Autozygome-based analysis of Pakistani families with monogenic conditions

Ur Rehman Atta

Ur Rehman Atta, 2020, Autozygome-based analysis of Pakistani families with monogenic conditions

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_A79893DCEE441

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

CHUV - Division of Genetic Medicine

Autozygome-based analysis of Pakistani families with monogenic conditions

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Atta Ur Rehman

Masters in Human Genetics from Quaid-i-Azam University, Pakistan

Jury

Prof. Jardena Puder, Président

Prof. Andrea Superti-Furga, Directeur de thèse

Prof. Carlo Rivolta, Co-directeur de thèse

Prof. Susanne Kohl, Expert

Dr Stéphanie Christen-Zaech, Expert

Lausanne

(2020)

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Madame	Prof.	Jardena	**Puder**
Directeur·trice de thèse	Monsieur	Prof.	Andrea	**Superti-Furga**
Co-directeur·trice	Monsieur	Prof.	Carlo	**Rivolta**
Expert·e·s	Madame	Dre	Susanne	**Kohl**
	Madame	Dre	Stéphanie	**Christen-Zäch**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Atta Ur Rehman

Master in Human Genetics, Quaid-i-Azam University, Pakistan

intitulée

Autozygome-based analysis of Pakistani families with monogenic conditions

Lausanne, le 30 juin 2020

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale

Acknowledgments

I pay my deepest gratitude and heartiest thanks to my thesis co-director Prof. Carlo Rivolta, whose kind supervision, in time guidance, all the time availability and rectification of my mistakes with smiley face made it possible for me to complete my PhD. Indeed, this would not have been possible without his support, constructive critique and continuous encouragement. I am also very much grateful to Prof. Andrea Superti-Furga for his willingness to supervise my thesis and for providing me with all the pre-requisites for this work including a smooth transition from my previous lab to the new one.

I pay tribute to all members of my thesis committee. It's worthwhile to mention and appreciate the support provided by the president of my thesis committee Prof. Jardena Puder. Likewise, the time dedication and scientific input received from my external examiners Prof. Susanne Kohl and Dr. Stéphanie Christen-Zaech worth a lot appreciation.

I extend my deepest thanks to Rosanna who trained me with the lab techniques. Her continuous guidance, and selfless support will always be remembered. Many thanks indeed go to my lab fellows Nicola, Katarina, Mathieu, Virginie, Ikram, Kostas, and Francesco for their nice company and for their availability to support me whenever I asked them. Likewise, I pay my gratitude to Carole, Catarina and Belinda for their untiring assistance, and amazing company.

My friends in Lausanne, Ijaz Anwar, Kiran Shehzadi, Jamal Nasir, Amir Shehzad and Muhammad Ansar worth a lot of acknowledgments. Last but not least, sacrifices of my family during this journey are unforgettable. Indeed, I have no words to pay them my tribute. I love them so much.

“Where there's a will there's a way”

English proverb

Résumé

Le Pakistan a un des taux de maladies génétiques Mendéliennes le plus élevé du monde, corrélé avec un taux exceptionnellement élevé de mariages consanguins (~65%). Dans cette thèse, nous cherchons à comprendre la cause génétique de différentes maladies génétiques, principalement liées à l'appareil oculaire et à la peau, dans une cohorte de familles pakistanaises, en exploitant la consanguinité comme moyen de faciliter l'identification de mutations récessives.

En tout, nous avons identifié des variantes causales pour de telles conditions dans 98 des 116 familles analysées (taux de réussite de ~84%). Indépendamment des catégories de maladies, nos analyses ont aussi démontré un haut degré d'autozygotie et des mutations fondatrices prévalentes dans la population pakistanaise. Plus de 90% des familles analysées dans ce travail étaient liées à des variantes pathogènes homozygotes dans des gènes connus pour causer des maladies Mendéliennes et plus de 40% d'entre elles avaient des mutations fondatrices.

Collectivement, 61 variantes génétiques distinctes ont été identifiées dans 41 gènes sur un total de 86 familles avec des maladies héréditaires de l'œil. Pareillement, 12 variantes distinctes ont été identifiées dans 5 gènes pour 12 familles avec des conditions héréditaires affectant la peau. En tout, plus de la moitié des variantes génétiques présentées dans cette thèse n'avait pas été publiée auparavant. *ABCA4* était le gène le plus fréquemment muté dans les familles souffrant de maladies oculaires (13 sur 86), tandis que *HR* était le plus communément muté dans les familles avec des troubles de la chute des cheveux. Donc, nos découvertes reflètent raisonnablement la pratique traditionnelle de l'endogamie et la présence de stratifications extensives dans la population pakistanaise. En plus d'étendre le

paysage génétique des maladies héréditaires de l'œil et de la peau, nos découvertes aideront de futurs chercheurs et cliniciens à faire un screening rapide de leurs patients pour des mutations fondatrices, avant de s'embarquer dans un séquençage complet de l'exome ou du génome, facilitant ainsi le conseil génétique au Pakistan. Finalement, nous prédisons que ces découvertes au sujet des troubles affectant la vision ou la peau pourront s'appliquer à d'autres maladies rares présentes chez des individus de cette région, comme la surdité, la déficience intellectuelle ou des défauts de développement.

Abstract

Pakistan has one of the world's highest rates of inherited genetic diseases, likely correlating with the exceptionally high rate of consanguineous marriages (~65%). In this work, we sought to understand the genetic etiology of different monogenetic diseases, principally related to eyes and skin, in a cohort of Pakistani families, in fact by exploiting consanguinity as a means to facilitate identification of recessive mutations.

Overall, we identified causative variants for such conditions in 98 out of the 116 families analyzed (84% success rate). Independently from individual disease categories, our data demonstrate a very high degree of autozygosity and prevalent founder mutations in the Pakistani population. Thus, more than 90% of the pedigrees analyzed in this thesis were linked to homozygous pathogenic variants in known Mendelian disease genes, and over 40% of them had founder mutations.

Collectively, 61 distinct genetic variants were identified in 41 genes in a total of 86 families with inherited eye diseases (IEDs). Likewise, 12 distinct variants were identified in five genes among 12 families with inherited skin conditions. As a whole, more than half of the genetic variants presented in this thesis were not previously published. *ABCA4* was the most frequently mutated gene in families with inherited eye diseases (13 out of 86), while *HR* was commonly mutated in families with hair loss disorders (8 out of 11). Thus, our findings reasonably reflect on the traditional practice of endogamy and the presence of extensive stratification within the Pakistani population. In addition to expanding the current genetic landscape of inherited eye/skin diseases, our findings will help future researchers and clinicians to rapidly screen their patients for known founder mutations before embarking on whole exome or genome sequencing and facilitate genetic counselling in Pakistan. Finally, we

predict that these insights on eye and skin disorders might apply to other rare conditions that affect individuals from this region, such as deafness, intellectual disabilities, and developmental defects.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Main context

An unprecedented progress has been made in the field of genomics since completion of the Human Genome Project (HGP) in 2003, which opened new windows for understanding the genome in a broader context of human health and disease [1-4]. The discovery that human genome consists of ~19,000 predicted protein-coding genes [5, 6], and that pathological changes in either of them could likely lead to phenotypic consequences in humans has considerably contributed to the advancement of knowledge in the field of Medical Genetics. As a result, numerous novel phenotypes as well as their underlying genetic components, have been identified over the last few years, and this trend continues even today. In addition to increasing our understanding of the human molecular genetics, these investigations would help in the future to develop novel therapeutics, as well as diagnostic, preventive, or prognostic measures for an ever-increasing number of both rare and common human diseases [7]. According to an estimate by Bittles (2019), there are approximately 7,000 rare diseases, the majority of which appear to be the consequence of founder mutations, especially in societies where community endogamy and close-kin marriages are common [8].

Monogenic or single-gene disorders represent a category of inherited human conditions that arise due to mutation(s) in a single gene. Such conditions are often called Mendelian disorders, due to the fact that they follow a classical Mendelian inheritance pattern (autosomal dominant, recessive or X-linked). Though considered rare as a whole, monogenic diseases affect millions of people worldwide with a global prevalence of approximately 10 in 1,000 persons (WHO; accessed March 17, 2020). As an example, over 25 million people in the

United States alone are affected by various forms of monogenic disorders, resulting in a large morbidity and mortality, which in turn poses a significant burden to the economy [9, 10].

As of April 2020, the Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders, has reported 3,873 genes that are linked with as many as 5,572 human monogenic disorders and traits (https://omim.org). However, these numbers never remain static and continue to rise as new phenotypes and genotypes are discovered [7]. According to the WHO, the estimated number of human monogenic diseases should exceed 10,000 [11]. This is not surprising, because the physiological function of most human genes is still unknown [12]. In other words, nearly 70% of the human predicted protein coding genes are yet to be associated with any human phenotype [13, 14].

The current diagnostic rate for all monogenic disorders is still rather low (generally around 50%) [7], despite the substantial technological advancements in DNA sequencing [15]. The reasons for this low diagnostic yield are manifold. First, the diagnostic rates vary greatly across different disease classes, and that such rates are negatively correlated with the genetic complexity in a monogenic disorder [7, 16, 17]. For example, a recent study suggests a diagnostic rate of only 17% for renal diseases, 42% for intellectual disability, 56% for IRDs and non-syndromic deafness, while 76% in case of ciliary dyskinesia [17]. Second, the identification of new phenotypes, discovery of novel genes associated with known human diseases, and the detection of unknown pathogenic variants further preclude the diagnostic process. Third, with the existing annotation tools, the current annotation of human genome is sub-optimal. There are incomplete gene models, and the precise structure and number of several human genes is still unknown. Fourth, robust analytical and computational

approaches would be required to keep pace with the rapidly changing and highly demanding variant and/or gene prioritization strategies [6, 12, 16, 18, 19].

Much remains to be understood. Our current knowledge of the human genome is largely based on its coding part (exome) which collectively represent only a small fraction of it (~1-2%), although the physiological function of many genes is still unknown [12]. In addition, the functional consequences of the non-coding part of the genome are far from being clear [7, 16]. Thus, the understanding of the non-coding part of the human genome (nick-named "grey matter") needs equal attention, due to the fact that ~80% of the non-coding genome may have a distinct biological function [20].

1.2 Impact of consanguinity on monogenic diseases

Globally, around one billion people live in populations where consanguinity is common, and the majority of them belong to countries in South and West Asia, the Middle East, and the North and sub-Saharan Africa [21, 22]. Nevertheless, high rates of consanguinity have also been reported in countries from South America, such as Northwestern Venezuela and Southeast Brazil [23]. In Pakistan specifically, the rate of consanguinity generally exceeds 50%, since marriages among first cousins are highly favored (Figure 1) [24]. As children of consanguineous couples are generally more likely than children of non-consanguineous couples to be affected by recessively inherited genetic anomalies, the incidence of rare autosomal recessive Mendelian diseases is relatively higher in countries like Pakistan [22, 25]. For instance, in children of first cousin parents, which share 12.5% of their genome, there is a higher likelihood of ancestral chromosomal segments carrying recent pathogenic mutations to reunite. Thus, consanguinity unmasks the adverse health effects of recessive mutations through bi-parental inheritance of the same mutation in a homozygous state (Figure 2).

Further, the rarer the mutation, the more likely it is to be detected through the analysis of consanguineous pedigrees [26].

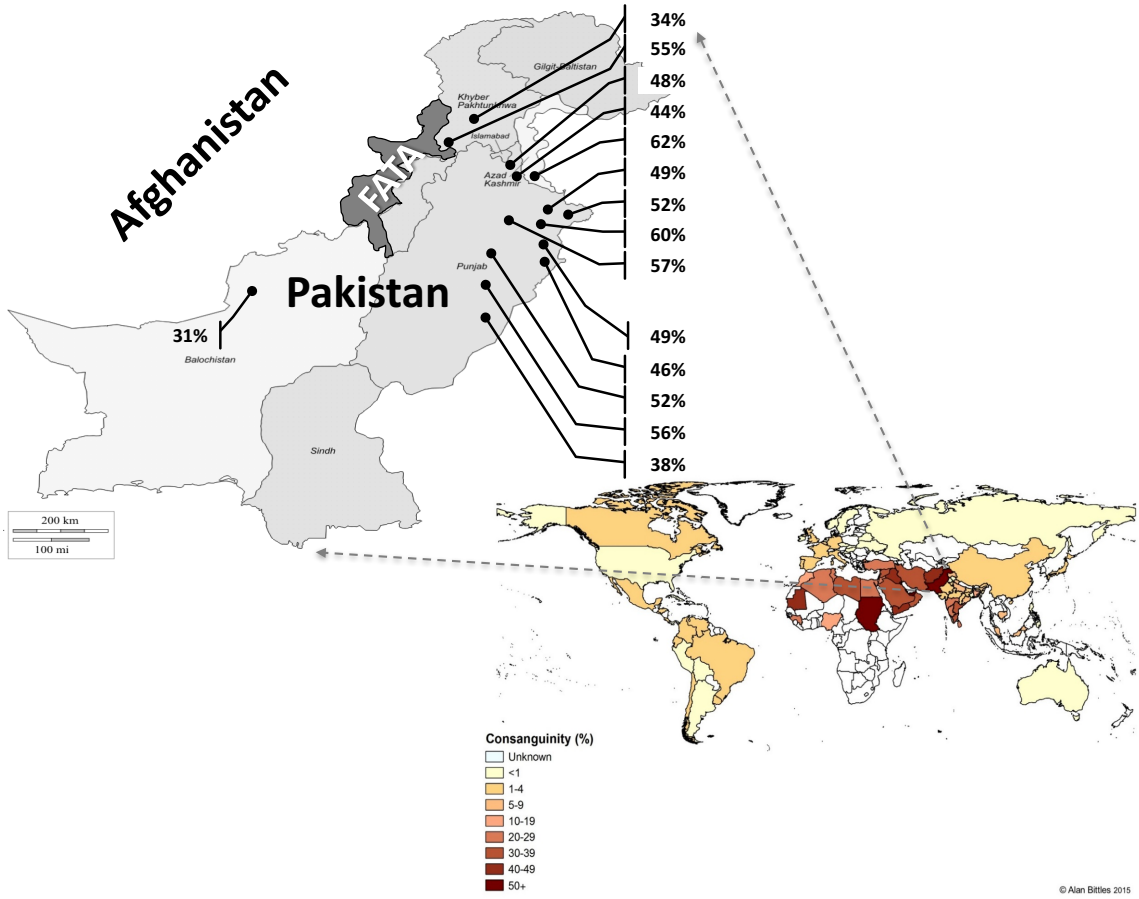


Figure 1. Global prevalence of consanguinity with Pakistan shown in the zoom-in.

Consanguinity in different parts of Pakistan are shown in percent, whereas the grey-color coding in Pakistan represent major administrative sub-units in the country. FATA = Federally Administered Tribal Areas. Adapted from published papers [22, 27].

1.3 Monogenic diseases of the eye

The development of the human eye is a complex and highly coordinated morphogenetic process, completed under an extensive network of genetic signaling, and ensured by a precise spatial and temporal gene regulation [28, 29]. Any perturbations (genetic or environmental) during this process may lead to irreversible malformations or dysfunctions of the eye [30]. Further, depending upon the underlying causes, these defects may be structural (for example: anophthalmia, microphthalmia, ocular coloboma, cataracts, corneal dystrophies, and glaucoma) or more functional in nature (for instance, inherited retinal dystrophies or IRDs) [31, 32]. A comprehensive list of genes implicated in various compartments of the eye could be obtained from "PanelApp", a publicly available tool developed by Genomics England [33]. This tool is regularly updated and curated manually. As of April 17, 2020, eight gene panels corresponding to "ophthalmological disorders" are listed in PanelApp. These panels include anophthalmia or microphthalmia (34 genes), cataracts (99 genes), corneal abnormalities (32 genes), developmental glaucoma (10 genes), infantile nystagmus (10 genes), ocular coloboma (29 genes), optic neuropathy (21 genes), and retinal disorders (201 genes) [33]. The latter constitute a genetically heterogeneous group of rare genetic conditions that are mainly characterized by the progressive loss of rods and cone photoreceptors, resulting in complete or nearly complete blindness at the end [34]. Globally, IRDs affect approximately one million people, with a frequency of 1 in 3,000 births. Clinically, IRDs may range from mild and usually non-progressive night blindness with or without difficulty in color vision to a more severe and degenerative phenotypes, such as retinitis pigmentosa or cone and cone-rod dystrophies [35]. According to Retinal Information Network (RetNet), mutations in over 270 genes have been so far implicated in various forms of IRDs (https://sph.uth.edu/retnet/; accessed on April 20, 2020)[36], and by sequencing the coding part of these genes it is possible to explain

the genetic basis of the disease in approximately 60% of the patients. IRDs are inherited as an autosomal recessive, autosomal dominant, or X-linked trait, with autosomal recessive being the most prominent one [37]. Recent technological advancements have significantly increased gene discovery rates [38], and have led to the development of gene replacement therapy for patients with biallelic mutations in the *RPE65* gene.

Although the prevalence of IRDs is not well-documented in Pakistan, a hospital-based study in Karachi, a metropolitan city, revealed that 1 in 800 patients that visited the ophthalmic outpatient department had retinal dystrophies, with retinitis pigmentosa (RP) being the most frequent type (64%) followed by Stargardt disease (14.7%), and cone dystrophies (6.7%). Unsurprisingly, more than half of these patients were born to consanguineous couples [39]. Interestingly, in IRD families of Pakistani origin, 90% of the mutations in non-syndromic cases have not been observed in other populations, and mutations in 35 different genes have been found to cause non-syndromic IRDs specifically in families of Pakistani origin [40]. Keeping in mind the high rates of consanguinity and the unique socio-cultural structure of the society, the occurrence of rare autosomal recessive Mendelian diseases in Pakistan is easily explained.

1.4 Monogenic diseases of the skin

The skin is the largest organ in the human body, and isolates and protects us from a wide range of external insults, be them of physical, chemical or microbial nature. Keratinocytes, which represent a continuously renewing cellular compartment of the human skin [41], perform most of these functions through a highly coordinated and dynamic balance between its differentiation, proliferation, and regeneration processes [42]. The list of inherited skin diseases is extensive, due to the cellular and structural complexity associated with derma-epidermal architecture [42]. However, consistent with the scope of my thesis, I will briefly

describe two major dermatological conditions that are attributable to genetic alterations. These include congenital hair loss disorders and inherited skin fragility.

Hereditary hair loss disorders or hypotrichosis are a group of conditions clinically characterized by sparse to complete loss of hairs on the scalp and the whole body. Both syndromic and non-syndromic forms of hypotrichosis exist, with the syndromic forms mostly accompanied by symptoms such as retinal degeneration, intellectual disability, and hearing impairment. Non-syndromic forms may be inherited in an autosomal dominant or recessive fashion, with mutations known in almost an equal number of genes for each category [43]. Mutation in genes such as *APCDD1*, *CDSN*, *KRT74*, *U2HR*, *EPS8L3*, *SNRPE*, and *RPL21* results in autosomal dominant hypotrichosis, while mutations in *HR*, *DSG4*, *LIPH*, *LPAR6/P2RY5*, *DSP*, and *DSC3* genes are known to cause autosomal recessive hair loss. Most interestingly, screening only four of these genes (*HR*, *DSG4*, *LIPH*, *LPAR6*) has been able to explain the genetic etiology of the disease in 87% cases of Pakistani origin [43]. Despite the fact that the majority of hypotrichosis cases are well-explained by one of these genes, there are still examples of genetically unresolved cases. The best examples include hypotrichosis type 9 and type 10, with associated loci known on chromosome 10 (10q11.23-q22.3) and chromosome 7 (7p22.3-p21.3), respectively. Nevertheless, no definitive gene has been yet associated with these disorders [43].

One type of inherited skin fragility, epidermolysis bullosa (EB), is characterized by the inability of the skin and mucous membranes to withstand external mechanical stress [42]. Thus, trauma-induced skin eruptions, blistering and painful wounds constitute the major clinical landmarks of this disorder, although the involvement of other tissues such as teeth, nails, hairs, and mucosal membranes further expands its clinical spectrum [42]. Affecting ~1 in

120,000 people worldwide [44], EB possess extensive clinical and genetic heterogeneity. For instance, there are over 30 known clinical variants of EB, with attributable sequence variants known in 18 distinct genes [45]. The majority of these genes encode structural proteins that play a significant role in keeping the dermal-epidermal junctions intact [42]. Consistent with the complexity of skin adhesion structures and cellular types, the number of skin fragility disorders continues to rise, in parallel with our knowledge about these structures [42].

1.5 Autozygosity mapping

Mapping all the autozygous intervals in an individual's genome is called autozygosity or homozygosity mapping [46]. Traditionally, it has been performed using minisatellites or microsatellite markers. With the advent of next-generation sequencing (NGS) technologies such as high-density SNPs, whole exome- and genome sequencing, tracking such intervals is now much faster and more reliable than previous methods [46]. If the disorder shows a recessive inheritance pattern, once the autozygous intervals are identified, screening of relevant candidate genes inside the region(s) is the next step. Since frequencies, length and genomic distribution of autozygous intervals vary considerably among individuals, it is crucial to scrutinize the list of intervals based on mutual sharing between patients but not healthy individuals within the same pedigree. Generally, the likelihood of finding a pathogenic mutation increases if two or more affected members of a family share the same autozygous interval. Similarly, the more closely related the parents are, the more likely for a homozygous pathogenic mutation to be found within one of the largest intervals and vice versa [46-48]. For example, Wakeling and colleagues [48], while assessing the predictive power of homozygosity mapping in recessive Mendelian diseases, have demonstrated that size and rank of the homozygous interval harboring the pathogenic variant provides an additional evidence for its causality.

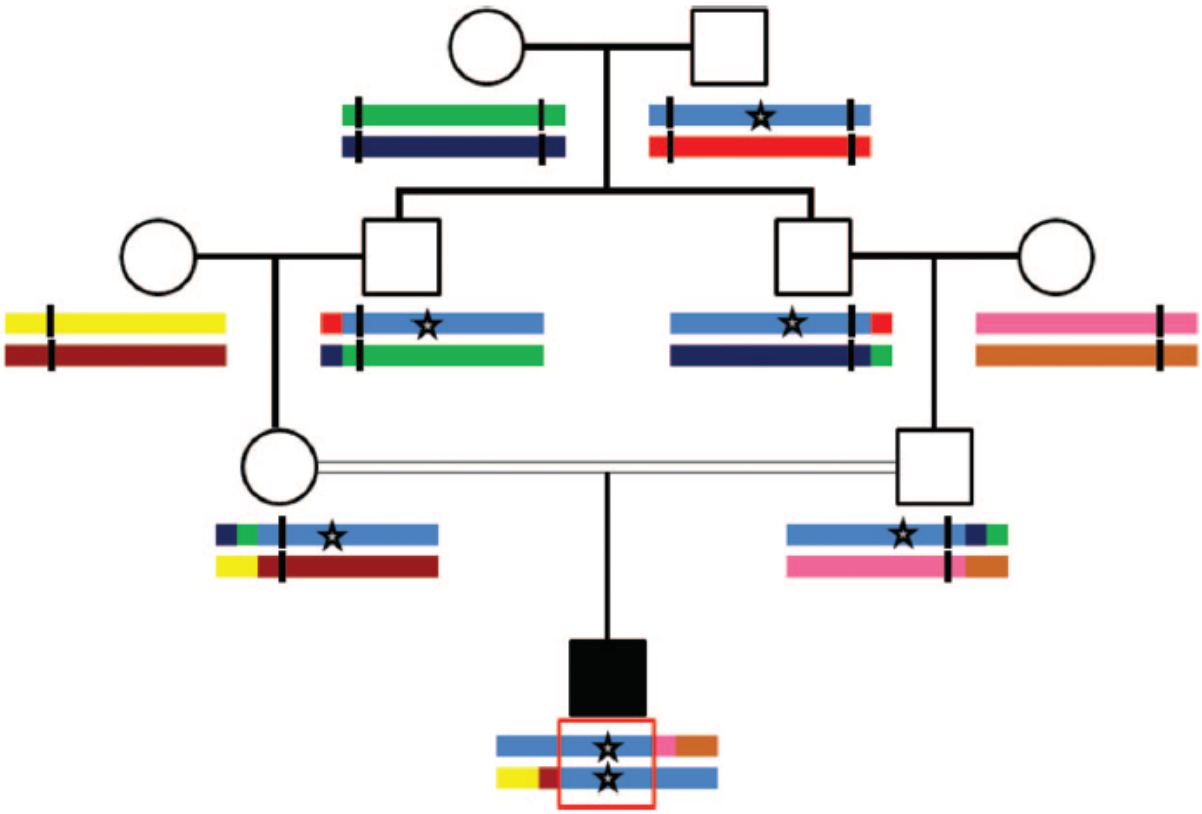


Figure 2. Schematic representation of autozygosity mapping.

Red box indicates an autozygous interval harboring a homozygous mutation (shown as star) inherited bi-parentally through consanguinity. Adapted from [26].

2. Results

2.1 An overview of the workflow

A total of 116 consanguineous families suffering from monogenic diseases were ascertained, mainly from the North-Western and central regions of Pakistan. These families were classified into two major clinical categories i.e. inherited eye diseases (IEDs) (100 families), and skin disorders (16 families). A combination of methods such as direct Sanger sequencing of known disease-associated gene or known founder variants, SNP-based autozygosity mapping followed by Sanger sequencing of candidate genes inside the autozygous intervals, and/or whole exome sequencing on one proband per family were applied for detection of the underlying genetic causes (Figure 3). For example, initial Sanger sequencing of candidate genes within autozygous intervals detected by SNP-based approaches in fourteen IEDs families resulted in the identification of causative variants in six families. Furthermore, Sanger sequencing-based screening of nineteen IEDs families for a known founder variant (p.Gly72Arg) in the *ABCA4* gene revealed the underlying genetic causes in eight families. Finally, whole exome sequencing (WES) was performed in 86 IEDs probands, thus enabling us to genetically characterize 72 additional families (83% diagnostic rate for WES) (Figure 3). Altogether, our approach led to the discovery of the genetic etiology in 86 families with IEDs. Similarly, Sanger sequencing of all exons and exon-intron boundaries of *HR* and *LIPH* genes revealed the underlying genetic causes in nine of the total fourteen families with either alopecia universalis congenita or wooly hairs/hypotrichosis analyzed. Likewise, the genetic components in three skin families were identified through whole exome sequencing. Consequently, 12 of the total 16 skin families (75%) were genetically characterized (Figure 3).

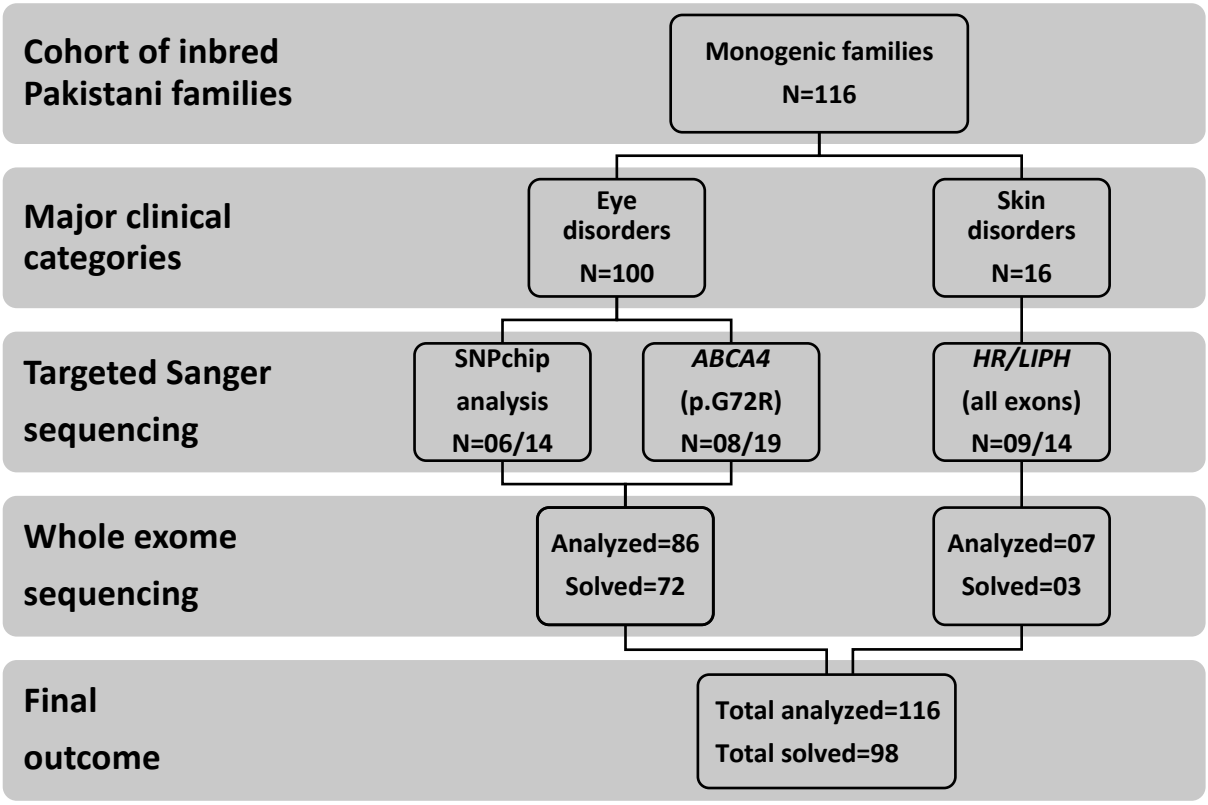


Figure 3. An overview of the workflow adopted in this study.

Numbers indicate the number of families.

2.2 Genetic findings in inherited eye disorders

2.2.1 An overview of the genetic and allelic heterogeneity associated with eye disorders

Among the 86 genetically characterized IEDs families, a total of 61 distinct genetic variants were identified in 41 different IED-associated genes (Table S1). Of the 41 genes carrying pathological changes, 32 were previously associated with inherited retinal diseases (IRDs), while 9 genes were already known to cause other forms of ocular diseases such as glaucoma, congenital cataract, microphthalmia/anophthalmia, and corneal dystrophy (Figure 4). These variants comprised 28 missenses (46%), 17 nonsense variants (28%), 10 small insertions or deletions (Indels) (16%), 5 splice site changes (8%), and one structural variant (2%) (Figure 5A). Of note, the structural variant was a large homozygous deletion (~11 Kb in size) encompassing the *RP1* gene and was identified in two unrelated consanguineous families from Northern Pakistan. Although we confirmed the presence of this deletion in both families with the help of PCR, the exact boundaries of this deletion are yet to be defined. As a whole, more than half (51%) of the disease-causing variants identified in our IEDs cohort were previously unpublished, and nearly half (49%) of the total variants were not reported in the public databases such as gnomAD database. In addition, those that were present in gnomAD were all very rare, with no occurrence of homozygous individuals. The variants were predominantly detected in a homozygous state (80 out of 86 families or 93%), although compound heterozygous (three families, 4%), heterozygous (two families, 2%), and hemizygous conditions (one family, 1%) were also observed (Figure 5B). Except for two alleles, all homozygous variants were found in genes that were located inside a large runs of homozygosity (ROH), typically more than 2 Mega bases (Mb) in size. An average homozygosity of 273 Mb (Range: 707.17 to 57.56 Mb) was recorded in our total cohort (Figure 6).

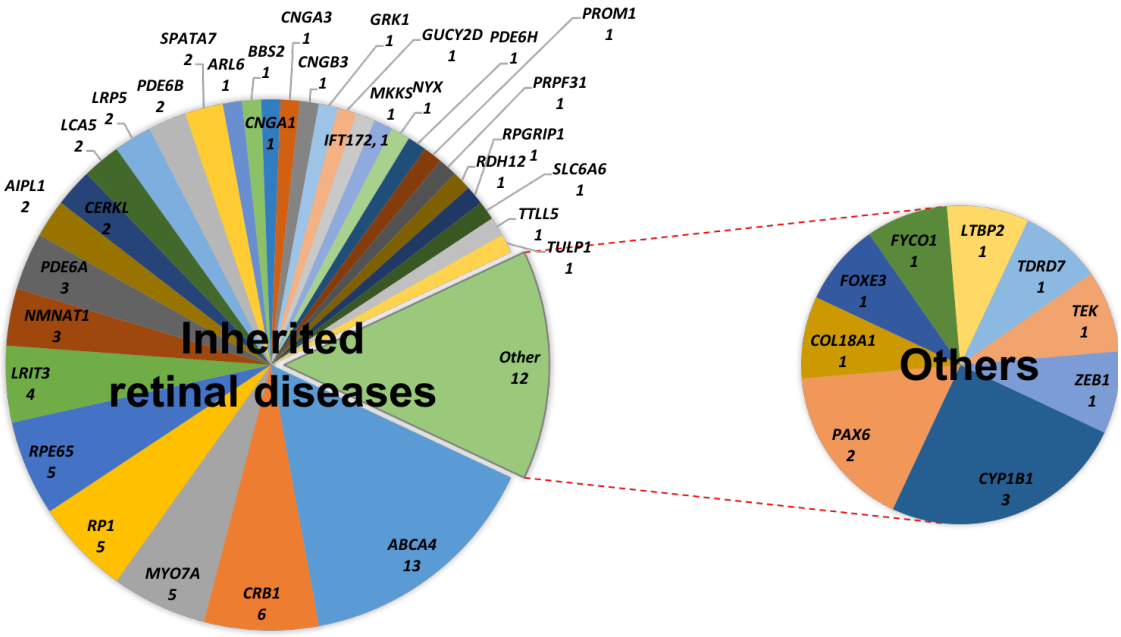


Figure 4. Pie chart showing genes harboring disease-causing variants in ocular cases.

Number indicates number of families associated with a specific gene. “Others” include cases of families with non-retinal disorders such as glaucoma, congenital cataract, microphthalmia/anophthalmia, and corneal dystrophy.

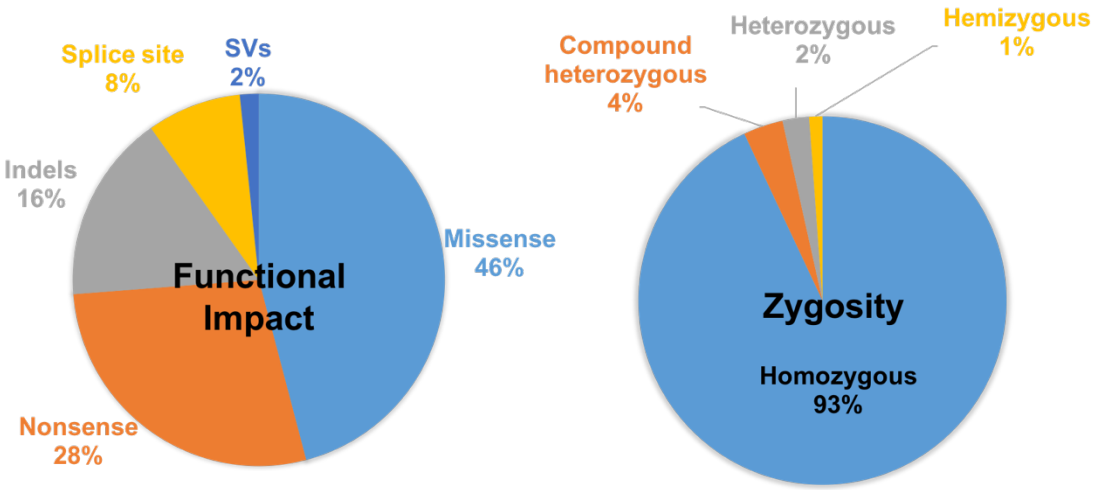


Figure 5. Classification of pathogenic variants based on their functional impact and zygosity.

(Indels = small insertions or deletions, SV = structural variants).

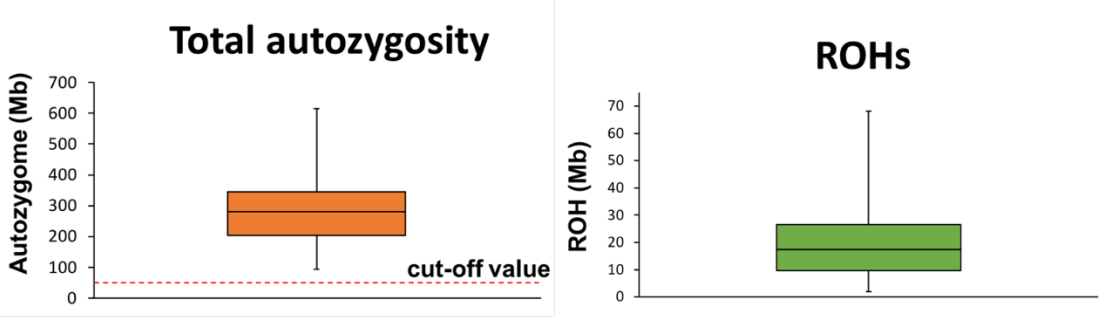


Figure 6. Box plots showing total autozygosity and runs of homozygosity (ROH). ROH refers to the autozygous intervals harboring disease-causing variants.

2.2.2 Founder variants in IRD associated genes account for 44% of IEDs

Among the total 61 disease-causing variants, ten variants spanning nine different IRD associated genes appeared in two or more independent families (Figure 7). Initially, we noticed that families harboring same genetic variants were all geographically clustered, thus pointing towards founder effect. Therefore, reanalysis of the WES data, through visualization of bam files on Integrative Genomics Viewer (IGV) tool, we have found common haplotypes surrounding the putative founder variants among the probands of the respective families. However, the exact length of the common haplotypes surrounding the founder variants could not be calculated. Our findings thus confirmed that these variants were inherited from a common ancestor, and thus truly represent founder alleles. Collectively, these variants appeared in 38 independent families, thus accounting for 44% of all IEDs cases. Genes carrying founder variants comprised *ABCA4*, *CRB1*, *LRIT3*, *RPE65*, *NMNAT1*, *PDE6A*, *CERKL*, *RP1*, and *MYO7A* (Figure 7). Of these ten founder alleles, a missense variant (p.Gly72Arg) in the *ABCA4* gene alone was present in 11 distinct families all originating from same geographic locality in North-Western Pakistan, and thus emerged as the most common variant in our study (responsible for 12% of IEDs cases).

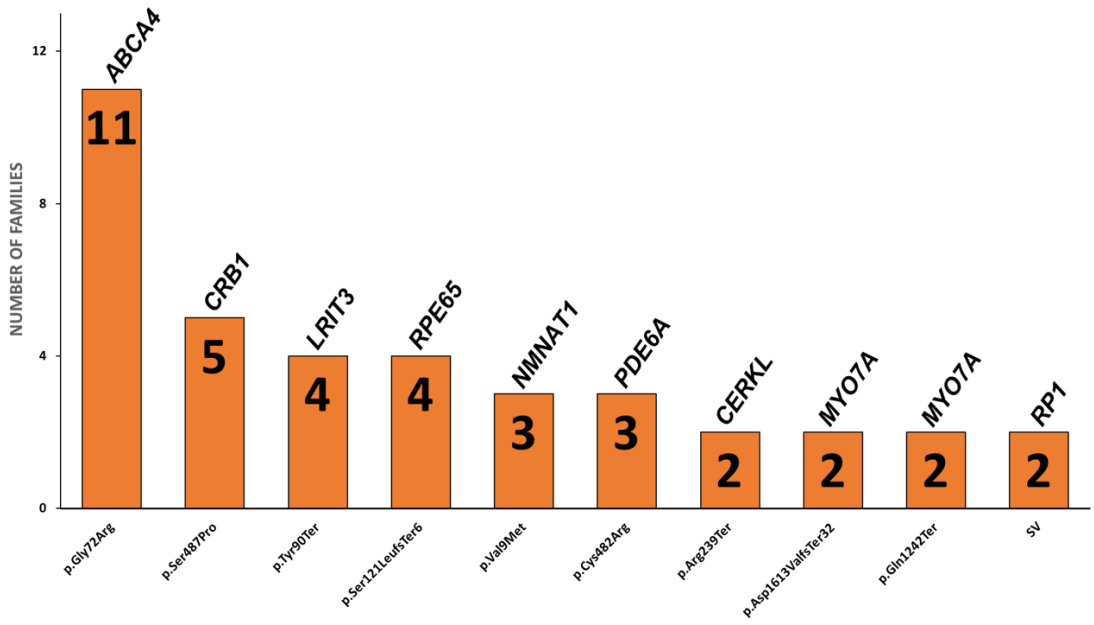


Figure 7. Putative founder variants found in IEDs.

Figures inside each bar represent number of families associated with a particular variant. (SV: structural variant).

2.2.3 *SLC6A6*-related taurine transporter deficiency disorder in a consanguineous family

We identified a consanguineous Pakistani family, of *Pashtuns* ethnicity, with an ultra-rare genetic condition known as taurine transporter (TauT) deficiency disorder. The family comprised of nine siblings (four affected, five clinically unaffected) who were born to a clinically healthy, consanguineous couple (Figure 10). Thus, inheritance of the disease was consistent with an autosomal recessive Mendelian pattern. According to the ophthalmologist report, best corrected visual acuity (BCVA) in both eyes of the index patient was hand movement (Good perception and projection of light in all quadrants). He was having nystagmus and was having hyperopia (3D) in both eyes. Anterior segment examination in both eyes was unremarkable. Fundus examination showed bilateral optic disc margin's blurring (somewhat lumpy margins suggesting disc drusen), retinal pigment epithelium (RPE)

changes in peripheral retina with few retinal pigment clumps. His macula was showing features of pucker/epi-retinal membrane in both eyes. Proband's electroretinography (ERG) examination showed severely reduced cones and rods response, suggesting severe retinal degeneration (Figure 8 and 9). In addition, the index patient showed dilated cardiomyopathy on electrocardiogram (ECG) examination. WES in the index patient revealed a novel homozygous missense variant (NM_001134367.3:c.1049C>T:p.Thr350Ile) in the *SLC6A6* gene. The variant (p.Thr350Ile) was unreported in any public databases, and predicted to substitute a nucleophilic residue in the highly conserved 5th transmembrane domain of the TauT. Furthermore, the variant was predicted to be pathogenic on majority of the online *in-silico* methods. Finally, a strict genotype-phenotype co-segregation within the family confirmed causality of p.Thr350Ile variant in the family (Figure 10).Though well studied in model animals, human TauT deficiency disorder constitute an extremely rare human genetic condition. Only few studies describing human TauT deficiency disorder are currently available in the literature and much remains to understand about this ultra-rare human genetic condition [49-51]. The *SLC6A6*-related early-onset retinal degeneration and cardiomyopathy seen in our patients not only well-matched with the recently published reports [49-51] (Table 1), but also supports the multisystemic phenotypes seen in the TauT knock-out mice [50].

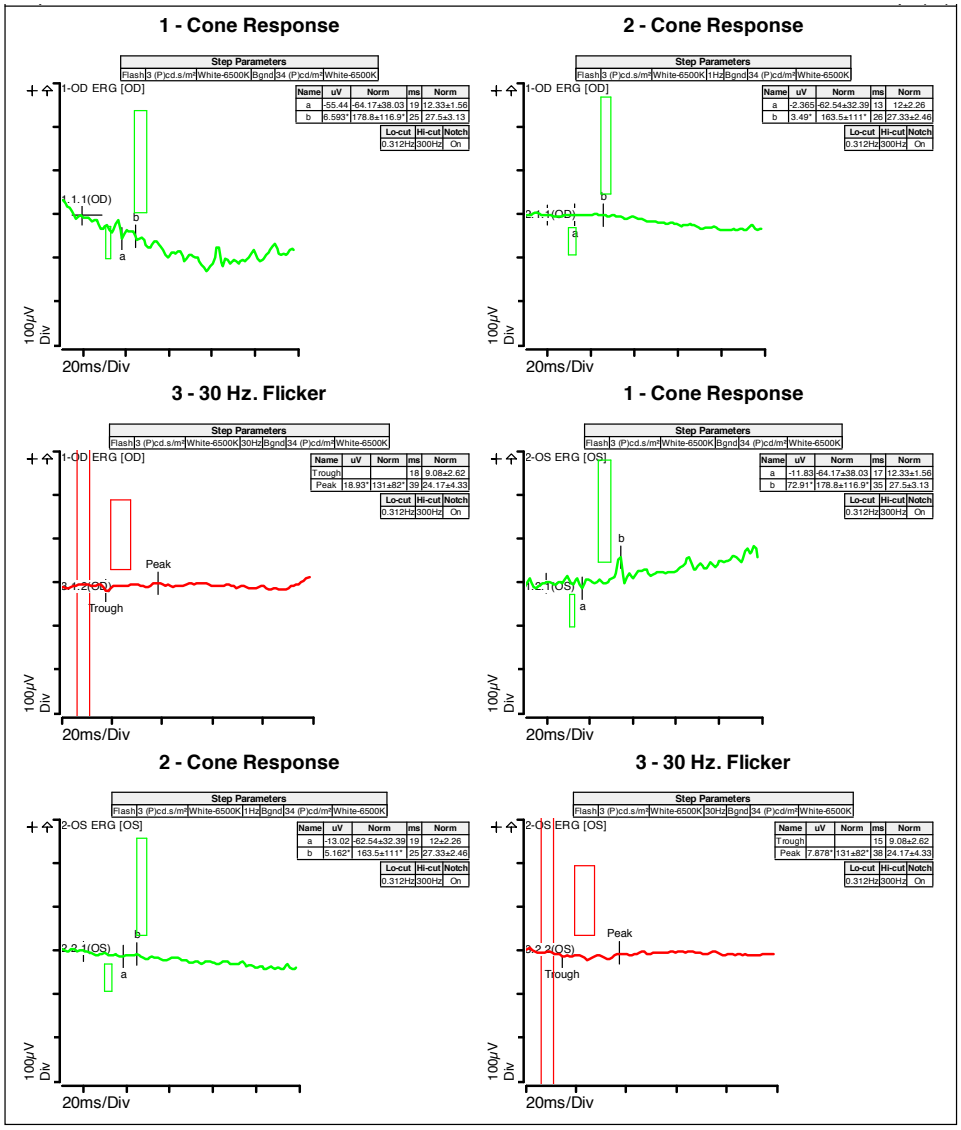


Figure 8. *SLC6A6* deficiency: Proband's ERG showing reduced cones response

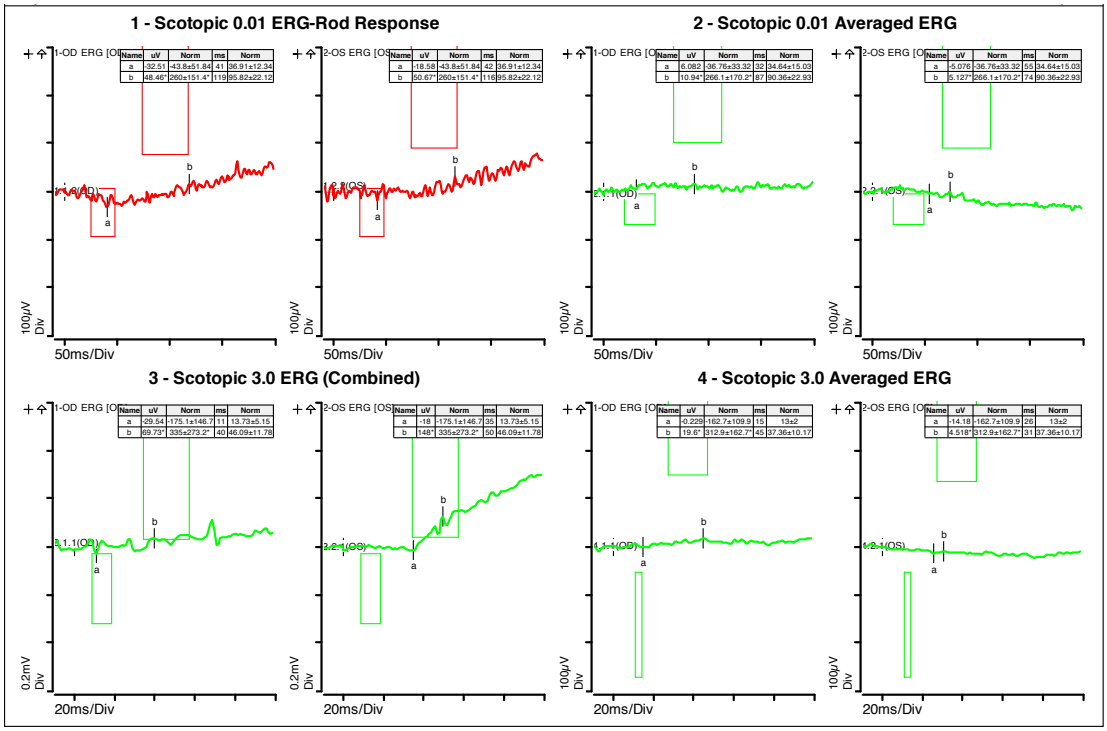


Figure 9. *SLC6A6* deficiency: Proband's ERG showing reduced rods response.

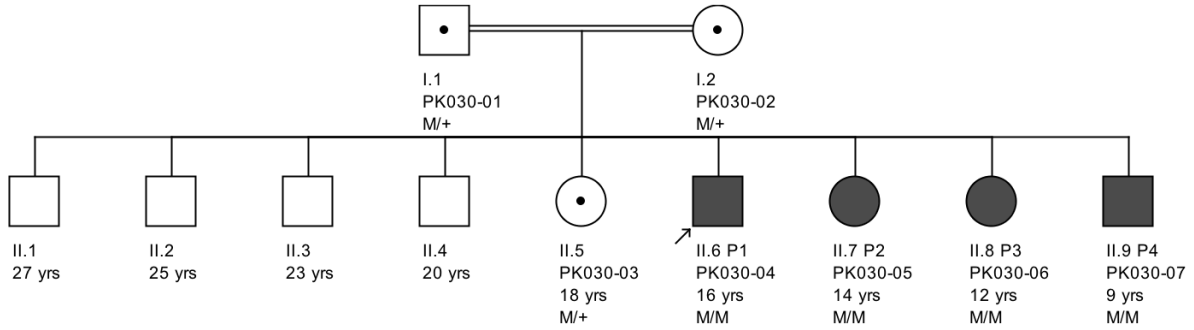


Figure 10. A consanguineous pedigree showing genotype-phenotype co-segregation for *SLC6A6* variant.

(M = NM_001134367.3:c.1049C>T:p.Thr350Ile). Arrow indicate proband.

Table 1. Literature survey about *SLC6A6*-related human TauT deficiency disorder

Phenotypes*	**cDNA**	**Protein**	**Zygosity**	**Ethnicity**	**Reference**
DCM	c.229+1delT	p.(?)	Homozygous	Pakistan	Shakeel et al. 2018
EORD	c.233C>A	p.(Ala78Glu)	Homozygous	Turkey	Preising et al. 2019
EORD+DCM	c.1196G>T	p.(Gly399Val)	Homozygous	Pakistan	Ansar et al. 2020
EORD+DCM	**c.746C>T**	**p.(Thr249Ile)**	**Homozygous**	**Pakistan**	**This study**

DCM = dilated cardiomyopathy; EORD = early-onset retinal degeneration.

Publications

These results have been either published in an article in “Genes” (Section 7.2.1) or are part of a submitted manuscript that is currently under review in “Human Mutation” (Section 7.3.2)

2.3 Genetic findings in skin cases

2.3.1 A hotspot mutation in the *COL7A1* gene causes dystrophic epidermolysis bullosa

This section focuses on the genetic findings in a consanguineous Pakistani family with epidermolysis bullosa (EB). The family was ascertained from North-Western Pakistan and includes a total of four affected children with recessive dystrophic EB, the most severe form of the disease. The patients, aged between 3 and 12 years, were all born to healthy parents who were mutually double-first cousins. Clinically, the proband presented with severe skin fragility and multiple skin injuries since birth. Although blistering of the body was generalized, the condition was more pronounced on her hands, feet, elbows, knees, and flexural areas. Additional symptoms included milia, atrophic scarring, anonychia, tissue granulation, blistering of oral cavity and the mucosa, anemia, and growth retardation. Pseudo-syndactyly of toes along with progressive contractures resulted in 'mitten-feet'. The spectrum of clinical manifestations was variable across the patients. For instance, the proband's brothers and one cousin were lacking symptoms like milia, atrophic scarring, chronic wounds, anonychia, tissue granulations, growth retardation, and pseudo-syndactylism of toes and fingers. Moreover, the percentage of damaged skin in these patients was significantly lower than in the proband on average (10% vs. 40%, respectively).

We performed whole exome sequencing (WES) in the proband, which revealed a homozygous missense mutation (c.8038C>T:p.Gly2680Ser) in exon 108 of the *COL7A1* gene. Homozygosity mapping confirmed that *COL7A1* lied inside a large autozygous genomic interval of 43.19 Mb, on chromosome 3. Residue Gly2680 is located within the triple helical domain of collagen VII and was found to be highly conserved across many vertebrate species, both at nucleotide and amino acid levels. Furthermore, Gly2680Ser is an extremely rare allele (allele frequency:

1.657e-05, in gnomAD) with no homozygotes reported so far. Finally, strict genotype-phenotype co-segregation of this mutation in the family confirmed its causality. Although glycine substitutions at this specific codon of collagen VII were previously reported in a heterozygous, as well as in a compound heterozygous state [52-55], to the best of our knowledge, it was never identified in a homozygous state. Interestingly, heterozygous individuals in our family did not show any clinical sign. Thus, we suggest that this mutation can be pathogenic per se, and possibly associated with generalized intermediate RDEB. Since glycine substitution at codon 2680 of collagen VII were previously reported in patients from diverse ethnic backgrounds such as China, UK, Poland, and Iran, we speculate that this mutation most likely constitutes a recurrent mutational hotspot in the *COL7A1* gene.

Publication

This project has been published in “Clinical Dysmorphology” as first author article. Section 7.2.2.

2.3.2 Clinical findings in families with hair loss disorders

A total of 16 consanguineous families with hereditary hair loss disorders were ascertained from Punjab province of Pakistan with the help of our collaborator in Islamabad. The clinical spectrum of the disorder in these families ranged from complete absence of hairs on the whole body (alopecia universalis congenita, OMIM # 203655) to congenital absence of scalp hairs with sparse eyebrows and eyelashes (hypotrichosis with/without woolly hairs, OMIM # 604379) (Figure 11). Nevertheless, secondary symptoms such as dermatitic lesions, excessive sweating, sensitivity to sun, itching, dry skin, delayed bleeding upon injuries/wound, rapid nails growth, and crowded teeth were also evident in some families (Table 2).

Figure 11. Clinical presentation of probands showing hair loss disorders

Table 2. Major clinical features observed in families with hereditary hair loss disorders

Family ID	Disease onset/status	Body hairs	Dermatitic lesions	Excessive sweating	Itching	Soft skin	Excessive bleeding	Rapid nails growth	Crowded teeth
PK-L	Congenital	-	+	-	-	+	-	-	-
PK-N	Congenital	-	-	-	-	-	-	-	-
F117	Congenital	-	+	+	+	+	+	+	+
F118	Progressive	-	-	+	+	+	+	+	+
F134	Progressive	-	-	+	+	+	+	+	+
F164	Progressive	-	-	+	+	+	+	+	+
F165	Progressive	-	-	+	+	+	+	+	+
F166	Progressive	-	-	+	+	+	+	+	+
F168	Progressive	-	-	+	+	+	+	+	+
F64	Congenital	WH	-	-	-	-	-	-	-
F65	Congenital	HT	-	-	+	+	+	+	+

(+) presence, or (-) absence of a clinical feature, WH = wooly hairs, HT = Hypotrichoses.

2.3.3 Genetic findings in families with hair loss disorders

Collectively, we identified disease-causing variants in 11 of the total 15 pedigrees analyzed through either direct Sanger sequencing of the *HR* and *LIPH* genes (nine families) or whole exome sequencing in one proband per family (two families). Thus far, genetic findings have been inconclusive in four probands with whole exome sequencing data available. The mutational spectrum in the 11 genetically solved families include nine distinct variants reported in four separate genes (*HR*, *LIPH*, *LPAR6*, *DSP*) (Table 3). For example, five alleles were reported in *HR* gene, two alleles in the *LIPH* gene, while one allele each was found in *LPAR6* and in *DSP* genes. Four of these nine variants were never published before. With the exception of one family with compound heterozygosity in the *LIPH* gene, all patients had homozygous variants in the *HR*, *LPAR6*, and *DSP* genes, thus reflecting on the practice of endogamy in the country. As expected, majority of the families (8 out of 11) had disease-causing variants in the *HR* gene, thus supporting data from existing literature [2, 56],

Of the five *HR* variants, four represented loss-of-function alleles and were all reported in the proximal part of HR protein, while one variant was a missense substitution (p.Arg1095Trp) and was located inside a functionally important Jmjc domain of HR protein that possess histone demethylase activity (Figure 9). Interestingly, two of the five alleles reported in the *HR* gene could be considered as founder variants, due to their recurrence in five independent families of our cohort. These include p.Trp612GlyfsTer4 (two families), and p.Pro144LeufsTer24 (three families). Since these variants were identified through Sanger sequencing, no further data were available to perform haplotype analysis. Similarly, two frameshift mutations in the *LIPH* gene

(p.Val437GlyfsTer4; p.Ile220ArgfsTer25), previously known to be founder alleles in Pakistan [57], were detected in compound heterozygosity in one family. WES analysis revealed a novel frameshift deletion (p.Phe24HisfsTer29) in *LPAR6*, co-segregating with hypotrichosis as a recessive trait in one family. Finally, a previously-known [58] pathogenic variant (p.Pro498Leu) in *DSP* gene was reported in a homozygous state in one family with hereditary hair loss disorder, through WES analysis (Table 3).

Table 3. Genetic variants identified in 11 families with hereditary hair loss disorders

Family	Gene	Transcript	cDNA Change	*Protein Change	PMID
PK-L	*LPAR6*	NM_001162498.1	c.66_69dup	p.(Phe24HisfsTer29)	Novel
PK-N	*DSP*	NM_001319034.1	c.1493C>T	p.(Pro498Leu)	26148547
F117	*HR*	NM_005144.4	c.3283C>T	p.(Arg1095Trp)	Novel
F118	*HR*	NM_005144.4	c.1837C>T	p.(Arg613Ter)	10674375
F134	*HR*	NM_005144.4	c.1834delT	**p.(Trp612GlyfsTer48)**	Novel
F164	*HR*	NM_005144.4	c.431delC	**p.(Pro144LeufsTer24)**	Novel
F165	*HR*	NM_005144.4	c.431delC	**p.(Pro144LeufsTer24)**	Novel
F166	*HR*	NM_005144.4	c.431delC	**p.(Pro144LeufsTer24)**	Novel
F168	*HR*	NM_005144.4	c.2455C>T	p.(Arg819Ter)	21919222
F64	*LIPH*	NM_139248.2	c.1303_1309dupGAAAACG	**p.(Val437GlyfsTer4)**	19262606
F64	*LIPH*	NM_139248.2	c.659_660delTA	**p.(Ile220ArgfsTer25)**	19262606
F65	*HR*	NM_005144.4	c.1834delT	**p.(Trp612GlyfsTer48)**	Novel

MAF = Minor allele frequency, PMID = PubMed ID, *Likely founder mutations shown in bold face, n.a = not available.

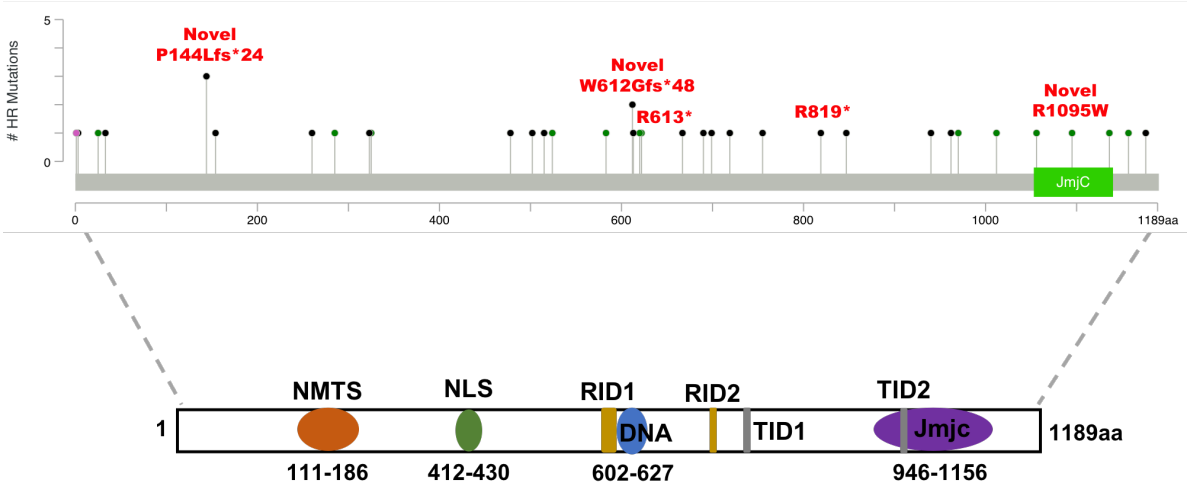


Figure 12. Genetic landscape of *HR* variants and functional domains.

Upper panel: *HR* variants reported in HGMD/ClinVar databases. Our variants are indicated in red-font. legends: green: missenses, black: loss-of-function variants, purple: Start-loss (created using MutationMapper). **Lower panel**: Different functional domains of the HR protein. legends: NMTS, nuclear matrix-targeting signal; NLS, nuclear localization signal; RID, RORα interaction domains; DNA binding domain; TIR, TR interaction domains, and Jmjc domain (adopted from Maatough et al. 2018).

2.4 *Subproject:* Genetic findings in Japanese patients with rare retinal ciliopathies

As our laboratory is heavily involved in research in the genetic basis of eye disorders, I had the opportunity to participate in a project related to a large cohort of Japanese patients with rare retinal ciliopathy. This provided me with the opportunity of comparing the findings in Pakistani families with those of a very different population, both geographically as well as in terms of population structure.

The exceptionally high missing heritability seen in Japanese individuals with genetic retinal degenerations is a distinguishing feature that is rarely observed in other populations. Following the genetic screening of 331 unrelated Japanese patients, we successfully linked the disease etiology to two predicted loss of function (pLoF) variants in the *RP1* gene. The variants included a disruptive *Alu* element insertion (c.4052_4053ins328/p.Tyr1352Alafs*9) and a nonsense mutation (p.Arg1933*). Unexpectedly, both variants were too frequent to be considered as causative individually for a rare retinal ciliopathy. The nonsense mutation (p.Arg1933*) is highly prevalent in Japan (frequency = 0.6%), and Japanese individuals carrying this mutation, both in homozygosis and in heterozygosis, remain completely asymptomatic. Interestingly, however, significant enrichment was seen for p.Arg1933* in HRDs patients compared to controls (frequency = 2.1%, i.e., a 3.5-fold enrichment; p-value = $9.2 \times 10{-}5$). Following familial co-segregation analysis and an association study, we showed that p.Arg1933* can act as a Mendelian mutation in trans with the *Alu* insertion. Nevertheless, p.Arg1933* might also associate with disease in combination with two other alleles in the *EYS* gene, in a non-Mendelian fashion. Thus, our findings suggest that rare retinal ciliopathies can be caused by relatively common variants in a quasi-Mendelian inheritance

pattern, thus intersecting monogenic and complex diseases. There is no indication yet of a similar phenomenon in the Pakistani population.

Publication

Findings of this project are published in “Nature Communications” and I co-authored this article. My contribution in the project included genotyping of Japanese IRDs cohort for selected mutations using Sanger sequencing. Section 7.3.1.

2.5 *Subproject:* Case report of a patient with Usher syndrome

We performed whole exome sequencing on the DNA of a Greek patient with Usher syndrome. The patient was a 28-year-old female with a history of RP and bilateral hearing impairment since childhood. Our genetic analysis revealed compound heterozygosity in the *USH2A* gene (OMIM 608400, NM_007123.5). The mutations included a nonsense change (c.100C>T; p.Arg34Ter) that was present in trans with a synonymous change (c.949C>A; p.Arg317=). The latter variant has been previously shown to activate a new splice site thus resulting in splicing defects [59]. Both variants were classified as pathogenic in the ClinVar database, by multiple submitters. Finally, segregation analysis in the family revealed strict genotype-phenotype correlation. The patient also underwent a successful closure of a full-thickness macular hole (FTMH) in her right eye. Of note, and as discussed below, we did not find any family with *USH2A* pathogenic variants in our Pakistani patient cohort.

Publication

A case report describing the surgical outcomes and genetic findings of this project is published in “Ophthalmology and Therapy”. My contribution to the project includes Sanger validation of the *USH2A* variants and segregation analysis in the family. Section 7.3.3.

2.6 Families and cases without molecular diagnosis

Despite the overall diagnostic success of 84%, genetic diagnosis was elusive in a subset of families with monogenic conditions (both ocular and skin cases). The reasons could be manifold, and possibly disease- or pedigree-specific. In general, comprehensive clinical data are required for correct initial diagnosis of patients suffering from phenotypically heterogeneous conditions or in cases where phenocopies/complex traits could overlap with true Mendelian phenotypes. Thus, correct initial diagnosis is always crucial for researchers to achieve a molecular diagnosis. By contrast, the minimal clinical data associated with Pakistani patients raises the possibility of misdiagnosis at the first place, thus hindering correct molecular diagnosis afterwards. For example, high myopia, a complex trait with enough familial aggregation, could easily be misclassified as inherited retinal disease if bona fide clinical data are not available. This is particularly true in the context of Pakistani population where consanguinity is very high and families have rather extended structures. Similarly, without sufficient clinical data, it is difficult to correctly discriminate between polygenic hair loss disorders (Examples: androgenetic alopecia or alopecia areata) and true monogenic hair loss disease (alopecia universalis congenita) due to the presence of clinical overlap between these two distinct disease types. As a proof of concept, molecular diagnosis in our data helped us to reclassify a few patients as suffering from congenital cataracts, and corneal dystrophy

who were initially misdiagnoses as retinal diseases. Thus, it is not unlikely that our cohort of monogenic families was also prone to such prejudice at the first place owing to the availability of minimal clinical information, possibly resulting in lack of molecular diagnosis in certain cases. In addition, we have observed that two or more clinically similar phenotypes with distinct underlying genetic mechanisms could simultaneously segregate in the same family owing to the high level of genomic homozygosity. Without a detailed clinical dissection of such families, it is rather difficult to establish a correct molecular diagnosis through genotype-phenotype co-segregation for a single variant. Alternatively, disease-causing variants in two or more genes could likely contribute to the same phenotype (so-called digenic or multigenic inheritance).

The majority of our unsolved cases comprised multiplex families; however, three out of four unsolved skin families were simplex which are generally more challenging to characterize genetically than their multiplex counterparts. Likewise, genome-wide autozygosity level in our unsolved cohort was, in principle, higher than our minimum threshold of 50 Mb for a consanguineous family; this difference was only nominal in a few unsolved cases. Thus, lack of molecular diagnosis in few cases could possibly be attributed to the low level of genomic autozygosity. Nevertheless, molecular diagnosis has never been complete despite all technological advancements, and improvements in the analytical skills as well as theoretical knowledge. This is because variation in a large fraction of the human genome is either intractable by the current sequencing methods or difficult to interpret. For example, with the existing exome sequencing technologies that provide non-uniform coverage throughout the genome, it is difficult to uncover disease-causing variants in the GC-rich regions that requires special strategies. *RPGR*, which explains more than 70% of the X-linked retinitis pigmentosa (RP) cases, constitute the most famous example of poor coverage for its exon ORF15 in exome

sequencing as well as panel-based next-generation sequencing methods due to its highly repetitive purine-rich content. Compatible with this idea, we have found, through visualization of WES data on Integrative Genomics Viewer (IGV), that exon ORF15 of the *RPGR* gene was poorly covered in our exome data. This made us to speculate that causative variants in four unresolved probands with a suspected X-linked RP are likely to reside inside the exon ORF15 of *RPGR* gene, and thus our computational pipeline was unable to detect them due to the poor sequence coverage. Though we are optimistic to find causative variants in these four probands, we have not yet screened them for ORF15 exon. Thus, we postulate that "intractable genomic regions" (such ORF15) could possibly contribute, at least to some extent, to the missing heritability seen in our cohort. Likewise, current computational and analytical approaches have their own limitations specially to detect large structural variants/complex chromosomal rearrangements such as large deletions, inversions or duplications. Therefore, we cannot completely exclude the possibility of missing such variants or other potentially disease-causing variants during filtering of our exome data. We also argue that variation in the non-coding genome, regulatory sequences, and deep-intronic variants impacting splicing likely explain the unknown genetic mechanisms in our unsolved cases. Our analysis mainly focused on the coding part of genome (exome), and the discovery and understanding the role of such variants was beyond the scope of this thesis. Finally, we anticipate that non-conventional modes of inheritance such as digenic, oligogenic, and mitochondrial inheritance, although rare, could lead to the missing heritability in our cohort.

3. Discussion

Consanguinity has been widely accepted as a major risk factor for the occurrence of rare recessive Mendelian disorders, yet it is a long-lived social practice in many Asian and African countries [60]. According to an estimate, there are approximately 7,000 rare diseases (RDs) affecting the human race, and the majority of them appear as the consequence of founder mutations in societies where community endogamy and close-kin marriages are common [8]. Pakistan has one of the world's highest rates of inherited genetic diseases likely correlating with the exceptionally high rate of consanguineous marriages (~65%) in the country [61, 62].

Pakistani families have been at the forefront in Medical Genetics research due to the highest rates of consanguinity in the country. While exploiting consanguinity as a means to identify recessive mutations, significant scientific findings in Pakistani families have been recently published [40, 61, 63-66]. Compatible with the same idea, we sought to understand the genetic etiology of consanguineous Pakistani families with different monogenetic diseases, mostly affecting the eyes and the skin. Our approach included targeted Sanger sequencing of founder mutations, SNP-based autozygosity mapping, candidate genes screening, and whole exome sequencing. Overall, we have been able to identify causative variants in 98 out of 116 monogenic families analyzed (84% success rate). Our data generally demonstrate the presence of a very high degree of autozygosity and of prevalent founder mutations in the Pakistani population, irrespective of the individual disease categorization. Thus, more than 90% of the families presented in this thesis were linked to homozygous pathogenic variants in known Mendelian disease genes. Our findings corroborate a previous study demonstrating 92% homozygous recessive mutations in a cohort of Pakistani probands [66]. Similarly, more than 40% of the pedigrees in our cohort were associated with founder mutations, thus

reflecting on the traditional practice of endogamy and the presence of extensive stratification within the Pakistani population, as described earlier [65, 67-70]. Consistent with a previous study [66], our cohort was enriched for homozygous loss-of-function alleles compared to missense mutations.

In Pakistan, an estimated 1.12 million people are blind, 1.09 million have severe vision loss while 6.79 million suffer from moderate vision loss. Surprisingly, the vision loss burden has continued to rise in the country in the last three decades [71]. In our cohort of inherited eye diseases (IEDs), we achieved a WES diagnostic rate of 83% (72 out of 86 families) which, to the best of our knowledge, represents a higher figure with respect to many previous studies [72-75]. We have also found marked genetic and allelic heterogeneity (41 genes carrying 61 distinct variants) associated with IEDs in our cohort. Thirty-two out of 41 genes (78%) were previously linked to inherited retinal diseases (IRDs), whereas nine genes (*COL18A1*, *CYP1B1*, *FOXE3*, *FYCO1*, *LTBP2*, *PAX6*, *TDRD7*, *TEK*, *ZEB1*) were previously known to cause various forms of structural abnormalities of the eye, notably: microphthalmia, primary congenital glaucoma, cataracts, and corneal dystrophies. Overall, more than half of the identified genetic variants were not previously published, and ~50% of them were absent from public databases such as gnomAD. Compatible with the heterogenous nature of IEDs, our findings are also in line with previous studies [73, 76-82]. Except for one family with X-linked disease and two families with autosomal dominant diseases, inheritance of the disorder in all remaining pedigrees was consistent with a recessive Mendelian pattern, and the variants discovered were indeed exclusively bi-allelic (homozygous or compound heterozygous). Homozygous pathogenic variants were mainly uncovered with the help of homozygosity mapping, again in line with the previous studies suggesting homozygosity mapping as an effective gene-discovery tool in consanguineous families [40, 72, 83]. In general, it is a lot easier to establish

pathogenicity for loss of function (LoF) alleles compared to missense mutations. Consistent with a previous study [64], more than half (54%) of the identified variants in our cohort were predicted to result in LoF alleles, i.e. nonsense, frameshift, or splice-site mutations, and all, except for three alleles, were found in homozygous state. Among them, a large homozygous deletion was likely to lead to, a complete knock out, of the *RP1* gene in two pedigrees. The deletion co-segregated in fact with the disease in both pedigrees. Thus, our findings support a recent case-control study showing that the occurrence of natural knock-outs for 1,317 human genes in a Pakistani cohort is attributable to homozygous LoF mutations [63]. Scientists therefore believe that "a human knockout project" will soon be initiated [63], therefore by focusing on families from endogamous populations, such as Pakistani [63].

In addition to frequently mutated IED-associated genes, we found variants in genes that were either newly associated with IEDs or were known to cause rare forms of ocular conditions. For example, we detected a homozygous missense variant in the *SLC6A6* gene (NM_003043.5:c.746C>T:p.Thr249Ile), co-segregating with a severe early-onset retinal degeneration and suggestive cardiomyopathy (as shown by abnormal ECGs) in four affected siblings of a consanguineous family. *SLC6A6* encode taurine transporter (TauT) which is a ubiquitously expressed osmolyte transporter that maintains the intracellular taurine content in many tissues including heart, brain, retina, kidney, liver and skeletal muscles [84]. A number of taurine deficiency disorders, in particular affecting kidney, retina and liver, were anticipated as early as 1990s [85]. However, a direct link between genetic alterations in *SLC6A6* gene and any human pathological condition has been established only recently in three independent studies. First, a homozygous deletion of a canonical splice site in the *SLC6A6* (c.229+1delT) was associated with dilated cardiomyopathy in a sporadic Pakistani patient born of consanguineous parents [49]. Secondly, a homozygous missense mutation in

the *SLC6A6* (p.Ala78Glu) was found to cause early-onset retinal degeneration without obvious extraocular findings in two brothers of Turkish descent [50]. More recently, our group has found a homozygous missense mutation in the *SLC6A6* (p.Gly399Val) causing progressive childhood retinal degeneration and cardiomyopathy in a consanguineous Pakistani family with two affected children. The later study has also shown beneficial effects of a long-term oral supplementation of taurine in their patients [51]. To sum up, our family with *SLC6A6*-related human phenotype is the third in Pakistan while fourth in the world.

We also report a rare form of congenital stationary night blindness (CSNB), commonly called Oguchi disease, in a consanguineous Pakistani family with four affected individuals. In this family, the disease was linked to a previously known pathogenic variant in the *GRK1* (p.Asp537ValfsTer7) [86, 87]. Thus far only twenty variants in the *GRK1* have been associated with Oguchi disease [88].

Overall, our study was successful in finding the molecular basis of IEDs and skin diseases in the majority of families analyzed. Interestingly, we did not uncover completely unknown disease genes but showed the presence of putative founder mutations in a number of known genes associated with disease. It is possible that based on our data and previous reports [64, 65], more extensive studies involving an even larger number of Pakistani families could be useful to uncover additional genes causing visual impairment, or other ultra-rare recessively-inherited monogenic disorders. The existence of both novel and previously known founder mutations is explained by the population sub-structuring in the country. To this end, our data may have practical significance as ten distinct founder mutations in nine separate genes explained the genetic etiologies in 44% of our cohort. Genes carrying founder mutations included *ABCA4*, *CRB1*, *LRIT3*, *RPE65*, *NMNAT1*, *PDE6A*, *CERKL*, *RP1*, and *MYO7A*. A founder

allele in the *ABCA4* (p.Gly72Arg) was the most commonly reported disease-causing variant in our cohort of IEDs collectively accounting for 11 independent families. Thus, our findings corroborate several previous studies demonstrating high founder effects in Pakistan, occasionally leading to the discovery of novel gene-disease associations [65, 69]. Interestingly, we did not uncover any instance of retinal disorder associated with the *USH2A* gene: this is surprising as variants in this gene are highly prevalent in other cohorts of IED individuals [89]. Indirectly, this observation underscores the role of founder mutations in Pakistan: no founder mutation seems to exist for *USH2A*.

The scope of this thesis also entails families with monogenic skin conditions, notably skin blistering and hair loss disorders. Through WES analysis in a consanguineous family with recessive dystrophic epidermolysis bullosa (EB), we identified a homozygous missense mutation in *COL7A1* (p.Gly2680Ser). The mutation was previously known to cause EB in patients from diverse ethnic backgrounds such China, UK, Poland, and Iran [54, 55, 90, 91]. This made us to speculate that this mutation could likely constitute a recurrent mutational hotspot in the *COL7A1* gene. EB is a rare skin condition that affect ~1 in 120,000 people worldwide, or one out of every 20,000 live births in the United States. Mainly characterized by extreme skin fragility and trauma-induced blistering, there are more than 30 clinical subtypes of EB and associated mutations are known in as many as 18 genes [44]. Recent studies suggest that various treatment options such as gene replacement or correction, protein replacement, and cell-based therapies for EB are currently in their early clinical trials [92].

Genetic analysis in probands from 11 families with hereditary hair loss disorders revealed a total of nine distinct genetic variants within four genes (*HR*, *LIPH*, *LPAR6*, *DSP*). Among them, *HR* was the most frequently mutated gene (8 out of 11 families, or 73%). The first genetic variant that linked the human hairless (*HR*) gene to alopecia universalis two decades ago was also reported in a Pakistani family [93]. Since then, pathological changes in *HR* have been mainly associated with three overlapping human disorders, namely; alopecia universalis congenita (ALUNC), atrichia with papular lesions (APL), and Marie-Unna Hereditary Hypotrichosis (MUHH). While the first two categories arise from recessive mutations in the coding part of *HR*, the latter is associated with dominant changes in the upstream open reading frame (U2HR) of the gene [94]. Thus far, at least 79 variants in *HR* are flagged as pathogenic in the ClinVar database [95]. *HR* encodes a nuclear transcription factor that possess functional domains essential for DNA binding, histone demethylation, nuclear translocation and protein–protein interactions [94]. Previous literature suggests that the majority (75%) of mutations associated with APL or ALUNC are located inside these functional domains [94]. Consistent with these findings, all five *HR* mutations reported in this thesis localized within the functional domains of the HR protein: p.Pro144LeufsTer24 (nuclear matrix targeting signal, NMTS domain), p.Trp612GlyfsTer48; p.Arg613Ter (DNA binding domain), p.Arg819Ter (TR interaction domain, TID1), and p.Arg1095Trp (JmjC domain). Except for the missense (p.Arg1095Trp), all remaining four mutations constitute predicted loss-of-function (LoF) alleles, and therefore we can easily postulate causality for them in our patients. We also report compound heterozygosity in the *LIPH* gene (p.Val437GlyfsTer4; p.Ile220ArgfsTer25), associated with wooly hair phenotype in one Pakistani family. These mutations were previously known to be founder alleles in Pakistani population [57]. Additional variants in two remaining families with hereditary hair loss disorders included a

novel homozygous frameshift deletion in *LPAR6* (p.Phe24HisfsTer29), and a previously-known pathogenic mutation in *DSP* (p.Pro498Leu).

Though largely similar to previous studies in the world populations, our study has a few distinctive features. First, the diagnostic yield (84%) achieved in our study was unprecedented. This is likely due to the high degree of consanguinity as observed from the pedigree-based estimations. Our analysis also confirmed this fact by showing a very high level of genome-wide autozygosity and significant enrichment for recessive genetic burden. This provides an evidence that consanguinity is the major contributing factor for recessive genetic diseases in Pakistan, and that autozygosity mapping is a powerful gene mapping tool when employed in consanguineous pedigrees. Second, the appearance of frequent founder variants in our study indicate that population stratification and community endogamy contribute significantly to the overall recessive genetic burden in Pakistan. This is in fact true for all kinds of Mendelian diseases and across all geographic and ethnic strata of Pakistan. Nevertheless, contribution of dominant and X-linked variants, though nominally present in our cohort, could not be completely ruled out while studying consanguineous pedigrees. Third, our data showed highest proportion of bi-allelic loss-of-function (LoF) variants, thus further strengthening the deleterious effects of null alleles. Since LoF alleles are predicted to inactivate/knock out protein coding genes, such variants could provide an insightful information about gene function, thus highlighting importance of Pakistani families for future medical research. Fourth, our analysis discovered ultra-rare retinal diseases while slightly depleted for some common retinal phenotypes. For example, we have found an extremely rare taurine-transporter deficiency disorder in a consanguineous Pakistani family associated with *SLC6A6* gene. By contrast, our data completely lacked for *USH2A*-associated human phenotypes although pathogenic variants in *USH2A* gene are considered to be the most

common cause of non-syndromic retinitis pigmentosa or Usher syndrome (retinitis pigmentosa and deafness) [89]. Our analysis therefore highlights the unique genetic architecture of Pakistani families, and warrants further studies to uncover the full spectrum of retinal diseases (including rare human phenotypes) by undertaking large number of consanguineous families. In summary, our data generally confirm our starting hypothesis: a high degree of autozygosity in Pakistan leads to the frequent clinical expression of prevalent founder mutations in Pakistan, across the panel of different disease categories, thus making Pakistani families distinctive for gene hunt studies using autozygosity mapping. As outlined above, this is likely the consequence of the extensive population stratification as well as the high prevalence of consanguinity in the society. Our data introduce novel disease-causing alleles in the literature and thus further expands the current genetic landscape of inherited eye and skin diseases.

We also anticipate that these insights on eye and skin disorders, and possibly the strategies derived to create novel targeted approaches for rapid diagnosis and prevention, might apply to other recessive conditions that affect Pakistan and its regions, such as deafness, intellectual disabilities, and perhaps metabolic disorders and developmental defects.

4. Perspectives

We hope that our findings, and specifically the identification of founder alleles in nine genes, will stimulate researchers and clinicians in Pakistan to screen their patients for known mutations in a time and cost-efficient way; indeed, an argument could be made to implement a panel of frequent founder mutations, both at a diagnostic level and, perhaps in the future, at a prevention level. This could be tailored to families or communities with a greater risk for one or another specific disease, who could then benefit from carrier testing and genetic counselling.

Furthermore, our data could be used as a useful genetic resource for future studies undertaking gene therapy trials for retinal diseases associated with these particular genes, excluding *RPE65* gene for which gene therapy has already been in the market. In addition, our data could provide a hope for individuals with certain genetic conditions to get benefits from existing gene or cell-based therapies. For example, patients with *RPE65*-associated visual impairment (five families in our study) could benefit from *Luxturna*, a gene therapy designed to treat patients with *RPE65*-related vision loss.

Finally, we hope that scientific findings like ours might encourage the Pakistani nation to enter into an era of personalized medicine. With the growing knowledge about genetic disorders and increasing affordability of genetic testing, we hope that diagnostic services will soon be available to the public. As a subsequent step following diagnostics, we suggest that Pakistan might develop a national database of genetic diseases following the example of public databases and registries in many developed countries. These infrastructures would be the prerequisite to assist professionals and to guide the public health care systems towards

disease-specific or community-specific prevention and treatment strategies in accordance with epidemiologic data and treatment availabilities.

5. Author contribution

This thesis document is solely written by me (Atta Ur Rehman) and reviewed by Prof. Carlo Rivolta and Prof. Andrea Superti-Furga and describes all the scientific findings that were produced during my three years PhD research work (September 2017 to August 2020). Except where necessary (Figure 1, 2, and 12 which are modified from other sources and thus cited), all the remaining data (including Tables and Figures) in this document are genuine, generated by myself and constitute part of my own thesis work.

Identification and recruitment of monogenic Pakistani families and blood/saliva sampling was done partly by me and partly by our collaborators in Pakistan.

All the laboratory ("wet lab") procedures associated with this thesis were done by myself.

Bioinformatic analysis of raw data (filtering of WES data and creation of autozygosity maps using in-house pipeline/AutoMap) was kindly done by Mathieu Quinodoz or Virginie Peter. All downstream analysis such as manual curation of the WES data for detection of pathogenic/likely pathogenic variants, Sanger validation of WES results, and segregation analysis was performed by myself.

My contribution to the first author articles (published in ***Genes***, and ***Clinical Dysmorphology***) includes project design/execution, sample acquisition, wet lab work, data analysis, and manuscript write-up as well correspondence (***Clinical Dysmorphology***). Except for Figure 2 in the ***"Genes"*** article which was drawn with the help of Mathieu Quinodoz, remaining Figures/Tables in both first-author articles were drawn by myself.

Of the co-author articles, my contribution to the project published in ***"Nature Communications"*** included genotyping of Japanese IRDs cohort for selected mutations using

Sanger sequencing. For the article published in **"*Ophthalmology and Therapy*"**, I performed segregation analysis in the family. Lastly, one co-author article is currently under review in ***"Human Mutation"*** where my contribution is the addition of one family with *GRK1*-related Oguchi disease to the manuscript.

6. References

1. Green, E.D., M.S. Guyer, and I. National Human Genome Research, *Charting a course for genomic medicine from base pairs to bedside.* Nature, 2011. **470**(7333): p. 204-13.
2. Ali, G., et al., *Identification of a recurrent nonsense mutation in HR gene responsible for atrichia with papular lesions in two Kashmiri families.* J Gene Med, 2020. **22**(5): p. e3167.
3. Wang, S., et al., *Atrichia with papular lesions in a chinese family caused by novel compound heterozygous mutations and literature review.* Dermatology, 2013. **226**(1): p. 68-74.
4. Ahmed, M.S., et al., *Identification of novel mutation in the HR gene responsible for atrichia with papular lesions in a Pakistani family.* J Dermatol, 2013. **40**(11): p. 927-8.
5. Flicek, P., et al., *Ensembl 2014.* Nucleic Acids Res, 2014. **42**(Database issue): p. D749-55.
6. Ezkurdia, I., et al., *Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes.* Hum Mol Genet, 2014. **23**(22): p. 5866-78.
7. Chong, J.X., et al., *The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities.* Am J Hum Genet, 2015. **97**(2): p. 199-215.
8. Bittles, A.H., *Precision medicine: Rare diseases and community genetics.* Digit Med, 2019. **5**(4): p. 54-61.
9. Christiano, A., C.P. Howson, and B. Modell, *March of Dimes Global Report on Birth Defects: The hidden toll of dying and disabled children.* 2006.
10. Angelis, A., D. Tordrup, and P. Kanavos, *Socio-economic burden of rare diseases: A systematic review of cost of illness evidence.* Health Policy, 2015. **119**(7): p. 964-79.

11. WHO, *World Health Organization. Genes and human diseases: Monogenic diseases*. p. https://www.who.int/genomics/public/geneticdiseases/en/index2.html (Accessed April 15, 2020).

12. Karczewski, K.J., et al., *The mutational constraint spectrum quantified from variation in 141,456 humans.* bioRxiv, 2020: p. 531210.

13. Boycott, K.M., et al., *International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases.* American Journal of Human Genetics, 2017. **100**(5): p. 695-705.

14. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.

15. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery.* Nature Reviews Genetics, 2011. **12**(11): p. 745-755.

16. Eilbeck, K., A. Quinlan, and M. Yandell, *Settling the score: variant prioritization and Mendelian disease.* Nat Rev Genet, 2017. **18**(10): p. 599-612.

17. Wright, C.F., D.R. FitzPatrick, and H.V. Firth, *Paediatric genomics: diagnosing rare disease in children.* Nat Rev Genet, 2018. **19**(5): p. 325.

18. Eilbeck, K., et al., *Quantitative measures for the management and comparison of annotated genomes.* Bmc Bioinformatics, 2009. **10**.

19. Pertea, M. and S.L. Salzberg, *Between a chicken and a grape: estimating the number of human genes.* Genome Biology, 2010. **11**.

20. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

21. Bittles, A.H., et al., *Reproductive behavior and health in consanguineous marriages.* Science, 1991. **252**(5007): p. 789-94.

22. Hamamy, H., et al., *Consanguineous marriages, pearls and perils: Geneva International Consanguinity Workshop Report.* Genet Med, 2011. **13**(9): p. 841-7.

23. Bronberg, R., et al., *Biosocial correlates and spatial distribution of consanguinity in South America.* Am J Hum Biol, 2016. **28**(3): p. 405-11.

24. NIPS, *Pakistan Demographic and Health Survey*, I. National Institute of Population Studies, Pakistan, Editor. 2013: Maryland, USA.

25. Chinthapalli, K., *First cousin marriage can double risk of birth defects, finds study.* Bmj-British Medical Journal, 2013. **347**.

26. Alkuraya, F.S., *Autozygome decoded.* Genet Med, 2010. **12**(12): p. 765-71.

27. Ahmad, B., A.U. Rehman, and S. Malik, *Consanguinity and Inbreeding Coefficient in Tribal Pashtuns Inhabiting the Turbulent and War-Affected Territory of Bajaur Agency, North-West Pakistan.* J Biosoc Sci, 2016. **48**(1): p. 113-28.

28. Fuhrmann, S., *Eye morphogenesis and patterning of the optic vesicle.* Curr Top Dev Biol, 2010. **93**: p. 61-84.

29. Sinn, R. and J. Wittbrodt, *An eye on eye development.* Mech Dev, 2013. **130**(6-8): p. 347-58.

30. Plaisancie, J., et al., *Genetics of anophthalmia and microphthalmia. Part 1: Non-syndromic anophthalmia/microphthalmia.* Hum Genet, 2019. **138**(8-9): p. 799-830.

31. Harding, P. and M. Moosajee, *The Molecular Basis of Human Anophthalmia and Microphthalmia.* J Dev Biol, 2019. **7**(3).

32. Markitantova, Y. and V. Simirskii, *Inherited Eye Diseases with Retinal Manifestations through the Eyes of Homeobox Genes.* Int J Mol Sci, 2020. **21**(5).

33. Martin, A.R., et al., *PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels.* Nat Genet, 2019. **51**(11): p. 1560-1565.

34. Hamel, C.P., *Cone rod dystrophies.* Orphanet J Rare Dis, 2007. **2**: p. 7.

35. Berger, W., B. Kloeckener-Gruissem, and J. Neidhardt, *The molecular basis of human retinal and vitreoretinal diseases.* Prog Retin Eye Res, 2010. **29**(5): p. 335-75.

36. RetNet. *Retinal Information Network*. [Accessed on April 17, 2020]; Available from: https://sph.uth.edu/retnet/.

37. Hamel, C.P., *Gene discovery and prevalence in inherited retinal dystrophies.* C R Biol, 2014. **337**(3): p. 160-6.

38. Traboulsi, E.I., *Hope and major strides for genetic diseases of the eye.* J Genet, 2009. **88**(4): p. 395-7.

39. Adhi, M.I. and J. Ahmed, *Frequency and Clinical Presentation of Retinaly Dystrophies - A Hospital Based Study.* Pak J Ophthalmology, 2002. **18**(4): p. 106-10.

40. Khan, M.I., et al., *The molecular basis of retinal dystrophies in pakistan.* Genes (Basel), 2014. **5**(1): p. 176-95.

41. Werner, S., T. Krieg, and H. Smola, *Keratinocyte-fibroblast interactions in wound healing.* J Invest Dermatol, 2007. **127**(5): p. 998-1008.

42. Has, C. and L. Bruckner-Tuderman, *The genetics of skin fragility.* Annu Rev Genomics Hum Genet, 2014. **15**: p. 245-68.

43. Basit, S., S. Khan, and W. Ahmad, *Genetics of human isolated hereditary hair loss disorders.* Clin Genet, 2015. **88**(3): p. 203-12.

44. Fine, J.D., *Epidemiology of Inherited Epidermolysis Bullosa Based on Incidence and Prevalence Estimates From the National Epidermolysis Bullosa Registry.* Jama Dermatology, 2016. **152**(11): p. 1231-1238.

45. McGrath, J.A., *Recently Identified Forms of Epidermolysis Bullosa.* Annals of Dermatology, 2015. **27**(6): p. 658-666.

46. Alkuraya, F.S., *Discovery of rare homozygous mutations from studies of consanguineous pedigrees.* Curr Protoc Hum Genet, 2012. **Chapter 6**: p. Unit6 12.

47. Pemberton, T.J., et al., *Genomic patterns of homozygosity in worldwide human populations.* Am J Hum Genet, 2012. **91**(2): p. 275-92.

48. Wakeling, M.N., et al., *Homozygosity mapping provides supporting evidence of pathogenicity in recessive Mendelian disease.* Genet Med, 2019. **21**(4): p. 982-986.

49. Shakeel, M., M. Irfan, and I.A. Khan, *Rare genetic mutations in Pakistani patients with dilated cardiomyopathy.* Gene, 2018. **673**: p. 134-139.

50. Preising, M.N., et al., *Biallelic mutation of human SLC6A6 encoding the taurine transporter TAUT is linked to early retinal degeneration.* FASEB J, 2019. **33**(10): p. 11507-11527.

51. Ansar, M., et al., *Taurine treatment of retinal degeneration and cardiomyopathy in a consanguineous family with SLC6A6 taurine transporter deficiency.* Hum Mol Genet, 2020. **29**(4): p. 618-623.

52. Yan, Y.S., et al., *Five Novel COL7A1 Gene Mutations in Three Chinese Patients with Recessive Dystrophic Epidermolysis Bullosa.* Annals of Clinical and Laboratory Science, 2018. **48**(1): p. 100-105.

53. Vahidnezhad, H., et al., *Dystrophic Epidermolysis Bullosa: COL7A1 Mutation Landscape in a Multi-Ethnic Cohort of 152 Extended Families with High Degree of Customary Consanguineous Marriages.* J Invest Dermatol, 2017. **137**(3): p. 660-669.

54. Wertheim-Tysarowska, K., et al., *Novel and recurrent COL7A1 mutation in a Polish population.* Eur J Dermatol, 2012. **22**(1): p. 23-8.

55. Almaani, N., et al., *Identical Glycine Substitution Mutations in Type VII Collagen May Underlie Both Dominant and Recessive Forms of Dystrophic Epidermolysis Bullosa.* Acta Dermato-Venereologica, 2011. **91**(3): p. 262-266.

56. Mehmood, S., et al., *Disease causing homozygous variants in the human hairless gene.* Int J Dermatol, 2016. **55**(9): p. 977-81.

57. Shimomura, Y., et al., *Founder mutations in the lipase h gene in families with autosomal recessive woolly hair/hypotrichosis.* J Invest Dermatol, 2009. **129**(8): p. 1927-34.

58. Jan, A., et al., *A novel homozygous variant in the dsp gene underlies the first case of non-syndromic form of alopecia.* Arch Dermatol Res, 2015. **307**(9): p. 793-801.

59. Vache, C., et al., *Nasal epithelial cells are a reliable source to study splicing variants in Usher syndrome.* Hum Mutat, 2010. **31**(6): p. 734-41.

60. Bittles, A.H. and J.V. Neel, *The costs of human inbreeding and their implications for variations at the DNA level.* Nat Genet, 1994. **8**(2): p. 117-21.

61. Riaz, M., et al., *Implementation of public health genomics in Pakistan.* Eur J Hum Genet, 2019. **27**(10): p. 1485-1492.

62. Manzoor, R., et al., *Consanguineous marriages: Effects on pregnancy outcomes in Pakistan.* Journal of Development Policy, Research & Practice, 2018: p. 78.

63. Saleheen, D., et al., *Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity.* Nature, 2017. **544**(7649): p. 235-239.

64. Harripaul, R., et al., *Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families.* Mol Psychiatry, 2018. **23**(4): p. 973-984.

65. Li, L., et al., *Mutation in the intracellular chloride channel CLCC1 associated with autosomal recessive retinitis pigmentosa.* PLoS Genet, 2018. **14**(8): p. e1007504.

66. Martin, H.C., et al., *Quantifying the contribution of recessive coding variation to developmental disorders.* Science, 2018. **362**(6419): p. 1161-1164.

67. Rehman, A.U., et al., *Exploring the Genetic Landscape of Retinal Diseases in North-Western Pakistan Reveals a High Degree of Autozygosity and a Prevalent Founder Mutation in ABCA4.* Genes (Basel), 2019. **11**(1).

68. Khan, R., et al., *A founder RDH5 splice site mutation leads to retinitis punctata albescens in two inbred Pakistani kindreds.* Ophthalmic Genet, 2020. **41**(1): p. 7-12.

69. Abdelhadi, O., et al., *Founder mutation in KCNJ10 in Pakistani patients with EAST syndrome.* Mol Genet Genomic Med, 2016. **4**(5): p. 521-6.

70. Ahmed, J., et al., *Genetic study of Khyber-Pukhtunkhwa resident Pakistani families presenting primary microcephaly with intellectual disability.* J Pak Med Assoc, 2019. **69**(12): p. 1812-1816.

71. Hassan, B., et al., *A comprehensive study capturing vision loss burden in Pakistan (1990-2025): Findings from the Global Burden of Disease (GBD) 2017 study.* PLoS One, 2019. **14**(5): p. e0216492.

72. Li, L., et al., *Homozygosity Mapping and Genetic Analysis of Autosomal Recessive Retinal Dystrophies in 144 Consanguineous Pakistani Families.* Invest Ophthalmol Vis Sci, 2017. **58**(4): p. 2218-2238.

73. Maranhao, B., et al., *Investigating the Molecular Basis of Retinal Degeneration in a Familial Cohort of Pakistani Decent by Exome Sequencing.* PLoS One, 2015. **10**(9): p. e0136561.

74. Gonzàlez-Duarte, R., et al., *Scaling New Heights in the Genetic Diagnosis of Inherited Retinal Dystrophies*, in *Retinal Degenerative Diseases*. 2019, Springer. p. 215-219.

75. Stone, E.M., et al., *Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease.* Ophthalmology, 2017. **124**(9): p. 1314-1331.

76. Khan, M., et al., *Identification and Analysis of Genes Associated with Inherited Retinal Diseases*, in *Retinal Degeneration*. 2019, Springer. p. 3-27.

77. Pierce, E., *Genome Editing for Inherited Retinal Degenerations.* Ophthalmology, 2018. **125**(9): p. 1431-1432.

78. Duncan, J.L., et al., *Inherited Retinal Degenerations: Current Landscape and Knowledge Gaps.* Transl Vis Sci Technol, 2018. **7**(4): p. 6.

79. Ghofrani, M., et al., *Homozygosity Mapping and Targeted Sanger Sequencing Identifies Three Novel CRB1 (Crumbs homologue 1) Mutations in Iranian Retinal Degeneration Families.* Iran Biomed J, 2017. **21**(5): p. 294-302.

80. Sahel, J.A., K. Marazova, and I. Audo, *Clinical characteristics and current therapies for inherited retinal degenerations.* Cold Spring Harb Perspect Med, 2014. **5**(2): p. a017111.

81. Beryozkin, A., et al., *Identification of Mutations Causing Inherited Retinal Degenerations in the Israeli and Palestinian Populations Using Homozygosity Mapping.* Investigative Ophthalmology & Visual Science, 2014. **55**(2): p. 1149-1160.

82. Thompson, D.A., et al., *Retinal degeneration associated with RDH12 mutations results from decreased 11-cis retinal synthesis due to disruption of the visual cycle.* Hum Mol Genet, 2005. **14**(24): p. 3865-75.

83. Maria, M., et al., *Homozygosity mapping and targeted sanger sequencing reveal genetic defects underlying inherited retinal disease in families from pakistan.* PLoS One, 2015. **10**(3): p. e0119806.

84. Ito, T., et al., *Cardiac and skeletal muscle abnormality in taurine transporter-knockout mice.* J Biomed Sci, 2010. **17 Suppl 1**: p. S20.

85. Sturman, J.A., *Taurine deficiency.* Prog Clin Biol Res, 1990. **351**: p. 385-95.

86. Skorczyk-Werner, A., et al., *The first case of Oguchi disease, type 2 in a Polish patient with confirmed GRK1 gene mutation.* Klin Oczna, 2015. **117**(1): p. 27-30.

87. Yamamoto, S., et al., *Defects in the rhodopsin kinase gene in the Oguchi form of stationary night blindness.* Nat Genet, 1997. **15**(2): p. 175-8.

88. Poulter, J.A., et al., *New pathogenic variants and insights into pathogenic mechanisms in GRK1-related Oguchi disease.* BioRxiv, 2020.

89. Lenassi, E., et al., *A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants.* Eur J Hum Genet, 2015. **23**(10): p. 1318-27.

90. Vahidnezhad, H., et al., *Multigene Next-Generation Sequencing Panel Identifies Pathogenic Variants in Patients with Unknown Subtype of Epidermolysis Bullosa: Subclassification with Prognostic Implications.* Journal of Investigative Dermatology, 2017. **137**(12): p. 2649-2652.

91. Yan, Y., et al., *Five Novel COL7A1 Gene Mutations in Three Chinese Patients with Recessive Dystrophic Epidermolysis Bullosa.* Ann Clin Lab Sci, 2018. **48**(1): p. 100-105.

92. Has, C., A. South, and J. Uitto, *Molecular Therapeutics in Development for Epidermolysis Bullosa: Update 2020.* Mol Diagn Ther, 2020.

93. Ahmad, W., et al., *Alopecia universalis associated with a mutation in the human hairless gene.* Science, 1998. **279**(5351): p. 720-4.

94. Maatough, A., et al., *Human Hairless Protein Roles in Skin/Hair and Emerging Connections to Brain and Other Cancers.* J Cell Biochem, 2018. **119**(1): p. 69-80.

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

7. Appendices

7.1 List of genetic variants identified in our cohort of 86 IEDs families

Family ID	Gene	Transcript ID	cDNA change	Protein change	gnomAD MAF	Zygosity	Method	PMID
PK-B	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	WES	10958763
PK-D	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	WES	10958763
PK-E	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Comphet	WES	10958763
PK-E	*ABCA4*	NM_000350.2	c.3081T>G	p.(Tyr1027Ter)	n.a	Comphet	WES	25312043
PK-F	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	WES	10958763
PK-H	*MKKS*	NM_170784.2	c.280T>C	p.(Phe94Leu)	n.a	Hom	SNP	Novel
PK-J	*RDH12*	NM_152443.2	c.609C>A	p.(Ser203Arg)	0.00002786	Hom	WES	22065924
PK-L	*NMNAT1*	NM_022787.3	c.25G>A	p.(Val9Met)	n.a	Hom	SNP	22842227
PK-M	*NMNAT1*	NM_022787.3	c.25G>A	p.(Val9Met)	n.a	Hom	SNP	22842227
PK-O	*CYP1B1*	NM_000104.3	c.1168C>A	p.(Arg390Ser)	0.00001602	Hom	SNP	14635112
PK-P	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	TSS	10958763
PK001	*TULP1*	NM_001289395.1	c.1307A>G	p.(Lys436Arg)	0.00002472	Hom	WES	9660588
PK002	*PDE6B*	NM_001145291.1	c.427del	p.(Ala143LeufsTer7)	n.a	Hom	WES	Novel
PK003	*FOXE3*	NM_012186.2	c.440T>C	p.(Leu147Pro)	n.a	Hom	SNP	Novel
PK004	*CNGA3*	NM_001298.2	c.847C>T	p.(Arg283Trp)	0.0001402	Hom	WES	9662398
PK006	*BBS2*	NM_031885.3	c.1438C>T	p.(Arg480Ter)	0.00001647	Hom	SNP	24608809
PK007	*CNGA1*	NM_001142564.1	c.1298G>A	p.(Gly433Asp)	n.a	Hom	WES	25775262
PK008	*TDRD7*	NM_014290.2	c.3036C>G	p.(Phe1012Leu)	n.a	Hom	WES	Novel
PK009	*CNGB3*	NM_019098.4	c.1574_1575del	p.(Phe525Ter)	0.000003993	Hom	WES	Novel
PK010	*RPE65*	NM_000329.2	c.550G>T	p.(Glu184Ter)	n.a	Hom	WES	Novel
PK026	*RP1*	NM_006269.1	c.?	p.?	n.a	Hom	WES	Novel
PK028	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	TSS	10958763

Continued.

PK029	*SPATA7*	NM_018418.4	c.1028G>A	p.(Arg343Lys)	n.a	Hom	WES	Novel
PK030	*SLC6A6*	NM_001134367.3	c.1049C>T	p.(Thr350Ile)	n.a	Hom	WES	Novel
PK031	*AIPL1*	NM_001285401.2	c.762G>A	p.(Trp254Ter)	0.000329	Hom	WES	30718709
PK032	*ABCA4*	NM_000350.2	c.1222C>T	p.(Arg408Ter)	0.0000159	Hom	WES	28947085
PK033	*LCA5*	NM_181714.3	c.664_669del	p.(Ala222_Lys223del)	n.a	Hom	WES	Novel
PK034	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	TSS	10958763
PK035	*CRB1*	NM_201253.2	c.2290C>T	p.(Arg764Cys)	0.0000757	Hom	WES	30718709
PK036	*CRB1*	NM_201253.2	c.1459T>C	p.(Ser487Pro)	n.a	Hom	WES	24265693
PK037	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	TSS	10958763
PK038	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	TSS	10958763
PK039	*PAX6*	NM_001604.5	c.510G>A	p.(Trp170Ter)	n.a	Hom	WES	28321846
PK040	*CRB1*	NM_201253.2	c.1459T>C	p.(Ser487Pro)	n.a	Hom	WES	24265693
PK041	*CRB1*	NM_201253.2	c.1459T>C	p.(Ser487Pro)	n.a	Hom	WES	24265693
PK042	*LCA5*	NM_001122769.2	c.1261C>T	p.(Gln421Ter)	0.00000413	Hom	WES	Novel
PK043	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	TSS	10958763
PK044	*MYO7A*	NM_000260.3	c.4838del	p.(Asp1613ValfsTer32)	0.000008284	Hom	WES	22135276
PK046	*PRPF31*	NM_015629.3	c.322+5G>A	p.?	n.a	Het	WES	Novel
PK049	*MYO7A*	NM_001127180.1	c.2525C>G	p.(Thr842Ser)	0.00002898	Hom	WES	Novel
PK051	*AIPL1*	NM_001285400.2	c.399G>T	p.(Gln133His)	0.00001627	Hom	WES	Novel
PK055	*CERKL*	NM_001160277.1	c.715C>T	p.(Arg239Ter)	0.0003319	Hom	WES	30718709
PK056	*IFT172*	NM_015662.2	c.3268G>A	p.(Val1090Met)	0.0001768	Comphet	WES	Novel
PK056	*IFT172*	NM_015662.2	c.4960A>G	p.(Met1654Val)	0.0001774	Comphet	WES	Novel
PK057	*CERKL*	NM_001160277.1	c.715C>T	p.(Arg239Ter)	0.0003319	Hom	WES	30718709
PK058	*TTLL5*	NM_015072.4	c.3744dup	p.(Ser1249ValfsTer15)	0.00005178	Hom	WES	Novel
PK059	*COL18A1*	NM_030582.3	c.4054_4055del	p.(Leu1352ValfsTer72)	0.0002961	Hom	WES	29977801
PK060	*PDE6H*	NM_006205.2	c.35C>G	p.(Ser12Ter)	0.000092	Hom	WES	22901948
PK061	*ABCA4*	NM_000350.2	c.214G>A	p.(Gly72Arg)	0.00002784	Hom	TSS	10958763

Continued.

PK100	*RP1*	NM_006269.1	c.4555del	p.(Arg1519GlufsTer2)	0.000004008	Hom	WES	28418496
PK102	*SPATA7*	NM_001040428.3	c.157C>T	p.(Arg53Ter)	0.00006744	Hom	WES	25133751
PK104	*MYO7A*	NM_001127180.1	c.3724C>T	p.(Gln1242Ter)	n.a	Hom	WES	24651602
PK105	*PDE6B*	NM_001350155.1	c.571G>A	p.(Gly191Ser)	0.000003988	Hom	WES	Novel
PK107	*PDE6A*	NM_000440.2	c.1444T>C	p.(Cys482Arg)	n.a	Hom	WES	Novel
PK108	*PDE6A*	NM_000440.2	c.1444T>C	p.(Cys482Arg)	n.a	Hom	WES	Novel
PK109	*PDE6A*	NM_000440.2	c.1444T>C	p.(Cys482Arg)	n.a	Hom	WES	Novel
PK110	*MYO7A*	NM_000260.3	c.4838del	p.(Asp1613ValfsTer32)	0.000008284	Hom	WES	22135276
PK111	*MYO7A*	NM_001127180.1	c.3724C>T	p.(Gln1242Ter)	n.a	Hom	WES	24651602
PK112	*ZEB1*	NM_030751.5	c.685-2A>G	p.?	n.a	Hom	WES	25441224
PK301	*ARL6*	NM_001323514.1	c.387_394del	p.(Asn130GlyfsTer3)	n.a	Hom	WES	Novel
PK303	*ABCA4*	NM_000350.2	c.6658C>T	p.(Gln2220Ter)	0.00005171	Hom	WES	28118664
PK304	*RPGRIP1*	NM_020366.3	c.2480G>T	p.(Arg827Leu)	n.a	Hom	WES	12920076
PK305	*RP1*	NM_006269.1	c.?	p.?	n.a	Hom	WES	Novel
PK310	*RP1*	NM_006269.1	c.615+1G>A	p.?	n.a	Hom	WES	Novel
PK312	*LRP5*	NM_002335.3	c.629A>G	p.(Tyr210Cys)	0.000003989	Hom	WES	Novel
PK315	*GUCY2D*	NM_000180.3	c.71del	p.(Pro24ArgfsTer61)	n.a	Hom	WES	Novel
PK316	*CYP1B1*	NM_000104.3	c.1018C>T	p.(Gln340Ter)	n.a	Hom	WES	Novel
PK317	*NMNAT1*	NM_022787.3	c.25G>A	p.(Val9Met)	n.a	Hom	WES	22842227
PK318	*CRB1*	NM_201253.2	c.1459T>C	p.(Ser487Pro)	n.a	Hom	TSS	24265693
PK320	*PROM1*	NM_001145847.1	c.1379_1380insT	p.(Thr461AspfsTer48)	n.a	Hom	WES	Novel
PK323	*PAX6*	NM_001258465.1	c.607C>T	p.(Arg203Ter)	n.a	Het	WES	7550230
PK327	*RP1*	NM_006269.1	c.3396G>A	p.(Trp1132Ter)	0.000007963	Hom	WES	22317909
PK331	*LTBP2*	NM_000428.2	c.5270G>A	p.(Cys1757Tyr)	n.a	Hom	WES	32165823
PK333	*TEK*	NM_001290078.1	c.2783G>C	p.(Arg928Pro)	n.a	Hom	WES	Novel
PK334	*CRB1*	NM_201253.2	c.1459T>C	p.(Ser487Pro)	n.a	Hom	WES	24265693
PK401	*RPE65*	NM_000329.2	c.361del	p.(Ser121LeufsTer6)	0.00000399	Hom	WES	Novel

Continued.

PK402	*RPE65*	NM_000329.2	c.361del	p.(Ser121LeufsTer6)	0.00000399	Hom	WES	23878505
PK404	*GRK1*	NM_002929.2	c.1610_1613del	(Asp537ValfsTer7)	n.a	Hom	WES	26349155
PK405	*RPE65*	NM_000329.2	c.361del	p.(Ser121LeufsTer6)	0.00000399	Hom	WES	23878505
PK406	*RPE65*	NM_000329.2	c.361del	p.(Ser121LeufsTer6)	0.00000399	Hom	WES	Novel
PK407	*FYCO1*	NM_024513.3	c.4127T>C	p.(Leu1376Pro)	0.000003994	Hom	WES	21636066
PK408	*LRP5*	NM_001291902.1	c.2745+1G>A	p.?	n.a	Comphet	WES	Novel
PK408	*LRP5*	NM_001291902.1	c.430G>A	p.(Val144Ile)	0.00003183	Comphet	WES	Novel
PK409	*CYP1B1*	NM_000104.3	c.1169G>A	p.(Arg390His)	0.0001032	Hom	WES	10655546
PK410	*LRIT3*	NM_198506.4	c.269dup	p.(Tyr90Ter)	0.000006387	Hom	WES	Novel
PK411	*LRIT3*	NM_198506.4	c.269dup	p.(Tyr90Ter)	0.000006387	Hom	WES	Novel
PK412	*LRIT3*	NM_198506.4	c.269dup	p.(Tyr90Ter)	0.000006387	Hom	WES	Novel
PK413	*LRIT3*	NM_198506.4	c.269dup	p.(Tyr90Ter)	0.000006387	Hom	WES	Novel
PK414	*NYX*	NM_022567.2	c.37+5G>T	p.?	n.a	Hemi	WES	Novel

CNV: copy number variation, Del: deletion, Hom: homozygous, Comphet: compound heterozygous, Het: heterozygous, ROH: runs of homozygosity, WES: whole exome sequencing, SNP: single nucleotide polymorphism, PMID: PubMed ID, n.a: not available.

7.2 First author publications

7.2.1 Exploring the Genetic Landscape of Retinal Diseases in North-Western Pakistan Reveals a High Degree of Autozygosity and a Prevalent Founder Mutation in *ABCA4*

genes

MDPI

Article

Exploring the Genetic Landscape of Retinal Diseases in North-Western Pakistan Reveals a High Degree of Autozygosity and a Prevalent Founder Mutation in *ABCA4*

Atta Ur Rehman [1], Virginie G. Peter [2,3,4], Mathieu Quinodoz [3,5], Abdur Rashid [6], Syed Akhtar Khan [7], Andrea Superti-Furga [1] and Carlo Rivolta [3,4,8,*]

1 Division of Genetic Medicine, Lausanne University Hospital and University of Lausanne, 1011 Lausanne, Switzerland; atta.rehman@unil.ch (A.U.R.); asuperti@unil.ch (A.S.-F.)
2 Institute of Experimental Pathology, Lausanne University Hospital and University of Lausanne, 1011 Lausanne, Switzerland; virginie.peter@unil.ch
3 Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, UK; mathieu.quinodoz@unil.ch
4 Clinical Research Center, Institute of Molecular and Clinical Ophthalmology Basel (IOB), 4031 Basel, Switzerland
5 Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland
6 Government Degree College Ara Khel, FR Kohat 26000, Khyber Pakhtunkhwa, Pakistan; rasheedkhan037@gmail.com
7 Department of Ophthalmology, Khalifa Gul Nawaz Hospital, Bannu 28100, Pakistan; drak8304@gmail.com
8 Department of Ophthalmology, University Hospital Basel, 4031 Basel, Switzerland
* Correspondence: carlo.rivolta@iob.ch

Received: 31 October 2019; Accepted: 17 December 2019; Published: 21 December 2019

Abstract: Variants in more than 271 different genes have been linked to hereditary retinal diseases, making comprehensive genomic approaches mandatory for accurate diagnosis. We explored the genetic landscape of retinal disorders in consanguineous families from North-Western Pakistan, harboring a population of approximately 35 million inhabitants that remains relatively isolated and highly inbred (~50% consanguinity). We leveraged on the high degree of consanguinity by applying genome-wide high-density single-nucleotide polymorphism (SNP) genotyping followed by targeted Sanger sequencing of candidate gene(s) lying inside autozygous intervals. In addition, we performed whole-exome sequencing (WES) on at least one proband per family. We identified 7 known and 4 novel variants in a total of 10 genes (*ABCA4*, *BBS2*, *CNGA1*, *CNGA3*, *CNGB3*, *MKKS*, *NMNAT1*, *PDE6B*, *RPE65*, and *TULP1*) previously known to cause inherited retinal diseases. In spite of all families being consanguineous, compound heterozygosity was detected in one family. All homozygous pathogenic variants resided in autozygous intervals ≥2.0 Mb in size. Putative founder variants were observed in the *ABCA4* (NM_000350.2:c.214G>A; p.Gly72Arg; ten families) and *NMNAT1* genes (NM_022787.3:c.25G>A; p.Val9Met; two families). We conclude that geographic isolation and sociocultural tradition of intrafamilial mating in North-Western Pakistan favor both the clinical manifestation of rare "generic" variants and the prevalence of founder mutations.

Keywords: hereditary retinal diseases; autozygosity mapping; consanguinity; Pakistan

1. Introduction

Inherited retinal dystrophies (IRDs) constitute a genetically heterogeneous group of rare conditions of the eye. They are mainly characterized by the progressive loss of rod and/or cone photoreceptors,

resulting in complete or nearly complete blindness at the end [1]. Globally, IRDs affect approximately one million people, with a frequency of 1 in 3000 births. Clinically, they may range from mild and non-progressive night blindness to more severe and degenerative phenotypes, including retinitis pigmentosa (RP) and cone or cone-rod dystrophies [2]. To date, mutations in over 271 genes have been linked to various forms of IRDs (RetNet; https://sph.uth.edu/RETNET/; accessed on 12 December 2019), and the sequencing of their coding parts has allowed the detection of pathogenic mutations in more than 60% of the patients [3]. IRDs are inherited as an autosomal recessive, autosomal dominant, X-linked, or mitochondrial trait, with autosomal recessive being the most prominent type [1,4]. Recent technological advancements, such as next-generation sequencing (NGS), have significantly increased gene discovery rates in a wide range of inherited ocular conditions [5], with a sensitivity value of ~75%, when applied to a clinically focused IRDs group [6]. Since consanguinity unmasks the adverse effects of recessive mutations through bi-parental inheritance of the same allele, it is possible to reveal the presence of disease-causing variants in consanguineous pedigrees by simply flagging large segments of consecutive homozygous genotypes surrounding the mutations, using a technique called "autozygosity mapping" [7–10]. For a more rapid and robust analysis, scientists usually combine autozygosity mapping with NGS to maximize the acquisition of relevant genetic information [10]. The combination of such information with targeted functional studies has provided significant insights into the molecular mechanisms of rare Mendelian diseases, including IRDs.

Since children of consanguineous couples are more likely than children of non-consanguineous parents to be affected by recessive genetic anomalies [11], the incidence of rare Mendelian diseases is higher in populations having a high degree of endogamy [12]. For example, Pakistan has one of the highest rates of inherited genetic diseases in the world, likely due to the fact that consanguinity is present in more than 50% of the population and marriages among first cousins are highly favored by the society [13,14]. According to a recent estimate, approximately 1.12 million people in Pakistan are blind, and the vision loss burden has continued to rise in the country since 1990 [15].

Although some information exists on blindness caused by cataracts or refractive errors, the prevalence of IRDs is not well documented in Pakistan at the level of the whole population. A hospital-based study in Karachi, a metropolitan city, revealed that 1 in 800 patients who visited the ophthalmic outpatient department had retinal dystrophies, with RP being the most frequent type (64%), followed by Stargardt disease (14.7%) and cone dystrophies (6.7%). Unsurprisingly, more than half of the patients from this study were born to consanguineous parents [16]. Recently, a few studies on IRDs have been published in Pakistan [17–20]. However, the majority of these reports were based on pedigrees from the Punjab and Sindh provinces, thus leaving North-Western Pakistan largely unexplored. Administratively known as Khyber Pakhtunkhwa (KP), this part of the country is predominantly a Pashtuns territory and includes a heterogeneous population of approximately 35 million inhabitants. Consanguinity in KP ranges between 22% and 66%, and the rate of consanguinity was found to have increased over time, possibly due to the growing violence and geo-political conflicts in the region (consanguineous marriages are believed to strengthen pre-existing intra-familial relationships and thus be advantageous in the context of civil unrest) [21–25]. To our knowledge, no comprehensive study on the genetic spectrum of IRDs has ever been undertaken in North-Western Pakistan, possibly because of socio-economic and cultural limitations, a lack of infrastructure, difficult terrain, and escalating conflicts in the region.

2. Materials and Methods

2.1. Enrollment of Families and Collection of Samples

Our study conforms to the standards of the Declaration of Helsinki and was approved by the Institutional Review Boards of the Hazara University, Mansehra, Pakistan (approval code: F.No:185/HU/Zool/2018/583) and of all our respective Institutions. Informed consents were provided in written form by all families prior to their participation and were signed by all members who

were enrolled. Families with at least two or more affected persons, a history of consanguinity, and a clear autosomal recessive inheritance pattern of the disease were selected for molecular analysis. All families participating in our study were ethnically Pashtuns and were geographically located in the KP province in North-Western Pakistan. Clinical and demographic information was obtained via a pre-designed questionnaire, and pedigrees were drawn and cross-checked through face-to-face interviews with patients and/or elder members of the family, in their native language. Clinical data were obtained either directly from the patients' medical reports (if available) or through consultation with the local clinicians/ophthalmologists who examined them on our request. Due to the limited medical infrastructure in the region, detailed clinical investigations, such as electroretinography (ERG) or fundus images, were not available. Furthermore, most of the clinical information obtained was derived from self-reported data, including, for example, difficulties in day/night vision, photophobia, disease onset and progression, response to medication, outcome of the Ishihara test, dark/light adaptation, ability to see/focus near and distant objects, loss of central or peripheral vision, and nystagmus. Patients were also investigated for other parameters, such as their ability to perform routine activities, e.g., reading, doing physical activities, and socializing with people. Electronic versions of pedigrees were created using the Pedigree Chart Designer (CeGaT, Tubingen, Germany). Saliva samples were collected by using the Oragene saliva kit (OG-500, DNA Genotek, Ottawa, ON, Canada) from patients as well as their clinically unaffected relatives, following the manufacturer's guidelines. DNA was extracted from these samples following standard protocols, e.g., by following the prepIT-L2P manual (DNA Genotek, Canada). Quantitative and qualitative assessments of DNA were made using the NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and electrophoresis on 1% agarose gels.

2.2. Genotyping and Homozygosity Mapping

Initially, nine families were subjected to genetic analysis using genome-wide high-density single-nucleotide polymorphism (SNP) arrays. For this purpose, genomic DNA of two or more individuals per family was genotyped by using the InfiniumCoreExome-24v1-1 array (Illumina, San Diego, CA, USA), which encompasses ~550,000 genome-wide SNP markers, at the iGE3 Genomics Platform of the University of Geneva, Switzerland. Arrays were processed using an iScan, according to the manufacturer's protocol. Genotype calls were generated using the GenomeStudio software by Illumina. PLINK was used to analyze the genotype data [26]. Following the identification of shared autozygous intervals among two or more patients from the same family, all exons and exon–intron boundaries of candidate gene(s) inside these intervals were sequenced using the Sanger method. Additionally, families that belonged to the same geographic area and had clinically overlapping phenotypes were also investigated for the presence of shared autozygous intervals. Using this method, putative disease-causing variants were identified in four consanguineous pedigrees segregating autosomal recessive IRDs, while the remaining unsolved families were subsequently characterized by whole-exome sequencing (WES).

2.3. Whole-Exome Sequencing

Overall, WES was performed for 10 pedigrees. For WES analysis, 2.0 µg of genomic DNA from index patients was initially processed by Novogene Co. Ltd (Hong Kong, China). Sequencing libraries were generated using the Agilent SureSelect Human All ExonV6 kit (Agilent Technologies, Santa Clara, CA, USA), while fragmentation was carried out by hydrodynamic shearing (Covaris, Massachusetts, MA, USA). Following adapter ligation, DNA fragments were selectively enriched in a PCR reaction. Products were purified using the AMPure XP system (Beckman Coulter, Beverly, CA, USA) and quantified using an Agilent high-sensitivity DNA assay on the Agilent Bioanalyzer 2100 system. Captured DNA libraries underwent paired-end sequencing on an Illumina Novaseq 6000 S4 platform, resulting in sequences of 150 bases (PE150 sequencing strategy). WES data were analyzed using our in-house computational pipeline [27], and autozygosity mapping was done

using AutoMap (unpublished). Finally, Sanger sequencing was performed to validate the potentially pathogenic variants detected and to confirm their causality, via genotype–phenotype cosegregation within the families.

3. Results

3.1. Clinical Synopsis

As mentioned earlier, detailed clinical investigation could not be achieved for all patients. However, a summary of the clinical information of one family is shown in Figure S1. On the basis of the few data available, mostly based on patients' symptoms, we could identify five major clinical IRD classes. Briefly, patients with severe early-onset blindness were tentatively categorized as individuals with Leber congenital amaurosis (LCA, two families), while patients presenting with IRD and extra-ocular symptoms such obesity, hypogonadism, learning/developmental disabilities, post-axial polydactyly of hands and/or feet, and renal abnormalities were examined by a local clinician who classified them as suffering from Bardet–Biedl syndrome (BBS) (two families). Patients with progressive loss of central vision were categorized as having macular dystrophy (eleven families), while those presenting with initial night blindness and progressive loss of peripheral vision were classified as suffering from RP (four families). Patients with RP were clinically evaluated with the help of a local ophthalmologist who reported the presence of bilateral bone spicules and peripheral retinal vascular attenuation, through fundus examination. Lastly, patients with a complete inability to discriminate between colors were classified as having achromatopsia (one family).

3.2. Molecular Findings

Collectively, we identified 11 disease-causing variants in 10 IRD-associated genes, in a total of 20 consanguineous IRDs pedigrees, all from North-Western Pakistan (Table 1). These variants were detected using a genome-wide SNP array followed by Sanger sequencing (four families), WES (ten families), and targeted Sanger sequencing alone (six families). Of these variants, seven were previously known IRD mutations, while four variants had never been identified before. Newly detected changes comprised three protein-truncating mutations and one nonsynonymous single-nucleotide variant (SNV). While the pathogenicity of protein-truncating variants can be easily postulated, the causality of the nonsynonymous SNV was inferred through in silico analysis and segregation studies. In spite of all families being consanguineous, compound heterozygosity was detected in one family (PK-E). All homozygous pathogenic variants were detected inside tractable autozygous intervals (≥2.0 Mb in size) and co-segregated with the disease in homozygosis in members from the remaining 19 families, including those who were not pre-ascertained by means of homozygosity mapping (Figure 1). Putative founder mutations were observed in the *ABCA4* (NM_000350.2:c.214G>A; p.Gly72Arg; 10 families) and in the *NMNAT1* genes (NM_022787.3:c.25G>A; p.Val9Met; 2 families), which together accounted for more than half of the IRD pedigrees analyzed in this study.

Table 1. Molecular findings in 20 consanguineous pedigrees from North-Western Pakistan segregating autosomal recessive inherited retinal dystrophies (IRDs).

Family ID	Tentative Diagnosis	Gene Name	Transcript ID	cDNA Change	Protein Change	GnomAD MAF	Mutation Number	Zyg-Osity	ROH Size	Method	References
PK004	ACHM	*CNGA3*	NM_001298.2	c.847C>T	p.Arg283Trp	0.0001402	M1	Hom	5-Mb	WES	[28]
PK-H	BBS	*MKKS*	NM_170784.2	c.280T>C	p.Phe94Leu	NA	M2	Hom	10-Mb	SNP array	This study
PK006	BBS	*BBS2*	NM_031885.3	c.1438C>T	p.Arg480Ter	0.00001647	M3	Hom	31-Mb	SNP array	[29]
PK-L	LCA	*NMNAT1*	NM_022787.3	c.25G>A	p.Val9Met	NA	M4	Hom	2-Mb	SNP array	[30]
PK-M	LCA	*NMNAT1*	NM_022787.3	c.25G>A	p.Val9Met	NA	M4	Hom	2-Mb	SNP array	[30]
PK001	RP	*TULP1*	NM_001289395.1	c.1307A>G	p.Lys436Arg	0.00002472	M5	Hom	17-Mb	WES	[31]
PK007	RP	*CNGA1*	NM_001142564.1	c.1298G>A	p.Gly433Asp	NA	M6	Hom	21-Mb	WES	[18]
PK009	MD	*CNGB3*	NM_019098.4	c.1574_1575del	p.Phe525Ter	NA	M7	Hom	39-Mb	WES	This study
PK010	RP	*RPE65*	NM_000329.2	c.550G>T	p.Glu184Ter	NA	M8	Hom	21-Mb	WES	This study
PK002	RP	*PDE6B*	NM_001145291.1	c.427del	p.Ala143LeufsTer7	NA	M9	Hom	4-Mb	WES	This study
PK-E	MD	*ABCA4*	NM_000350.2	c.3081T>G	p.Tyr1027Ter	NA	M10	Het	NA	SNP-WES	[32]
PK-E	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Het	NA	SNP-WES	[33]
PK-B	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	7-Mb	SNP-WES	[33]
PK-P	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	NA	TSS	[33]
PK-F	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	24-Mb	SNP-WES	[33]
PK-D	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	24-Mb	SNP-WES	[33]
PK028	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	NA	TSS	[33]
PK034	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	NA	TSS	[33]
PK037	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	NA	TSS	[33]
PK038	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	NA	TSS	[33]
PK043	MD	*ABCA4*	NM_000350.2	c.214G>A	p.Gly72Arg	0.00002784	M11	Hom	NA	TSS	[33]

MAF: Minor allele frequency; NA: Not available; Hom: Homozygous; Het: Heterozygous; ROH: Runs of homozygosity; WES: Whole-exome sequencing; TSS: Targeted Sanger sequencing. ACHM: Achromatopsia; BBS: Bardet–Biedl syndrome; LCA: Leber congenital amaurosis; RP: Retinitis pigmentosa; MD: Macular dystrophy.

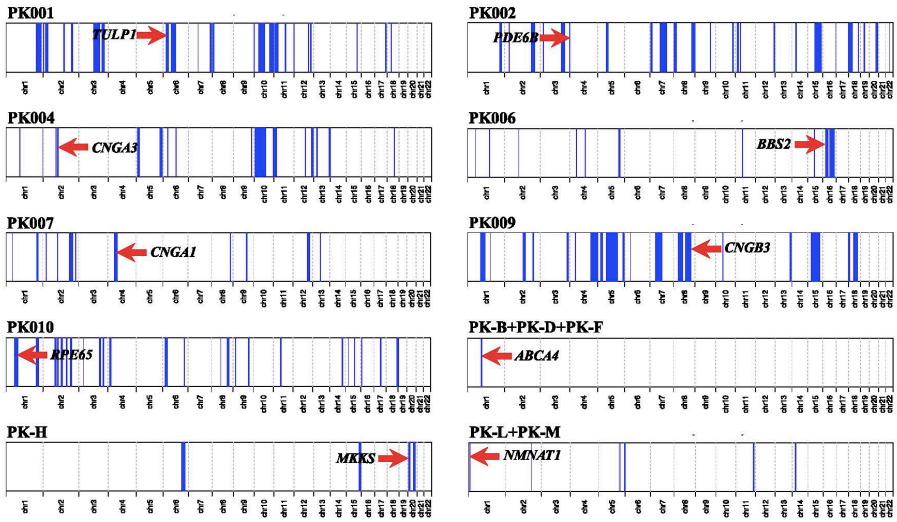


Figure 1. Homozygosity mapping showing genome-wide autozygous intervals as blue peaks. Red arrows indicate autozygous intervals harboring genes in which mutations were identified.

3.3. Macular Dystrophy (Possibly Including Stargardt Disease and Cone-Rod Degeneration)

Following SNP-based autozygosity mapping in three apparently unrelated families (PK-B, PK-D, and PK-F), we initially identified a ~2.0 Mb autozygous interval on chromosome 1, which was shared by three probands belonging to these families (Figure 1). Interestingly, the *ABCA4* gene was residing inside this interval, and an approximately 100 kb haplotype flanking this gene was identical in all three patients. WES analysis revealed a homozygous missense variant (NM_000350.2:c.214G>A:p.Gly72Arg) in *ABCA4*. The same variant (p.Gly72Arg) was also present in a compound heterozygous state with a nonsense mutation (NM_000350.2:c.3081T>G:p.Tyr1027Ter) in an additional family (PK-E) belonging to the same geographic location. Both of these variants have previously been identified to cause Stargardt disease [32,33]. Next, we performed targeted Sanger sequencing for p.Gly72Arg in a cohort of 18 previously uncharacterized consanguineous pedigrees from the region and identified the p.Gly72Arg mutation in six of them, in homozygosis. In total, p.Gly72Arg was found to cause disease in at least 10 independent pedigrees from a small town in North-Western Pakistan, collectively accounting for 37 patients (Figure 2). Geographically, these families belonged to Darra Adam Khel in North-Western Pakistan, an area which is mainly inhabited by the Afridi clan of Pashtuns ethnicity. Since these families were from the same geographic region and had a common ethnic affiliation, and an identical haplotype around *ABCA4* was detected in the three patients who were investigated for it, we believe that p.Gly72Arg constitutes a founder mutation.

Furthermore, WES analysis in family PK009 revealed a nonsense variant (NM_019098.4:c.1574_1575del:p.Phe525Ter) in the *CNGB3* gene, which co-segregated with the disease in homozygosis (Figure 2). This variant has never been reported in any public databases and constitutes a loss-of-function allele, therefore likely representing the molecular cause of disease in this family.

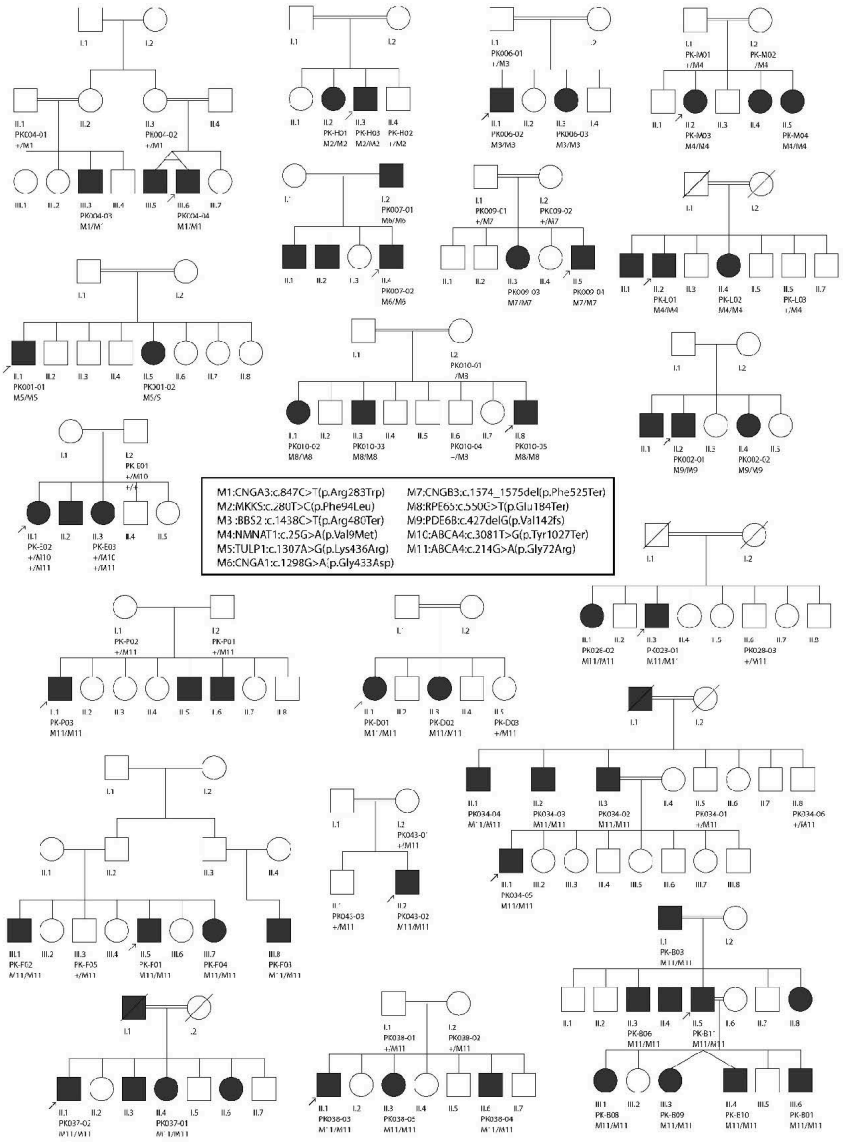


Figure 2. Segregation analysis of the mutations detected in a total of 20 consanguineous pedigrees from North-Western Pakistan. Probands in each pedigree are indicated by arrows. Due to space limitations, all pedigrees were trimmed.

3.4. Retinitis Pigmentosa

Through whole-exome sequencing in four consanguineous families with autosomal recessive RP, we identified disease-causing variants in four IRD-associated genes. These included: *TULP1* (NM_001289395.1:c.1307A>G:p.Lys436Arg; family PK001), *PDE6B* (NM_001145291.1:

c.427del:p.Ala143LeufsTer7; family PK002), *CNGA1* (NM_001142564.1: c.1298G>A:p.Gly433Asp; family PK007), and *RPE65* (NM_000329.2:c.550G>T:p.Glu184Ter; family PK010). All genes harboring these variants were located inside autozygous intervals with a genomic size of ≥2 Mb (Figure 1). All variants were validated by Sanger sequencing and confirmed to co-segregate with the disease in all affected family members (Figure 2). While variants identified in the *TULP1* (NM_001289395.1:c.1307A>G:p.Lys436Arg) and in the *CNGA1* (NM_001142564.1: c.1298G>A:p.Gly433Asp) genes were formerly known to cause RP [18,31], DNA changes in *PDE6B* and in *RPE65* have never been reported in any public databases. Since they are both loss-of-function variants (nonsense and frame-shift variants), they can be assumed to represent bona fide mutations.

3.5. Leber Congenital Amaurosis (Early-Onset Retinal Blindness)

Using SNP-based autozygosity mapping in two families with early-onset visual problems (PK-L, PK-M), we identified an autozygous interval on chromosome 1 that was shared by both probands from these families (Figure 1). Since *NMNAT1*, residing in this region, was a suitable candidate gene for LCA, we screened all exons and exon–intron boundaries of this gene, using Sanger sequencing. We found a missense variant (NM_022787.3:c.25G>A:p.Val9Met) in exon 2 that co-segregated with the disease in both families (Figure 2). The same variant (p.Val9Met) was previously reported to cause LCA in a pedigree of Pakistani descent [30]. However, we could not establish whether this previously identified family had any relationship with the pedigrees analyzed in our study. Considering the geographic proximity of these families, we suggest that p.Val9Met in *NMNAT1* constitutes another example of a founder mutation in North-Western Pakistan.

3.6. Bardet–Biedl Syndrome

Again, autozygosity mapping identified common intervals on chromosomes 16 and 20 in two families (PK006 and PK-H) with BBS (Figure 1); these intervals included the *BBS2* and the *MKKS* genes, respectively. In affected members of family PK006, the previously reported homozygous nonsense mutation *BBS2*:NM_031885.3:c.1438C>T:p.Arg480Ter was found [29], whereas family PK-H segregated a novel missense mutation (*MKKS*:NM_170784.2:c.280T>C:p.Phe94Leu). Both *BBS2* and *MKKS* variants co-segregated with the disease in a homozygous state and were present heterozygously in healthy individuals from both families (Figure 2). The *MKKS* variant was found to result in the disruption of a highly conserved amino acid (Phe94) and was predicted to be highly deleterious by numerous online prediction tools such as PolyPhen-2, PROVEAN, Mutation Taster, SIFT, MutationTaster2, and LRT [34–39]. Additionally, the variant was never identified in any public databases, including The Genome Aggregation Database (gnomAD), The Exome Aggregation Consortium (ExAC), and The Human Gene Mutation Database (HGMD) [40–42].

3.7. Achromatopsia

Through exome sequencing in a consanguineous pedigree (PK004) with three affected children suffering from putative complete achromatopsia, we identified a homozygous nonsynonymous single-nucleotide variant (NM_001298.2:c.847C>T:p.Arg283Trp) in the *CNGA3* gene. Autozygosity mapping revealed the *CNGA3* gene to lie within a 5 Mb autozygous interval on chromosome 2 (Figure 1). The mutation is a known cause of achromatopsia [28].

4. Discussion

Pakistan has one of the highest prevalence of inherited genetic diseases in the world [13], likely due to the high consanguinity rate of its population, generally exceeding 50% [11,12,14]. In this country, marriages of first cousins are highly favored, and families from Pakistan are considered a valuable resource for medical genetics research, which has led to significant scientific findings in the recent past [13,43]. Several studies on IRDs have been conducted in Pakistan during the last few years, but the majority of them were based on pedigrees from the Punjab and Sindh provinces [17–20]. Khan et

al. [20] showed that 90% of the mutations for non-syndromic IRDs and 100% of the mutations for syndromic IRDs were specific to families of Pakistani origin and that mutations in 35 different genes were found to cause non-syndromic IRDs specifically in families of Pakistani descent. In our study, we observed a different trend: out of the 11 mutations identified, only 4 were novel, whereas the remainder had previously been reported in patients from China, the UK, and Germany, and only two of them were found in Pakistani residents [28,29,31–33]. This probably reflects the fact that we focused our analysis on pedigrees from the North-Western part of the country, which remains largely isolated from populations inhabiting central Pakistan for cultural, linguistic, and geographic reasons. Further, North-Western Pakistan is predominantly a Pashtuns territory and, in contrast to central Pakistan, these populations trace their lineage to Pashtuns living in Western Afghanistan, with whom they share, even today, linguistic and sociocultural affiliations.

Our data therefore provide a first insight into the genetic landscape of IRDs in this peculiar part of the world, slightly extending the known mutational spectrum of IRDs. In particular, we have reported two founder mutations (*ABCA4*:c.214G>A:p.Gly72Arg and *NMNAT1*:c.25G>A:p.Val9Met) that, together, were responsible for disease in more than half of the patients in our study. At least 37 patients from 10 different families were found to be affected by the mutation in *ABCA4*, while the *NMNAT1* mutation touched at least 6 people in two unrelated families. Owing to the high rates of traditional intra-familial marriages, strong socio-cultural and ethnic divides, as well as geographic barriers, we predict that the number of patients/families affected by these two founder mutations may be even higher. Furthermore, a considerable number of variants identified in our study were detected inside large autozygous intervals (≥10 Mb), thus reflecting recent endogamy in the population. Similarly, our data support the previous notion that the likelihood for a homozygous pathogenic variant to be found within one autozygous interval is much higher in consanguineous pedigrees [10,44–46].

In summary, our study explored the genetic landscape of inherited retinal diseases in North-Western Pakistan, a large but relatively ignored part of the country. In addition to detecting a high degree of autozygosity and relevant founder mutations in the region, our data further expand the mutational spectrum of IRD-associated genes by adding four new variants to it. These findings will help future researchers/clinicians in their rapid screenings of patients from this region and will assist families seeking genetic counselling. We also predict that these insights on eye disorders might apply to other rare conditions that affect individuals from this region, such as deafness, intellectual disabilities, and developmental defects.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/1/12/s1, Figure S1: Clinical features from the proband of family PK-H (*MKKS*:NM_170784.2:c.280T>C:p.Phe94Leu). Fundus images of the patient's right (A) and left (B) eye. Post-axial polydactyly of the right (C) and left (D) foot, indicated by red arrows.

Author Contributions: Conceptualization, A.U.R., C.R., and A.S.-F.; Methodology, A.U.R., A.R., and S.A.K.; Software, V.G.P. and M.Q.; Validation, A.U.R., A.R., and S.A.K., Investigation, V.G.P. and M.Q.; Writing—original draft preparation, A.U.R.; Writing—review and editing, V.G.P., M.Q., C.R., and A.S.-F.; Supervision, C.R. and A.S.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the Swiss National Science Foundation (Grant #176097 to CR.)

Acknowledgments: We are highly indebted to all families for their volunteer participation in our study. We are also grateful to the Swiss Confederation for the award of Swiss Government Excellence PhD Fellowship to Atta Ur Rehman.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hamel, C.P. Gene discovery and prevalence in inherited retinal dystrophies. *C. R. Biol.* **2014**, *337*, 160–166. [CrossRef] [PubMed]
2. Berger, W.; Kloeckener-Gruissem, B.; Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Res* **2010**, *29*, 335–375. [CrossRef] [PubMed]

3. Duncan, J.L.; Pierce, E.A.; Laster, A.M.; Daiger, S.P.; Birch, D.G.; Ash, J.D.; Iannaccone, A.; Flannery, J.G.; Sahel, J.A.; Zack, D.J.; et al. Inherited Retinal degenerations: Current Landscape and knowledge gaps. *Transl. Vis. Sci. Technol.* **2018**, *7*, 6. [CrossRef] [PubMed]
4. Brown, M.D.; Voljavec, A.S.; Lott, M.T.; MacDonald, I.; Wallace, D.C. Leber's hereditary optic neuropathy: A model for mitochondrial neurodegenerative diseases. *Faseb. J.* **1992**, *6*, 2791–2799. [CrossRef]
5. Traboulsi, E.I. Hope and major strides for genetic diseases of the eye. *J. Genet.* **2009**, *88*, 395–397. [CrossRef] [PubMed]
6. Stone, E.M.; Andorf, J.L.; Whitmore, S.S.; DeLuca, A.P.; Giacalone, J.C.; Streb, L.M.; Braun, T.A.; Mullins, R.F.; Scheetz, T.E.; Sheffield, V.C.; et al. Clinically focused molecular investigation of 1000 consecutive families with inherited retinal disease. *Ophthalmology* **2017**, *124*, 1314–1331. [CrossRef]
7. Abu-Safieh, L.; Alrashed, M.; Anazi, S.; Alkuraya, H.; Khan, A.O.; Al-Owain, M.; Al-Zahrani, J.; Al-Abdi, L.; Hashem, M.; Al-Tarimi, S.; et al. Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. *Genome Res.* **2013**, *23*, 236–247. [CrossRef]
8. Khalak, H.G.; Wakil, S.M.; Imtiaz, F.; Ramzan, K.; Baz, B.; Almostafa, A.; Hagos, S.; Alzahrani, F.; Abu-Dhaim, N.; Abu Safieh, L.; et al. Autozygome maps dispensable DNA and reveals potential selective bias against nullizygosity. *Genet. Med.* **2012**, *14*, 515–519. [CrossRef]
9. Alkuraya, F.S. Autozygome decoded. *Genet. Med.* **2010**, *12*, 765–771. [CrossRef]
10. Alkuraya, F.S. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. *Hum. Genet.* **2013**, *132*, 1197–1211. [CrossRef]
11. Chinthapalli, K. First cousin marriage can double risk of birth defects, finds study. *Br. Med. J.* **2013**, *347*, f4374. [CrossRef] [PubMed]
12. Hamamy, H.; Antonarakis, S.E.; Cavalli-Sforza, L.L.; Temtamy, S.; Romeo, G.; Kate, L.P.; Bennett, R.L.; Shaw, A.; Megarbane, A.; van Duijn, C.; et al. Consanguineous marriages, pearls and perils: Geneva international consanguinity workshop report. *Genet. Med.* **2011**, *13*, 841–847. [CrossRef] [PubMed]
13. Riaz, M.; Tiller, J.; Ajmal, M.; Azam, M.; Qamar, R.; Lacaze, P. Implementation of public health genomics in Pakistan. *Eur. J. Hum. Genet.* **2019**, *27*, 1485–1492. [CrossRef] [PubMed]
14. National Institute of Population Studies. *Pakistan Demographic and Health Survey*; Pakistan, I., Ed.; NIPS: Maryland, MD, USA, 2013.
15. Hassan, B.; Ahmed, R.; Li, B.; Noor, A.; Hassan, Z.U. A comprehensive study capturing vision loss burden in Pakistan (1990–2025): Findings from the global burden of disease (GBD) 2017 study. *PLoS ONE* **2019**, *14*, e0216492. [CrossRef] [PubMed]
16. Adhi, M.I.; Ahmed, J. Frequency and clinical presentation of retinaly dystrophies—A Hospital based study. *Pak. J. Ophthalmol.* **2002**, *18*, 106–110.
17. Li, L.; Chen, Y.; Jiao, X.; Jin, C.; Jiang, D.; Tanwar, M.; Ma, Z.; Huang, L.; Ma, X.; Sun, W.; et al. Homozygosity mapping and genetic analysis of autosomal recessive retinal dystrophies in 144 consanguineous pakistani families. *Invest. Ophthalmol. Vis. Sci.* **2017**, *58*, 2218–2238. [CrossRef] [PubMed]
18. Maria, M.; Ajmal, M.; Azam, M.; Waheed, N.K.; Siddiqui, S.N.; Mustafa, B.; Ayub, H.; Ali, L.; Ahmad, S.; Micheal, S.; et al. Homozygosity mapping and targeted sanger sequencing reveal genetic defects underlying inherited retinal disease in families from pakistan. *PLoS ONE* **2015**, *10*, e0119806. [CrossRef]
19. Maranhao, B.; Biswas, P.; Gottsch, A.D.; Navani, M.; Naeem, M.A.; Suk, J.; Chu, J.; Khan, S.N.; Poleman, R.; Akram, J.; et al. Investigating the molecular basis of retinal degeneration in a familial cohort of pakistani decent by exome sequencing. *PLoS ONE* **2015**, *10*, e0136561. [CrossRef]
20. Khan, M.I.; Azam, M.; Ajmal, M.; Collin, R.W.; den Hollander, A.I.; Cremers, F.P.; Qamar, R. The molecular basis of retinal dystrophies in pakistan. *Genes* **2014**, *5*, 176–195. [CrossRef]
21. Pervaiz, R.; Faisal, F.; Serakinci, N. Practice of consanguinity and attitudes towards risk in the pashtun population of khyber pakhtunkhwa, pakistan. *J. Biosoc. Sci.* **2018**, *50*, 414–420. [CrossRef]
22. Ahmad, B.; Rehman, A.U.; Malik, S. Consanguinity and inbreeding coefficient in tribal pashtuns inhabiting the turbulent and war-affected territory of bajaur agency, north-west pakistan. *J. Biosoc. Sci.* **2016**, *48*, 113–128. [CrossRef] [PubMed]
23. Rehman, A.U.; Ahmad, I.; Zaman, M.; Malik, S. Transition in consanguinity in dir lower district, a victim of war, natural disaster and population displacement, in north-west pakistan—A response to sthanadar et al. (2015). *J. Biosoc. Sci.* **2016**, *48*, 421–426. [CrossRef] [PubMed]

24. Sthanadar, A.A.; Bittles, A.H.; Zahid, M. Civil unrest and the current profile of consanguineous marriage in Khyber Pakhtunkhwa province, Pakistan. *J. Biosoc. Sci.* **2014**, *46*, 698–701. [CrossRef]
25. Sthanadar, A.A.; Bittles, A.H.; Zahid, M. Increasing prevalence of consanguineous marriage confirmed in Khyber Pakhtunkhwa province, Pakistan. *J. Biosoc. Sci.* **2016**, *48*, 418–420. [CrossRef] [PubMed]
26. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef] [PubMed]
27. Royer-Bertrand, B.; Castillo-Taucher, S.; Moreno-Salinas, R.; Cho, T.J.; Chae, J.H.; Choi, M.; Kim, O.H.; Dikoglu, E.; Campos-Xavier, B.; Girardi, E.; et al. Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. *Sci. Rep.* **2015**, *5*, 17154. [CrossRef] [PubMed]
28. Kohl, S.; Marx, T.; Giddings, I.; Jagle, H.; Jacobson, S.G.; Apfelstedt-Sylla, E.; Zrenner, E.; Sharpe, L.T.; Wissinger, B. Total colourblindness is caused by mutations in the gene encoding the alpha-subunit of the cone photoreceptor cGMP-gated cation channel. *Nat. Genet.* **1998**, *19*, 257–259. [CrossRef] [PubMed]
29. Xing, D.J.; Zhang, H.X.; Huang, N.; Wu, K.C.; Huang, X.F.; Huang, F.; Tong, Y.; Pang, C.P.; Qu, J.; Jin, Z.B. Comprehensive molecular diagnosis of Bardet-Biedl syndrome by high-throughput targeted exome sequencing. *PLoS ONE* **2014**, *9*, e90599. [CrossRef]
30. Falk, M.J.; Zhang, Q.; Nakamaru-Ogiso, E.; Kannabiran, C.; Fonseca-Kelly, Z.; Chakarova, C.; Audo, I.; Mackay, D.S.; Zeitz, C.; Borman, A.D.; et al. NMNAT1 mutations cause Leber congenital amaurosis. *Nat. Genet.* **2012**, *44*, 1040–1045. [CrossRef]
31. Gu, S.; Lennon, A.; Li, Y.; Lorenz, B.; Fossarello, M.; North, M.; Gal, A.; Wright, A. Tubby-like protein-1 mutations in autosomal recessive retinitis pigmentosa. *Lancet* **1998**, *351*, 1103–1104. [CrossRef]
32. Fujinami, K.; Zernant, J.; Chana, R.K.; Wright, G.A.; Tsunoda, K.; Ozawa, Y.; Tsubota, K.; Robson, A.G.; Holder, G.E.; Allikmets, R.; et al. Clinical and molecular characteristics of childhood-onset Stargardt disease. *Ophthalmology* **2015**, *122*, 326–334. [CrossRef] [PubMed]
33. Rivera, A.; White, K.; Stohr, H.; Steiner, K.; Hemmrich, N.; Grimm, T.; Jurklies, B.; Lorenz, B.; Scholl, H.P.; Apfelstedt-Sylla, E.; et al. A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. *Am. J. Hum. Genet.* **2000**, *67*, 800–813. [CrossRef] [PubMed]
34. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **2013**, *7*, 7–20. [CrossRef] [PubMed]
35. Choi, Y.; Chan, A.P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747. [CrossRef]
36. Schwarz, J.M.; Rodelsperger, C.; Schuelke, M.; Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **2010**, *7*, 575–576. [CrossRef]
37. Sim, N.L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452–W457. [CrossRef]
38. Schwarz, J.M.; Cooper, D.N.; Schuelke, M.; Seelow, D. MutationTaster2: Mutation prediction for the deep-sequencing age. *Nat. Methods* **2014**, *11*, 361–362. [CrossRef]
39. Chun, S.; Fay, J.C. Identification of deleterious mutations within three human genomes. *Genome Res.* **2009**, *19*, 1553–1561. [CrossRef]
40. Stenson, P.D.; Ball, E.V.; Mort, M.; Phillips, A.D.; Shiel, J.A.; Thomas, N.S.; Abeysinghe, S.; Krawczak, M.; Cooper, D.N. Human gene mutation database (HGMD): 2003 update. *Hum. Mutat.* **2003**, *21*, 577–581. [CrossRef]
41. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *536*, 285–291. [CrossRef]
42. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alföldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv* **2019**, 531210.
43. Saleheen, D.; Natarajan, P.; Armean, I.M.; Zhao, W.; Rasheed, A.; Khetarpal, S.A.; Won, H.H.; Karczewski, K.J.; O'Donnell-Luria, A.H.; Samocha, K.E.; et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **2017**, *544*, 235–239. [CrossRef] [PubMed]

44. Wakeling, M.N.; Laver, T.W.; Wright, C.F.; De Franco, E.; Stals, K.L.; Patch, A.M.; Hattersley, A.T.; Flanagan, S.E.; Ellard, S.; Study, D.D.D. Homozygosity mapping provides supporting evidence of pathogenicity in recessive Mendelian disease. *Genet. Med.* **2019**, *21*, 982–986. [CrossRef]
45. Szpiech, Z.A.; Xu, J.; Pemberton, T.J.; Peng, W.; Zollner, S.; Rosenberg, N.A.; Li, J.Z. Long runs of homozygosity are enriched for deleterious variation. *Am. J. Hum. Genet.* **2013**, *93*, 90–102. [CrossRef] [PubMed]
46. Alkuraya, F.S. Discovery of rare homozygous mutations from studies of consanguineous pedigrees. *Curr. Protoc. Hum. Genet.* **2012**, *6*, 6–12. [CrossRef] [PubMed]

7.2.2 Whole-exome sequencing in a consanguineous Pakistani family identifies a mutational hotspot in the *COL7A1* gene, causing recessive dystrophic

Whole-exome sequencing in a consanguineous Pakistani family identifies a mutational hotspot in the *COL7A1* gene, causing recessive dystrophic epidermolysis bullosa

Atta Ur Rehman,[a] Virginie G. Peter,[a] Mathieu Quinodoz,[a] Muhammad Dawood[b] and Carlo Rivolta[a,c,d,e]

Dystrophic epidermolysis bullosa is a major form of epidermolysis bullosa and may be inherited as an autosomal dominant or recessive trait, with associated mutations in the *COL7A1* gene. Here, we describe a consanguineous Pakistani family with four affected individuals suffering from recessive dystrophic epidermolysis bullosa. Exome sequencing of the proband's DNA revealed a homozygous missense variant (c.8038G>A:p.Gly2680Ser) in *COL7A1* which cosegregated with disease in the family. The emergence of this particular glycine substitution in patients from diverse ethnic backgrounds such as China, United Kingdom, Poland, Iran, and Pakistan indicates that this variant most likely constitutes a recurrent mutational hotspot in the *COL7A1* gene, rather than a germline mutation present at low levels in the general population. *Clin Dysmorphol* 29:86–89 Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Clinical Dysmorphology 2020, 29:86–89

Keywords: consanguinity, dystrophic epidermolysis bullosa, Pakistan

[a]Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Switzerland; [b]Department of Dermatology, District Headquarter Hospital Khar, District Bajaur, Khyber Pakhtunkhwa, Pakistan; [c]Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom; [d]Clinical Research Center, Institute of Molecular and Clinical Ophthalmology Basel (IOB) and [e]Department of Ophthalmology, University Hospital Basel, Basel, Switzerland

Correspondence to Atta Ur Rehman, Mphil, Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Lausanne, Switzerland
Tel: +41 21 692 5469; fax: +41 21 692 5455; e-mail: atta.rehman@unil.ch

Received 13 May 2019 Accepted 19 August 2019

Introduction

Epidermolysis bullosa is a rare inherited blistering condition of the skin that affects approximately 1 in 120000 people (Fine, 2016). It is heterogeneous both clinically and genetically as, to date, there are more than 30 known clinical variants and at least 18 disease associated genes have been reported (McGrath, 2015). With a prevalence of 1 in ≈350000 individuals, dystrophic epidermolysis bullosa (DEB) is one of the major forms of epidermolysis bullosa (Fine, 2016). DEB may be inherited as a recessive (RDEB, MIM 226600) or dominant (DDEB, MIM 131750) disease, each further classified into different clinical subtypes. All major forms of DEB, whether dominant or recessive, are attributable to mutations in the *COL7A1* gene, which encodes type VII collagen (Pfendner and Lucky, 2018). This gene is relatively large (≈32 kb, 118 exons) and, according to the *COL7A1* gene variants database, it has been found to carry over 800 individual mutations (http://www.col7a1-database.info/ accessed 2019-04-03). While the majority of DEB-associated mutations are family specific, some recurrent mutations in *COL7A1* gene have also been reported (Wertheim-Tysarowska *et al.*, 2012).

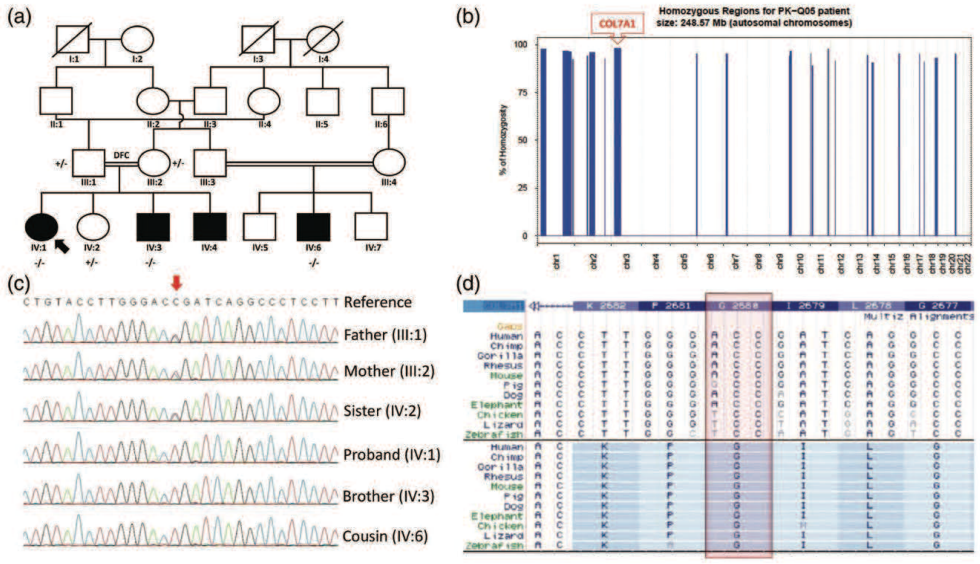
Clinical report

This study conforms to the standards of the Declaration of Helsinki and was approved by the Institutional Review Board of the Hazara University, Mansehra, Pakistan. Written informed consent of the patients was obtained prior to their participation in the study.

A consanguineous Pakistani family with four affected individuals suffering from RDEB was enrolled for molecular diagnosis. Patients were 3 to 12 years old, all born to consanguineous and healthy parents (Fig. 1a). The proband is the first child of double first cousins. She was presented with severe skin fragility and multiple skin injuries at birth, followed by mechanically induced generalized blistering of the skin. Blisters with atrophic scarring were, however, more frequent on hands, feet, knees, elbows, and flexural areas (Fig. 2c and d). Occasionally, the proband experienced chronic wounds at the blistering sites that lasted for more than 6 months. Anonychia, mild dental decay, and milia were also observed, and tissue granulation was restricted to the abdomen. Blistering of the oral mucosa caused severe swallowing difficulties in the proband, especially in her early childhood (Fig. 2a). At the time of examination (12 years of age), the proband was suffering from anemia and growth retardation. Progressive contractures of hands and feet were also noticed. Excessive blistering, contractures, and pseudosyndactylism of toes resulted in a 'mitten-feet' (Fig. 2d). Generally, more than 40% of her skin was affected by blisters and associated symptoms. In contrast, the proband's two affected brothers and one first cousin showed less severe symptoms. For instance, although generalized blistering of the skin and involvement of the oral mucosa was a common feature (Fig. 2b), these patients lacked symptoms like milia, atrophic scarring, chronic wounds, anonychia, tissue granulation, growth retardation, and

DOI: 10.1097/MCD.0000000000000299

Fig. 1



Molecular diagnosis of patients with RDEB. (a) Pedigree of the family. Filled symbols represent affected individuals, whereas open symbols indicate healthy subjects. The proband is designated by a black arrow. Double lines between symbols indicate consanguineous marriages. (+): wild-type allele, (–): mutant allele, p.Gly2680Ser. (b) Homozygosity map displaying ROH across the genome as vertical peaks. The position of *COL7A1* is shown on the top of a large 43-Mb ROH on chromosome 3. (c) Sanger sequencing results for all available family members. The missense variant (c.8038G>A:p.Gly2680Ser) is indicated by a red arrow. While both parents and a healthy sister of the proband are heterozygous, all patients are homozygous for this mutation. (d) Output from the UCSC genome browser, showing conservation of the missense variant c.8038G>A:p.Gly2680Ser (highlighted in red box) across vertebrates. DFC, double first cousin; RDEB, recessive dystrophic epidermolysis bullosa; ROH, runs of homozygous.

pseudosyndactylism of toes. The proportion of affected skin in these patients was also significantly lower. Taken together, the clinical presentation of the proband was clearly indicative of dystrophic epidermolysis bullosa. We further characterized the clinical subtype as 'RDEB-generalized intermediate' by using an electronic version of a clinical diagnostic tool for epidermolysis bullosa (Yenamandra *et al.*, 2017).

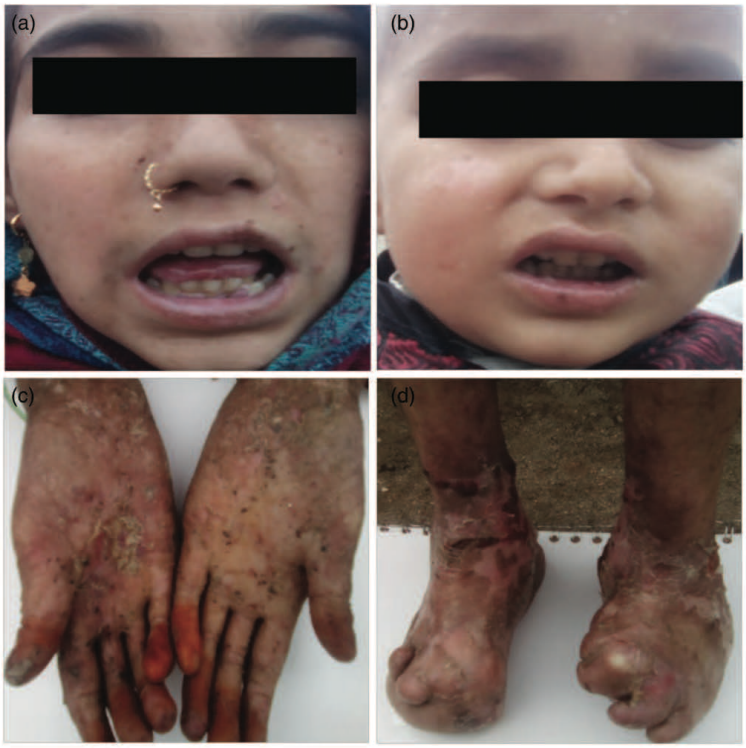
Results

Because of the marked genetic heterogeneity displayed by epidermolysis bullosa, we performed whole-exome sequencing (WES) on the proband's DNA. We first performed homozygosity mapping on the WES data, which enabled us to identify numerous runs of homozygosity (ROHs), globally accounting for 248.57 Mb of the autosomal exome (Fig. 1b). Notably, chromosomes 1, 2, and 3 alone contributed to more than 72% of the total homozygosity (Fig. 1b). Following a stringent computational analysis of the WES data, we retained 15 high-quality and rare (minor allele frequency, < 1%) homozygous variants that had an impact on the protein level (not synonymous nor intronic). Among them, a homozygous missense variant (NM_000094.3:c.8038G>A:p.Gly2680Ser, genome build hg19) in exon 108 of the *COL7A1* gene was considered as the potential candidate for the disease etiology. Upon Sanger validation, this variant perfectly cosegregated with the phenotype in the family (Fig. 1a and c). In silico analysis predicted this Glycine-to-Serine substitution as highly pathogenic. The functional significance of this variant could also be recognized by the fact that codon 2680 is highly conserved across vertebrate species, both at nucleotide and amino acid levels (Fig. 1d). Finally, c.8038G>A is an extremely rare allele (allele frequency = 0.00001992 in gnomAD, absent in South Asians and with highest population-specific allele frequency of 0.000123 in Africans) with no homozygotes identified so far (Lek *et al.*, 2016).

Discussion

A clear genotype-phenotype correlation has never been established in DEB because identical mutations in *COL7A1* have been found to cause different clinical variants of the disease (Almaani *et al.*, 2011; Knöpfel *et al.*, 2018). In particular, glycine substitutions at codon 2680 of *COL7A1* gene have been found to cause different clinical entities of DEB. For instance, c.8039G>A:p.Gly2680Asp was reported in a proband from the United Kingdom to be associated with DDEB, subtype

Fig. 2



Clinical presentation of patients with RDEB. (a) Proband showing oral blisters, mild teeth decay, and facial milia. (b) Proband's brother showing oral blisters. (c) Proband's hands showing excessive skin eruption as a consequence of frequent blisters. (d) Proband's feet, displaying progressive contractures coupled with pseudosyndactylism of toes, resulting in 'mitten-feet.' RDEB, recessive dystrophic epidermolysis bullosa.

pruriginosa (Almaani *et al.*, 2011) while the same amino acid, changed to serine (c.8038G>A:p.Gly2680Ser), was found in a Polish patient with severe generalized RDEB, in a compound heterozygous state with another missense variant (c.425A>G:p.Lys142Arg) (Wertheim-Tysarowska *et al.*, 2012). Likewise, the identical glycine substitution (c.8038G>A:p.Gly2680Ser) was reported in a compound heterozygous state with a nonsense mutation (c.6994C>T:p.Arg2332Ter) in an Iranian family with severe generalized RDEB (Vahidnezhad *et al.*, 2017). Last, this variant was reported in conjunction with a novel frameshift mutation (c.4871delC:p.Pro1624GlnfsTer86) in a Chinese family with Bart's Syndrome, a mild form of RDEB (Yan *et al.*, 2018). This highlights the fact that the absence of genotype-phenotype correlation in epidermolysis bullosa is not uncommon, which may be due to the involvement of unknown genetic or environmental factors (Knöpfel *et al.*, 2018).

We describe here the occurrence of a missense variant c.8038G>A:p.Gly2680Ser in the *COL7A1* gene detected homozygously in a consanguineous Pakistani family with RDEB. Consistent with previous findings, the clinical heterogeneity observed in this family is probably the consequence of the complete penetrance but the markedly variable expressivity of this mutation. Although this variant was previously reported to be associated with DEB, to the best of our knowledge, it was never identified in a homozygous state. Our findings suggest that this mutation can be pathogenic *per se* and possibly associated with generalized intermediate RDEB.

Conclusion

The emergence of c.8038G>A:p.Gly2680Ser mutation in patients from diverse ethnic backgrounds such as China, Poland, Iran, and now Pakistan indicates that this variant most likely constitutes a recurrent mutational hotspot in

the *COL7A1* gene, rather than a mutation present at low levels in the general population.

Acknowledgements

We are highly indebted to the members of the family for their participation in this study. We are also thankful to the Swiss Confederation for the award of 'Swiss Government Excellence Fellowship for PhD' to the first author of this article.

Conflicts of interest

There are no conflicts of interest.

References

Almaani N, Liu L, Dopping-Hepenstal PJ, Lai-Cheong JE, Wong A, Nanda A, *et al.* (2011). Identical glycine substitution mutations in type VII collagen may underlie both dominant and recessive forms of dystrophic epidermolysis bullosa. *Acta Derm Venereol* **91**:262–266.

Fine JD (2016). Epidemiology of inherited epidermolysis bullosa based on incidence and prevalence estimates from the national epidermolysis bullosa registry. *JAMA Dermatol* **152**:1231–1238.

Knöpfel N, Noguera-Morel L, Hernández-Martin A, García-Martin A, García M, Mencía Á, *et al.* (2018). Identical COL71A1 heterozygous mutations resulting in different dystrophic epidermolysis bullosa phenotypes. *Pediatr Dermatol* **35**:e94–e98.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, *et al.*; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**:285–291.

McGrath JA (2015). Recently identified forms of epidermolysis bullosa. *Ann Dermatol* **27**:658–666.

Pfendner EG, Lucky AW (2018). Dystrophic Epidermolysis Bullosa. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, *et al.*, editors. *GeneReviews(R)*. Seattle, WA: University of Washington.

Vahidnezhad H, Youssefian L, Zeinali S, Saeidian AH, Sotoudeh S, Mozafari N, *et al.* (2017). Dystrophic epidermolysis bullosa: COL7A1 mutation landscape in a multi-ethnic cohort of 152 extended families with high degree of customary consanguineous marriages. *J Invest Dermatol* **137**:660–669.

Wertheim-Tysarowska K, Sobczyńska-Tomaszewska A, Kowalewski C, Kutkowska-Kaźmierczak A, Woźniak K, Niepokój K, *et al.* (2012). Novel and recurrent COL7A1 mutation in a polish population. *Eur J Dermatol* **22**:23–28.

Yan Y, Meng Z, Hao S, Wang F, Jin X, Sun D, *et al.* (2018). Five novel COL7A1 gene mutations in three Chinese patients with recessive dystrophic epidermolysis bullosa. *Ann Clin Lab Sci* **48**:100–105.

Yenamandra VK, Moss C, Sreenivas V, Khan M, Sivasubbu S, Sharma VK, Sethuraman G. (2017). Development of a clinical diagnostic matrix for characterizing inherited epidermolysis bullosa. *Br J Dermatol* **176**:1624–1632.

7.3 Co-author publications

7.3.1 A frequent variant in the Japanese population determines quasi-Mendelian inheritance of rare retinal ciliopathy

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-10746-4 OPEN

A frequent variant in the Japanese population determines quasi-Mendelian inheritance of rare retinal ciliopathy

Konstantinos Nikopoulos[1,2,20], Katarina Cisarova[1,20], Mathieu Quinodoz[1,20], Hanna Koskiniemi-Kuendig[1], Noriko Miyake[3], Pietro Farinelli[1], Atta Ur Rehman[1], Muhammad Imran Khan[4,5], Andrea Prunotto[1], Masato Akiyama[6], Yoichiro Kamatani[6], Chikashi Terao[6], Fuyuki Miya[7], Yasuhiro Ikeda[8], Shinji Ueno[9], Nobuo Fuse[10], Akira Murakami[11], Yuko Wada[12], Hiroko Terasaki[9], Koh-Hei Sonoda[8], Tatsuro Ishibashi[8], Michiaki Kubo[13], Frans P.M. Cremers[4,5], Zoltán Kutalik[14], Naomichi Matsumoto[3], Koji M. Nishiguchi[15,16], Toru Nakazawa[15,16] & Carlo Rivolta[1,17,18,19]

Hereditary retinal degenerations (HRDs) are Mendelian diseases characterized by progressive blindness and caused by ultra-rare mutations. In a genomic screen of 331 unrelated Japanese patients, we identify a disruptive *Alu* insertion and a nonsense variant (p.Arg1933*) in the ciliary gene *RP1*, neither of which are rare alleles in Japan. p.Arg1933* is almost polymorphic (frequency = 0.6%, amongst 12,000 individuals), does not cause disease in homozygosis or heterozygosis, and yet is significantly enriched in HRD patients (frequency = 2.1%, i.e., a 3.5-fold enrichment; p-value = 9.2×10^{-5}). Familial co-segregation and association analyses show that p.Arg1933* can act as a Mendelian mutation in *trans* with the *Alu* insertion, but might also associate with disease in combination with two alleles in the *EYS* gene in a non-Mendelian pattern of heredity. Our results suggest that rare conditions such as HRDs can be paradoxically determined by relatively common variants, following a quasi-Mendelian model linking monogenic and complex inheritance.

[1] Unit of Medical Genetics, Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland. [2] Service of Medical Genetics, Lausanne University Hospital (CHUV), 1011 Lausanne, Switzerland. [3] Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan. [4] Department of Human Genetics, Radboud University Medical Center, 6500 HB Nijmegen, The Netherlands. [5] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, 6525 GA Nijmegen, The Netherlands. [6] Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan. [7] Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 113-8510, Japan. [8] Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan. [9] Department of Ophthalmology, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan. [10] Department of Integrative Genomics, Tohoku Medical Megabank Organization, Sendai 980-8573, Japan. [11] Department of Ophthalmology, Juntendo University School of Medicine, Tokyo 113-8421, Japan. [12] Yuko Wada Eye Clinic, Sendai 980-0011, Japan. [13] RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan. [14] Institute of Social and Preventive Medicine, Lausanne University Hospital, 1011 Lausanne, Switzerland. [15] Department of Ophthalmology, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan. [16] Department of Advanced Ophthalmic Medicine, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan. [17] Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, UK. [18] Institute of Molecular and Clinical Ophthalmology Basel (IOB), 4031 Basel, Switzerland. [19] University of Basel, 4001 Basel, Switzerland. [21] These authors contributed equally: Konstantinos Nikopoulos, Katarina Cisarova, Mathieu Quinodoz. Correspondence and requests for materials should be addressed to C.R. (email: carlo.rivolta@iob.ch)

Together with intellectual disabilities, hereditary retinal degenerations (HRDs, comprising retinitis pigmentosa and allied diseases) represent a group of conditions for which both genetic and allelic heterogeneity is the highest in humans[1,2]. To date, almost 300 genes and thousands of mutations have been identified as causative of HRD, and the detection of novel disease genes and variants continues at a steady pace (RetNet database: https://sph.uth.edu/retnet/). Considering that the overall prevalence of HRDs does not exceed 1 in 2000 individuals, the average contribution of any given HRD gene to the disease is incredibly small. Similarly, apart from two DNA variants that appear to be relatively frequent in the general population and determine a specific form of the disease (p.Asn1868Ile and p.Gly863Ala in *ABCA4*)[3,4], most mutations are so rare that they are seldom detected in more than one pedigree, worldwide. In addition, although HRDs affect people from the five continents, their specific allelic assortment seems to be population-specific[5,6]. For instance, similar to other islanders or groups of people that have experienced relative historical isolation, Japanese carry certain alleles, including pathogenic ones, which are not found elsewhere in the world[7]. Furthermore, lack of significant reduction in fitness before the reproductive age, associated with such an elevated heterogeneity, have led to the consequence that the number of recessive mutations that are detected heterozygously in the general, unaffected population is remarkably high and may affect up to one person in two[8].

Despite this extraordinary variability and abundance of mutations, HRDs are almost invariantly inherited as a monogenic, Mendelian trait, for which the presence of only one (dominant) or two (recessive) mutations in the same gene, genome-wide, is at the same time a necessary and sufficient condition for pathogenicity[9]. At the other end of the spectrum of ocular conditions having a genetic component lies age-related macular degeneration (AMD), another retinal disease affecting people aged 50 and over. AMD is a bona fide complex disease with a relatively high prevalence (1 in 13 individuals), favored by the presence of polymorphic SNPs, highly penetrant rare variants, and environmental factors[10]. Between these two pillars of inheritance, there is an intermediate zone, consisting in a few examples for which extremely rare mutations in more than one gene are associated with Bardet–Biedl syndrome, a retinal ciliopathy displaying sometimes digenic triallelic inheritance[11–13].

RP1 is one of the several HRD genes identified to date, and one of the few causing disease by more than one Mendelian pattern of inheritance. Originally described as linked to autosomal dominant retinitis pigmentosa (adRP)[14–16], a subtype of HRD, it was later shown to be associated with a recessive form of the same disease (arRP)[17]. To date, at least 60 mutations have been reported in *RP1*, most of which cluster within its last exon (exon 4), cumulatively accounting approximately for 5.5% and up to 4.5% of all adRP and arRP cases, respectively[18,19]. However, some DNA variants at the far 3' end of the gene, including nonsense variants, appear not to cause disease, at least not according to a dominant or recessive pattern of inheritance[20,21]. *RP1* encodes a multi-modular protein of 2156 amino acids, which is a member of the doublecortin family and is present in the ciliary axoneme of both rods and cones, the light-sensing neurons of the retina[22,23]. Mutations in *RP1* thus determine visual loss as a consequence of a ciliopathic phenotype affecting these specialized cell types.

Following the screen of a large set of Japanese HRD patients, we identify three mutations in the *RP1* gene: a mobile *Alu* element insertion in exon 4, a novel frameshift mutation, and a nonsense variant in the far 3' part of the coding sequence. While the first two variants behave as classical recessive Mendelian alleles, NM_006269.1:c.5797C>T/p.Arg1933* appears to cause disease according to a more complex pattern of inheritance. When present in *trans* with respect to the *Alu* insertion, it acts as a Mendelian mutation. Furthermore, despite being enriched in patients vs. controls, p.Arg1933* is completely benign in homozygosis or in heterozygosis. By performing an association test between 28 HRD patients, heterozygous carriers of this nonsense allele, and 3554 controls, we find that p.Arg1933* may be pathogenic not only as a Mendelian allele, but also in association with variants elsewhere in the genome, and in particular with two DNA changes in another ciliary gene, *EYS*.

Results

An *Alu* insertion in *RP1* is a prevalent cause of HRD in Japan. In the framework of a screening effort of 331 unrelated Japanese patients, we identified a novel, unusual mutation by whole-genome sequencing, consisting in the insertion of a mobile *Alu* element in exon 4 of the *RP1* gene (m1, or NM_006269.1: c.4052_4053ins328/p.Tyr1352Alafs*9) in a female individual from a recessive HRD family. This insertion caused the disruption of the reading frame by introducing 328 additional nucleotides, including a premature termination codon in the canonical *RP1* coding sequence. The mother of the proband was heterozygous for this variant and the proband's affected brother was also a homozygote, in support of the notion that this was indeed a recessive HRD mutation (Fig. 1a). Targeted screening for this *Alu* insertion in the remaining 330 patients (all forms of HRDs, isolate or recessive cases, not genetically pre-screened), as well as in 524 Japanese controls, available for direct testing of m1, identified 15 other affected and unrelated individuals and one heterozygous control carrying this insertion. In total, six patients were homozygous for the mutation (12 alleles), which co-segregated with the disease as a classical Mendelian, recessive allele, whenever this could be tested, while 10 carried it heterozygously. Altogether, these findings indicate that this *Alu* insertion is not only clearly pathogenic, but it is also a rather prevalent cause of retinal degeneration within the Japanese islands at the level of a single allele (1.8% of all HRD Japanese patients), possibly second only to the most frequent mutation so far identified in this country, i.e., NM_001142800.1:c.4957dup in *EYS*[24–26].

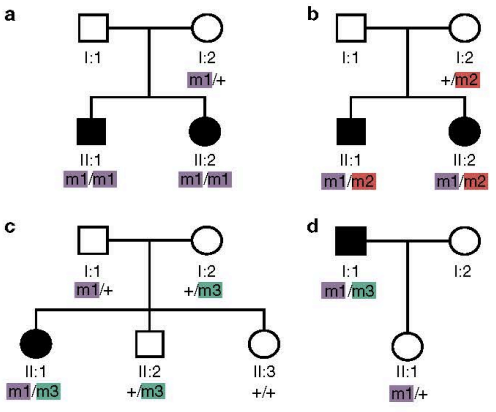


Fig. 1 Segregation analysis of the *RP1* mutations found in this study. Pedigrees of representative families are shown

Remarkably, 6 of the 10 individuals who carried the *Alu* insertion heterozygously were in fact compound heterozygotes for either of two other changes in *RP1*: a novel frameshift mutation (c.4196del/p.Cys1399Leufs*5, m2, two unrelated individuals) and a nonsense variant c.5797 C > T/p.Arg1933* (m3, four unrelated individuals) that was previously identified in the general population and is present in dbSNP as entry #rs118031911 (Supplementary Table 1). Again, both variants, detected by direct Sanger sequencing, co-segregated with the disease in relevant families, according to an autosomal recessive pattern of inheritance (Fig. 1b–d).

m3 is enriched in patients, but does cause HRD per se. Frameshift c.4196del/p.Cys1399Leufs*5 (m2) was absent from 3480 Japanese control chromosomes and was reported in the gnomAD database[27] to have an allele frequency of 5.44×10^{-5} in East Asia, indicating that this DNA variant is a very rare allele, as it is the case for most HRD mutations.

In contrast, the rs118031911/T allele (m3), despite being virtually absent in many world populations, was found to be relatively frequent in East Asians (Supplementary Fig. 1), and probably too frequent to be a Mendelian allele for HRD, according to the Hardy–Weinberg model. In particular, our direct screening of 12,379 Japanese individuals with no retinal degeneration showed the presence of rs118031911/T in 145 subjects, 142 heterozygotes and 3 homozygotes (148 alleles), validating the notion that this DNA variant is in fact almost polymorphic in Japan (allele frequency = 0.6%). All these subjects were examined by fundoscopy and, in addition, we evaluated clinically one of the three homozygotes (the only one who could be re-assessed, in agreement with our Institutional Review Board protocol) by a very thorough ophthalmological examination. At age 28 years old, she had no visual symptoms and displayed no ocular abnormalities: she had normal visual acuity (20/20 in both eyes), intact visual field (Goldmann perimetry), and no evidence of retinal degeneration through slit lamp examination and fundoscopy. Furthermore, optical coherence tomography imaging, used to assess detailed retinal structures, showed no signs of retinal thinning and electroretinogram, a test allowing objective detection of minimal retinal dysfunction even in the absence of subjective symptoms, showed normal responses. Finally, absence of late-onset HRD, who could have escaped detection in a 28-year-old individual, was confirmed by the assessment of the fundi of the other two rs118031911/T homozygotes, who displayed no signs of retinal degeneration at ages of 78 and 79 years, respectively. Overall, both population based-data and direct clinical assessments confirm that rs118031911/T does not cause HRD per se, in heterozygosis or in homozygosis.

However, specific screening for the rs118031911/T allele in the same cohort of 331 Japanese HRD patients mentioned above led to the identification of 10 additional heterozygotes (14 alleles in total) showing that its frequency in HRD patients was 2.1% (14 alleles out of 662) (Supplementary Fig. 1). The 3.5-fold enrichment of rs118031911/T in patients vs. controls (148 alleles out of 24,758 = 0.6%) was highly significant [p-value $= 9.2 \times 10^{-5}$, threshold = $0.05/N$, $N = 1$, by Fisher's exact test (24,610:148 vs. 648:14)], indicating that this relatively common variant has in fact an effect on retinal health. WES analysis of these 10 patients detected no mutations in HRD genes that could explain their phenotype, according to a Mendelian fashion of inheritance. Considering that rs118031911/T introduces a nonsense codon in the *RP1* open reading frame and was found in *trans* with respect to the *Alu* insertion in some patients, it is not unlikely that it could represent a hypomorphic variant contributing to the mutational load of genes involved in retinal homeostasis. In other words, despite being benign when considered as a Mendelian allele (monoallelically or biallelically), rs118031911/T could exert a pathogenic function in conjunction with DNA changes in other known HRD genes, according to an oligogenic pattern of inheritance that was previously modeled for hereditary ciliopathies[28–30].

A non-Mendelian pattern of inheritance for m3. We tested this hypothesis by assessing enrichment of nonsynonymous, rare, and low-frequency variants (minor allele frequency between 0.1% and 5%, according to published literature; further details in Methods) in m3 carriers. We analyzed the 10 patients mentioned above, as well as 18 additional patients with the same genotype (heterozygous for rs118031911/T, with no other recognized mutations in *RP1* and no mutations in HRD genes that could explain their phenotype), identified following a targeted screening of 713 Japanese HRD cases from another internal cohort (Supplementary Table 1). Specifically, we performed an association test between these 28 individuals and 3554 Japanese controls from the 3.5KJPN database[31] by considering all 228 bona fide HRD genes[32] from the RetNet database (Supplementary Table 2) that could produce multiallelic inheritance of HRD in m3 heterozygotes, in line with previous protocols involving similar analyses[33–38]. Cryptic relatedness among patients, as well as the presence of additional, undetected *RP1* mutations in *trans* with respect to rs118031911/T were excluded prior to performing the test (Supplementary Table 3 and Supplementary Data 1). The association analysis identified two variants that were significantly enriched in patients vs. controls (Table 1, Fig. 2). Interestingly, although they were not in linkage disequilibrium, both variants belonged to the gene *EYS* (NM_001142800.1:c.2528 G > A;p.Gly843Glu and c.8805 C > A;p.Tyr2935*, p-values = 5.6×10^{-5} and 1.9×10^{-4}, respectively; threshold = $2.8 \times 10^{-4} = 0.05/N$, where $N = 178$, by Fisher's exact test), possibly highlighting a mechanism of pathogenesis directly involving the proteins EYS and RP1. Indeed, a third DNA change within *EYS* (c.4957del;p.Ser1653Valfs*26) ranked 3rd in the list of associated variants, even if its p-value did not reach statistical significance after Bonferroni correction. Furthermore, we performed the same analysis by considering not only variants from RetNet sequences, but from the whole human exome, in both cohorts. In support of the data obtained, the two significant hits detected in the HRD gene set were also the two top hits detected exome-wide, even though no variant reached the threshold for exome-wide significance. Altogether, these results indicate that the rs118031911/T nonsense could act in concert with at least two DNA changes (and possibly with more) to determine a pathological phenotype in a non-Mendelian fashion.

Based on previous data on Bardet-Biedl syndrome[11], we tested a digenic diallelic vs. triallelic mode of action for rs118031911/T on HRD, by comparing the frequency of this variant in patients for whom the molecular causes of retinal degenerations were identified (i.e., solved cases) vs. unsolved HRD cases vs. controls. As expected, a comparison of unsolved vs. controls showed significance, as reported above, whereas solved vs. controls did not show any significant enrichment for rs118031911/T (6 rs118031911/T variants over 722 alleles for solved vs. 148 over 24,758 for controls, p-value = 0.46, OR = 1.4, CI = 0.50–3.14, by Fisher's exact test). Comparison of solved vs. unsolved HRD cases showed borderline non-significant enrichment for rs118031911/T in unsolved cases (p-value = 0.07, OR = 2.30, CI = 0.94–6.76, by Fisher's exact test), possibly indicating that either well-defined triallelism does

Table 1 Summary of the results of the association study

Gene	Variant	Frequency in cases	Frequency in controls	OR	95% CI (OR)	*p*-value
EYS	NM_001142800.1:c.2528G>A; p.Gly843Glu	0.125	0.017	8.23	3.08–18.78	5.6E-05*
EYS	NM_001142800.1:c.8805C>A; p.Tyr2935*	0.054	0.002	33.30	5.86–128.76	1.9E-04*
EYS	NM_001142800.1:c.4957del; p.Ser1653Valfs*26	0.054	0.004	13.85	2.62–46.83	1.9E-03
USH2A	NM_206933.2:c.15355C>T; p.Arg5119Trp	0.036	0.002	20.15	2.16–92.44	5.9E-03
USH2A	NM_206933.2:c.2802T>G; p.Cys934Trp	0.036	0.003	14.55	1.60–63.22	0.010
PDZD7	NM_024895.4:c.1267G>A; p.Ala423Thr	0.036	0.003	13.38	1.48–57.70	0.012
EYS	NM_001142800.1:c.7394C>G; p.Thr2465Ser	0.089	0.029	3.28	1.01–8.30	0.024
CC2D2A	NM_001080522.2:c.501G>T; p.Lys167Asn	0.036	0.004	8.44	0.96–34.65	0.027
RPGRIP1L	NM_015272.4:c.171G>T; p.Leu57Phe	0.054	0.011	4.90	0.96–15.63	0.028
BBIP1	NM_001243783.2:c.112T>C; p.Ser38Pro	0.036	0.005	7.70	0.87–31.35	0.032

Note: The top 10 hits from this test are shown
OR odds ratio, *CI* confidence interval
*p-values retaining statistical significance

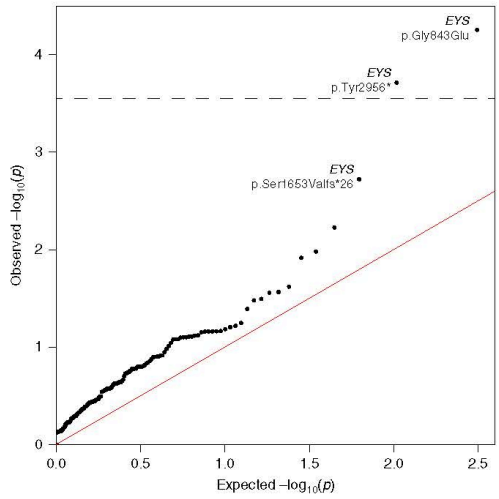


Fig. 2 Results of the association study. Quantile-quantile (Q-Q) plot of rare/low-frequency non-synonymous variants in HRD genes in 28 patients heterozygous for rs118031911/T vs. 3554 Japanese controls. The significance threshold is indicated by the dotted line

not take place for this variant or that we did not have enough power to detect it.

Discussion

The extreme genetic heterogeneity of retinal degenerations, together with the elevated number of pathogenic and hypomorphic changes in HRD genes that are detected in the unaffected population, have evoked the theoretical possibility that non-Mendelian, oligogenic inheritance could be responsible for these conditions[9]. Digenic heredity has been clearly demonstrated for specific combinations of mutations[39–41] in particular pedigrees or in individual cases, including digenic triallelic transmission of Bardet–Biedl syndrome[11,42]. For these patients, the presence of two (diallelic) or three (triallelic) mutations at two different loci (digenism) causes disease, presumably by compromising the overall function of gene products that belong to the same complex or are part of the same biochemical pathway. This model seems to be particularly true for genes encoding for proteins that form or play a role within the cell primary cilium, according to the paradigm of mutational load put forward by N. Katsanis and coworkers[43]. In these instances, accumulation of rare variants (which individually may have a little effect) in multiple ciliary genes can produce a pathological phenotype that is connected to ciliary function and result in a ciliopathy, including retinal ciliopathies[44–46].

Intriguingly, despite our association test was not limited to ciliary genes, both our significant hits lie within a ciliary gene, *EYS*. In primates, the EYS protein has been shown to physically co-localize with RP1 in the ciliary axoneme of photoreceptors and is thought to play a role in the structural organization and maintenance of these cells' apical part, the outer segment (OS)[47]. This functional role is further supported by studies in zebrafish, where *EYS* knockouts show progressive retinal degeneration due to mis-localization of specific OS proteins and the disruption of F-actin filaments[48,49], a key component not only for the integrity, but also for the morphogenesis of the OS[50]. In a similar fashion, targeted disruption of the *RP1* gene in mice leads to defects of the OS, because of the incorrect stacking of its discs[51]. The co-localization of RP1 and EYS, as well as their common role in the homeostasis of the OS, strongly indicates that they may have synergic functions and that pathogenesis could occur in a digenic fashion.

In this work we show that two specific *RP1* alleles are responsible for a relatively large number of Mendelian HRD cases in Japan. Interestingly, none of these two changes is a rare allele at all, compared to the average frequencies of classical HRD mutations. The first, the c.4052_4053ins328/p.Tyr1352Alafs*9 *Alu* element insertion in *RP1* (m1), seems to be the second most common HRD recessive mutation described so far in Japan, and its frequency may even be underestimated, since insertional events of mobile elements are difficult to detect by conventional screening techniques. The second variant, c.5797 C > T/p.Arg1933* or rs118031911/T (m3), is even more frequent, and by far more interesting. Despite introducing a premature stop codon in the *RP1* open reading frame, this DNA change is almost polymorphic in East Asia and does not cause disease either in heterozygous or homozygous carriers. However, this same change may act as pathogenic allele in a Mendelian fashion (with another *RP1* mutation in *trans*), or in association with rare variants in at least another gene, according to a non-Mendelian, possibly oligogenic pattern of inheritance. Although we currently ignore the molecular mechanisms leading to this unusual model of pathogenicity, it is probably the consequence of an increased global mutational load with threshold effect, determined by the accumulation of variants with different pathogenic potential. The presence of one or of two rs118031911/T alleles likely produces a load that is below this pathological threshold, while the co-

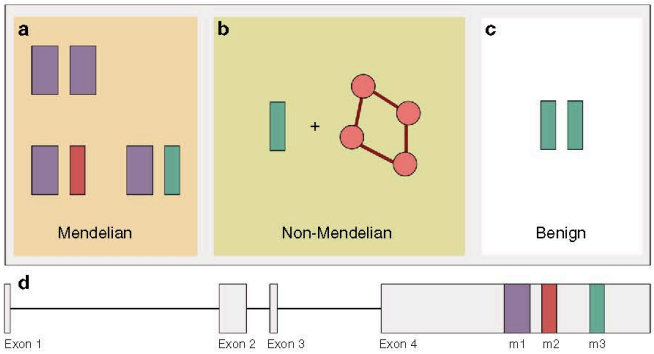


Fig. 3 Schematic representation of the inheritance pattern of the identified mutations in *RP1*, highlighting the concept of rs118031911/T-mediated quasi-Mendelian inheritance of HRDs. **a** In *trans* with respect to the *Alu* element insertion (m1, or c.4052_4053ins328/p.Tyr1352Alafs*9), m3 (rs118031911/T, or c.5797C>T/p.Arg1933*) results in autosomal recessive inheritance of the disease, similar to m1 in a homozygous state or in a compound heterozygous combination with m2 (c.4196del/p.Cys1399Leufs*5). **b** Combinations of the hypomorphic m3 allele with additional hypomorphs and/or heterozygous recessive alleles in other genes result in disease following a non-Mendelian pattern, whereas (**c**) homozygosis for m3 has no pathological consequences. **d** Structure of *RP1*: exons are represented by boxes, connected by solid lines (introns). The relative positions of m1, m2, and m3 are also indicated

occurrence of extra variants could result in the crossing of such a limit for normal retinal homeostasis. This hypothesis is supported by the evidence that rs118031911/T is pathogenic in conjunction with a very severe mutation, i.e., the insertion of an *Alu* element in *RP1*'s exon 4 mentioned above, which completely ablates the open reading frame of the gene. We term this model of inheritance quasi-Mendelian, to define the differential behavior (Mendelian or non-Mendelian) that specific alleles may have with respect to different genotypes at the same locus or elsewhere in the genome.

In conclusion, it seems that, at least for *RP1*-associated HRD, disorders displaying a Mendelian pattern of inheritance may also genetically behave like multigenic conditions, for which both polymorphic (having a low effect) and rare (having a rather high effect) variants can determine pathogenesis (Fig. 3). The findings from our study require further replication, ideally in other East Asian cohorts. However, the low prevalence of HRD and the even lower percentage of HRD patients carrying rs118031911/T limit our ability to assemble cohorts of sufficient power at this time, which would not only enable us to validate our findings but also to propose more defined models of pathogenicity. To better illustrate this: the identification of the 28 heterozygotes reported in this study corresponds to the screening of roughly 5–10 million Japanese individuals. Nevertheless, our work provides a proof of concept that a non-negligible proportion of HRDs can be caused by inheritance mechanisms that transcend the Mendelian model, to be investigated in detail by future, very large-scale and population-specific sequencing endeavors, such as for instance the 100,000 genomes project[52]. Furthermore, our findings suggest that oligogenic heredity of human diseases (and perhaps of other traits) may not be limited to a low number of cases with ultra-rare conditions, as shown up to now[34,35,53,54], but could extend to more frequent phenotypes and represent a bridge between monogenic and complex inheritance.

Methods

Subjects. The study was initiated following the approval by the Institutional Review Boards of our respective Institutions (University of Lausanne, Yokohama City University Graduate School of Medicine, Radboud University Medical Center, Kyushu University, Nagoya University Graduate School of Medicine, Tohoku Medical Megabank Organization, Juntendo University School of Medicine, Tohoku University Graduate School of Medicine, and Tokyo Medical and Dental University). All subjects provided written informed consent, and the study was conducted in adherence with the Declaration of Helsinki.

Tohoku University School of Medicine, Kyushu University School of Medicine, and Nagoya University School of Medicine, all based in Japan, were the centers where all Japanese patients with HRD were recruited. HRD was diagnosed clinically after excluding possible secondary causes of retinal degeneration such as toxicity and uveitis. Final diagnosis required the presence of reduced electroretinogram (ERG) responses, visual field loss, and funduscopic abnormalities consistent with retinal degeneration (retinal vascular narrowing and abnormalities of the retinal pigment epithelium etc.) symmetrically in both eyes.

Genotypes from individuals without HRD were collected from both published and unpublished databases [the BioBank Japan Project ($N = 12,379$), the ToMMo Japanese Reference Panel Project (3.5KJPN release, $N = 3554$)[31], the Tohoku University School of Medicine ($N = 95$), the Yokohama City University Graduate School of Medicine ($N = 429$)] or were obtained experimentally by direct genotyping of genomic DNA, with standard molecular biology techniques.

Summary phenotypes of carriers of m1, m2, or m3 mutations are listed in Supplementary Table 1.

Whole genome sequencing and analysis. Genome sequencing of the first index patient was performed using the sequencing platform by Complete Genomics[55]. Sequence reads were mapped to the human reference genome (NCBI build 37) and variants were called genome-wide. These included: single-nucleotide variants (SNVs), copy-number variations (CNVs), as well as structural variations (SVs) such as *Alu* element insertions and/or chromosomal rearrangements. Data were extracted from MasterVar files and other relevant matrices by ad hoc Perl, bash, and R scripts, available upon request. Assessment of pathogenic variants was performed as previously described[56].

Screening for the *Alu* element insertion. In order to screen for the presence of the *Alu* element in exon 4 of *RP1* distinct pair of primers were designed (forward: 5′-AGGCTTGTTTCCTAGGAGAGGT-3′, reverse: 5′-TTCTGCTTCTTTTTCACTTAGGC-3′) using the CLCbio Genomics Workbench (Qiagen, Hilden, Germany).

PCR amplification was performed in a 20 µl total volume containing 20 ng genomic DNA, 1× GoTaq buffer, 0.5 mM dNTPs, 10 µM of each primer, and 2 units (5 U/µl) of GoTaq polymerase (Promega, Madison, Wisconsin). PCR products were separated following agarose gel electrophoresis. PCR products displaying abnormal size profiles were purified (ExoSAP-IT, USB, Cleveland Ohio) and a sequencing reaction was performed in a total volume of 5 µl using 1 µl primer 3.3 µM, 0.5 µl BigDye Terminator v1.1, and 1 µl of the provided Buffer (Applied Biosystems, Foster City, California) Big Dye terminator cycle sequencing kit on an ABI 3130xl Genetic Analyzer (Applied Biosystems).

For this screening, we used 524 controls from the Tohoku University School of Medicine ($N = 95$) and the Yokohama City University Graduate School of Medicine ($N = 429$).

Whole exome sequencing and analysis. Paired-end DNA sequencing libraries of 28 individuals were generated using Agilent SureSelect Human All ExonV6 kit (Agilent Technologies, CA, USA) by Novogene Co., Ltd., Hong Kong. One microgram of genomic DNA per sample was fragmented into 180–280 bp fragments by hydrodynamic shearing (Covaris, Massachusetts, USA). After the

reparation of the 3' and 5' ends and the adenylation of the 3' ends, paired-end adaptors were ligated to the DNA fragments. DNA fragments with ligated adaptors on both ends were enriched by PCR. PCR products were further purified using the AMPure XP system (Beckman Coulter, Beverly, USA) and quantified using the Agilent high sensitivity DNA assay on the Agilent Bioanalyzer 2100 system. Captured DNA fragments were sequenced on an Illumina NovaSeq 6000 platform (Illumina, San Diego, California).

Raw sequence files were assessed, trimmed, and mapped to the human genome reference sequence (UCSC hg19) using Novoalign V3.08.02 (Novocraft, Selangor, Malaysia). Variants were called jointly by GATK 3.8[57] and annotation was performed using EPACTS (http://genome.sph.umich.edu/wiki/EPACTS) and Annovar[58]. The nomenclature of all variants studied was validated by using VariantValidator[59].

All single nucleotide variants were further filtered to obtain only high-quality variants. Briefly, quality control was carried out using the following parameters: (1) remove individual calls if Depth (DP) < 8 or GenotypeQuality (GQ) < 20, (2) exclude variants if the average GQ value ≤35, (3) exclude variants if call-rate value ≤ 0.9, (4) keep only variants with no deviation from Hardy-Weinberg equilibrium ($p > 0.05$ after Bonferroni correction), (5) keep variants passing GATK VQSR (VQSRTranche of 90.0), (6) final hard filtering step with Quality by Depth (QD) ≥ 2, FisherStrand (FS) ≤ 60, RMSMappingQuality (MQ) ≥ 40, MappingQualityRankSumTest (MQRankSum) ≥ −12.5, ReadPosRankSumTest (ReadPosRankSum) ≥ −8, StrandOddsRatio (SOR) ≤ 3 and ExcessHet ≤ 20.

Association study on rare and low-frequency variants in RetNet genes. An association test was performed on rare and low-frequency non-synonymous variants from a curated RetNet list ($N = 228$ genes, Suppl. Table 2) to test possible association of variants in the 28 patients carrying rs118031911/T heterozygously compared to 3554 controls from the ToMMo database[31]. More specifically, variants were retained if they had a frequency comprised between 0.1 and 5% in controls, with the exclusion of rs118031911/T itself, according to published methods[60–62], and in particular to the paper by Marouli and coworkers[60]. This resulted in the selection of 178 variants in 84 different genes, which were used to test association in rs118031911/T carriers vs. controls by Fisher's exact test with an experiment-wide Bonferroni-corrected threshold of 2.81×10^{-4}, for $\alpha = 0.05$ ($0.05/178 = 2.81 \times 10^{-4}$). *P*-values and odds ratios were obtained by the fisher.test function with default parameters in R (v3.5.1) and the Q–Q plot (Fig. 2) was obtained by using the qqman package[63].

Relatedness analyses. PLINK (v1.90b5)[64] was used to compute PI_HAT values between all pairs of the 28 rs118031911/T carriers using calls from whole exome sequencing. This analysis showed no relatedness, with PI_HAT values between 0.00 and 0.07 (threshold for relatedness = 0.2)[65], for all 378 possible pairwise combinations (Supplementary Data 1).

SNP genotyping of the *RP1* locus. To detect haplotypes in *trans* with respect to rs118031911/T, we genotyped 23 SNPs encompassing the *RP1* locus over ~260 kb, by standard techniques.

Data availability

The data supporting the findings of this study, as a whole, contain information that could compromise the privacy/consent of the participants, therefore we provide the genotypes of the cases analyzed as a summary statistics file containing aggregated data (Supplementary Data 2). Genotypes of control Japanese individuals from the ToMMo Japanese Reference Panel Project (3.5KJPN release, v20181105open) can be accessed at https://jmorp.megabank.tohoku.ac.jp/201905/downloads. Data from the BioBank Japan Project and the Tohoku University School of Medicine and Yokohama City University Graduate School of Medicine is available upon request to the corresponding author, pending authorization of the Centers that generated them and in agreement with their specific IRB approvals.

Code availability

Processing of raw sequence files, variant calling, variant annotation, as well as all the other analyses described in this work were performed using available and open source software, as detailed in the Methods section. No custom code was therefore used, aside from simple scripts integrating commands for such programs.

Received: 21 September 2018 Accepted: 23 May 2019
Published online: 28 June 2019

References

1. den Hollander, A. I., Black, A., Bennett, J. & Cremers, F. P. Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. *J. Clin. Invest.* **120**, 3042–3053 (2010).
2. Berger, W., Kloeckener-Gruissem, B. & Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Eye Res.* **29**, 335–375 (2010).
3. Runhart, E. H. et al. The common ABCA4 Variant p.Asn1868Ile shows nonpenetrance and variable expression of stargardt disease when present in trans with severe variants. *Invest. Ophthalmol. Vis. Sci.* **59**, 3220–3231 (2018).
4. Zernant, J. et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J. Med. Genet.* **54**, 404–412 (2017).
5. Jin, Z. B. et al. Identifying pathogenic genetic background of simplex or multiplex retinitis pigmentosa patients: a large scale mutation screening study. *J. Med. Genet.* **45**, 465–472 (2008).
6. Khan, M. I. et al. The molecular basis of retinal dystrophies in pakistan. *Genes* **5**, 176–195 (2014).
7. Yamaguchi-Kabata, Y. et al. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
8. Nishiguchi, K. M. & Rivolta, C. Genes associated with retinitis pigmentosa and allied diseases are frequently mutated in the general population. *PLoS One* **7**, e41902 (2012).
9. Rivolta, C., Sharon, D., DeAngelis, M. M. & Dryja, T. P. Retinitis pigmentosa and allied diseases: numerous diseases, genes, and inheritance patterns. *Hum. Mol. Genet.* **11**, 1219–1227 (2002).
10. den Hollander, A. I. & de Jong, E. K. Highly penetrant alleles in age-related macular degeneration. *Cold Spring Harb. Perspect. Med.* **5**, a017202 (2014).
11. Katsanis, N. et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256–2259 (2001).
12. Katsanis, N. The continuum of causality in human genetic disorders. *Genome Biol.* **17**, 233 (2016).
13. Badano, J. L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat. Rev. Genet.* **3**, 779–789 (2002).
14. Pierce, E. A. et al. Mutations in a gene encoding a new oxygen-regulated photoreceptor protein cause dominant retinitis pigmentosa. *Nat. Genet.* **22**, 248–254 (1999).
15. Guillonneau, X. et al. A nonsense mutation in a novel gene is associated with retinitis pigmentosa in a family linked to the RP1 locus. *Hum. Mol. Genet.* **8**, 1541–1546 (1999).
16. Sullivan, L. S. et al. Mutations in a novel retina-specific gene cause autosomal dominant retinitis pigmentosa. *Nat. Genet.* **22**, 255–259 (1999).
17. Khaliq, S. et al. Novel association of RP1 gene mutations with autosomal recessive retinitis pigmentosa. *J. Med. Genet.* **42**, 436–438 (2005).
18. Avila-Fernandez, A. et al. Identification of an RP1 prevalent founder mutation and related phenotype in Spanish patients with early-onset autosomal recessive retinitis. *Ophthalmology* **119**, 2616–2621 (2012).
19. El Shamieh, S. et al. Targeted next generation sequencing identifies novel mutations in RP1 as a relatively common cause of autosomal recessive rod-cone dystrophy. *Biomed. Res. Int.* **2015**, 485624 (2015).
20. Baum, L. et al. RP1 in Chinese: Eight novel variants and evidence that truncation of the extreme C-terminal does not cause retinitis pigmentosa. *Hum. Mutat.* **17**, 436 (2001).
21. Ziviello, C. et al. Molecular genetics of autosomal dominant retinitis pigmentosa (ADRP): a comprehensive study of 43 Italian families. *J. Med. Genet.* **42**, e47 (2005).
22. Liu, Q. et al. Identification and subcellular localization of the RP1 protein in human and mouse photoreceptors. *Invest. Ophthalmol. Vis. Sci.* **43**, 22–32 (2002).
23. Liu, Q., Zuo, J. & Pierce, E. A. The retinitis pigmentosa 1 protein is a photoreceptor microtubule-associated protein. *J. Neurosci.* **24**, 6427–6436 (2004).
24. Iwanami, M., Oshikawa, M., Nishida, T., Nakadomari, S. & Kato, S. High prevalence of mutations in the EYS gene in Japanese patients with autosomal recessive retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **53**, 1033–1040 (2012).
25. Suto, K. et al. Clinical phenotype in ten unrelated Japanese patients with mutations in the EYS gene. *Ophthalmic Genet.* **35**, 25–34 (2014).
26. Arai, Y. et al. Retinitis Pigmentosa with EYS mutations is the most prevalent inherited retinal dystrophy in Japanese Populations. *J. Ophthalmol.* **2015**, 819760 (2015).
27. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
28. Hildebrandt, F., Benzing, T. & Katsanis, N. Ciliopathies. *N. Engl. J. Med.* **364**, 1533–1543 (2011).
29. Badano, J. L. et al. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. *Nature* **439**, 326–330 (2006).
30. Lindstrand, A. et al. Copy-number variation contributes to the mutational load of Bardet-Biedl syndrome. *Am. J. Hum. Genet.* **99**, 318–336 (2016).
31. Nagasaki, M. et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. *Nat. Commun.* **6**, 8018 (2015).

32. Hanany, M. & Sharon, D. Allele frequency analysis of variants reported to cause autosomal dominant inherited retinal diseases question the involvement of 19% of genes and 10% of reported pathogenic variants. *J. Med. Genet.* (2019).
33. Ji, W. et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat. Genet.* **40**, 592–599 (2008).
34. Sykiotis, G. P. et al. Oligogenic basis of isolated gonadotropin-releasing hormone deficiency. *Proc. Natl Acad. Sci. USA* **107**, 15140–15144 (2010).
35. Kousi, M. et al. Evidence for secondary-variant genetic burden and non-random distribution across biological modules in a recessive ciliopathy. Preprint at https://www.biorxiv.org/content/10.1101/362707v1.full (2018).
36. Momozawa, Y. et al. Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. *Nat. Genet.* **43**, 43–47 (2011).
37. Klassen, T. et al. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. *Cell* **145**, 1036–1048 (2011).
38. Stessman, H. A. et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49**, 515–526 (2017).
39. Ebermann, I. et al. PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. *J. Clin. Invest.* **120**, 1812–1823 (2010).
40. Kajiwara, K., Berson, E. L. & Dryja, T. P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604–1608 (1994).
41. Yoshimura, H. et al. Massively parallel DNA sequencing facilitates diagnosis of patients with Usher syndrome type 1. *PLoS ONE* **9**, e90688 (2014).
42. Beales, P. L. et al. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. *Am. J. Hum. Genet.* **72**, 1187–1199 (2003).
43. Zaghloul, N. A. & Katsanis, N. Functional modules, mutational load and human genetic disease. *Trends Genet.* **26**, 168–176 (2010).
44. Tory, K. et al. High NPHP1 and NPHP6 mutation rate in patients with Joubert syndrome and nephronophthisis: potential epistatic effect of NPHP6 and AHI1 mutations in patients with NPHP1 mutations. *J. Am. Soc. Nephrol.* **18**, 1566–1575 (2007).
45. Leitch, C. C. et al. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. *Nat. Genet.* **40**, 443–448 (2008).
46. Louie, C. M. et al. AHI1 is required for photoreceptor outer segment development and is a modifier for retinal degeneration in nephronophthisis. *Nat. Genet.* **42**, 175–180 (2010).
47. Alfano, G. et al. EYS is a protein associated with the ciliary axoneme in rods and cones. *PLoS ONE* **11**, e0166397 (2016).
48. Lu, Z. et al. Ablation of EYS in zebrafish causes mislocalisation of outer segment proteins, F-actin disruption and cone-rod dystrophy. *Sci. Rep.* **7**, 46098 (2017).
49. Messchaert, M. et al. Eyes shut homolog is important for the maintenance of photoreceptor morphology and visual function in zebrafish. *PLoS ONE* **13**, e0200789 (2018).
50. Goldberg, A. F., Moritz, O. L. & Williams, D. S. Molecular basis for photoreceptor outer segment architecture. *Prog. Retin. Eye Res.* **55**, 52–81 (2016).
51. Liu, Q., Lyubarsky, A., Skalet, J. H., Pugh, E. N. Jr. & Pierce, E. A. RP1 is required for the correct stacking of outer segment discs. *Invest. Ophthalmol. Vis. Sci.* **44**, 4171–4183 (2003).
52. Genomics_England. The 100,000 Genomes Project Protocol v3. (2017).
53. Priest, J. R. et al. De novo and rare variants at multiple loci support the oligogenic origins of atrioventricular septal heart defects. *PLoS Genet.* **12**, e1005963 (2016).
54. Katsanis, N. The oligogenic properties of Bardet-Biedl syndrome. *Hum. Mol. Genet.* **13**, R65–R71 (2004).
55. Drmanac, R. et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
56. Nishiguchi, K. M. et al. Whole genome sequencing in patients with retinitis pigmentosa reveals pathogenic DNA structural changes and NEK2 as a new disease gene. *Proc. Natl Acad. Sci. USA* **110**, 16139–16144 (2013).
57. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
58. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
59. Freeman, P. J., Hart, R. K., Gretton, L. J., Brookes, A. J. & Dalgleish, R. VariantValidator: accurate validation, mapping, and formatting of sequence variation descriptions. *Hum. Mutat.* **39**, 61–68 (2018).
60. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).
61. Chang, J. et al. Exome-wide analyses identify low-frequency variant in CYP26B1 and additional coding variants associated with esophageal squamous cell carcinoma. *Nat. Genet.* **50**, 338–343 (2018).
62. Do, R. et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102–106 (2015).
63. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Preprint at https://www.biorxiv.org/content/early/2014/05/14/005165 (2014).
64. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
65. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

Acknowledgements

This work was supported by the following agencies. The Swiss National Science Foundation (grant #176097, to C.R.); the Rotterdamse Stichting Blindenbelangen, the Stichting voor Ooglijders and the Nelly Reef fund (to M.I.K. and F.P.M.C.); the Japan Agency for Medical Research and Development (grant #17ek0109213h0001) and the Japan Society for the Promotion of Science (grant #16K11315, to K.M.N.); Research on Measures for Intractable Diseases, Comprehensive Research on Disability Health and Welfare, the Strategic Research Program for Brain Science, and the Initiative on Rare and Undiagnosed Diseases (all to N.Ma.); the Japanese Agency for Medical Research and Development, Grants-in-Aid for Scientific Research (to N.Ma. and N.Mi.); the BioBank Japan Project, by the Japanese Ministry of Education, Culture, Sport, Science and Technology and the Japanese Agency for Medical Research and Development (to M.K.); the PhD Fellowship in Life Sciences (Faculty of Biology and Medicine, University of Lausanne), to M.Q. Most HRD patients were recruited from the Japanese Retinitis Pigmentosa Registry Project (JRPRP). The authors would also like to warmly thank Dr. Matthew Robinson for his expert advice on this work.

Author contributions

K.N., K.C., M.Q., and C.R. designed the research, wrote the manuscript, analyzed all the data, and interpreted the results. H.K.K. performed the analyses of the WGS sample and segregation analyses of the families. N.Mi. screened control individuals for the *RP1* mutations. P.F., A.U.R., and K.N. performed Sanger sequencing analyses. A.P. contributed to the statistical analyses, while M.A., F.M., and K.M.N. analyzed the data. Under the coordination of K.M.N. and T.N., Y.I., S.U., N.F., A.M., Y.W., H.T., K.-H.S., T. I., K.M.N., and T.N. examined the patients and collected biological samples for genetic analysis. Y.K., C.T., N.Ma., and M.K. provided access to databases of control individuals. M.I.K. and F.P.M.C. provided part of the sequencing data. Z.K. supervised the statistical analysis and wrote part of the manuscript. C.R. was the principal investigator of this study. All authors read and commented on the article.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10746-4.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

7.3.2 New pathogenic variants and insights into pathogenic mechanisms in *GRK1*-related Oguchi disease

New pathogenic variants and insights into pathogenic mechanisms in GRK1-related Oguchi disease

James A. Poulter[1,§,*], Molly S. C. Gravett[2,*], Rachel L. Taylor[3], Kaoru Fujinami[4-7], Julie De Zaeytijd[8], James Bellingham[6], Atta Ur Rehman[9], Takaaki Hayashi[10], Mineo Kondo[11], Abdur Rehman[12], Muhammad Ansar[13], Dan Donnelly[14], Carmel Toomes[1], Manir Ali[1], UK Inherited Retinal Disease Consortium, Genomics England Research Consortium, Elfride De Baere[8], Bart P. Leroy[8,15], Nigel P. Davies[16], Robert H. Henderson[17], Andrew R. Webster[5,6], Carlo Rivolta[18-20], Omar A. Mahroo[5,6], Gavin Arno[4-6], Graeme C. Black[3,21], Martin McKibbin[1,22], Sarah A. Harris[23,**], Kamron N. Khan[1,22,**] and Chris F. Inglehearn[1,**].

[1] Division of Molecular Medicine, Leeds Institute of Medical Research, University of Leeds, Leeds, UK.

[2] School of Molecular and Cellular Biology, University of Leeds, UK.

[3] Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicines and Health, University of Manchester, Manchester, UK.

[4] National Institute of Sensory Organs, National Hospital Organization Tokyo Medical Centre, Tokyo, Japan.

[5] Moorfields Eye Hospital, London, UK.

[6] UCL Institute of Ophthalmology, London, UK.

[7] Keio University School of Medicine, Tokyo, Japan.

[8] Ghent University, Ghent, Belgium.

[9] Division of Genetic Medicine, Centre Hospitalier Universitaire Vaudois (CHUV) University of Lausanne, Switzerland.

[10] The Jikei University School of Medicine, Tokyo, Japan.

[11] Mie University Graduate School of Medicine, Mie, Japan.

[12] Department of Genetics, Faculty of Science, Hazara University Mansehra, Pakistan.

[13] Clinical Research Center, Institute of Molecular and Clinical Ophthalmology Basel (IOB), Basel, Switzerland.

[14] School of Biology, University of Leeds, UK.

[15] Children's Hospital of Philadelphia, Philadelphia, PA, USA.

[16] St Thomas's Hospital, London, UK.

[17] Department of Ophthalmology, Great Ormond Street Hospital, London, UK.

[18] Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom.

[19] Clinical Research Center, Institute of Molecular and Clinical Ophthalmology Basel (IOB), Basel, Switzerland.

[20] Department of Ophthalmology, University Hospital Basel, Switzerland.

[21] Manchester Centre for Genomic Medicine, Saint Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, UK.

[22] Leeds Teaching Hospitals NHS Trust, St James' University Hospital, Leeds, UK.

[23] School of Physics, University of Leeds, UK.

[*] These authors contributed equally to this work.

[**]These authors contributed equally to this work.

[§]**Corresponding author**

Dr James Poulter

Division of Molecular Medicine

Leeds Institute of Medical Research

University of Leeds

St James' University Hospital

Beckett Street, Leeds,

LS9 7TF, UK.

E-mail: J.A.Poulter@leeds.ac.uk

Abstract

Purpose: Biallelic mutations in G-Protein coupled receptor kinase 1 (GRK1) cause Oguchi disease, a rare subtype of congenital stationary night blindness (CSNB). The purpose of this study was to identify pathogenic GRK1 variants and use in-depth bioinformatic analyses to evaluate how their impact on protein structure could lead to pathogenicity.

Methods: Patients' genomic DNA was sequenced by whole genome, whole exome or focused exome sequencing. Pathogenic variants, published and novel, were compared to non-disease associated missense variants. The impact of *GRK1* missense variants at the protein level were then predicted using a series of computational tools.

Results: We identified eleven previously unpublished cases with biallelic pathogenic GRK1 variants, including seven novel variants, and reviewed all *GRK1* pathogenic variants. Further structure-based scoring revealed a hotspot for missense variants in the kinase domain. Additionally, to aid future clinical interpretation, we identified the bioinformatics tools best able to differentiate pathogenic from non-pathogenic variants.

Conclusion: We identified new *GRK1* pathogenic variants in Oguchi disease patients and investigated how disease-causing variants may impede protein function, giving new insights into the mechanisms of pathogenicity. All pathogenic GRK1 variants described to date have been collated into a Leiden Open Variation Database (http://dna2.leeds.ac.uk/GRK1_LOVD/genes/GRK1).

Key words

GRK1, Rhodopsin, Oguchi disease, CSNB

Introduction

The first member of the G protein-coupled receptor kinase (GRK) family was discovered when enzymatic activity was observed in rod membranes that phosphorylated rhodopsin in a light-dependent manner (Kuhn & Dreyer, 1972). The enzyme, now known as GRK1 (MIM 180381), was found to be essential for quenching and recycling light-activated rhodopsin. To be recycled, light-activated rhodopsin is phosphorylated multiple times by GRK1, followed by binding of Arrestin-1 to activated-phosphorylated rhodopsin to block further transducin activation by steric exclusion (Krupnick, Gurevich, & Benovic, 1997; Wilden, Hall, & Kuhn, 1986). Failure of this process results in a build-up of activated rhodopsin and a lack of light-sensitive rhodopsin available for detection of light. Importantly, a build-up of activated rhodopsin appears to be well tolerated by photoreceptors, with no obvious cell death or structural consequences in the retina, meaning any future therapies leading to the deactivation of rhodopsin could restore visual acuity. Activated cone opsin is likely to be more reliant on GRK7 than GRK1 phosphorylation, as although both enzymes are expressed in cones, loss of GRK1 function impacts minimally on human photopic vision.

GRK1 is a serine/threonine protein kinase with a central catalytic AGC protein kinase (PK) domain that sits within a regulator of G protein signalling homology (RH) domain (Fig. 1). The PK domain binds ATP and the polypeptide substrate which is subsequently phosphorylated (Arencibia, Pastor-Flores, Bauer, Schulze, & Biondi, 2013). The RH domain is thought to have a key role in receptor binding, with loss of this domain preventing binding and phosphorylation of rhodopsin (He et al., 2017). While the N-terminus encodes a short alpha-helical domain, the C-terminus encodes a lipid-binding region, which is crucial for prenylation-dependent docking of GRK1 into the outer segment membranes of rod photoreceptors (Komolov & Benovic, 2018).

Oguchi disease (MIM 613411), a rare form of congenital stationary night blindness (CSNB), results from biallelic variants in either *SAG* (encoding Arrestin-1) or *GRK1*. *SAG*-mediated disease is most prevalent in Japanese patients, whilst pathogenic variants in *GRK1* are more common in South East Asians. Typically, Oguchi disease is characterised clinically by the Mizuo-Nakamura phenomenon – the presence of a golden-yellow colouration to the retina that disappears in the dark-adapted state and reappears shortly after light exposure (Miyake, Horiguchi, Suzuki, Kondo, & Tanikawa, 1996). Electrophysiologically there is normal cone function, delayed rod dark adaptation and marked rod desensitisation to a bright flash. Most

reported cases have this distinctive phenotype and no significant variation in disease expression has been reported, with the exception of a few rare cases (Hayashi, Gekka, Takeuchi, Goto-Omoto, & Kitahara, 2007; Nishiguchi et al., 2019). Distinguishing between Oguchi disease and other forms of CSNB is important as they may have different prognoses - some patients with Oguchi disease report very slowly progressive visual dysfunction, whereas those with classical CSNB do not. Minimally progressive retinal degeneration has been observed in *SAG*-associated Oguchi disease, and for one patient, this resulted in their disease being re-classified as retinitis pigmentosa 26 years after being diagnosed with a stationary rod dysfunction syndrome (Oguchi disease)(Nishiguchi et al., 2019). As most of our knowledge of Oguchi disease stems from our understanding of loss of SAG function further research is required in order to better understand *GRK1*-related Oguchi disease.

To date, thirteen pathogenic variants in *GRK1* have been implicated in Oguchi disease, eight of which are predicted to be null variants (Azam et al., 2009; Cideciyan et al., 1998; Godara et al., 2012; Hayashi et al., 2007; Jespersgaard et al., 2019; Li et al., 2017; Mucciolo et al., 2018; Oishi et al., 2007; Skorczyk-Werner, Kociecki, Wawrocka, Wicher, & Krawczyniski, 2015; Teke, Citirik, Kabacam, Demircan, & Alikasifoglu, 2016; Yamamoto, Sippel, Berson, & Dryja, 1997; Zhang et al., 2005). The molecular mechanisms by which these contribute to disease are poorly understood. Here we present seven further pathogenic variants in *GRK1*, and review new and known variants causing Oguchi disease. Using *in-silico* techniques we compare known disease-associated, with likely non-pathogenic, *GRK1* missense variants to identify features that could be contributing to the phenotype. Together this will help define disease mechanisms and assist in predicting likelihood of pathogenicity for variants identified in *GRK1*.

Methods

Editorial Policies and Ethical Considerations

Ethical approval for this study was obtained from the Yorkshire and The Humber – Leeds East Research Ethics Committee (REC reference: 17/YH/0032). The study adhered to the tenets of the Declaration of Helsinki. Blood samples were taken with informed consent from each participant, or with parental informed consent on behalf of children.

Clinical assessment

Study participants were ascertained from St James' University Hospital, Leeds, England; Manchester Centre for Genomic Medicine, Manchester, England; Moorfields Eye Hospital, London, England; Department of Ophthalmology, Ghent University Hospital, Ghent, Belgium; The Jikei University School of Medicine, Tokyo, Japan; Mie University Graduate School of Medicine, Mie, Japan; and National Institute of Sensory Organs, National Hospital Organization Tokyo Medical Centre, Tokyo Japan; Individuals were recruited following clinical examination by an experienced ophthalmologist, with the exception of Families 10 and 11 who were ascertained by local doctors in Pakistan.

Sequencing

Exome and clinical exome sequencing was performed on 3μg of genomic DNA using the Agilent SureSelectXT Human All Exon (V6) and Focussed Exome respectively (Agilent Technologies, CA, USA), according to the manufacturer's instructions. Captured libraries were pooled and sequenced on an Illumina HiSeq3000 sequencer (Illumina, CA, USA). The resulting fastq files were aligned to the Human genome assembly GRCh37 using Burrows-Wheeler Aligner (BWA) and reads processed using samtools, Picard and the Genome Analysis Toolkit. Variants within the captured regions were called and annotated using variant effect predictor (NCBI) and further filtered for those within genes known to cause retinal disease. This analysis pipeline is described in more detail elsewhere (Panagiotou et al., 2017). Copy number variation (CNV) analysis was performed using ExomeDepth, comparing patient BAM files with BAM files from unrelated individuals in the same sequencing run. All novel variants were confirmed by Sanger sequencing using standard methods.

The affected individual and unaffected parents of Family 1 underwent Genome sequencing as part of the 100,000 Genome Project, as previously detailed by Taylor *et al* (Taylor et al., 2017).

Identifying non-pathogenic variants

To identify protein-level characteristics influencing a disease phenotype we compared the 9 pathogenic missense variants with a group of 5 assumed non-pathogenic missense variants. For the non-pathogenic group, we identified all homozygous *GRK1* missense variants in gnomAD and only included those with a MAF >0.00035, the level above which variants are calculated to be too common to cause Oguchi disease (Whiffin et al., 2017). The five missense variants in gnomAD with frequencies above this level, p.Leu54Phe, p.Thr97Met, p,Ile241Thr, p.Glu464Gln and p.Ser536Leu, also all occur at least once as homozygotes in gnomAD. These are therefore assumed to be non-disease associated polymorphisms (further details in Supplementary Table 1), which for the purposes of this paper we will call non-pathogenic variants.

Structural analysis

A homology model of human wild-type GRK1 based on the bovine structure (PDB: 4PNI) was produced using SwissModel (Waterhouse et al., 2018). The locations of the RH and PK domains, ATP binding site, polypeptide substrate binding site and activation loop were identified using InterPro (Mitchell et al., 2019). The structure was visualised, and regions of interest labelled using PyMOL (Schrodinger, 2015). Residue depth and connectivity within the structure was determined using Site Directed Mutator (SDM) (Pandurangan, Ochoa-Montano, Ascher, & Blundell, 2017). To further explain variant impact, changes in residue physical characteristics; molecular weight (Mw), isoelectric point (pI) and hydrophobicity (Laskowski, Stephenson, Sillitoe, Orengo, & Thornton, 2020) were reviewed.

Scoring missense variants

A number of bioinformatics tools were reviewed to identify which approach best classified the variants. Scoring methods included: SDM (Pandurangan et al., 2017), VarSite (Laskowski et al., 2020), SIFT (Sim et al., 2012), Polyphen-2 (Adzhubei et al., 2010), CADD (Kircher et al., 2014), Rhapsody (Ponzoni, Oltvai, & Bahar, 2019) and Consurf (Ashkenazy et al., 2016). All statistical analyses were performed in GraphPad Prism 8.0.2 (for macOS, GraphPad Software, San Diego, California USA, www.graphpad.com) using Welch's T-test.

Results

Identification of novel *GRK1* pathogenic variants

Screening of inherited retinal disease patients by an international consortium of retinal genetics laboratories identified eleven cases with Oguchi disease harbouring eight different pathogenic variants in *GRK1*, of which seven are previously unreported (Table 1). All cases are of South Asian ethnicity except for Family 2 who are East Asian (Japanese) and Family 5 who are Eastern European (Polish). All variants were confirmed and segregated in available family members. Each patient had a history of non/minimally-progressive night blindness, and retinal examination revealed the Mizuo-Nakamura phenomenon, consistent with a clinical diagnosis of Oguchi disease.

A literature search revealed a further thirteen published Oguchi disease causing *GRK1* variants from twelve studies (Azam et al., 2009; Cideciyan et al., 1998; Godara et al., 2012; Hayashi et al., 2007; Jespersgaard et al., 2019; Li et al., 2017; Mucciolo et al., 2018; Oishi et al., 2007; Skorczyk-Werner et al., 2015; Teke et al., 2016; Yamamoto et al., 1997; Zhang et al., 2005). The nomenclature for these variants was updated based on the current HGVS nomenclature guidelines and the GRCh37 version of the Human Genome, and together with the new variants reported here, these have been included in a Leiden Open Variation Database (LOVD) of *GRK1* variants (http://dna2.leeds.ac.uk/GRK1_LOVD/genes/GRK1).

Pathogenic variant distribution and domain structure

The twenty Oguchi disease causing *GRK1* variants identified to date include missense (n=9), frameshift (n=6) and nonsense (n=4) variants (Table 1) as well as a large deletion encompassing exon 5 (Fig. 1A). Each variant has been identified in only a single family with the exception of p.Val380Asp (2 families), p.Glu48ProfsTer82 (2 families), the deletion of exon 5 (p.(Phe358GlyfsTer18)) (3 families) and the C-terminal frameshift variant p.(Asp537ValfsTer6) (6 families), with the latter variant having the highest allele frequency in gnomAD (89/172,958). Of the 89 alleles in gnomAD, 42 are in European (non-Finnish) and 35 in South Asian populations, which is consistent with this mutation only being identified in cases from Eastern Europe and South Asia.

Fifteen of the variants lie within the protein kinase domain, three within the RH domain and one at each of the N- and C-termini (Fig. 1B). While the frameshift and nonsense variants are present throughout the protein, all except one of the missense variants (n=8 out of 9) are

present within the protein kinase domain, and five cluster within a 30 amino-acid region between Glu362 and Pro391. These are likely to impair phosphorylation of light-activated rhodopsin.

Three variants located after this cluster introduce frameshifts in the final exon, giving rise to transcripts which would not be expected to undergo nonsense-mediated decay (Thermann et al., 1998). The extreme C-terminus of GRK1, which these three variants would disrupt, is prenylated to facilitate anchoring of GRK1 in the membranes of the rod photoreceptor outer segments (Pitcher, Freedman, & Lefkowitz, 1998), which maximizes the likelihood of interaction with activated rhodopsin. Lack of this domain is therefore likely to reduce the level of anchored GRK1, which will in turn reduce the quantity of rhodopsin being bound and phosphorylated.

Pathogenic missense variants affect key regions in *GRK1* active site

To further understand the mechanism by which the missense variants affected GRK1 function, we performed structure-based scoring on the SDM server using our homology model (Fig. 2A-C). Although SDM's primary function is predicting thermal stability, the most interesting results came from the scoring of wildtype (WT) residue depth and occluded surface packing (OSP), with both scores showing a significant difference between pathogenic and non-pathogenic variants ($p<0.01$ and $p<0.05$ respectively). This demonstrated that pathogenic variants were located deeper within the 3D structure of *GRK1* than non-pathogenic variants (residue depth = 3.5-11.8 and 3-6.2 respectively) (Fig. 2A, Supplementary Table 2) and were more connected/densely packed (OSP = 0.3-0.57 and 0.04-0.48 respectively) (Fig. 2B, Supplementary Table 2). Exceptions to these rules were seen in p.Gly199Arg (pathogenic) which is located at the surface of the protein and p.Ile241Thr (non-pathogenic) which is located deeper in GRK1. Gly199 is positioned within the conserved glycine rich loop of the PK domain and is part of the ATP binding site (Fig. 3). Mutation to a bulky positively charged arginine (Supplementary Table 2) is likely to disrupt both the shape and charge of the ATP binding pocket (Fig. 3). Although Ile241 is buried in the binding pocket of the PK domain (Fig. 3), there is little change in size and charge (Supplementary Table 2) which could explain why this variant is non-pathogenic.

Further investigation of mutation location in the protein structure demonstrated pathogenic variants were located proximal to key sites for protein function, including the substrate and

ATP binding domains, while non-pathogenic variants were not (Fig. 3A). Pathogenic variants were predicted to be likely to disrupt the shape and charge of the active site either directly as in the case of p.Gly199Arg, or allosterically through introduction of proline kinks into alpha helices (e.g. p.Leu308Pro & p.Ala377Pro), intramolecular charged groups (classically exposed) (e.g. p.Val380Asp & p.Pro391His), and bulky side groups within densely packed regions (e.g. p.Val380Phe) close to key residues (Fig. 3B, Supplementary Table 2). In the case of p.Arg438Cys and p.Glu362Lys, the equivalent arginine and glutamic acid residues are conserved and form a salt bridge in the bovine protein (PDB: 4PNI). It is therefore likely that substitution of Glu362 and Arg438 changes the charge of these residues in the human protein and is disrupting this salt bridge interaction (Supplementary Figure 2). Meanwhile non-pathogenic variants were often subtler changes, and not local to key residues (Fig. 3B, Supplementary Table 2). We hypothesise pathogenic variants within the protein kinase domain are inhibiting protein function through disruption of the shape and dynamics of this key region, while non-pathogenic variants prove less disruptive to shape as they are not internal and/or have subtler changes in residue characteristics. This is supported by Consurf's prediction that most of the pathogenic variants occur at loci important for protein structure or function (Fig. 2I, Supplementary Table 2).

Only one pathogenic missense variant, p.Leu157Pro, lies outside of the kinase domain, instead being present within the RH domain of GRK1, the domain primarily responsible for rhodopsin binding (He et al., 2017). The introduction of a proline residue within an α-helix is likely to disrupt the secondary structure and introduce a characteristic "kink" due to its rigidity and constrained phi angle. This could potentially impact on the ability of Rhodopsin to correctly bind GRK1 in order for phosphorylation to occur.

To determine which bioinformatic tools were most successful in differentiating between pathogenic and non-pathogenic variants, we scored missense variants using: VarSite, SDM, PolyPhen-2, SIFT, CADD, Rhapsody and Consurf. Predictions were compared between pathogenic and non-pathogenic variant sets, with significance assessed by Welch's T-test (Fig 2, Table 1, Supplementary Table 2). VarSite disease propensity, WT OSP, WT residue depth, PolyPhen-2, CADD, Rhapsody and Consurf, were able to distinguish between pathogenic and non-pathogenic variants with statistical significance. Consurf's structure/function prediction and Rhapsody, which both take the variant's location in the protein 3D structure into

consideration, were able to differentiate between pathogenic and non-pathogenic most successfully ($p < 0.001$) (Fig. 2H&I, Supplementary Table 2).

Rhapsody has the additional feature of being able to predict the impact of substituting any residue in a protein with all of the 19 other amino acids (*in silico* site directed mutagenesis), making it extremely valuable diagnostically, as it has the potential to inform future work by identifying residue changes which are likely to be poorly tolerated. We performed protein-wide site directed mutagenesis and reviewed the top scoring variants (Rhapsody score ≥ 0.896). This analysis revealed that the top scoring variants were present at just 22 loci across the protein, thus identifying a number of key residues with the potential to cause Oguchi disease, if mutated. Of the 22 amino acid residues that Rhapsody predicts could harbour the most damaging variants, 20 are located in the protein kinase domain (Fig. 4).

Discussion

Using a combined genetics and structural biology approach we have identified individuals with Oguchi disease due to biallelic *GRK1* variants and inferred the likely functional consequences of these and other published variants. We identified seven new pathogenic variants, taking the total number of *GRK1* variants associated with Oguchi disease to twenty. While eleven result in premature termination codons, nine result in single amino acid substitutions.

It is likely that transcripts encoding nonsense and frameshift variants, except those in the last exon, will be subject to nonsense mediated decay (NMD)(Thermann et al., 1998). This has not been confirmed in patient cells, but even if a truncated GRK1 lacking the residues encoded by exon 5 were produced, function has been shown to be significantly reduced, resulting in protein that was not able to phosphorylate Rhodopsin (Cideciyan et al., 1998). This implies that lack of functional GRK1 protein is one mechanism by which Oguchi disease can arise.

eight of the nine observed missense variants lie within the protein kinase domain, which is critical for GRK1 function. The kinase domain is the largest within GRK1, and variants in this region are likely to have a negative impact on phosphorylation of light-activated rhodopsin. Indeed, a previous study has shown the p.Val380Asp variant has no kinase activity for Rhodopsin compared to wild-type GRK1 (Khani, Nielsen, & Vogt, 1998). Five missense variants appear to be clustered within a 30 amino-acid region between Glu362 and Pro391, which may imply that amino-acid substitutions in this region are more likely to disrupt protein function than those located elsewhere. These variants were predicted to affect the structure of GRK1 and therefore kinase activity. As the structural analysis relies on a homology model of human GRK1, it is not always feasible to quantify atomic level changes (e.g. individual H-bond or salt bridge interactions within non-conserved residues). However, it is possible to draw general conclusions about the surrounding residue location and impact on secondary and tertiary structure. Therefore, loss of kinase activity through disruption of its shape and dynamics is a second mechanism by which Oguchi disease can occur.

Three variants have been identified in the last exon of *GRK1*, all of which result in frameshift transcripts that potentially escape nonsense mediated decay, leading to a GRK1 protein with a disrupted C-terminus. This region plays a crucial role in embedding GRK1 within the

photoreceptor outer segment membrane; GRK1 localises directly to the outer segment membranes due to a short prenylation sequence within the C-terminus that interacts with the prenyl-binding protein, delta (PrBP/δ), which facilitates attachment to the membranes (Huang, Orban, Jastrzebska, Palczewski, & Tesmer, 2011; Roosing, Collin, den Hollander, Cremers, & Siemiatkowska, 2014). This cellular localisation is essential for GRK1 function as it brings active rhodopsin into contact with docked GRK1 protein through a two-dimensional search rather than three-dimensional diffusion, which is essential for ultra-rapid signal termination (Sato, Chuprun, Schwartz, & Koch, 2015). Lack of this sequence is likely to cause failure of prenylation, leading to abolished binding of PrBP/δ and incorrect localisation of GRK1, a third mechanism by which disease can arise.

Our study therefore highlights three potential molecular mechanisms by which GRK1 variants lead to disease, each of which ultimately reduce or abolish the phosphorylation of light-activated rhodopsin. These are a) lack of GRK1 protein due to frame shift or nonsense variants in all exons but the last, resulting in NMD; b) missense variants leading to inability of GRK1 to phosphorylate rhodopsin due to a dysfunctional kinase domain or failure to bind rhodopsin, ATP or Mg2+; or c) a frameshift in the final exon, which would be expected to escape nonsense mediated decay and produce a protein retaining the ability to bind and phosphorylate light-activated rhodopsin, but which cannot localise to the outer segment membranes. We would therefore predict that, while variants in the first category lead to no protein being produced, those in the second produce proteins that correctly localise within the outer segments but have inhibited kinase activity, and those in the third produce a GRK1 protein likely to maintain kinase activity but which fails to localise and so rarely comes into contact with activated rhodopsin. Interestingly, the kinase activity of the p.(Asp537ValfsTer6) variant was found to be significantly reduced in transfected cos7 cells, suggesting c-terminal mutations may affect kinase activity as well as protein localisation. However, this lack of activity may be explained by difficulties in expressing the mutation containing GRK1 protein rather than on protein function (Khani et al., 1998).

Our comparison of GRK1 pathogenic variants to non-pathogenic variants identified from gnomAD facilitated an assessment of pathogenicity scores, which could inform future variant interpretation. We found that VarSite disease propensity, Polyphen2, CADD, Rhapsody and Consurf were able to significantly differentiate between pathogenic and non-pathogenic variants, as were biophysical scores of residue depth and occluded surface packing. Rhapsody

was the best of all the tools at differentiating between pathogenic and non-pathogenic variants. Rhapsody incorporates dynamics information from elastic network models of protein structure, as well as PolyPhen-2 and EVMutation (conservation) scores using a machine learning approach. This broad range of sequence-, structure- and dynamics-based information could account for its success since the impact of a variant can arise from a multitude of different physio-chemical effects. Despite producing the best separation (Rhapsody) and describing features (SDM), structure-based tools are limited by the availability of the protein's structure or a homologous structure. Although homology models are not always accurate for atomic level information, coarser detail such as whether residues are buried is likely to be reliable. Similarly, although Consurf can be used without a pdb structure it is dependent on homologous proteins having a structure. In this study we found using Consurf without a pdb structure was sufficient to predict pathogenicity based on whether a residue is likely to be buried or exposed, relating to its location in homologous proteins with known structures. Ultimately, these scores can only be used as a guide, and variant pathogenicity should be confirmed by segregation analysis and functional assays wherever possible, as rare non-pathogenic variants can also score highly.

Interrogation of gnomAD revealed a further three rare variants that were homozygous in at least one individual and could not be excluded as a cause of Oguchi disease based on allele frequency alone (Supplementary Table 3). We therefore calculated the same set of pathogenic prediction scores to assess the likely pathogenicity of these variants of unknown significance. The p.Met185Val variant, although rare in gnomAD, produced scores that were comparable with the non-pathogenic variants for each tool, suggesting it does not contribute to disease. However, both p.Ala353Ser and p.Ala387Val, produced scores that were compatible with pathogenicity, suggesting that the 2 individuals in gnomAD that are homozygous for these variants may have Oguchi disease. gnomAD contains exome and genome sequence data from over 130,000 unrelated individuals, with only those individuals affected by a severe paediatric disorder being removed. It is therefore possible that individuals with CSNB or Oguchi disease are present in the gnomAD database which could explain the presence of two homozygous, likely pathogenic variants.

In summary, we present seven novel variants in *GRK1* as a cause of Oguchi disease and perform an in-depth analysis of all *GRK1* variants in order to understand their contribution to disease. We describe three mechanisms likely to account for all *GRK1*-related disease, each of

which ultimately result in a failure to phosphorylate light-activated rhodopsin. We assessed the ability of different *in silico* pathogenicity scores to differentiate pathogenic and non-pathogenic variants and found that Rhapsody out-performed all tools tested for this dataset. Finally, we have created an LOVD database into which all the known pathogenic variants and several common non-pathogenic variants described herein have been entered, in order to guide future variant interpretation in *GRK1*-associated disease.

Accession Numbers

The variants reported in this paper have been submitted to the ClinVar database at the National Centre for Biotechnology Information (submission ref. SUB6701858).

Supplemental Data

Supplemental Data includes two figures, three tables and the 3D homology model of Human GRK1 based upon the bovine GRK1 crystal structure (PDB: 4PNI).

Acknowledgements

The authors would like to thank all the families for agreeing to participate in this study, and Roman A. Laskowski and Arun Prasad Pandurangan for assistance with VarSite and SDM respectively, and Ivet Bahar, Luca Ponzoni and the rest of the Bahar group for their help with Rhapsody. This study was supported by the European Retinal Disease Consortium (ERDC), the European Reference Network for Rare Eye Disease (ERN-EYE) and the Japan Eye Genetics Consortium (https://www.jegc.org/).

This study was funded by RP Fighting Blindness and Fight for Sight as part of the UK Inherited Retinal Disease Consortium (RP Genome Project GR586) and the UK National Health Service (NHS) Foundation Trust, the National Institute for Health Research Biomedical Research Centre (NIHR-BRC) at Moorfields Eye Hospital and the NIHR-BRC at Great Ormond Street Institute of Child Health, Moorfields Eye Charity and the Great Britain Sasakawa Foundation. MG is a PhD student on the Wellcome Trust 4-year PhD student in The Astbury Centre funded by The University of Leeds. GA is supported by a Fight For Sight (UK) Early Career Investigator Award. Sarah Harris acknowledges the EPSRC for their support (EP/S0306971/1). This research was made possible through access to the data and findings generated by the 100,000 Genomes Project managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

Members of the UK Inherited Retinal Disease Consortium include: Graeme C. Black, DPhil, FRCOphth (study chair); Georgina Hall, MSc; Stuart Ingram, BSc; Rachel L. Taylor, PhD; Panagiotis Sergouniotis, PhD; Andrew R. Webster, MD(Res), FRCOphth; Alison J. Hardcastle, PhD; Michel Michaelides, MD(Res), FRCOphth; Nikolas Pontikos, PhD; Michael Cheetham, PhD; Gavin Arno, PhD; Alessia Fiorentino, PhD; Chris F. Inglehearn, PhD; Carmel Toomes, PhD; Manir Ali, PhD; Martin McKibbin, FRCOphth; James A. Poulter, PhD; Kamron N. Khan, PhD, FRCOphth; Claire E.L. Smith, PhD; Susan Downes, MD, FRCOphth; Jing Yu, PhD and Veronica van Heyningen, PhD.

Members of the Genomics England Research Consortium include: Ambrose J. C., Arumugam P., Baple E. L., Bleda M., Boardman-Pretty F., Boissiere J. M., Boustred C. R., Brittain H., Caulfield M. J., Chan G. C., Craig C. E. H., Daugherty L. C., de Burca A., Devereau, A., Elgar G., Foulger R. E., Fowler T., Furió-Tarí P., Hackett J. M., Halai D., Hamblin A., Henderson S., Holman J. E., Hubbard T. J. P. , Ibáñez K., Jackson R., Jones L. J., Kasperaviciute D., Kayikci M., Lahnstein L., Lawson K., Leigh S. E. A., Leong I. U. S., Lopez F. J., Maleady-Crowe F., Mason J., McDonagh E. M., Moutsianas L., Mueller M., Murugaesu N., Need A. C., Odhams C. A., Patch C., Perez-Gil D., Polychronopoulos D., Pullinger J., Rahim T., Rendon A., Riesgo-Ferreiro P., Rogers T., Ryten M., Savage K., Sawant K., Scott R. H., Siddiq A., Sieghart A., Smedley D., Smith K. R., Sosinsky A., Spooner W., Stevens H. E., Stuckey A., Sultana R., Thomas E. R. A., Thompson S. R., Tregidgo C., Tucci A., Walsh E., Watters, S. A., Welland M. J., Williams E., Witkowska K., Wood S. M., Zarowiecki M.

Legends to Figures

Figure 1 – Distribution of pathogenic variants within GRK1. All variants shown are annotated according to human genome GRCh37/hg19, using GRK1 gene and protein accession numbers NM_002929 and NP_002920 respectively. Loss of function variants are shown above the gene/protein and missense variants are given below. (A) Genomic organisation of *GRK1* showing the location of all novel and published Oguchi disease causing variants. (B) Domain structure of GRK1 showing the location of all GRK1 variants. The start and end amino-acid positions of each domain are based on Lodowski *et al*. (2006)(Lodowski, Tesmer, Benovic, & Tesmer, 2006). NT = N-Terminus, RH = regulator of G-protein signalling homology.

Figure 2 – Bioinformatic prediction scores of pathogenic and non-pathogenic GRK1 variants. Comparison of predicted pathogenicity scores for all novel or published pathogenic missense GRK1 variants, with likely non-pathogenic missense variants identified in gnomAD. The whiskers in the box plots show the minimum and maximum values. Statistical significance was calculated using Welch's T-test in GraphPad prism 8.0.2. ns = $p \geq 0.05$, * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$. (A) Wildtype (WT) residue depth from Site Directed Mutator (SDM), (B) WT occluded surface packing (OSP) from SDM, (C) Thermostability score SDM, (D) Disease propensity scores as predicted in VarSite, (E) SIFT scores, (F) PolyPhen-2 scores, (G) CADD (v1.3) scores, (H) Rhapsody scores and (I) Consurf predicted roles (1 = structural/functional role predicted, 0 = no role predicted).

Figure 3 – Pathogenic and non-pathogenic variants location in homology model of GRK1. Homology model of human GRK1 based upon bovine GRK1 structure (PDB: 4PNI). Wildtype residues of pathogenic variants are labelled as red spheres (Leu157, Gly199, Leu308, Glu362, Ala377, Val380, Pro391 and Arg438) and non-pathogenic as blue spheres (Leu54, Thr97, Ile241, Glu464 and Ser536). The protein kinase domain is coloured in purple, RH domain in orange, activation loop in green, ATP binding residues are represented as yellow sticks and polypeptide substrate binding residues as cyan sticks. (a) A visualisation of variant location in the overall structure. The smaller representations to the right, with the ATP and polypeptide binding surface displayed demonstrate how intramolecular residue changes may deform the binding site surface. (b) A focus on residue location and local interactions, displaying only features within 15Å of the residue of interest.

Figure 4 – Rhapsody top scoring substitutions mapped onto the homology model of GRK1. Homology model of human GRK1 based upon bovine GRK1 structure (PDB: 4PNI). Wildtype residues of top-scoring (($\geq$0.896) rhapsody variants are labelled as spheres, coloured and numbered based on residue number. The colours and labels within the bar chart of top-scoring Rhapsody variants mimic this format. Locations with >4 variants in the top 93 scores include: G68, L253, L308, Y315, L318, Y383, L430 and G440.

References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., . . . Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, *7*(4), 248-249. doi:10.1038/nmeth0410-248

Arencibia, J. M., Pastor-Flores, D., Bauer, A. F., Schulze, J. O., & Biondi, R. M. (2013). AGC protein kinases: from structural mechanism of regulation to allosteric drug development for the treatment of human diseases. *Biochim Biophys Acta*, *1834*(7), 1302-1321. doi:10.1016/j.bbapap.2013.03.010

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*, *44*(W1), W344-350. doi:10.1093/nar/gkw408

Azam, M., Collin, R. W., Khan, M. I., Shah, S. T., Qureshi, N., Ajmal, M., . . . Cremers, F. P. (2009). A novel mutation in GRK1 causes Oguchi disease in a consanguineous Pakistani family. *Mol Vis*, *15*, 1788-1793. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19753316

Cideciyan, A. V., Zhao, X., Nielsen, L., Khani, S. C., Jacobson, S. G., & Palczewski, K. (1998). Null mutation in the rhodopsin kinase gene slows recovery kinetics of rod and cone phototransduction in man. *Proc Natl Acad Sci U S A*, *95*(1), 328-333. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9419375

Godara, P., Cooper, R. F., Sergouniotis, P. I., Diederichs, M. A., Streb, M. R., Genead, M. A., . . . Carroll, J. (2012). Assessing retinal structure in complete congenital stationary night blindness and Oguchi disease. *Am J Ophthalmol*, *154*(6), 987-1001 e1001. doi:10.1016/j.ajo.2012.06.003

Hayashi, T., Gekka, T., Takeuchi, T., Goto-Omoto, S., & Kitahara, K. (2007). A novel homozygous GRK1 mutation (P391H) in 2 siblings with Oguchi disease with markedly reduced cone responses. *Ophthalmology*, *114*(1), 134-141. doi:10.1016/j.ophtha.2006.05.069

He, Y., Gao, X., Goswami, D., Hou, L., Pal, K., Yin, Y., . . . Xu, H. E. (2017). Molecular assembly of rhodopsin with G protein-coupled receptor kinases. *Cell Res*, *27*(6), 728-747. doi:10.1038/cr.2017.72

Huang, C. C., Orban, T., Jastrzebska, B., Palczewski, K., & Tesmer, J. J. (2011). Activation of G protein-coupled receptor kinase 1 involves interactions between its N-terminal region and its kinase domain. *Biochemistry*, *50*(11), 1940-1949. doi:10.1021/bi101606e

Jespersgaard, C., Fang, M., Bertelsen, M., Dang, X., Jensen, H., Chen, Y., . . . Gronskov, K. (2019). Molecular genetic analysis using targeted NGS analysis of 677 individuals with retinal dystrophy. *Sci Rep*, *9*(1), 1219. doi:10.1038/s41598-018-38007-2

Khani, S. C., Nielsen, L., & Vogt, T. M. (1998). Biochemical evidence for pathogenicity of rhodopsin kinase mutations correlated with the oguchi form of congenital stationary night blindness. *Proc Natl Acad Sci U S A*, *95*(6), 2824-2827. doi:10.1073/pnas.95.6.2824

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, *46*(3), 310-315. doi:10.1038/ng.2892

Komolov, K. E., & Benovic, J. L. (2018). G protein-coupled receptor kinases: Past, present and future. *Cell Signal*, *41*, 17-24. doi:10.1016/j.cellsig.2017.07.004

Krupnick, J. G., Gurevich, V. V., & Benovic, J. L. (1997). Mechanism of quenching of phototransduction. Binding competition between arrestin and transducin for

phosphorhodopsin. *J Biol Chem*, *272*(29), 18125-18131. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9218446

Kuhn, H., & Dreyer, W. J. (1972). Light dependent phosphorylation of rhodopsin by ATP. *FEBS Lett*, *20*(1), 1-6. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11946367

Laskowski, R. A., Stephenson, J. D., Sillitoe, I., Orengo, C. A., & Thornton, J. M. (2020). VarSite: Disease variants and protein structure. *Protein Sci*, *29*(1), 111-119. doi:10.1002/pro.3746

Li, L., Chen, Y., Jiao, X., Jin, C., Jiang, D., Tanwar, M., . . . Hejtmancik, J. F. (2017). Homozygosity Mapping and Genetic Analysis of Autosomal Recessive Retinal Dystrophies in 144 Consanguineous Pakistani Families. *Invest Ophthalmol Vis Sci*, *58*(4), 2218-2238. doi:10.1167/iovs.17-21424

Lodowski, D. T., Tesmer, V. M., Benovic, J. L., & Tesmer, J. J. (2006). The structure of G protein-coupled receptor kinase (GRK)-6 defines a second lineage of GRKs. *J Biol Chem*, *281*(24), 16785-16793. doi:10.1074/jbc.M601327200

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., . . . Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*, *47*(D1), D351-D360. doi:10.1093/nar/gky1100

Miyake, Y., Horiguchi, M., Suzuki, S., Kondo, M., & Tanikawa, A. (1996). Electrophysiological findings in patients with Oguchi's disease. *Jpn J Ophthalmol*, *40*(4), 511-519. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9130055

Mucciolo, D. P., Sodi, A., Murro, V., Passerini, I., Palchetti, S., Pelo, E., . . . Rizzo, S. (2018). A novel GRK1 mutation in an Italian patient with Oguchi disease. *Ophthalmic Genet*, *39*(1), 137-138. doi:10.1080/13816810.2017.1323341

Nishiguchi, K. M., Ikeda, Y., Fujita, K., Kunikata, H., Akiho, M., Hashimoto, K., . . . Abe, T. (2019). Phenotypic Features of Oguchi Disease and Retinitis Pigmentosa in Patients with S-Antigen Mutations: A Long-Term Follow-up Study. *Ophthalmology*, *126*(11), 1557-1566. doi:10.1016/j.ophtha.2019.05.027

Oishi, A., Akimoto, M., Kawagoe, N., Mandai, M., Takahashi, M., & Yoshimura, N. (2007). Novel mutations in the GRK1 gene in Japanese patients with Oguchi disease. *Am J Ophthalmol*, *144*(3), 475-477. doi:10.1016/j.ajo.2007.03.025

Panagiotou, E. S., Sanjurjo Soriano, C., Poulter, J. A., Lord, E. C., Dzulova, D., Kondo, H., . . . Toomes, C. (2017). Defects in the Cell Signaling Mediator beta-Catenin Cause the Retinal Vascular Condition FEVR. *Am J Hum Genet*, *100*(6), 960-968. doi:10.1016/j.ajhg.2017.05.001

Pandurangan, A. P., Ochoa-Montano, B., Ascher, D. B., & Blundell, T. L. (2017). SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res*, *45*(W1), W229-W235. doi:10.1093/nar/gkx439

Pitcher, J. A., Freedman, N. J., & Lefkowitz, R. J. (1998). G protein-coupled receptor kinases. *Annu Rev Biochem*, *67*, 653-692. doi:10.1146/annurev.biochem.67.1.653

Ponzoni, L., Oltvai, Z. N., & Bahar, I. (2019). Rhapsody: Pathogenicity prediction of human missense variants based on protein sequence, structure and dynamics. *bioRxiv*. doi:https://doi.org/10.1101/737429

Roosing, S., Collin, R. W., den Hollander, A. I., Cremers, F. P., & Siemiatkowska, A. M. (2014). Prenylation defects in inherited retinal diseases. *J Med Genet*, *51*(3), 143-151. doi:10.1136/jmedgenet-2013-102138

Sato, P. Y., Chuprun, J. K., Schwartz, M., & Koch, W. J. (2015). The evolving impact of g protein-coupled receptor kinases in cardiac health and disease. *Physiol Rev*, *95*(2), 377-404. doi:10.1152/physrev.00015.2014

Schrodinger, LLC. (2015). *The PyMOL Molecular Graphics System, Version 1.8.*

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*, *40*(Web Server issue), W452-457. doi:10.1093/nar/gks539

Skorczyk-Werner, A., Kociecki, J., Wawrocka, A., Wicher, K., & Krawczyniski, M. R. (2015). The first case of Oguchi disease, type 2 in a Polish patient with confirmed GRK1 gene mutation. *Klin Oczna*, *117*(1), 27-30. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26349155

Taylor, R. L., Arno, G., Poulter, J. A., Khan, K. N., Morarji, J., Hull, S., . . . the, G. P. (2017). Association of Steroid 5alpha-Reductase Type 3 Congenital Disorder of Glycosylation With Early-Onset Retinal Dystrophy. *JAMA Ophthalmol*, *135*(4), 339-347. doi:10.1001/jamaophthalmol.2017.0046

Teke, M. Y., Citirik, M., Kabacam, S., Demircan, S., & Alikasifoglu, M. (2016). A novel missense mutation of the GRK1 gene in Oguchi disease. *Mol Med Rep*, *14*(4), 3129-3133. doi:10.3892/mmr.2016.5620

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., . . . Kulozik, A. E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J*, *17*(12), 3484-3494. doi:10.1093/emboj/17.12.3484

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, *46*(W1), W296-W303. doi:10.1093/nar/gky427

Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A. H., Karczewski, K., Ing, A. Y., . . . Ware, J. S. (2017). Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med*, *19*(10), 1151-1158. doi:10.1038/gim.2017.26

Wilden, U., Hall, S. W., & Kuhn, H. (1986). Phosphodiesterase activation by photoexcited rhodopsin is quenched when rhodopsin is phosphorylated and binds the intrinsic 48-kDa protein of rod outer segments. *Proc Natl Acad Sci U S A*, *83*(5), 1174-1178. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3006038

Yamamoto, S., Sippel, K. C., Berson, E. L., & Dryja, T. P. (1997). Defects in the rhodopsin kinase gene in the Oguchi form of stationary night blindness. *Nat Genet*, *15*(2), 175-178. doi:10.1038/ng0297-175

Zhang, Q., Zulfiqar, F., Riazuddin, S. A., Xiao, X., Yasmeen, A., Rogan, P. K., . . . Hejtmancik, J. F. (2005). A variant form of Oguchi disease mapped to 13q34 associated with partial deletion of GRK1 gene. *Mol Vis*, *11*, 977-985. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16319817

Figure 1.

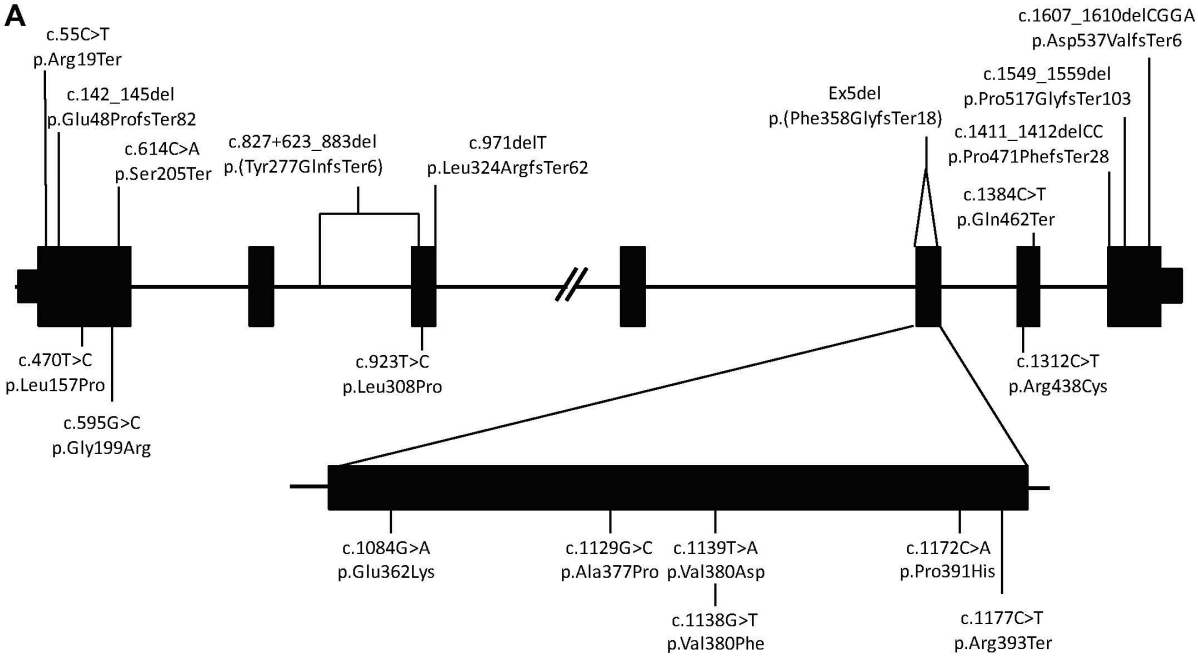


Figure 2.

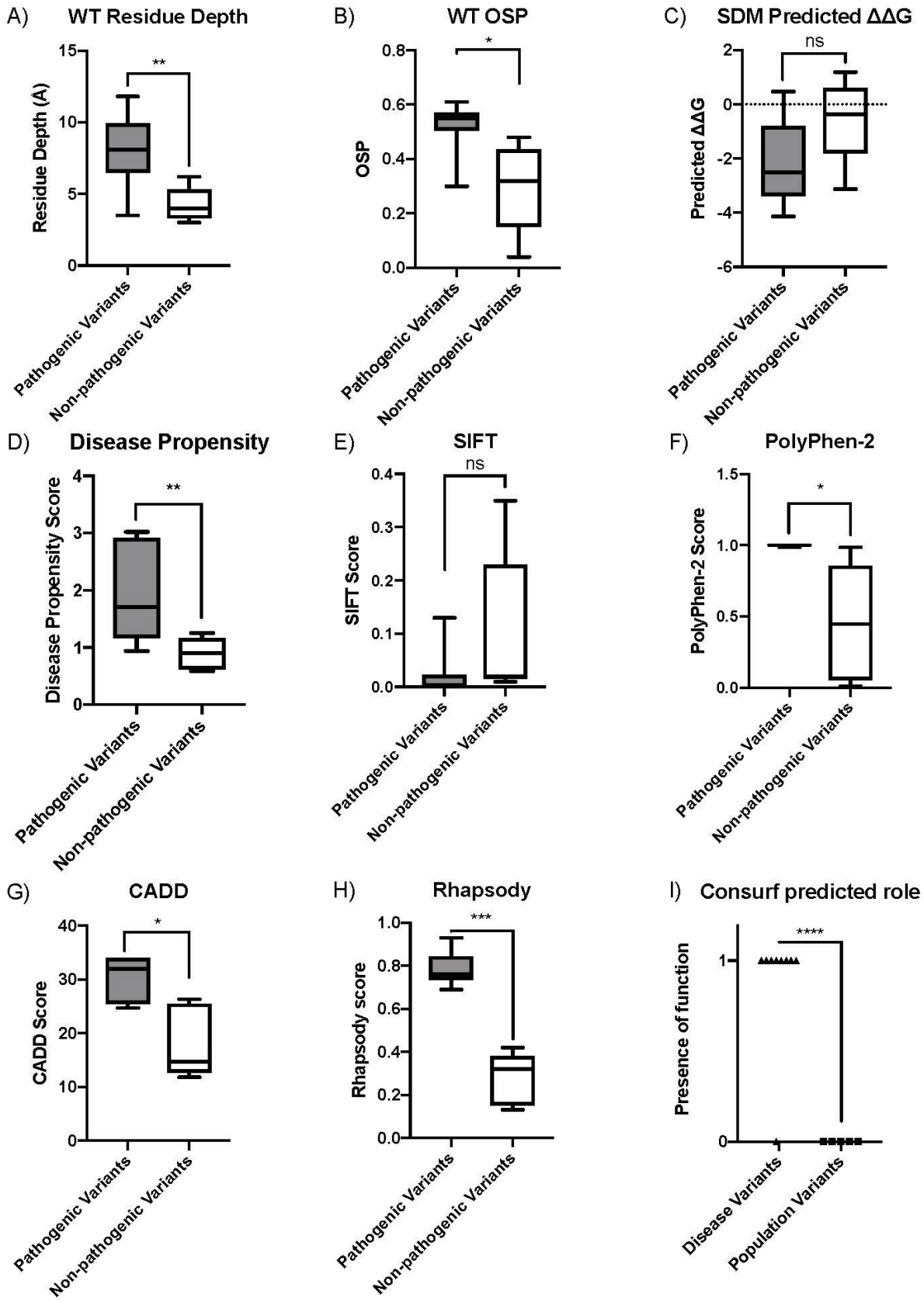

A)
WT Residue Depth
Residue Depth (A)
15
10
5
0
**
Pathogenic Variants
Non-pathogenic Variants
B)
WT OSP
OSP
0.8
0.6
0.4
0.2
0.0
*
Pathogenic Variants
Non-pathogenic Variants
C)
SDM Predicted ΔΔG
Predicted ΔΔG
2
0
-2
-4
-6
ns
Pathogenic Variants
Non-pathogenic Variants
D)
Disease Propensity
Disease Propensity Score
4
3
2
1
0
**
Pathogenic Variants
Non-pathogenic Variants
E)
SIFT
SIFT Score
0.4
0.3
0.2
0.1
0.0
ns
Pathogenic Variants
Non-pathogenic Variants
F)
PolyPhen-2
PolyPhen-2 Score
1.5
1.0
0.5
0.0
*
Pathogenic Variants
Non-pathogenic Variants
G)
CADD
CADD Score
40
30
20
10
0
*
Pathogenic Variants
Non-pathogenic Variants
H)
Rhapsody
Rhapsody score
1.0
0.8
0.6
0.4
0.2
0.0
***
Pathogenic Variants
Non-pathogenic Variants
I)
Consurf predicted role
Presence of function
1
0
****
Disease Variants
Population Variants

Figure 3.

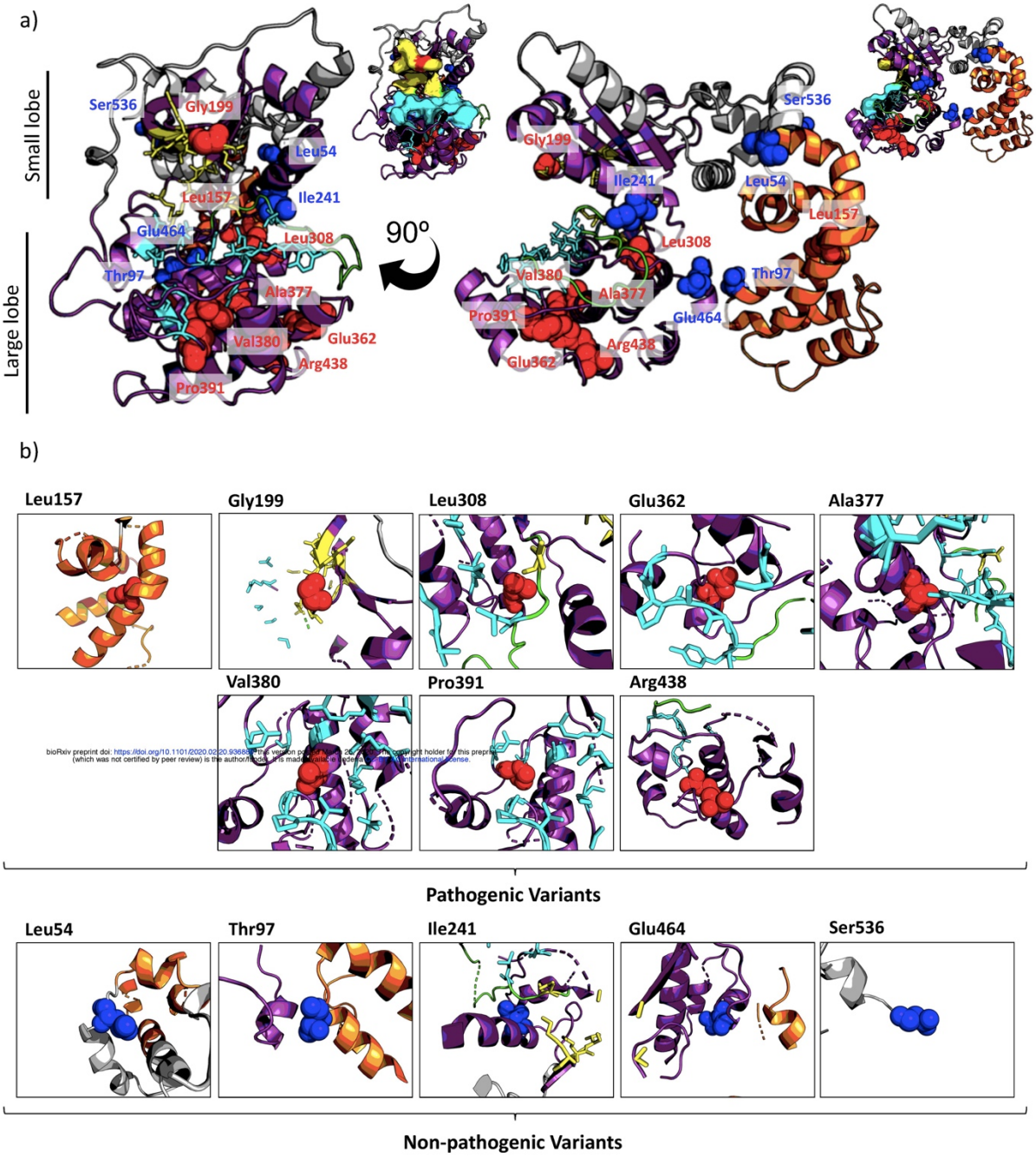


Figure 4.

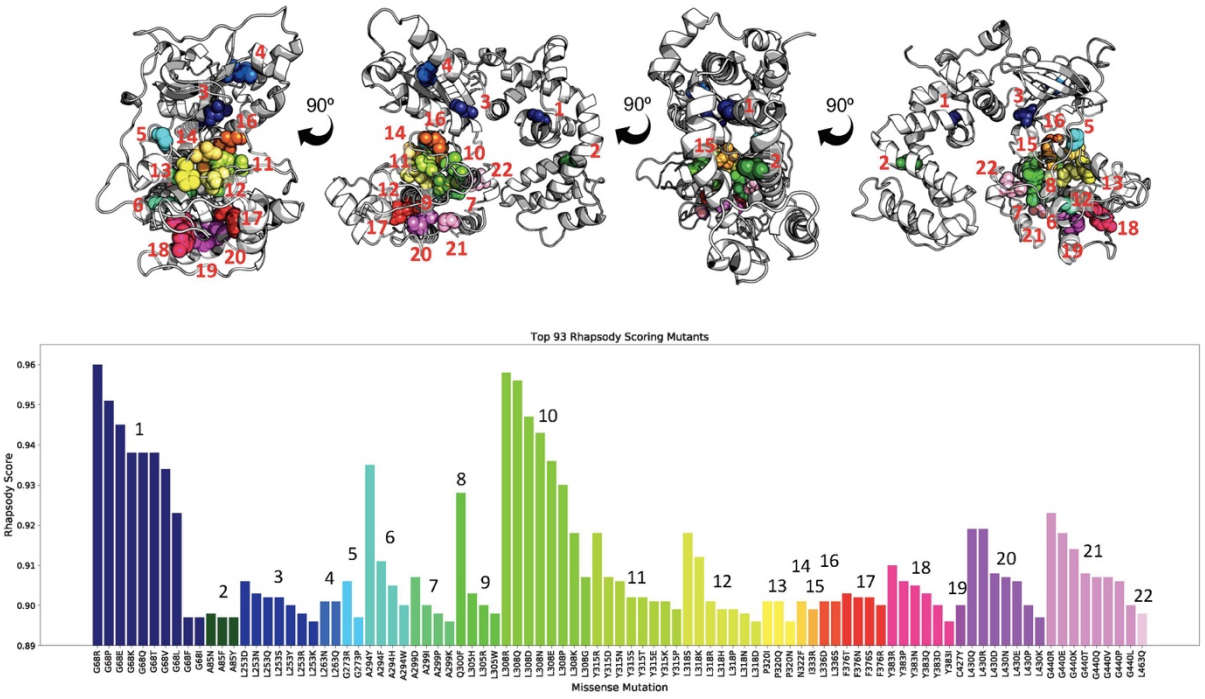

90°
90°
90°
Top 93 Rhapsody Scoring Mutants
Rhapsody Score
Missense Mutation

Table 1 – Summary of all novel and published mutations in GRK1.

DNA variant	Exon	Protein variant	rs number	gnomAD Frequency	Domain	PolyPhen-2	CADD (v1.3)	References
c.55C>T	1	p.Arg19Ter	rs370713047	17/239,700 (1 homozygous)	N-Terminus	N/A	37.0	Li et al (2017)
c.142_145del	1	p.Glu48ProfsTer82	rs748680704	6/248,384	RH domain	N/A	28.5	This study (**F1, F8**)
c.470T>C	1	p.Leu157Pro	N/A	0	RH domain	1.00 (Prob. D)	26.0	Mucciolo et al (2018)
c.595G>C	1	p.Gly199Arg	N/A	0	Kinase	1.00 (Prob. D)	24.7	This study (**F2**)
c.614C>A	1	p.Ser205Ter	N/A	0	Kinase	N/A	36	Azam et al (2009)
c.827+623_883del	3	p.(Tyr277GlnfsTer6)	N/A	0	Kinase	N/A	27.8	Zhang et al (2005)
c.923T>C	3	p.Leu308Pro	N/A	0	Kinase	0.996 (Prob. D)	29.2	Teke et al. (2016)
c.971del	3	p.Leu324ArgfsTer62	N/A	0	Kinase	N/A	35	Oishi et al (2007)
Ex5 del	5	p.(Phe358GlyfsTer18)	N/A	0	Kinase	N/A	25.1[a]	Yamamoto et al. (1997); Cideciyan et al. (1998)
c.1084G>A	5	p.Glu362Lys	N/A	0	Kinase	1.00 (Prob. D)	34	This study (**F3**)
c.1129G>C*	5	p.Ala377Pro	N/A	0	Kinase	0.991 (Prob. D)	25	Godara et al (2012)[1]
c.1138G>T	5	p.Val380Phe	N/A	2/173,226	Kinase	0.998 (Prob. D)	32	This study (**F4**)
c.1139T>A*	5	p.Val380Asp	rs777094000	5/173,124	Kinase	0.999 (Prob. D)	34	Yamamoto et al. (1997)[2]; Godara et al. (2012)[1]
c.1172C>A	5	p.Pro391His	rs570621429	1/141,802	Kinase	1.00 (Prob. D)	34	Hayashi et al (2007)
c.1177C>T	5	p.Arg393Ter	rs137877289	6/141,764	Kinase	N/A	46	This study (**F5**)[3]
c.1312C>T	6	p.Arg438Cys	rs765070399	2/141,922	Kinase	1.00 (Prob. D)	34	This study (**F9**)
c.1384C>T	6	p.Gln462Ter	N/A	0	Kinase	N/A	42	Jespersgaard et al (2019)
c.1411_1412del	7	p.Pro471PhefsTer28	N/A	0	Kinase	N/A	35	Oishi et al (2007)
c.1549_1559del	7	p.Pro517GlyfsTer130	N/A	0	RH domain	N/A	35	This study (**F6**)
c.1607_1610del	7	p.Asp537ValfsTer7	rs756235051	89/172,958	C-Terminus	N/A	26.7	Yamamoto et al (1997)[2]; Skorczyk-Werner et al (2015); This study (**F5**[2]**,F7, F10, F11**)

All mutations are shown relative to Human genome reference GRCh37, and GRK1 references NM_002929 and NP_002920. The precise start and end locations for each domain in GRK1 were obtained from Lodowski et al. (2006). All variants identified are all homozygous except for [1,2,3], which are compound heterozygous, * Predicted – DNA mutation not provided in publication, Prob. D = Probably Damaging as scored by PolyPhen2. [a]Based on deletion of whole of exon 5 and surrounding splice sites as exact coordinates are unknown.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.20.936980. this version posted March 25, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY 4.0 International license.

7.3.3 Management of full-thickness macular hole in a patient with Usher syndrome

Ophthalmol Ther
https://doi.org/10.1007/s40123-020-00276-4

CASE REPORT

Management of Full-Thickness Macular Hole in A Genetically Confirmed Case with Usher Syndrome

Evangelia S. Panagiotou · Thomas Papathomas · Konstantinos Nikopoulos · Stavrenia Koukoula · Mathieu Quinodoz · Atta Ur Rehman · Theodoros Giannopoulos · Carlo Rivolta · Anastasios G. Konstas

Received: May 12, 2020

ABSTRACT

Introduction: Full-thickness macular hole (FTMH) formation is rarely seen in patients with retinitis pigmentosa (RP) and can have an adverse impact on their residual visual function. The underlying mechanisms are unknown, and clinical experience is limited regarding surgical outcomes. Here, we describe the surgical management of FTMH in a young patient with genetically confirmed Usher syndrome, the most common form of syndromic RP.

Case Report: A 28-year-old woman presented with blurred vision in her right eye (RE). She had a history of RP and bilateral hearing impairment since childhood. Fundoscopy and spectral-domain optical coherence tomography revealed a FTMH in the RE along with typical RP features bilaterally. After pars plana vitrectomy (PPV) with internal limiting membrane peel and gas tamponade, the FTMH closed. Six months after PPV the patient underwent cataract surgery in the affected eye, and the visual acuity remained stable compared to baseline. The clinical diagnosis of Usher syndrome was genetically confirmed by whole exome sequencing (WES), which revealed the presence of two pathogenic nucleotide variants *in trans* (compound heterozygosity) in the gene *USH2A*.

Conclusion: We report a rare case of successful closure of a FTMH in a patient with Usher

Digital Features To view digital features for this article go to https://doi.org/10.6084/m9.figshare.12459272

E. S. Panagiotou · T. Papathomas · T. Giannopoulos · A. G. Konstas (✉)
1st Department of Ophthalmology, Aristotle University of Thessaloniki, AHEPA Hospital, Thessaloniki, Greece
e-mail: agkonstas@gmail.com

K. Nikopoulos
Laboratory of Oncogenomics, Department of Hematology, Lausanne University Hospital (CHUV), Lausanne, Switzerland

S. Koukoula
Ophthalmica Institute, Thessaloniki, Greece

M. Quinodoz · C. Rivolta
Department of Genetics and Genome Biology, University of Leicester, Leicester, UK

M. Quinodoz · C. Rivolta
Institute of Molecular and Clinical Ophthalmology Basel (IOB), Basel, Switzerland

M. Quinodoz · C. Rivolta
Department of Ophthalmology, University of Basel, Basel, Switzerland

A. U. Rehman
Division of Genetic Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

A. G. Konstas
3rd Department of Ophthalmology, Aristotle University of Thessaloniki, AHEPA Hospital, Thessaloniki, Greece

Adis

Published online: 21 June 2020

syndrome. Surgical treatment of FTMH can help preserve the central vision in RP patients, whose peripheral vision is severely affected.

Keywords: Full-thickness macular hole; Pars plana vitrectomy; Retinitis pigmentosa; Usher syndrome

Key Summary Points

Full-thickness macular hole (FTMH) formation is rarely seen in patients with retinitis pigmentosa (RP) and can have an adverse impact on their residual visual function, while clinical experience is limited regarding surgical outcomes.

We report a rare case of successful closure of a FTMH in a patient with Usher syndrome, the most common form of syndromic RP, which was genetically confirmed.

This is the first case of FTMH in Usher syndrome reporting surgical outcomes combined with genetic data.

Vitrectomy appears to be effective in these cases and should be performed to try to preserve the central vision of RP patients who usually have severely impaired peripheral vision.

INTRODUCTION

Usher syndrome is an autosomal recessive disorder characterised by retinitis pigmentosa (RP), sensorineural hearing loss and in some cases vestibular dysfunction [1] with a prevalence estimated to be between 6.2 and 16.67/100,000 [2, 3]. It is the most common form of syndromic RP and accounts for about 10% of inherited blindness caused by retinal dystrophies and optic neuropathies [4]. Usher syndrome is genetically heterogeneous and to date more than ten genes have been identified as causative of the disease [5, 6].

Although in RP the peripheral visual field is typically constricted with initial sparing of the central retina, macular abnormalities are often observed, leading to reduction of the visual acuity (VA). Cystoid macular oedema and epiretinal membrane are the most common macular anomalies in these patients, while the presence of a macular hole is rare [7, 8]. Since in RP patients the peripheral vision is affected, it is important to detect and treat any potential macular abnormalities to preserve central vision. Due to the limited experience in cases of macular hole in RP, it is useful to study the outcomes and prognosis of surgical treatment. Here, we report the anatomical and functional outcomes of surgical management in a young patient with Usher syndrome and full-thickness macular hole (FTMH). The genetic pathogenic variants underlying the patient's condition are also presented.

CASE PRESENTATION

A 28-year-old woman was referred to our department with a complaint of blurred vision during a period of 1 month in her right eye (RE). She reported nyctalopia and restricted visual field in both eyes since childhood and had been diagnosed RP with no family history. She also had had bilateral hearing impairment since childhood and wore hearing aids.

At presentation, her best-corrected VA was 2.5/10 (− 1.00sph) in the RE and 3.5/10 (− 1.00cyl × 90°) in the left eye (Snellen). Anterior segment examination was unremarkable with bilateral clear lenses. Dilated fundoscopy revealed a full-thickness macular hole and posterior vitreous detachment in the RE. In addition, both eyes showed typical signs of RP including peripheral retinal atrophy with bone spicule-shaped pigmentation in the mid periphery, waxy pallor of the optic nerve head and attenuation of retinal vessels (Fig. 1). Spectral domain optical coherence tomography (SD-OCT) confirmed the presence of a FTMH stage IV in the RE along with atrophy of the outer retinal layers and absence of the ellipsoid zone (Fig. 2a). In the fellow eye, thinning of the outer retinal layers was also observed, while the

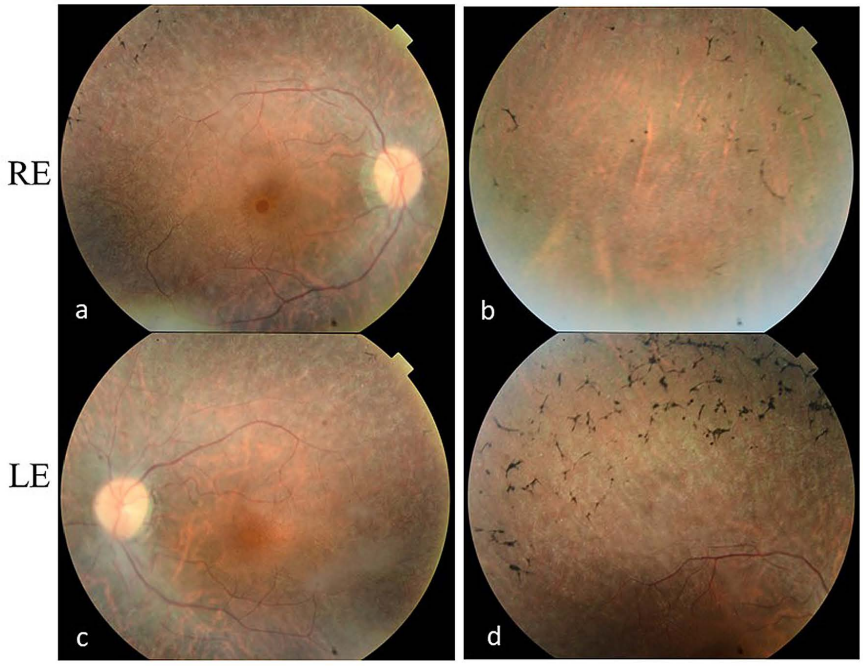


Fig. 1 Pre-operative fundus photographs. **a**, **b** Fundus of the right eye showing peripheral retinal atrophy with a few intraretinal pigment deposits in a bone-spicule configuration, retinal vessel attenuation and optic nerve head pallor, while a full-thickness macular hole can be seen in the macula. **c**, **d** The left eye also has a typical RP appearance with multiple bone spicules in the periphery and a relatively spared macula. *RE* right eye, *LE* left eye

ellipsoid zone was only preserved in the fovea (Fig. 2b, d, f).

The patient underwent 23-gauge three-port pars plana vitrectomy (PPV) with internal limiting membrane (ILM) peeling assisted by membrane dual blue and tamponade with 16% hexafluoroethane (C_2F_6) gas. One month after surgery, OCT (RE) confirmed the successful closure of the macular hole with severe disruption of the ellipsoid zone (Fig. 2c). Due to posterior subcapsular cataract development in the RE, the patient subsequently underwent successful uncomplicated cataract surgery with phacoemulsification and intraocular lens implantation 6 months after PPV. At final follow-up, 11 months after PPV, the BCVA remained stable at 2.5/10 (RE) while the macular hole remained closed and a small part of the ellipsoid zone could be seen (Fig. 2e).

The scotopic full-field electroretinogram (ffERG) was non-recordable in either of the eyes indicating mostly loss of rod system sensitivity (Fig. 3a, b). The responses obtained from the photopic ERG, which entirely assesses cone function, were reduced and delayed in both eyes (Fig. 3c, d). These findings are in keeping with rod-cone dystrophy [1]. Due to the co-existence of severe bilateral sensorineural hearing impairment (data not shown), a clinical diagnosis of Usher syndrome was made.

After written informed consent had been obtained from the patient and family members

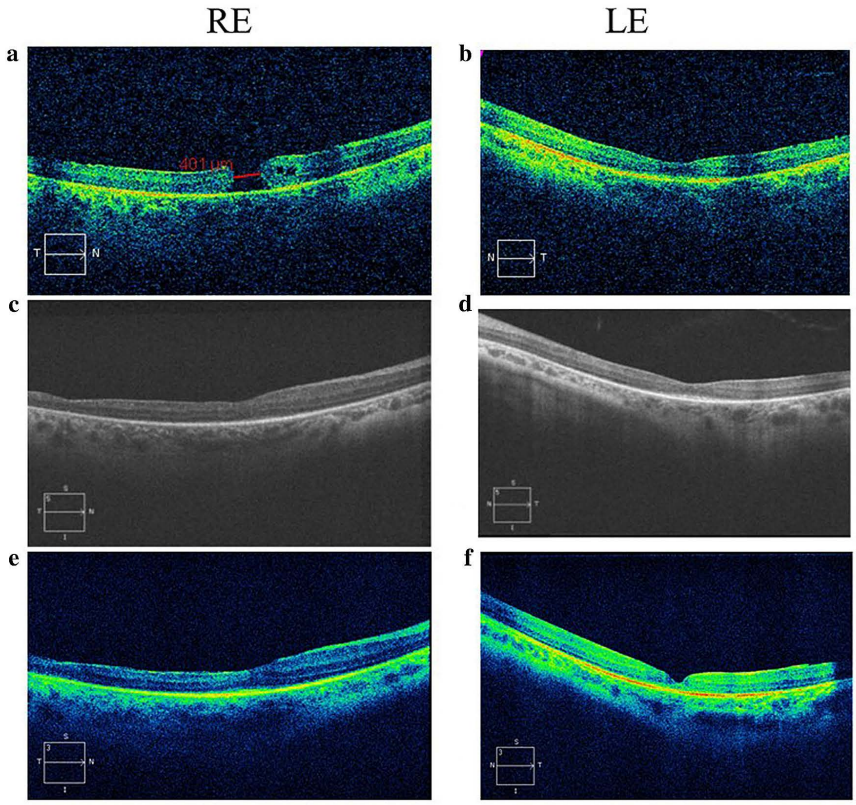


Fig. 2 SD-OCT of the macula of both eyes. **a** Preoperatively, in the right eye a FTMH (width 401 µm) and a few intraretinal cystic spaces can be seen. There is also thinning of the outer retinal layers and no ellipsoid zone can be seen. **b**, **d**, **f** In the left eye, thinning of the outer retinal layers is observed, while only the foveal part of the ellipsoid zone is preserved. **c** One month after vitrectomy, the FTMH is closed in the right eye. **e** Eleven months after vitrectomy, the macular hole in the right eye remains closed with only few photoreceptors remaining in the temporal part of the fovea. *RE* right eye, *LE* left eye

(approved by the Bioethics Committee of the School of Medicine, Aristotle University of Thessaloniki), saliva samples were obtained from the patient and family members for subsequent genomic DNA extraction (Oragene DNA OG-500 saliva kits, DNA Genotek). Genetic testing following whole exome sequencing (WES) analyses revealed that the patient carried two compound heterozygous pathogenic variants in the *USH2A* gene (OMIM 608400, NM_007123.5) and in particular: a nonsense mutation, c.100C>T, p.(Arg34Ter), and a synonymous change c.949C>A, p.(Arg317=), which has been shown to lead to a new splice site [9]. The two variants, classified as pathogenic in ClinVar by multiple submitters, segregated with the phenotype in the patient's

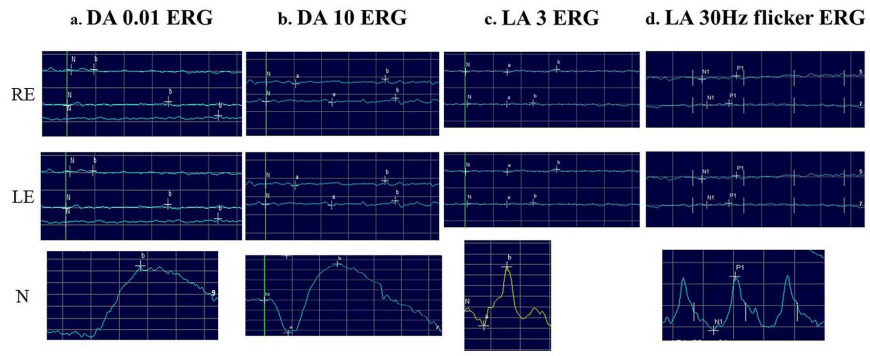


Fig. 3 Full-field ERG findings indicating the presence of rod-cone dystrophy. **a**, **b** Under scotopic conditions, the ERG is non-detectable in both eyes. **c**, **d** Under photopic conditions, the responses are reduced and delayed bilaterally. The x axis represents time in ms (20 ms/div). The *y* axis represents potential in μV (**a**, **b** 50 μV/div, **c**-d 20 μV/div). *DA* dark-adapted, *LA* light-adapted, *ERG* electroretinogram, *RE* right eye, *LE* left eye, *N* normal

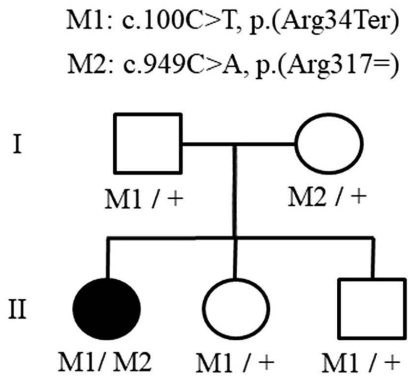


Fig. 4 Genetic analysis of the patient's family showing the pedigree and segregation data for the pathogenic variants in *USH2A*. The genotypes for all tested family members are shown below each individual, with M1 representing the first mutant allele, c.100C>T, p.(Arg34Ter), M2 the second mutant allele, c.949C>A, p.(Arg317=) and + representing the wild-type allele. The affected individual is shaded black

family (Fig. 4). Genetic counselling was given to the patient and family members.

DISCUSSION

Macular abnormalities are relatively frequent in RP patients, the most prevalent being cystoid macular oedema, epiretinal membrane and vitreomacular traction [7, 8, 10]. FTMH in RP has been reported to be rare (≤ 1%), although most studies have focused on non-syndromic RP patients [7, 10–12]. The only study including exclusively Usher patients was performed by Testa et al. [13] who demonstrated a high prevalence of macular abnormalities (47%). In this Italian cohort of 134 genetically confirmed Usher cases, FTMH was found only in one eye of a patient carrying *USH2A* mutations and thus showing a prevalence of 0.4% [13].

The mechanism underlying the formation of FTMH in RP remains unclear and is considered multifactorial. Principally, abnormalities of the vitreomacular interface, such as vitreomacular traction and epiretinal membrane, are considered to be a major mechanism [11, 14, 15]. Moreover, atrophy due to chronic cystoid macular oedema and macular schisis in highly myopic eyes could potentially lead to FTMH [12, 15]. Of note, our patient had posterior vitreous detachment in the affected eye, so our hypothesis is that vitreomacular traction is the underlying mechanism in this case.

The role of vitrectomy is established in the treatment of idiopathic FTMH leading to a high closure rate and significant VA improvement compared to observation [16]. However, due to the rarity of FTMH cases in the context of RP, surgical outcomes have only been reported in a few small case series including patients with heterogeneous characteristics such as RP stage, symptoms duration, baseline VA, FTMH size and ocular comorbidities [10, 12, 14, 15, 17, 18]. Eight RP cases with FTMH have been reported to have a successful anatomic outcome with seven of them showing higher post-operative VA and one of them having unaltered VA [10, 14, 15, 18]. Another four cases with high myopia have been reported, two of which were accompanied by retinal detachment. All four underwent vitrectomy resulting in successful closure, with VA improvement in three patients, while in the fourth one the VA deteriorated [12, 15, 18]. Moreover, in three patients the FTMH failed to close after one vitrectomy and resulted in worse VA [10, 14, 18]. Of note, FTMH reopening after successful surgical intervention has been reported in one patient [17]. Based on these studies, a successful closure of the macular hole is achieved in the majority of the cases, whereas the visual outcomes seem to be mostly positive but can be variable. Thus, the surgical treatment of these cases is usually considered beneficial, notwithstanding the potential risks of vitrectomy, including phototoxicity [19] and visual field defects [20]. The presence of long-standing retinal degeneration may affect the visual outcomes in RP patients. In our patient the FTMH remained closed until the most recent follow-up 11 months post-operatively and the VA remained stable compared to baseline.

Importantly, most of these studies have included patients with either non-syndromic RP or no clear clinical statement (syndromic vs. non-syndromic RP) [10, 12, 15, 18]. The only study reporting surgical outcomes of a FTMH case with Usher syndrome was performed by Vingolo et al. [21]. The authors report the results from three cases including two RP patients with lamellar holes and one Usher subject with FTMH who underwent combined microincision vitrectomy and cataract surgery. The latter case was successful showing a VA improvement from 2/20 to 4/20. However, the underlying mechanism may be different from our report because the macular holes were secondary to "chronic CMO and tangential vitreoretinal tractions". No OCT data were presented for this patient [21].

No genetic data were presented in the aforementioned reports of similar cases. To the best of our knowledge, this is the first case of surgical management of a FTMH in a patient with genetically confirmed Usher syndrome. In our patient, WES revealed two compound heterozygous mutations in *USH2A*. Pathogenic variants in this gene account for the majority among the entire spectrum of genetic aberrations present in patients with Usher syndrome type 2 [4]. *USH2A* encodes usherin, localised to the periciliary membrane complex, a region of the photoreceptor inner segment that surrounds the connecting cilium between the outer and inner segment [22]. The stop mutation, p.(Arg34Ter), has been previously reported in several patients with Usher syndrome [23, 24]. The second pathogenic variant, p.(Arg317=), is predicted to activate a cryptic donor splice site and has also been previously reported [25, 26]. Its effect has been functionally assessed in vitro at the mRNA level [r.(949C>A, 951_1143del)], where it was shown to lead to the deletion of the last 193 bases of exon 6 and thus result in the introduction of a premature termination codon in exon 7 [9]. We recommend that the genetic defects underlying these patients' RP should be reported, when possible. Apart from the typical benefits of genetic testing, such as obtaining an accurate diagnosis, offering genetic counselling and identifying patients eligible for gene therapy or clinical trials, it could help us establish useful genotype-phenotype correlations, e.g. while examining whether specific genes or mutations are associated with the development of macular hole in RP.

CONCLUSION

We report a rare case of FTMH in a patient with genetically confirmed Usher syndrome that was

treated successfully with surgery. Our case report supports the notion that vitrectomy is an effective approach in these cases as demonstrated by the successful anatomical outcome, which stabilised central vision. Surgical management should therefore be selected to try to preserve the central vision of these patients who already may have severely impaired peripheral vision. Further research is needed to enrich our understanding and surgical experience in such cases and potentially identify clinical and/or genetic features related to the development and prognosis of FTMH in the context of RP.

ACKNOWLEDGEMENTS

Funding. Genetic analyses were supported by the Swiss National Science Foundation (Berne, Switzerland), grant no. 176097, to Carlo Rivolta. No funding or sponsorship was received for the publication of this article.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Prior Presentation. This manuscript is based on work that was presented at the 15th Meeting of the Greek Vitreoretinal Society, 30 January-1 February 2020, Athens, Greece.

Disclosures. Evangelia S. Panagiotou, Thomas Papathomas, Konstantinos Nikopoulos, Stavrenia Koukoula, Mathieu Quinodoz, Atta Ur Rehman, Theodoros Giannopoulos and Carlo Rivolta declare that they have no conflict of interest. Anastasios G. Konstas is a member of the journal's Editorial Board.

Compliance with Ethics Guidelines. Written informed consent for the publication of this report and for genetic testing was obtained from the patient and family members (approved by the Bioethics Committee of the School of Medicine, Aristotle University of Thessaloniki).

Data Availability. Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

REFERENCES

1. Leroy BP. Usher Syndromes. In: Puech B, De Laey J-J, Holder GE, editors. Inherited chorioretinal dystrophies. 1st ed. Berlin: Springer; 2014. p. 143–149.
2. Hope CI. Usher syndrome in the city of Birmingham—prevalence and clinical classification. Br J Ophthalmol. 1997;81(1):46–53.
3. Kimberling WJ, Hildebrand MS, Shearer AE, et al. Frequency of Usher syndrome in two pediatric populations: implications for genetic screening of deaf and hard of hearing children. Genet Med. 2010;12(8):512–6.
4. Bocquet B, Lacroux A, Surget MO, et al. Relative frequencies of inherited retinal dystrophies and optic neuropathies in Southern France: assessment of 21-year data management. Ophthalm Epidemiol. 2013;20(1):13–25.
5. Mathur P, Yang J. Usher syndrome: hearing loss, retinal degeneration and associated abnormalities. Biochim Biophys Acta Mol Basis Dis [Internet]. 2015;1852(3):406–20.

6. RETNET (The Retinal Information Network) [Internet]. https://sph.uth.edu/RetNet/

7. Testa F, Rossi S, Colucci R, et al. Macular abnormalities in Italian patients with retinitis pigmentosa. Br J Ophthalmol. 2014;98(7):946–50.

8. Liew G, Strong S, Bradley P, et al. Prevalence of cystoid macular oedema, epiretinal membrane and cataract in retinitis pigmentosa. Br J Ophthalmol. 2019;103(8):1163–6.

9. Vaché C, Besnard T, Blanchet C, et al. Nasal epithelial cells are a reliable source to study splicing variants in Usher syndrome. Hum Mutat. 2010;31(6):734–41.

10. Hagiwara A, Yamamoto S, Ogata K, et al. Macular abnormalities in patients with retinitis pigmentosa: prevalence on OCT examination and outcomes of vitreoretinal surgery. Acta Ophthalmol. 2011;89(2):122–5.

11. Fragiotta S, Rossi T, Carnevale C, et al. Vitreomacular interface disorders in retinitis pigmentosa. Graefe's Arch Clin Exp Ophthalmol. 2019;257(10):2137–46.

12. Liu J, Lyu J, Zhang X, Zhao P. Lamellar hole-associated epiretinal membrane is a common feature of macular holes in retinitis pigmentosa. Eye [Internet]. 2019;34(4):643–9.

13. Testa F, Melillo P, Rossi S, et al. Prevalence of macular abnormalities assessed by optical coherence tomography in patients with Usher syndrome. Ophthalmic Genet [Internet]. 2018;39(1):17–211.

14. Yan F, Xia FJ, Jiang F, Yu HG. Visual and morphological outcomes of vitreomacular traction syndrome in retinitis pigmentosa treated by vitrectomy. Int J Ophthalmol. 2018;11(8):1411–5.

15. Jin ZB, Gan DK, Xu GZ, Nao-I N. Macular hole formation in patients with retinitis pigmentosa and prognosis of pars plana vitrectomy. Retina. 2008;28(4):610–4.

16. Parravano M, Giansanti F, Cm E, Yc Y, Rizzo S, Virgili G. Vitrectomy for idiopathic macular hole. Cochrane Database Syst Rev. 2015;1:5.

17. García-Fernández M, Castro-Navarro J, Bajo-Fuente A. Unilateral recurrent macular hole in a patient with retinitis pigmentosa: a case report. J Med Case Rep. 2013;7:2–5.

18. Ratra D, Raval V. Surgery for macular holes associated with unusual concomitant pathologies. Oman J Ophthalmol. 2013;6(2):112–5.

19. Charles S. Illumination and phototoxicity issues in vitreoretinal surgery. Retina. 2008;28(1):1–4.

20. Yonemura N, Hirata A, Hasumura T, Negi A. Fundus changes corresponding to visual field defects after vitrectomy for macular hole. Ophthalmology. 2001;108(9):1638–43.

21. Vingolo EM, Valente S, Gerace E, Spadea L, Nebbioso M. Macular hole in retinitis pigmentosa patients: microincision vitrectomy with polydimethylsiloxane as possible treatment. Eye. 2015;29(5):699–702.

22. Maerker T, van Wijk E, Overlack N, et al. A novel Usher protein network at the periciliary reloading point between molecular transport machineries in vertebrate photoreceptor cells. Hum Mol Genet. 2008;17(1):71–86.

23. Dreyer B, Tranebjærg L, Rosenberg T, Weston MD, Kimberling WJ, Nilssen Ø. Identification of novel USH2A mutations: implications for the structure of USH2A protein. Eur J Hum Genet. 2000;8(7):500–6.

24. Bernal S, Medà C, Solans T, et al. Clinical and genetic studies in Spanish patients with Usher syndrome type II: description of new mutations and evidence for a lack of genotype-phenotype correlation. Clin Genet. 2005;68(3):204–14.

25. Pennings RJE, Te Brinke H, Weston MD, et al. USH2A mutation analysis in 70 Dutch families with Usher syndrome type II. Hum Mutat. 2004;24(2):185.

26. Dreyer B, Brox V, Tranebjaerg L, et al. Spectrum of USH2A mutations in Scandinavian patients with Usher syndrome type II. Hum Mutat. 2008;29(3):451.