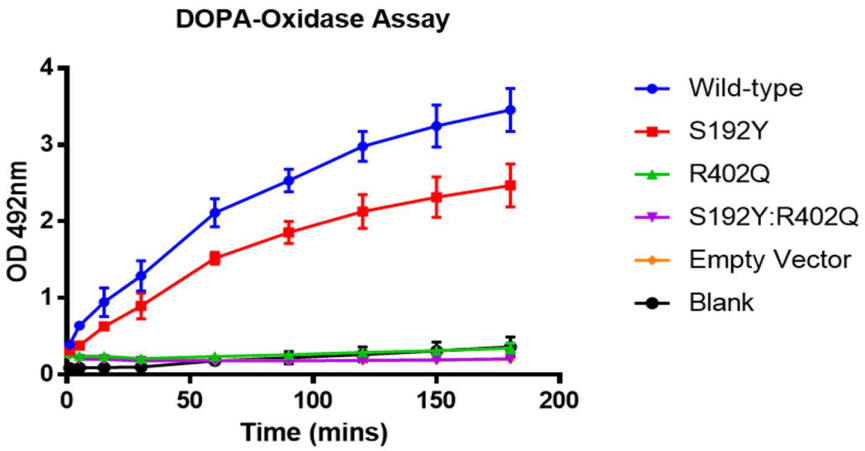


Figure 33. DOPA-Oxidase assay time course measuring the production of Dopachrome by tyrosinase from the substrate L-DOPA. The optical density was recorded at a wavelength of 492 nm every 5 minutes for the first 15 minutes, followed by every 30 minutes for a total of 3 hours. Cell lysates containing wild-type, S192Y, R402Q, and S192Y:R402Q tyrosinase were assayed at the same time as empty vector cell lysate and a buffer-only blank (n= 3). The assay plate was incubated at 37˚C between measurements.

The percentage activity at the final time point shows a highly significant difference between wild-type and S192Y DOPA-oxidase activity ($P < 0.001$) and a highly significant difference between S192Y and R402Q activity ($P < 0.001$). There was no significant difference between R402Q activity and the activity of the double mutant S192Y:R402Q ($P = 0.827$). Both the mutants S192Y:R402Q and R402Q show no detectable activity in the DOPA-only assay as their percentage activity is the same the (tyrosinase-negative samples) empty vector and blank (*Figure 34*).

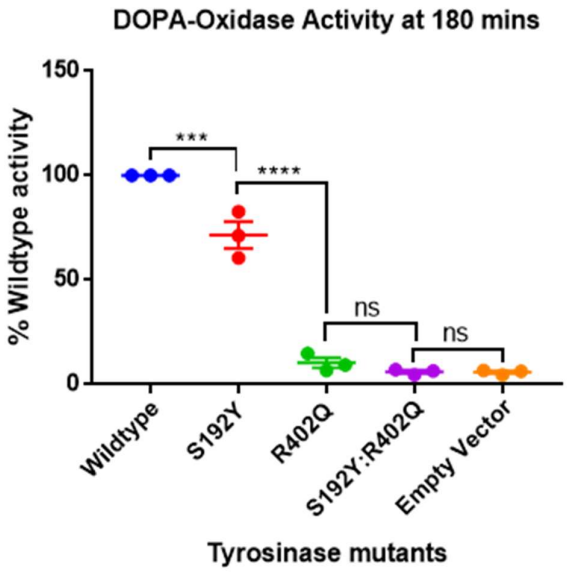


Figure 34. The DOPA-oxidase activity of each tyrosinase mutant at the final 180 minute time point as a percentage of wild-type activity (n= 3). A one-way ANOVA has been used to test for significant difference in activity of mutant tyrosinase, followed by Sidak's multiple comparisons test between wild-type and S192Y ($P < 0.001$***), between S192Y and R402Q ($P < 0.001$****), between R402Q and S192Y:R402Q ($P = 0.827$), and between S192Y:R402Q and empty vector ($P = 1.000$).

4.5.1.2 **MBTH-coupled DOPA-oxidase assay**

The MBTH-coupled DOPA-Oxidase assay provides a fast increase in absorbance for both wild-type tyrosinase and S192Y-mutant tyrosinase. The rate of reaction for S192Y is lower than wild-type but the absorbance for the S192Y reaction is fully saturated two hours later. S192Y:R402Q does not show any greater absorbance than the buffer-only blank, even at the final time point. However, R402Q appears to have a low but noticeable activity with a gradual increase in absorbance over time, diverging away from the blank and S192Y:R402Q double mutant (*Figure 35*).

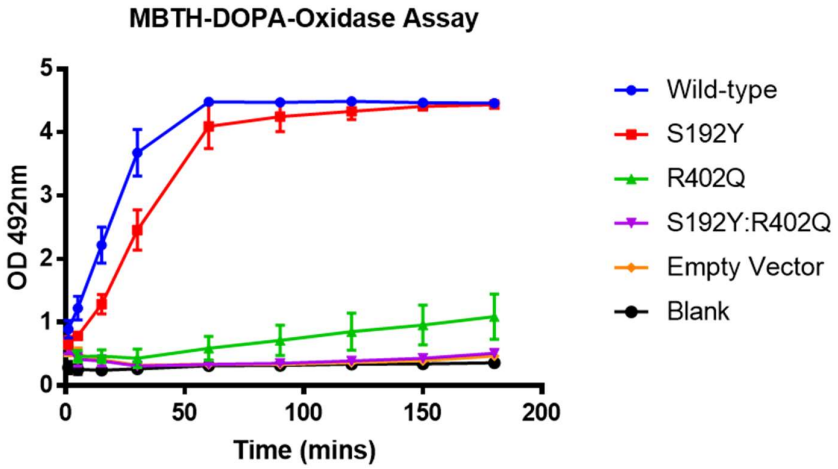


Figure 35. MBTH-coupled DOPA-Oxidase assay time course measuring the production of dopaquinone by tyrosinase. The dopaquinone product is sequestered by MBTH to create a stable MBTH-dopaquinone adduct. The optical density was recorded at a wavelength of 492 nm every 5 minutes for the first 15 minutes, followed by every 30 minutes for a total of 3 hours. Cell lysates containing wild-type, S192Y, R402Q, and S192Y:R402Q tyrosinase were assayed at the same time as empty vector cell lysate and a buffer-only blank (n= 3). The assay plate was incubated at 37˚C between measurements.

The percentage activity at the final time point shows no difference between wild-type and S192Y activity due to the increased sensitivity of the assay and early saturation. There is a highly significant difference between S192Y and R402Q activity ($P < 0.001$). There was no significant difference between R402Q activity and the activity of the double mutant S192Y:R402Q ($P = 0.082$) or against the negative control ($P = 1.000$) (tyrosinase-negative empty vector). However, there is a non-significant trend towards the R402Q mutant having greater activity than the double S192Y:R402Q mutant, becoming more exaggerated towards the final time point (*Figure 36*).

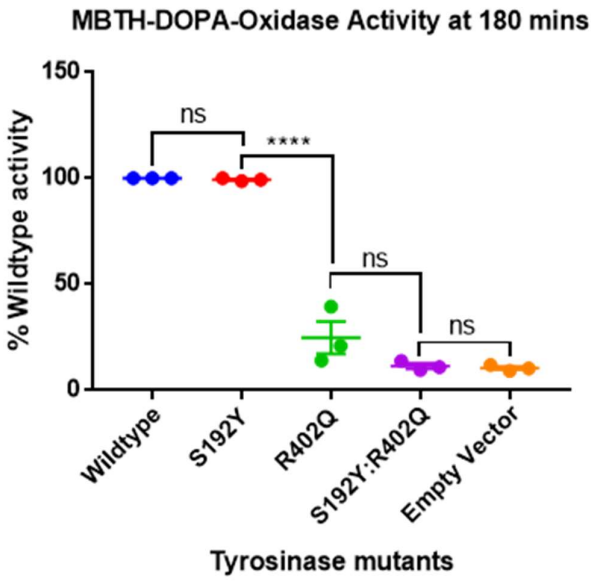


Figure 36. The DOPA-oxidase activity of each tyrosinase mutant at the final 180 minute time point as a percentage of wild-type activity (n= 3) A one-way ANOVA has been used to test for significant difference in activity of mutant tyrosinase, followed by Sidak's multiple comparisons test between wild-type and S192Y (P = 1.000), between S192Y and R402Q (P < 0.001****), between R402Q and S192Y:R402Q (P = 0.082), and between S192Y:R402Q and empty vector (P = 1.000).

When correcting the DOPA-oxidase values for the empty vector control the percentage of remaining activity becomes clearer. The S192Y mutant ranges between 70% activity and wild-type activity, in stark contrast with R402Q which shows an average of 16% activity in the MBTH-coupled assay but less than 5% activity when the assay was measuring the slower formation of dopachrome. The tyrosinase containing both mutations is consistently low for both assays with near to zero DOPA-oxidase activity (*Table 15*).

Table 15. Mean activity for each tyrosinase mutation at the final (180 minute) time point studied. The activity is calculated as a percentage of wild-type activity for each assay repeat. The percentage of remaining activity was then averaged. Results are for both the DOPA-only and MTBH-coupled DOPA-oxidase assays.

	Percentage of wild-type tyrosinase activity	
Tyrosinase mutation	**DOPA only**	**MBTH + DOPA**
S192Y	69.7 %	99.3 %
R402Q	4.7 %	15.05 %
S192Y:R402Q	0.2 %	1.1 %

4.5.2 Retention of tyrosinase in the endoplasmic reticulum

The digestion of tyrosinase by endo H shows there is a high ratio of immature (fully deglycosylated, endo H sensitive) 65 KDa protein to mature 75 KDa protein for all tyrosinase variants. From the western blot it is clear that wild-type tyrosinase had the greatest percentage of mature protein followed by the S192Y mutant. There appears to be no mature tyrosinase when the protein is carrying the R402Q and S192Y:R402Q mutations as nearly all protein was digested. There is a distinct secondary band forming beneath the undigested protein suggesting only partial digestion occurred here as the protein had been released from the ER but had not fully matured (*Figure 37*). The amount of wild-type tyrosinase produced is also higher than all mutants. This is clear as the total protein concentration was normalised across each well for gel loading yet the overall amount of tyrosinase detected in the wild-type lane is greater (*Figure 37*).

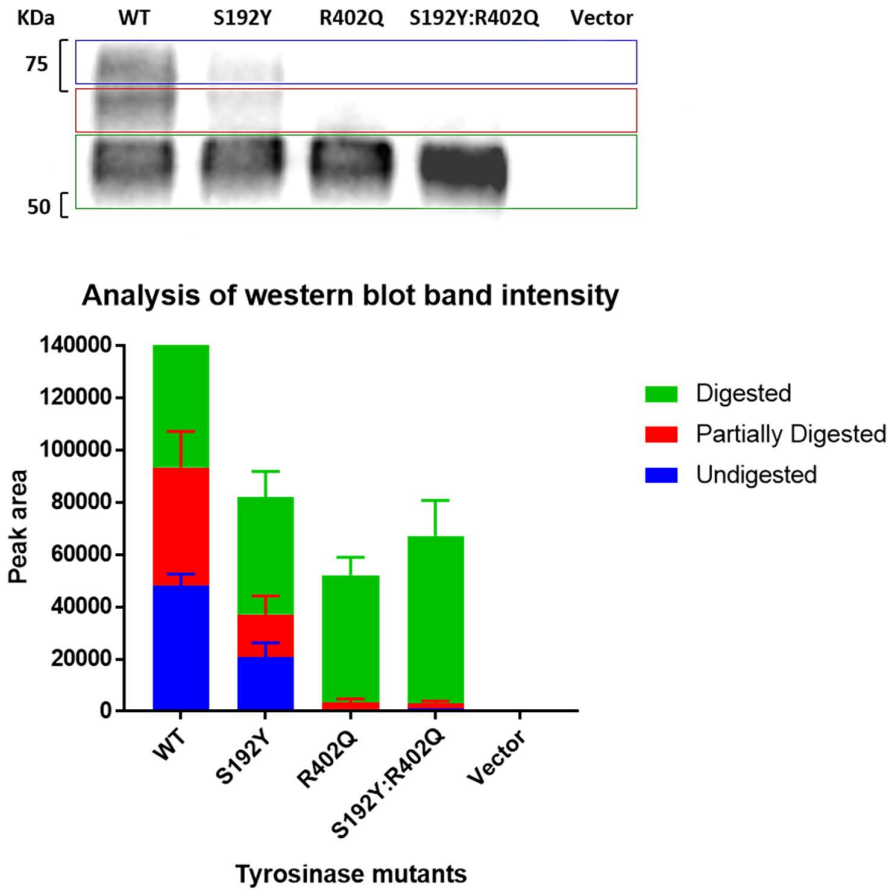


Figure 37. Western blot analysis of endo H digested tyrosinase wild-type and mutants. Analysis of western blot band and intensity using imageJ to compare digested and undigested protein. Cell lysates from transfected HEK293F cells expressing recombinant tyrosinase protein were digested with endo H for 1 hour and denatured for western blot. They were run for 1 hour at 125 V in a 10% SDS-PAGE gel, transferred for 1 hour at 35 V and blocked for 90 mins. A single HRP conjugated anti-DDDDK antibody was used to detect the recombinant tyrosinase with C-terminal FLAG-tag (n= 3). The band intensity was plotted as a graph with the area under the peak representing the size and intensity of the band. The ratio of 75 KDa band to 65 KDa band shows the ratio of mature to immature protein (digested and undigested) respectively. In between these two bands there is a partially digested band of roughly 70 KDa.

Analysis of the undigested/mature tyrosinase shows that a smaller percentage of the S192Y-mutant is mature compared with wild-type, however the difference is not statistically significant ($P = 0.095$). The percentage of mature S192Y-mutant is greater than that of R402Q-mutant with high statistical significance ($P < 0.001$) but there is no difference between the R402Q and S192Y:R402Q mutants ($P = 1.000$) which are the same as the vector-only negative control (*Figure 38*). The percentage of 'partially digested' tyrosinase follows the same pattern as undigested for each of the wild-type/mutant proteins. There is significant difference between the percentage of fully digested wild-type tyrosinase and fully digested S192Y-tyrosinase ($P = 0.024$) and the amount of fully digested S192Y mutant is visibly higher. The amount of R402Q

and S192Y:R402Q proteins that have been fully digested are higher still, with no significant difference between them. The amount of mature protein decreases from wild-type to S192Y followed by R402Q and S192Y:R402Q with no significant difference between the final two mutants (*Figure 38*). Statistical analysis is detailed in (*Table 16*).

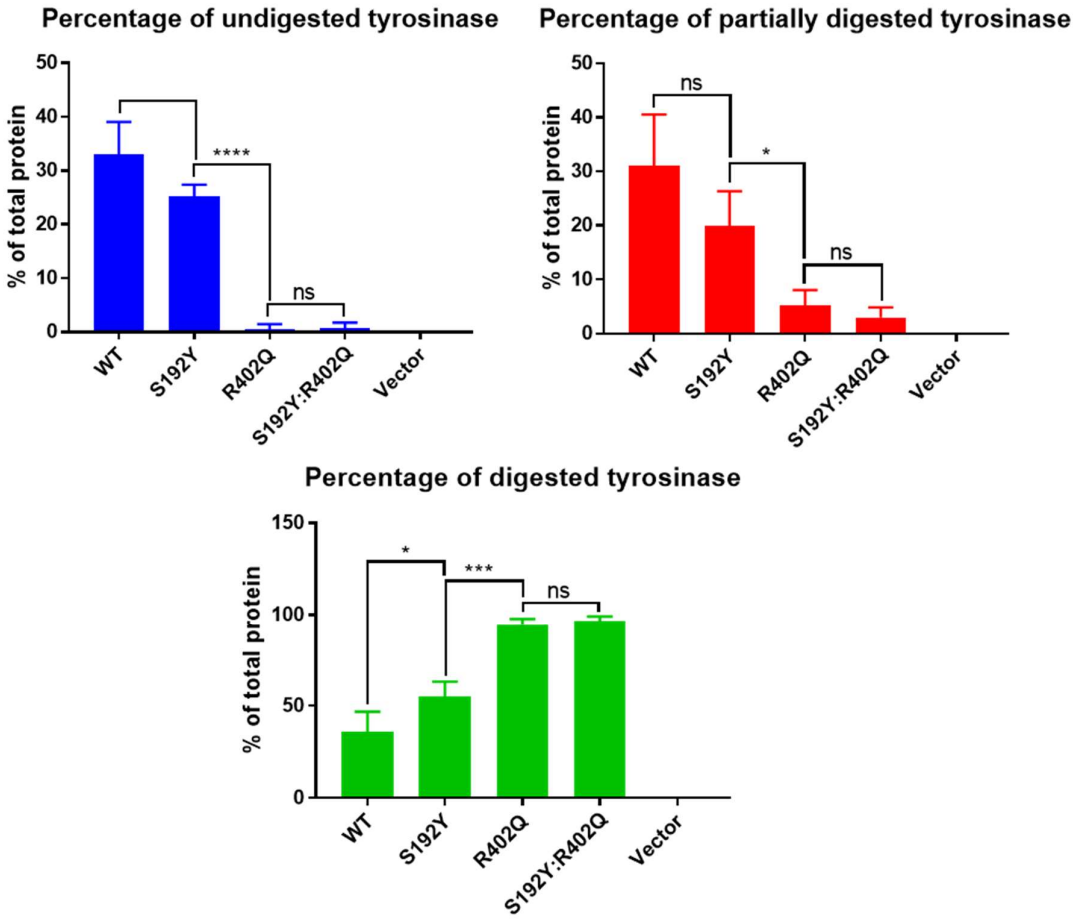


Figure 38. Western blot analysis of endo H digested tyrosinase wild-type and mutants. The ratio of undigested 75 KDa band to digested 65 KDa band shows the ratio of mature to immature protein (blue and green bar charts) respectively (n= 3). In between these two bands there is a partially digested band of roughly 70 KDa (red bar chart, n= 3)). The intensity of each band has been plotted as a percentage of the total protein. A one-way ANOVA has been used to compare the difference in percentage of undigested, digested, and partially digested protein between the different tyrosinase mutations.

Table 16. Sidak's multiple comparisons test following a one-way ANOVA to assess statistical significance of percentage undigested/ partially digested/ digested tyrosinase wild type and mutants. The summary values are included in *Figure 38* where appropriate.

	Sidak's multiple comparisons test	n	Mean Diff. [A-B]	SE of diff.	95% CI of diff.	P Value	Summary
Undigested	WT vs. S192Y	3	7.882	2.72	-1.191 to 16.95	0.095	ns
	WT vs. R402Q	3	32.47	2.72	23.39 to 41.54	<0.001	****
	WT vs. S192Y:R402Q	3	32.36	2.72	23.29 to 41.43	<0.001	****
	S192Y vs. R402Q	3	24.58	2.72	15.51 to 33.66	<0.001	****
	R402Q vs. S192Y:R402Q	3	-0.1062	2.72	-9.179 to 8.967	1.000	ns
Partially digested	WT vs. S192Y	3	11.09	4.36	-2.691 to 24.86	0.138	ns
	WT vs. R402Q	3	25.92	4.36	12.14 to 39.7	0.001	***
	WT vs. S192Y:R402Q	3	28.02	4.36	14.24 to 41.79	0.000	***
	S192Y vs. R402Q	3	14.84	4.36	1.059 to 28.61	0.033	*
	R402Q vs. S192Y:R402Q	3	2.094	4.36	-11.68 to 15.87	0.994	ns
Digested	WT vs. S192Y	3	-18.97	5.27	-35.61 to -2.329	0.024	*
	WT vs. R402Q	3	-58.39	5.27	-75.03 to -41.75	<0.001	****
	WT vs. S192Y:R402Q	3	-60.38	5.27	-77.01 to -43.74	<0.001	****
	S192Y vs. R402Q	3	-39.42	5.27	-56.06 to -22.78	<0.001	***
	R402Q vs. S192Y:R402Q	3	-1.988	5.27	-18.63 to 14.65	0.998	ns

4.5.3 Melanin production within transfected HEK293F cells

When pelleted, the transfected HEK293F were visibly different in colour and apparent pigmentation (*Figure 39*). Transmission electron microscopy was used to confirm the presence of melanin inside the cell.

Figure 39. Image of pigmented cell pellet expressing tyrosinase wild-type and mutants. HEK293F expressing recombinant wild-type and mutant tyrosinase were pelleted 96 hours after transfection and media was discarded. Each square image is a sample taken from the centre of the pellet photograph. The empty vector is negative control with no tyrosinase present.

4.5.3.1 Transmission electron microscopy images of transfected cells

Transmission electron microscopy of HEK293F cells transfected with wild-type tyrosinase revealed dark deposits strongly resembling melanosomes. There appears to be some normal melanosome structures as well as many larger vesicles containing multiple dark deposits and

some deposits with less defined structures (*Figure 40*). The abnormal multi-granular deposits resemble compound melanosomes that have been shown to occur within keratinocytes before lysosomal degradation occurs [190]. The less defined structures may also resemble neuromelanin as it is unclear whether neuromelanin is contained within a melanosome-like membrane [191]. The optimum time to pellet cells for maximal pigmentation appears to be just after 72 hours. At this point melanin has begun to accumulate but has not been eradicated from cells. The TEM images show cells under stress, as melanogenesis is not a normal process for this cell line and the cell therefore using lysosome-like structures to remove the melanin.

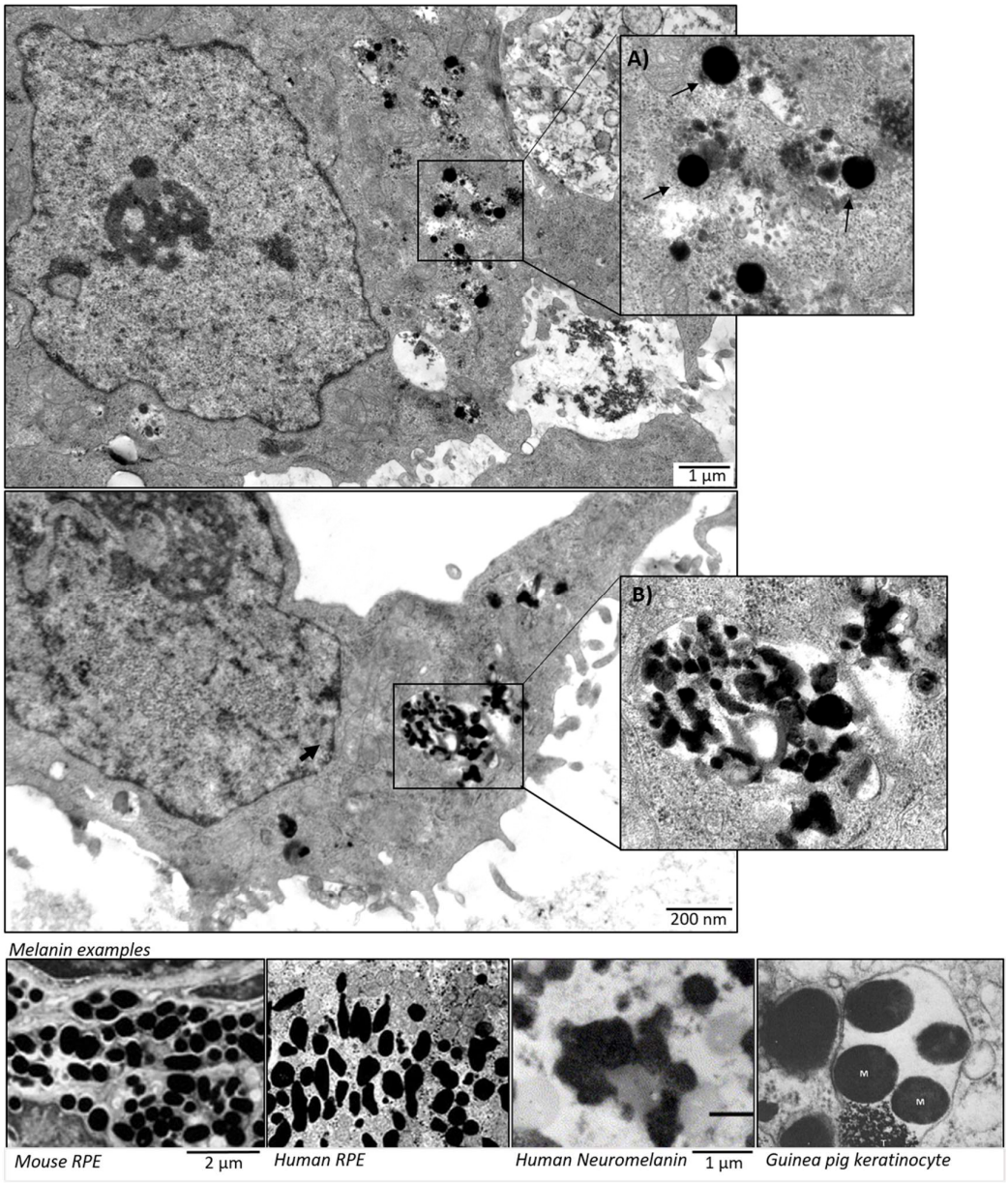


Figure 40. Transmission electron micrograph images of melanin deposits within HEK293F cells transfected with tyrosinase. The cells were transfected with wild-type tyrosinase, selected for using the neomycin resistance gene within the vector, and pelleted 72 hours post transfection. The cell pellet was fixed and prepared as ultra thin slices mounted on a copper grid for imaging by TEM. A) Transfected HEK293F cell containing dark vesicles consistent with mature melanosomes of mouse[192], and human RPE[91], but some dark deposits have less defined edges similar to human neuromelanin[191] (example images across the bottom). B) A large lysosomal-like structure within the cell containing multiple deposits of melanin, consistent with the compound melano-lysosome image of the Guinea pig keratinocyte [190].

4.6 Discussion

4.6.1 The functional effect of the tyrosinase common variants

The assays used to determine the functional effect of each variant improved upon previous work as the full length enzyme was produced human cells and was even able to exhibit the native function of melanin production in non-melanogenic cells. As the tyrosinase protein was not purified from whole lysates the functional studies were able to more accurately represent *in vivo* activity taking into account the possible consequences of expression and degradation of mutant tyrosinase.

The common tyrosinase variant S192Y has been implicated in OCA1B but only when in addition to the common variant R402Q and a rare disease-causing mutation and this hypothesis is still in debate. Both mutations are located on the surface of the protein. The serine to tyrosine mutation at position 192 is not a considered a conservative mutation as the tyrosine introduces a large aromatic group that could protrude from the protein or cause a conformational change (*Figure 41b*). At position 402, the mutation of the basic arginine to an acidic glutamine would cause a change in surface charge and potentially cause the residue to shift position, burying itself within the protein (*Figure 41c*).

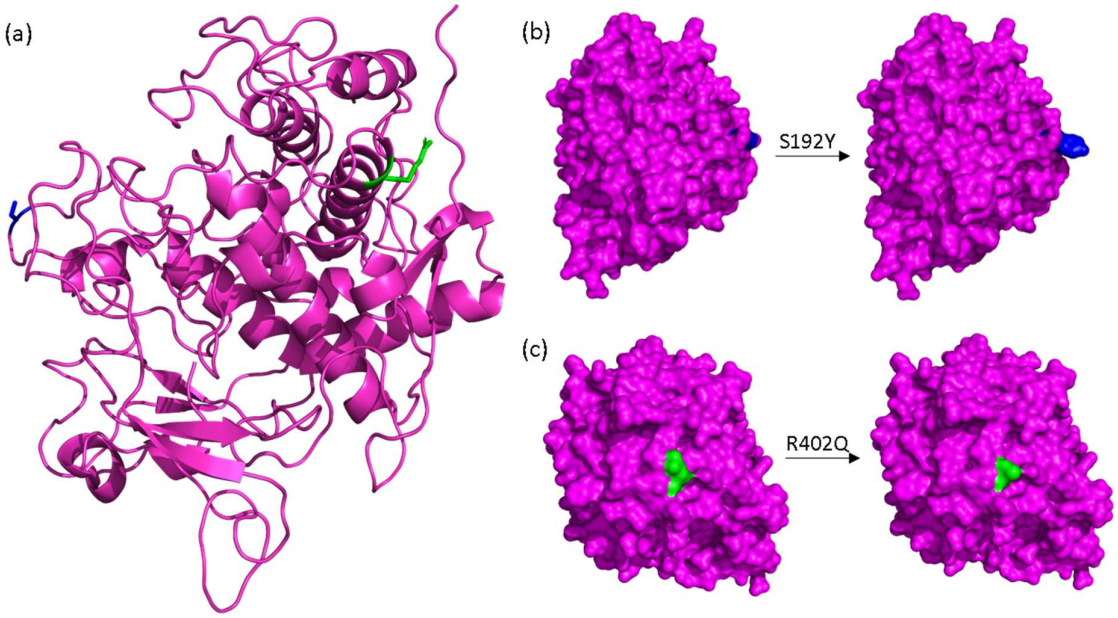


Figure 41. A homology model of the tyrosinase protein. (a) The tyrosinase protein was modelled from the closely related TYRP1 protein (PDB: 5M8L) using the Modeller program and figure was produced in open source PyMOL. The serine residue at position 192 is highlighted in blue and the arginine residue at position 402 is highlighted in green. (b) A surface map of the tyrosinase protein with the serine residue at position 192 exchanged for a tyrosine. (b) The arginine residue at position 402 exchanged for a glutamine.

A functional study of the effect of S192Y on the tyrosinase enzyme has shown a subtle knock down of activity. Previously it has been stated that the S192Y variant causes a 40% loss of activity which is in line with these results, however variability affected the significance of statistical analysis. Further to this, the maturity of overall S192Y-mutant was not affected when comparing the percentage breakdown of total protein with wild-type percentage breakdown. Although the overall amount of tyrosinase produced was lower which may be due to early degradation if the S192Y mutant causes misfolding within the protein. It could be argued that the loss of activity is minor as the S192Y mutant produced visible pigmentation and the knock down of activity would equate the differences between pigmentation in populations[59]. This supports the assertion that S192Y alone is very unlikely to be sufficient for a causal mutation.

The common tyrosinase variant R402Q has also been implicated in OCA1B, both as a combined variant in a tri-allelic genotype and an independent causal variant when in combination with a rare variant on the trans allele. A functional study of the R402Q-mutant tyrosinase revealed a close to complete loss of DOPA-oxidase activity with only 5% to 16% remaining. Even though the percentage of remaining activity is so low, it has been stated that only 5%-8% activity is required to influence an OCA1B phenotype [193] and the R402Q-mutant still appears to produce a low level of pigmentation as seen in the cell pellet (*Figure 39*). The R402Q-mutant activity was greater than that of the double S192Y:R402Q-mutant (at 0.2% to 0.4%), however the difference was not statistically significant for either DOPA-oxidase assay according to the time-point used in this study. The results of R402Q differ from those previously published as functional analysis showed a remaining enzymatic activity of close to 25% [60, 146].

This large loss of R402Q-mutant activity is not reflected in the phenotype as homozygous R402Q has never been reported as a causal genotype. It has been argued that even R402Q in *trans* with a known mutation can be insufficient for albinism [148],[166]. The difference in the functional findings is this study and those seen in a homozygous R402Q phenotype may be due to the mutant's temperature sensitivity. Berson *et al*. produced a fully active R402Q mutant at 31˚C compared with a completely inactive mutant at 37˚C. The loss in tyrosinase activity was determined a result of ER retention. When expressed in non-melanogenic cells, wild-type tyrosinase folds and exits the ER inefficiently and it appears that this is exaggerated at physiological temperatures [180]. Further to this, Halaban *et al.* have shown that co-expression of wild-type tyrosinase with R402Q-mutant helps translocation of the mutant protein, reducing ER retention [188]. It would suggest that an exaggeration of ER retention would occur if a second ER retention mutation was expressed alongside the R402Q-mutant. Such a compound heterozygous genotype may be more likely to be result in a severe albinism phenotype. It is interesting that R422W has also shown temperature sensitivity in previous studies. There are a

few possible explanations for the stark differences in activity reported by opposing research groups [61, 62, 186]. Firstly, the protein maturation pathway of insect cells used by the Dolinska group may be causing both tyrosinase to undergo abnormal glycosylation, consequently wild-type activity may also be hindered [61, 186, 194]. Secondly, the effect of protein retention and degradation is not accounted for by Dolinska *et al.* as the tyrosinase is purified before any enzymatic assay. A final cause of inconsistency may be due to the removal of the C-terminal transmembrane domain. The domain was removed for easier purification of tyrosinase but this may be negatively effecting tyrosinase activity, wild-type included, as the two cytoplasmic serine residues are no longer available for phosphorylation by PKC-β [61, 186].

4.6.2 The tri-allelic theory of OCA1 inheritance

The tri-allelic theory of OCA1 inheritance states that the two common *TYR* mutations, S192Y and R402Q, coinherited with a pathogenic disease-causing *TYR* mutation will result in hypomorphic albinism. As the pathogenic disease-causing mutation would cause a complete loss of active tyrosinase, the most damaging haplotype would likely be R402Q and S192Y in *cis*. The data for maturity and activity of this double mutant was equal to that of the negative control (empty vector), showing a total loss of active tyrosinase. For a hypomorphic OCA1B phenotype there needs to be some remaining tyrosinase activity as pigment can still accumulate. The complete loss of activity exhibited is likely due to the temperature sensitivity of R402Q already discussed and the negative effect on translocation produced by expression of tyrosinase in non-melanogenic cells.

The tri-allelic genotype is used to explain the lack of phenotype reported by Oetting *et al.* in volunteers with a compound heterozygous R402Q and known disease-causing mutation genotype [148]. It was suggested that S192Y is required in addition to produce the necessary knock down of tyrosinase activity for detectable features of albinism. The results of this study do not find a statistically significant difference between the availability of active R402Q-tyrosinase and of S192Y:R402Q-tyrosinase. There is a trend towards greater activity in the R402Q mutant which may be underestimated in these conditions as the HEK293F cells used are not a traditionally melanogenic cell line. Furthermore, the temperature sensitivity reported in R402Q can mean that pigmentation may occur in the periphery, as measurements of skin temperature cover a range from 31˚C to 39˚C with an average of roughly 33˚C [195]. The addition of S192Y may therefore decrease activity enough to negate the effect of permissible temperatures and the difference in activity between R402Q and R402Q:S192Y-tyrosinase may only be detectable in more native cells.

4.6.3 Development of a tyrosinase functional assay

The assay described here analysed the DOPA-oxidase activity of tyrosinase but did not analyse the enzyme's tyrosine hydroxylase activity. Previous studies have applied the same simple dopachrome production assay to tyrosine hydroxylase only replacing the substrate L-DOPA with tyrosine. However, the two activities cannot be distinguished as DOPA-oxidase is still forming part of the pathway to dopachrome production. Tyrosine hydroxylase activity can be distinguished using [^{3}H]-tyrosine and measuring the release of ^3H_2O, however the aim of this study was to create a simple a robust measure that could be used with relative speed for novel mutations, making a radioactive assay unfeasible.

Including the 3-methyl-2-benzothiazoline MBTH in the DOPA-oxidase activity assay is perhaps an improvement on the non-MBTH DOPA-chrome assay. It allows measurement of the direct product of L-DOPA oxidation, trapping it as a stable dopaquinone-MBTH adduct. Moreover, DOPA-chrome itself is still not a stable product so MBTH is increasing the assay's sensitivity. The sensitivity is also increased because the molar absorption coefficient of the pink pigment is seven to eight times greater than that of dopachrome.

During this study it was noted that the transient transfection of HEK293F cells with tyrosinase can cause the cells to produce melanin after a fairly short time period. The graded difference in melanin production between wild-type and mutant tyrosinase expression is possibly the most valid measure of available active tyrosinase. The measure of melanogenesis in a cell line with no background melanin could be used to create an extremely quick and simple assay of *TYR* mutants, the only downfall being that the mechanism of melanin production within HEK293F cells is not understood at present. The cellular origin of the HEK293 line is unclear as embryonic kidney cells consist of a heterogeneous mixture of almost all the types of cells present in the body. It is difficult to determine the cellular characteristics since adenovirus 5 used in transformation may have significantly altered cell morphology and expression profile. Most likely the cells are endothelial, epithelial or fibroblast in origin, however, it has been speculated that they are neuronal due to the presence of some neuronal mRNA[196].

Non-melanogenic cells are chosen as they are tyrosinase-negative, however, melanogenic cells could be utilised if the native *TYR* is knocked out using a technique such as CRISPR/cas9. A knock-out cell line could be transfected with mutated tyrosinase to assess the protein activity in its native environment and could be further developed to assess the other genes involved in albinism. Another option is to use a cell line known for melanin production but only under certain conditions. For example, the human RPE cell line, ARPE-19, can produce melanin after 2-3 months of culture as a confluent monolayer, but the cells do not readily produce melanin

under other conditions [197]. As this cell line is highly relevant to the ocular deficits characteristic of albinism it may be a prime candidate for a melanogenic assay of tyrosinase mutants. Currently, iPSC cells generated from skin biopsies are the best option for screening therapeutic candidates for albinism [198]. If a simple cell line was utilised instead the screening of candidate therapeutics would be much quicker and more cost effective. Furthermore, it is known that tyrosinase is the only essential protein for the process of melanogenesis [79]. It is, therefore, unnecessary to understand exact mechanism of melanin production as we can assume it is a direct measure of tyrosinase activity.

4.6.4 Testing potential therapeutics

The described robust functional assay may be used to test potential therapeutics. For example, the drug nitisinone has been used to improve the ocular and skin hypopigmentation of OCA1B in albino mice. Nitisinone inhibits tyrosine degradation, increasing the plasma concentration of tyrosine which appears to upregulate the melanogenic activity of tyrosinase. This drug is specific for OCA1B mutations and is already FDA approved for treatment of hereditary tyrosinemia type 1 (elevated levels of tyrosine in the blood) so could be prime candidate for treating the tri-allelic genotype [199].

Other potential therapeutics include ERAD inhibitors such as kifunensine which blocks recognition of the protein or eeyarestatin I which blocks retrotranslocation of misfolded proteins and enhances folding of unstable proteins in the ER [200]. The application of palmitic acid has been shown to decrease ER retention of tyrosinase in melanoma cells and could be tested in the context of albinism [201]. Also, a recent publication has suggested that OCA1 may be treated with a low-dose tyrosinase inhibitor that acts as a chemical chaperone helping tyrosinase leave the ER[202].

4.6.5 Conclusion

The functional study of these controversial tyrosinase mutations aids the difficult diagnosis of albinism by validating mutations and linking them to their resulting phenotypes. The identification of causative genotypes will inform the best possible clinical and genetic workflow to be followed for efficient diagnosis. In addition, the production of a robust and accurate assay of tyrosinase mutations will enable the screening on possible therapeutics bringing us closer to finding a treatment of which there currently is none available.

Chapter 5: The UKGTN approved albinism and nystagmus 31 gene panel as a diagnostic tool

5.1 Introduction

5.1.1 Challenges in diagnosis for nystagmus and albinism

Diagnosis of infantile nystagmus syndrome (INS) can be very challenging as the nystagmus can be an isolated trait or may occur in association with a diverse range of underlying visual sensory and systemic disorders. Infants with very atypical nystagmus or with clear neurologic signs are frequently investigated with an early MRI brain scan and multiple additional tests in order to identify severe neurological or metabolic disease [48]. However, most cases of INS are due to inherited ocular disorders and should be approached with detailed ocular work up, including VEP, ERG and OCT when possible [48, 203]. The diagnosis of idiopathic INS is mostly one of exclusion, given when investigations have revealed no obvious aetiology. Therefore, young patients are frequently given a diagnosis of idiopathic INS because of limited clinical resources and the subtlety of some disorders or a delay in the onset of phenotypic features. This means that many children with underlying neurological or ophthalmic conditions are inaccurately 'diagnosed' and in some cases, opportunities for intervention are missed. Subtle pathologies such as those seen in diseases with wide expressivity and heterogeneity, e.g. albinism and isolated foveal hypoplasia, may go undetected for many years, without a thorough detailed ocular investigation. Genetic confirmation can be used to determine a correct diagnosis when features are not pathognomic for a single disorder, sometimes avoiding further unnecessary investigation as well as providing a confident diagnosis for the patient and their family and permitting genetic counselling.

There are few treatments for the various ophthalmic conditions associated with INS, nevertheless, a molecular diagnosis is required for genetic counselling, correct management of the condition and identification (or exclusion) of disorders with significant systemic associations. Over 80% of nystagmus in childhood is inherited, but when this project started in 2015, only a small proportion was diagnosable with NHS-available (UKGTN approved) genetic tests [204]. There were three main UKGTN panels relevant to nystagmus; a cataract panel, a retinal dystrophy panel, and a small albinism panel. The albinism panel was particularly limited as it only contained two of the main causal genes for OCA. It was also possible to order sequencing of the *FRMD7* gene associated with X-linked idiopathic nystagmus; however, this is only suitable for known X-linked to ensure high likelihood of a positive result. Furthermore, many conditions associated

with INS have overlapping and heterogeneous features that may be missed due to their subtlety. Within the OMIM database there are 820 gene entries associated with nystagmus [OMIM, Jan 2019] but these can be narrowed down to a more select few that can result in a clinical diagnosis when mutations are located.

5.1.2 The UKGTN approved albinism and nystagmus 31 gene panel

In 2017 Luke Michaels was the lead in submitting a suggested UKGTN gene panel for nystagmus and albinism. The panel was partly based upon some initial findings from this cohort and preliminary analysis of the TruSight One clinical exome. The first gene on this list is *FRMD7*, the gene for X-linked idiopathic INS (*Table 17*).

Tarpey *et al.* reported that 57% of patients with plausible X-linked inheritance and 94% of those where X-linked inheritance carry mutations in the *FRMD7* gene [33]. There is yet to be a gene associated with autosomal idiopathic INS but there are numerous disorders with subtle features that may have INS as the only visible trait. Therefore, the panel includes genes such as *CACNA1A* which is known to cause spinocerebellar ataxia and which can present as apparent isolated INS but will develop into late onset ataxia [205].

Albinism is a term that covers a wide range of disorders affecting pigmentation and vision, often resulting in nystagmus. Defects to the melanogenesis pathway can result in Oculocutaneous albinism (OCA), affecting skin, hair and eyes, or Ocular albinism (OA), affecting eyes only. The genes *TYR*, *OCA2*, *TYRP1*, *SLC45A2*, *SLC24A5* and *LRDMA* (*C10orf11*) are responsible for OCA subtypes 1-4 and 6-7 respectively [141]. The gene *GPR143* is responsible with X-linked OA (OA type 1). A second subtype of Ocular albinism was proposed under the term autosomal recessive ocular albinism (AROA) [144], however, this term refers to the same causal genes as in OCA. The distinction is not necessary and is no longer used as it is referring to one end of the wide phenotypic spectrum covered by OCA [166]. Albinism can also be the result of systemic disorders that cause defective melanogenesis as part of a wider problem, such as the melano-lysomal disorder Elejalde syndrome which affects the central nervous system [77]. Syndromic causes of albinism have been included on the panel due to the difficulty in diagnosis in very young children (*Table 17*).

The features associated with albinism may be missed or may not be indicative of albinism due to the subtlety and heterogeneity of some phenotypes. For example, mutations in the *PAX6* gene are known to cause a phenotype similar to ocular albinism without visible aniridia that is characteristic of this gene [43, 206]. The X-linked gene *CACNA1F* can cause congenital stationary night blindness (CSNB) with great clinical variability and may present with early onset

nystagmus[45]. Leber congenital amaurosis (LCA) is another ocular disorder with a heterogeneous phenotype, often characterised by severe vision loss, nystagmus and amaurotic pupils [207]. As LCA is the most common cause of inherited childhood blindness and some causative genes can produce phenotypes initially indistinguishable from isolated INS, some LCA genes have been included in the UKGTN panel (*Table 17*).

Table 17. UKGTN approved 31 gene panel for albinism and nystagmus. HGNC approved gene names for genes are listed with their associated conditions according to OMIM and assumed pattern of inheritance. AR = Autosomal recessive, AD = Autosomal dominant, XL = X-linked, XLD = X-linked dominant.

Gene	Pattern of inheritance	OMIM name of associated condition
AP3B1	AR	Hermansky-pudlak syndrome 2
AP3D1	AR	Hermansky-pudlak syndrome 10
BLOC1S3	AR	Hermansky-pudlak syndrome 8
BLOC1S6	AR	Hermansky-pudlak syndrome 9
CACNA1A	AD	Epileptic encephalopathy; episodic ataxia type 2; familial hemiplegic migraine; spinocerebellar ataxia 6
CACNA1F	XL	Aland island eye disease, CSNB
CASK	XLD	Fg syndrome 4
DTNBP1	AR	Hermansky-pudlak syndrome 7
FRMD7	XL	Nystagmus 1, congenital, x-linked
GPR143	XL	Albinism ocular type i; nystagmus, congenital, x-linked
HPS1	AR	Hermansky-pudlak syndrome 1
HPS3	AR	Hermansky-pudlak syndrome 3
HPS4	AR	Hermansky-pudlak syndrome 4
HPS5	AR	Hermansky-pudlak syndrome 5
HPS6	AR	Hermansky-pudlak syndrome 6
LRMDA	AR	Albinism, oculocutaneous, type 7
LYST	AR	Chèdiak-Higashi syndrome
MANBA	AR	Mannosidosis, beta a, lysosomal
MITF	AR	Albinism, ocular, with sensorineural deafness; tietz albinism-deafness syndrome; coloboma, osteopetrosis, microphthalmia, macrocephaly, albinism, and deafness; Waardenburg syndrome, type 2A
MLPH	AR	Griscelli syndrome, type 3
MYO5A	AR	Griscelli syndrome, type 1
OCA2	AD	Albinism, oculocutaneous, type 2
PAX6	AR	Foveal hypoplasia 1
RAB27A	AR	Griscelli syndrome, type 2
SACS	AR	Spastic ataxia, charlevoix-saguenay type
SETX	AR	Spinocerebellar ataxia, autosomal recessive 1
SLC24A5	-	Albinism, oculocutaneous, type 6
SLC45A2	AR	Albinism, oculocutaneous, type 4
TULP1	AR	Retinitis pigmentosa 14
TYR	AR	Albinism, oculocutaneous, type 1A; albinism, oculocutaneous, type 1B; Waardenburg syndrome/albinism, digenic

TYRP1	AR	Albinism, oculocutaneous, type 3

5.1.3 Importance of phenotyping

There are two main prerequisites for a genetic test to have a successful outcome: accurate clinical diagnosis and a careful family history. As nystagmus is such a highly heterogeneous disorder in both genetic cause and phenotype, overall clinical diagnosis is not always clear and instead the specific phenotypic features must be used to guide which genes to study. This has been exemplified in a recent study by Thomas *et al.* where a small cohort of 15 nystagmus patients were sequenced on panel containing 336 genes [208]. The researchers were ultimately able to diagnose 80% of patients with mutations found in only five genes, two albinism genes, *CACNA1F*, *FRMD7*, and the *CRYBA1* gene for cataracts. They also showed the clinical diagnosis was consistent with the genetic diagnosis 9/12 times and the mismatched clinical diagnoses were changed to match on further investigation of ocular phenotype. When the group first carried out analysis, they were blinded to phenotype and were only able to diagnose half as many as when phenotype information was available. Currently, a debate is underway with NHS England regards streamlining and standardising approaches to genetic testing including eye disorders including albinism and nystagmus. Our panel is being proposed as one option but other options include a much broader 'eye panel' which might be employed prior to detailed phenotyping. One argument is that this may provide an efficient workflow if phenotyping is subsequently directed towards proving or disproving putative genetic diagnoses. Clearly, the most efficient workflow is dependent on many factors and this is a topic of much debate currently, but what is clear is that phenotyping will always be a necessary part of any diagnostic workflow.

5.1.4 Variant prioritisation and analysis

The ACMG standards and guidelines are the overarching parameters generally used for interpreting sequence variants of Mendelian disorders for diagnostic clinical laboratories. The guidelines describe a detailed classification system based on different criteria such as population data, computational predictive data, segregation data, and functional studies. According to these parameters, common variants are benign and defined as such by having an allele frequency of greater than 5% in a population database such as 1000 genomes. However, there is evidence within this project to suggest two common variants may provide a causal genotype for OCA1B when combined with another pathogenic variant [166]. Therefore, strict adherence to the guidelines would likely miss genetic diagnosis for some patients with hypomorphic albinism.

5.2 Aim and hypothesis

The aim of this study was to evaluate the clinical utility of a UKGTN approved panel of 31 genes for genetic diagnosis of nystagmus and albinism. The real-world limitations of phenotyping were investigated by genotyping both completely and incompletely phenotyped patients. The overall goal of this study was to provide a cost effective, efficient and accurate workflow which is currently needed for an improved and standardised diagnosis that can actually be implemented in a large healthcare system.

5.3 Choosing a sequencing panel

There are numerous factors to consider when choosing the most appropriate NGS technique including cost, coverage, ease of use and timescale. Whole genome sequencing (WGS) produces the highest volume of data with the greatest coverage uniformity. However, WGS tends to be the most costly technique as not only does it sequence the ~1% of protein-coding DNA but also the ~99% of non-coding sequences. The problem of cost and data storage can be reduced by targeting only the coding sequences in WES. Furthermore, the greater read depth achievable by targeted sequencing can overcome the problems of coverage uniformity, allowing indels and single nucleotide variants to be called accurately.

In 2015 Sun *et al*. carried out both WGS and WES of 9 patients with intellectual disability and then compared the data. WES did not miss any of the 500 genes of interest and WGS had 99% coverage, however, the WES depth of coverage was better and the technique lower cost, making it the preferred option [209]. Dewey *et al*. published a study in 2014 analysing the whole genome sequence of 12 volunteers and there was inadequate coverage of 10% to 19% of variants in genes known to play important roles in inherited diseases [210]. Therefore, a targeted NGS approach enriching the exonic regions of genes of interest currently appears to be the best for coverage depth and for cost if the goal is identifying clinically 'callable' mutations.

One possible targeted NGS panel is Agilent's HaloPlex custom panel, which can be custom designed to cover genes of interest. Previous work used this system to target the coding regions for 55 genes associated with nystagmus, albinism and related ocular disorders (method described in section 2.3.2). A group of 33 samples with nystagmus were run on the panel and a brief analysis of this data suggests a likely diagnosis in roughly 28%, variants of interest in 33%, with a remaining 36% in need of further interrogation. The pick-up rate is good, but the panel is also attractive from a research point of view due to the large amount of data that can be further analysed in the cases with no obvious diagnosis. However, the custom HaloPlex panel cost

approximately £700 per person, whereas larger all-encompassing exome panels are becoming less costly due to wider use and standardised workflow for wet lab and downstream bioinformatics analysis.

There are two main targeted exome sequencing panels on the market, Agilent's SureSelect Focused Exome and Ilumina's TruSight One Sequencing Panel, both for use with the Illumina NextSeq 500/550 sequencer. The SureSelect system requires a large amount of gDNA (200 ng to 3 µg) and purchase of a shearing system, whereas the TruSight One panel requires a much smaller amount of gDNA (50 ng) and uses enzymatic shearing method provided as part of the kit reagents. The SureSelect system has the advantage of a greater number of genes covered (6,110 compared with TruSight One's 4813) and Agilent claims the sequencing can be carried out in a single day. However, the TruSight One panel has a much clearer workflow and is in use in the Southampton WISH (Wessex Investigational Science Hub) lab, meaning local expertise and equipment was available. It is additionally used cross-specialty for diagnostic purposes in a number of laboratories. Due to this and the ability to use a smaller amount of DNA the Illumina TruSight One panel was chosen to sequence probands from the nystagmus and albinism cohort.

5.4 Method

5.4.1 Proband phenotyping and clinical data storage

Patient recruitment and DNA collection was carried out as described in section 2.1. The patient clinical information was stored in a large excel file containing numerous fixed drop-down options to describe the results for each clinical test (methods section 2.4.3).

Probands were included if nystagmus and normal ERG results were present during clinical phenotyping. In total, there were 81 probands divided into four phenotype groups: clinically idiopathic nystagmus (with full phenotyping), clinically idiopathic nystagmus (with incomplete phenotyping), clinically consistent with albinism (with full phenotyping), and clinical features suggestive of albinism (with incomplete phenotype or missing phenotype data) (*Table 18*). Group 1, clinically idiopathic nystagmus with full phenotyping, was defined by the presence of nystagmus with a predominantly horizontal waveform as determined through eye tracking, normal electroretinogram (ERG) findings, a normal fovea determined via optical coherence tomography (OCT), visual evoked potentials (VEP) that suggests normal chaismal routing, and no iris transillumination as determined via slit lamp microscope (n=18). Group 2 was defined as the same features of group one, but allowing for unknowns where the phenotyping has not been performed (n=15). Group 3, clinically consistent with albinism with full phenotyping, was defined

as probands with nystagmus, normal ERG findings, foveal hypoplasia determined via OCT, a VEP suggestive of chiasmal misrouting, and the presence of iris transillumination (n= 20). Group 4 was the same as group 3 in requiring both nystagmus and a normal ERG, but then differed to allow for a partial phenotype by only requiring just one of further feature suggestive of albinism (i.e. foveal hypoplasia, suggested misrouting, and/or iris transillumination), whether through negative ('normal') findings or unknowns (n= 28).

Table 18. Clinical features used as the selection criteria for each of the four phenotype sub-groups. Unknown refers to no phenotyping attempted due to time limits or patient age, whereas as unclear refers to inconclusive results. Where multiple options are within a box the phenotype can be either or. Group 4 has the further criterion that one of the three features highlighted in red must be present, e.g. OCT: FH/ VEP-misrouting: Yes/ Iris transillumination: Yes.

Cohort sub-group number	**Cohort sub-group**	**Proband**	**Nystagmus**	**Predominant waveform direction**	**ERG**	**OCT**	**VEP-misrouting suggested**	**Iris trans-illumination**
1	Clinically idiopathic nystagmus (with full phenotyping)	Yes	Yes	Horizontal	Normal	Normal	No	No
2	Clinically idiopathic nystagmus (with incomplete phenotyping)	Yes	Yes	Horizontal/ Unknown	Normal/ Unknown	Normal/ Unknown/ Unclear	No/ Unknown/ Unclear	No/ Unknown/ Unclear
3	Clinically consistent with Albinism (with full phenotyping)	Yes	Yes	-	Normal	FH	Yes	Yes
4	Clinical features suggestive of Albinism (with incomplete phenotype or missing phenotype data) *must include at least one of the features highlighted in red	Yes	Yes	-	Normal/ Unknown	FH*/ Unknown/ Normal	Yes*/ Unknown/ No	Yes*/ Unknown/ No

5.4.2 NGS data

Proband DNA samples were processed for NGS sequencing following the protocols described in methods section 2.3 for the Illumina TruSight One v1.0 targeted sequencing panel and the Illumina NextSeq 550 sequencer. Following sequencing, the raw data output underwent processing and quality checks by the bioinformatician Luke O'Gorman (following the protocol in section 2.4). Sample data was compiled from five separate sequencing batches, with only samples passing quality control thresholds admitted (as described in section 2.4).

5.4.3 Analysis of variants

The UKGTN approved gene panel for albinism and nystagmus was used to prioritise genes for identification of pathogenic and likely pathogenic causal mutations (*Table 17*).

The two categories of 'pathogenic' and 'likely pathogenic' were used to prioritise variants. A variant was categorised as 'pathogenic' if it had a 'pathogenic' annotation in ClinVar or in InterVar or was listed as a 'disease causing mutation (DM)' in HGMD. A variant was categorised as 'likely pathogenic' if it fit the following criteria: not synonymous, an allele frequency of less than or equal to 5% in 1000 Genomes Project all populations (1000g_all), Exome Sequencing Project 6500 all populations (esp6500_all) and Exome Aggregation Consortium all populations (ExAC_all), and had either a CADDPhred score greater than or equal to 15 or an absolute MaxEntScan score greater than or equal to 3. The pathogenic or likely pathogenic variants were then considered causal if the known inheritance pattern was compatible with the zygosity of the variant(s) (within the same gene) e.g. the inheritance of OCA3 is autosomal recessive so two variants are required within the TYRP1 gene for a causal genotype.

5.4.4 Sanger sequencing of variants

Sanger sequencing was performed for a region of the *PAX6* gene containing a variant with a quality control failure rate greater than 5% across the cohort (methods section 2.5).The parents of NG420 were also Sanger sequenced to determine the presence of the variants *OCA2* p.V433I, *TYR* p.S129Y, and *TYR* R402Q (methods section 2.5). Results were then used to a create a pedigree tree to determine the segregation of variants.

5.5 Results

5.5.1 NGS run metrics

Sequencing metrics are useful in determining the likelihood of reliable data produced from the sequencing run. Within the Illumina online tool BaseSpace, two particular metrics of note are the cluster density on the flow cell and the Q score distribution (methods section 2.3.3). Due to a limited number of samples per sequencing kit the sequencing was carried out across multiple runs. All runs bar one had a good cluster density averaging roughly 230 K/mm^2 and a high Q30 generally around 80%. One run appeared to fail, showing very poor run metrics (*Figure 42*).

The much higher cluster density in lanes 1 and 2 of run B has negatively affected the Q score distribution of this run (*Figure 42*). The cause was a faulty index primer used to label and pool 12 samples. The data was too poor to pass bioinformatics quality checks and these 12 samples were excluded from analysis within this project. The mean read depth across all samples was determined by the bioinformatician Luke O'Gorman and was 127.5X with 90.6% of all target regions achieving a depth of 20X or greater (Supplementary table A.2.1).

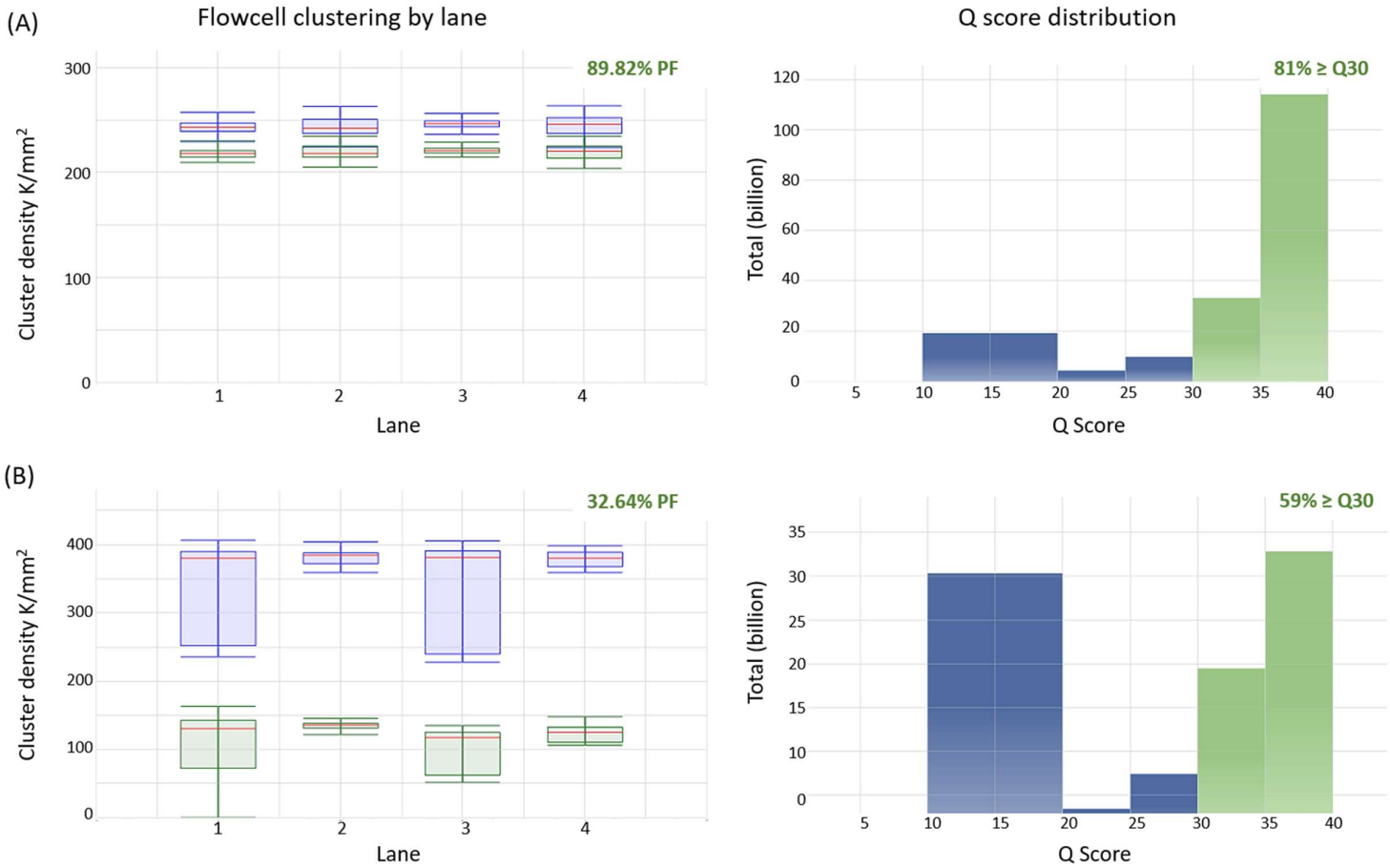


Figure 42. Comparison of run metrics between a high quality run (A) and poor quality run (B) recorded on Basespace sequencing hub (Illumina online tool). The cluster density on each lane of the flowcell is shown for each run on the left and the percentage of clusters passing filters (%PF) is printed in green on each boxplot. The Phred quality score (Q score) distribution for each run is on the right. If a base has a Q score of 30, the probability of the base call being accurate is 99.9%. The percentage of bases with a Q ≥ 30 is printed in green on each histogram.

5.5.2 Sanger sequencing of false positives within the *PAX6* gene

During the bioinformatic quality control steps, two *PAX6* variants were identified with high failure rates (miscalls by GATK haplotype caller in 27.7% and 24.7% of the entire cohort respectively). The queried *PAX6* variants, c.G1308T and c.C1306A, were not present when this codon (p.Q436) was Sanger sequenced in the six probands harbouring these variants (NG296, NG299, NG327, NG333, NG399 and NG429). The sequencing data was slightly poor for NG327 and NG429 however the chromatogram still appears to show wildtype sequence (*Figure 43*). For this reason, these *PAX6* variants were excluded as false positives, alongside a *HPS6* variant with a failure rate greater than 50% (GATK haplotype caller 53.1%).

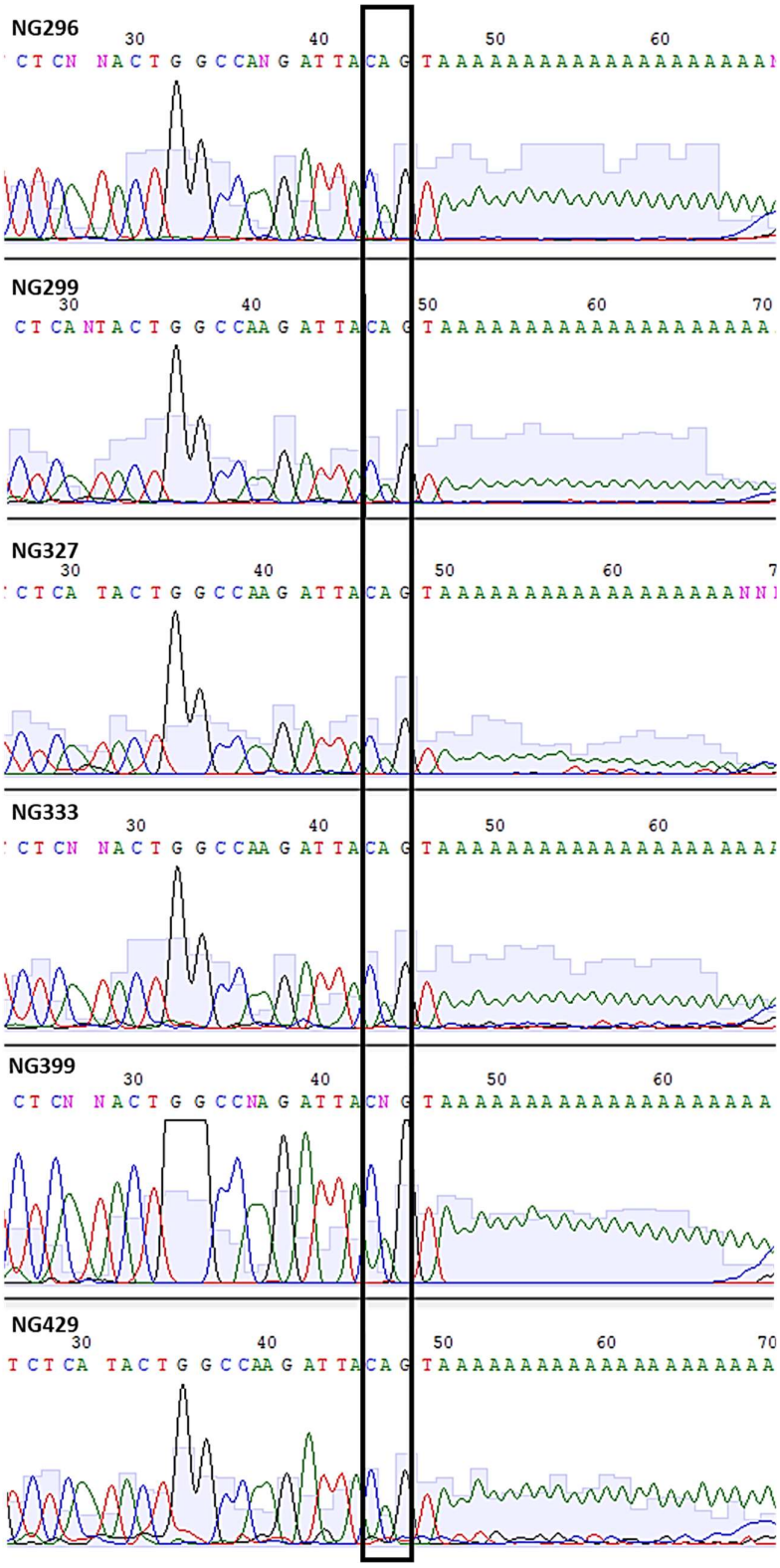


Figure 43. Sanger sequencing chromatograms for the samples queried for the presence of c.G1308T and c.C1306A in the *PAX6* gene. The codon containing c.G1308T (p.Q436H) and c.C1306A (p.Q436K) is highlighted in the black box. The variants are not present in NG296, NG299, NG327, NG333 NG399 and NG429.

5.5.3 Genetic resolution of a nystagmus and albinism cohort

A total of 46 variants identified in this cohort of 81 probands met the criteria for assumed 'pathogenic' genetic variants. A further 70 variants met the criteria for assumed 'likely pathogenic' genetic variants. The results have been split by their nystagmus and albinism phenotype groups.

5.5.3.1 Pathogenic variants identified in clinically idiopathic nystagmus phenotype groups

'Pathogenic' variants were identified in 18 probands from the clinically idiopathic nystagmus phenotype groups 1 and 2 (total n=33) (*Table 19*). Of these 18, six probands were found to have causal genotypes that followed the known inheritance pattern, four in the full phenotype group and two in the partial phenotype group. The causal genotypes were identified across four genes from the 31 gene panel (*TYR*, *OCA2*, *CACNA1F* and *FRMD7*). The high ratio of known pathogenic variants to identified causal genotype is partly due to the inclusion of the common *TYR* variant R402Q, which has a MAF of 28% (1000g_EUR). Even though the variant is common within European populations it is rated as pathogenic by ClinVar as it is thought to be causal when in a tri-allelic albinism genotype [166]. Due to the low penetrance of the R402Q variant it has been excluded from initial genotype analysis but included within the data to allow for analysis later on.

Two probands (group 1), NG335 and NG443, carry known pathogenic hemizygous mutations within the X-linked IIN gene *FRMD7*. One proband (group 1) harbours a hemizygous mutation in the X-linked gene *CACNA1F*. The proband NG258 is within the completely phenotyped clinically idiopathic nystagmus group but has two compound heterozygous mutations within *TYR* which is a causal genotype of OCA1. The partially phenotyped proband NG315 (group 2) harbours two compound heterozygous mutations in *OCA2.* The final causal genotype is due to a second hemizygous mutation in *CACNA1F* carried by NG381 (group 2) (*Table 19*).

Table 19. Pathogenic variants found within samples from the clinically idiopathic nystagmus phenotype groups (groups 1 and 2) . Table shows the 18 samples identified as carrying 'pathogenic' variants. Pathogenic variants were determined to be 'causal' based on the known inheritance pattern and marked with a cross (X). Yellow indicates heterozygous variants, red indicates homozygous variants or variant on the X chromosome in a male, and grey highlights the R402Q *TYR* variants involved in the tri-allelic genotype investigated by Norman *et al*. Male probands are in highlighted in blue and female probands are highlighted in pink. NCBI transcript accessions numbers: *TYR* NM_000372, *MLPH* NM_001281474, *OCA2* NM_000275, *TULP1* NM_003322, *CACNA1F* NM_001256790, *FRMD7* NM_194277, *SACS* NM_014363.

Chr.	Gene	Mutation type	Variant	ClinVar	InterVar	HGMD	Phenotype group 1									Phenotype group 2									
							NG172	NG264	NG335	NG433	NG456	NG477	NG506	NG528	NG534	NG296	NG306	NG315	NG327	NG367	NG381	NG411	NG449	NG472
11	*TYR*	nonsynonymous	c.C1217T p.P406L	P		DM								X											
11	*TYR*	nonsynonymous	c.C1291A p.P431T			DM								X											
11	*TYR*	nonsynonymous	c.G1205A p.R402Q	P																					
2	*MLPH*	nonsynonymous	c.G104A p.R35Q			DM																			
15	*OCA2*	nonsynonymous	c.G1327A p.V443I	P		DM												X							
15	*OCA2*	nonsynonymous	c.A1025G p.Y342C			DM																			
15	*OCA2*	splicing	c.574-19A>G			DM												X							
6	*TULP1*	splicing	c.1112+2T>G		P																				
X	*CACNA1F*	nonsynonymous	c.G1556A p.R519Q			DM						X													
X	*CACNA1F*	splicing	c.1309+1G>C		P																X				
X	*FRMD7*	stopgain	c.C1003T p.R335X	P	P	DM			X																
X	*FRMD7*	nonsynonymous	c.G70A p.G24R	P		DM				X															
13	*SACS*	nonsynonymous	c.A4466G p.N1489S			DM																			

5.5.3.2 Likely pathogenic variants identified in clinically idiopathic nystagmus phenotype group 1

On addition of 'likely pathogenic' variants to the completely phenotyped clinically idiopathic nystagmus group, there were 18 further variants but only two additional likely causal genotypes. Proband NG299 was found to harbour a likely pathogenic variant within the *FRMD7* gene and a second heterozygous *HPS5* variant was identified in NG445 (*Table 20*). There was also a hemizygous *CACNA1F* mutation in NG528 which already carried two mutations in the *TYR* gene (*Table 19*). The *CACNA1F* gene could be a better candidate for the apparent IIN phenotype exhibited by this proband, however, the likely pathogenic variant is a novel variant and is not considered as dependable as the proven *TYR* mutations.

Table 20. Pathogenic and likely pathogenic variants identified in samples from the clinically idiopathic nystagmus phenotype group 1 (16 probands) . Pathogenic (X) and likely pathogenic (\) variants were determined to be 'causal' based on the known inheritance pattern and marked with a cross or a dash respectively. Yellow indicates pathogenic heterozygous variants, red indicates homozygous or hemizygous variants, and pale yellow/red indicates the same for likely pathogenic variants. Male probands are highlighted in blue and female probands are highlighted in pink. The common R402Q *TYR* variant is highlighted in grey. NCBI transcript accessions numbers: *LYST* NM_000081, *HPS5* NM_007216, *AP3B1* NM_001271769, *DTNBP1* NM_032122, *TYR* NM_000372, *OCA2* NM_000275, *SACS* NM_014363, *MYO5A* NM_000259, *BLOC1S3* NM_212550, *HPS3* NM_032383, *CASK* NM_001126055, *CACNA1F* NM_001256790, *FRMD7* NM_194277.

Chr.	Gene	Mutation type	Variant	1000g	esp6500	ExAC	CADD-phred	MaxEnt-Scan	NG172	NG264	NG288	NG289	NG299	NG335	NG338	NG433	NG445	NG456	NG477	NG506	NG528	NG534	NG554	NG558
chr1	*LYST*	nonsynonymous	c.C8960G p.P2987R		0.000	0.000	26.3																	
chr1	*LYST*	synonymous	c.C6078T p.Y2026Y	0.001		0.001																		
chr11	*HPS5*	nonsynonymous	c.C2951T p.T984I	0.010	0.027	0.024	18.61										\							
chr11	*HPS5*	nonsynonymous	c.G413A p.G138E			0.000	32										\							
chr11	*TYR*	nonsynonymous	c.C1217T p.P406L	0.002	0.004	0.004	32														X			
chr11	*TYR*	nonsynonymous	c.G1205A p.R402Q	0.081	0.203	0.177	34																	
chr11	*TYR*	nonsynonymous	c.C1291A p.P431T		0.000		28.3														X			
chr13	*SACS*	nonsynonymous	c.A13409G .N4470S			0.000	22.7																	
chr13	*SACS*	nonsynonymous	c.C11032G .P3678A	0.044	0.017	0.040	25.9																	
chr13	*SACS*	nonsynonymous	c.C8393A p.P2798Q	0.000	0.003	0.002	26.5																	
chr13	*SACS*	nonsynonymous	c.G5848A p.D1950N	0.001	0.000	0.000	17.11																	
chr13	*SACS*	nonsynonymous	c.A4466G p.N1489S	0.007	0.007	0.009	0.002																	
chr13	*SACS*	nonsynonymous	c.G2080A p.A694T	0.008	0.026	0.023	15																	
chr13	*SACS*	nonsynonymous	c.G304C p.D102H				27.1																	
chr15	*OCA2*	nonsynonymous	c.G1897C p.V633L				29.6																	
chr15	*OCA2*	splicing	c.574-19A>G	0.006	0.007	0.006		0.718																
chr15	*MYO5A*	nonsynonymous	c.A3960T p.R1320S	0.012	0.030	0.034	21.7																	
chr19	*CACNA1A*	nonsynonymous	c.C3409G p.P1137A	0.000	0.001	0.000	23.1																	
chr19	*BLOC1S3*	nonsynonymous	c.C322G p.L108V	0.007		0.012	23.9																	
chr3	*HPS3*	nonsynonymous	c.C2692T p.R898C	0.001		0.001	28.4																	
chr5	*AP3B1*	nonsynonymous	c.G875A p.R292H	0.000	0.001	0.000	35																	
chr6	*DTNBP1*	nonsynonymous	c.C814T p.P272S	0.016	0.037	0.044	19.87																	
chrX	*CASK*	nonsynonymous	c.C1700T p.T567I		0.000	0.000	23.8																	
chrX	*CACNA1F*	nonsynonymous	c.G1361A p.R454Q	0.008	0.017	0.030	33																	
chrX	*CACNA1F*	nonsynonymous	c.T566C p.I189T	0.001	0.000	0.000	26.1																	
chrX	*FRMD7*	stopgain	c.C1003T p.R335X			0.000	39							X										

chrX	*FRMD7*	nonsynonymous	c.T790G p.C264G				25.5																
chrX	*FRMD7*	nonsynonymous	c.G782A p.R261Q				34																
chrX	*FRMD7*	nonsynonymous	c.G70A p.G24R				29.6																

5.5.3.3 Likely pathogenic variants identified in clinically idiopathic nystagmus phenotype group 2

Pathogenic and likely pathogenic variants for the clinically idiopathic group with incomplete phenotyping are shown in *Table 21*. In addition to the pathogenic variants in *Table 19* there are 30 likely pathogenic variants across a total of 15 probands. However, according to assumed inheritance patterns only three more causal genotypes have been identified. The probands NG318 and NG327 have a very similar genotype of two compound heterozygous S*ACS* likely pathogenic variants plus a likely pathogenic variant in the dominant gene *CACNA1A*. The variants between the probands different but the similar overall genotype strongly suggests spastic/spinocerebellar ataxia for both probands. A likely pathogenic variant was also found within *FRMD7* in proband NG383. A second *TYRP1* variant was identified as likely pathogenic in NG296 suggesting an OCA3 genotype when in combination with the pathogenic *TYRP1* mutation first shown in *Table 19*. After inclusion of likely pathogenic variants the UKGTN gene panel is able to diagnose 6/18 probands with partial phenotyping and the same number with the full phenotype. In total 12 causal genotypes have been identified in 36 probands with clinically idiopathic nystagmus phenotypes.

Table 21. Pathogenic and likely pathogenic variants identified in samples from the clinically idiopathic nystagmus phenotype group 2 (16 probands). Pathogenic (X) and likely pathogenic (\) variants were determined to be 'causal' based on the known inheritance pattern and marked with a cross or a dash respectively. Yellow indicates pathogenic heterozygous variants, red indicates homozygous or hemizygous variants, and pale yellow/red indicates the same for likely pathogenic variants. Male probands are highlighted in blue and female probands are highlighted in pink. The common R402Q *TYR* variant is highlighted in grey. NCBI transcript accessions numbers: *LYST* NM_000081, *HPS1* NM_000195, *AP3B1* NM_001271769, *DTNBP1* NM_032122, *TYR* NM_000372, *OCA2* NM_000275, *SACS* NM_014363, *MYO5A* NM_000259, *BLOC1S3* NM_212550, *HPS3* NM_032383, *HPS4* NM_152841, *CACNA1F* NM_001256790, *TYRP1* NM_000550, *MLPH* NM_001281474, *TULP1* NM_003322, *SETX* NM_015046, *CACNA1A* NM_001127221, *FRMD7* NM_194277.

Chr.	Gene	Mutation type	Variant	1000g	esp6500	ExAC	CADDphred	MaxEntScan	NG190	NG191	NG232	NG296	NG306	NG315	NG318	NG327	NG367	NG381	NG383	NG411	NG438	NG449	NG472
chr1	*LYST*	nonsynonymous	c.A9017G p.K3006R	0.000799	0.001	0.001	23.9																	
chr1	*LYST*	nonsynonymous	c.C5945T p.T1982I	0.001597	0.008	0.006	17.06																	
chr10	*HPS1*	nonsynonymous	c.G1915A p.G639S	0.000399	0.001	0.001	33																	
chr11	*TYR*	nonsynonymous	c.G1205A p.R402Q	0.08127	0.203	0.177	34																	
chr13	*SACS*	nonsynonymous	c.A13717C p.N4573H	0.002995	0.004	0.003	25.5									\								
chr13	*SACS*	nonsynonymous	c.G11324A p.S3775N				23.8																	
chr13	*SACS*	nonsynonymous	c.C11032G p.P3678A	0.04353	0.017	0.040	25.9									\								
chr13	*SACS*	nonsynonymous	c.C3739A p.Q1247K				21.8								\									
chr13	*SACS*	nonsynonymous	c.G2658T p.Q886H			0.000	24.3																	
chr13	*SACS*	nonsynonymous	c.G1941T p.K647N				23.6								\									
chr15	*OCA2*	nonsynonymous	c.A2242Tp.M748L			0.000	26.2																	
chr15	*OCA2*	nonsynonymous	c.G1327A:p.V443I	0.000799	0.005	0.003	34							X										
chr15	*OCA2*	splicing	c.574-19A>G	0.00619	0.007	0.006		0.718						X										
chr15	*MYO5A*	splicing	c.4480+2C>T		0.000	0.000	24.5																	
chr15	*MYO5A*	nonsynonymous	c.A3960T p.R1320S	0.011582	0.030	0.034	21.7																	
chr19	*CACNA1A*	nonsynonymous	c.G3040A p.E1014K	0.001198	0.004	0.003	16.78								\									
chr19	*CACNA1A*	nonsynonymous	c.A2192C p.E731A	0.003994	0.010	0.010	24.8									\								
chr19	*BLOC1S3*	nonsynonymous	c.C322G p.L108V	0.007388		0.012	23.9																	
chr19	*BLOC1S3*	nonsynonymous	c.G335A p.R112Q				26.5																	
chr2	*MLPH*	nonsynonymous	c.G104A p.R35Q				33																	
chr22	*HPS4*	nonsynonymous	c.G537T p.Q179H				27.6																	
chr3	*HPS3*	nonsynonymous	c.G2215A p.G739R	0.003794	0.012	0.010	22.9																	
chr5	*AP3B1*	nonsynonymous	c.G2848A p.V950M	0.001597	0.007	0.004	16.41																	
chr6	*DTNBP1*	nonsynonymous	c.C814T p.P272S	0.015575	0.037	0.044	19.87									\								
chr6	*TULP1*	splicing	c.1112+2T>G				23.8	7.647																
chr9	*TYRP1*	nonsynonymous	c.G70A p.A24T	0.000799	0.002	0.002	15.91					\												

chr9	*TYRP1*	nonsynonymous	c.C457T p.R153C		0.001	0.001	33																
chr9	*SETX*	nonsynonymous	c.T4660G p.C1554G	0.005791	0.003	0.006	21.6																
chr9	*SETX*	nonsynonymous	c.T472G p.L158V	0.001597	0.004	0.004	24.3																
chr9	*SETX*	nonsynonymous	c.A381C p.E127D				23.2																
chrX	*CACNA1F*	nonsynonymous	c.G1361A p.R454Q	0.007682	0.017	0.030	33																
chrX	*CACNA1F*	splicing	c.1114+1G>C				23.2																
chrX	*FRMD7*	nonsynonymous	c.G2044A p.G682S				28.8																

5.5.3.4 **Pathogenic variants in clinically consistent with albinism phenotype groups**

'Pathogenic' variants were identified across 39 probands from the clinically consistent with albinism phenotype groups 3 and 4 (total n= 48) (*Table 22*). In the fully phenotyped group 3 there were two causal genotypes identified. NG280 and NG540 were both found to be carrying compound heterozygous mutations in the *OCA2* gene. A third pathogenic mutation was also identified in NG280 in the same gene.

In the albinism group with only a partial phenotype there were nine causal genotypes identified (*Table 22*). NG195, NG391 and NG551 were each found to harbour heterozygous mutations in the dominant gene *PAX6* which is known to cause some features similar to albinism phenotypes. Two heterozygous *TYR* mutations were identified in NG395. One of these mutations is p.P406L which is considered an OCA1B mutation and may be why the proband presents a partial albinism phenotype [186]. There are two more OCA2 causal genotypes in this group, found in probands NG340 and NG416, both genotypes contain the *OCA2* mutation p.V443I. NG394 was identified as harbouring a homozygous pathogenic mutation in the *HPS5* gene. NG498 is carrying a dominant mutation in the *CACNA1A* gene for ataxia that can present similar to IIN in young children. The final causal genotype is in NG543 who is carrying a pathogenic variant in the *CACNA1F* gene providing a causal genotype for X-linked congenital stationary night-blindness.

Table 22. Pathogenic variants found within samples from the clinically consistent with albinism phenotype groups (groups 3 and 4). Table shows the 39 probands identified as carrying 'pathogenic' variants. Pathogenic variants were determined to be 'causal' based on the known inheritance pattern and marked with a cross (X). Yellow indicates heterozygous variants, red indicates homozygous variants or variant on the X chromosome in a male, and grey highlights the R402Q *TYR* variants involved in the tri-allelic genotype investigated by Norman *et al*. Male probands are in highlighted in blue and female probands are highlighted in pink. NCBI transcript accessions numbers: *HPS5* NM_007216, *PAX6* NM_001258464 *TYR* NM_000372, *OCA2* NM_000275, *SLC24A5* NM_205850, *CACNA1A* NM_001127221, *TULP1* NM_003322, *CACNA1F* NM_001256790.

Chr.	Gene	Mutation type	Variant	ClinVar	InterVar	HGMD	Phenotype group 3																Phenotype group 4																						
							NG167	NG178	NG198	NG263	NG265	NG280	NG356	NG420	NG454	NG483	NG492	NG512	NG521	NG524	NG530	NG540	NG559	NG195	NG270	NG309	NG333	NG340	NG348	NG361	NG391	NG394	NG395	NG406	NG416	NG429	NG430	NG441	NG498	NG514	NG536	NG543	NG545	NG548	NG551
chr11	*HPS5*	splicing	c.135+1G>A		P																										X														
chr11	*PAX6*	splicing	c.10+1G>A		P	DM																								X														
chr11	*PAX6*	nonsynonymous	c.C382T p.R128C	P		DM																																						X
chr11	*PAX6*	nonsynonymous	c.G295C p.A99P			DM																	X																					
chr11	TYR	nonsynonymous	c.C649T p.R217W	P		DM																																						
chr11	*TYR*	nonsynonymous	c.G816T p.W272C			DM																	X																					
chr11	*TYR*	nonsynonymous	c.C1217T p.P406L	P		DM																																						
chr11	*TYR*	nonsynonymous	c.G1336A p.G446S	P		DM																																						
chr11	*TYR*	nonsynonymous	c.G1205A p.R402Q	P																																								
chr15	*OCA2*	nonsynonymous	c.C593T p.P198L			DM																																						
chr15	*OCA2*	nonsynonymous	c.A1393G p.N465D	P		DM																																						
chr15	*OCA2*	nonsynonymous	c.G1327A p.V443I	P		DM																																						
chr15	*OCA2*	nonsynonymous	c.A1025G p.Y342C			DM																																						
chr15	*OCA2*	splicing	c.574-19A>G			DM																																						
chr15	*OCA2*	synonymous	c.C1113T p.G371G			DM																																						
chr15	*SLC24A5*	stopgain	c.T216G p.Y72X			DM																																						
chr19	*CACNA1A*	nonsynonymous	c.G1360 :p.A454T			DM																																						
chr6	*TULP1*	splicing	c.1112+2T>G		P																																						
chrX	*CACNA1F*	nonsynonymous	c.A2042C p.N681T			DM																																						
chrX	*CACNA1F*	nonsynonymous	c.G1361A p.R454Q			DM																																						

5.5.3.5 Likely pathogenic variants identified in clinically consistent with albinism phenotype group 3

On addition of 'likely pathogenic' variants to the clinically consistent with albinism group with full phenotyping, there are 30 more variants found within this group though only providing two further likely causal genotypes. NG521 harbours two likely pathogenic mutations within the *TYRP1* gene, which is a likely causal genotype for OCA3, and NG512 harbours two mutations within *OCA2*, a pathogenic and a likely pathogenic variant. There is a further *OCA2* variant considered likely pathogenic in NG280, adding to the already established OCA2 genotype (*Table 23*).

There also appears to be a pattern of probands carrying heterozygous mutations in more than one gene, particularly in the genes causing albinism. Compound heterozygous mutations across genes do not provide a traditional genotype but may be considered for their epistatic activity. NG198, NG213 and NG454 all have a mutation in both *TYR* and *OCA2*, and NG167 and NG530 each have a mutation in *HPS5* and *TYR* (*Table 23*). Furthermore, NG265 carries four heterozygous variants in different gene associated with forms of albinism; *LYST* (Chèdiak-Higashi syndrome), *TYRP1*, *OCA2*, *MYO5A* (Griscelli syndrome type 1). In total there are thirteen unresolved probands in group 3 that carry at least one pathogenic or likely pathogenic mutation in an albinism gene.

Table 23. Pathogenic and likely pathogenic variants identified in samples from the clinically consistent with albinism phenotype group 3 (20 probands). Pathogenic (X) and likely pathogenic (\) variants were determined to be 'causal' based on the known inheritance pattern and marked with a cross or a dash respectively. Yellow indicates pathogenic heterozygous variants, red indicates homozygous or hemizygous variants, and pale yellow/red indicates the same for likely pathogenic variants. Male probands are highlighted in blue and female probands are highlighted in pink. The common R402Q *TYR* variant is highlighted in grey. NCBI transcript accessions numbers: *LYST* NM_000081, *HPS5* NM_007216, *HPS3* NM_032383, *SLC45A2* NM_001012509, *AP3B1* NM_001271769, *DTNBP1* NM_032122, *PAX6* NM_001258464, *TYRP1* NM_000550, *SETX* NM_015046, *C10orf11* NM_001305581, *HPS1* NM_000195, *TYR* NM_000372, *OCA2* NM_000275, *SACS* NM_014363, *CACNA1A* NM_001127221, *SLC24A5* NM_205850, *MYO5A* NM_000259, *MLPH* NM_001281474, *TULP1* NM_003322, *BLOC1S3* NM_212550, *CASK* NM_001126055, *CACNA1F* NM_001256790, *FRMD7* NM_194277.

Chr.	Gene	Mutation type	Variant	1000g	esp6500	ExAC	CADD-phred	MaxEnt-Scan	NG167	NG178	NG198	NG213	NG263	NG265	NG280	NG356	NG420	NG454	NG483	NG492	NG508	NG512	NG518	NG521	NG524	NG530	NG536	NG540	NG559
chr1	*LYST*	nonsynonymous	c.C7870T p.R2624W	0.000	0.003	0.003	31																						
chr1	*LYST*	nonsynonymous	c.A650T p.D217V			0.000	23.6																						
chr11	*HPS5*	nonsynonymous	c.C2951T p.T984I	0.010	0.027	0.024	18.61																						
chr11	*HPS5*	nonsynonymous	c.G1138T p.D380Y			0.000	27.8																						
chr3	*HPS3*	nonsynonymous	c.G2215A p.G739R	0.004	0.012	0.010	22.9																						
chr5	*AP3B1*	nonsynonymous	c.G2848A p.V950M	0.002	0.007	0.004	16.41																						
chr6	*DTNBP1*	nonsynonymous	c.C814T p.P272S	0.016	0.037	0.044	19.87																						
chr9	*TYRP1*	nonsynonymous	c.C80G p.P27R		0.000	0.000	24.5																		\				
chr9	*TYRP1*	nonsynonymous	c.C1037G p.P346R		0.000	0.000	31																		\				
chr9	*SETX*	nonsynonymous	c.T4660G p.C1554G	0.006	0.003	0.006	21.6																						
chr9	*SETX*	nonsynonymous	c.G2216A p.G739E	0.004	0.003	0.001	25.2																						
chr11	*TYR*	nonsynonymous	c.G230A p.R77Q	0.000	0.000	0.000	33																						
chr11	*TYR*	nonsynonymous	c.G529T p.V177F	0.000		0.000	26.7																						
chr11	*TYR*	nonsynonymous	c.G816T p.W272C				29.6																						
chr11	*TYR*	nonsynonymous	c.G896A p.R299H	0.000		0.000	33																						
chr11	*TYR*	splicing	c.1037-1G>A			0.000	25.2																						
chr11	*TYR*	nonsynonymous	c.C1099T p.H367Y			0.000	29.5																						
chr11	*TYR*	stopgain	c.C1204T p.R402X			0.000	51																						

chr11	*TYR*	nonsynonymous	c.C1264T p.R422W			0.000	34	
chr11	*TYR*	nonsynonymous	c.G1205A p.R402Q	0.081	0.203	0.177	34	
chr15	*OCA2*	nonsynonymous	c.A2480G p.Y827C				24.5	
chr15	*OCA2*	nonsynonymous	c.A2341G p.N781D				26.7	
chr15	*OCA2*	nonsynonymous	c.C2228T p.P743L			0.000	32	
chr15	*OCA2*	nonsynonymous	c.C2020G p.L674V	0.001		0.000	27.3	
chr15	*OCA2*	nonsynonymous	c.A1465G p.N489D		0.001	0.000	28.2	
chr15	*OCA2*	nonsynonymous	c.G1327A p.V443I	0.001	0.005	0.003	34	
chr15	*OCA2*	nonsynonymous	c.A1025G p.Y342C		0.000	0.000	24.3	
chr13	*SACS*	nonsynonymous	c.C8393A p.P2798Q	0.000	0.003	0.002	26.5	
chr15	*OCA2*	nonsynonymous	c.A868G p.R290G			0.000	22.6	
chr15	*OCA2*	nonsynonymous	c.C593T p.P198L	0.002	0.000	0.000	29	
chr15	*OCA2*	splicing	c.574-19A>G	0.006	0.007	0.006		
chr15	*SLC24A5*	stopgain	c.T216G p.Y72X		0.000	0.000	35	
chr15	*MYO5A*	nonsynonymous	c.C4279T p.R1427C		0.000	0.000	35	
chr15	*MYO5A*	nonsynonymous	c.A3960T p.R1320S	0.012	0.030	0.034	21.7	
chr15	*MYO5A*	nonsynonymous	c.G3136A p.V1046M	0.000	0.001	0.001	20.8	
chr2	*MLPH*	nonsynonymous	c.C415T p.R139W	0.014	0.001	0.012	34	
chr6	*TULP1*	splicing	c.1112+2T>G				23.8	
chr19	*BLOC1S3*	nonsynonymous	c.C322G p.L108V	0.007		0.012	23.9	
chrX	*CACNA1F*	nonsynonymous	c.G1361A p.R454Q	0.008	0.017	0.030	33	

5.5.3.6 Likely pathogenic variants identified in clinically consistent with albinism phenotype group 4

There were 47 pathogenic and likely pathogenic variants shared across 27 probands in the partially phenotyped clinically consistent with albinism group (group 4). There was only one proband in both the clinically consistent with albinism groups that had no pathogenic or likely pathogenic variants identified. There were a further in three cases where the only variant identified was the common SNP p.R402Q (NG284, NG524 and NG548).

In addition to the initial nine causal genotypes identified, there is a single further causal genotype in NG386. NG386 is homozygous for a likely pathogenic *CACNA1A* variant. Within group 4 there are nine probands with partially resolved albinism genotypes. Each of these nine samples were found to harbour a pathogenic or likely pathogenic mutation in at least one albinism gene but did not have a causal genotype (*Table 24*). For example, NG270 has an assumed pathogenic mutation in both *HPS5* and *TYR*, as well as a likely pathogenic mutation in *TYRP1*. NG348 also has a multiple variants in albinism genes with a heterozygous variant in *SLC45A2* and another variant in the syndromic albinism gene *MLPH* (Griscelli Syndrome type 3).

Table 24. Pathogenic and likely pathogenic variants identified in samples from the clinically consistent with albinism phenotype group 4 (27 probands). Pathogenic (X) and likely pathogenic (\) variants were determined to be 'causal' based on the known inheritance pattern and marked with a cross or a dash respectively. Yellow indicates pathogenic heterozygous variants, red indicates homozygous or hemizygous variants, and pale yellow/red indicates the same for likely pathogenic variants. Male probands are highlighted in blue and female probands are highlighted in pink. The common R402Q *TYR* variant is highlighted in grey. NCBI transcript accessions numbers: *LYST* NM_000081, *HPS5* NM_007216, *HPS3* NM_032383, *SLC45A2* NM_001012509, *AP3B1* NM_001271769, *DTNBP1* NM_032122, *PAX6* NM_001258464, *TYRP1* NM_000550, *SETX* NM_015046, *C10orf11* NM_001305581, *HPS1* NM_000195, *TYR* NM_000372, *OCA2* NM_000275, *SACS* NM_014363, *CACNA1A* NM_001127221, *SLC24A5* NM_205850, *MYO5A* NM_000259, *MLPH* NM_001281474, *TULP1* NM_003322, *BLOC1S3* NM_212550, *CASK* NM_001126055, *CACNA1F* NM_001256790, *FRMD7* NM_194277.

Chr.	Gene	Mutation type	Variant	1000g	esp6500	ExAC	CADD-phred	MaxEnt-Scan	NG195	NG270	NG272	NG309	NG333	NG340	NG348	NG361	NG386	NG391	NG394	NG395	NG399	NG406	NG416	NG423	NG429	NG430	NG441	NG474	NG498	NG514	NG536	NG543	NG545	NG548	NG551
chr1	*LYST*	nonsynonymous	c.A8426G p.E2809G	0.000	0.000	0.000	22													\															
chr11	*HPS5*	nonsynonymous	c.C2951T p.T984I	0.010	0.027	0.024	18.61			\			\																					
chr11	*HPS5*	nonsynonymous	c.G2703A p.M901I	0.002	0.007	0.005	20.9												X					\										
chr11	*HPS5*	splicing	c.135+1G>A				27.2												X															
chr3	*HPS3*	nonsynonymous	c.G2215A p.G739R	0.004	0.012	0.010	22.9									\					\													
chr5	*SLC45A2*	nonsynonymous	c.G686A p.C229Y			0.000	25.3								\																			
chr5	*AP3B1*	nonsynonymous	c.C2514A p.F838L	0.007	0.008	0.008	22.4		\																									
chr6	*DTNBP1*	nonsynonymous	c.C814T p.P272S	0.016	0.037	0.044	19.87		\																							\	\		
chr11	*PAX6*	nonsynonymous	c.C382T p.R128C				34																											X
chr11	*PAX6*	nonsynonymous	c.G295C p.A99P				27.8		X																									
chr9	*TYRP1*	nonsynonymous	c.C1097T p.T366M			0.000	24.3														\													
chr9	*TYRP1*	nonsynonymous	c.G1114A p.A372T				15.97			\																									
chr9	*SETX*	nonsynonymous	c.T472G p.L158V	0.002	0.004	0.004	24.3														\													
chr9	*SETX*	nonsynonymous	c.A431G p.N144S			0.000	23																					\						
chr10	*C10orf11*	nonsynonymous	c.G480T p.L160F	0.004		0.002	18.46																					\						
chr10	*HPS1*	nonsynonymous	c.G317A p.R106Q			0.000	27.3																									\		
chr10	*HPS1*	nonsynonymous	c.T11C p.V4A	0.041	0.036	0.033	24.5											\										\						
chr11	*PAX6*	splicing	c.10+1G>A				27.2											X																
chr11	*TYR*	nonsynonymous	c.C649T p.R217W		0.000	0.000	25.1																								X			
chr11	*TYR*	nonsynonymous	c.C1064T p.A355V		0.000	0.000	26.8																				\							
chr11	*TYR*	nonsynonymous	c.C1217T p.P406L	0.002	0.004	0.004	32			X		X								X														

chr11	*TYR*	nonsynonymous	c.G1205A p.R402Q	0.081	0.203	0.177	34	
chr11	*TYR*	nonsynonymous	c.G1336A p.G446S			0.000	31	
chr15	*OCA2*	nonsynonymous	c.C2228T p.P743L			0.000	32	
chr15	*OCA2*	splicing	c.1842+2T>C				23.5	
chr15	*OCA2*	nonsynonymous	c.A1465G p.N489D		0.001	0.000	28.2	
chr15	*OCA2*	nonsynonymous	c.G1327A p.V443I	0.001	0.005	0.003	34	
chr15	*OCA2*	nonsynonymous	c.A1025G p.Y342C		0.000	0.000	24.3	
chr13	*SACS*	nonsynonymous	c.C11032G p.P3678A	0.044	0.017	0.040	25.9	
chr13	*SACS*	nonsynonymous	c.T8339G p.F2780C	0.001	0.004	0.003	26.1	
chr13	*SACS*	nonsynonymous	c.T8022A p.F2674L				23.3	
chr13	*SACS*	nonsynonymous	c.C5639T p.T1880I				27.6	
chr13	*SACS*	nonsynonymous	c.G2080A p.A694T	0.008	0.026	0.023	15	
chr15	*OCA2*	synonymous	c.C1113T:p.G371G	0.019	0.009	0.021		
chr19	*CACNA1A*	nonsynonymous	c.G2977A:p.E993K				25.7	
chr19	*CACNA1A*	nonsynonymous	c.A2195C p.E732A	0.004	0.010	0.010	24.8	
chr19	*CACNA1A*	nonsynonymous	c.G1360A p.A454T	0.002	0.004	0.005	28.2	
chr15	*MYO5A*	nonsynonymous	c.G5065A p.V1689I	0.001	0.003	0.002	22.7	
chr15	*MYO5A*	nonsynonymous	c.A3960T p.R1320S	0.012	0.030	0.034	21.7	
chr2	*MLPH*	nonsynonymous	c.C415T p.R139W	0.014	0.001	0.012	34	
chr6	*TULP1*	splicing	c.1112+2T>G				23.8	
chr19	*BLOC1S3*	nonsynonymous	c.C322G:p.L108V	0.007		0.012	23.9	
chrX	*CASK*	nonsynonymous	c.G2210A p.R737Q	0.001	0.001	0.000	23.9	
chrX	*CACNA1F*	nonsynonymous	c.G3604A p.E1202K				23.2	
chrX	*CACNA1F*	nonsynonymous	c.A2042C p.N681T	0.002	0.002	0.002	26.1	
chrX	*CACNA1F*	nonsynonymous	c.G1361A p.R454Q	0.008	0.017	0.030	33	
chrX	*FRMD7*	stopgain	c.C910T p.R304X				37	

5.5.3.7 **Tri-allelic OCA1B theory helps to resolve partial albinism genotypes**

Across the clinically consistent with albinism phenotype groups there are 31 cases with at a single mutation in an autosomal recessive albinism gene, providing only a partially resolved genotype. There is a theory of tri-allelic inheritance to explain the high level of missing heritability in OCA1 [166]. Evidence has suggested the common variants p.R402Q and p.S192Y can provide a causal genotype when combined with another pathogenic/likely pathogenic mutation. In this cohort, the incidence of S192Y is 51.2% and the incidence of R402Q is 27.7%. A split of incidence by phenotype reveals a particularly high occurrence of R402Q in the albinism groups, with a MAF of 35.4%, and slight enrichment of S192Y, with a MAF of 55.2%. NG420 is a further case with a genotype of interest. This probands harbours a single pathogenic *OCA2* variant but is also homozygous both R402Q and S192Y. NG420 is not included in *Table 25* as no other *TYR* variants were identified in this proband but is shown in *Figure 44*.

When the tri-allelic genotype is included in the variant prioritisation step, nine of the twelve probands carrying a single *TYR* mutation are resolved (*Table 25*). All probands identified as partially resolved, carrying a single heterozygous *TYR* variant were from the clinically consistent with albinism phenotype groups.

Table 25. Of the twelve probands carrying a single pathogenic of likely pathogenic mutation in the *TYR* gene, nine can be resolved with the tri-allelic genotype described by Norman *et al.*. Probands carrying a single TYR mutation from 'clinically consistent with albinism' phenotype groups 3 and 4 are shown, as well as, the common *TYR* variants S192Y and R402Q (highlighted in dark grey and pale grey respectively). Yellow indicates heterozygous variants, red indicates homozygous variants. The presence of both common variants plus another *TYR* variant indicates a causal tri-allelic OCA1 genotype, which is signified by the letter 'T'. NCBI transcript accessions number: *TYR* NM_000372.

Gene	Mutation type	Variant	1000g	esp6500	ExAC	CADD-phred	MaxEnt-Scan	ClinVar	InterVar	HGMD	Phenotype group 3									Phenotype group 4		
											NG198	NG213	NG263	NG356	NG454	NG483	NG530	NG559	NG270	NG309	NG441	NG536
TYR	nonsynonymous	c.G230A p.R77Q	0.000	0.000	0.000	33		P		DM												
TYR	nonsynonymous	c.G529T p.V177F	0.000		0.000	26.7				DM												
TYR	nonsynonymous	c.C575A p.S192Y	0.123	0.275	0.252	25							T	T	T	T	T	T		T	T	T
TYR	nonsynonymous	c.C649T p.R217W		0.000	0.000	25.1		P		DM												T
TYR	nonsynonymous	c.G816T p.W272C				29.6				DM								T				
TYR	nonsynonymous	c.G896A p.R299H	0.000		0.000	33		P		DM					T							
TYR	splicing	c.1037-1G>A			0.000	25.2			P	DM							T					
TYR	nonsynonymous	c.C1064T p.A355V		0.000	0.000	26.8		P		DM											T	
TYR	nonsynonymous	c.C1099T p.H367Y			0.000	29.5				DM						T						
TYR	stopgain	c.C1204T p.R402X			0.000	51		P	P	DM			T									
TYR	nonsynonymous	c.C1217T p.P406L	0.002	0.004	0.004	32		P		DM										T		
TYR	nonsynonymous	c.C1264T p.R422W			0.000	34				DM				T								
TYR	nonsynonymous	c.G1205A p.R402Q	0.081	0.203	0.177	34		P					T	T	T	T	T	T		T	T	T
TYR	nonsynonymous	c.G1336A p.G446S			0.000	31		P		DM												

5.5.3.8 Segregation of variants in family #420

The proband NG420 was categorised as having a complete albinism phenotype. They were found to carry the known *OCA2* variant V433I, as well as being homozygous for the common *TYR* variants R402Q and S192Y. Sanger sequencing of these variants was carried out within parents of the proband to reveal how the variants segregate with phenotype (*Figure 44*). Both parents appeared to exhibit transillumination upon phenotyping but no other features of albinism (a VEP was not carried out). Father carries the common *TYR* variants and mother has the common *TYR* variants in combination with the *OCA2* mutation.

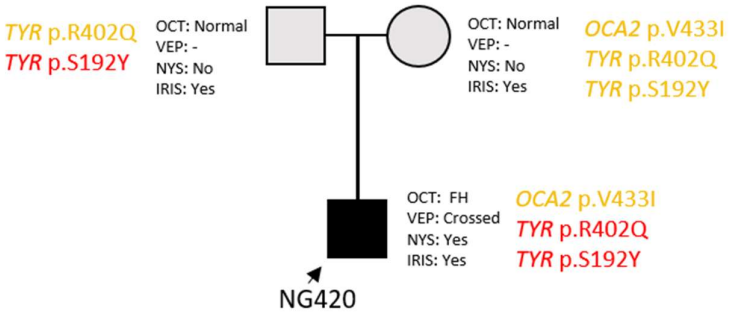


Figure 44. The segregation of a rare *OCA2* variant and the common *TYR* variants in NG420 through Sanger sequencing of parents.

5.5.4 Summary of diagnostic results

The number of samples assigned a causal and likely casual genotype is summarised in

Table 26. Overall, the diagnostic yield from this method of genetic analysis if 43.2% with a higher diagnostic rate in the albinism phenotype groups compared with the idiopathic nystagmus groups.

Table 26. Summary of diagnostic results outlining the number of causal genotypes identified in each phenotype sub-group and the genes responsible.

Group No.	**Phenotype Sub-group**	**No. of probands**	**Probands with 'pathogenic' genotypes**	**Probands with 'likely pathogenic' genotypes**	**Probands with *TYR* tri-allelic genotype**	**% causal genotype**	**Causal genes**
1	Clinically idiopathic nystagmus (with full phenotyping)	18	4	3	0	**38.9**	*CACNA1A, CACNA1F, FRMD7, HPS5, TYR.*
2	Clinically idiopathic nystagmus (with incomplete phenotyping)	15	2	3	0	**33.0**	*CACNA1A, CACNA1F, FRMD7, OCA2, SACS*
3	Clinically consistent with Albinism (with full phenotyping)	20	2	2	6	**50.0**	*OCA2, TYR, TYRP1*
4	Clinical features suggestive of Albinism (with incomplete phenotype/phenotype data)	28	9	1	3	**46.4**	*TYR, OCA2, PAX6, CACNA1A, CACNA1F, HPS5*

5.6 Discussion

In this study, detailed phenotyping methods have been combined with a clinical diagnostic gene panel to investigate their clinical utility and effect on diagnostic yield. The application of the UKGTN albinism and nystagmus 31 gene panel to the TruSight One sequencing data was able to provide a diagnostic yield of 43.2% in this cohort of 81 patients. A yield substantially higher than the majority of exome diagnostic analyses with the TruSight One panel (reported at close to 26%)[211]. Of the 31 genes on the panel, nine provided causal genotypes in this cohort with variants in the *TYR* gene accounting for the most at 13.5%, followed by *OCA2* with 7.4% and *CACNA1A* and *FRMD7* both accounting for 4.9% each. The greatest diagnostic yield came from patients with a complete phenotype characteristic of albinism (50% were assigned a causal genotype) and this was followed by patients with a partial albinism phenotype (either fully phenotyped with partial features or incompletely phenotyped) (46.4% were assigned a causal genotype).

The UKGTN 31 gene panel is described as a panel for nystagmus and albinism but it is particularly suited to albinism and apparent idiopathic INS as these phenotypes can have a large crossover. In the majority of cases, the patient phenotype and likely causal genotype matched. However, six patients were identified as having likely causal genotypes in genes that had not been previously associated with their phenotype group. For example, NG528 from phenotype group 1 (clinically idiopathic nystagmus with full phenotyping) was found to harbour compound heterozygous *TYR* mutations, providing a causal genotype for OCA1. Similarly, NG315 harboured a compound heterozygous OCA2 genotype and was also from group 1. These two examples reflect the heterogeneous and overlapping phenotypes often exhibited by young children with nystagmus and suggest that albinism cannot be ruled out when searching for the aetiology of apparent IIN. Even though, splitting the probands into phenotype subgroups has not been completely definitive, it does provide further confidence when the genotype coincides with the clinical phenotype. It is therefore important for infant patient phenotyping to be an ongoing process as the nystagmus develops, as features can be brought to light that provide further direction for the mutation search.

When in young children, apparent idiopathic INS or subtle feature of albinism may be indicative of congenital ataxias. For example, NG318 and NG327 were placed in the clinically idiopathic nystagmus phenotype group but both carried causal genotypes for spinocerebellar and spastic ataxia. Furthermore, NG386 had a phenotype clinically consistent with albinism but

was found to be homozygous for a likely pathogenic *CACNA1A* variant that provides a causal genotype for ataxia.

5.6.1 Effectiveness of the TruSight One enrichment sequencing panel

The TruSight One (TSO) enrichment panel has significant benefits over a custom-built capture panel such as the HaloPlex system, as the range of genes covered make it suitable for broad diagnostic use. Therefore, a large institution such as the NHS benefits from economies of scale if the panel were to be broadly implemented. The TSO sequencing panel has already been put forward and approved for use with other UKGTN screening panels e.g. 'NF1 and other disorder of the Ras/MAPK pathway' and 'NGS panel for Disorder of Sex Development'. For the genetic diagnosis of nystagmus, the broad gene coverage is particularly useful as varied and overlapping phenotypes mean a causal gene is difficult to pin point with clinical phenotyping alone. The data analysis is then optimised by only investigating the UKGTN approved 31 genes list, which streamlines analysis to the most likely candidates. The remaining genes covered by the TSO panel make this method of sequencing suited to research in that it provides data for genes that have potential to cause nystagmus but have not yet been identified as such. The TSO panel also lends itself to retrospective expansion as further genes become clinically relevant.

However, there are also disadvantages to using this pre-made capture panel. Firstly, Illumina NGS is most accurate and powerful when determining single nucleotide variants but cannot always be relied upon for correct determination of copy number variation[212]. It is possible to examine the data for copy number variants i.e. using software such as FishingCNV that compares coverage depth of the test sample with control samples (used to create a background distribution) to determine rare deletions or duplications in the test sample exome data [213]. Such software is still new and relatively untested so has been used in parallel with the more trusted MLPA method[208]. One further shortfall of the NGS technology used (and most NGS technologies) is that it produces short sequencing reads (≤151 bases), thus inaccuracies can occur when assembling the reads against a reference genome. This is particularly true in homopolymeric regions and these regions may suffer from false positives such as the *PAX6* p.Q286 seen here. Longer-read sequencing technology, such as Pacific Bioscience's SMRT sequencing, may greatly reduce some of this error. In 2013 Sikkema-Raddatz *et al*. suggested targeted NGS reads are just as sensitive and specific as Sanger sequencing [214], however Sanger sequencing remains the gold standard and is still used to confirm NGS data when reporting mutations to patients [110]. Finally, though the TSO panel appears to be the best targeted exome panel available for the required purpose there are still some improvements that could be made. The TSO panel was first designed based on OMIM data in 2015 which is continually updated to

include newly identified genes. One such gene, *SLC38A8*, has been shown as a likely candidate for isolated foveal hypoplasia and has recently been included in TSO version 3.0, but is yet to be included within the UKGTN 31 gene list[215].

As not all genes may be included on cheaper targeted panels, whole exome sequencing (WES) could be the best current solution for finding a genetic diagnosis unresolved patients. However, whole genome sequencing (WGS) is becoming even more popular than WES with a greater capacity for reduced costs and more power[216]. WGS may be a viable option in the future but the cost of greater data storage and in depth analysis must be considered. Furthermore, to optimise the search for causal mutations the parents of proband should also be sequenced to create data trios, following the same method as the 1000 genomes project[145].

5.6.2 Determining causal variants within the UKGTN 31 gene panel

It is important to streamline analysis of an extensive list of variants so only putative causal variants are investigated further. The criteria described in the ACMG guidelines is strict but does help to provide a confident, clinically call-able diagnosis. However, the ACMG guidelines are sometimes inappropriate for an albinism cohort as common variants are considered a causal genotype for OCA1B[166].

In this study, pathogenic variants were those that have been interpreted as significant either clinically or functionally and thus entered into one of the three databases: ClinVar, InterVar, or HGMD. Of these databases, HGMD is possibly considered the most prudent due to manual curation of variants from literature, however it is still known to misclassify some benign/low penetrance variants as disease-causing mutations[217]. ClinVar is often used a resource for clinical practise and has been found to have a good classification system when variants are reported after clinical testing[218]. However, the ClinVar database a high level of misclassification of low-penetrance variants as there is no single 'master' classification if different variant submissions are conflicting[218]. This is of particular importance for albinism phenotypes where penetrance is variable, and the resource interprets R402Q as 'likely pathogenic' with no caveat to its complex genotype. The InterVar database combines ratings (strong/ moderate/ supporting) for each factor conferring pathogenicity in an attempt to follow the ACMG guidelines. However, InterVar identified the lowest number of 'pathogenic' variants in this cohort. Including all three databases gives the greatest coverage of reported pathogenic variants, the only downside is the greater potential for falsely identifying benign variants as pathogenic.

The likely pathogenic variants were determined based upon their population frequency and pathogenicity score. CADD Phred was used as it is an overarching method combining information from 63 different annotations. The simpler pathogenicity scores Sift, PolyPhen and GERP are all included within this single tool [115]. The filter parameters of MAF ≤ 5% and CADD Phred ≥ 15 could be relaxed to lower the threshold for likely pathogenic variants and potentially increase the diagnostic yield. However, the number of false positives may also increase which is not suitable for clinical reporting. The answer is to resolve variants with functional and segregation studies to provide more tangible evidence of their effects.

5.6.3 Complex albinism genotypes

The tri-allelic OCA1B genotype accounted for 18.8% of the resolved cases with phenotypes clinically consistent with albinism (11.1% of the total cases resolved). Surprisingly, the tri-allelic genotype was more prevalent in group 3 with full phenotyping than in group 4 with partial phenotyping. However, whether the triallelic genotype results in OCA1A or OCA1B cannot be commented upon as only ocular phenotyping was performed skin pigmentation was not taken into account. Three probands (NG198, NG213 and NG270) with a pathogenic or likely pathogenic *TYR* variants have missing heritability but interestingly a second variant in another OCA gene has been identified in each proband. NG198 has a pathogenic variant in *OCA2*, NG213 has a likely pathogenic variant in *OCA2* and NG270 has a likely pathogenic variant in *TYRP1*. Genetic mutations throughout the pigmentary pathway are known to modify the phenotype of albinism. For example, mutations in the pigment gene *MC1R* can modify the OCA2 phenotype to produce red hair[219]. Similarly, Chiang *et al.* reported an OCA phenotype with red hair produced by mutations in the *TYRP1* and *OCA2* genes[143].

5.6.4 Conclusion

Currently the TruSight One clinical exome is a cost effective, efficient technology that can be utilised for genetic diagnosis and implemented throughout a large health service such as the NHS. The UKGTN albinism and nystagmus gene panel guides variant analysis to gene that would be clinically reportable and are suited to the patient phenotypes. The method of filtering for causal variants is currently strict but still provides a high diagnostic yield. The current method will improve as the predictive tools are perfected and more information is submitted to the databases.

The findings suggest that the diagnostic techniques used are effective in guiding diagnosis but only to a certain degree as this phenotypic group of INS patients is notoriously difficult to

pinpoint. Clinical phenotyping is important for exclusion of retinal dystrophy, known prematurity at birth, and neurological disorder. Further details are not necessary for use of this method, though may guide diagnosis in instances such as the tri-allelic genotype and its association with hypomorphic albinism.

Implementation of the method described here may significantly reduce the time and number of investigations that many children with INS undergo. It also enables a greater ability for family counselling in genetics and with coming to terms with the diagnosis. The study has also emphasised a need for greater research into apparent missing heritability in two main instances. Firstly, there are apparent INS phenotypes that seem to have no causal gene, and secondly, hypomorphic albinism phenotypes sometimes exhibit complex genotypes reflecting the important synergistic relationships between pigment genes.

Chapter 6: Discussion

Clinical evaluation of childhood Nystagmus can be very challenging as it can occur as either an isolated trait or in association with a very wide variety of ocular and systemic disorders. Isolated (or idiopathic) Infantile Nystagmus Syndrome has only been associated with one gene, *FRMD7*. However, there is a broad range of ocular and systemic disorders associated with INS (where other ophthalmic or systemic features are present) and the number of associated genes exceeds 800. The phenotypic heterogeneity of the associated disorders can cause an inflated number of 'idiopathic INS' diagnoses, where features indicative of the causative ocular disorder are too subtle to detect. The problem is exaserbated when patients are also too young to undergo robust clinical phenotyping. As idiopathic INS and albinism are the largest causative groups, improvements to diagnosis of these phenotypes is likely to have the greatest impact on INS diagnosis.

Hypomorphic albinism is a particularly difficult group to diagnose without a combination of detailed ophthalmic examination and genotyping. This is due to the disorder presenting with a range of features, the combination of which differs for patient to patient, and even within family groups carrying the same mutations[41, 166]. There is a particular phenotypic crossover with albinism, *PAX6* disease phenotypes and FHONDA. Often, a specific diagnosis may only be achieved through sequencing the associated causal gene(s). Furthermore, there is a high level of missing heritability in albinism. This can partially be accounted for by a tri-allelic genotype in the *TYR* gene, though many cases of missing heritability still remain unresolved and 'partial' diagnoses are common[166].

6.1 The tri-allelic inheritance of OCA1B

The segregation of *TYR* variants through family pedigrees (described in Chapter 3) provides positive evidence for a causal tri-allelic genotype involving a pathogenic variant in combination with two common SNPs. The theory that two common SNPs can be pathogenic is further validated by the occurrence of a proband with albinism (NG420) that is homozygous for both SNPs and carries no other *TYR* mutations (*Figure 44* in Chapter 5). It could therefore be argued that the tri-allelic theory extends to a causal double-variant haplotype (p.R402Q:S192Y) that is pathogenic when homozygous. The homozygous double-variant genotype has been previously noted in subjects with hypopigmentation but ophthalmic data was not reported[60]. On the other hand, the example of NG420 genotype is further complicated by the presence of an assumed 'pathogenic' *OCA2* mutation that could allude to epistasis between *TYR* and *OCA2*.

Since we first published evidence for tri-allelic inheritance of OCA1B there have been three other studies commenting on the genotype. The first, by Lasseaux *et al*., suggests the S192Y variant is not necessary for an OCA1B phenotype and describes 17 cases as compound heterozygous with only R402Q as the second variant (versus 31 with a tri-allelic genotype). However, they also state that the number of albinism cases with R402Q and S192Y in *cis* is much higher than the estimated frequency in the general population, suggesting that the p.[S192Y:R402Q] haplotype plays a causal role[167]. This was followed by a study looking closely at the possible additive effect of S192Y on the OCA1B phenotype. The group concluded there was no significant difference between the phenotype of patients with and without the S192Y variant. However, this is an extremely difficult distinction to make due to the high level of variability exhibited in OCA1. A point that that was further highlighted in this study's cohort[149]. The apparent lack of need for S192Y here may be due other unreported background variants that are not reported (i.e. in other OCA genes or filtered out in the bioinformatics process), though it may be possible that a highly pathogenic variant in combination with R402Q is sufficient to push the level of haploinsufficiency over a threshold. It could be that variable penetrance of the R402Q variant is the reason behind this long debate regarding its pathogenicity. The third publication, by Grønskov *et al.*, reports the identification of two putative pathogenic haplotypes, both of which include the allele p.[S192Y:R402Q]. Both haplotypes they describe contain three further variants, though the pathogenicity of these non-coding variants is unknown. The authors suggest the most likely explanation for these pathogenic haplotypes is the presence of S192Y and R402Q. They also describe a case of partial albinism caused by a homozygous p.[S192Y;R402Q] genotype. The overall conclusion from each of these studies is that either one of both of the common variants is likely to play a causal role in albinism when in a complex genotype, and consequently the prevalence of albinism has been substantially underestimated.

6.2 Functional assessment of potentially pathogenic variants

Understanding the functional effect of an ever-growing number of identified human variants is a significant challenge for bioinformatic processing and subsequent genetic counselling. Functional assays provide essential information on a variants pathogenicity and mechanism of action. This information can be necessary for making clinical decisions of diagnosis and determining an appropriate course of action. The tyrosinase gene has been studied in relation to skin cancer, albinism and general pigmentation. However, the encoding *TYR* gene still contains numerous variants of unknown, but often assumed, pathogenicity. It is important to determine the functional effect of these variants for the assignation of clinical genetic diagnosis.

Here, the functional consequences of the common TYR variants R402Q and S192Y were assessed. It was determined that both variants cause a reduction in enzymatic function as well as hinder the maturation of tyrosinase, likely causing ER retention and ER associated degradation. The reduction in bioavailable S192Y-*TYR* was significant but minimal; a finding reflected in the unaffected phenotypes of those homozygous for this variant. The activity of the R402Q-*TYR* was much lower than that of wild-type and lower than initially expected. This common variant had a remaining activity between 4% and 15%, which is at least 10% lower than previously report by Tripathi *et al*.[146]. The difference demonstrates the large variability between methods of analysis and suggests there is a need for a uniform, robust, and easily repeatable approach. Previously it has been suggested that only 5-8% of tyrosinase activity is required for low-level pigmentation[193], though 'low-level of pigmentation' does not equate to the (sometimes slightly pale) unaffected phenotype exhibited by individuals homozygous for R402Q. The difference between phenotype and genotype may be due to several reasons. Firstly, there is the inevitable aberrant processing of tyrosinase when produced *in vitro* which may affect the mutant protein more significantly than the wild-type protein, particularly if the variant confers ER retention. Secondly, the thermosensitive mutant-*TYR* is being expressed at a constant 37˚C, which does not reflect the temperature variability throughout the human body. Perhaps the variable presence of clinical features seen in patients with hypomorphic albinism is due to the variable activity of thermosensitive variants. In Chapter 4, it was hypothesised that the combination of R402Q and S192Y on the same allele would have an additive negative effect. The results showed a slight trend towards this end but were not significant. It appears the assays were not sensitive enough to low level activity due to the reasons listed above. An assessment of the mutant protein activity after synthesis over a range of lower temperatures may provide a clearer answer for this hypothesis.

During tyrosinase protein production is was noted that expression of tyrosinase within a non-melanogenic cell line (HEK293F) can lead to pigmentation. The level of cellular pigmentation between those transduced with the different *TYR*-mutants was visibly different and has the possibility to provide the most physiologically relevant measure of tyrosinase activity. However, the mechanism behind this pigmentation is unknown as the cellular machinery required for the classical pathway are not likely to be present. In 2010 Beisemeier *et al.* demonstrated melanin synthesis can occur in an RPE cell line via a non-classical pathway. Cells transduced with tyrosinase were able to pigment through a mechanism that bypassed the normal machinery required for melanosome maturation. The cells were transduced with wild-type tyrosinase which led to the production of DOPA-positive granules and larger melanin granules without the

presence of traditional melanosomes[81]. The ability of HEK293F cells to produce melanin may be further utilised as a tool for evaluating tyrosinase mutations.

Overall, the functional assessment of the two common *TYR* variants provides evidence for the pathogenicity of a tri-allelic genotype by showing that each variant can contribute to the haploinsufficiency of tyrosinase. Furthermore, if only 5-8% of tyrosinase activity is required for low-level of pigmentation then low-dose therapeutics become plausible for improving phenotypes resulting from both OCA1A and OCA1B mutations. The method used in Chapter 4 provides both an assay for *TYR* variants and a potential model for therapeutic testing.

6.3 Gene panel for diagnosis of nystagmus and albinism

The difficulty with the diagnosis of nystagmus is amplified when working with infants and very young children, for example carrying out an MRI, VEP, or ERG can be near to impossible. Furthermore, determining basics such as visual acuity and night vision is difficult until the infant is old enough to clearly communicate. More recently, advances in the field of OCT have enabled the use of handheld devices such as the Spectral domain HH-OCT Bioptigen Envisu system which has proved reliable in infants both with and without nystagmus[220]. Clinical phenotyping of adults can be much easier, not only due to compliance but also due to symptoms being more apparent. Therefore, an efficient workflow for the genetic diagnosis of nystagmus differs greatly depending on the age of the patient. A diagnosis is clinically most important when patients are very young in order to plan their management and support.

When trying to determine the molecular diagnosis of nystagmus in children, the extensive phenotype heterogeneity makes the use of single gene sequencing unrealistic. This was discussed in Chapter 5 where the two main phenotype groups, IIN/albinism, were further split into those with a full phenotype and those that either had (1) a partial phenotype or (2) it was unknown whether the feature was present. Even with these limitations, the assigned phenotype mostly correlated with the identified causal genotype (a total to six probands had a genotype not expected for their phenotypic group). Several cases were found to have a genotype consistent with albinism even though they presented with apparent IIN in clinic. This demonstrates that an IIN phenotype in a very young patient (or when limited clinical equipment is available) can mask another ocular disorder such as albinism, spinocerebellar ataxia, or *PAX6* disease[45, 167, 205, 208]. Moreover, it has been suggested that the most sensitive ocular feature for albinism is foveal hypoplasia[221]. This feature was the most consistent throughout the affected probands and family (described in Chapter 3). Fortunately, the fovea can be imaged reliably in both infants and adults, although it is not specific to albinism group disorders.

Following detailed phenotyping, an efficient genetic diagnosis can be achieved with a targeted NGS panel. Whether the targeting of genes comes from the sequencing panel itself or during data analysis does not impact the genetic diagnosis and therefore a broad/targeted sequencing panel can be chosen depending upon cost and availability. Premade, readily available gene panels are the most sensible and cost effective option. A broad panel such as the TSO can then be manipulated after sequencing to provide an efficient targeted approach. Using the TSO in combination with analysis of 31 targeted genes, a cohort of young and highly heterogeneous patients with frequently subtle phenotypes had a diagnostic yield of 43.2% (Chapter 5).

The effectiveness of the TSO and UKGTN gene panel is further evidenced by a study by the Gottlob group describing a NimbleGen Human custom array NGS panel specifically designed for the purpose of diagnosing nystagmus[208]. The custom array included 336 genes associated with nystagmus in the OMIM database and was used to sequence 15 patients. The paper claims a high rate of diagnosis (80%) with causal mutations. However, most mutations are found in four genes included in the UKGTN panel; *FRMD7, TYR, TYRP1, and CACNA1F*. The final gene, *CRYBA1*, causes cataracts which would have been easily determined through basic phenotyping alone. Additionally, this study included cases enriched for likely known genetic diagnoses and does not represent a broad clinical group.

The method of variant identification described in Chapter 5 is likely to overlook large deletions that would be found with a technique such as MLPA. Of the genes included within the UKGTN panel, only *OCA2* is known to be affected by a common large deletion. Most OCA2 patients of Sub-Saharan African heritage carry a 2.7 Kb homozygous deletion within *OCA2*[222]. However, the large deletion mutation is much less common in African Americans and rare in Caucasian populations[222]. Even though large deletions are uncommon they may still account for some missing heritability. A potential improvement to the gene panel may be to use computational analysis of read depth to estimate copy number but is not yet possible due to inconsistent coverage. A technique such as MLPA could be used instead or in combination to confirm or disprove any predictions of erroneous copy number.

6.3.1 Is phenotyping crucial for genetic diagnosis?

There is a current discussion of whether the genetic diagnosis of disease requires clinical phenotyping as the first step, or whether it would be more efficacious to immediately employ a large diagnostic gene panel and direct phenotyping towards putative genetic diagnoses. When considering nystagmus, there are some crucial features that need to be determined in the first

instance as early-onset nystagmus could be either congenital or acquired. Previous diagnostic work-ups have included detailed eye-tracking an MRI scanning as standard practise[2]. However, due to the costly nature of MRI scans and the necessity for general anaesthesia in young infants and difficulty to implement eye-tracking in the very young, the most common first step is to assess the nystagmus and utilise other, more accessible, eye tests. A thorough examination of patient history and at least a minimal level of clinical phenotyping may then lead to a decision on whether the likely cause could be genetic and if so, if it is likely due to retinal disease or neurological disease etc. in order to direct gene panel testing or prioritisation. For example, an abnormal ERG would suggest the patient requires a retinal gene investigation, whereas, a normal ERG and a phenotype with features consistent with albinism suggests that albinism and nystagmus genes should be investigated.

The high level of diagnosis achieved from only four genes in the Gottlob gene panel gives further evidence to support the use of a UKGTN gene panel[208]. A readymade commercially available panel is lower in cost and easy to implement throughout the NHS because it is suited to many different genetic diseases. A tailored filtering process, suited to specific phenotypes, makes the TSO panel suitable for nystagmus. The UKGTN panel works well at doing this, covering the most important genes associated with the subtle albinism and nystagmus phenotypes described in this cohort.

6.3.2 Bioinformatic analysis of NGS data

High-throughput NGS sequencing on such a large scale can generate an immense amount of data e.g. roughly 12,000 variants are identified per person when run on the TSO clinical exome. This provides a great opportunity to identify known diagnostic variants, identify new causal variants and even new complex genotypes. Conversely, it makes the prioritisation of relevant variants much more challenging. Phenotyping information can be very helpful in guiding the initial genetic search as a pathway or known causal genes can be selected. Following this, variants need to be prioritised through filtering steps. The ACMG standards for clinical reporting are complex, convoluted and sometimes too restrictive but they are helpful as a guide for assessing the pathogenicity of variants. These guidelines give a greater weighting to functional evidence and segregation of variant in *trans* with a known mutation, than to *in silico* tools. However, sometimes there is insufficient evidence and *in silico* tools can be particularly helpful in identifying putative pathogenic mutations for investigation.

According to the CMG guidelines, moderate evidence of pathogenicity includes variants located in a mutational hotspot or a conserved functional domain, and this is often analysed by

in silico predictive tools. CADD is considered one of the current best tools for predicting deleterious single nucleotide variants though it has difficulty in scoring large deletions and insertions. Often, predictive tools can fair well when compared to databases of known pathogenic coding mutations, but are less effective when scoring non-coding variants[223]. Intronic and synonymous exonic variants remain a problem for the time being due to the lack of current knowledge about their potential regulatory and splicing effects. MaxEntScan is considered a good unbiased model for splice site prediction but there are many models to choose from, each approaching the problem in a different way[224]. *In silico* predictive tools will become more reliable as more is discovered about the human genome and as genomic machine/deep learning continues to develop[225].

6.4 Remaining missing heritability in nystagmus genotypes

In Chapter 3 there is an initial high level of missing heritability in patients with hypomorphic albinism phenotypes. The majority of the partial genotypes can be resolved by inclusion a causal tri-allelic genotype in the *TYR* gene, taking the percentage of resolved cases from 22% to 56%. Similarly, the tri-allelic genotype accounted for 9 patients out of 48 in a cohort comprised of either complete or hypomorphic albinism (findings reported in Chapter 5). The identification of a causal genotype including previously 'benign' common variants, alongside description of a hypomorphic albinism that may sometimes be misdiagnosed, has accounted for some of the missing heritability seen in the genetic diagnosis of nystagmus. However, there is still a high level of missing heritability that is unaccounted for.

In Chapter 5, 57% of the cohort did not have a completely resolved causal genotype. This may be in part due to the moderate thresholds used to the filter through variants, as some variants may be novel but do not score highly in the *in silico* predictive models. Furthermore, as demonstrated within this body of work, some common population variants may be having a causal effect but are considered benign by universal and respected guidelines, such as those by ACMG. To tackle this problem, large datasets are required. Either in the form of a large cohort, to allow for wide scale analysis of patterns of inheritance, or in the form of large family groups such as the Pakistani populations mentioned in Chapter 3. The tool GenoPy, recently developed in Southampton, is designed to predict the additive effect of multiple variants[226]. For example, such a tool may elucidate the epistatic relationship between variants throughout the pigmentary pathway. Analysis of variants in a cohort or nystagmus and albinism patients in Chapter 5 identified four individual cases where digenic inheritance is a potential causal genotype (NG198, NG213, NG270, and potentially NG420) with a further single case identified in Chapter 3 (NG265). Digenic inheritance consists of mutations in two different genes contributing to the

causal genotype and has been previously reported across OCA subtypes 1-4[143, 227]. It is already known that the phenotype of OCA2 can be altered by mutations in another pigmentary gene, *MC1R,* resulting in an uncommon presentation of OCA2 with red hair[219]. However, mutations across two different genes is not yet widely reported or accepted as a causal genotype for albinism.

Another potential reason for missing heritability is that some causal genes remain undiscovered and a broader sequencing tool needs to be used such as whole exome sequencing or whole genome sequencing. Both of these techniques are only truly helpful if they are part of a large cohort, family group, or completed for trios of parents and child, due to the masses of sequencing variants. The 100,000 genomes project (Genomic England) is utilising trios to diagnose rare disorders and the successful identification of the causal genotype in thousands of cases has been recently published[228].

6.5 Summary

In summary, detailed clinical phenotyping is crucial for evaluation of any patient with nystagmus as it may inform the choice of gene panel and the prioritisation of variants during analysis. With the phenotyping, gene panel, variant prioritisation and the tri-allelic genotype described in this study we were able to provide a high diagnostic yield for patients with nystagmus (*Figure 45*). For albinism there is a more complex story than simple autosomal recessive inheritance of OCA. This work provides evidence for a novel tri-allelic genotype as well as providing some evidence of epistatic relationships between other OCA subtypes (*Figure 45*c). During the course of this PhD (with part credit to the publication of Chapter 3) more data is emerging to suggest there is a subclinical form of albinism where a patient is fair but has no other signs of the disorder without detailed ocular phenotyping[166, 229]. There seems to be a threshold at which the albinism reaches clinical status and this might be based partly on the presence 'causal mutations' but also the individual's pigmentary background. It is becoming clear than the inheritance of albinism can be very complex as SNPs throughout the entire pigmentation pathway (>30 genes) may have an effect on the penetrance of each 'pathogenic' mutation.

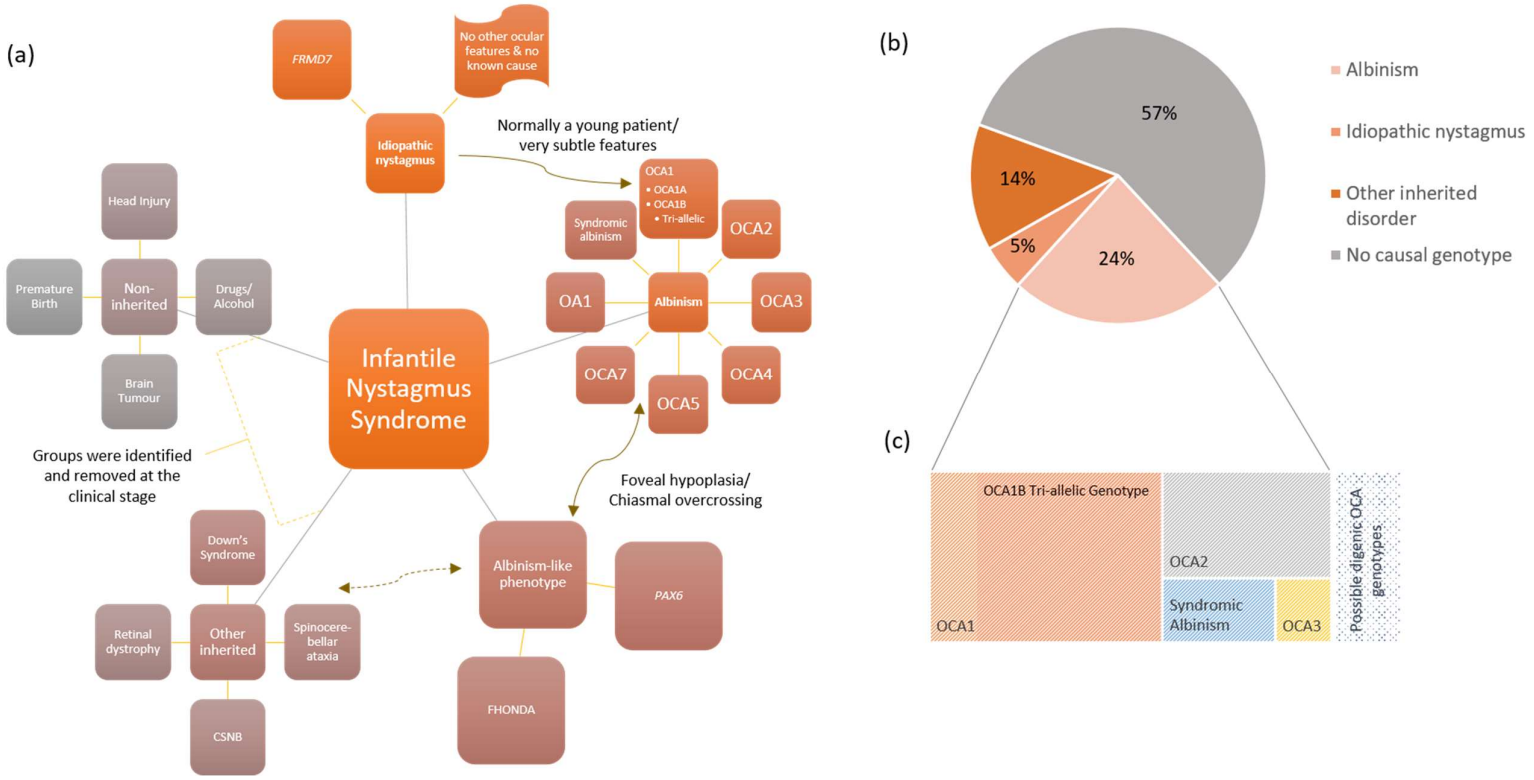


Figure 45. Causes of infantile nystagmus syndrome can be separated into clinically distinguishable phenotype groups to aid diagnosis.(a) There was a diagnostic yield of 43% following sequencing with TruSight One gene panel and prioritisation of the 31 gene on the UKGTN albinism and nystagmus panel (Chapter 5). (b) The albinism category can be further split into subgroups. The causal genotypes in OCA1 are largely accounted for by the tri-allelic genotype.

Ultimately, the broad clinical use of sequencing data should reduce the cost of care. Once the analysis of NGS data is fine-tuned to be timely and precise, identifying the causal genotype can lower the need for costly diagnostic and therapeutic procedures. A clear workflow created from the combination of clinical knowledge and genetic diagnoses in syndromes with overlapping symptoms will reduce variance between medical centres. In addition, accurate and specific genetic diagnosis is a prerequisite for meaningful management plans and any future clinical trials. The lack of definitive diagnosis has been a major limiting factor in the development of therapeutics for both nystagmus and albinism. Interestingly, the interpretation and implications of complex genotypes and missing heritability offers an emerging dilemma for clinicians. This scenario is likely to become more commonplace and complex if epistasis between melanogenesis pathway genes is shown to include those known to cause syndromic, systemic features. For example, a patient with an albinism phenotype and one mutation in a gene causing Hermansky-Pudlak syndrome (with no other albinism gene mutation identified) does not meet threshold for clinical diagnosis. However, this suggests that HPS is more likely for this individual. In clinical practice this 'non-confirmatory clue' may result in haematology testing to seek features of HPS as this diagnosis holds significant systemic implications.

Genetic counselling of families and early diagnosis makes treatment more plausible when therapeutic options become available. Furthermore, genetic diagnosis may be the first step towards providing personalised medicine where the most efficacious therapeutics are chosen to target the causative mutations. Examples include, therapeutics to prevent ER retention in the case of the *TYR* p.R402Q mutation (or other ER retained mutation), premature termination codon read-through therapies, such as Ataluren (first demonstrated with a *PAX6* mutation[230]), and gene therapy, in the cases where small molecules are not helpful. The functional study has examined methods for analysing tyrosinase mutations and informed on a potential all-encompassing assay that may be used to analyse numerous OCA genes in future. The efficacy and mechanism of different targeted therapeutics may be tested in the simple cell model described here and personalised to each patient's genotype.

6.6 Future work

As a continuation of the functional evaluation of tyrosinase, the recombinant protein could be expressed at 31˚C to determine the temperature sensitivity of each mutant. It may be that the R402Q-mutant and S192Y:R402Q present a greater difference in maturity, and therefore activity, if R402Q-*TYR* can evade ER-associated degradation at the lower temperature. The ability of tyrosinase-transfected HEK293F cells to produce melanin may be utilised in a high-

throughput approach for novel therapeutics. Different types of mutations can be treated with multiple therapeutics known to affect the melanin pathway, using the visible pigment produced as a measure of tyrosinase activity. Nitisinone is a small molecule drug that has been shown to improve pigmentation in mice with OCA1B by blocking tyrosine metabolism and increasing the level available for the melanin synthesis pathway[199]. Nitisinone may work to increase melanin production if there is small amount of remaining *TYR* activity. However, it would not be effective if all tyrosinase is degraded at the RNA level or before it can leave the ER. Ataluren is a translation read-through inducing drug that has improved the phenotype of mice with *Pax6* nonsense mutations by lowering nonsense mediated decay[230]. Such a translation read-through drug may only be effective if the protein is able to fold correctly and avoid ER-associated degradation. Palmitic acid has been shown to diminish the proteasomal degradation of tyrosinase leading to increased melanin accumulation in melanocytes and melanoma cells[231]. It is possible that certain drugs may be more effective when used in combination. Hence, a high throughput assay is necessary to trial numerous therapeutics on various different mutations.

When considering the remaining missing heritability in nystagmus and albinism there are a several strategies that can be taken. Firstly, the NGS data could be further interrogated for large indels. To avoid the added cost of an additional MLPA panel, software such as FishingCNV could be used to assess the copy number of genes and determine the likely presence of any large deletions or insertions. Secondly, the recently developed GenePY tool may be used to statistically infer the pathogenicity of variants across different OCA genes and assess the validity of an epistatic genotype[226]. It may be necessary to sequence family trios to resolve the remaining cases of missing heritability, particularly if causative genes are yet to be identified as causal. Another course of action would be to reanalyse the data using different (possibly improved) bioinformatic methods and filtering strategies. A recent paper showed how reanalysis can have a genuine high pick up rate through a simple change to methods[232], which is particularly relevant in a paediatric cohort as the phenotype may progress and become clear with age. A final point to consider is the identification of splicing variants. Often splicing variants will be filtered out due to the current lack of understanding of intronic regions. To combat this, transcriptomics such as RNA-Seq could be employed to identify difficult splicing mutations.

Appendix A Supplementary materials

A.1 Supplementary materials for Chapter 2: Materials and methods.

A.1.1 HaloPlex custom designed gene panel

Table A.1.1. Genes included on the custom designed HaloPlex target enrichment panel

Gene	Justification
ADAM17	Responsible for variation in pigmentation phenotype by association study of populations. (Inflammatory skin and bowel disease, neonatal)
ADAMTS20	Required for Melanoblast Survival. Responsible for variation in pigmentation phenotype by association study of populations.
AP3B1	Hermansky-Pudlak Syndrome type 2.
AP3D1	Involved in intracellular pigment trafficking. Mouse KO has Hermansky-Pudlak phenotype.
ASIP	Responsible for variation in population pigmentation.
ATRN	Possible role in melanocortin signalling pathways that regulate energy homeostasis and hair colour. Responsible for variation in pigmentation phenotype by association study of populations.
BLOC1S3	Hermansky-Pudlak Syndrome type 8.
BLOC1S6	Hermansky-Pudlak Syndrome type 9.
CACNA1A	Episodic Ataxia type 2 and Familial Hemiplegic Migraine.
CASK	Interacts with the FRMD7 locus. Mutations have been shown to interrupt binding between CASK and FRMD7. (X-linked mental retardation)
FRMD7 **(chrX:131210821-131267250:** **5000bp upstream, exons and introns, 200bp downstream)**	X-linked idiopathic nystagmus.
GPR143 **(chrX:9693253-9734205: exons and introns, 200bp downstream)**	X-linked ocular albinism (OA1)
CRB1	Leber's congenital amaurosis. Known to be modified when mutated by other genes.

DCT	Regulator of melanocyte proliferation. Interacts closely with *TYR* and *TYRP1*.
DTNBP1	Hermansky-Pudlak Syndrome type 7.
EDN3	Waardenburg syndrome subtype with significant pigmentary changes. Role in melanocyte formation.
HERC2	Prader-Willi syndrome. Responsible for variation in pigmentation phenotype by association study of populations. Linkage with OCA2, otherwise weak associated with pigment.
HPS1	Hermansky-Pudlak Syndrome type 1.
HPS3	Hermansky-Pudlak Syndrome type 3.
HPS4	Hermansky-Pudlak Syndrome type 4.
HPS5	Hermansky-Pudlak Syndrome type 5.
HPS6	Hermansky-Pudlak Syndrome type 6.
IRF4	Interacts with *MITF* to regulate the *TYR* promoter. Responsible for variation in pigmentation phenotype by association study of populations.
KITLG	Familial, progressive hyperpigmentation. Responsible for variation in pigmentation phenotype by association study of populations.
LYST	Chèdiak-Higashi Syndrome.
MC1R	Population pigmentation and modifies OCA2 phenotype.
MITF	Waardenburg and Tietz syndromes. Regulates expression of *TYR* and *TYRP1*.
MLPH	Griscelli syndrome. Responsible for variation in pigmentation phenotype by association study of populations. Tethers melanosomes to actin in mice.
MYO5A	Griscelli syndrome type 3 and neuroectodermal melanolysosomal disease. Role in transport and anchorage in melanocytes. Responsible for variation in pigmentation phenotype by association study of populations.
OCA2	Oculocutaneous albinism type 2. Involved in Prader-willi syndrome and Angelman syndrome.
PAX6	Aniridia, foveal hypoplasia and chiasmal over-crossing.
***PMEL* (AKA *SILV*)**	Also called SILV gene. Interacts with melanosome. Responsible for variation in pigmentation phenotype by association study of populations.

RAB27A	Griscelli Type 2. Interacts with *MLPH*. In mice, tethers melanosomes. Responsible for variation in pigmentation phenotype by association study of populations.
SEMA3A	Thought to interact with *FRMD7*. Effects on growth cone and neurite development.
SEMA3C	Related to *SEMA3A* which may interact with *FRMD7* to induce neurite outgrowth changes. Causes hypopigmentation in mice.
SHROOM2	Regulates melanosome biogenesis and is localised in the pigmented retina. Has been associated with ocular albinism.
SLC12A3	Responsible for variation in pigmentation phenotype by association study of populations. (Gitelman syndrome, a kidney disorder).
SLC24A4	Responsible for variation in pigmentation phenotype by association study of populations.
SLC24A5	SNPs in this gene are known to affect skin pigmentation.
SLC45A2	Oculocutaneous albinism type 4 and population skin tone differences.
SNAI2	Role in the formation and survival of melanocytes. Piebaldism and Waardenburg syndrome type 2.
SOX10	Waardenburg syndrome subtype and homologue of mouse pigmentary change gene.
TPCN2	SNPs in this gene strongly associated with hair and skin pigmentation.
TRPM1	Congenital stationary night blindness type 1C. Responsible for variation in pigmentation phenotype by association study of populations.
TYR	Oculocutaneous albinism type 1.
TYRP1	Oculocutaneous albinism type 3
ZIC2	Required for organisation of decussation at chiasm and homologue of mouse pigmentary change gene.
FARP1	High homology with *FRMD7* in humans and known role in neurite extension.
FARP2	Highly homology to *FRMD7* in humans and known role in neurite extension.
CDR1	Associated with paraneoplastic cerebellar degeneration.
TBR1	Strong binding partner of CASK and involved in brain development.

CASKIN1	Strongly and stably bound to CASK in the brain involved in coating tails of neurexins.
CASKIN2	Strong interactor with CASK and forms a tripartite complex.
APBA1	Strong interactor with CASK and forms a tripartite complex.
Lin7A	Forms tripartite complex with CASK for exocytosis and cell adhesion.
***XPO1* (encodes CRM1 protein)**	CRM1-dependent nuclear export sequence in FRMD7 is used to export FRMD7 out of the nucleus to its usual place in the cytoplasm.
C10orf11	Oculocutaneous albinism type 7. Required for melanocyte-differentiation.

A.1.2 SNP fingerprint panel.

Table A.1.2. SNPs included on the SNP fingerprinting panel designed by Pengelly et al. [126]. dbSNP is defined on the negative strand.

Chromosome	Position in GRCh37 (hg19) genome	dbSNP rsID	Gene	Alleles
1	179520506	rs1410592	NPHS2	A/C
1	67861520	rs2229546	IL12RB2	A/G
2	169789016	rs497692	ABCB11	A/G[b]
2	227896976	rs10203363	COL4A4	C/T
3	4403767	rs2819561	SUMF1	A/G[b]
4	5749904	rs4688963	EVC	A/G[b]
5	82834630	rs309557	VCAN	A/G[b]
6	146755140	rs2942	GRM1	C/T
7	48450157	rs17548783	ABCA13	C/T
8	94935937	rs4735258	PDP1	C/T
9	100190780	rs1381532	TDRD7	A/G[b]
10	100219314	rs10883099	HPSE2	A/G
11	16133413	rs4617548	SOX6	C/T
12	993930	rs7300444	WNK1	A/G
13	39433606	rs9532292	FREM2	A/G

14	50769717	rs2297995	L2HGDH	A/G
15	34528948	rs4577050	SLC12A6	C/T
16	70303580	rs2070203	AARS	A/G[b]
17	71197748	rs1037256	COG1	C/T
18	21413869	rs9962023	LAMA3	A/G
19	10267077	rs2228611	DNMT1	C/T[b]
20	6100088	rs10373	FERMT1	G/T[b]
21	44323590	rs4148973	NDUFV3	C/T
22	21141300	rs4675	SERPIND1	A/C

A.2 Supplementary materials from Chapter 5: The UKGTN approved albinism and nystagmus 31 gene panel as a diagnostic tool

A.2.1 Clinical exome sequencing quality

Table A.2.1. The depth, coverage and number of coding variants for the 81 samples in batches 1-6. Sequencing data analysis carried out by Luke O'Gorman.

Batch	No. samples	Mean depth	Min depth	Max depth	Mean coverage at 20x	Min coverage at 20x	Max coverage at 20x	Mean number of coding variants
1	16	166.3	114.2	222.3	0.98	0.972	0.987	12714
2	13	143.3	80.7	187.4	0.97	0.939	0.984	12063
3	21	123.5	61.4	214	0.956	0.911	0.986	11944
4	2	110.3	99	121.6	0.957	0.952	0.963	12145
5	13	107.7	89.5	129.5	0.94	0.915	0.968	10520
6	16	91.8	51	129.7	0.943	0.906	0.972	11849
Total	**81**							

Appendix B Publications

B.1 Publication from Chapter 3: Missing heritability in albinism

Norman, C. S., O'Gorman, L., Gibson, J., Pengelly, R. J., Baralle, D., Ratnayaka, J. A., Griffiths, H., Rose-Zerilli, M., Ranger, M., Bunyan, D., Lee, H., Page, R., Newall, T., Shawkat, F., Mattocks, C., Ward, D., Ennis, S., Self, J. E. (2017). **Identification of a functionally significant tri-allelic genotype in the Tyrosinase gene (*TYR*) causing hypomorphic oculocutaneous albinism (OCA1B).** *Scientific Reports*, *7*, 4415.

Arshad, M.W., Harlalka, G.V., Lin, S., D'Atri, I., Mehmood, S., Shakil, M., Hassan, M.J., Chioza, B.A., Self, J.E., Ennis, S., O'Gorman, L., Norman, C., Aman, T., Ali, S.S., Kaul, H., Baple, E.L., Crosby, A.H., Ullah, M.I. and Shabbir, M.I. (2018) **Mutations in TYR and OCA2 associated with oculocutaneous albinism in Pakistani families.** *Meta Gene*, 17, 48-55.

B.2 Publications from Chapter 5: The UKGTN approved albinism and nystagmus 31 gene panel as a diagnostic tool

O'Gorman, L., Norman, C.S., Michaels, L., Newall, T., Crosby, A.H., Mattocks, C., Cree, A.J., Lotery, A.J., Baple, E.L., Ratnayaka, J.A., Baralle, D., Lee, H., Osborne, D., Shawkat, F., Gibson, J., Ennis, S. & Self J.E., (2019). **A small gene sequencing panel realises a high diagnostic rate in patients with congenital nystagmus following basic phenotyping**. *Scientific reports*, 9(1), pp.1-8.

Glossary of Eye Disorders

Achromatopsia is a hereditary visual disorder also known as congenital colour blindness. The disorder is characterised by a complete absence of cone cell activity when measured with electroretinography, resulting in decreased vision, light sensitivity, and the absence of color vision.

Acquired Nystagmus is the result of damage to the parts of the brain involved in oculomotor functions or visual pathways. Those with acquired nystagmus tend to suffer from oscillopsia (objects in the visual field appear to be moving).

Aland Islands eye disease is an X-linked recessive retinal disease that can be caused by mutations to the *CACNA1F* gene. The disease is characterized by fundus hypopigmentation, nystagmus, astigmatism, progressive myopia, night-blindness and reduced visual acuity.

Albinism is a congenital disorder characterised by the complete or partial loss of pigment in the skin, hair and eyes, as well as affecting vision. Albinism covers a broad spectrum of pigment/ocular disorders consisting of ocular albinism, affecting eyes and vision only, and oculocutaneous albinism subtypes 1 to 7 (in more detail below) .

(Isolated) **Aniridia** is a congenital disorder characterised by defective iris development but can also cause foveal and optic nerve hypoplasia, cataract, and corneal changes. Mutations in the gene *PAX6* are associated with aniridia.

Congenital Nystagmus, also known as early onset nystagmus, is a general term for inherited that can be caused by another ocular disorder or may be isolated. Isolated cases (**congenital idiopathic nystagmus**) are associated with mutations in the *FRMD7* gene. Congenital nystagmus and infantile nystagmus are often used interchangeably (though IN does not specify that the disease is inherited).

Congenital Stationary Night Blindness refers to a group of eye disorders predominantly characterised by abnormal function of rod cells, detected through electroretinography. The range of vision is wide with one commonality being greatly reduced vision in low light.

Chèdiak-Higashi Syndrome (CHS) is a rare, inherited, complex, immune disorder that is characterized by oculocutaneous albinism, immune deficiencies and a tendency to bruise and bleed easily.

Cross-McKusick-Breen Syndrome (CMBS) is rare disorder characterized by oculocutaneous albinism, small eyes with cloudy corneas, gingival fibromatosis (fibrous nodules in the gums) and severe mental and physical retardation.

Elejalde Syndrome (ES) is a subtype of Griscelli syndrome (in more detail below), characterised by oculocutaneous albinism and neurologic defects.

Foveal Hypoplasia, also known as macular hypoplasia, is the underdevelopment of the fovea (and indentation centred in the retina used for high acuity vision) and is often associated with albinism and *PAX6* mutations.

FHONDA Syndrome is rare inherited disorder distinct from albinism that shares some phenotypic similarities. FHONDA stands for foveal hypoplasia, optic nerve decussation defects and anterior dysgenesis.

Griscelli Syndrome (GS) is a rare inherited disorder characterised by oculocutaneous albinism and silvery hair. There are 3 subtypes which differ on whether or not mental retardation and immune deficiencies are present.

Hermansky-Pudlak Syndrome (HPS) is an extremely rare inherited disorder characterised by oculocutaneous albinism, platelet abnormality (leading to bleeding problems), and lysosomal storage of the abnormal compound ceroid lipofuscin.

Leber Congenital Amaurosis (LCA) is one of the most common form of inherited sight loss in children. LCA is a retinal dystrophy causing nystagmus, sluggish pupillary response, photophobia and severe vision loss.

Noonan Syndrome** is a genetic disorder that prevents normal development in various parts of the bod, including unusual facial characteristics, short stature, heart defects, and possible developmental delays. ****not an associated eye disorder, but is present in our cohort in combination with albinism.

Oculocutaneous Albinism (OCA) Subtypes consist of OCA1 to OCA7. The most well documented forms are OCA1 to OCA4, each having a known associated gene and population. OCA5 to OCA7 are more recent discoveries so not as much is known about these subtypes. The subtypes differ in pigment levels but are all characterised by reduced pigment in skin, hair and eyes.

Spinocerebellar Ataxia is a degenerative, inherited, sometimes fatal disorder characterised by slow progressive incoordination of gait and often poor coordination of hands,

speech, and eye movements. There are over 20 types of spinocerebellar ataxia each differing in age of onset, symptoms and expected life span.

Retinitis Pigmentosa is one of the most common forms of inherited retinal degeneration. Retinitis pigmentosa describes a group of closely related retinal dystrophies that gradually cause vision loss as the photoreceptors deteriorate.

List of References

1. Penix, K., M.W. Swanson, and D.K. DeCarlo, *Nystagmus in pediatric patients: interventions and patient-focused perspectives.* Clinical Ophthalmology (Auckland, N.Z.), 2015. **9**: p. 1527-36.

2. Leigh, R.J. and D.S. Zee, *The neurology of eye movements*. Vol. 90. 2015. Oxford University Press, USA.

3. Theodorou, M. and R. Clement, *Classification of infantile nystagmus waveforms.* Vision Research, 2016. **123**: p. 20-5.

4. Pilling, R.F., J.R. Thompson, and I. Gottlob, *Social and visual function in nystagmus.* British Journal of Ophthalmology, 2005. **89**(10): p. 1278-81.

5. Stayte, M., B. Reeves, and C. Wortham, *Ocular and vision defects in preschool children.* British Journal of Ophthalmology, 1993. **77**(4): p. 228-32.

6. Abel, L.A., *Infantile nystagmus: current concepts in diagnosis and management.* Clinical & Experimental Optometry, 2006. **89**(2): p. 57-65.

7. Braddick, O. and J. Atkinson, *Development of human visual function.* Vision Research, 2011. **51**(13): p. 1588-609.

8. Gottlob, I. and F.A. Proudlock, *Aetiology of infantile nystagmus.* Current Opinion in Neurology, 2014. **27**(1): p. 83-91.

9. Snell, R.S. and M.A. Lemp, *Clinical anatomy of the eye*. 2013: John Wiley & Sons.

10. Lee, H., R. Purohit, V. Sheth, et al., *Retinal development in albinism: a prospective study using optical coherence tomography in infants and young children.* The Lancet, 2015. **385**: p. S14.

11. Leigh, R.J. and D.S. Zee, *The neurology of eye movements*. Vol. 55. 1999. Oxford University Press, USA.

12. Yee, R.D., R.W. Baloh, and V. Honrubia, *Study of congenital nystagmus: optokinetic nystagmus.* British Journal of Ophthalmology, 1980. **64**(12): p. 926-32.

13. Halmagyi, G.M., M.A. Gresty, and J. Leech, *Reversed optokinetic nystagmus (OKN): mechanism and clinical significance.* Annals of Neurology, 1980. **7**(5): p. 429-35.

14. Hertle, R.W. *Nystagmus in infancy and childhood*. in *Seminars in Ophthalmology*. 2008. Taylor & Francis, UK.

15. Serra, A. and R.J. Leigh, *Diagnostic value of nystagmus: spontaneous and induced ocular oscillations.* Journal of Neurology, Neurosurgery, and Psychiatry, 2002. **73**(6): p. 615-8.

16. Costa, A.C., M.C. Lopes, and C.R. Nakanami, *Influence of head posture on the visual acuity of children with nystagmus.* Arquivos Brasileiros de Oftalmologia, 2014. **77**(1): p. 8-11.

17. Cesarelli, M., P. Bifulco, L. Loffredo, et al., *Relationship between visual acuity and eye position variability during foveations in congenital nystagmus.* Documenta Ophthalmologica, 2000. **101**(1): p. 59-72.

18. Dunn, M.J., D. Wiggins, J.M. Woodhouse, et al., *The Effect of Gaze Angle on Visual Acuity in Infantile Nystagmus.* Investigative Ophthalmology and Visual Science, 2017. **58**(1): p. 642-650.

19. Dunn, M.J., T.H. Margrain, J.M. Woodhouse, et al., *Grating visual acuity in infantile nystagmus in the absence of image motion.* Investigative Ophthalmology and Visual Science, 2014. **55**(4): p. 2682-6.

20. Kumar, A., S. Shetty, P. Vijayalakshmi, et al., *Improvement in visual acuity following surgery for correction of head posture in infantile nystagmus syndrome.* Journal of Pediatric Ophthalmology and Strabismus, 2011. **48**(6): p. 341-6.

21. Huber-Reggi, S.P., C.C. Chen, L. Grimm, et al., *Severity of infantile nystagmus syndrome-like ocular motor phenotype is linked to the extent of the underlying optic nerve projection defect in zebrafish belladonna mutant.* Journal of Neuroscience, 2012. **32**(50): p. 18079-86.

22. Huang, M.Y.Y., C.-C. Chen, C.J. Bockisch, et al., *Doomed to Move the Eyes: Infantile Nystagmus-Like Eye Movements in Healthy Human Subjects.* Investigative Ophthalmology and Visual Science, 2011. **52**(14): p. 3016-3016.

23. Brodsky, M.C. and L.F. Dell'Osso, *A unifying neurologic mechanism for infantile nystagmus.* JAMA Ophthalmology, 2014. **132**(6): p. 761-8.

24. Dell'Osso, L.F. and R.B. Daroff, *Congenital nystagmus waveforms and foveation strategy.* Documenta Ophthalmologica, 1975. **39**(1): p. 155-82.

25. Dell'Osso, L.F., R.W. Hertle, and R.B. Daroff, *"Sensory" and "motor" nystagmus: erroneous and misleading terminology based on misinterpretation of David Cogan's observations.* Archives of Ophthalmology, 2007. **125**(11): p. 1559-61.

26. Abadi, R.V. and A. Bjerre, *Motor and sensory characteristics of infantile nystagmus.* British Journal of Ophthalmology, 2002. **86**(10): p. 1152-60.

27. Abadi, R.V. and C.M. Dickinson, *Waveform characteristics in congenital nystagmus.* Documenta Ophthalmologica, 1987. **64**(2): p. 153-167.

28. Clement, R.A., J.P. Whittle, M.R. Muldoon, et al., *Characterisation of congenital nystagmus waveforms in terms of periodic orbits.* Vision Research, 2002. **42**(17): p. 2123-30.

29. Optican, L.M. and D.S. Zee, *A hypothetical explanation of congenital nystagmus.* Biological Cybernetics, 1984. **50**(2): p. 119-34.

30. Harris, C.M., *Problems in modelling congenital nystagmus: Towards a new model*, in *Eye Movement Research: Mechanisms, Processes, and Applications*, R.W. John M. Findlay and W.K. Robert, Editors. 1995. North-Holland.

31. Broomhead, D.S., R.A. Clement, M.R. Muldoon, et al., *Modelling of congenital nystagmus waveforms produced by saccadic system abnormalities.* Biological Cybernetics, 2000. **82**(5): p. 391-9.

32. Casteels, I., C.M. Harris, F. Shawkat, et al., *Nystagmus in infancy.* British Journal of Ophthalmology, 1992. **76**(7): p. 434-7.

33. Tarpey, P., S. Thomas, N. Sarvananthan, et al., *Mutations in FRMD7, a newly identified member of the FERM family, cause X-linked idiopathic congenital nystagmus.* Nature Genetics, 2006. **38**(11): p. 1242-4.

34. Kerrison, J.B., V.J. Arnould, M.M. Barmada, et al., *A gene for autosomal dominant congenital nystagmus localizes to 6p12.* Genomics, 1996. **33**(3): p. 523-6.

35. Kerrison, J.B., M.R. Vagefi, M.M. Barmada, et al., *Congenital motor nystagmus linked to Xq26-q27.* American Journal of Medical Genetics Part A, 1999. **64**(2): p. 600-7.

36. Self, J.E., F. Shawkat, C.T. Malpas, et al., *Allelic variation of the FRMD7 gene in congenital idiopathic nystagmus.* Archives of Ophthalmology, 2007. **125**(9): p. 1255-63.

37. Yonehara, K., M. Fiscella, A. Drinnenberg, et al., *Congenital Nystagmus Gene FRMD7 Is Necessary for Establishing a Neuronal Circuit Asymmetry for Direction Selectivity.* Neuron, 2016. **89**(1): p. 177-93.

38. Watkins, R.J., M.G. Thomas, C.J. Talbot, et al., *The role of FRMD7 in idiopathic infantile nystagmus.* Journal of Ophthalmology, 2011. **2012**.

39. Self, J. and A. Lotery, *A review of the molecular genetics of congenital Idiopathic Nystagmus (CIN).* Ophthalmic Genetics, 2007. **28**(4): p. 187-91.

40. Sarvananthan, N., M. Surendran, E.O. Roberts, et al., *The prevalence of nystagmus: the Leicestershire nystagmus survey.* Investigative Ophthalmology and Visual Science, 2009. **50**(11): p. 5201-6.

41. Liu, J.Y., X. Ren, X. Yang, et al., *Identification of a novel GPR143 mutation in a large Chinese family with congenital nystagmus as the most prominent and consistent manifestation.* Journal of Human Genetics, 2007. **52**(6): p. 565-570.

42. Holmstrom, G., M.L. Bondeson, U. Eriksson, et al., *'Congenital' nystagmus may hide various ophthalmic diagnoses.* Acta Ophthalmologica, 2014. **92**(5): p. 412-6.

43. Thomas, S., M.G. Thomas, C. Andrews, et al., *Autosomal-dominant nystagmus, foveal hypoplasia and presenile cataract associated with a novel PAX6 mutation.* European Journal of Human Genetics, 2014. **22**(3): p. 344-9.

44. Kohl, S., H. Jagle, B. Wissinger, et al., *Achromatopsia*, in *GeneReviews[Internet]*, M.P. Adam, et al., Editors. 1993. University of Washington: Seattle (WA).

45. Boycott, K.M., W.G. Pearce, and N.T. Bech-Hansen, *Clinical variability among patients with incomplete X-linked congenital stationary night blindness and a founder mutation in CACNA1F.* Canadian Journal of Ophthalmology, 2000. **35**(4): p. 204-13.

46. Paulson, H.L., *The spinocerebellar ataxias.* Journal of Neuro-Ophthalmology, 2009. **29**(3): p. 227-37.

47. Self, J., C. Mercer, E.M.J. Boon, et al., *Infantile nystagmus and late onset ataxia associated with a CACNA1A mutation in the intracellular loop between s4 and s5 of domain 3.* Eye, 2009. **23**(12): p. 2251-2255.

48. Bertsch, M., M. Floyd, T. Kehoe, et al., *The clinical evaluation of infantile nystagmus: What to do first and why.* Ophthalmic Genetics, 2017. **38**(1): p. 22-33.

49. Hendrickson, A.E. and C. Yuodelis, *The morphological development of the human fovea.* Ophthalmology, 1984. **91**(6): p. 603-12.

50. Poulter, J.A., M. Al-Araimi, I. Conte, et al., *Recessive mutations in SLC38A8 cause foveal hypoplasia and optic nerve misrouting without albinism.* American Journal of Medical Genetics Part A, 2013. **93**(6): p. 1143-50.

51. Gronskov, K., J. Ek, and K. Brondum-Nielsen, *Oculocutaneous albinism.* Orphanet Journal of Rare Diseases, 2007. **2**(43): p. 43.

52. Simeonov, D.R., X. Wang, C. Wang, et al., *DNA variations in oculocutaneous albinism: an updated mutation list and current outstanding issues in molecular diagnostics.* Human Mutation, 2013. **34**(6): p. 827-35.

53. Kausar, T., M.A. Bhatti, M. Ali, et al., *OCA5, a novel locus for non-syndromic oculocutaneous albinism, maps to chromosome 4q24.* Clinical Genetics, 2013. **84**(1): p. 91-3.

54. Morice-Picard, F., E. Lasseaux, S. Francois, et al., *SLC24A5 mutations are associated with non-syndromic oculocutaneous albinism.* Journal of Investigative Dermatology, 2014. **134**(2): p. 568-571.

55. Bertolotti, A., E. Lasseaux, C. Plaisant, et al., *Identification of a homozygous mutation of SLC24A5 (OCA6) in two patients with oculocutaneous albinism from French Guiana.* Pigment Cell Melanoma Research, 2016. **29**(1): p. 104-6.

56. Gronskov, K., C.M. Dooley, E. Ostergaard, et al., *Mutations in c10orf11, a melanocyte-differentiation gene, cause autosomal-recessive albinism.* American Journal of Medical Genetics Part A, 2013. **92**(3): p. 415-21.

57. Hutton, S.M. and R.A. Spritz, *Comprehensive analysis of oculocutaneous albinism among non-Hispanic caucasians shows that OCA1 is the most prevalent OCA type.* Journal of Investigative Dermatology, 2008a. **128**(10): p. 2442-2450.

58. Lewis, R.A., *Oculocutaneous Albinism Type 1*, in *GeneReviews[Internet]*, M.P. Adam, et al., Editors. 1993. University of Washington: Seattle (WA).

59. Chaki, M., M. Sengupta, M. Mondal, et al., *Molecular and functional studies of tyrosinase variants among Indian oculocutaneous albinism type 1 patients.* Journal of Investigative Dermatology, 2011. **131**(1): p. 260-2.

60. Jagirdar, K., D.J. Smit, S.A. Ainger, et al., *Molecular analysis of common polymorphisms within the human Tyrosinase locus and genetic association with pigmentation traits.* Pigment Cell Melanoma Research, 2014. **27**(4): p. 552-64.

61. Dolinska, M.B., E. Kovaleva, P. Backlund, et al., *Albinism-causing mutations in recombinant human tyrosinase alter intrinsic enzymatic activity.* PloS One, 2014. **9**(1): p. e84494.

62. Mondal, M., M. Sengupta, and K. Ray, *Functional assessment of tyrosinase variants identified in individuals with albinism is essential for unequivocal determination of genotype-to-phenotype correlation.* British Journal of Dermatology, 2016. **175**(6): p. 1232-1242.

63. Dolinska, M.B., N. Kus, K. Farney, et al., *Oculocutaneous Albinism Type 1: Link between Mutations, Tyrosinase Conformational Stability, and Enzymatic Activity.* Pigment Cell Melanoma Research, 2016. **30**(1): p.41-52.

64. Gargiulo, A., F. Testa, S. Rossi, et al., *Molecular and clinical characterization of albinism in a large cohort of Italian patients.* Investigative Ophthalmology and Visual Science, 2011. **52**(3): p. 1281-9.

65. Bassi, M.T., M.V. Schiaffino, A. Renieri, et al., *Cloning of the gene for ocular albinism type 1 from the distal short arm of the X chromosome.* Nature Genetics, 1995. **10**(1): p. 13-9.

66. Pearce, W.G., G.J. Johnson, and J.G. Gillan, *Nystagmus in a female carrier of ocular albinism.* Journal of Medical Genetics, 1972. **9**(1): p. 126-9.

67. Preising, M.N., H. Forster, M. Gonser, et al., *Screening of TYR, OCA2, GPR143, and MC1R in patients with congenital nystagmus, macular hypoplasia, and fundus hypopigmentation indicating albinism.* Molecular Vision, 2011. **17**: p. 939.

68. Zou, X., H. Li, L. Yang, et al., *Molecular genetic and clinical evaluation of three Chinese families with X-linked ocular albinism.* Scientific Reports, 2017. **7**: p. 33713.

69. Preising, M., J.P. Op de Laak, and B. Lorenz, *Deletion in the OA1 gene in a family with congenital X linked nystagmus.* British Journal of Ophthalmology, 2001. **85**(9): p. 1098-103.

70. Faugere, V., S. Tuffery-Giraud, C. Hamel, et al., *Identification of three novel OA1 gene mutations identified in three families misdiagnosed with congenital nystagmus and carrier status determination by real-time quantitative PCR assay.* BMC Genetics, 2003. **4**: p. 1.

71. Hu, J., D. Liang, J. Xue, et al., *A novel GPR143 splicing mutation in a Chinese family with X-linked congenital nystagmus.* Molecular Vision, 2011. **17**: p. 715-22.

72. Charles, S.J., A.T. Moore, J.W. Grant, et al., *Genetic counselling in X-linked ocular albinism: clinical features of the carrier state.* Eye, 1992. **6 (Pt 1)**(1): p. 75-9.

73. O'Donnell, F.E., Jr., R.A. King, W.R. Green, et al., *Autosomal recessively inherited ocular albinism. A new form of ocular albinism affecting females as severely as males.* Archives of Ophthalmology, 1978. **96**(9): p. 1621-5.

74. Hutton, S.M. and R.A. Spritz, *A comprehensive genetic study of autosomal recessive ocular albinism in Caucasian patients.* Investigative Ophthalmology and Visual Science, 2008. **49**(3): p. 868-72.

75. Chiang, P.W., E. Spector, and A.C. Tsai, *Oculocutaneous albinism spectrum.* American Journal of Medical Genetics Part A, 2009. **149A**(7): p. 1590-1.

76. Chiang, P.W., J.M. Drautz, A.C. Tsai, et al., *A new hypothesis of OCA1B.* American Journal of Medical Genetics Part A, 2008. **146A**(22): p. 2968-70.

77. Scheinfeld, N.S., *Syndromic albinism: a review of genetics and phenotypes.* Dermatology Online Journal, 2003. **9**(5): p. 5.

78. Spritz, R.A. and J. Oh, *HPS gene mutations in Hermansky-Pudlak syndrome.* American Journal of Medical Genetics Part A, 1999. **64**(2): p. 658.

79. Prota, G., *Melanins and melanogenesis*. 2012. Academic Press, USA.

80. Boulton, M. and P. Dayhaw-Barker, *The role of the retinal pigment epithelium: topographical variation and ageing changes.* Eye, 2001. **15**(Pt 3): p. 384-9.

81. Biesemeier, A., F. Kreppel, S. Kochanek, et al., *The classical pathway of melanogenesis is not essential for melanin synthesis in the adult retinal pigment epithelium.* Cell and Tissue Research, 2010. **339**(3): p. 551-560.

82. Toyofuku, K., J.C. Valencia, T. Kushimoto, et al., *The etiology of oculocutaneous albinism (OCA) type II: the pink protein modulates the processing and transport of tyrosinase.* Pigment Cell Melanoma Research, 2002. **15**(3): p. 217-24.

83. Brilliant, M.H., *The mouse p (pink-eyed dilution) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH.* Pigment Cell Melanoma Research, 2001. **14**(2): p. 86-93.

84. Kobayashi, T., G. Imokawa, D.C. Bennett, et al., *Tyrosinase stabilization by Tyrp1 (the brown locus protein).* Journal of Biological Chemistry, 1998. **273**(48): p. 31801-5.

85. Bin, B.H., J. Bhin, S.H. Yang, et al., *Membrane-Associated Transporter Protein (MATP) Regulates Melanosomal pH and Influences Tyrosinase Activity.* PloS One, 2015. **10**(6): p. e0129273.

86. Roffler-Tarlov, S., J.H. Liu, E.N. Naumova, et al., *L-Dopa and the albino riddle: content of L-Dopa in the developing retina of pigmented and albino mice.* PloS One, 2013. **8**(3): p. e57184.

87. Goshima, Y., S. Watanabe, E. Seki, et al., *Immunoreactivity of a G protein-coupled l-DOPA receptor GPR143, in Lewy bodies.* Neuroscience Research, 2019. **148**: p. 49-53.

88. Masukawa, D., K. Yamada, and Y. Goshima, *Overexpression of the gene product of ocular albinism 1 (GPR143/OA1) but not its mutant forms inhibits neurite outgrowth in PC12 cells.* Journal of Pharmacological Sciences, 2019. **141**(1): p. 41-48.

89. Thurtell, M.J. and R.J. Leigh, *Therapy for nystagmus.* Journal of Neuro-Ophthalmology, 2010. **30**(4): p. 361-71.

90. Jacobs, J.B. and L.F. Dell'Osso, *Congenital nystagmus: hypotheses for its genesis and complex waveforms within a behavioral ocular motor system model.* Journal of Vision, 2004. **4**(7): p. 604-25.

91. Helga Kolb, E.F. and N. Ralph, *Webvision:the organization of the retina and visual system*. 2007. Bethesda, National Center for Biotechnology Information, USA.

92. Gabriele, M.L., G. Wollstein, H. Ishikawa, et al., *Optical coherence tomography: history, current status, and laboratory work.* Investigative Ophthalmology and Visual Science, 2011. **52**(5): p. 2425-36.

93. Seo, J.H., Y.S. Yu, J.H. Kim, et al., *Correlation of visual acuity with foveal hypoplasia grading by optical coherence tomography in albinism.* Ophthalmology, 2007. **114**(8): p. 1547-51.

94. Guillery, R.W., A.N. Okoro, and C.J. Witkop, Jr., *Abnormal visual pathways in the brain of a human albino.* Brain Research, 1975. **96**(2): p. 373-7.

95. Dorey, S., M. Neveu, L. Burton, et al., *The clinical features of albinism and their correlation with visual evoked potentials.* British Journal of Ophthalmology, 2003. **87**(6): p. 767-772.

96. von dem Hagen, E.A., M.B. Hoffmann, and A.B. Morland, *Identifying human albinism: a comparison of VEP and fMRI.* Investigative Ophthalmology and Visual Science, 2008. **49**(1): p. 238-49.

97. Hingorani, M., K.A. Williamson, A.T. Moore, et al., *Detailed ophthalmologic evaluation of 43 individuals with PAX6 mutations.* Investigative Ophthalmology and Visual Science, 2009. **50**(6): p. 2581-90.

98. Hoffmann, M.B. and S.O. Dumoulin, *Congenital visual pathway abnormalities: a window onto cortical stability and plasticity.* Trends in Neurosciences, 2015. **38**(1): p. 55-65.

99. Greven, M.A. and L.B. Nelson, *Four-muscle tenotomy surgery for nystagmus.* Current Opinion in Ophthalmology, 2014. **25**(5): p. 400-5.

100. Averbuch-Heller, L., R.J. Tusa, L. Fuhry, et al., *A double-blind controlled study of gabapentin and baclofen as treatment for acquired nystagmus.* Annals of Neurology, 1997. **41**(6): p. 818-25.

101. Starck, M., H. Albrecht, W. Pollmann, et al., *Drug therapy for acquired pendular nystagmus in multiple sclerosis.* Journal of Neurology, 1997. **244**(1): p. 9-16.

102. McLean, R., F. Proudlock, S. Thomas, et al., *Congenital nystagmus: randomized, controlled, double-masked trial of memantine/gabapentin.* Annals of Neurology, 2007. **61**(2): p. 130-8.

103. Baloh, R.W. and A. Winder, *Acetazolamide-responsive vestibulocerebellar syndrome: clinical and oculographic features.* Neurology, 1991. **41**(3): p. 429-33.

104. Nelwan, M., *Treat oculocutaneous albinism with gene therapy.* Journal of Advances in Biology & Biotechnology, 2018. **16**(3): p. 1-12

105. Ikawa, Y., R. Hess, H. Dorward, et al., *In vitro functional correction of Hermansky-Pudlak Syndrome type-1 by lentiviral-mediated gene transfer.* Molecular Genetics and Metabolism, 2015. **114**(1): p. 62-5.

106. Yoshimi, K., T. Kaneko, B. Voigt, et al., *Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR-Cas platform.* Nature Communications, 2014. **5**(1): p. 4240.

107. Wetterstrand, K.A. *The Cost of Sequencing a Human Genome: data from the NHGRI Genome Sequencing Program (GSP)*, 2016 [Accessed: 02 Feb 2017]; Available from: https://www.genome.gov/sequencingcosts/.

108. International HapMap, C., *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.

109. Shah, N., Y.C. Hou, H.C. Yu, et al., *Identification of Misclassified ClinVar Variants via Disease Population Prevalence.* American Journal of Medical Genetics Part A, 2018. **102**(4): p. 609-619.

110. Richards, S., N. Aziz, S. Bale, et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* Genetics in Medicine, 2015. **17**(5): p. 405-24.

111. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Research, 2010. **38**(16): p. e164.

112. Ng, P.C. and S. Henikoff, *Predicting deleterious amino acid substitutions.* Genome Research, 2001. **11**(5): p. 863-74.

113. Davydov, E.V., D.L. Goode, M. Sirota, et al., *Identifying a high fraction of the human genome to be under selective constraint using GERP++.* PLoS Computational Biology, 2010. **6**(12): p. e1001025.

114. Adzhubei, I.A., S. Schmidt, L. Peshkin, et al., *A method and server for predicting damaging missense mutations.* Nature Methods, 2010. **7**(4): p. 248-9.

115. Kircher, M., D.M. Witten, P. Jain, et al., *A general framework for estimating the relative pathogenicity of human genetic variants.* Nature Genetics, 2014. **46**(3): p. 310-5.

116. Rentzsch, P., D. Witten, G.M. Cooper, et al., *CADD: predicting the deleteriousness of variants throughout the human genome.* Nucleic Acids Research, 2019. **47**(D1): p. D886-D894.

117. Yeo, G. and C.B. Burge, *Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals.* Journal of Computational Biology, 2004. **11**(2-3): p. 377-94.

118. Arshad, M.W., G.V. Harlalka, S. Lin, et al., *Mutations in TYR and OCA2 associated with oculocutaneous albinism in Pakistani families.* Meta Gene, 2018. **17**: p. 48-55.

119. Kriss, A. and I. Russell-Eggitt, *Electrophysiological assessment of visual pathway function in infants.* Eye, 1992. **6 (Pt 2)**: p. 145-53.

120. McCulloch, D.L., M.F. Marmor, M.G. Brigell, et al., *ISCEV Standard for full-field clinical electroretinography (2015 update).* Documenta Ophthalmologica, 2015. **130**(1): p. 1-12.

121. Odom, J.V., M. Bach, M. Brigell, et al., *ISCEV standard for clinical visual evoked potentials: (2016 update).* Documenta Ophthalmologica, 2016. **133**(1): p. 1-9.

122. Sabina, J. and J.H. Leamon, *Bias in Whole Genome Amplification: Causes and Considerations.* Methods in Molecular Biology, 2015. **1347**: p. 15-41.

123. Li, H., B. Handsaker, A. Wysoker, et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

124. Stenson, P.D., M. Mort, E.V. Ball, et al., *The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine.* Human Genetics, 2014. **133**(1): p. 1-9.

125. Thankaswamy-Kosalai, S., P. Sen, and I. Nookaew, *Evaluation and assessment of read-mapping by multiple next-generation sequencing aligners based on genome-wide characteristics.* Genomics, 2017. **109**(3-4): p. 186-191.

126. Pengelly, R.J., J. Gibson, G. Andreoletti, et al., *A SNP profiling panel for sample tracking in whole-exome sequencing studies.* Genome Medicine, 2013. **5**(9): p. 89.

127. Schouten, J.P., C.J. McElgunn, R. Waaijer, et al., *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.* Nucleic Acids Research, 2002. **30**(12): p. e57.

128. Cooper, G.M., D.L. Goode, S.B. Ng, et al., *Single-nucleotide evolutionary constraint scores highlight disease-causing mutations.* Nature Methods, 2010. **7**(4): p. 250-1.

129. Ioannidis, N.M., J.H. Rothstein, V. Pejaver, et al., *REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants.* American Journal of Medical Genetics Part A, 2016. **99**(4): p. 877-885.

130. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5463-7.

131. Smith, L.M., J.Z. Sanders, R.J. Kaiser, et al., *Fluorescence detection in automated DNA sequence analysis.* Nature, 1986. **321**(6071): p. 674-9.

132. Wang, X., Y. Zhu, N. Shen, et al., *Mutation analysis of a Chinese family with oculocutaneous albinism.* Oncotarget, 2016. **7**(51): p. 84981-84988.

133. Kalahroudi, V.G., B. Kamalidehghan, A.A. Kani, et al., *Two Novel Tyrosinase (TYR) Gene Mutations with Pathogenic Impact on Oculocutaneous Albinism Type 1 (OCA1).* PloS One, 2014. **9**(9): p. e106656.

134. Halaban, R., E. Cheng, and D.N. Hebert, *Coexpression of wild-type tyrosinase enhances maturation of temperature-sensitive tyrosinase mutants.* Journal of Investigative Dermatology, 2002. **119**(2): p. 481-8.

135. Maley, F., R.B. Trimble, A.L. Tarentino, et al., *Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases.* Analytical Biochemistry, 1989. **180**(2): p. 195-204.

136. Winder, A.J. and H. Harris, *New assays for the tyrosine hydroxylase and dopa oxidase activities of tyrosinase.* European Journal of Biochemistry, 1991. **198**(2): p. 317-26.

137. Oetting, W.S. and R.A. King, *Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism.* Human Mutation, 1999. **13**(2): p. 99-115.

138. King, R.A., J. Pietsch, J.P. Fryer, et al., *Tyrosinase gene mutations in oculocutaneous albinism 1 (OCA1): definition of the phenotype.* Human Genetics, 2003. **113**(6): p. 502-13.

139. McCafferty, B.K., M.A. Wilk, J.T. McAllister, et al., *Clinical Insights Into Foveal Morphology in Albinism.* Journal of Pediatric Ophthalmology and Strabismus, 2015. **52**(3): p. 167-72.

140. Wolf, A.B., S.E. Rubin, and S.R. Kodsi, *Comparison of clinical findings in pediatric patients with albinism and different amplitudes of nystagmus.* Journal of American Association for Pediatric Ophthalmology and Strabismus, 2005. **9**(4): p. 363-8.

141. Montoliu, L., K. Gronskov, A.H. Wei, et al., *Increasing the complexity: new genes and new types of albinism.* Pigment Cell Melanoma Research, 2014. **27**(1): p. 11-8.

142. Chiang, P.W., E. Spector, and A.C. Tsai, *Evidence suggesting the inheritance mode of the human P gene in skin complexion is not strictly recessive.* American Journal of Medical Genetics Part A, 2008. **146A**(11): p. 1493-6.

143. Chiang, P.W., A.B. Fulton, E. Spector, et al., *Synergistic interaction of the OCA2 and OCA3 genes in a family.* American Journal of Medical Genetics Part A, 2008. **146A**(18): p. 2427-30.

144. Fukai, K., S.A. Holmes, N.J. Lucchese, et al., *Autosomal recessive ocular albinism associated with a functionally significant tyrosinase gene polymorphism.* Nature Genetics, 1995. **9**(1): p. 92-5.

145. Genomes Project, C., A. Auton, L.D. Brooks, et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

146. Tripathi, R.K., L.B. Giebel, K.M. Strunk, et al., *A polymorphism of the human tyrosinase gene is associated with temperature-sensitive enzymatic activity.* Gene Expression, 1991. **1**(2): p. 103-10.

147. Toyofuku, K., I. Wada, J.C. Valencia, et al., *Oculocutaneous albinism types 1 and 3 are ER retention diseases: mutation of tyrosinase or Tyrp1 can affect the processing of both mutant and wild-type proteins.* FASEB Journal, 2001. **15**(12): p. 2149-61.

148. Oetting, W.S., J. Pietsch, M.J. Brott, et al., *The R402Q tyrosinase variant does not cause autosomal recessive ocular albinism.* American Journal of Medical Genetics Part A, 2009. **149A**(3): p. 466-9.

149. Monferme, S., E. Lasseaux, C. Duncombe-Poulet, et al., *Mild form of oculocutaneous albinism type 1: phenotypic analysis of compound heterozygous patients with the R402Q variant of the TYR gene.* British Journal of Ophthalmology, 2019. **103**(9): p. 1239-1247.

150. Lewis, R.A., *Oculocutaneous Albinism Type 2*, in *GeneReviews [Online]*, M.P. Adam, et al., Editors. 2003 [Updated August 2012], University of Washington, Seattle.

151. Azuma, N. and M. Yamada, *Missense mutation at the C terminus of the PAX6 gene in ocular anterior segment anomalies.* Investigative Ophthalmology and Visual Science, 1998. **39**(5): p. 828-30.

152. Kikkawa, T., C.R. Casingal, S.H. Chun, et al., *The role of Pax6 in brain development and its impact on pathogenesis of autism spectrum disorder.* Brain Research, 2019. **1705**: p. 95-103.

153. Lee, S.T., R.D. Nicholls, S. Bundey, et al., *Mutations of the P gene in oculocutaneous albinism, ocular albinism, and Prader-Willi syndrome plus albinism.* New England Journal of Medicine, 1994. **330**(8): p. 529-34.

154. Gronskov, K., J. Ek, A. Sand, et al., *Birth prevalence and mutation spectrum in danish patients with autosomal recessive albinism.* Investigative Ophthalmology and Visual Science, 2009. **50**(3): p. 1058-64.

155. Mondal, M., M. Sengupta, S. Samanta, et al., *Molecular basis of albinism in India: evaluation of seven potential candidate genes and some new findings.* Gene, 2012. **511**(2): p. 470-4.

156. Oetting, W.S., S.S. Garrett, M. Brott, et al., *P gene mutations associated with oculocutaneous albinism type II (OCA2).* Human Mutation, 2005. **25**(3): p. 323.

157. Hawkes, J.E., P.B. Cassidy, P. Manga, et al., *Report of a novel OCA2 gene mutation and an investigation of OCA2 variants on melanoma risk in a familial melanoma pedigree.* Journal of Dermatological Science, 2013. **69**(1): p. 30-7.

158. Spritz, R.A., J. Oh, K. Fukai, et al., *Novel mutations of the tyrosinase (TYR) gene in type I oculocutaneous albinism (OCA1).* Human Mutation, 1997. **10**(2): p. 171-4.

159. Oetting, W.S., J.P. Fryer, and R.A. King, *Mutations of the human tyrosinase gene associated with tyrosinase related oculocutaneous albinism (OCA1). Mutations in brief no. 204. Online.* Human Mutation, 1998. **12**(6): p. 433-4.

160. Opitz, S., B. Käsmann-Kellner, M. Kaufmann, et al., *Detection of 53 novel DNA variations within the tyrosinase gene and accumulation of mutations in 17 patients with albinism.* Human Mutation, 2004. **23**(6): p. 630-631.

161. Oetting, W.S., J.P. Fryer, S. Shriram, et al., *Oculocutaneous albinism type 1: the last 100 years.* Pigment Cell Melanoma Research, 2003. **16**(3): p. 307-11.

162. Rooryck, C., F. Morice-Picard, N.H. Elcioglu, et al., *Molecular diagnosis of oculocutaneous albinism: new mutations in the OCA1-4 genes and practical aspects.* Pigment Cell Melanoma Res, 2008. **21**(5): p. 583-7.

163. Oetting, W.S., M.M. Mentink, C.G. Summers, et al., *Three different frameshift mutations of the tyrosinase gene in type IA oculocutaneous albinism.* American Journal of Medical Genetics Part A, 1991. **49**(1): p. 199-206.

164. Fryer, J.P., W.S. Oetting, and R.A. King, *Identification and characterization of a DNase hypersensitive region of the human tyrosinase gene.* Pigment Cell Melanoma Research, 2003. **16**(6): p. 679-84.

165. Eichers, E.R., R.A. Lewis, N. Katsanis, et al., *Triallelic inheritance: a bridge between Mendelian and multifactorial traits.* Annals of Medicine, 2004. **36**(4): p. 262-72.

166. Norman, C.S., L. O'Gorman, J. Gibson, et al., *Identification of a functionally significant tri-allelic genotype in the Tyrosinase gene (TYR) causing hypomorphic oculocutaneous albinism (OCA1B).* Scientific Reports, 2017. **7**(1): p. 4415.

167. Lasseaux, E., C. Plaisant, V. Michaud, et al., *Molecular characterization of a series of 990 index patients with albinism.* Pigment Cell Melanoma Research, 2018. **31**(4): p. 466-474.

168. Farooqi, M.S., M. Mitui, E.R. Londin, et al., *High Concentration Capture Probes Enhance Massively Parallel Sequencing Assays.* Clinical Chemistry, 2016. **62**(7): p. 1032-4.

169. Nan, H., P. Kraft, D.J. Hunter, et al., *Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians.* International Journal of Cancer, 2009. **125**(4): p. 909-17.

170. Ibarrola-Villava, M., H.H. Hu, M. Guedj, et al., *MC1R, SLC45A2 and TYR genetic variants involved in melanoma susceptibility in southern European populations: results from a meta-analysis.* European Journal of Cancer, 2012. **48**(14): p. 2183-91.

171. Puri, N., J.M. Gardner, and M.H. Brilliant, *Aberrant pH of melanosomes in pink-eyed dilution (p) mutant melanocytes.* Journal of Investigative Dermatology, 2000. **115**(4): p. 607-13.

172. Kobayashi, T. and V.J. Hearing, *Direct interaction of tyrosinase with Tyrp1 to form heterodimeric complexes in vivo.* Journal of Cell Science, 2007. **120**(Pt 24): p. 4261-8.

173. Jung, S.W., K. Lee, and A.E. Cho, *Computational approaches to predict binding interactions between mammalian tyrosinases and (S)-(+)-decursin and its analogues as potent inhibitors.* RSC Advances, 2016. **6**(52): p. 46765-46774.

174. Morell, R., R.A. Spritz, L. Ho, et al., *Apparent digenic inheritance of Waardenburg syndrome type 2 (WS2) and autosomal recessive ocular albinism (AROA).* Human Molecular Genetics, 1997. **6**(5): p. 659-64.

175. Lai, X., H.J. Wichers, M. Soler-Lopez, et al., *Structure of Human Tyrosinase Related Protein 1 Reveals a Binuclear Zinc Active Site Important for Melanogenesis.* Angew Chem Int Ed Engl, 2017. **56**(33): p. 9812-9815.

176. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* Journal of Molecular Biology, 1993. **234**(3): p. 779-815.

177. D'Mello, S.A., G.J. Finlay, B.C. Baguley, et al., *Signaling Pathways in Melanogenesis.* International Journal of Molecular Sciences, 2016. **17**(7): p. 1144.

178. Dolinska, M.B. and Y.V. Sergeev, *The consequences of deglycosylation of recombinant intra-melanosomal domain of human tyrosinase.* Biological Chemistry, 2017. **399**(1): p. 73-77.

179. Wang, N. and D.N. Hebert, *Tyrosinase maturation through the mammalian secretory pathway: bringing color to life.* Pigment Cell Melanoma Research, 2006. **19**(1): p. 3-18.

180. Berson, J.F., D.W. Frank, P.A. Calvo, et al., *A common temperature-sensitive allelic form of human tyrosinase is retained in the endoplasmic reticulum at the nonpermissive temperature.* Journal of Biological Chemistry, 2000. **275**(16): p. 12281-12289.

181. Park, H.Y., J.M. Perez, R. Laursen, et al., *Protein kinase C-beta activates tyrosinase by phosphorylating serine residues in its cytoplasmic domain.* Journal of Biological Chemistry, 1999. **274**(23): p. 16470-8.

182. Cronin, C.A., A.B. Ryan, E.M. Talley, et al., *Tyrosinase expression during neuroblast divisions affects later pathfinding by retinal ganglion cells.* Journal of Neuroscience, 2003. **23**(37): p. 11692-7.

183. Giménez, E., A. Lavado, P. Giraldo, et al., *A Transgenic Mouse Model with Inducible Tyrosinase Gene Expression Using the Tetracycline (Tet-on) System Allows Regulated Rescue of Abnormal Chiasmatic Projections Found in Albinism.* Pigment Cell Melanoma Research, 2004. **17**(4): p. 363-370.

184. Lopez, V.M., C.L. Decatur, W.D. Stamer, et al., *L-DOPA is an endogenous ligand for OA1.* PLoS Biology, 2008. **6**(9): p. e236.

185. McKay, B.S., *Pigmentation and vision: Is GPR143 in control?* Journal of Neuroscience Research, 2019. **97**(1): p. 77-87.

186. Dolinska, M.B., N.J. Kus, S.K. Farney, et al., *Oculocutaneous albinism type 1: link between mutations, tyrosinase conformational stability, and enzymatic activity.* Pigment Cell Melanoma Research, 2017. **30**(1): p. 41-52.

187. Halaban, R., E. Cheng, Y. Zhang, et al., *Aberrant retention of tyrosinase in the endoplasmic reticulum mediates accelerated degradation of the enzyme and contributes to the dedifferentiated phenotype of amelanotic melanoma cells.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(12): p. 6210-5.

188. Halaban, R., S. Svedine, E. Cheng, et al., *Endoplasmic reticulum retention is a common defect associated with tyrosinase-negative albinism.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(11): p. 5889-5894.

189. Tarentino, A.L., T.H. Plummer, Jr., and F. Maley, *The release of intact oligosaccharides from specific glycoproteins by endo-beta-N-acetylglucosaminidase H.* Journal of Biological Chemistry, 1974. **249**(3): p. 818-24.

190. Wolff, K. and H. Honigsmann, *Are melanosome complexes lysosomes?* Journal of Investigative Dermatology, 1972. **59**(2): p. 170-6.

191. Fedorow, H., F. Tribl, G. Halliday, et al., *Neuromelanin in human dopamine neurons: comparison with peripheral melanins and relevance to Parkinson's disease.* Progress in Neurobiology, 2005. **75**(2): p. 109-24.

192. Zhang, Q., B. Zhao, W. Li, et al., *Ru2 and Ru encode mouse orthologs of the genes mutated in human Hermansky-Pudlak syndrome types 5 and 6.* Nature Genetics, 2003. **33**(2): p. 145-53.

193. King, R.A. and C.J. Witkop, *Detection of heterozygotes for tyrosinase-negative oculocutaneous albinism by hairbulb tyrosinase assay.* American Journal of Medical Genetics Part A, 1977. **29**(2): p. 164-8.

194. Marchal, I., D.L. Jarvis, R. Cacan, et al., *Glycoproteins from insect cells: sialylated or not?* Biological Chemistry, 2001. **382**(2): p. 151-9.

195. Mehnert, P., J. Malchaire, B. Kampmann, et al., *Prediction of the average skin temperature in warm and hot environments.* European Journal of Applied Physiology, 2000. **82**(1-2): p. 52-60.

196. Shaw, G., S. Morse, M. Ararat, et al., *Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells.* FASEB Journal, 2002. **16**(8): p. 869-71.

197. Boulton, M.E., *Studying melanin and lipofuscin in RPE cell culture models.* Experimental Eye Research, 2014. **126**: p. 61-7.

198. Baulier, E., A. Garcia Diaz, B. Corneo, et al., *Generation of a human Ocular Albinism type 1 iPSC line, SEIi001-A, with a mutation in GPR143.* Stem Cell Research & Therapy, 2018. **33**: p. 274-277.

199. Onojafe, I.F., D.R. Adams, D.R. Simeonov, et al., *Nitisinone improves eye and skin pigmentation defects in a mouse model of oculocutaneous albinism.* Journal of Clinical Investigation, 2011. **121**(10): p. 3914-23.

200. Wang, F., W. Song, G. Brancati, et al., *Inhibition of endoplasmic reticulum-associated degradation rescues native folding in loss of function protein misfolding diseases.* Journal of Biological Chemistry, 2011. **286**(50): p. 43454-64.

201. Ando, H., Z.M. Wen, H.Y. Kim, et al., *Intracellular composition of fatty acid affects the processing and function of tyrosinase through the ubiquitin-proteasome pathway.* Biochemical Journal, 2006. **394**(Pt 1): p. 43-50.

202. Teramae, A., Y. Kobayashi, H. Kunimoto, et al., *The Molecular Basis of Chemical Chaperone Therapy for Oculocutaneous Albinism Type 1A.* Journal of Investigative Dermatology, 2019. **139**(5): p. 1143-1149.

203. Holmstrom, G., U. Eriksson, K. Hellgren, et al., *Optical coherence tomography is helpful in the diagnosis of foveal hypoplasia.* Acta Ophthalmologica, 2010. **88**(4): p. 439-42.

204. Ehrt, O., *Infantile and acquired nystagmus in childhood.* European Journal of Paediatric Neurology, 2012. **16**(6): p. 567-72.

205. Self, J., C. Mercer, E.M. Boon, et al., *Infantile nystagmus and late onset ataxia associated with a CACNA1A mutation in the intracellular loop between s4 and s5 of domain 3.* Eye, 2009. **23**(12): p. 2251-5.

206. Vincent, M.C., A.L. Pujo, D. Olivier, et al., *Screening for PAX6 gene mutations is consistent with haploinsufficiency as the main mechanism leading to various ocular defects.* European Journal of Human Genetics, 2003. **11**(2): p. 163-9.

207. Mataftsi, A., D.F. Schorderet, L. Chachoua, et al., *Novel TULP1 mutation causing leber congenital amaurosis or early onset retinal degeneration.* Investigative Ophthalmology and Visual Science, 2007. **48**(11): p. 5160-7.

208. Thomas, M.G., G.D. Maconachie, V. Sheth, et al., *Development and clinical utility of a novel diagnostic nystagmus gene panel using targeted next-generation sequencing.* European Journal of Human Genetics, 2017. **25**(6): p. 725.

209. Sun, Y., C.A. Ruivenkamp, M.J. Hoffer, et al., *Next-generation diagnostics: gene panel, exome, or whole genome?* Human Mutation, 2015. **36**(6): p. 648-55.

210. Dewey, F.E., M.E. Grove, C. Pan, et al., *Clinical interpretation and implications of whole-genome sequencing.* The Journal of American Medical Association, 2014. **311**(10): p. 1035-45.

211. Pajusalu, S., T. Kahre, H. Roomere, et al., *Large gene panel sequencing in clinical diagnostics-results from 501 consecutive cases.* Clinical Genetics, 2018. **93**(1): p. 78-83.

212. Wang, M., I. Newsham, Y.Q. Wu, et al., *High-Throughput Next Generation Sequencing Methods and Applications.* The Journal of Biomolecular Techniques 2011. **22**(Suppl): p. S7-S7.

213. Shi, Y. and J. Majewski, *FishingCNV: a graphical software package for detecting rare copy number variations in exome-sequencing data.* Bioinformatics, 2013. **29**(11): p. 1461-2.

214. Sikkema-Raddatz, B., L.F. Johansson, E.N. de Boer, et al., *Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics.* Human Mutation, 2013. **34**(7): p. 1035-42.

215. Perez, Y., L. Gradstein, H. Flusser, et al., *Isolated foveal hypoplasia with secondary nystagmus and low vision is associated with a homozygous SLC38A8 mutation.* European Journal of Human Genetics, 2014. **22**(5): p. 703-6.

216. Belkadi, A., A. Bolze, Y. Itan, et al., *Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants.* Proceedings of the National Academy of Sciences, 2015. **112**(17): p. 5473-5478.

217. Vail, P.J., B. Morris, A. van Kan, et al., *Comparison of locus-specific databases for BRCA1 and BRCA2 variants reveals disparity in variant classification within and among databases.* Journal of Community Genetics, 2015. **6**(4): p. 351-9.

218. Yang, S., S.E. Lincoln, Y. Kobayashi, et al., *Sources of discordance among germ-line variant classifications in ClinVar.* Genetics in Medicine, 2017. **19**: p. 1118.

219. King, R.A., R.K. Willaert, R.M. Schmidt, et al., *MC1R mutations modify the classic phenotype of oculocutaneous albinism type 2 (OCA2).* American Journal of Medical Genetics Part A, 2003. **73**(3): p. 638-45.

220. Lee, H., F. Proudlock, and I. Gottlob, *Is handheld optical coherence tomography reliable in infants and young children with and without nystagmus?* Investigative Ophthalmology and Visual Science, 2013. **54**(13): p. 8152-9.

221. Kruijt, C.C., G.C. de Wit, A.A. Bergen, et al., *The Phenotypic Spectrum of Albinism.* Ophthalmology, 2018. **125**(12): p. 1953-1960.

222. Puri, N., D. Durbam-Pierre, R. Aquaron, et al., *Type 2 oculocutaneous albinism (OCA2) in Zimbabwe and Cameroon: distribution of the 2.7-kb deletion allele of the P gene.* Human Genetics, 1997. **100**(5-6): p. 651-6.

223. Drubay, D., D. Gautheret, and S. Michiels, *A benchmark study of scoring methods for non-coding mutations.* Bioinformatics, 2018. **34**(10): p. 1635-1641.

224. Jian, X., E. Boerwinkle, and X. Liu, *In silico tools for splicing defect prediction: a survey from the viewpoint of end users.* Genetics in Medicine, 2014. **16**(7): p. 497-503.

225. Zou, J., M. Huss, A. Abid, et al., *A primer on deep learning in genomics.* Nature Genetics, 2019. **51**(1): p. 12-18.

226. Mossotto, E., J.J. Ashton, L. O'Gorman, et al., *GenePy - a score for estimating gene pathogenicity in individuals using next-generation sequencing data.* BMC Bioinformatics, 2019. **20**(1): p. 254.

227. Wei, A.H., D.J. Zang, Z. Zhang, et al., *Exome sequencing identifies SLC24A5 as a candidate gene for nonsyndromic oculocutaneous albinism.* Journal of Investigative Dermatology, 2013. **133**(7): p. 1834-40.

228. Ouwehand, W.H., *Whole-genome sequencing of rare disease patients in a national healthcare system.* bioRxiv, 2020: p. 507244.

229. Potjer, T.P., S. Bollen, A. Grimbergen, et al., *Multigene panel sequencing of established and candidate melanoma susceptibility genes in a large cohort of Dutch non-CDKN2A/CDK4 melanoma families.* International Journal of Cancer, 2019. **144**(10): p. 2453-2464.

230. Wang, X., K. Gregory-Evans, K.M. Wasan, et al., *Efficacy of Postnatal In Vivo Nonsense Suppression Therapy in a Pax6 Mouse Model of Aniridia.* Molecular Therapy - Nucleic Acids, 2017. **7**: p. 417-428.

231. Ando, H., H. Watabe, J.C. Valencia, et al., *Fatty acids regulate pigmentation via proteasomal degradation of tyrosinase: a new aspect of ubiquitin-proteasome function.* Journal of Biological Chemistry, 2004. **279**(15): p. 15427-33.

232. Wright, C.F., J.F. McRae, S. Clayton, et al., *Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders.* Genetics in Medicine, 2018. **20**(10): p. 1216-1223.