Comprehensive rare variant analysis using whole genome sequencing to determine the molecular pathology of inherited retinal disease

Authors

Keren J Carss,[1,2,18] Gavin Arno,[2,3,4,18] Marie Erwood,[1,2] Jonathan Stephens,[1,2] Alba Sanchis-Juan,[1,2] Sarah Hull,[3,4] Karyn Megy,[1,2] Detelina Grozeva,[2,5] Eleanor Dewhurst,[1,2] Samantha Malka,[3,4] Vincent Plagnol,[6] Christopher Penkett,[1,2] Kathleen Stirrups,[1,2] Roberta Rizzo,[4] Genevieve Wright,[4] Dragana Josifova,[2,7] Maria Bitner-Glindzicz,[2,8] Richard H Scott,[2,9] Emma Clement,[2,10] Louise Allen,[2,11] Ruth Armstrong,[2,12] Angela F Brady,[2,13] Jenny Carmichael,[2,12] Manali Chitre,[2,12] Robert HH Henderson,[2,4,10] Jane Hurst,[2,10] Robert E MacLaren,[2,4,14] Elaine Murphy,[2,15] Joan Paterson,[2,12] Elisabeth Rosser,[2,10] Dorothy A Thompson,[2,16] Emma Wakeling,[2,13] Willem H Ouwehand,[1,2] Michel Michaelides,[2,3,4] Anthony T Moore,[2,3,4,17] NIHR-BioResource Rare Diseases Consortium,[2] Andrew R Webster,[2,3,4,19] F Lucy Raymond[2,5,19]

Affiliations

1. Department of Haematology, University of Cambridge, NHS Blood and Transplant Centre, Cambridge, CB2 0PT, UK

2. NIHR BioResource - Rare Diseases, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK

3. UCL Institute of Ophthalmology, University College London, London, EC1V 9EL, UK

4. Moorfields Eye Hospital, London, EC1V 2PD, UK

5. Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, CB2 0XY, UK

6. University College London Genetics Institute, London, WC1E 6BT, UK

7. Clinical Genetics Department, Guy's Hospital, Great Maze Pond, London, SE1 9RT, UK

8. UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK

9. North East Thames Regional Genetics Service, Great Ormond Street Hospital, , London, WC1N 3JH, UK

10. Great Ormond Street Hospital for Children, Great Ormond Street, London, WC1N 3JH, UK

11. Department of Ophthalmology, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK

12. Department of Medical Genetics, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK

13. North West Thames Regional Genetics Service, London North West Healthcare NHS Trust, Watford Road, Harrow, HA1 3UJ, UK

14. Nuffield Laboratory of Ophthalmology, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK

15. Charles Dent Metabolic Unit, National Hospital for Neurology and Neurosurgery, Queen Square, London, WC1N 3BG, UK

16. Clinical and Academic Department of Ophthalmology, Great Ormond Street Hospital for Children, London, WC1N 3JH, UK

17. Ophthalmology Department, UCSF School of Medicine, University of California San Francisco, San Francisco, CA 94158, USA

18. These authors contributed equally to this work

19. These authors contributed equally to this work

Correspondence

Corresponding author: Professor F. Lucy Raymond. flr24@cam.ac.uk

Abstract

Inherited retinal disease is a common cause of visual impairment, and represents a highly heterogeneous group of conditions. Here, we present findings from a cohort of 722 individuals with inherited retinal disease, who have had whole genome sequencing (n=605), whole exome sequencing (n=72), or both (n=45) performed, as part of the NIHR-BioResource Rare Diseases research study. We identified pathogenic variants (single nucleotide variants, indels, or structural variants) for 404/722 (56%) individuals. Whole genome sequencing gives unprecedented power to detect three categories of pathogenic variants in particular: structural variants, variants in GC-rich regions, which have significantly improved coverage compared to whole exome sequencing, and variants in non-coding regulatory regions. In addition to previously reported pathogenic regulatory variants, we have identified a previously

unreported pathogenic intronic variant in the *CHM* in two males with choroideremia. We have also identified 19 genes not previously known to be associated with inherited retinal disease, which harbour biallelic predicted protein-truncating variants in unsolved cases. Whole genome sequencing is an increasingly important comprehensive method with which to investigate the genetic causes of inherited retinal disease.

Introduction

Inherited retinal disease (IRD) describes a phenotypically heterogeneous group of conditions consequent upon dysfunction and/or degeneration of the neural retina or retinal pigment epithelium, resulting in visual impairment. IRD is the most common cause of severe visual impairment among working age individuals in the UK.[1] Determining the genetic cause of IRD in affected individuals allows accurate assessment of risk of the disease to other family members, provides useful prognostic information for affected individuals, and can provide much needed insight and understanding of the disease for those affected. Furthermore, genetic stratification of affected individuals is increasingly being used to direct specific treatment options, including clinical trials of medicines and restoration of protein function through gene therapy trials.[2]

Whole exome sequencing (WES) of large cohorts of individuals has transformed research into rare Mendelian diseases in recent years by facilitating discovery of pathogenic variants, newly described disease-associated genes, and other insights into the genetic architecture of rare diseases.[3–5] Whole genome sequencing (WGS)

has thus far typically been employed on a smaller scale, and has demonstrated advantages over other methods.[6,7] Large-scale projects such as the Genomics England 100,000 Genome Project are also beginning to use WGS to investigate rare diseases.[8]

IRD represents a particularly good clinical phenotype to provide comprehensive investigation by high-throughput sequencing technologies because it exhibits a high degree of phenotypic, genetic and allelic heterogeneity, with over 250 genes and loci associated with autosomal dominant, autosomal recessive, X-linked and mitochondrial inheritance (RetNet), and it is often difficult to predict the disease-associated gene from an individual's phenotype alone.[9] IRD is therefore among the many rare diseases that are increasingly being investigated through high-throughput sequencing.[10–13] These technologies are also transforming clinical practice by being incorporated into diagnostic services.[14,15] Most of these studies have used targeted pull-down of a predetermined panel of known genes, which are limited not only in that they only cover certain genes, but also because they cannot reliably detect structural variants (SVs). WES, on the other hand, is not limited to known disease-associated genes, but coverage is generally variable across the exome, negatively affecting variant detection compared to targeted gene panel testing.[16] PCR-free WGS should overcome the disadvantages of WES and targeted gene panels, although the remaining challenge will be in handling the large amount of data generated, and variant interpretation. Indeed, a recent study compared WGS to targeted gene panels and found that WGS improves the pathogenic variant detection rate by facilitating detection of SVs and variants in regulatory regions.[17]

The aims of the current study were threefold. First, to identify pathogenic variants in a large cohort of individuals with IRD, including intractable cases in which previous investigations had not yielded a diagnosis. Second, to explore advantages and disadvantages of WGS as a tool with which to investigate IRD. Third, to identify

candidate IRD-associated genes, and provide new insights into phenotypes and the genetic architecture of IRD. Here we present a cohort of 722 individuals with IRD, who have had high-throughput sequencing as part of the NIHR-BioResource Rare Diseases research study.

Methods

Cohort

722 individuals with IRD were recruited to the NIHR BioResource Rare Diseases research study. All participants provided written informed consent and the study was approved by the East of England Cambridge South national institutional review board (13/EE/0325). Most of the individuals (n= 657) were recruited at the Inherited Eye Disease clinics, Moorfields Eye Hospital NHS Foundation Trust (London, UK). The remainder were recruited at Cambridge University Hospitals NHS Foundation Trust (Cambridge, UK), Great Ormond Street Hospital For Children NHS Foundation Trust (London, UK), Guy's And St Thomas' NHS Foundation Trust (London, UK), London North West Healthcare NHS Trust (London, UK), and University College London Hospitals NHS Foundation Trust (London, UK).

Most individuals had undergone some previous genetic testing using routine diagnostic approaches (direct Sanger sequencing of highly suggestive genes based on clinical characteristics, direct sequencing of candidate genes in gene discovery projects and arrayed primer extension (APEX, Asper Biotech Ltd, Tartu, Estonia) assays for retinitis pigmentosa [MIM: 268000] and Leber congenital amaurosis (LCA [MIM: 204000]). Individuals with no likely pathogenic variant detected by these methods were recruited to the study although systematic documentation of pre-

screening of all cases was not available. 152 individuals of the cohort had no pre-screening performed.

High throughput sequencing

WES and WGS were performed as previously described.[18,19] Genome build GRCh37 was used for mapping and variant calling. SVs were identified in the individuals who had WGS using two independent algorithms: Isaac Copy Number Variant Caller (Canvas, Illumina), which identifies copy number gains and deletions based on read depth, and Isaac Structural Variant Caller (Manta, Illumina), which identifies translocations, deletions, tandem duplications, insertions, and inversions based on both paired read fragment spanning and split read evidence.[20]

To identify the likely ethnicity of each individual, a principal component analysis was performed on WGS data from individuals of various ethnicities in the 1000 genomes project with 20,000 single nucleotide variants (SNVs) using PLINK.[21] The location of the centre of the cluster of each ethnicity was then calculated, and a likely ethnicity assigned to each of the WGS samples based on the closest cluster.

Variant interpretation

To facilitate variant interpretation, a list of reported IRD-associated genes was assembled, including genes associated with syndromic forms of IRD or albinism, from various sources including RetNet and literature searches. This list was manually curated according to published evidence of pathogenicity to compile a shortlist of 224 high-confidence IRD-associated genes (table S1).[22]

To identify pathogenic variants, a two-step variant filtering protocol was designed, utilising automated filtering followed by manual review. For SNVs and indels,

automated filtering identified variants that fulfil the following criteria: passes standard Illumina quality filters in >80% of the whole NIHR BioResource Rare Diseases cohort (n = 6688); predicted to be a high impact, medium impact, or splice region variant, or present in the HGMD Pro database;[23] and has minor allele frequency (MAF) <0.01 in control datasets including the NIHR BioResource Rare Diseases cohort and the Exome Aggregation Consortium (ExAC) database.[24] If a variant is present in the HGMD Pro database a higher MAF threshold of 0.1 was used. Finally, we identified just those variants that affect an IRD-associated gene. In one individual (G002628) with hypomagnesemia and retinitis pigmentosa, we interrogated known hypomagnesemia-associated genes in addition to IRD-associated genes. For SVs in the WGS data, automated filtering identified variants that fulfil the following criteria: passes standard Illumina quality filters; overlaps at least one exon; does not overlap known benign SVs in healthy cohorts;[25] has MAF <0.01 in the NIHR BioResource Rare Diseases cohort, and affects an IRD-associated gene.

Manual review of all the variants that passed the automated filtering was then performed. The variant is considered to be pathogenic if it fulfils the following criteria. First, the genotype and frequency of the variant is consistent with the expected mode of inheritance of the individual's family (if known), and with the expected mode of inheritance of the gene. Second, the phenotype of the individual is consistent with the phenotype known to be associated with the gene. Third, the variant is predicted to result in truncation of the protein, or it is predicted to be damaging to the protein using scores such as CADD,[26] or it has previously been reported as pathogenic in HGMD Pro. Fourth, the variant appears to be of good quality upon examination of the sequencing reads with Integrative Genomics Viewer (IGV).[27] Fifth, the variant affects the Ensembl canonical transcript or a known retinal transcript.

Variant confirmations

A subset of the pathogenic SNVs and indels was confirmed by Sanger sequencing using standard protocols. For SVs, Sanger sequencing was performed across the predicted breakpoints to generate a unique junction fragment sequence. Genotyping using HumanCoreExome-24 v1.0 BeadChip (Illumina) was also performed, followed by SV identification using GenomeStudio and cnvPartition software (Illumina). Sequences of all primers are available on request.

Comparing coverage of WGS and WES datasets

Protein-coding regions of the Ensembl canonical transcript of each autosomal IRD-associated gene were split into 50bp bins. The mean GC% of each bin was calculated using data available from the UCSC genome browser. The mean coverage of each bin was calculated across a sample of 100 individuals with IRD in the NIHR BioResource Rare Diseases cohort (WGS data), and also in the ExAC cohort, for which per-base coverage has been previously published on 10% of the cohort.[24] We used the ExAC WES dataset for the comparison of the coverage of IRD-associated genes, rather than the WES data generated as part of this study, because ExAC was used as a frequency control set in this study, and because it is based on > 60,000 individuals.

Coverage data are presented as relative to the mean coverage of all bins (34.4X in the WGS dataset, and 65.6X in ExAC). Coverage for each observed range of GC content (20-30%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, and 80-90%) was compared using one-tailed Mann-Whitney U tests. P values were corrected for testing of 7 observed ranges of GC content.

Variant analysis in the *CHM*

Two individuals with a clinical diagnosis of choroideremia [MIM: 303100] in whom no pathogenic coding variant was found, underwent intronic variant analysis of *CHM* ([MIM: 300390]; ENST00000357749.2, GRCh37/hg19 chrX:85,116,185-85,302,566). Intronic variants in *CHM* that passed standard quality filters and had a MAF ≤0.01 in the UK10K WGS cohort and the NIHR BioResource Rare Diseases cohort were prioritised for analysis. Splice site prediction of the reference and alternate sequences were compared using Human Splice Finder V3.0 (HSF) and NNSPLICE0.9.[28,29] We identified a candidate variant that was confirmed by direct Sanger sequencing of genomic DNA from affected members of both families and an obligate carrier. Reverse transcription PCR (RT-PCR) spanning *CHM* exons 4-5 was carried out on total RNA isolated from peripheral blood mononuclear cells.

Identification of genes containing likely biallelic predicted protein-truncating variants in unsolved cases.

Variants (SNVs, indels, and large deletions) were identified that fulfil the following criteria: in an unsolved case where an individual has undergone WGS, are high impact and autosomal, the genotype is consistent with autosomal recessive inheritance, have MAF <0.001 in the control datasets described, are in the Ensembl canonical transcript, are in a gene that contains no homozygous predicted protein-truncating variants in the ExAC database, and appear to be real on examination in IGV. This analysis was not limited to IRD-associated genes, but included all protein-coding genes.

Results

In a highly heterogeneous cohort of individuals with inherited retinal disease, high-throughput sequencing achieves a pathogenic variant detection rate of 56%

High-throughput sequencing was performed on a cohort of 722 individuals with IRD, as part of the NIHR-BioResource Rare Diseases research study. 605 of the individuals had WGS, 72 had WES, and an additional 45 had both. For WGS average coverage was 37X (SD = 2.7), with a minimum of 95% of each genome covered to at least 15X. For WES average target coverage was 43X (SD = 14.9), with 83% of target bases covered to at least 15X.

The cohort is phenotypically heterogeneous, and the most frequent phenotypes are retinitis pigmentosa (n=311), retinal dystrophy (n=101), cone-rod dystrophy (n=53), Stargardt disease (n=45), macular dystrophy (n=37), and Usher syndrome (n=37). The majority of individuals are unrelated probands (714/722), and there are three parent-offspring pairs and one pair of siblings. The majority of the cohort had a negative result for a genetic test, such as single gene Sanger sequencing or targeted gene panel sequencing, prior to enrolment in the current study.

To identify the pathogenic variants, rare, coding, and high-quality variants, including large deletions, in 224 genes that are known to be associated with IRD, were considered. The likely cause of IRD was identified in 404/722 individuals, corresponding to a pathogenic variant detection rate of 56% (table 1, table S2). A further 5% of individuals (36/722) have been classified as partially solved. These individuals either have a single likely pathogenic variant in a known gene associated with recessive IRD and in keeping with the individual's phenotype, more than two heterozygous variants in a gene associated with recessive IRD, or a variant that only explains part of the phenotype (table S2). The remaining 282/722 (39%) of individuals remains without a complete or partial molecular diagnosis.

By using WES alone we were able to identify pathogenic variants in 59/117 (50%) individuals. Subsequently, 45 of the 58 individuals who were unsolved by WES underwent WGS, and we identified or confirmed pathogenic variants in a further 14 (table 1). For three individuals the bait for the variant location was absent from the capture kit (Roche NimbleGen, SeqCap EZ Exome v3), for two individuals each had a large deletion not called by WES and, in one individual, a large indel was also not called by WES. For three more individuals, the variant was called by WES but the quality was poor. For the remaining five individuals, although the variants were identified by WES, they were not the expected mode of inheritance for the family, therefore WGS was performed to exclude more plausible cause of disease by including analysis of copy number variants and exons of high GC content.

The pathogenic variant detection rate also varies depending on the phenotype of the individual (table 2). For example, 168/311 (54%) of individuals with retinitis pigmentosa have been solved, which is similar to the overall rate; 31/37 (84%) of individuals with Usher syndrome have been solved but only 6/21 (29%) of individuals with cone dystrophy have been solved. Also, individuals who had not had any genetic test prior to this study had a notably higher pathogenic variant detection rate than the overall rate. Of 152 individuals who had no pre-screening, including individuals with a diagnosis of retinitis pigmentosa, cone dystrophy or congenital stationary night blindness, 96 were solved (63%), suggesting that our cohort is enriched for intractable cases.

The pathogenic variant detection rate within our cohort also varies depending on the ethnicity of the individual (table 3). Likely ethnicity was estimated from the WGS data. Only 13/43 (30%) of individuals of African ancestry were solved, compared to 259/467 (55%) of individuals of European ancestry and 70/123 (57%) of individuals of South Asian ancestry. Higher genetic diversity in African populations, combined with underrepresentation of non-European populations in control datasets, result in

an excess of rare and apparently rare variation in these individuals, rendering variant interpretation more challenging.[24,30] There was a median of 12 rare coding variants in IRD-associated genes for manual review in individuals of European ancestry and 30 in individuals of African ancestry. Interestingly, in individuals of South Asian ancestry, 59/89 pathogenic variants (66%) were homozygous, compared to only 82/446 (18%) in individuals of European ancestry. This is likely due to increased rates of consanguinity in South Asian populations, and probably explains why they have comparable pathogenic variant detection rates to individuals of European ancestry in this study, despite also being underrepresented in control databases of allele frequencies.

The genes in which pathogenic variants are most frequently found are summarised in table 4. *ABCA4* [MIM: 601691] is implicated in 73/440 solved or partially solved cases (17%). In total, 796 pathogenic alleles were identified in 95 different known IRD-associated genes. Of these alleles, 687 (86.3%) are autosomal recessive, 72 (9%) are autosomal dominant, 35 (4.4%) are X-linked hemizygous, and 2 (0.3%) are X-linked monoallelic. Some of the pathogenic alleles are homozygous, and some occur in more than one individual in the cohort, thus there are 537 different pathogenic alleles. Of these, 499 (92.9%) are exonic SNVs or indels, 30 (5.6%) are large deletions overlapping at least one exon, 6 (1.1%) are SNVs or indels that are synonymous or in regulatory regions, and 2 (0.4%) are tandem duplications overlapping at least one exon. 291/537 (54.2%) are previously unreported in publically available databases emphasizing the allelic heterogeneity of IRD, and the rest have been previously reported and are in the HGMD Pro database.[23]

We performed confirmatory Sanger sequencing and segregation in families in a subset of variants. Of 177 alleles Sanger sequenced, all were confirmed. Of these, 52 were tested in at least one additional informative family member, and segregated as expected. In five of the individuals whose case is unsolved , a pathogenic variant

was identified by a different method concurrent to this study. Four of these are variants in highly repetitive regions requiring specific optimised Sanger sequencing protocols, for example *RPGR* [MIM: 312610] exon *ORF15*, and one is a heteroplasmic variant in the mitochondrial genome, which was apparent retrospectively in the WGS data, but was not identified by the variant caller.

Pathogenic structural variants are an important cause of IRD, and their detection is facilitated by WGS

There are 33 SVs (31 deletions and 2 tandem duplications) that are pathogenic in 31 individuals (table S2). Each SV was further investigated using either SNP genotyping array, direct Sanger sequencing of a unique PCR amplified product spanning the predicted breakpoints, or both techniques. Twenty-three (70%) have been confirmed by Sanger sequencing of the breakpoint, 6 (18%) have been confirmed by both Sanger sequencing the breakpoint and SNP genotyping array, 2 (6%) have been confirmed by genotyping only and 2 (6%) have not yet been confirmed by an alternative method, although visual inspection of the IGV plots predict a deletion. In seven SVs, confirmatory Sanger sequencing of PCR generated products over the breakpoint also revealed an insertion (1-51 bps) at the breakpoint of the variants predicted by the algorithms.

The Manta algorithm, which identifies structural variants based on both paired read fragment spanning and split read evidence, predicts the correct breakpoints with 100% accuracy in 13/26 SVs that were called by Manta and have been confirmed by Sanger sequencing. It predicts the correct breakpoints to within 3bp in 21/26 SVs. Interestingly, we have identified one identical SV in two unrelated individuals. G001296 and W000139 both have 15:89750128-89757489del, which overlaps exons 7-9 of *RLBP1* [MIM: 180090]. In G001296 the deletion is heterozygous and occurs in

conjunction with a second pathogenic variant, and in W000139 it is homozygous. It does not occur in any other individual in the NIHR BioResource Rare Diseases cohort.

Case study one illustrates the power of WGS to detect SVs. Two likely pathogenic heterozygous variants were identified in *EYS* [MIM: 612424] in individual W000325, who presented with typical retinitis pigmentosa. The variants are a missense variant ENST00000503581.1:c.6473T>C (p.Leu2158Pro) that is predicted to be damaging by CADD (phred-scaled CADD score = 23.5) and has not been reported before in public databases, and a 55kb deletion (chr6:65602819-65658187del), confirmed by Sanger sequencing, which overlaps exons 15-18 (c.2260-2380_2847-6084del). If transcribed, this large deletion is predicted to lead to a frameshift and premature termination in exon 19 (p.R949Ifs*4). This individual had undergone WES analysis prior to WGS analysis, from which it was not possible to identify the pathogenic deletion (figure 1).

Case study two is an interesting example of likely uniparental isodisomy discovered through analysis of SVs. Individual W000170 presented with an atypical early-onset form of retinal dystrophy. A homozygous in-frame combination indel in *KCNV2* [MIM: 607604] was first identified: ENST00000382082.3:c.222_232delGGACCAGCAGGinsGGTCACCACCACCTTGG (ENSP00000371514.3:p.Asp75_Gln77delinsValThrThrThrLeu). The mother of the individual is heterozygous for this indel, and the father is homozygous for the reference allele. By visual inspection of the surrounding region using the IGV software,[27] a homozygous tandem duplication was identified (chr9:2717844-2718030dup), flanking the indel (figure 2), which was confirmed by Sanger sequencing. Further investigation revealed a long run of homozygosity with approximate coordinates chr9:2100000-27400000. Taken together, we consider that

the most likely explanation for these observations is homozygosity for the *KCNV2* disease allele due to partial maternal uniparental isodisomy of chromosome 9.

Case study three is an interesting example of overlapping compound heterozygous deletions. Individual W000164 presented with typical retinitis pigmentosa. Two likely pathogenic overlapping heterozygous deletions were identified using WGS. The first (chr6:64475599-64501270del) encompasses exons 38-40 of *EYS*, and the second (chr6:64491812-64513698del) encompasses exons 38-39 of *EYS* (figure 3). Sanger sequencing confirmed both deletions. This example illustrates the value of algorithms that use paired read fragment spanning evidence such as Manta. Only these are able to provide characterisation of breakpoints to a single base pair resolution. WES or indeed read depth analysis alone of WGS (e.g. the Canvas algorithm) would fail to characterise the breakpoints precisely, rendering Sanger sequencing confirmation or family analysis more difficult.

WGS improves coverage of GC-rich regions compared to WES

In the current study, we have observed increased power to detect variants in GC-rich regions. To assess the possible impact of this, we calculated the average coverage of exons of known IRD-associated genes, split into 50bp bins, in our WGS dataset and in the ExAC WES dataset. For exons of IRD-associated genes with a GC content of 45-65% the average coverage by read depth of sequence using WES is higher than for WGS. The calling of SNV exonic variants in this range is not significantly different for WES and WGS. Comparing WES from ExAC, we find significantly higher coverage in our WGS dataset in bins with GC <30% or >70% (figure 4). Additionally, the variability of coverage in our WGS data is much less than that of the ExAC dataset. This uniformity of coverage is one of the main factors that

make WGS particularly powerful for SV detection. Calling SVs in WES where the variance of read depth is so great is unreliable (figure 1A).

Case study four illustrates the clinical relevance of this improvement in coverage. Individual G004991 presented with Leber's congenital amaurosis. Pathogenic compound heterozygous variants were identified in the first coding exon of *GUCY2D* [MIM: 600179], which is 76% GC (figure 5). The variants are: an in-frame deletion ENST00000254854.4:c.238_252delGCCGCCGCCCGCCTG (p.Ala80_Leu84del), not previously reported in publically available databases and a previously reported missense variant ENST00000254854.4:c.307G>A (p.Glu103Lys).[31] This exon is not covered in the capture kit used for WES, and had WES been used rather than WGS to investigate this individual, neither pathogenic variant would have been identified. Furthermore, visual inspection of the sequencing reads using IGV demonstrates biallelic inheritance of the two variants, since they occur within 70 nucleotides and are never observed on the same 150bp read.

WGS allows identification of pathogenic variants in non-coding regions

In our cohort, three different pathogenic non-coding SNVs were identified. For example, in G008165, who presented with Stargardt disease, one intronic SNV was identified in *ABCA4* in *trans* with a previously reported synonymous SNV, 1:94466602C>T (ENST00000370225.3:c.6342G>A) that results in a premature donor splice site that truncates exon 46.[32] The heterozygous intronic variant was 1:94476951A>G (ENST00000370225.3:c.5461-10T>C), which has also been previously reported and causes aberrant splicing.[33] This variant was found in 16 individuals, all of whom presented with typical features of *ABCA4*-retinopathy, and thus represents a significant portion of disease alleles. In two of these individuals this variant is homozygous, in a further 9 individuals it occurs in conjunction with a

second heterozygous likely pathogenic allele, and in 5, we have identified no second likely pathogenic allele and the individuals remain partially solved. This variant is rare in both ExAC (MAF = 2.2×10^{-4}, no homozygotes) and the whole NIHR BioResource Rare Diseases cohort (MAF = 1.8×10^{-3}, no homozygotes other than the two described here).

In another example, individual G001035, who presented with Usher syndrome, the deep intronic *USH2A* [MIM: 608400] variant 1:216064540T>C (ENST00000307340.3:c.7595-2144A>G) was identified. This is heterozygous and occurs in conjunction with a second heterozygous likely pathogenic allele. It has been previously reported, and results in the retention of a pseudoexon, which causes a frameshift and premature truncation of the protein.[34]

Finally, in two unrelated males with a clinical diagnosis of choroideremia (G001372 and G007713) we identified a deep intronic variant in *CHM* that has not been previously reported. Both individuals had previously undergone *CHM* screens, which were negative, and no convincing pathogenic coding variant was found in any IRD-associated gene in either individual upon analysis of the WGS data. Given the unequivocal diagnosis of choroideremia in these individuals and the single genetic cause of choroideremia, the sequence of these cases was selected for interrogation of rare non-coding variants in *CHM* (chrX:85,116,185-85,302,566). We hypothesised that a deep intronic variant may act as a null variant by altering splicing as previously described.[35]

A rare (MAF≤0.01) deep intronic hemizygous variant was identified in G001372 in the genomic region of *CHM*: chrX:85,220,593T>C (ENST00000357749.2:c.315-1536A>G). Individual G007713 had three rare intronic variants, including the same chrX:85,220,593T>C variant. This variant is absent in the UK10K genome project dataset and the 1000 genomes dataset.[30,36] The variant was confirmed in both

individuals by direct Sanger sequencing, and it was also confirmed in an affected male cousin of G007713, and in the heterozygous state in the mother of G007713.

Splice prediction analysis using HSF[28] and NNSPLICE0.9[29] demonstrated the likely introduction of a cryptic splice acceptor site by this variant (HSF consensus value: 92.66, NNSPLICE0.9 score 0.94). The presence of a strong donor site (HSF consensus value: 84.71, NNSPLICE0.9 score: 0.89, sequence atgcaag**gt**aaactg) 224bp downstream of the cryptic acceptor site suggested that a cryptic exon could arise from this variant (figure 6).

To confirm this, we used RT-PCR amplification spanning exons 4-5 of *CHM* from G001372, G007713, and the heterozygous mother of G007713. This demonstrated a fragment approximately 200bp larger than expected (2 fragments in the heterozygous mother). Upon sequencing of the fragments, the predicted cryptic exon was confirmed. The variant is predicted to lead to a premature termination codon after 9 altered amino acid residues (p.S105Rfs*10).

Monoallelic variants in genes associated with a recessive mode of inheritance

Interpreting single heterozygous variants in recessive IRD-associated genes presents a challenge, as it is typically difficult to distinguish between a case in which a second pathogenic allele in the same gene has been missed, from a case in which the individual is just a carrier of the single heterozygous variant and the real cause of disease lies elsewhere. In our cohort there are only 16 individuals (2%) who have a single heterozygous likely pathogenic variant in a recessive IRD-associated gene. Of these, 12 are in *ABCA4*, and one is in each of *CEP290* [MIM: 610142], *CNGB1* [MIM: 600724], *GUCY2D*, and *CRB1* [MIM: 604210]. These individuals all have phenotypes that are strongly indicative of variants in the gene in question, and the identified

variant in each case is either a predicted protein-truncating variant or a missense previously reported as being pathogenic. We excluded the possibility that these individuals have an additional previously reported pathogenic deep intronic variant. The probability is high that at least some of these 16 partially solved individuals harbour a second pathogenic variant in the same gene, perhaps in a regulatory region, that remains elusive. This cohort is under continuing investigation to identify further variants, including variants in regulatory regions. The proportion of individuals who have a single likely pathogenic variant in an autosomal recessive gene in this study is lower than previously reported (2%).[17] This may reflect a combination of detailed specialist phenotyping to reduce phenocopies and the comprehensive exon and non-coding coverage achieved using WGS.

Identification of genes containing likely biallelic predicted protein-truncating variants in unsolved cases.

We performed further investigation of 247 individuals, in whom no pathogenic variants were detected in the known IRD-associated genes, and whose family history is not inconsistent with recessive disease.. These were screened for any gene containing ≥2 predicted protein-truncating alleles, including SVs. This yielded a list of 19 genes in 16 individuals (table 5 and table S3). None occurred in more than one individual, but three individuals have variants in two genes. Segregation has not been performed on the double heterozygous variants. None of these genes contain any homozygous predicted protein-truncating variants in the ExAC dataset, suggesting that they may not tolerate biallelic loss of function variation.

Discussion

Overall, we have identified pathogenic variants for 404/722 individuals with IRD in the NIHR-BioResource Rare Diseases study, which is a pathogenic variant detection rate of 56%. Factors that influence this rate within our cohort include the phenotype of the individual, for example individuals with Usher syndrome are substantially more likely to receive a molecular diagnosis than individuals with retinitis pigmentosa. There are several possible reasons for this. First, it may be that a higher proportion of Usher syndrome genes overall have been identified than retinitis pigmentosa genes. Second, more specific phenotypes can suggest a smaller number of candidate genes; retinitis pigmentosa is far more genetically heterogeneous than Usher syndrome. Third, it is more difficult to distinguish pathogenic monoallelic variants from the many rare benign inherited monoallelic variants, than it is to distinguish pathogenic biallelic variation, therefore phenotypes that are predominantly recessive tend to have higher pathogenic variant detection rates. Also, it may be that individuals with some phenotypes were more likely to undergo some pre-screening prior to enrolment in the project, which would exclude individuals with variants in known genes and likely reduce the pathogenic variant detection rate.

Our pathogenic variant detection rate of 56% with WGS is comparable to those previously reported in other similar studies of IRD, which ranges between 39-70%.[10–15,17] Factors that influence differences between these rates include the technology used and the degree of pre-screening performed on the cohort, as well as the phenotypes included and the ethnic distribution of the cohort as discussed. Our detection rate of 50% using WES in 117 individuals is perhaps lower than expected due to the WES depth of coverage of 43x used compared to diagnostic laboratory median coverage of >80x that is recommended [39]. Our observation that the subset of our cohort who had no pre-screening had a higher pathogenic variant detection rate than the overall rate, suggests that our cohort is enriched for intractable cases, and

that the overall rate of 56% is an underestimate compared with what would be expected if WGS was used as the first-line test.

We have observed that using primarily WGS instead of WES or targeted gene panels has improved our power to detect three categories of variants in particular: SVs, variants in GC-rich regions, and variants in regulatory regions. Previous studies have also observed that WGS is superior to WES for the detection of SVs, particularly small deletions (e.g. single exon deletions).[6,37,38] This is due to high uniformity of coverage achieved with WGS, and the very high probability that the breakpoints of a SV will be covered by WGS reads. With WGS, therefore, precise characterisation of the SV to single base pair resolution is often possible, without any further investigation.

Our study also supports previous reports that PCR-free sequencing protocols, such as WGS, capture genomic regions that are particularly high or low in GC content much more effectively than methods such as WES and targeted gene panels, which require a PCR amplification step.[37,39] GC-rich regions tend to be poorly covered by methods that require a PCR amplification step due to their high stability and consequent resistance to standard denaturation protocols, and variant calling suffers as a consequence.[40–42]

The WGS analysis identified three intronic variants that would not have been identified by standard WES without specific prior knowledge. The interpretation of these variants required the availability of sequence from a large number of internal controls (~13,000 alleles from the NIHR BioResource Rare Diseases consortium) and external WGS controls (~8,000 alleles from UK10K). Frequency data on non-coding variants comparable to the ExAC dataset for coding variants is needed and will transform rare disease non-coding variant interpretation. Clinical WES platforms can be designed to target known pathogenic variants in regulatory regions in addition

to the exome footprint.[43] However, they require regular redesign to capture newly reported non-coding pathogenic variants, and they do not give the option of identifying pathogenic deep intronic variants that have not been previously reported.

Regarding limitations of WGS, highly repetitive regions generally still remain poorly covered in our WGS data, due to the difficulty of uniquely mapping the reads in these regions to the genome. This has previously been reported,[17] and is also a problem for targeted gene panels and WES.[13,14] This is an important issue for IRD, due to the existence of clinically important repetitive regions such as *RPGR ORF15*, which is highly repetitive, and is a mutational hotspot that constitutes one of the most common causes of X-linked retinitis pigmentosa.[44] While in the future, increasing read lengths and improved read mapping algorithms are likely to result in improved coverage over these regions, currently they must still be sequenced separately using optimised PCR amplification and Sanger sequencing protocols to exclude the possibility of them harbouring pathogenic variants. WGS also remains costlier than targeted gene panels or WES, and so currently costs more per diagnosis.[45] With the cost of WGS continuing to fall and the emergence of large-scale WGS projects such as the Genomics England 100,000 Genome Project,[8] WGS is nevertheless an increasingly important tool for investigating genetic causes of rare diseases, including IRD, in clinical practice as well as research, and is likely to become the preferred first-line test in the future.[47]

This project has already contributed to novel insights into the genetic architecture and phenotypic spectrum of IRD. We have identified a distinctive electroretinogram phenotype, predominantly involving the cone pathways, in two individuals with variants in *CACNA2D4* [MIM: 608171].[48] Additionally, an individual from this study with biallelic *IFT140* [MIM: 614620] variants, along with additional cases, expanded the known phenotypic spectrum of *IFT140*-associated disease, which was originally reported as a severe syndromic ciliopathy, and is now known to also cause non-

syndromic retinitis pigmentosa.[19] Several other individuals in our cohort with variants in both *RGR* [MIM: 600342] and *CDHR1* [MIM: 609502] led to the discovery that a recessive retinal disorder previously associated with a homozygous variant in *RGR* is more likely to be caused by the variant in *CDHR1*, which is in complete linkage disequilibrium with the originally reported variant. [46]

There are many possible reasons why 282/722 (39%) of our cohort remain as yet unsolved. Some may have pathogenic variants that have not been called by the variant calling software because, for example, they are in regions of poor coverage such as repetitive regions. Others may have pathogenic variants that were called but not manually reviewed because they did not pass one of our filters. Some are likely to have pathogenic variants in genes, associated with IRD that are not on our list of IRD-associated genes. For some unsolved individuals, the cause of their disease may be more complex than Mendelian inheritance, but may be oligogenic, or influenced by environmental factors. Finally, it is likely that a proportion of unsolved individuals have pathogenic variants in regulatory regions, rather than in coding exons. Regulatory variation is known to be an important cause of IRD.[12,32,34,49,50] Having used WGS for the majority of our cohort, we are well placed to investigate this class of variation. However, interpretation of regulatory variation remains a significant challenge. Experimental verification of specific variants is particularly difficult as many IRD-associated genes are restricted to expression in the retina, making functional confirmation of the effects of any candidate regulatory variation challenging. Our investigation of potentially pathogenic regulatory variation is ongoing.

We have identified 19 genes that contain likely biallelic, predicted protein-truncating variants in unsolved individuals with IRD, including *CROCC* [MIM: 615776], *IRX5* [MIM: 606195] and *NUMB* [MIM: 603728] that have known roles in retinal development and function.[51,52] These may represent strong candidates for previously

unreported recessive IRD-associated genes (table S3). Identification of at least two additional individuals with likely pathogenic variants is necessary to provide sufficient supportive evidence of any of these genes being associated with IRD.

In conclusion, we present here a large cohort of individuals with IRD who have had WGS, and we have achieved a pathogenic variant detection rate of 56%. We have identified three categories of pathogenic variant that WGS substantially improves the detection of: SVs, variants in GC-rich regions, and variants in regulatory regions. Studying a cohort of this size using WGS provides new insights into phenotypes and the genetic architecture of IRD. In the future, WGS is likely to become the preferred choice of test with which to investigate the genetic causes of IRD.

Supplemental data

Supplemental data includes three tables.

Consortia

Members of the NIHR-BioResource Rare Diseases Consortium:

Timothy Aitman, Hana Alachkar, Sonia Ali, Louise Allen, David Allsup, Gautum Ambegaonkar, Julie Anderson, Richard Antrobus, Ruth Armstrong, Gavin Arno, Gururaj Arumugakani, Sofie Ashford, William Astle, Antony Attwood, Steve Austin, Chiara Bacchelli, Tamam Bakchoul, Tadbir K Bariana, Helen Baxendale, David Bennett, Claire Bethune, Shahnaz Bibi, Maria Bitner-Glindzicz, Marta Bleda, Harm Boggard, Paula Bolton-Maggs, Claire Booth, John R Bradley, Angie Brady, Matthew Brown, Michael Browning, Christine Bryson, Siobhan Burns, Paul Calleja, Natalie Canham, Jenny Carmichael, Keren Carss, Mark Caulfield, Elizabeth Chalmers, Anita

Chandra, Patrick Chinnery, Manali Chitre, Colin Church, Emma Clement, Naomi Clements-Brod, Virginia Clowes, Gerry Coghlan, Peter Collins, Nichola Cooper, Amanda Creaser-Myers, Rosa DaCosta, Louise Daugherty, Sophie Davies, John Davis, Minka De Vries, Patrick Deegan, Sri VV Deevi, Charu Deshpande, Lisa Devlin, Eleanor Dewhurst, Rainer Doffinger, Natalie Dormand, Elizabeth Drewe, David Edgar, William Egner, Wendy N Erber, Marie Erwood, Tamara Everington, Remi Favier, Helen Firth, Debra Fletcher, Frances Flinter, James C Fox, Amy Frary, Kathleen Freson, Bruce Furie, Abigail Furnell, Daniel Gale, Alice Gardham, Michael Gattens, Neeti Ghali, Pavandeep K Ghataorhe, Rohit Ghurye, Simon Gibbs, Kimberley Gilmour, Paul Gissen, Sarah Goddard, Keith Gomez, Pavel Gordins, Stefan Gräf, Daniel Greene, Alan Greenhalgh, Andreas Greinacher, Sofia Grigoriadou, Detelina Grozeva, Scott Hackett, Charaka Hadinnapola, Rosie Hague, Matthias Haimel, Csaba Halmagyi, Tracey Hammerton, Daniel Hart, Grant Hayman, Johan WM Heemskerk, Robert Henderson, Anke Hensiek, Yvonne Henskens, Archana Herwadkar, Simon Holden, Muriel Holder, Susan Holder, Fengyuan Hu, Aarnoud Huissoon, Marc Humbert, Jane Hurst, Roger James, Stephen Jolles, Dragana Josifova, Rashid Kazmi, David Keeling, Peter Kelleher, Anne M Kelly, Fiona Kennedy, David Kiely, Nathalie Kingston, Ania Koziell, Deepa Krishnakumar, Taco W Kuijpers, Dinakantha Kumararatne, Manju Kurian, Michael A Laffan, Michele P Lambert, Hana Lango Allen, Allan Lawrie, Sara Lear, Melissa Lees, Claire Lentaigne, Ri Liesner, Rachel Linger, Hilary Longhurst, Lorena Lorenzo, Rajiv Machado, Rob Mackenzie, Robert MacLaren, Eamonn Maher, Jesmeen Maimaris, Sarah Mangles, Ania Manson, Rutendo Mapeta, Hugh S Markus, Jennifer Martin, Larahmie Masati, Mary Mathias, Vera Matser, Anna Maw, Elizabeth McDermott, Coleen McJannet, Stuart Meacham, Sharon Meehan, Karyn Megy, Sarju Mehta, Michel Michaelides, Carolyn M Millar, Shahin Moledina, Anthony Moore, Nicholas Morrell, Andrew Mumford, Sai Murng, Elaine Murphy, Sergey Nejentsev, Sadia Noorani, Paquita Nurden, Eric Oksenhendler, Willem H Ouwehand, Sofia Papadia, Soo-Mi Park,

Alasdair Parker, John Pasi, Chris Patch, Joan Paterson, Jeanette Payne, Andrew Peacock, Kathelijne Peerlinck, Christopher J Penkett, Joanna Pepke-Zaba, David J Perry, Val Pollock, Gary Polwarth, Mark Ponsford, Waseem Qasim, Isabella Quinti, Stuart Rankin, Julia Rankin, F Lucy Raymond, Karola Rehnstrom, Evan Reid, Christopher J Rhodes, Michael Richards, Sylvia Richardson, Alex Richter, Irene Roberts, Matthew Rondina, Elisabeth Rosser, Catherine Roughley, Kevin Rue-Albrecht, Crina Samarghitean, Alba Sanchis-Juan, Richard Sandford, Saikat Santra, Ravishankar Sargur, Sinisa Savic, Sol Schulman, Harald Schulze, Richard Scott, Marie Scully, Suranjith Seneviratne, Carrock Sewell, Olga Shamardina, Debbie Shipley, Ilenia Simeoni, Suthesh Sivapalaratnam, Kenneth Smith, Aman Sohal, Laura Southgate, Simon Staines, Emily Staples, Hans Stauss, Penelope Stein, Jonathan Stephens, Kathleen Stirrups, Sophie Stock, Jay Suntharalingam, R Campbell Tait, Kate Talks, Yvonne Tan, Jecko Thachil, James Thaventhiran, Ellen Thomas, Moira Thomas, Dorothy Thompson, Adrian Thrasher, Marc Tischkowitz, Catherine Titterton, Cheng-Hock Toh, Mark Toshner, Carmen Treacy, Richard Trembath, Salih Tuna, Wojciech Turek, Ernest Turro, Chris Van Geet, Marijke Veltman, Julie Vogt, Julie von Ziegenweldt, Anton Vonk Noordegraaf, Emma Wakeling, Ivy Wanjiku, Timothy Q Warner, Evangeline Wassmer, Hugh Watkins, Andrew Webster, Steve Welch, Sarah Westbury, John Wharton, Deborah Whitehorn, Martin Wilkins, Lisa Willcocks, Catherine Williamson, Geoffrey Woods, John Wort, Nigel Yeatman, Patrick Yong, Tim Young, Ping Yu

Acknowledgments

This work was supported by The National Institute for Health Research England (NIHR) for the NIHR BioResource – Rare Diseases project (grant number RG65966). The Moorfields Eye Hospital cohort of patients, clinical and imaging data, were

ascertained and collected with the support of grants from the National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust and UCL Institute of Ophthalmology, Moorfields Eye Hospital Special Trustees, Moorfields Eye Charity, the Foundation Fighting Blindness (USA), and Retinitis Pigmentosa Fighting Blindness. MM is a recipient of an FFB Career Development Award. EM's is supported by UCLH/UCL NIHR Biomedical Research Centre. FLR and DG are supported by Cambridge NIHR Biomedical Research Centre.

The authors declare no conflict of interest.

Accession Numbers

The accession number for the high throughput sequencing data reported in this paper is EGAD00001002656 as a subset of the NIHR BioResource Rare Disease Consortium Data.

Web Resources

Ensembl Genome Browser, http://www.ensembl.org

ExAC, http://exac.broadinstitute.org

IGV, http://www.broadinstitute.org/igv/home

Ocular Genome Institute, https://oculargenomics.meei.harvard.edu/index.php/ret-trans/110-human-retinal-transcriptome

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

PLINK, http://pngu.mgh.harvard.edu/purcell/plink/

RetNet, http://www.sph.uth.tmc.edu/RetNet

UCSC genome browser, http://genome.ucsc.eduUK10K, http://www.uk10k.org

References

1. Liew, G., Michaelides, M., and Bunce, C. (2014). A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. BMJ Open *4*, e004015.

2. Ellingford, J.M., Sergouniotis, P.I., Lennon, R., Bhaskar, S., Williams, S.G., Hillman, K.A., O'Sullivan, J., Hall, G., Ramsden, S.C., Lloyd, I.C., et al. (2015). Pinpointing clinical diagnosis through whole exome sequencing to direct patient care: A case of Senior-Loken syndrome. Lancet *385*, 1916.

3. Fitzgerald, T.W., Gerety, S.S., Jones, W.D., van Kogelenberg, M., King, D.A., McRae, J., Morley, K.I., Parthiban, V., Al-Turki, S., Ambridge, K., et al. (2014). Large-scale discovery of novel genetic causes of developmental disorders. Nature *519*, 223–228.

4. Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P. a, Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N. Engl. J. Med. *369*, 1502–1511.

5. Beaulieu, C.L., Majewski, J., Schwartzentruber, J., Samuels, M.E., Fernandez, B.A., Bernier, F.P., Brudno, M., Knoppers, B., Marcadier, J., Dyment, D., et al. (2014). FORGE Canada consortium: Outcomes of a 2-year national rare-disease

gene-discovery project. Am. J. Hum. Genet. *94*, 809–817.

6. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W.M., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature *511*, 344–347.

7. Stavropoulos, D.J., Merico, D., Jobling, R., Bowdin, S., Monfared, N., Thiruvahindrapuram, B., Nalpathamkalam, T., Pellecchia, G., Yuen, R.K.C., Szego, M.J., et al. (2016). Whole-genome sequencing expands diagnostic utility and improves clinical management in paediatric medicine. Npj Genomic Med. *1*, 15012.

8. The 100,000 Genomes Project [Internet]. [cited 2016 Jul 22]. Available from: https://www.genomicsengland.co.uk/the-100000-genomes-project/

9. Berger, W., Kloeckener-Gruissem, B., and Neidhardt, J. (2010). The molecular basis of human retinal and vitreoretinal diseases. Prog. Retin. Eye Res. *29*, 335–375.

10. Huang, X.-F., Huang, F., Wu, K.-C., Wu, J., Chen, J., Pang, C.-P., Lu, F., Qu, J., and Jin, Z.-B. (2015). Genotype-phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing. Genet. Med. *17*, 271–278.

11. Patel, N., Aldahmesh, M. a., Alkuraya, H., Anazi, S., Alsharif, H., Khan, A.O., Sunker, A., Al-mohsen, S., Abboud, E.B., Nowilaty, S.R., et al. (2015). Expanding the clinical, allelic, and locus heterogeneity of retinal dystrophies. Genet. Med. 1–9.

12. Eisenberger, T., Neuhaus, C., Khan, A.O., Decker, C., Preising, M.N., Friedburg, C., Bieg, A., Gliem, M., Charbel Issa, P., Holz, F.G., et al. (2013). Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: The example of retinal dystrophies. PLoS One *8*,

e78496.

13. Tiwari, A., Bahr, A., Bähr, L., Fleischhauer, J., Zinkernagel, M.S., Winkler, N., Barthelmes, D., Berger, L., Gerth-Kahlert, C., Neidhardt, J., et al. (2016). Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. Sci. Rep. *6*, 28755.

14. O'Sullivan, J., Mullaney, B.G., Bhaskar, S.S., Dickerson, J.E., Hall, G., O'Grady, A., Webster, A., Ramsden, S.C., and Black, G.C. (2012). A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. J Med Genet *49*, 322–326.

15. Khan, K.N., Chana, R., Ali, N., Wright, G., Webster, A.R., Moore, A.T., and Michaelides, M. (2016). Advanced diagnostic genetic testing in inherited retinal disease: experience from a single tertiary referral centre in the UK National Health Service. Clin. Genet. 1–8.

16. Consugar, M.B., Navarro-Gomez, D., Place, E.M., Bujakowska, K.M., Sousa, M.E., Fonseca-Kelly, Z.D.Z.D., Taub, D.G., Janessian, M., Wang, D.Y., Au, E.D., et al. (2015). Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. Genet. Med. *17*, 253–261.

17. Ellingford, J.M., Barton, S., Bhaskar, S., Williams, S.G., Sergouniotis, P.I., O'Sullivan, J., Lamb, J.A., Perveen, R., Hall, G., Newman, W.G., et al. (2016). Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. Ophthalmology *123*, 1143–1150.

18. Westbury, S.K., Turro, E., Greene, D., Lentaigne, C., Kelly, A.M., Bariana, T.K., Simeoni, I., Pillois, X., Attwood, A., Austin, S., et al. (2015). Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with

unexplained bleeding and platelet disorders. Genome Med. *7*, 36.

19. Hull, S., Owen, N., Islam, F., Tracey-White, D., Plagnol, V., Holder, G.E., Michaelides, M., Carss, K., Raymond, F.L., Rozet, J.-M., et al. (2016). Nonsyndromic Retinal Dystrophy due to Bi-Allelic Mutations in the Ciliary Transport Gene *IFT140*. Investig. Opthalmology Vis. Sci. *57*, 1053.

20. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2015). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220–1222.

21. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

22. Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D. a, Ambridge, K., Barrett, D.M., Bayzetinova, T., et al. (2014). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet *6736*, 1–10.

23. Stenson, P.D., Mort, M., Ball, E. V., Shaw, K., Phillips, A.D., and Cooper, D.N. (2014). The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum. Genet. *133*, 1–9.

24. 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. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

25. Zarrei, M., Macdonald, J.R., Merico, D., and Scherer, S.W. (2015). A copy number variation map of the human genome. Nat. Publ. Gr. *16*, 172–183.

26. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

27. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

28. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Bëroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: An online bioinformatics tool to predict splicing signals. Nucleic Acids Res. *37*, 1–14.

29. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved Splice Site Detection in Genie. J. Comput. Biol. *4*, 311–323.

30. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

31. Henderson, R.H., Waseem, N., Searle, R., van der Spuy, J., Russell-Eggitt, I., Bhattacharya, S.S., Thompson, D.A., Holder, G.E., Cheetham, M.E., Webster, A.R., et al. (2007). An assessment of the apex microarray technology in genotyping patients with Leber congenital amaurosis and early-onset severe retinal dystrophy. Invest. Ophthalmol. Vis. Sci. *48*, 5684–5689.

32. Braun, T. a., Mullins, R.F., Wagner, A.H., Andorf, J.L., Johnston, R.M., Bakall, B.B., Deluca, A.P., Fishman, G. a., Lam, B.L., Weleber, R.G., et al. (2013). Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt

disease. Hum. Mol. Genet. *22*, 5136–5145.

33. Sangermano, R., Bax, N.M., Bauwens, M., van den Born, L.I., De Baere, E., Garanto, A., Collin, R.W.J., Goercharn-Ramlal, A.S.A., den Engelsman-van Dijk, A.H.A., Rohrschneider, K., et al. (2016). Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. Ophthalmology *123*, 1375–1385.

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

35. Van Den Hurk, J.A.J.M., Van De Pol, D.J.R., Wissinger, B., Van Driel, M.A., Hoefsloot, L.H., De Wijs, I.J., Van Den Born, I., Heckenlively, J.R., Brunner, H.G., Zrenner, E., et al. (2003). Novel types of mutation in the choroideremia (CHM) gene: A full-length L1 insertion and an intronic mutation activating a cryptic exon. Hum. Genet. *113*, 268–275.

36. Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R.B., Xu, C., Futema, M., Lawson, D., et al. (2015). The UK10K project identifies rare variants in health and disease. Nature *526*, 82–90.

37. Turner, T.N., Hormozdiari, F., Duyzend, M.H., McClymont, S.A., Hook, P.W., Iossifov, I., Raja, A., Baker, C., Hoekzema, K., Stessman, H.A., et al. (2016). Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. Am. J. Hum. Genet. *98*, 58–74.

38. Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q.B., Antipenko, A., Shang, L., Boisson, B., Casanova, J.-L., and Abel, L. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc.

Natl. Acad. Sci. U. S. A. *112*, 5473–5478.

39. Lelieveld, S.H., Spielmann, M., Mundlos, S., Veltman, J. a, and Gilissen, C. (2015). Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. Hum. Mutat. *36*, 815–822.

40. Hoischen, A., Gilissen, C., Arts, P., Wieskamp, N., Van Vliet, W. Der, Vermeer, S., Steehouwer, M., De Vries, P., Meijer, R., Seiqueros, J., et al. (2010). Massively parallel sequencing of ataxia genes after array-based enrichment. Hum. Mutat. *31*, 492–499.

41. Benjamini, Y., and Speed, T.P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. *40*, 1–14.

42. Chilamakuri, C.S.R., Lorenz, S., Madoui, M.-A., Vodák, D., Sun, J., Hovig, E., Myklebost, O., and Meza-Zepeda, L.A. (2014). Performance comparison of four exome capture systems for deep sequencing. BMC Genomics *15*, 449.

43. Patwardhan, A., Harris, J., Leng, N., Bartha, G., Church, D.M., Luo, S., Haudenschild, C., Pratt, M., Zook, J., Salit, M., et al. (2015). Achieving high-sensitivity for clinical applications using augmented exome sequencing. Genome Med. *7*, 71.

44. Pusch, C.M., Broghammer, M., Jurklies, B., Besch, D., and Jacobi, F.K. (2002). Ten novel ORF15 mutations confirm mutational hot spot in the RPGR gene in European patients with X-linked retinitis pigmentosa. Hum. Mutat. *20*, 405.

45. McRae, J.F., Clayton, S., Fitzgerald, T.W., Kaplanis, J., Prigmore, E., Rajan, D., Sifrim, A., Aitken, S., Akawi, N., Alvi, M., et al. (2016). Prevalence, phenotype and architecture of developmental disorders caused by de novo mutation. bioRxiv.

46. Arno, G., Hull, S., Carss, K., Dev-Borman, A., Chakarova, C., Bujakowska, K.,

van den Born, I., Robson, A.G., Holder, G.E., Michaelides, M., et al. (2016). Reevaluation of the Retinal Dystrophy Due to Recessive Alleles of RGR With the Discovery of a Cis-Acting Mutation in CDHR1. Invest. Ophthalmol. Vis. Sci. *57*, 4806–4813.

47. Chiang, J. (Pei-W., Lamey, T., McLaren, T., Thompson, J. a, Montgomery, H., and De Roach, J. (2015). Progress and prospects of next-generation sequencing testing for inherited retinal dystrophy. Expert Rev. Mol. Diagn. *15*, 1269–1275.

48. Ba-Abbad, R., Arno, G., Carss, K., Stirrups, K., Penkett, C.J., Moore, A.T., Michaelides, M., Raymond, F.L., Webster, A.R., and Holder, G.E. (2016). Mutations in CACNA2D4 cause distinctive retinal dysfunction in humans. Ophthalmology *123*, 668–671.e2.

49. Mayer, A.K., Rohrschneider, K., Strom, T.M., Glöckle, N., Kohl, S., Wissinger, B., and Weisschuh, N. (2015). Homozygosity mapping and whole-genome sequencing reveals a deep intronic PROM1 mutation causing cone-rod dystrophy by pseudoexon activation. Eur J Hum Genet. doi: 10.1038/ejhg.2015.144.

50. den Hollander, A.I., Koenekoop, R.K., Yzer, S., Lopez, I., Arends, M.L., Voesenek, K.E.J., Zonneveld, M.N., Strom, T.M., Meitinger, T., Brunner, H.G., et al. (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. Am. J. Hum. Genet. *79*, 556–561.

51. Yang, J., Liu, X., Yue, G., Adamian, M., Bulgakov, O., and Li, T. (2002). Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. J. Cell Biol. *159*, 431–440.

52. Cheng, C.W., Chow, R.L., Lebel, M., Sakuma, R., Cheung, H.O.L., Thanabalasingham, V., Zhang, X., Bruneau, B.G., Birch, D.G., Hui, C.C., et al. (2005). The Iroquois homeobox gene, Irx5, is required for retinal cone bipolar cell

development. Dev. Biol. *287*, 48–60.

Figures

Figure 1: Case study 1, WGS increases power to detect SVs, compared to WES. A: One individual (W000325) with retinitis pigmentosa has a pathogenic heterozygous deletion within *EYS* (6:65602819-65658187). IGV plots showing the deleted region in both WGS and WES data in the individual. The deletion was identified by WGS due to the drop in coverage and increased distance between mate-paired reads. However, it was not identified from WES data. B: The individual also has a pathogenic missense variant ENST00000503581.1:c.6473T>C (p.Leu2158Pro).

Figure 2: Case study 2, identification of pathogenic tandem duplication and likely UPD by WGS. One individual (W000170) with atypical, early-onset retinal dystrophy, has a homozygous pathogenic tandem duplication within *KCNV2*, likely due to partial maternal uniparental isodisomy of chromosome 9. A: IGV plot showing the 184bp tandem duplication of 9:2717844-2718028 in WGS data. B: The tandem duplication was confirmed by PCR and Sanger sequencing. The arrows represent the positions of the primers. C: There is a ~25Mb region of homozygosity in chromosome 9 in this individual, which encompasses *KCNV2*. Approximate coordinates of this region of homozygosity are 9:2100000-27400000. D: Sanger sequencing also confirms a homozygous combined in-frame insertion/deletion in *KCNV2*, within the tandem duplication: ENST00000382082.3:c.222_232delGGACCAGCAGGinsGGTCACCACCACCTTGG (ENSP00000371514.3:p.Asp75_Gln77delinsValThrThrThrLeu). The mother of the affected individual is heterozygous for this variant, but the father is homozygous for the reference allele (not shown).

Figure 3: Case study 3, identification and characterisation of compound heterozygous deletions in *EYS* by WGS. A: one individual (W000164) with retinitis pigmentosa has compound heterozygous deletions in *EYS*, shown by the green dashed lines. B: Sanger sequencing confirmed both deletions. Sequencing across the breakpoint (vertical dashed line) of deletion 2 is shown.

Figure 4: Regions of IRD-associated genes that are low or high in GC content have significantly higher coverage in WGS data than in the ExAC WES project. Coverage was calculated across protein-coding regions of autosomal IRD-associated genes (Ensembl canonical transcript), split into 50bp bins. Coverage shown is relative to the average coverage of each dataset. Error bars show standard deviation. * = $p < 1 \times 10^{-15}$.

Figure 5: Case study 4, Identification of pathogenic variants by WGS in GC-rich regions not covered by WES. One individual (G004991) with Leber's congenital amaurosis has pathogenic compound heterozygous variants ENST00000254854.4:c.238_252delGCCGCCGCCCGCCTG (p.Ala80_Leu84del) and ENST00000254854.4:c.307G>A (p.Glu103Lys) in exon 1 of *GUCY2D*, which is 76% GC-rich. This exon is not covered by WES in our cohort, as demonstrated by the WES data of a control sample shown here. It is also not well covered in ExAC.

Figure 6: The deep intronic *CHM* variant c.315-1536A>G results in the inclusion of a cryptic exon. A: Sequence of the cryptic exon included by the deep intronic

CHM variant chrX:85,220,593T>C (ENST00000357749.2:c.315-1536A>G). B: RT-PCR showing increased size of fragment due to inclusion of the cryptic exon.

Tables

Sequencing method	Total cases	Cases solved	Cases partially solved	Cases unsolved
WGS	605	331 (55%)	31 (5%)	243 (40%)
WES	72	59 (82%)	3 (4%)	10 (14%)
WES and WGS	45	14 (31%)	2 (4%)	29 (64%)
TOTAL	722	404 (56%)	36 (5%)	282 (39%)

Table 1: Pathogenic variant detection rates by sequencing technology. WES alone solved 59/117 (50%) cases. Subsequently, 45 of the 58 cases which are unsolved by WES also underwent WGS. Partially solved cases either have one likely pathogenic variant in a gene with biallelic inheritance, or more than two heterozygous variants in a gene with biallelic inheritance, or a variant that only appears to explain part of the phenotype.

Phenotype	Total cases	Cases solved	Cases partially solved	Cases unsolved
RP	311	168 (54%)	11 (4%)	132 (42%)
RD	101	55 (54%)	5 (5%)	41 (41%)
CRD	53	29 (55%)	3 (6%)	21 (40%)
Stargardt	45	27 (60%)	10 (22%)	8 (18%)
Macular dystrophy	37	18 (49%)	1 (3%)	18 (49%)
Usher	37	31 (84%)	2 (5%)	4 (11%)
Other	27	10 (37%)	0 (0%)	17 (63%)
CSNB	26	23 (88%)	0 (0%)	3 (12%)
Cone dystrophy	21	6 (29%)	1 (5%)	14 (67%)
Multiple	21	6 (29%)	3 (14%)	12 (57%)
LCA	18	16 (89%)	0 (0%)	2 (11%)
Achromatopsia	9	6 (67%)	0 (0%)	3 (33%)
Albinism	8	6 (75%)	0 (0%)	2 (25%)
FEVR	8	3 (38%)	0 (0%)	5 (62%)
TOTAL	722	404 (56%)	36 (5%)	282 (39%)

Table 2: Pathogenic variant detection rate by phenotype. RP = retinitis pigmentosa, RD = retinal dystrophy, CRD = cone-rod dystrophy, Other = any phenotype with frequency of less than eight, CSNB = congenital stationary night blindness, Multiple = more than one phenotype including syndromic cases, LCA = Leber congenital amaurosis, FEVR = Familial exudative vitreoretinopathy.

Likely ethnicity	Total cases	Cases solved	Cases partially solved	Cases unsolved
EUR	467	259 (55%)	23 (5%)	185 (40%)
SAS	123	70 (57%)	4 (3%)	49 (40%)
AFR	43	13 (30%)	4 (9%)	26 (60%)
EAS	13	1 (8%)	1 (8%)	11 (85%)
AMR	4	2 (50%)	1 (25%)	1 (25%)
TOTAL	650	345 (53%)	33 (5%)	272 (42%)

Table 3: Pathogenic variant detection rate by ethnicity. Likely ethnicity estimated from WGS data using principal component analysis. Table includes individuals who had WGS only. EUR = European, SAS = South Asian, AFR = African, EAS = East Asian, AMR = Ad Mixed American.

Gene	Total cases	Cases solved	Cases partially solved
ABCA4	73	57	16
USH2A	61	50	11
EYS	16	15	1
RP1	16	16	0
CACNA1F	13	13	0
RPGR	13	13	0
CRB1	12	11	1
CNGB1	9	8	1
MYO7A	8	8	0
PDE6B	8	8	0
BBS1	7	7	0
CERKL	7	7	0
CNGB3	7	7	0
PROM1	7	7	0
RHO	7	7	0
CDHR1	6	6	0
CLN3	6	6	0
PRPF31	6	6	0
PRPH2	6	6	0
RP2	5	5	0
TRPM1	5	5	0

Table 4: Number of solved and partially solved cases by gene. Only genes with pathogenic variants in five or more individuals are shown.

Individual	Phenotype	Sex	Ethnicity	Gene	Variant genomic	Variant HGVSc	Variant HGVSp	GT	Consequence	Description
G001284	Multiple	F	SAS	*SCAPER*	15:76998312G>A	ENST00000563290.1:c.2179C>T	p.Arg727Ter	0/1	stop_gained	S-phase cyclin A-associated protein in the ER
G001284	Multiple	F	SAS	*SCAPER*	15:77064214CA>C	ENST00000563290.1:c.1116delT	p.Val373SerfsTer21	0/1	frameshift	S-phase cyclin A-associated protein in the ER
G001298	RD	F	SAS	*FUT5*	19:5867071G>T	ENST00000252675.5:c.666C>A	p.Tyr222Ter	1/1	stop_gained	fucosyltransferase 5 (alpha (1,3) fucosyltransferase)
G001298	RD	F	SAS	*PODNL1*	19:14046820C>CAGCT	ENST00000339560.5:c.374_377dupAGCT	p.Gln127AlafsTer119	1/1	frameshift	podocan-like protein 1-like
G001411	RP	M	AFR	*NAALADL1*	11:64812774G>GC	ENST00000358658.3:c.2191dupG	p.Ala731GlyfsTer9	0/1	frameshift	N-acetylated alpha-linked acidic dipeptidase-like 1
G001411	RP	M	AFR	*NAALADL1*	11:64822078G>A	ENST00000358658.3:c.736C>T	p.Arg246Ter	0/1	stop_gained	N-acetylated alpha-linked acidic dipeptidase-like 1
G005002	CRD	M	SAS	*WASF3*	13:27216381GTGTTTTCAATTTTCAGATTGTGAACCA>G	ENST00000335327.5:c.-10-14_3delTTTTCAATTTTCAGATTGTGAACCATG	NA	1/1	splice_acceptor	WAS protein family, member 3
G005019	Usher	F	SAS	*PLD4*	14:105395186GC>G	ENST00000392593.4:c.388delC	p.Gln130ArgfsTer108	0/1	frameshift	phospholipase D family, member 4
G005019	Usher	F	SAS	*PLD4*	14:105398102CTGTCCCCA>C	ENST00000392593.4:c.937_944delTGTCCCCA	p.Cys313GlyfsTer167	0/1	frameshift	phospholipase D family, member 4
G005203	RP	F	SAS	*FAM71A*	1:212799206T>TGCAG	ENST00000294829.3:c.991_994dupGGCA	p.Thr332ArgfsTer88	1/1	frameshift	family with sequence similarity 71, member A
G005251	Cone dystrophy	M	SAS	*POMZP3*	7:76240807T>TG	ENST00000310842.4:c.538dupC	p.Gln180ProfsTer14	1/1	frameshift	POM121 and ZP3 fusion
G005492	RP	F	EUR	*IRX5*	16:54967694CTAAAG>C	ENST00000394636.4:c.1362_1366delTAAAG	p.Lys455ProfsTer19	1/1	frameshift	iroquois homeobox 5
G005513	Stargardt	M	EUR	*ITIH2*	10:7776934CT>C	ENST00000358415.4:c.1838delT	p.Leu613ArgfsTer5	0/1	frameshift	inter-alpha-trypsin inhibitor heavy chain 2
G005513	Stargardt	M	EUR	*ITIH2*	10:7780709CAT>C	ENST00000358415.4:c.2084_2085delAT	p.His695ArgfsTer5	0/1	frameshift	inter-alpha-trypsin inhibitor heavy chain 2
G005514	RP	M	AFR	*SLC37A3*	7:140064249G>A	ENST00000326232.9:c.334C>T	p.Arg112Ter	1/1	stop_gained	solute carrier family 37, member 3
G007696	CRD	M	SAS	*NUMB*	14:73746066G>GC	ENST00000355058.3:c.1162dupG	p.Ala388GlyfsTer6	1/1	frameshift	numb homolog (Drosophila)
G007696	CRD	M	SAS	*FAM57B*	16:30038139GC>G	ENST00000380495.4:c.234delG	p.Gln79AsnfsTer43	1/1	frameshift	family with sequence similarity 57, member B
G007723	RP	M	SAS	*FOXI2*	10:129536034AC>A	ENST00000388920.4:c.498delC	p.Asp166GlufsTer87	1/1	frameshift	forkhead box I2
G008152	RP	M	EUR	*CROCC*	1:17292217G>A	ENST00000375541.5:c.4406-1G>A	NA	1/1	splice_acceptor	ciliary rootlet coiled-coil, rootletin
W000146	RP	F	EUR	*CCZ1B*	7:6844600C>A	ENST00000316731.8:c.1075G>TENST00000375541.5:c.4406-1G>A	p.Glu359Ter	1/1	stop_gained	CCZ1 vacuolar protein trafficking and biogenesis associated homolog B (S. cerevisiae)
W000278	Other	F	SAS	*OR2M7*	1:248487484AG>A	ENST00000317965.2:c.386delC	p.Pro129LeufsTer2	1/1	frameshift	olfactory receptor, family 2, subfamily M, member 7
W000375	RP	M	SAS	*PRTFDC1*	10:25231367T>C	ENST00000320152.6:c.49-2A>G	NA	1/1	splice_acceptor	phosphoribosyl transferase domain containing 1
W000375	RP	M	SAS	*UBAP1L*	15:65395024T>G	ENST00000559089.1:c.121-2A>C	NA	1/1	splice_acceptor	ubiquitin associated protein 1-like

Table 5: High-impact, likely biallelic variants in 19 genes in unsolved cases. Genomic coordinates refer to genome build GRCh37. Analysis is limited to cases that underwent WGS in whom pathogenic variants in known genes were not detected, and whose family history is not inconsistent with recessive inheritance of disease. RP = retinitis pigmentosa, RD = retinal dystrophy, CRD = cone-rod dystrophy, Other = any phenotype with frequency of less than eight individuals, Multiple = more than one phenotype including syndromic cases.