Multi-disciplinary team directed analysis of whole genome sequencing reveals pathogenic non-coding variants in molecularly undiagnosed inherited retinal dystrophies

Malena Daich Varela[1,2], James Bellingham,[1] Fabiana Motta,[1,3] Neringa Jurkute,[1,2] Jamie M Ellingford,[4,5] Mathieu Quinodoz,[6,7,8] Kathryn Oprych,[1] Michael Niblock,[1] Lucas Janeschitz-Kriegl,[6,7] Karolina Kaminska,[6,7] Francesca Cancellieri,[6,7] Hendrik PN Scholl,[6,7] Eva Lenassi,[4,5] Elena Schiff,[2] Hannah Knight,[2] Graeme Black,[4,5] Carlo Rivolta,[6,7,8] Michael E Cheetham,[1] Michel Michaelides,[1,2] Omar A Mahroo,[1,2] Anthony T Moore,[1,2,9] Andrew R Webster,[1,2] Gavin Arno*[1,2,10]

[1]UCL Institute of Ophthalmology, London, UK

[2]Moorfields Eye Hospital, London, UK

[3]Department of Ophthalmology, Universidade Federal de Sao Paulo, Sao Paulo, Brazil

[4]North West Genomic Laboratory Hub, Manchester Centre for Genomic Medicine, Manchester University Hospitals NHS Foundation Trust, St Mary's Hospital, Manchester, UK.

[5]Division of Evolution and Genomic Sciences, Neuroscience and Mental Health Domain, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.

[6]Institute of Molecular and Clinical Ophthalmology Basel, 4031 Basel, Switzerland. [7]Department of Ophthalmology, University of Basel, 4031 Basel, Switzerland. [8]Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, UK.

[9]University of California, San Francisco, San Francisco, California, USA.

[10]North Thames Genomic Laboratory Hub, Great Ormond Street Hospital For Children, London, UK

*Corresponding author: Gavin Arno, UCL Institute of Ophthalmology, 11-43 Bath Street, London EC1v 9EL, UK, g.arno@ucl.ac.uk, +44(0)2076086971

Abstract

Purpose: To identify, using genome sequencing (GS), likely pathogenic non-coding variants in inherited retinal dystrophy (IRD) genes

Methods: Patients with IRD were recruited to the study and underwent comprehensive ophthalmological evaluation and GS. The results of GS were investigated through virtual gene panel analysis and plausible pathogenic variants and clinical phenotype evaluated by multi-disciplinary team (MDT) discussion. For unsolved patients in whom a specific gene was suspected to harbour a missed pathogenic variant, targeted re-analysis of non-coding regions was performed on GS data. Candidate variants were functionally tested including by mRNA analysis, minigene and luciferase reporter assays.

Results: Previously unreported, likely pathogenic, non-coding variants, in 7 genes (*PRPF31*, *NDP*, *IFT140*, *CRB1*, *USH2A*, *BBS10*, and *GUCY2D*), were identified in 11 patients. These

were shown to lead to mis-splicing (*PRPF31*, *IFT140*, *CRB1*, *USH2A*) or altered transcription levels (*BBS10*, *GUCY2D*).

Conclusion: MDT-led, phenotype driven, non-coding variant re-analysis of GS is effective in identifying missing causative alleles.

Introduction

Inherited retinal dystrophy (IRD) is a heterogeneous group of rare diseases that result in visual impairment caused by retinal dysfunction.[1] Over 300 genes and loci have been associated with IRD to date (https://sph.uth.edu/retnet/), with a carrier frequency estimated to be up to approximately 1 in 2.5 individuals,[2] and a prevalence of around 1 in 2000 people.[3] IRD can be classified in many ways taking into account inheritance pattern, age at onset, rate of progression, main retinal cell type affected (rods, cones, retinal pigment epithelium, retinal ganglion cells or choroid), and extra-ocular features.[4]

Genetic diagnosis of IRD is essential for effective clinical management and is more relevant now with the approval of the first gene therapy for *RPE65*-retinal dystrophy (Luxturna), and with many other clinical trials for these disorders in progress and development.[5] Although some of the approaches such as optogenetics, retinal cell transplantation, and artificial retinal prostheses) are gene agnostic ,[6–8] many of the current efforts focus on gene and RNA supplementation and editing, making genetic diagnosis a key inclusion criterion.[9–12]

Genetic testing in the ophthalmic genetics' clinic has evolved dramatically over the last decades, as technologies have become more accessible and inexpensive. Single-candidate gene approach by Sanger sequencing, next generation sequencing (NGS)-based panels,

exome sequencing (ES), and genome sequencing (GS) are currently employed by clinical diagnostic, commercial, and research labs worldwide.[13] GS represents the most comprehensive test, with a significant improvement in diagnostic yield compared to NGS panels[14] and ES.[15] Yet, the pathogenic variant detection rate in IRD ranges between 50% to 85% depending on patient cohorts.[15–18] The missing genetic etiology has been attributed to undiscovered genes, novel damaging variants, copy number variants (CNV), synonymous changes that affect transcript processing, and changes in non-coding regions (introns and regulatory) that are particularly difficult to interpret and functionally assay, especially for retina-specific genes.[15,19–21] However, for molecular geneticists to truly capitalise on this technology, improved interpretation and understanding of how or which non-protein-altering variants cause disease is required.

Here we present findings from large-scale GS studies in the United Kingdom (UK): the National Institute for Health Research Rare Disease (NIHR-RD) and 100,000 Genomes Project (100KGP). We demonstrate that targeted investigations of non-coding variants in a case-led manner is effective for identification of missing variants in IRD.

Results

Patient information including clinical findings and genetic details are summarized in Tables 1 and 2. Candidate gene non-coding variant analysis, following negative clinical testing, and multidisciplinary team (MDT) discussion highlighting specific candidate genes in each case led us to identify 2 patients with a candidate variant in *PRPF31*, 1 with a variant in *NDP*, 2 with a recurrent variant in *CRB1*, 1 variant in *USH2A*, 1 in *IFT140*, 1 in *BBS10* and 1 in *GUCY2D*.

Patient 1 (GC 763) is a 34-year-old male with typical fundus features of rod-cone dystrophy (RCD), no extra-ocular manifestations suggesting syndromic disease, and a family history indicative of autosomal dominant RCD (adRCD, Figure 1A), with incomplete penetrance due to the observed unaffected obligate carrier mother. *PRPF31*, a common cause of ADRP showing incomplete penetrance, was screened by direct Sanger sequencing and was followed by NGS panel genetic testing for all known autosomal dominant rod-cone dystrophy genes; these were negative. Subsequently, he was recruited for GS through the 100KGP with his unaffected parents.

Following negative coding variant analysis through the clinical pipeline and MDT discussion, *PRPF31* was deemed the most likely gene to harbour a pathogenic variant. A single rare variant (MAF<0.001) was found in intron 10 of the *PRPF31* gene: (GRCh38) chr19:54129939C>G NM_015629.4: c.1374+569C>G. This variant was absent from the gnomAD dataset and predicted to cause a C>G change at the +3 position of a deep intronic donor site, creating a new splice donor site (nnsplice score 0.61: TCTgt**c**agt > TCTgt**g**agt). Further analysis of the upstream 500bp revealed only weak splice acceptor sites with the closest being 88bp upstream of the new splice donor site (nnsplice score 0.37; Figure 2).

Reverse transcription polymerase chain reaction (RT-PCR) of the *PRPF31* transcript using oligonucleotide primers spanning exon 8-14 resulted in two distinct amplicons from whole blood RNA of patient 1, compared to a control sample. Direct sequencing of the two amplicons showed a wild-type spliced transcript and a larger fragment incorporating a deep intronic cryptic exon of 88bp, matching the by *in silico* predicted pseudoexon. Inclusion of 88bp into the transcript after exon 10 would lead to a reading-frame shift and a premature termination codon: p.(Gly459Serfs*46) indicating that this is likely to represent a loss of function (LOF) allele. This variant was also found in the patient's three affected siblings, carrier mother, and affected maternal grandfather.

One additional candidate variant (c.-9+1G>A) was identified in *PRPF31* in an individual with adRCD (**Patient 2**, GC14595, table 1 and 2) and his similarly affected mother. This variant is absent from the gnomAD dataset and is located at a canonical splice donor site in the 5' untranslated region (UTR). It was not possible to obtain a fresh sample from the patient for further analysis.

Patient 3 (GC 21538) reported decreased vision since his late 20s, followed by night blindness and loss of peripheral vision. He was diagnosed with cone-rod dystrophy (CRD). His acuity slowly decreased over a follow up period of 17 years, and rod-derived symptoms such as nyctalopia started to arise during his 30s. He did not present extra-ocular manifestations that would suggest syndromic disease at the age of 45 years.

Retinal dystrophy panel genetic testing identified a heterozygous previously reported missense variant in *IFT140*, c.2611C>T, p.(Arg871Cys). This is present in 4/253206 alleles in gnomAD v2.1 and reported 3 times in ClinVar (2 pathogenic, 1 variant of uncertain significance, VUS). Non-coding region analysis revealed the intronic variant c.2577+4_2577+5del, predicted to abolish the canonical splice donor site of intron 20 AAGgtg**ag** > AAGgtg**gg** (nnsplice score: 0.00). This variant was also found in the unaffected patient's mother. Nested RT-PCR of the patient's sample showed multiple fragments amplified: wild-type, partial skipping of exon 20, and complete skipping of exon 20 (Figure 3).

Patient 4 (GC 18850) was reviewed shortly after a full-term birth due to horizontal nystagmus and poor fixation. At 6 months old, he was noticed to have bilateral retinal folds with subretinal fluid accumulating around the optic disc (Figure 1B). His retinal phenotype and poor visual acuity (VA; 1.6 LogMAR in the right eye and No Light Perception in the left eye) were stable throughout the years. His maternal grandfather had a history of childhood

exudative vitreoretinopathy and bilateral retinal detachment in his 20s, implying an X-linked inheritance pattern. Taken together, the clinical findings were suggestive of Norrie disease or X-linked familial exudative retinopathy, secondary to a damaging variant in *NDP*. His case remained unsolved following coding variant analysis as part of the NIHR-RD study. Rare variant analysis of the entire *NDP* gene region, revealed a single variant, chrX:43817961C>T: NM_000266.4: c.-70G>A. The was absent from the gnomAD dataset and unique to the proband within the NIHR-RD (approx.13,000 alleles) and absent from the 100KGP dataset. This variant, although outside the promoter region, is located in a transcription factor binding region and DNase hypersensitive region spanning the UTR of exon 2 (UCSC ChIP-seq TFB clusters track [strongest binding: CTCF, SMC3, RAD21] and UCSC DNase ChIP-seq metadata tracks [116/125 cell types], ENCODE datasets). However, we were unable to demonstrate any effect of the variant on the transcriptional activity of a luciferase reporter gene in HEK293 cells (data not shown). Subsequently, SpliceAI prediction of the variant effect showed a donor gain (score 0.40) GAGgt**g**aa > GAGgt**a**aa at position c.-72 which may introduce a splice donor site upstream of the start codon and therefore disrupt the correct transcript. RNA samples were unavailable to examine this.

Patients 5 (GC 17009) **and 6** (GC 3671) are two unrelated males, born to unaffected non-consanguineous parents, who had onset of visual symptoms in infancy and were diagnosed with early-onset severe retinal dystrophy (EOSRD). There was no family history of eye disease. Each had a retinal phenotype highly suggestive of *CRB1*-retinopathy (Figure 1C). Patient 5 was recruited to the 100KGP along with his unaffected parents, and patient 6 was recruited to the NIHR-RD study as a singleton. Coding variant analysis (including splice regions) of all IRD genes identified only a single heterozygous pathogenic or likely pathogenic variant in *CRB1* in each case (patient 5: NM_201253.3: c.2290C>T, p.(Arg764Cys), patient 6: c.2842+5G>A). Non-coding region analysis of the *CRB1*-locus

revealed an identical single candidate second variant in both cases (c.3879-1203C>G). This variant was predicted to create a deep intronic splice donor site (+1 position C>G, nnsplice score 0.96: CAG**c**tatg > CAG**g**tatg). Analysis of the single-nucleotide variants (SNVs) co-inherited with the variant of interest showed that the haplotype was different in each case, demonstrating that the variant was likely to have arisen independently. Minigene analysis using a fragment of *CRB1* exon 10-11, including the 3.49kb intron in a splicing minigene vector construct, and introducing the c.3879-1203C>G variant by site-directed mutagenesis demonstrated that a cryptic exon of 156bp was included into transcripts when transfected into HEK293 cells (Figure 4). This would be expected to lead to the frameshift consequence, p.(Cys1294Phefs*2). The near complete loss of normal splicing demonstrated by the minigene assay may suggest that this represents a LOF allele. An additional patient was found to harbor the same deep intronic *CRB1* variant *in trans* with a LOF variant (chr1:197427726A>T NM_201253.3*:* c.2401A>T p.Lys801*) through data sharing as part of the European Retinal Dystrophy Consortium (Patient 10, table 1 and 2, supplement). Her symptoms started at the age of 2.5 years and due to nyctalopia and field constriction she was diagnosed with EOSRD at the age 4 years. On exam, she had para-arteriolar preservation of the RPE, pigmented bone spicules, attenuated vessels, and generalized atrophy. Optical Coherence Tomography was typical of *CRB1*-retinopathy, with de-laminated inner retinal layers, retinal thickening, and loss of outer layers.

Patient 7 (GC 3769) was diagnosed with Usher syndrome type II, with hearing loss identified in early childhood and RCD diagnosed in her teenage years. The patient presented with a mild RCD phenotype, retaining a visual field of around 25 degrees at the age of 41 years with preserved visual acuity of 0.2 logMAR in the right and left eyes. Her fundus examination showed classic triad of RCD: pigmented peripheral bone spicules, vessel thinning and pale optic discs (Figure 1D). OCT imaging revealed oedema and loss of the EZ

line nasal and temporal to the fovea. Routine genetic testing identified the single heterozygous pathogenic variant in *USH2A,* chr1:216325412T>G, c.1036A>C, p.(Asn346His) (NM_206933.4). She was recruited to the 100KGP with her unaffected mother and unaffected brother. Phasing of the missense variant in the unaffected family members demonstrated that the missing variant should have been inherited from her mother, non-coding variant analysis of the *USH2A* locus focused on the maternal allele revealed 6 variants rare in the gnomAD dataset (MAF<0.01) with only one, Chr1:216261980T>C, c.4885+375A>G, surviving additional filtering (MAF<0.001, manual curation for low-complexity region variants) and having a strong prediction for splice site effect, which strengthens a deep intronic splice donor site at the +5 position: AAAgtaa**a** > AAAgtaa**g**). Direct Sanger sequencing of PCR amplicons spanning exons 22-24 of *USH2A* from nasal epithelial cell brushings showed an alternate splice product comprising inclusion of a pseudoexon (130bp), predicted to lead to a frameshift and premature stop codon, p.(Gly1629Valfs*52), consequent upon *USH2A* c.4885+375A>G (Figure 5).

Patient 8 (GC 18582) had been diagnosed as an adult with CRD; she developed loss of central vision and photophobia in her 4th decade. She had bilateral macular and peripheral retinal atrophy which progressed during follow up (Figure 1E. Her disease was confined to the retina with no syndromic features.

Singleton GS as part of the 100KGP revealed a single pathogenic variant in the coding region of the *BBS10* gene: GRCh38 chr12:76345866_76345867del, c.2119_2120delGT, p.Val707* (NM_024685.4). To identify a potential pathogenic trans allele, an interrogation of the non-coding regions of the *BBS10* gene was performed, revealing a variant in the upstream region, Chr12:76348438dupG c.-80dupC (NM_024685.4), absent from the gnomAD v2.1 dataset. Given the position of this variant in the 5'UTR, we hypothesised a possible regulatory effect. Inspection of the *BBS10* promoter region at the Eukaryotic Promoter Database (EPD;

https://epd.epfl.ch/EPDnew_database.php), indicated the presence of two probable regulatory elements. EPD#1 is located at c.-106_c.-58, and EPD#2 is located at c.-829_c.-782, with the c.-80dupC variant located within EPD#1.

A series of *BBS10* promoter constructs driving firefly luciferase expression based on the pGL3 vector were made (supplementary methods and Figure S1). The BBS10(500bp) promoter construct contained only EPD#1, whilst BBS10(1kb) contained both EPD#1 and 2. When transfected into HEK293 cells, robust firefly luciferase expression was observed from both wild-type *BBS10* promoters - typically >5x that observed for the SV40 based pGL3-Control (data not shown). Relative to the BBS10(500bp) promoter, the BBS10(1kb) promoter displayed ~80 % activity (Figure 6). Introduction of the c.-80dupC variant led to a ~70 % decrease in promoter activities, compared to their respective wild-type counterparts. This suggested that EPD#1 has a significant role in the expression of *BBS10* that is affected by c.-80dupC. Analysis of other variants around the c.-80dupC position, especially c.-83T>G, further support the importance of this region as transcriptionally active (see Figure S1). This was further confirmed by complete deletion of EPD#1 (ΔEPD#1) in both the BBS10(500bp) and BBS10(1kb) constructs, with firefly luciferase levels falling to ~2.5 % and ~1.5 % of the respective wild-type (background activity of pGL3-Basic is ~0.6 %). Interestingly, EPD#2 appears to have an inhibitory effect on the BBS10 promoter in this system, with the BBS10(1kb) ΔEPD#2 variant exhibiting firefly luciferase expression at nearly double that of the wild-type BBS10(1kb).

Oxford Nanopore Technologies (ONT) long-read sequencing of PCR amplicons spanning the entire genomic region (3kb) enabled phasing of the two variants to demonstrate they are *in trans*, in the absence of any family members for segregation analysis.

Patient 9 (GC 21400) was recruited for GS with both unaffected parents through the 100KGP. Clinically, he was noticed to have poor vision and nystagmus in his first year of life, which remained stable over his lifetime. He had good general health. His retinal exam showed a featureless fundus and macular atrophy (Figure 1F). Full-field electroretinogram revealed poor rod and cone responses, and he was diagnosed with EOSRD. A single heterozygous missense c.2837C>A, p.(Ala946Glu) (chr17:8015395C>A, NM_000180.3) variant was identified in the *GUCY2D* gene. All other IRD genes had been excluded based on complete coding variant analysis. Non-coding variant analysis revealed a SNV upstream of the start codon of the *GUCY2D* gene (chr17:8002596T>C, *GUCY2D* c.-148T>C) shared with an unrelated affected individual recruited to the study independently (patient 11, see supplementary data). Both individuals had a trans allele harbouring a candidate pathogenic variant in the gene (c.2837C>A, p.(Ala946Glu) and c.3043+5G>A, respectively). Therefore, the upstream variant was considered a good candidate in these unrelated cases. Analysis of transcription factor binding site motifs in the region (http://alggen.lsi.upc.es) showed 4 CRX binding sites in the vicinity of c.-148 (CBE#1-4, supplementary data). The variant falls within the core 5'-TAAT'-3 sequence of CBE#1 (CTCT**A**ATTA > CTCT**G**ATTA) so we hypothesised that the variant may disrupt CRX binding to the *GUCY2D* regulatory region. Therefore, a series of *GUCY2D* upstream constructs were created comprising a 1kb upstream region including the c.-148C>T variant, and deletions of the CBE#1-4 (supplementary data). Luciferase reporter assays in HEK293 cells required co-transfection with a *CRX* expression plasmid to induce expression from the *GUCY2D* promoter region, which demonstrated that the variant reduced expression by ~25% compared to wild-type and the effect was greater still in constructs with CBE#2-4 deleted with a reduction of ~40% expression with CBE#1 in isolation (see supplementary methods and Figure S2).

Discussion

Application of NGS for investigation of Mendelian disorders is now widely accepted. When restricted to the coding regions of known genes, this fails to identify the complete causative genotype in many individuals. We report the likely disease-associated genotype of 9 individuals who underwent GS and virtual gene-panel testing, followed by MDT-led targeted candidate gene non-coding variant analysis based on clinical and genetic data. In addition, two patients harbouring recurrent variants in *CRB1* and *GUCY2D* were identified in other centres (patients 10 and 11).

Intronic mutations can alter canonical donor and acceptor splice sites and create cryptic splice sites, causing mis-splicing events such as exon skipping, pseudoexon inclusion, or intron retention.[22] Unless already proven pathogenic or found within canonical splice motifs at the intron/exon boundary, these variants may escape detection by standard sequencing approaches, requiring GS re-analysis and representing a diagnostic challenge.[23] Nevertheless, due to the high prevalence of IRD variant carriers,[2] analysing full genes to detect the second hit in IRD patients with monoallelic variants does not seem the optimal approach.[24] We demonstrate here that narrowing the search by an MDT-prompted diagnostic awareness can help target the analysis to a small number of candidate genes and variants.

This study led not only to the finding of novel, intronic, damaging variants, but also to molecular diagnosis for the 11 patients studied here. We hypothesise that the spectrum of non-coding pathogenic variants may be somewhat limited. For example, for a pseudoexon to become active, the complete splicing architecture must be present; meaning that creating a deep intronic donor or acceptor site alone may have a limited effect in the absence of a partner site and thus the possibility of a single variant to accomplish this is indeed limited. A clue towards this from the current study is that the deep intronic variant in *CRB1,* c.3879-

1203C>G, was found to be recurrent on independent genetic backgrounds. Previous findings in *ABCA4* suggest that pathogenic variants may occur at splice junctions of potential pseudoexons, detectable by RNA studies.[25,26] Indeed, intronic pathogenic changes have been reported in many genes, with some recurrent or prevalent variants in *ABCA4*, *CEP290* and *USH2A*,[26–31] further suggesting that the spectrum of intronic mutations is far more limited than that of coding mutations. Thus, it is possible that the intronic variants which are able to disrupt splicing are clustered at certain hotspots within the introns and many seemingly strong deep intronic splice sites may not lead to pseudoexon inclusion at all. Advanced splice prediction tools such as SpliceAI will prove useful to demonstrate this.[32,33]

Proving that non-coding variants are damaging requires input from an expert MDT, supportive clinical data and functional studies. Undertaking such work is currently possible only through research projects and for high priority cases, but should thereby provide enough evidence for the variant or region to be included in future diagnostic testing, ultimately leading to improved diagnostic yield for IRD. The relevance/prevalence of non-coding variants has been highlighted by Sangermano *et al*. and Khan *et al*., who found intronic *ABCA4* changes in 67% and 25% of the probands within their cohorts of unsolved Stargardt (STGD) and STGD-like individuals, respectively.[34,35] *CEP290* c.2991+1655A>G is the most commonly found pathogenic variant in this gene, with 60% to 90% of the individuals with *CEP290*-related Leber Congenital Amaurosis (LCA) having it in at least one allele (active clinical trials are using CRISPR/Cas9 - NCT03872479 - and antisense oligonucleotide - NCT03913143 - technologies to target this particular variant).[36] Similarly, *USH2A* c.7595-2144A>G was reported with a frequency of 4% among an Usher syndrome type 2A (USH2A)-patient cohort, representing the second most frequent *USH2A*-causing variant.[31] Identifying and understanding these changes allows us to develop targeted therapies, which could correct aberrant transcripts.[37]

Another potential site for pathogenic variants is the cis-acting regulatory sequence. The upstream region and 5′ UTR of any gene usually contains the promoter, enhancer, regulatory sites, and transcription factor binding motifs, thus being important for gene regulation.[38] Variants located in these regions have been found to be damaging in IRD genes such as *CHM*, *NMNAT1*, *EYS*, *LCA5*, *PRPF31*, *PRPF4*, and *PCDH15*.[39–45] In particular, *NMNAT1* has been found to harbour a hotspot for 5'UTR variants.[40] In this report, we have identified a damaging upstream variant in *BBS10* and an upstream variant in *GUCY2D* that disrupts binding of the transcription factor CRX. The trans BBS10 allele, p.Val707Ter, has previously been associated with classical BBS,[46] and it would appear that the c.-80dupC allele is a hypomorph. As such, it would appear that ~15% of 2N levels of wild-type BBS10 are sufficient to prevent manifestation of extraocular BBS10 related features. Similarly, reduced levels of wild-type GUCY2D expression may be related to the unusual presentation associated with the expected pathogenic p.Ala946Glu trans allele.[47] The 3'UTR region, located downstream from the stop codon, serves as a binding site for micro RNAs, which can also affect gene expression.[48] Variants located in these regions were found to be damaging in Tourette's syndrome (*SLITRK1*)[49] and in Phosphoglycerate kinase 1 deficiency (*PGK1*),[50] however these events are certainly rare and supporting evidence regarding their pathogenicity is hard to gather. Although such sites for pathogenic variants are likely to be even more limited than coding and cryptic splicing variants, these might still represent an important cause of disease.

Key to the analysis pipeline and discovery of candidate variants is the MDT role, which facilitates the interaction between the clinical team, genetic counsellors, clinical scientists, and specialists in ophthalmic genetics. Interpretation of genomic data in isolation can miss vital clues for the elucidation of complex cases. An integrated clinical/genomic data analysis

pipeline as a second stage following simple variant discovery/interpretation, including detailed medical history, retinal imaging, and functional testing, in combination with targeted variant interrogation, and input of experts in ophthalmology and genomics, can lead to a diagnostic uplift in cases that would otherwise remain unsolved.[20,51] This synergy between clinicians and specialists in genetics was also noted to increase the diagnostic yield in other areas such as pulmolonology,[52] neurogenomics,[53] nephrology,[54] and prenatal diagnosis.[55] Regular scheduled MDT meetings to discuss potential clinical diagnostic tests and analyse genetic results is undoubtedly of good practice and leads to an increased quality of patient care, with faster, more efficient diagnoses, and contributing to new disease-causing gene discovery. Nonetheless, many molecular laboratories, in particular commercial and service laboratories, may present difficulties exploiting detailed clinical data, therefore missing out on critical information that can highlight gene targets, and drive the analysis.

In conclusion, our work provides novel non-coding variants in several IRD genes, that will enrich retinal panels and hopefully contribute to decrease the missing genetic etiology in this field. It also underscores the immense benefit of MDT, decreasing diagnostic time and cost. Furthermore, we pose a question with regard to the diversity of viable intronic splice changes. Future worldwide collaborations will help determine if said hypothesis stands.

Materials and Methods

This study adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board and ethics committee of Moorfields Eye Hospital (REC 12/LO/0141). Informed consent was obtained from all participants prior to the inclusion in the study. For patients and relatives recruited for the 100KGP, informed consent for GS was obtained in accordance with approval from the HRA committee East of England-Cambridge

south (REC 14/EE/1112). For patients recruited to the NIHR-RD study, participants provided written informed consent and the study was approved by the East of England Cambridge South national institutional review board (13/EE/0325).

All patients underwent clinical variant interpretation using virtual gene panel analysis and MDT discussion. Patients were selected for study based on having a clinical presentation that was likely to correspond to a particular gene and, in the case of recessive diseases, a heterozygous pathogenic or likely pathogenic variant in a gene that was deemed highly likely to harbour the disease variant.

Subsequent non-coding variant analysis was performed on the gene of interest in all available family members' GS data to establish phase, where possible. Rare variants (allele frequency ≤0.001) were identified and considered to be candidate pathogenic changes, variants were further manually curated for quality (using the Integrative genome viewer, IGV[56]), low complexity, repetitive and polymorphic regions (using gnomAD). *In silico* analysis was performed on surviving variants including splice prediction (nnsplice, SpliceAI), conservation score, and regulatory/promoter region prediction, depending on the variant position in the gene. Where a single variant was identified, no specific prediction threshold was applied for splice prediction analysis.

Compelling variants were functionally assayed using PAXgene stabilised whole blood for mRNA transcript analysis to identify mis-splicing in ubiquitously expressed genes (*PRPF31, IFT140*) or mRNA transcript analysis on total RNA extracted from nasal epithelial cells (*USH2A*), where expression is absent in whole blood. Minigene analysis was performed for retina-specific genes (*CRB1*) and luciferase reporter gene assays was undertaken for putative regulatory region variants (*BBS10, GUCY2D*). Methods can be found in the supplementary data.

ONT single molecule sequencing was performed on PCR amplified gDNA to phase variants. Briefly, PCR amplification was performed with oligonucleotide primers spanning the *BBS10* gene and reactions were purified for ONT library preparation with AMPure XP magnetic beads (Beckman Coulter). Up to 50ng of purified PCR product was used for library preparation using the SQK-LSK109 kit (Oxford Nanopore Technologies, Oxford, UK). Libraries were run on Flongle (Oxford Nanopore Technologies, Oxford, UK) flowcells for approximately 12 hours. Reads were basecalled with Guppy v2 (Oxford Nanopore Technologies, Oxford, UK) and aligned to the human genome build GRCh38 using Minimap2 (https://github.com/lh3/minimap2). Samtools (https://samtools.github.io) was used to generate indexed sorted BAM files for visualisation of individual read data using IGV.

Ophthalmic examination included visual acuity (using logMAR visual acuity charts), spectral domain optical coherence tomography (Spectralis, Heidelberg Engineering Ltd, Heidelberg, Germany), ultra-widefield colour fundus photography (200°, Optos plc, Dunfermline, UK), and fundus autofluorescence imaging, performed with 55° Spectralis or ultra-widefield Optos. Four patients had visual electrophysiology testing, including full-field and pattern electroretinography, which incorporated the International Society for Clinical Electrophysiology of Vision (ISCEV) standards.[57]

Data availability

Further details of the GS and ES data presented in the study are available via direct contact with the corresponding author. Data accessibility information for the 100KGP is available online (www.genomicsengland.co.uk/join-a-gecip-domain).

Acknowledgements

This work was supported by Moorfields Eye Charity (GR001203 and Stephen and Elizabeth Archer in memory of Marion Woods), National Eye Research Centre (SAC051), Fight for Sight (UK), National Institute of Health Research Biomedical Research Centre (NIHR-BRC) at Moorfields Eye Hospital and UCL Institute of Ophthalmology. GA is funded by a Fight For Sight UK Early Career Investigator Award (5045/46) and NIHR-BRC at Great Ormond Street Hospital Institute for Child Health. This research was made possible through access to the data and findings generated by The 100,000 Genomes Project. The 100,000 Genomes Project is 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. This work was also supported by the Swiss National Science Foundation (grants # 176097 and # 204285 to C.R.).

Competing Interests

The authors alone are responsible for the content and writing of this article. The authors have no conflict of interests.

References

1. Pontikos N, Arno G, Jurkute N, Schiff E, Ba-Abbad R, Malka S, Gimenez A, Georgiou M, Wright G, Armengo M, et al. Genetic Basis of Inherited Retinal Disease in a Molecularly Characterized Cohort of More Than 3000 Families from the United Kingdom. *Ophthalmology*. 2020;127(10):1384-1394. doi:10.1016/j.ophtha.2020.04.008

2. Hanany M, Rivolta C, Sharon D. Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases. *Proc Natl Acad Sci U S A*. 2020;117(5):2710-2716. doi:10.1073/pnas.1913179117

3. Galvin O, Chi G, Brady L, Hippert C, Del Valle Rubido M, Daly A, Michaelides M. The Impact of Inherited Retinal Diseases in the Republic of Ireland (ROI) and the United Kingdom (UK) from a Cost-of-Illness Perspective. *Clin Ophthalmol*. 2020;14:707-719. doi:10.2147/OPTH.S241928

4. Tsui I, Song BJ, Lin CS, Tsang SH. A Practical Approach to Retinal Dystrophies. *Adv Exp Med Biol*. 2018;1085:245-259. doi:10.1007/978-3-319-95046-4_51

5. Daich Varela M, Guimaraes T, Georgiou M, Michaelides M. Leber Congenital Amaurosis/Early-Onset Severe Retinal Dystrophy: Current Management and Clinical Trials. *Br J Ophthalmol*. Published online 2021.

6. Stingl K, Schippert R, Bartz-Schmidt KU, Besch D, Cottriall CL, Edwards TL, Gekeler F, Greppmaier U, Kiel K, Koitschev A, et al. Interim Results of a Multicenter Trial with the New Electronic Subretinal Implant Alpha AMS in 15 Patients Blind from Inherited Retinal Degenerations. *Front Neurosci*. 2017;11:445. doi:10.3389/fnins.2017.00445

7. Singh MS, Park SS, Albini TA, Canto-Soler MV, Klassen H, MacLaren RE, Takahashi M, Nagiel A, Schwartz SD, Bharti K. Retinal stem cell transplantation: Balancing safety and potential. *Prog Retin Eye Res*. 2020;75:100779. doi:10.1016/j.preteyeres.2019.100779

8. Scholl HPN, Strauss RW, Singh MS, Dalkara D, Roska B, Picaud S, Sahel J. Emerging therapies for inherited retinal degeneration. *Sci Transl Med*. 2016;8(368):368rv6. doi:10.1126/scitranslmed.aaf2838

9. Bainbridge JWB, Mehat MS, Sundaram V, Robbie SJ, Barker SE, Ripamonti C, Georgiadis A, Mowat FM, Beattie SG, Gardneret PJ, et al. Long-term effect of gene therapy on Leber's congenital amaurosis. *N Engl J Med*. 2015;372(20):1887-1897. doi:10.1056/NEJMoa1414221

10. Russell S, Bennett J, Wellman JA, Chung DC, Yu Z, Tillman A, Wittes J, Pappas J, Elci O, McCague S, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet (London, England)*. 2017;390(10097):849-860. doi:10.1016/S0140-6736(17)31868-8

11. Peddle CF, Fry LE, McClements ME, MacLaren RE. CRISPR Interference-Potential Application in Retinal Disease. *Int J Mol Sci*. 2020;21(7). doi:10.3390/ijms21072329

12. Xue K, MacLaren RE. Antisense oligonucleotide therapeutics in clinical trials for the treatment of inherited retinal diseases. *Expert Opin Investig Drugs*. 2020;29(10):1163-1170. doi:10.1080/13543784.2020.1804853

13. Oprych K, Silva RS, Pontikos N, Arno G. 12 Genome Analysis for Inherited Retinal Disease: The State of the Art. *Adv Vis Res Vol III Genet Eye Res around Globe*. Published online 2021:153.

14. Ellingford JM, Barton S, Bhaskar S, Williams SG, Sergouniotis PI, O'Sullivan J, Lamb JA, Perveen R, Hall G, Newman WG, et al. Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. *Ophthalmology*. 2016;123(5):1143-1150. doi:10.1016/j.ophtha.2016.01.009

15. Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S, et al. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet*. 2017;100(1):75-90. doi:10.1016/j.ajhg.2016.12.003

16. Perea-Romero I, Gordo G, Iancu IF, Del Pozo-Valero M, Almoguera B, Blanco-Kelly F, Carreño E, Jimenez-Rolando B, Lopez-Rodriguez R, Lorda-Sanchez I, Martin-Merida I, et al. Genetic landscape of 6089 inherited retinal dystrophies affected cases in Spain and their therapeutic and extended epidemiological implications. *Sci Rep*. 2021;11(1):1526. doi:10.1038/s41598-021-81093-y

17. Stone EM, Andorf JL, Whitmore SS, DeLuca AP, Giacalone JC, Streb LM, Braun TA, Mullins RF, Scheetz TE, Sheffield VC, et al. Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease. *Ophthalmology*. 2017;124(9):1314-1331. doi:10.1016/j.ophtha.2017.04.008

18. Gonzàlez-Duarte R, de Castro-Miró M, Tuson M, Ramírez-Castañeda V, Gils RV, Marfany G. Scaling New Heights in the Genetic Diagnosis of Inherited Retinal Dystrophies. *Adv Exp Med Biol*. 2019;1185:215-219. doi:10.1007/978-3-030-27378-1_35

19. Jamshidi F, Place EM, Mehrotra S, Navarro-Gomez D, Maher M, Branham KE, Valkanas E, Cherry TJ, Lek M, MacArthur D, et al. Contribution of noncoding

pathogenic variants to RPGRIP1-mediated inherited retinal degeneration. *Genet Med*. 2019;21(3):694-704. doi:10.1038/s41436-018-0104-7

20. Zeitz C, Michiels C, Neuillé M, Friedburg C, Condroyer C, Boyard F, Antonio A, Bouzidi N, Milicevic D, Veaux R, et al. Where are the missing gene defects in inherited retinal disorders? Intronic and synonymous variants contribute at least to 4% of CACNA1F-mediated inherited retinal disorders. *Hum Mutat*. 2019;40(6):765-787. doi:10.1002/humu.23735

21. Van Cauwenbergh C, Van Schil K, Cannoodt R, Bauwens M, Van Laethem T, De Jaegere S, Steyaert W, Sante T, Menten B, Leroy BP, et al. arrEYE: a customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. *Genet Med*. 2017;19(4):457-466. doi:10.1038/gim.2016.119

22. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2016;17(1):19-32. doi:10.1038/nrg.2015.3

23. Khan AO, Becirovic E, Betz C, Neuhaus C, Altmüller J, Riedmayr LM, Motameny S, Nürnberg G, Nürnberg P, Bolz HJ. A deep intronic CLRN1 (USH3A) founder mutation generates an aberrant exon and underlies severe Usher syndrome on the Arabian Peninsula. *Sci Rep*. 2017;7(1):1411. doi:10.1038/s41598-017-01577-8

24. González-Del Pozo M, Martín-Sánchez M, Bravo-Gil N, Méndez-Vidal C, Chimenea A, Rodríguez-de la Rúa E, Borrego S, Antiñolo G. Searching the second hit in patients with inherited retinal dystrophies and monoallelic variants in ABCA4, USH2A and CEP290 by whole-gene targeted sequencing. *Sci Rep*. 2018;8(1):13312. doi:10.1038/s41598-018-31511-5

25. Braun TA, Mullins RF, Wagner AH, Andorf JL, Johnston RM, Bakall BB, Deluca AP,

Fishman GA, Lam BL, Weleber RG, et al. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet*. 2013;22(25):5136-5145. doi:10.1093/hmg/ddt367

26. Khan M, Arno G, Fakin A, Parfitt DA, Dhooge PPA, Albert S, Bax NM, Duijkers L, Niblock M, Hau KL, et al. Detailed Phenotyping and Therapeutic Strategies for Intronic ABCA4 Variants in Stargardt Disease. *Mol Ther Nucleic Acids*. 2020;21:412-427. doi:10.1016/j.omtn.2020.06.007

27. Bauwens M, Garanto A, Sangermano R, Naessens S, Weisschuh N, De Zaeytijd J, Khan M, Sadler F, Balikova I, Van Cauwenbergh C, et al. ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet Med*. 2019;21(8):1761-1771. doi:10.1038/s41436-018-0420-y

28. den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, Voesenek KEJ, Zonneveld MN, Strom TM, Meitinger T, Brunner HG, et al. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet*. 2006;79(3):556-561. doi:10.1086/507318

29. Vaché C, Besnard T, le Berre P, García-García G, Baux D, Larrieu L, Abadie C, Blanchet C, Jörn Bolz H, Millan J, et al. Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy. *Hum Mutat*. 2012;33(1):104-108. doi:10.1002/humu.21634

30. Liquori A, Vaché C, Baux D, Blanchet C, Hamel C, Malcolm S, Koenig M, Claustres M, Roux A. Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. *Hum Mutat*. 2016;37(2):184-193. doi:10.1002/humu.22926

31. Slijkerman RW, Vaché C, Dona M, García-García G, Claustres M, Hetterschijt L,

Peters TA, Hartel BP, Je Pennings R, Millan JM, et al. Antisense Oligonucleotide-based Splice Correction for USH2A-associated Retinal Degeneration Caused by a Frequent Deep-intronic Mutation. *Mol Ther Nucleic Acids*. 2016;5(10):e381. doi:10.1038/mtna.2016.89

32. Rowlands C, Thomas HB, Lord J, Wai HA, Arno G, Beaman G, Sergouniotis P, Gomes-Silva B, Campbell C, Gossan N, et al. Comparison of in silico strategies to prioritize rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. *Sci Rep*. 2021;11(1):20607. doi:10.1038/s41598-021-99747-2

33. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Fazel Darbandi S, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019;176(3):535-548.e24. doi:10.1016/j.cell.2018.12.015

34. Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, van den Born LI, Imran Khan M, Cornelis SS, Verheij JBGM, et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med*. 2019;21(8):1751-1760. doi:10.1038/s41436-018-0414-9

35. Khan M, Cornelis SS, Pozo-Valero M Del, Whelan L, Runhart EH, Mishra K, Bults F, AlSwaiti Y, AlTalbishi A, De Baere E, et al. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet Med*. 2020;22(7):1235-1246. doi:10.1038/s41436-020-0787-4

36. Dulla K, Aguila M, Lane A, Jovanovic K, Parfitt DA, Schulkens I, Lam Chan H, Schmidt I, Beumer W, Vorthoren L, et al. Splice-Modulating Oligonucleotide QR-110 Restores CEP290 mRNA and Function in Human c.2991+1655A>G LCA10 Models.

Mol Ther Nucleic Acids. 2018;12:730-740. doi:10.1016/j.omtn.2018.07.010

37. Hammond SM, Wood MJA. Genetic therapies for RNA mis-splicing diseases. *Trends Genet*. 2011;27(5):196-205. doi:10.1016/j.tig.2011.02.004

38. Davuluri R V, Grosse I, Zhang MQ. Computational identification of promoters and first exons in the human genome. *Nat Genet*. 2001;29(4):412-417. doi:10.1038/ng780

39. Radziwon A, Arno G, K Wheaton D, McDonagh EM, Baple EL, Webb-Jones K, Birch DG, Webster AR, MacDonald IM. Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum Mutat*. 2017;38(6):704-715. doi:10.1002/humu.23212

40. Coppieters F, Todeschini AL, Fujimaki T, Baert A, De Bruyne M, Van Cauwenbergh C, Verdin H, Bauwens M, Ongenaert M, Kondo M, et al. Hidden Genetic Variation in LCA9-Associated Congenital Blindness Explained by 5'UTR Mutations and Copy-Number Variations of NMNAT1. *Hum Mutat*. 2015;36(12):1188-1196. doi:10.1002/humu.22899

41. Van Schil K, Naessens S, Van de Sompele S, Carron M, Aslanidis A, Van Cauwenbergh C, Mayer AK, Van Heetvelde M, Bauwens M, Verdin H, et al. Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. *Genet Med*. 2018;20(2):202-213. doi:10.1038/gim.2017.97

42. Eisenberger T, Neuhaus C, Khan AO, Decker C, Preising MN, Friedburg C, Bieg A, Gliem M, Issa PC, Holz FG, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PLoS One*. 2013;8(11):e78496. doi:10.1371/journal.pone.0078496

43. den Hollander AI, Koenekoop RK, Mohamed MD, Arts HH, Boldt K, Towns KV, Sedmak T, Beer M, Nagel-Wolfrum K, McKibbin M, et al. Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis. *Nat Genet*. 2007;39(7):889-895. doi:10.1038/ng2066

44. Rose AM, Shah AZ, Waseem NH, Chakarova CF, Alfano G, Coussa RG, Ajlan R, Koenekoop RK, Bhattacharya SS. Expression of PRPF31 and TFPT: regulation in health and retinal disease. *Hum Mol Genet*. 2012;21(18):4126-4137. doi:10.1093/hmg/dds242

45. Chen X, Liu Y, Sheng X, Tam POS, Zhao K, Chen X, Rong W, Liu Y, Liu X, Pan X, et al. PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum Mol Genet*. 2014;23(11):2926-2939. doi:10.1093/hmg/ddu005

46. Scheidecker S, Hull S, Perdomo Y, Studer F, Pelletier V, Muller J, Stoetzel C, Schaefer E, Defoort-Dhellemmes S, Drumare I, et al. Predominantly Cone-System Dysfunction as Rare Form of Retinal Degeneration in Patients With Molecularly Confirmed Bardet-Biedl Syndrome. *Am J Ophthalmol*. 2015;160(2):364-372.e1. doi:10.1016/j.ajo.2015.05.007

47. Taylor RL, Parry NRA, Barton SJ, Campbell C, Delaney CM, Ellingford JM, Hall G, Hardcastle C, Morarji J, Nichol EJ, et al. Panel-Based Clinical Genetic Testing in 85 Children with Inherited Retinal Disease. *Ophthalmology*. 2017;124(7):985-991. doi:10.1016/j.ophtha.2017.02.005

48. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116(2):281-297. doi:10.1016/s0092-8674(04)00045-5

49. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin M, Gunel M, et al. Sequence variants in SLITRK1 are associated

with Tourette's syndrome. *Science*. 2005;310(5746):317-320. doi:10.1126/science.1116502

50. Behlmann AM, Goyal NA, Yang X, Chen PH, Ankala A. A Hemizygous Deletion Within the PGK1 Gene in Males with PGK1 Deficiency. *JIMD Rep*. 2019;45:105-110. doi:10.1007/8904_2018_147

51. Di Scipio M, Tavares E, Deshmukh S, Audo I, Green-Sanderson K, Zubak Y, Zine-Eddine F, Pearson A, Vig A, Yu Tang C, et al. Phenotype Driven Analysis of Whole Genome Sequencing Identifies Deep Intronic Variants that Cause Retinal Dystrophies by Aberrant Exonization. *Invest Ophthalmol Vis Sci*. 2020;61(10):36. doi:10.1167/iovs.61.10.36

52. Grimes HL, Holden S, Babar J, Karia S, Ta Wetscherek M, Barker A, Herre J, Knolle MD, Maher ER, Genomics England Research Consortium; et al. Combining clinical, radiological and genetic approaches to pneumothorax management. *Thorax*. Published online June 2021. doi:10.1136/thoraxjnl-2021-217210

53. McLean A, Tchan M, Devery S, Smyth R, Kumar K, Tomlinson S, Tisch S, Wu K. 087 An integrated neurogenomics clinic–28-months experience and outcome of a tertiary referral centre. BMJ Neurology Open 2021;3:doi: 10.1136/bmjno-2021-ANZAN.87

54. Jayasinghe K, Stark Z, Kerr PG, Gaff C, Martyn M, Whitlam J, Creighton B, Donaldson E, Hunter M, Jarmolowicz A, et al. Clinical impact of genomic testing in patients with suspected monogenic kidney disease. *Genet Med*. 2021;23(1):183-191. doi:10.1038/s41436-020-00963-4

55. Mone F, O'Connor C, Hamilton S, Quinlan-Jones E, Allen S, Marton T, Williams D, Kilby MD. Evolution of a prenatal genetic clinic-A 10-year cohort study. *Prenat*

Diagn. 2020;40(5):618-625. doi:10.1002/pd.5661

56. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26. doi:10.1038/nbt.1754

57. McCulloch DL, Marmor MF, Brigell MG, Hamilton R, Holder GE, Tzekov R, Bach M. ISCEV Standard for full-field clinical electroretinography (2015 update). *Doc Ophthalmol*. 2015;130(1):1-12. doi:10.1007/s10633-014-9473-7

Legends

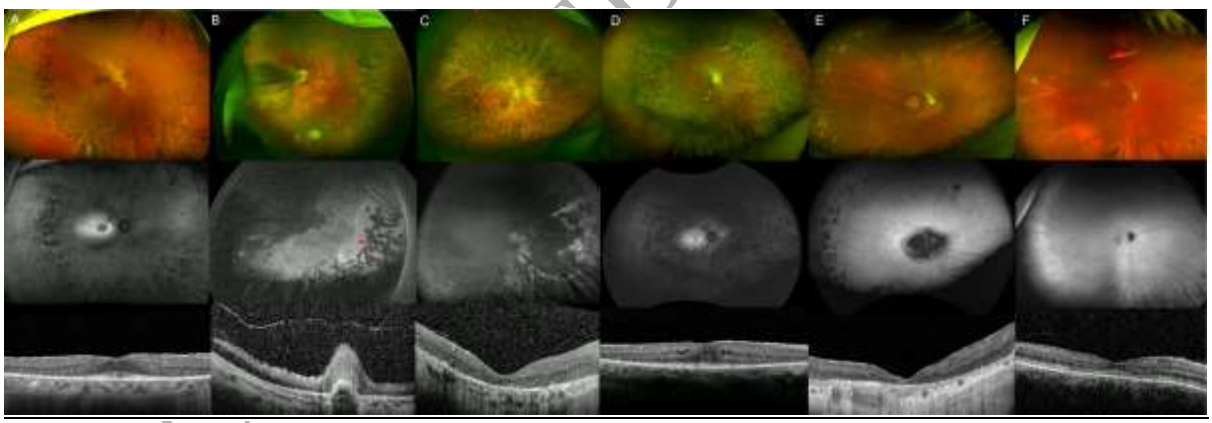

Figure 1: Retinal imaging from individuals with IRD within the analysed cohort. A) Ultrawide-field (UWF) colour fundus image showing pigmented bone spicules-like lesions and vessel thinning in patient 1, with *PRPF31*-associated rod-cone dystrophy (RCD). Fundus autofluorescence (FAF) imaging is positive for a macular hyperautofluorescent ring, characteristic of rod-cone dystrophies (RCD), and generalized decreased hypoAF. Macular optical coherence tomography (OCT) shows a subfoveal island of outer layers. B) UWF

colour and FAF images from patient 4, a child with Norrie disease and bilateral retinal folds. Prophylactic bilateral pan-retinal photocoagulation spots are also visible, marked with red arrows. Macular OCT shows a retinal fold and poor retinal architecture. C) UWF fundus and OCT imaging from patient 6, who has *CRB1*-early onset severe retinal dystrophy. Retinal images are positive for nummular, dense, deep pigment deposition, preserved autofluorescence adjacent to few peripheral retinal arterioles, and a poorly laminated retina. D) UWF colour and autofluorescence imaging of patient 7, showing typical RCD features associated with *USH2A* retinopathy. We can see dense pigmented deposits in the mid-periphery, a perifoveal hyperautofluorescent ring demarcating the area of functioning retina, and decreased peripheral autofluorescence. Macular OCT shows a subfoveal island of outer segments and cystoid macular oedema. E) UWF colour, autofluorescence, and OCT imaging from patient 8, with *BBS10*-associated retinopathy. Retinal appearance shows a CRD pattern, with demarcated posterior pole and peripheral areas of retinal pigment epithelium atrophy. Macular OCT appears well correlated, with generalised loss of the outer layers. F) UWF colour, autofluorescence, and OCT images of patient 9, diagnosed with *GUCY2D*-related EOSRD. Of note is the featureless fundus, mild vessel narrowing, macular atrophy, and disrupted ellipsoid zone centrally.

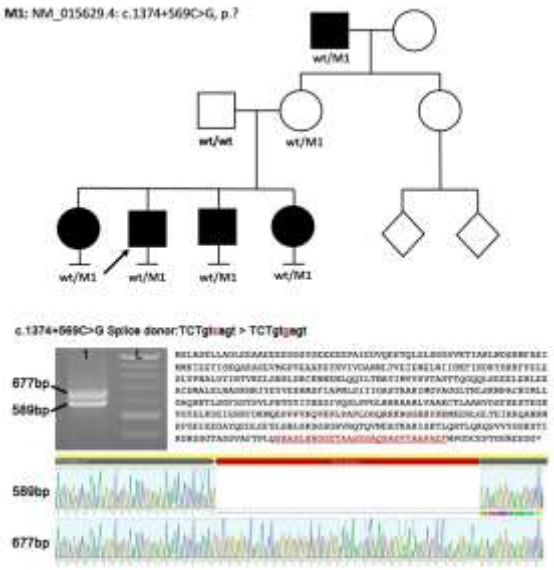

Figure 2: Analysis of mis-splicing due to *PRPF31* c.1374+569C>G. Pedigree of patient 1 showing dominant family history with incomplete penetrance. Reverse transcription polymerase chain reaction (RT-PCR) and direct sequencing of patient 1 *PRPF31* transcript (exon 8-14) derived from PAXgene RNA. Agarose gel analysis showing two distinct bands from RT-PCR amplification in patient sample. L: ladder. Direct sequencing of amplicons shows inclusion of 88bp pseudoexon leading to a frameshift and premature termination.

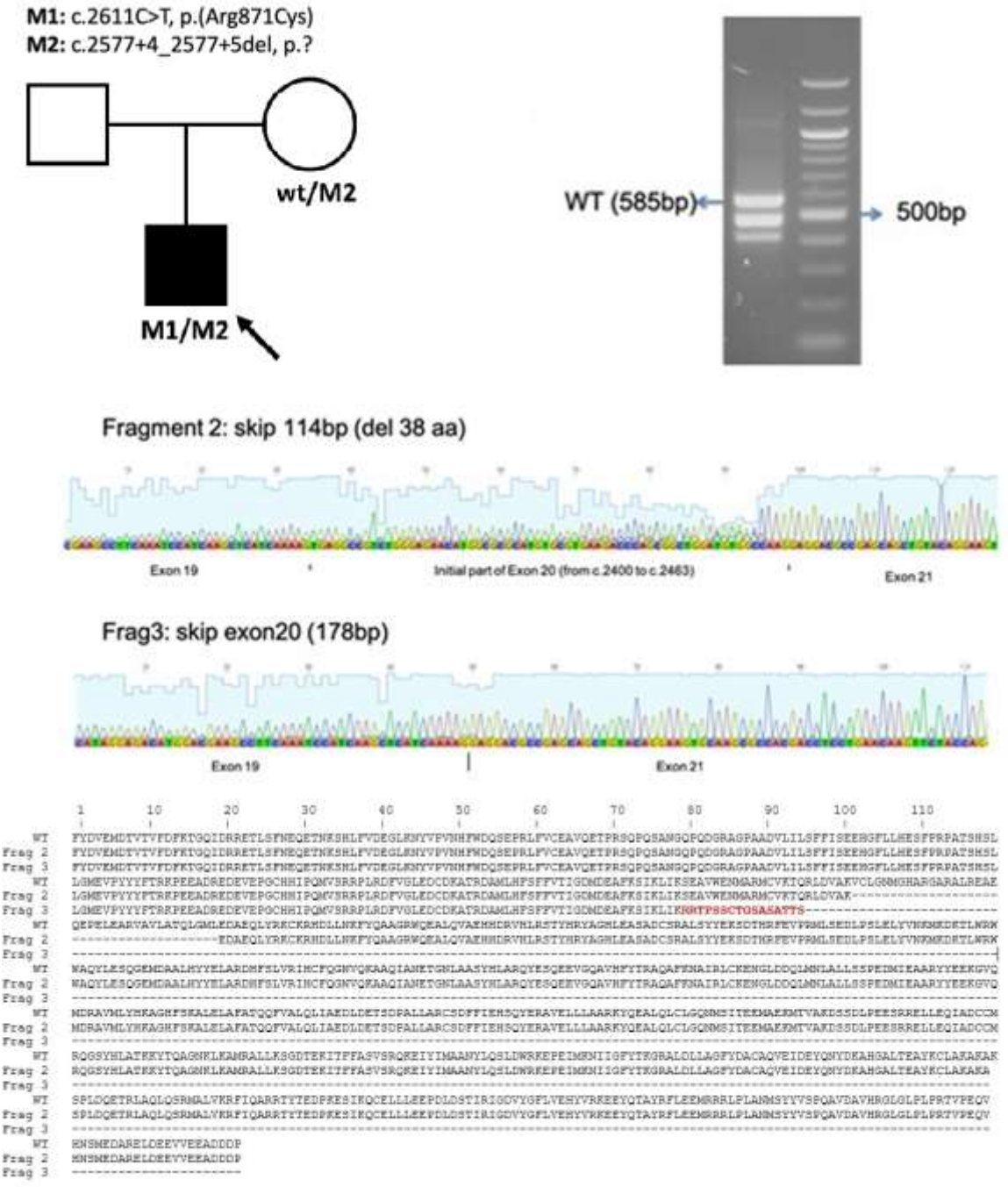

Figure 3: Analysis of mis-splicing due to *IFT140* c.2577+4_2577+5del. Pedigree of patient 3. Nested RT-PCR showing multiple fragments amplified from patient 3 cDNA derived from PAXgene RNA sample. Bands [lane 3 (patient) fragment 1 (wild-type, WT), 2 and 3, L: ladder] were purified and sequenced by direct Sanger sequencing. Fragment 2 and 3 showed

partial skipping of exon 20 and complete skipping of exon 20 respectively with clustal alignment of the fragments shown.

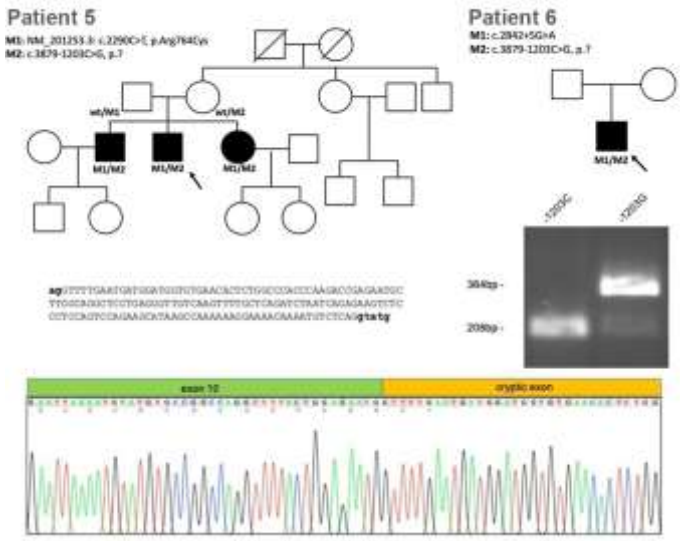

Figure 4: Minigene analysis of *CRB1* c.3879-1203C>G. Pedigrees of patient 5 and patient 6. HEK293 cells transfected with CRB1-wt or CRB1-1203C>G containing minigene constructs. The PCR analysis shows splicing in of a 156bp cryptic exon included in the CRB1-1203C>G transfected cells not present in the wild-type cells.

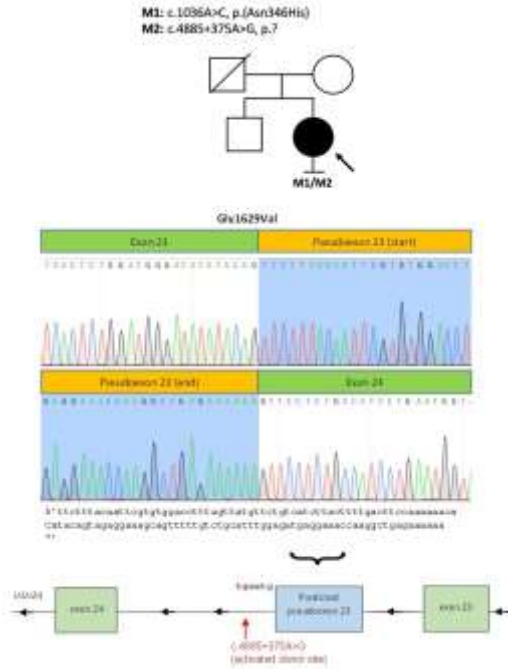

Figure 5: Cryptic exon inclusion consequent upon *USH2A* c.4885+375A>G. Prediction of pseudoexon inclusion due to a strengthened deep intronic splice donor site (*USH2A* is on the reverse strand). Direct Sanger sequencing of PCR amplicons showed an alternate splice product comprising inclusion of the predicted pseudoexon (130bp), frameshift and premature stop codon (p.Gly1629ValfsTer52).

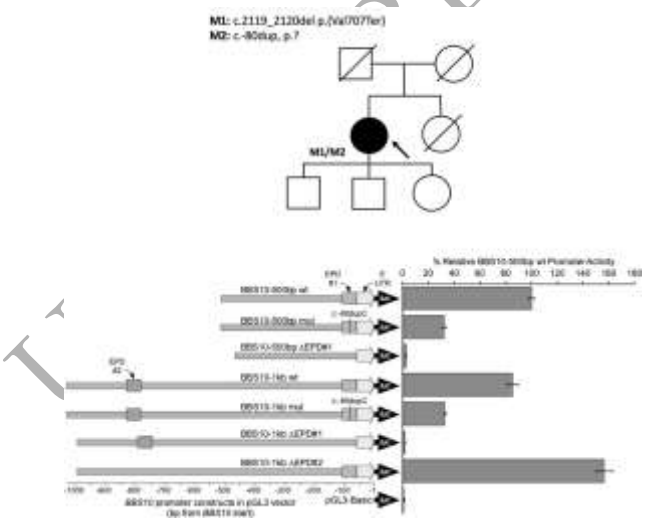

Figure 6: *BBS10* promoter constructs and activity assayed by firefly luciferase expression in HEK293 cells. The various *BBS10* promoter constructs are represented showing the relative positions of the c.-80dupC variant, 5'UTR, and the EPD#1 and EPD#2 elements. The firefly luciferase reporter (*luc*) is also indicated. Expression levels are depicted relative to the wild-type BBS10(500bp) promoter with 95% confidence intervals indicated. BBS10(500bp) promoter variants are indicated by the fine crosshatching, whilst BBS10(1kb) promoter variants are indicated by bold crosshatching. Deletion of the EPD#1 promoter region resulted in complete loss of expression, while deletion of EPD#2 resulted in increased transcription (>1.5-fold). The mutant constructs showed ~70% reduction in expression levels.

Table 1. Genetic and analysis details of our cohort of patients.

ID	Gene	GRCh38 coordinates	Variant	gnomAD	*In silico* predictions	Experiment	Result
1	*PRPF31*	Chr19:54129939C>G	c.1374+569C>G Het	Absent	New splice donor site (nnsplice 0.00 > 0.61, Splice AI	Direct mRNA	88bp pseudoexon insertion, p.(Gly459Serfs*46)

					donor gain score 0.74 - -3 bp)		
2	*PRPF31*	Chr19:54115798G>A	c.-9+1G>A Het	Absent	Abolish splice donor site (nnsplice 0.99 > 0.00, SpliceAI donor loss 0.99 - -1bp-; donor gain 0.66 - -44bp-)	N/A	N/A
3	*IFT140*	Chr16:1526044G>A Chr16:1526614_15	c.2611C>T p.(Arg871Cys) c.2577+4_25	0.0000 1580 Absent	Pathogenic x2, VUS x1	- Direct mRNA	- Two cryptic splice transcripts:

		26615del	77+5del		(ClinVar ID 987304) Abolish splice donor site (nnsplice 1.00 > 0.00, Splice AI donor loss 0.71 – 5bp-; donor gain 0.1 - -16bp -)		Exon 20 truncation p.(Val822_Leu859del) Exon 20 skipping p.(Ser800Argfs*16)
4	*NDP*	ChrX: 43958715C>T	c.-70G>A hemi	Absent	Splice AI donor gain (score	Luciferase assay	Negative

					0.40 – 3bp -)		
5	*CRB1*	Chr1:197427615C>T Chr1:197440963C>G	c.2290C>T p.(Arg764Cys) c.3879-1203C>G	Absent Absent	N/A New splice donor site (+1 position C>G, nnsplice 0.00 > 0.96, Splice AI donor gain 0.29 - - 1bp -)	- Minigene analysis	- 156bp cryptic exon inclusion, p.(Trp1293_Cys1294insPhe*)
6	*CRB1*	Chr1:197429619G>A Chr1:197440963C>G	c.2842+5G>A c.3879-1203C>G	0.00003188 Absent	Likely pathogenic (ClinVar ID 438078) New	- Minigene analysis	- 156bp cryptic exon inclusion, p.(Trp1293_Cys1294insPhe*)

					splice donor site (+1 position C>G, nnsplice 0.00 > 0.96, Splice AI donor gain 0.29- -1bp -)		
7	*USH2A*	Chr1:216325412T>G Chr1:216088638T>C	c.1036A>C p.(Asn346His) c.4885+375A>G	0.00006729 Absent	Pathogenic (ClinVar ID 48347) Strengthen cryptic splice donor site (+5 positio	- Direct mRNA assay (nasal epithelial)	- 130bp pseudoexon insertion, p.(Gly1629Valfs*52)

					n A>G, nnsplice 0.76 > 0.99, Splice AI donor gain 0.67 – 5bp -)		
8	*BBS10*	Chr12:76345866_76345867del Chr12:76348438dup	c.2119_2120del p.(Val707*) c.-80dup	0.00006 Absent	Pathogenic (Clinvar ID 406221) Located in core promoter region	- Luciferase assay	- 70% reduced expression
9	*GUCY2D*	Chr17:8015395C>A Chr17:8002596T>	c.2837C>A, p.(Ala946Glu) c.-148T>C	Absent Absent	Uncertain significance	- Luciferase	- 30% reduced expression

		C			(ClinVar ID 848290) Located in CRX binding site	assay	
10	*CRB1*	Chr1:197427726A>T Chr1:197440963C>G	c.2401A>T p.(Lys801*) c.3879-1203C>G	0.0001183 0.0000972	Pathogenic (ClinVar ID 5736) New splice donor site (+1 position C>G, nnsplice 0.00 > 0.96, Splice AI donor gain	- Minigene analysis	- 156bp cryptic exon inclusion, p.(Cys1294Phefs*2)

					0.29 - -1bp -)		
11	*GUCY2D*	Chr17:8015846G>A Chr17:8002596T>C	c.3043+5G>A GUCY2D c.-148T>C	Absent Absent	Likely pathogenic (ClinVar ID 1184616) Located in CRX binding site	- Luciferase assay	- 30% reduced expression

Abbreviations: mRNA: messenger RNA

Table 2. Clinical details of our cohort of patients.

ID	Phenotype	Sex - Race	Symptoms and age of onset (years)	BCVA logMAR (age)	Visual field (age)	Fundus features	Macular OCT features	Family history
1	RCD	M	Nyctalopia (20)	0.2 OD & 0 OS	~25° (41)	Pigmented bone spicules,	Subfoveal island of outer	Yes, AD non penetrant

				(26)		RPE mottling, vessel thinning	segments, oedema	
2	RCD	M – Asian Indian	-	1.5 OD & 1.3 OS (41)	~10° (41)	Pigmented bone spicules, white spots, vessel thinning	Thin, overall loss of EZ line, few cysts	AD
3	CRD	M - White	Decreased acuity (28)	HM OD & 0.25 OS (45)	-	RPE atrophy and pigment clumping on posterior pole	Thin, loss of EZ line	No
4	Bilateral retinal folds/FEVR	M - White	Nystagmus and decreased acuity (birth)	1.6 OD & NLP OS (6)	-	Bilateral retinal folds	Thin retina, schisis	Yes (maternal grandfather)
5	EOSRD	M -	Constricted	LP OD	~10°	Diffuse	Right	Yes

		Whit e	visual field (10)	& OS (45)	(35)	pigment; peripheral telangiecta sia, retinal exudate and tractional retinal detachment in OD	macular schisis, atrophy and loss of EZ line OU, blurred layers	(brother and sister – no consanguini ty)
6	EOSRD	M	Constricted visual field (5)	LP OD & OS (41)	-	Diffuse widespread pigment in bone spicules and nummules, peripheral atrophy	Thin, loss of EZ line, blurred layers	No
7	Usher syndrome type II	F - Whit e	Nyctalopia (14) Hearing loss since small child (aids)	0.2 OD & 0.2 OS (48)	~25° (41)	Pigmented bone spicules, vessel thinning, pale optic	Subfove al island of outer segment s, oedema,	No

						disc	right epiretinal membrane	
8	CRD	F - White	Photoaversion (40)	1 OD & 0.7 OS (69)	-	RPE atrophy on posterior pole and peripheral patches	Subfoveal island of outer segments	No
9	EOSRD	M - White	Nystagmus and poor VA (birth)	1.8 OD & 1.5 OS (44)	-	Featureless fundus and macular atrophy	Disrupted subfoveal EZ line	Yes (sister)
10	EOSRD	F – White	Nyctalopia, constricted visual field	1 OD, 1.1 OS (16)	20° (16)	Para-arteriolar preservation of the RPE (FAF), bone spicules, attenuated vessels,	Blurred inner retinal layers and retinal thickening, loss of outer retinal	No

						generalized atrophy	layers	

Abbreviations: RCD: rod-cone dystrophy; CRD: cone-rod dystrophy; FEVR: familial exudative vitreoretinopathy; EOSRD: early onset severe retinal dystrophy; F: feminine; M: masculine; BCVA: best corrected visual acuity; OD: right eye; OS: left eye; HM: hand movements; LP: light perception; NLP: no light perception; RPE: retinal pigment epithelium; FAF: fundus autofluorescence; OCT: optical coherence tomography; EZ: ellipsoid zone; AD: autosomal dominant.