



Diagnostic genome sequencing improves diagnostic yield: a prospective single-centre study in 1000 patients with inherited eye diseases

Nicole Weisschuh,¹ Pascale Mazzola,² Theresia Zuleger,² Karin Schaeferhoff,² Laura Kühlewein,³ Friederike Kortüm,³ Dennis Witt,² Alexandra Liebmann,² Ruth Falb,² Lisa Pohl,³ Milda Reith,³ Lara G Stühn,² Miriam Bertrand,² Amelie Müller,² Nicolas Casadei ,² Olga Kelemen,² Carina Kelbsch,³ Christoph Kernstock,³ Paul Richter,³ Françoise Sadler,³ German Demidov,² Leon Schütz,² Jakob Admard,² Marc Sturm,² Ute Grasshoff,² Felix Tonagel,³ Tilman Heinrich,^{2,4} Fadi Nasser,³ Bernd Wissinger,¹ Stephan Ossowski,^{2,5} Susanne Kohl,¹ Olaf Riess,^{2,6} Katarina Stingl,³ Tobias B Haack ^{2,6}

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jmg-2023-109470>).

For numbered affiliations see end of article.

Correspondence to

Dr Tobias B Haack, University of Tübingen, Tübingen 72074, Germany; tobias.haack@med.uni-tuebingen.de
Dr Katarina Stingl; Katarina.Stingl@med.uni-tuebingen.de

NW, PM and TZ contributed equally.

NW, PM and TZ are joint first authors.
SK, OR, KS and TBH are joint senior authors.

Received 20 June 2023
Accepted 10 September 2023
Published Online First 21 September 2023



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Weisschuh N, Mazzola P, Zuleger T, et al. *J Med Genet* 2024;**61**:186–195.

ABSTRACT

Purpose Genome sequencing (GS) is expected to reduce the diagnostic gap in rare disease genetics. We aimed to evaluate a scalable framework for genome-based analyses ‘beyond the exome’ in regular care of patients with inherited retinal degeneration (IRD) or inherited optic neuropathy (ION).

Methods PCR-free short-read GS was performed on 1000 consecutive probands with IRD/ION in routine diagnostics. Complementary whole-blood RNA-sequencing (RNA-seq) was done in a subset of 74 patients. An open-source bioinformatics analysis pipeline was optimised for structural variant (SV) calling and combined RNA/DNA variation interpretation.

Results A definite genetic diagnosis was established in 57.4% of cases. For another 16.7%, variants of uncertain significance were identified in known IRD/ION genes, while the underlying genetic cause remained unresolved in 25.9%. SVs or alterations in non-coding genomic regions made up for 12.7% of the observed variants. The RNA-seq studies supported the classification of two unclear variants.

Conclusion GS is feasible in clinical practice and reliably identifies causal variants in a substantial proportion of individuals. GS extends the diagnostic yield to rare non-coding variants and enables precise determination of SVs. The added diagnostic value of RNA-seq is limited by low expression levels of the major IRD disease genes in blood.

INTRODUCTION

Although protein-coding regions represent only 1–2% of the human genome, they harbour an estimated 85% of annotated pathogenic variants.^{1,2} Despite these numbers, genome sequencing (GS) usually achieves a higher diagnostic yield than sequencing approaches that focus on exonic regions, not least because of its more homogeneous coverage^{3,4} and higher efficiency in capturing genomic regions that are particularly high or low

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Genome sequencing has the potential to boost diagnostic sensitivity, but requires structural and bioinformatic adaptations that can be challenging to implement in routine care.

WHAT THIS STUDY ADDS

⇒ Using inherited eye diseases, we prospectively showcase the clinical utility and added diagnostic value of comprehensive genome analyses ‘beyond the exome’. We define the genes and types of genomic variation involved in the largest genome-sequenced cohort of inherited retinal degeneration or inherited optic neuropathy to date.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study suggests that genome-based diagnostics should be considered as an appropriate first-line diagnostic tool that improves our ability to detect different types of disease-causing DNA variants, providing the technical basis for defining the genetic architecture of inherited eye diseases.

inguanine or cytosine content.^{3,5} The real advantage of GS is of course its ability to detect variants outside the protein-coding regions. In addition, the homogeneous coverage by GS aids the detection of CNVs by semiquantitative algorithms and callers based on the discovery of discordant pair and split-read alignments. Even more important, GS is able to detect copy-neutral rearrangements such as inversions and translocations. Numerous studies have found that the diagnostic yield can be improved by the use of GS, for instance in intellectual disability,^{6,7} paediatric disease,^{8,9} neurological disorders^{10,11} and inherited retinal degeneration (IRD).^{5,12,13}

The latter disease group comprises a range of disorders characterised by progressive degeneration or stationary dysfunction of the outer retina and/or the retinal pigment epithelium. IRD affects approximately 1 in 3000 individuals in North America and Europe.^{14–17} Genetic diagnostics is hampered by the considerable clinical overlap of disease entities and the complexity in the genetic causes (>270 ‘disease genes’ (RetNet, <https://sph.uth.edu/retnet>)). Typical for Mendelian disorders, diagnostic rates vary widely by phenotype and are inversely correlated with the level of genetic heterogeneity.¹⁸ The genetic diagnostic rate can exceed 90% in certain clinically well-defined phenotypes of IRDs (eg, choroideraemia),¹⁹ while the most common subtype of IRD, retinitis pigmentosa, shows an extreme level of genetic heterogeneity, resulting in lower diagnostic rates.^{20,21}

Another group of inherited eye diseases with less pronounced clinical heterogeneity is inherited optic neuropathy (ION). This disease group mainly affects visual acuity, central visual fields and colour vision due to a progressive loss of retinal ganglion cells and their axons that form the optic nerve. Among IONs, Leber hereditary optic neuropathy (LHON) and dominant optic atrophy (DOA) are the two most common disorders seen in clinical practice.²² Three variants in the mitochondrial DNA account for ~95% of all cases of LHON, whereas about 70% of DOA cases harbour pathogenic variants in *OPA1*.²³ Undoubtedly, next-generation sequencing (NGS) approaches have accelerated the identification of the underlying disease-causing variants in IRD and ION.^{24,25} Nevertheless, 24–52% of IRD cases and up to 78% of ION cases remain genetically undiagnosed despite rather comprehensive work-up such as targeted sequencing applying specific capture panels or exome sequencing (ES).^{24–29}

Since 2019, individuals affected with IRD or ION recruited at the University Eye Hospital Tübingen receive genetic diagnostic testing based on GS. Having reached a significant number of 1000 datasets, the aim of the present study was to provide the mutational spectrum observed in an unselected cohort of IRD and ION cases and, in particular, to evaluate the advantages of GS compared with targeted approaches. Given the notoriously challenging annotation of intronic variants with respect to their functional consequences, the potential diagnostic added value of complementary RNA sequencing was additionally tested in a subgroup.

MATERIALS AND METHODS

Cohort

All individuals in this study have been exclusively examined at a specialised outpatient clinic for IRD and ION established at the University Eye Hospital Tübingen. Enrolment of the entire cohort included consecutive admissions from January 2019 to September 2021. Only individuals who did not have previous genetic confirmation of the cause of their disease were recruited. Sampling was random and solely based on the patient’s interest in genetic testing and consent to a scientific data use of the results. Blood samples were sent to the genetic testing facility along with the patient’s informed consent and clinical diagnosis.

A total of 1000 affected individuals underwent genetic testing with GS. Among them, 921 were tested as singletons (ie, no other family member was sequenced). In 26 families, 2 affected individuals were tested, and in 3 families, 3 affected individuals were tested. Eighteen cases were sequenced as trios (affected child and unaffected parents), but their parents were not included in the 1000 cases.

Each patient underwent a comprehensive ophthalmological examination. The decision as to which examination was

performed in each patient was made on an individual basis, considering the patient’s compliance and the course of the disease. Individual examinations were as follows: assessment of best-corrected visual acuity using early treatment diabetic retinopathy study charts, semiautomated 90° kinetic visual field examination using the III4e and I4e isopters (Octopus 900; Haag-Streit, Köniz, Switzerland), colour vision testing using panel D15 tests and full-field electroretinography testing according to International Society for Clinical Electrophysiology of Vision standards (Espion; Diagnosys, Lowell, Massachusetts, USA). In addition, fundus and fundus autofluorescence photography, as well as spectral-domain optical coherence tomography (Spectralis HRA+OCT; Heidelberg Engineering, Heidelberg, Germany), were performed.

Prior to genetic testing and depending on their clinical phenotype and self-reported family history, patients were assigned a specific family code, namely ACHM (achromatopsia), ADRP (autosomal dominant retinitis pigmentosa), ARRP (autosomal recessive retinitis pigmentosa), BBS (Bardet-Biedl syndrome), BCM (blue cone monochromacy), BVMD (Best vitelliform macular dystrophy), CACD (central areolar choroidal dystrophy), CD (cone dystrophy), CHM (choroideraemia), CRD (cone-rod dystrophy), CSNB (congenital stationary night blindness), LCA (Leber congenital amaurosis), LHON, MDS (macular dystrophy), MISC (miscellaneous diagnoses; including, but not limited to: Bietti’s crystalline dystrophy, exudative vitreoretinopathy and Wagner’s disease), NYS (nystagmus), OA (ocular albinism), DOA, SCHI (retinoschisis), SRP (simplex retinitis pigmentosa), STGD (Stargardt disease), UD (unclear diagnosis; for example, conflicting phenotypical features), USH I (Usher syndrome type 1), USH II (Usher syndrome type 2) and XRP (X linked retinitis pigmentosa). For the sake of better legibility, these abbreviations are only used in figures and tables, but not in the main text. Note that family codes were not adjusted after genetic testing (ie, if the molecular diagnosis suggested a change or refinement of the clinical diagnosis).

Two hundred and six individuals had previously undergone first-tier genetic testing without discovering the cause of the disease, either by Sanger sequencing of recurrent genes (eg, *RHO* in ADRP or *CNGB3* in ACHM), targeted sequencing of 108 genes associated with IRD using molecular inversion probes³⁰ or targeted sequencing using a comprehensive diagnostic panel of genes associated with eye diseases.²⁸

Familial co-segregation analysis was performed either by conventional Sanger sequencing or multiplex ligation dependent probe amplification whenever possible.

Genome sequencing

Diagnostic genetic testing based on GS was performed at the Institute of Medical Genetics and Applied Genomics (IMGAG), University Hospital Tübingen, Germany. Clinical GS has been accredited at the IMGAG by the DAkkS according to DIN EN ISO 15189:2014. Accreditation is effective for the scope of activities as defined in the certification annex (D-ML-13130-04-00).

Genomic DNA was extracted from whole blood using the FlexiGene DNA kit (Qiagen, Hilden, Germany) and quantified using the Qubit Fluorometer (Thermo Fisher Scientific, Dreieich, Germany). One microgram of genomic DNA was further processed using the TruSeq PCR-Free Library Prep kit (Illumina, Berlin, Germany) and generated libraries were sequenced on a NovaSeq6000 System (Illumina) as 2×150bp paired-end reads to an average 49× coverage.

Mapping and variant calling

The conversion of the sequence data into FASTQ format was done with Illumina bcl2fastq. Adapter sequences were removed using SeqPurge³¹ and the remaining reads were mapped against the human reference genome (GRCh37, hg19) with Burrow Wheeler Aligner (BWA-MEM).³² Optical duplicates were removed with samblaster.³³ Insertions and deletions were realigned using ABRA2.³⁴ Variants were detected with freebayes³⁵ and annotated with Ensembl VEP³⁶ and various internal and external databases. CNVs and structural variants (SVs) were called using ClinCNV and Manta, respectively.^{37,38} For details, refer to the megSAP pipeline (<https://github.com/imgag/megSAP>) developed at the IMGAG, University Hospital Tübingen, Germany.

Variant filtering and interpretation

Various filtering steps were performed to prioritise potentially clinically relevant DNA variants. Filtering was mainly based on the predicted consequences of identified alterations, their listing in disease databases (specifically ClinVar,³⁹ HGMD⁴⁰ and LOVD⁴¹), and their allele frequency ($\leq 1.00\%$). Allele frequency was estimated using an in-house database, 1000 Genomes,⁴² dbSNP,⁴³ the Exome Aggregation Consortium browser⁴⁴ and the Genome Aggregation Database (gnomAD).⁴⁵ In a phenotype-based prioritisation, variants that have been previously associated with the individual's disease or phenotypical characteristics were evaluated. Furthermore, variants that have been predescribed as clinically relevant (ClinVar status 'pathogenic' or 'likely pathogenic', HGMD annotation, 'pathogenic' or 'likely pathogenic' variants from our in-house database) were prioritised. To ensure an efficient diagnosis with high sensitivity, the search criteria were individually adjusted in the context of the research question based on the available additional information (eg, incidence and inheritance of the disease, ethnicity and family history, pathomechanism of candidate genes to be assessed). Variant nomenclature in this study is in accordance with Human Genome Variation Society recommendations. Variant classification in this manuscript was performed using the classification tool from Franklin (<https://franklin.genoox.com>—Franklin by Genoox) which is based on American College of Medical Genetics guidelines.⁴⁶ The following classifications were used: pathogenic (P), likely pathogenic (LP), variant of uncertain significance (VUS), likely benign (LB), benign (B).

In this study, a case was considered unsolved if the identified variants were classified as benign or likely benign or if only a single pathogenic or likely pathogenic allele was identified in a gene associated with autosomal recessive inheritance. A case was considered possibly solved if one or more of the identified variants were classified as variants of uncertain significance. A case was considered solved if the identified variants were classified as pathogenic or likely pathogenic and consistent with the patient's reported phenotype. Some exceptions were made to this rule if expert opinion deviated from the automated variant classification (see also Explanations in the online supplemental table). Note that validation of biallelism was not a prerequisite for the classification as solved. Information on familial co-segregation in individual cases can be found in online supplemental table 1 in the column 'transmission/phase'.

Previously undescribed pathogenic and likely pathogenic variants have been submitted to ClinVar. Accession numbers can be found in online supplemental table 1.

Resequencing of exon 15 of *RPGR*

Due to the highly repetitive sequence of exon 15 of *RPGR*, which is not sufficiently covered by genome short-read sequencing, all unsolved cases with a clinical diagnosis of IRD were subjected to *RPGR* exon 15 (ORF15) resequencing. To this end, a long-range PCR (Roche, Expand Long Template PCR) covering exon 15 was performed. The amplicon was subsequently processed using the Nextera XT DNA library preparation kit (Illumina) and sequenced as 2×100 bp paired-end reads on a NovaSeq6000 System (Illumina).

RNA-sequencing and combined RNA/DNA data analysis

RNA-sequencing (RNA-seq) was performed on 74 cases. RNA was extracted from PAXgene Blood RNA Tube (Qiagen) with QIAasympy PAXgene Blood RNA kits on a QIAasympy SP with the protocol PAXgen RNA V5. From 500 ng of total RNA, mRNAs were enriched using polyA capture on a NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). Libraries were prepared on a Biomek i7 (Beckman Sequencing) using Next Ultra II Directional RNA Library Prep Kits for Illumina (NEB) and NEBNext Globin & rRNA Depletion Kits according to the manufacturer's instructions. The fragment sizes were determined with a Fragment Analyzer (High NGS Fragment 1–6000 bp assay (Agilent)) and the library concentration (approximately 5 ng/μL) was analysed with an Infinite 200Pro (Tecan) and the Quant-iT HS Assay Kit (Thermo Fisher Scientific). The 270 pM cDNA libraries were sequenced as 2×100 bp paired-end reads on an Illumina NovaSeq6000 (Illumina, San Diego, California, USA) with approximately 50 million clusters per sample.

Generated RNA sequences were analysed with respect to aberrant expression, aberrant splicing and allelic imbalance using the megSAP pipeline (V.2022_08, <https://github.com/imgag/megSAP>). In brief, the ngs-bits tool collection (V.2022_07-80, <https://github.com/imgag/ngs-bits>) was used for quality control (ReadQC) and preprocessing (SeqPurge) of fastq files. STAR (V.2.7.10a, <https://www.ncbi.nlm.nih.gov/pubmed/23104886>, <https://github.com/alexdobin/STAR/>) was used for read alignment and detection of splice junctions, which were post-processed with SplicingToBed. After mapping, MappingQC was used for quality control and Subread (V.2.0.3, <https://pubmed.ncbi.nlm.nih.gov/30783653/>, <https://sourceforge.net/projects/subread/>) for read counting based on an Ensembl gene annotation file (GRCh38, release 107, <http://www.ensembl.org/index.html>). Upon normalisation (megSAP) and quality assessment (RnaQC), expression values of genes and exons were compared with an in-house cohort (same tissue and processing system) using NGSDAnnotateRNA.

Clinical interpretation was done with GSvar, enabling filtering for expression of genes and exons by gene, biotype, expression value, read counts and Z-score, compared with the cohort and the splice junctions by gene, type, read count and motif. Integrative Genomics Viewer (IGV, V.2.11.9, <https://www.nature.com/articles/nbt.1754>, <https://software.broadinstitute.org/software/igv/>) was used for visual inspection.

RESULTS

Cohort characteristics

A total of 1000 individuals were enrolled in this study, of whom 941 were affected by IRD and 59 by ION. Males and females were equally represented (50.4% females, 49.6% males). In X linked disease, the proportion of affected males was naturally higher, for example, 100% in retinoschisis, and 91.7% in choroideraemia.

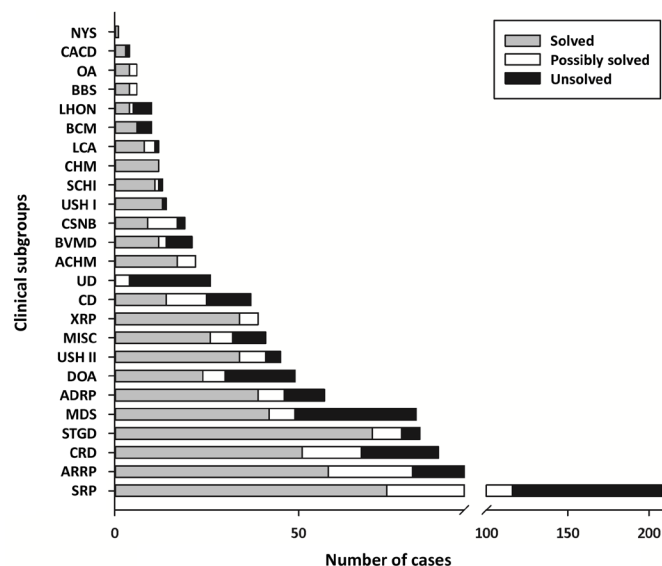


Figure 1 Distribution of solved cases, possibly solved cases and unsolved cases among the 25 clinical subgroups. ACHM, achromatopsia; ADRP, autosomal dominant retinitis pigmentosa; ARRP, autosomal recessive retinitis pigmentosa; BBS, Bardet-Biedl syndrome; BCM, blue cone monochromacy; BVMD, Best vitelliform macular dystrophy; CACD, central areolar choroidal dystrophy; CD, cone dystrophy; CHM, choroideraemia; CRD, cone-rod dystrophy; CSNB, congenital stationary night blindness; DOA, dominant optic atrophy; LCA, Leber congenital amaurosis; LHON, Leber hereditary optic neuropathy; MDS, macular dystrophy; MISC, miscellaneous diagnosis; NYS, nystagmus; OA, ocular albinism; SCHI, retinoschisis; SRP, simplex retinitis pigmentosa; STGD, Stargardt disease; UD, unclear diagnosis; USH I, Usher syndrome type I; USH II, Usher syndrome type 2; XRP, X linked retinitis pigmentosa.

The median age at time of testing was 39 years (range: 1–85 years), while 10.7% of participants were minors. In early-onset diseases, genetic testing was mostly performed in childhood and young adult age, for example, in achromatopsia (median age: 19 years; range: 3–55 years) and Leber congenital amaurosis (median age: 15.5 years; range: 2–63 years). In contrast, genetic testing was performed later in life in participants with late-onset diseases, for example, in macular dystrophy (mean age: 48 years; range: 9–80 years) and Best vitelliform macular dystrophy (mean age: 46 years; range: 6–75 years). An overview of the age distribution in all clinical subgroups is given in online supplemental figure 1.

The largest phenotypical group within the IRD cohort ($n=941$) was retinitis pigmentosa (42.4%, $n=399$), followed by cone-rod dystrophy (9.4%, $n=88$), Stargardt disease (8.8%, $n=83$) and macular dystrophy (8.7%, $n=82$). The smaller ION cohort included 59 cases, 83% of whom were diagnosed with dominant optic atrophy ($n=49$). **Figure 1** shows the number of cases in all clinical subgroups.

Diagnostic yield

In 57.4% of the participants ($n=574$), a definite genetic diagnosis could be made. Another 16.7% of participants ($n=167$) were shown to carry variants of unknown significance or variants with unconfirmed biallelism in genes consistent with their phenotype, but which require further functional validation. The remaining 25.9% ($n=259$) of the probands received a negative report. Thirty-one per cent of unsolved cases (8.0% of the entire cohort) carried one or more single heterozygous likely

pathogenic or pathogenic variants in a recessive gene, with *ABCA4* ($n=19$), *USH2A* ($n=5$) and *EYS* ($n=5$) being the most frequently affected.

The overall diagnostic yield (defined as solved and possibly solved cases) was 74.1% for the entire cohort. When considering only IRD cases ($n=941$), the overall diagnostic yield was 75.1% compared with 59.3% for the smaller ION cohort ($n=59$). Thirty per cent of participants reported other affected family members. In participants with a positive family history, the overall diagnostic yield increased to 85%.

The rate of solved cases varied widely among disease entities and was highest in diseases with little or no genetic heterogeneity, such as choroideraemia (100%), X linked retinitis pigmentosa (89.4%) and retinoschisis (84.6%) (**figure 1**). The rate of solved cases was considerably lower in macular dystrophy (51.2%) and simplex retinitis pigmentosa (36.1%). Overall, disease-causing and possibly disease-causing variants were identified in 190 genes (**table 1**). The most frequently implicated gene was *ABCA4* (16.3%), followed by *USH2A* (6.3%) and *RPGR* (4.7%). These three genes accounted for 30.7% of all solved cases. Accordingly, *ABCA4* accounted for 5 of the 10 most frequent alleles in the cohort and *USH2A* for 3 (online supplemental table 2).

Genes and variant types

In total, 1097 different variants were identified. A total of 53.2% of the variants ($n=584$) have already been described (ie, have an entry in HGMD), while 40.4% of the variants ($n=443$) were novel (ie, have no entry in HGMD; **figure 2A**). In addition, 6.5% ($n=71$) were SVs for which, with one exception, no matching HGMD entries are listed, as most HGMD entries do not specify exact breakpoints.

Among the 1097 unique variants, 1026 represent single nucleotide variants or small insertions and deletions, including missense variants ($n=548$; 49.9%), nonsense variants ($n=139$; 12.6%), frameshift variants ($n=173$; 15.7%), in-frame insertion or deletion variants ($n=27$; 2.4%), canonical splice site variants ($n=70$; 6.3%) and non-canonical splice site variants ($n=44$; 4.0%). Less commonly detected variants included intronic variants acting on splicing ($n=13$), variants in regulatory regions ($n=6$), small nuclear RNA variants ($n=2$), start loss ($n=1$), stop loss ($n=1$) and tRNA variants ($n=2$). In addition, 71 unique SVs were identified in 77 individuals (including several members of families with the same SV and individuals with two biallelic SVs), representing 6.5% of variants. **Figure 2** shows total numbers of all variant types.

Table 2 provides an overview of SVs. Their size ranged from 118 bp to 2.4 Mb. Multiple exon events were the most frequent ($n=35$), followed by single exon events ($n=25$). Larger SVs involving one or more genes were observed in 15 alleles. Few variants involved intronic or upstream regions and one case was shown to carry an unbalanced translocation. In terms of single genes, most SVs were identified in *EYS* (10 variants) and *PRPF31* (9 variants). Most SVs could be characterised with nucleotide resolution of breakpoints. An example of compound heterozygous SVs in *EYS* is shown in **figure 3**.

Inheritance patterns

Among cases with a definite genetic diagnosis, pathogenic or likely pathogenic variants were found in autosomal genes (83.5%), X linked (15.3%) and in mitochondrial genes (1.2%). Most solved cases (56.3%) carried variants in genes in which all so far described variants act exclusively recessively compared with 13.7% where variants were found in genes with solely

Table 1 Distribution of genes with causal variants according to clinical subgroup

Clinical subgroup	Number of solved cases (total cases)	Genes and number of cases with pathogenic and likely pathogenic variants
ACHM	17 (22)	<i>CNGB3</i> (9), <i>CNGA3</i> (8)
ADRP	39 (57)	<i>PRPF31</i> (12), <i>RHO</i> (7), <i>RP1</i> (4), <i>C1QTNF5</i> (3), <i>PRPF8</i> (3), <i>KLHL7</i> (2), <i>PRPH2</i> (2), <i>BEST1</i> (1), <i>CRX</i> (1), <i>IMPDH1</i> (1), <i>NR2E3</i> (1), <i>SNRNP200</i> (1), <i>TOPORS</i> (1)
ARRP	58 (95)	<i>EYS</i> (7), <i>CRB1</i> (4), <i>PDE6A</i> (4), <i>USH2A</i> (4), <i>CEP290</i> (3), <i>FLVCR1</i> (3), <i>PRCD</i> (3), <i>RHO</i> (3), <i>CNGA1</i> (2), <i>MERTK</i> (2), <i>NR2E3</i> (2), <i>PDE6B</i> (2), <i>RP1</i> (2), <i>TULP1</i> (2), <i>ABCA4</i> (1), <i>BBS1</i> (1), <i>CACNA1F</i> (1), <i>CDHR1</i> (1), <i>CERKL</i> (1), <i>CNGB1</i> (1), <i>MAK</i> (1), <i>PCARE</i> (1), <i>PRPF31</i> (1), <i>RDH12</i> (1), <i>REEP6</i> (1), <i>RLBP1</i> (1), <i>RP1L1</i> (1), <i>RP2</i> (1), <i>SLC24A1</i> (1), <i>TMEM126B</i> (1), <i>TTC8</i> (1)
BBS	4 (6)	<i>BBS2</i> (1), <i>BBS7</i> (1), <i>BBS10</i> (1), <i>MKKS</i> (1)
BCM	8 (11)	<i>OPN1MW/OPN1LW</i> (8)
BVMD	12 (21)	<i>BEST1</i> (10), <i>IMPG2</i> (1), <i>PRPH2</i> (1)
CACD	3 (4)	<i>ABCA4</i> (3)
CD	14 (37)	<i>CRX</i> (3), <i>ABCA4</i> (2), <i>GUCA1A</i> (2), <i>KCNV2</i> (2), <i>RPGR</i> (2), <i>CLN3</i> (1), <i>CNGA3</i> (1), <i>GUCY2D</i> (1)
CHM	12 (12)	<i>CHM</i> (12)
CRD	52 (88)	<i>ABCA4</i> (22), <i>PRPH2</i> (5), <i>PCARE</i> (4), <i>CERKL</i> (2), <i>CRX</i> (2), <i>GUCY2D</i> (2), <i>POC1B</i> (2), <i>PROM1</i> (2), <i>RP2</i> (2), <i>RPGR</i> (2), <i>AIPL1</i> (1), <i>CEP78</i> (1), <i>CLN3</i> (1), <i>CRB1</i> (1), <i>IMPG2</i> (1), <i>RP1L1</i> (1), <i>SCLT1</i> (1)
CSNB	9 (19)	<i>CACNA1F</i> (4), <i>NYX</i> (2), <i>GPR179</i> (1), <i>SAG</i> (1), <i>TRPM1</i> (1)
LCA	8 (12)	<i>RPE65</i> (4), <i>CEP290</i> (2), <i>ABCA4</i> (1), <i>IQCB1</i> (1)
LHON	3 (10)	<i>MT-ND4</i> (2), <i>MT-ND1</i> (1)
MDS	42 (82)	<i>ABCA4</i> (11), <i>PRPH2</i> (6), <i>RP1L1</i> (5), <i>CDHR1</i> (4), <i>PRDM13</i> (3), <i>CRX</i> (2), <i>GUCA1A</i> (2), <i>ABCC6</i> (1), <i>BEST1</i> (1), <i>CLN3</i> (1), <i>CRB1</i> (1), <i>IMPG2</i> (1), <i>KLHL7</i> (1), <i>MT-TL1</i> (1), <i>POC1B</i> (1), <i>RPGR</i> (1)
MISC	26 (41)	<i>IKBKG</i> (3), <i>FZD4</i> (2), <i>MFSD8</i> (2), <i>MT-ATP6</i> (2), <i>ABCA4</i> (1), <i>ARID1B</i> (1), <i>CNGB3</i> (1), <i>COL18A1</i> (1), <i>CYP4V2</i> (1), <i>GPR143</i> (1), <i>GRK1</i> (1), <i>HADHA</i> (1), <i>KCNV2</i> (1), <i>MMP19</i> (1), <i>MT-TL1</i> (1), <i>NPHP1</i> (1), <i>NR2E3</i> (1), <i>PHYH</i> (1), <i>PRPH2</i> (1), <i>RNU4ATAC</i> (1), <i>TUBB4B</i> (1)
NYS	1 (1)	<i>FRMD7</i> (1)
OA	4 (6)	<i>TYR</i> (4)
DOA	24 (49)	<i>OPA1</i> (18), <i>WFS1</i> (2), <i>ACO2</i> (1), <i>DNAJC30</i> (1), <i>NDUFA12</i> (1), <i>SSBP1</i> (1)
SCHI	11 (13)	<i>RS1</i> (11)
SRP	75 (208)	<i>EYS</i> (9), <i>USH2A</i> (9), <i>ABCA4</i> (5), <i>PRPF31</i> (6), <i>RP1</i> (5), <i>CRB1</i> (4), <i>FAM161A</i> (4), <i>MERTK</i> (3), <i>PRPF8</i> (3), <i>HK1</i> (2), <i>IQCB1</i> (2), <i>NR2E3</i> (2), <i>RPE65</i> (2), <i>TULP1</i> (2), <i>BEST1</i> (1), <i>C1QTNF5</i> (1), <i>CDHR1</i> (1), <i>CEP290</i> (1), <i>CFAP410</i> (1), <i>CHM</i> (1), <i>CWC27</i> (1), <i>MAK</i> (1), <i>NPHP1</i> (1), <i>PDE6A</i> (1), <i>PDE6B</i> (1), <i>PRPF4</i> (1), <i>PRPH2</i> (1), <i>RHO</i> (1), <i>RPGR</i> (1), <i>RPGRIP1</i> (1), <i>SPATA7</i> (1)
STGD	70 (83)	<i>ABCA4</i> (64), <i>CACNA1F</i> (1), <i>IFT140</i> (1), <i>IMPG1</i> (1), <i>PROM1</i> (1), <i>PRPH2</i> (1), <i>RP1L1</i> (1), <i>SPG11</i> (1)
UD	0 (26)	None
USH I	13 (14)	<i>MYO7A</i> (5), <i>PCDH15</i> (4), <i>CDH23</i> (1), <i>CEP250</i> (1), <i>CLRN1</i> (1), <i>USH1C</i> (1)
USH II	34 (54)	<i>USH2A</i> (21), <i>MYO7A</i> (4), <i>ADGRV1</i> (3), <i>CEP78</i> (2), <i>ARSG</i> (1), <i>CDH23</i> (1), <i>CLRN1</i> (1), <i>RS1</i> (1), <i>USH1G</i> (1)
XRP	35 (39)	<i>RPGR</i> (26), <i>RP2</i> (9)

Note that family codes were assigned prior to genetic testing and not adjusted after testing. Also note that six cases had dual molecular diagnoses resulting in a discrepancy of numbers for cases and pathogenic variants in the clinical subgroups ARRP, STGD and USH II.

ACHM, achromatopsia; ADRP, autosomal dominant retinitis pigmentosa; ARRP, autosomal recessive retinitis pigmentosa; BBS, Bardet-Biedl syndrome; BCM, blue cone monochromacy; BVMD, Best vitelliform macular dystrophy; CACD, central areolar choroidal dystrophy; CD, cone dystrophy; CHM, choroideraemia; CRD, cone-rod dystrophy; CSNB, congenital stationary night blindness; DOA, dominant optic atrophy; LCA, Leber congenital amaurosis; LHON, Leber hereditary optic neuropathy; MDS, macular dystrophy; MISC, miscellaneous diagnosis; NYS, nystagmus; OA, ocular albinism; SCHI, retinoschisis; SRP, simplex retinitis pigmentosa; STGD, Stargardt disease; UD, unclear diagnosis; USH I, Usher syndrome type I; USH II, Usher syndrome type 2; XRP, X linked retinitis pigmentosa.

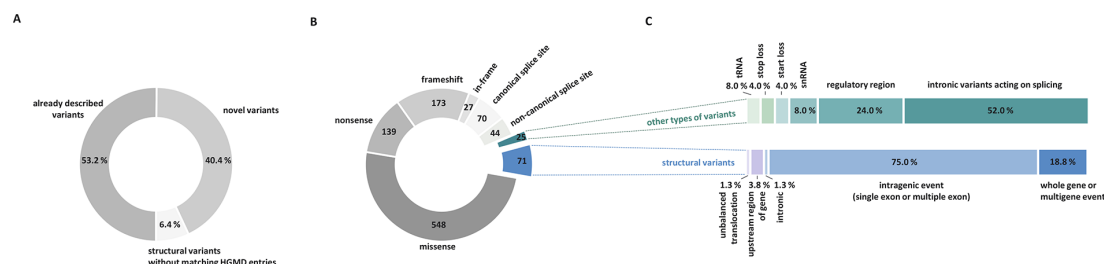


Figure 2 Characteristics of identified disease-causal genomic variation. (A) Variants with HGMD entry are described as already described variants. Novel variants are those with no HGMD entry so far. The fraction of structural variants is shown for which no matching HGMD entries are listed, as most HGMD entries do not specify exact breakpoints. (B) Distribution of variant types among the 1097 unique variants identified in the cohort. (C) Subcategories of variants from subgroups 'other types of variants' (n=25) and 'structural variants' (n=71) from (B). snRNA, small nuclear RNA.

Table 2 Types of structural variants

Structural variant	Number of alleles
Whole-gene or multigene event	15
Multiple genes	4 (1 deletion, 2 duplications, 1 inversion)
Whole gene	11 (11 deletions)
Intragenic event	60
Two or more exons	35 (30 deletions, 1 indel, 4 duplications)
One exon	25 (21 deletions, 1 indel, 1 duplication, 1 inversion, 1 inverted duplication)
Other	5
Intronic	1 (insertion)
Upstream region of gene	3 (2 deletions, 1 triplication)
Unbalanced translocation	1
Total	80
indel, insertion and deletion.	

dominant disease-causing alleles described so far. In several genes (in particular *RP1*, *BEST1* and *NR2E3*), variants can act either dominantly or recessively. Variants in these genes were identified in 13.4% of solved cases. Although 6.4% of cases (n=64) with a definite genetic diagnosis had pathogenic or likely pathogenic variants in more than one gene, the majority of these cases (n=57) had a single molecular diagnosis in addition to being a carrier of a pathogenic or likely pathogenic variant in a gene associated with an autosomal recessive disorder. Note that six cases (namely ARRP 441, ARRP 453, USH II 348, STGD 465 and two affected individuals from family CRD 844) had dual molecular diagnoses. Considering the large overlap of clinical subtypes in IRD, it is difficult to determine the contribution of each gene to the phenotype.

Prescreened cases

We obtained data on the presence or absence of previous genetic testing results in all participants. The majority of the cohort (n=794) had not received any previous genetic testing (cohort A), while in 69 individuals, Sanger sequencing of single or few recurrent genes had been negative (cohort B). Twenty individuals had received a prescreening with a mid-size research gene panel (108 genes)³⁰ with reduced coverage (cohort C), whereas

117 individuals had received genetic testing with a comprehensive diagnostic gene panel (cohort D).²⁸ Diagnostic yield was improved in all prescreened cohorts, with 47 additional definite genetic diagnoses made in cohort B (68.1%), 5 in cohort C (25.0%) and 42 in cohort D (35.9%).

RNA-seq in selected cases

We investigated a subset of 74 patients by RNA-seq to evaluate its added diagnostic value in IRD (71) or ION (3) as a genome-wide tool for improved variant interpretation and detection of transcript-deleterious alterations not expected from genome data analysis alone. An average of 11 335.75±2115.13 Mb of sequences was generated per sample with 52.78%±2.26% of the reads mapping to the coding region. The coverage of a set of 2176 housekeeping genes (see online supplemental table 3) was documented as an additional quality control parameter. In 58% of the samples, more than 65% of the housekeeping genes were covered at >20× or more than 80% at >10×. Evaluation of the expression levels of the 28 IRD/ION genes most frequently affected in our cohort indicated that only 5 of them (*OPA1*, *RP2*, *BEST1*, *PRPF8* and *PRFP31*) are stably expressed in blood (defined as transcript per million in >95% of individuals; online supplemental figure 2).

The cohort of 74 individuals included 21 solved cases, 17 unclear cases and 36 unsolved cases. In a first step, we investigated suspected disease-related variants with a predicted effect on splicing. Of the total of 31 prioritised changes, 77.4% (24 of 31) could not be evaluated at the transcript level due to generally low expression of the affected genes, 16.1% (5 of 31) remained non-conclusive as transcript alterations were either not statistically significant in quantitative analyses or the specific regions had insufficient read coverage.

In 6.5% (2 of 31) of cases, the transcriptome analysis provided additional evidence of the functional consequences of identified splice site changes in the genes *SCLT1* (individual CRD 671) and *SDCCAG8* (individual ARRP 285). Proband CRD 671 carried a heterozygous canonical splice site change predicted to alter the splice acceptor site of intron 15 in trans with a 1.4 kb deletion of the last coding exon, exon 21 (*SCLT1* (NM_144643.2): c.[1294-2A>G];[2005-1052_*363del]). While overall *SCLT1* expression levels were comparable with

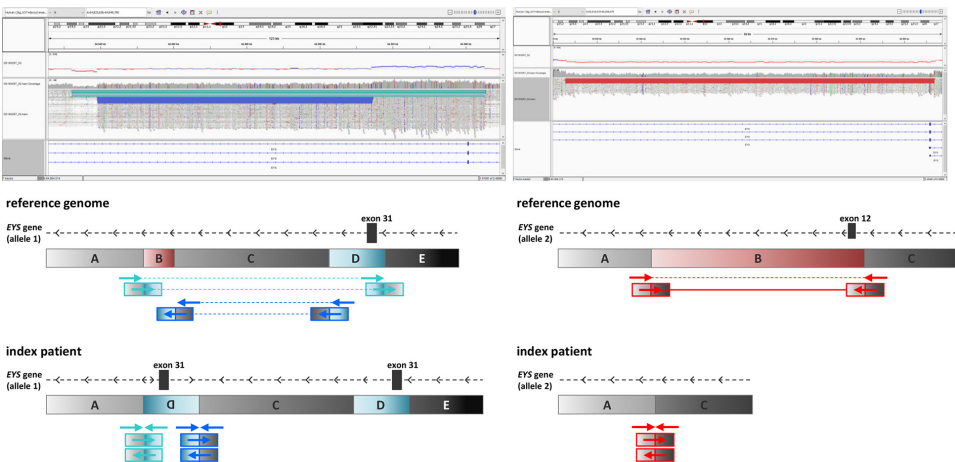


Figure 3 Illustrative example for biallelic structural variants. Genome sequencing with comprehensive bioinformatic analysis revealed a structural variant with breakpoints in intronic regions and only partial copy number alteration together with a deletion in trans in the *EYS* gene. The upper part shows the corresponding IGV screenshots, the lower part a schematic representation summarising the structural changes. The HGVS nomenclature is NC_000006.11:g.[65922918_66006755delinsGTTTTCTTTTA]; [64832337_64839052delins64914341_64945399inv]. Biallelism of the structural variants was confirmed by carrier testing using shallow genome sequencing. HGVS, Human Genome Variation Society; IGV, Integrative Genomics Viewer.

controls, there was evidence of multiple aberrant splicing events (online supplemental figure 3), consistent with a functional relevance of the identified *SCLT1* variants. Proband ARRP 285 was found to carry a heterozygous predicted frameshift variant in trans with a deep intronic alteration in *SDCCAG8* (NM_006642.3): [c.1946_1949del];[740+267C>T]]. Transcripts with the frameshift were not observed in the RNA-seq reads, consistent with reduced expression due to nonsense-mediated decay of the mutant transcript. The intronic variant c.740+267C>T is predicted to affect a splice enhancer site, possibly resulting in aberrant splicing comparable with the nearby c.740+356C>T variant.⁴⁷ In line with this hypothesis, we did not observe any reads supporting the usage of the canonical exon 7–8 splice junction in the patient's RNA-seq but the presence of reads with predominant inclusion of an alternative coding exon as well as other non-canonical transcripts (see online supplemental figure 4). While these findings are supportive of a functional relevance of the c.740+267C>T, we are aware of the technical limitations associated with overall rather low *SDCCAG8* expression levels and read length of 2×100bp, hindering quantitative assignments of the distribution of specific aberrant transcripts.

The transcript alterations in *SDCCAG8* and *SCLT1* were subsequently also identified in an unbiased systematic transcriptome-wide prioritisation for aberrant expression, mono-allelic expression or aberrant splicing. Interestingly, the aberrant expression of *SDCCAG8* was not significant at the gene level but was evident at the single-exon level, which also allows detection of altered expression of individual isoforms. However, no additional disease-related changes were detected in the untargeted transcriptome-wide analysis.

DISCUSSION

Although many IRD cohort studies have now been published, those involving GS are still scarce. In 2017, Carss and colleagues published a cohort study of 722 individuals with IRD.⁵ While 605 of the individuals were analysed using GS, 72 had ES, and an additional 45 had both. The composition of their cohort in terms of clinical subgroups and the proportion of singletons, duos and trios is very similar to ours. The main difference is that in their cohort, almost all participants had received some sort of prescreening, whereas in ours, this fraction was only 20.6%. Carss and colleagues achieved a pathogenic variant detection rate of 56%, which almost exactly matches our rate of 57.4%.⁵ In their study, 31% of cases that remained unsolved after ES could be solved by GS. Another IRD cohort study with 562 individuals was performed by Ellingford and colleagues in 2016, but only 46 cases underwent GS.¹³ They established a molecular diagnosis in 50% of cases which is consistent with the present study and that of Carss and colleagues. Ellingford and colleagues hypothesised that diagnostic yield could have been increased by 29% if their GS pipeline had been applied to all 562 patients in their cohort, rather than just 46.¹³ However, they concede that much of this hypothetical uplift could also be achieved by improved variant calling in targeted NGS pipelines and the inclusion of non-coding regions that are known to harbour pathogenic variants. Recently, Biswas and colleagues have analysed 409 individuals from 108 unrelated pedigrees with IRD using GS and achieved a molecular diagnosis rate of 57%.¹² In summary, our detection rate is very close or equal to that of other studies. In the following, we will discuss several aspects of our study in more detail.

Prescreening

In our cohort, 206 individuals had received prior genetic testing, either based on Sanger sequencing of single or few recurrent genes (cohort B, n=69), research-based testing using molecular inversion probes targeting 108 IRD genes (cohort C, n=20) or diagnostic-based testing using a comprehensive gene panel for inherited eye disorders (cohort D, n=117). Individuals who had not had any genetic test prior to this study (cohort A) had a notably higher rate of likely molecular diagnoses (76.2%) than those of cohorts B, C and D (66.0%). Nevertheless, GS improved diagnostic yield in all prescreened cohorts: 47 definite diagnoses were made in cohort B, 5 in cohort C and 42 in cohort D. This was achieved primarily through the detection of variants that reside in genes that have not been included in the first-tier testing. However, GS also improved diagnostic yield by the delineation of balanced⁴⁸ or complex SVs that were undetectable by prior testing strategies (see examples table 2 and figure 3).

Structural variants

We previously examined a cohort of 2158 individuals diagnosed with IRD using a targeted gene panel and found that the genotypes of 91 individuals (4.2%) comprised CNVs.²⁸ These variants were mainly recurrent deletions of single or multiple exons in four different genes (*PRPF31*, *USH2A*, *EYS* and *CHM*). In the present study, clinically relevant SVs were detected in 77 out of 1000 individuals (7.7%). These variants were found in 37 different genes and comprised not only CNVs but also copy-neutral rearrangements (see table 2). The nearly doubled detection rate of SVs in the present study can certainly be explained by the improved detection of SVs by GS, especially copy number neutral ones. In addition, GS allows breakpoint resolution to the nucleotide level in many cases. This facilitates segregation in family members where breakpoint PCR can be easily performed instead of a more laborious qPCR.

CNVs were identified in three genes not previously reported to have large deletions or duplications (*CNGA1*, *CWC27*, *SCLT1*), expanding their mutational spectrum.

Trios

Trio analysis (ie, including the parents of an affected child in the analysis) aids in the filtering of familial benign variants and allows the direct assessment of the inheritance pattern of candidate variants, as well as their phase. It also reliably identifies de novo variants. In this study, we have analysed 18 IRD trios. Of these, 13 are considered solved with pathogenic or likely pathogenic variants, while 1 case was shown to be compound heterozygous for a likely pathogenic variant and a variant of uncertain significance. Four trios remained unsolved. Hence, the diagnostic yield among the trios was 72.2% which basically equals the overall diagnostic yield of the entire cohort (74.1%). Therefore, we cannot conclude that trio analysis has a major benefit in IRD diagnostics, although the number of trios is probably too small to draw a definite conclusion.

De novo variants

Since parental DNA was available only in a subset of cases, we could not assess the actual frequency of de novo variation in our cohort. A de novo status of variants could be confirmed in three cases. While de novo variants have previously been found in 3.2–14.3% of cases diagnosed with retinitis pigmentosa,^{20 49} the proportion is generally unknown in IRD because most cohort studies have not assessed the de novo status. In any case, most variants of IRD can be expected to be inherited, as reproductive

fitness is unlikely to be affected—in contrast to developmental disorders where the proportion of de novo variants is high.⁶

Cases solved by additional measures

The inclusion of difficult-to-sequence genomic regions is essential in IRD, since two clinical subgroups are solely caused by X-chromosomal genes that are characterised by highly repetitive sequences. X linked retinitis pigmentosa is most commonly caused by mutations in *RPGR* and, to a lesser extent, *RP2*. The *RPGR* gene includes a C-terminal exon, termed ORF15, which is characterised by a highly repetitive glutamic acid/glycine-rich domain. While ORF15 is difficult to sequence, it is also a known mutation hot-spot that comprises 60% of disease-causing variants in *RPGR*.^{50–51} Based on clinical examination and self-reported family history, 39 cases in our cohort were diagnosed with X linked retinitis pigmentosa. Among these, 35 cases could be solved with pathogenic or likely pathogenic variants in *RP2* (n=9) or *RPGR* (n=26), leading to a diagnostic rate of 89.7%. This rate is only slightly lower than that of our former study (93.7%) based on targeted sequencing.²⁸ To ensure a high diagnostic sensitivity for variants in ORF15, all unsolved IRD cases were resequenced for this region. This was done for all these cases because small core families, which are prevalent in Germany, make it challenging to determine the mode of inheritance accurately. Resequencing led to a molecular diagnosis in only one case, suggesting that GS is capable of detecting most variants in the notoriously difficult-to-sequence region of ORF15.

Another under-represented target region is the red/green cone opsin (*OPN1LW/OPN1MW*) gene cluster associated with cone dysfunction disorders, such as blue cone monochromacy. Short-read sequencing is not sufficient to reveal the complexity of the opsin gene cluster and distinguish between copies since the variable number of *OPN1LW/OPN1MW* genes within the gene cluster shares an identity of 98% both for coding and non-coding nucleotide sequence.⁵² Using IGV, the red/green opsin gene cluster (*OPN1LW/OPN1MW*) was manually inspected in cases diagnosed with blue cone monochromacy and eventually analysed using a customised genotyping strategy in a research set-up.⁵² With this approach, 8 out of 11 cases with blue cone monochromacy could be solved.

Thirteen different intronic variants were identified in this study, seven of which are already known to cause a splicing defect, in particular the c.4253+43G>A variant in *ABCA4* and the c.2991+1655A>G variant in *CEP290*.^{53–54} The interpretation of non-coding variants is challenging and needs complementary methods to decipher their functional impact. Of the six novel deep intronic variants, we have functionally characterised one, namely the c.1033–327T>A variant in *POC1B*.⁵⁵ In addition, we have analysed a silent exonic variant (c.750A>G) in *MFSD8*.⁵⁶ Using in vitro splice assays and direct cDNA analysis, we could demonstrate a pathogenic effect for the c.1033–327T>A variant in *POC1B* identified in patient MDS 438 and the c.750A>G variant in *MFSD8* identified in family MISC 272.^{55–56}

Unsolved cases with monoallelic variants and diagnostic value of RNA-seq

Eighty cases were found to harbour pathogenic or likely pathogenic variants on one allele in genes associated with recessive disease. While it cannot be ruled out that these individuals are just incidental carriers, it is entirely possible that an elusive pathogenic variant resides on the second allele. Searching for these elusive variants and determining their potentially deleterious

effects, for example, through a systematic functional evaluation of rare deep-intronic variants, might prove crucial to improving the diagnostic rate. As for the latter, the results of our pilot study on 74 cases suggest that the added diagnostic value of RNA-seq from blood is rather limited in IRD. At first glance, this observation contrasts with RNA-seq studies suggesting improved diagnostic variant interpretation in up to 13–16% of cases, but this is likely to depend largely on the clinical diagnosis and expression of associated genes in accessible tissue.^{57–58} Apart from the disease phenotypes and tissues studied, the extent of concomitant DNA sequencing studies and bioinformatic analyses may contribute to the differences in the percentage of additional cases solved. For example, most studies were performed on cases that remained unresolved after exome sequencing, so RNA-seq could contribute significantly to the technical accessibility of non-coding regions and additional types of genomic variation. We expect that this benefit will decrease over the next years with improved detection and annotation of intronic genomic variation and diagnostic-grade genome analyses becoming widely available as a first-line diagnostic assay. However, we also agree on the importance of additional biosamples such as fibroblast cell lines as well as induced pluripotent stem cells and thereof derived organoids or specialised cells as a resource for second-line analyses in a research context.

Age at molecular diagnosis

In this study, age at onset was not assessed. However, we used the age at genetic testing as a variable. The mean age at molecular diagnosis for solved and possibly solved individuals was 38.2 years compared with 45.4 years for unsolved cases. This difference in the age distribution is statistically significant (Mann-Whitney U test, $p < 0.001$). Remarkably, these figures are almost identical to those from our previous study, where the average age of participants with a molecular diagnosis was 39.8 years, while it was 46.3 years for unsolved cases.²⁸ The differences in age distribution between solved and unsolved cases are reflected in the differences in age distribution within clinical subgroups. The median age at genetic testing in early-onset diseases such as achromatopsia or Leber congenital amaurosis was significantly lower than that for late-onset diseases such as macular dystrophy or central areolar choroidal dystrophy (online supplemental figure 1). For the latter, the rate of solved cases was significantly lower than for the former. A skewing of diagnostic yield towards an earlier age of onset/testing in IRD cohorts has been reported previously²⁵ and has been attributed to a potential multigenic or multifactorial aetiology in participants with an age of onset >50 years.⁵⁹

Benefits of GS

Although it is difficult to estimate numbers, a considerable proportion of undiagnosed cases in IRD may be attributed to variant types which can only be detected using GS. One of the major benefits of GS is the more uniform coverage compared with other technologies. This can be seen, for example, in gnomAD which contains both ES and GS datasets: while 89.4% of the exome was covered with ≥ 20 read depth, this value was exceeded by GS with a 97.1% coverage.⁶⁰ Since the ES datasets were generated before the GS datasets, improvements in the general sequencing process may account for some of the differences, but the more uniform coverage of GS has been observed repeatedly.^{3–4} Moreover, GS requires lower average coverage to obtain the same sensitivity and accuracy in variant calling compared with ES.^{61–62} Along this line, we aimed to estimate

the added diagnostic value of genome versus exome sequencing using an in silico approach focused on sequence coverage. From at least 10 representative diagnostic grade exome datasets generated with Agilent Human SureSelect Human All Exon Kits V6 and V7 as well as a Twist-based custom enrichment kit, the coverage of the genomic positions of the reported variants was calculated. Depending on testing system investigated, the diagnostic coverage cut-off applied (below 20× or 6× coverage) and the types of variation included (small variants with/without structural variants), the rate of exome failures ranged from 1.7% (<6×, small variants only, Twist) to 8.5% (<20×, including structural variants, SureSelect Human All Exon V6). Depending on the quality of an exome or panel experiment and extent of the initial diagnostic work-up, we assume that an estimated 5–10% of additional diagnoses is achievable using a bioinformatic pipeline optimised for data analysis beyond the exome and a constantly growing list of clinically annotated DNA variants in the non-coding regions. For inherited eye diseases, this estimate is also supported empirically by a diagnostic yield of 70.8% documented in a cohort recruited in the same clinical setting but investigated by targeted approaches,²⁸ compared with 76.8% in the current study, corresponding to 7.8% of additional diagnoses.

In conclusion, GS can provide specific genetic diagnoses for a considerable proportion of patients with IRD and ION. A molecular diagnosis provides individuals with improved genetic counselling, redirection of clinical management, monitoring for potential systemic manifestations, and preparedness for existing and future precision therapies. GS avoids serial testing of unsolved cases and is the only platform interrogating non-coding regions. Whereas ES currently is the favoured first-tier diagnostic tool, an increasing number of studies have used GS as a first-tier test.^{9,63–65} Given the continuous decrease of sequencing costs and increase in sequencing capacity, the number of GS analyses in a diagnostic is expected to increase further. The full extent of the benefits of GS will probably become more apparent when bioinformatics predictions have improved and allow for a better selection of candidate intronic variants. However, functional work-up of these intronic variants through functional follow-up studies will likely remain a challenge. While the targeted approaches have been shown to be effective for specific variants in selected genes,^{55,66} they are rather difficult to scale. Generic approaches with putatively broad clinical implementation, such as RNA-seq, currently lack adequate models, and readily accessible tissues such as blood are informative for only a small subset of disease genes underlying IRD or ION. This makes documentation of the results of such studies in publicly available databases all the more important as a community effort to improve the diagnosis of future patients.

Author affiliations

¹Institute for Ophthalmic Research, Centre for Ophthalmology, University of Tübingen, Tübingen, Germany

²Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany

³University Eye Hospital, Centre for Ophthalmology, University of Tübingen, Tübingen, Germany

⁴MVZ für Humangenetik und Molekularpathologie, Rostock, Germany

⁵Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Tübingen, Germany

⁶Center for Rare Disease, University of Tübingen, Tübingen, Germany

Twitter Nicolas Casadei @NLPCasadei

Acknowledgements All authors thank the patients and their families for participating in the study.

Contributors Conceptualisation—NW, SK, OR, KSt and TBH. Data curation—NW, PM, TZ and DW. Formal analysis—NC, OK, GD, LS, JA, MS and SO. Funding acquisition—OR. Investigation—NW, PM, TZ, KSc, LK, FK, DW, AL, RF, LP, MR, LGS, MB, AM, Ckel, Cke, PR, FS, UG, FT, TH, FN, BW, SK, KS and TBH. Supervision—SK, OR, KS and TBH. Visualisation—NW, PM, TZ, LS and TBH. Writing (original draft)—NW, PM, TZ and KSt. Author acting as guarantor—TBH. Writing (review and editing)—all authors.

Funding The study was supported by Illumina within the 'genome first study'. This work was in part supported by a grant of the Deutsche Forschungsgemeinschaft (Wi1189/12-1 to BW; 418081722 and 433158657 to TBH) and by the German Ministry of Health via genomDE to OR.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study involves human participants and was approved by the Ethics Committee of the University of Tübingen (367/2019BO1, 116/2015BO2, 637/2017BO1) and adhered to the principles of the Declaration of Helsinki. Written informed consent for genetic testing was obtained for all individuals. For probands who were minors at the time of blood collection, informed consent was obtained from the parents or legal guardians.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Sequence datasets have not been deposited in a public repository due to local consent regulations. Depersonalised data and additional experimental data that this study is based on can be provided upon request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Nicolas Casadei <http://orcid.org/0000-0003-2209-0580>

Tobias B Haack <http://orcid.org/0000-0001-6033-4836>

REFERENCES

- Stranneheim H, Wedell A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *J Intern Med* 2016;279:3–15.
- Stenson PD, Ball EV, Howells K, et al. The human gene mutation database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics* 2009;4:69–72.
- Lelieveld SH, Spielmann M, Mundlos S, et al. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum Mutat* 2015;36:815–22.
- Bekadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* 2015;112:5473–8.
- Carss KJ, Arno G, Erwood M, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am J Hum Genet* 2017;100:75–90.
- Gillissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014;511:344–7.
- Lindstrand A, Eisefeldt J, Pettersson M, et al. From cytogenetics to cytogenomics: whole-exome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability. *Genome Med* 2019;11:68.
- Costain G, Jobling R, Walker S, et al. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur J Hum Genet* 2018;26:740–4.
- Lionel AC, Costain G, Monfared N, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med* 2018;20:435–43.
- Kim A, Kumar KR, Davis RL, et al. Increased diagnostic yield of spastic paraplegia with or without cerebellar ataxia through whole-genome sequencing. *Cerebellum* 2019;18:781–90.

- 11 Palmer EE, Sachdev R, Macintosh R, *et al.* Diagnostic yield of whole genome sequencing after nondiagnostic exome sequencing or gene panel in developmental and epileptic encephalopathies. *Neurology* 2021;96:e1770–82.
- 12 Biswas P, Villanueva AL, Soto-Hermida A, *et al.* Deciphering the genetic architecture and ethnographic distribution of IRD in three ethnic populations by whole genome sequence analysis. *PLoS Genet* 2021;17:e1009848.
- 13 Ellingford JM, Barton S, Bhaskar S, *et al.* Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. *Ophthalmology* 2016;123:1143–50.
- 14 Congdon N, O'Colmain B, Klaver CCW, *et al.* Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol* 2004;122:477–85.
- 15 Holtan JP, Selmer KK, Heimdal KR, *et al.* Inherited retinal disease in Norway - a characterization of current clinical and genetic knowledge. *Acta Ophthalmol* 2020;98:286–95.
- 16 Liew G, Michaelides M, Bunce C. 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* 2014;4:e004015.
- 17 Puech B, Kostrubiec B, Hache JC, *et al.* Epidemiology and prevalence of hereditary retinal dystrophies in the northern France. *J Fr Ophthalmol* 1991;14:153–64.
- 18 Chong JX, Buckingham KJ, Jhangiani SN, *et al.* The genetic basis of mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet* 2015;97:199–215.
- 19 Simunovic MP, Jolly JK, Xue K, *et al.* The spectrum of CHM gene mutations in choroideremia and their relationship to clinical phenotype. *Invest Ophthalmol Vis Sci* 2016;57:6033–9.
- 20 Gao F-J, Li J-K, Chen H, *et al.* Genetic and clinical findings in a large cohort of Chinese patients with suspected retinitis pigmentosa. *Ophthalmology* 2019;126:1549–56.
- 21 Martin-Merida I, Avila-Fernandez A, Del Pozo-Valero M, *et al.* Genomic landscape of sporadic retinitis pigmentosa: findings from 877 Spanish cases. *Ophthalmology* 2019;126:1181–8.
- 22 Yu-Wai-Man P, Griffiths PG, Chinnery PF. Mitochondrial optic neuropathies - disease mechanisms and therapeutic strategies. *Prog Retin Eye Res* 2011;30:81–114.
- 23 Lenaers G, Neutznier A, Le Dantec Y, *et al.* Dominant optic atrophy: culprit mitochondria in the optic nerve. *Prog Retin Eye Res* 2021;83:100935.
- 24 Charif M, Bris C, Goudenège D, *et al.* Use of next-generation sequencing for the molecular diagnosis of 1,102 patients with a autosomal optic neuropathy. *Front Neurol* 2021;12:602979.
- 25 Stone EM, Andorf JL, Whitmore SS, *et al.* Clinically focused molecular investigation of 1000 consecutive families with inherited retinal disease. *Ophthalmology* 2017;124:1314–31.
- 26 Dockery A, Stephenson K, Keegan D, *et al.* Target 5000: target capture sequencing for inherited retinal Degenerations. *Genes (Basel)* 2017;8:304.
- 27 Sharon D, Ben-Yosef T, Goldenberg-Cohen N, *et al.* A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). *Hum Mutat* 2020;41:140–9.
- 28 Weisschuh N, Obermaier CD, Battke F, *et al.* Genetic architecture of inherited retinal degeneration in Germany: a large cohort study from a single diagnostic center over a 9-year period. *Hum Mutat* 2020;41:1514–27.
- 29 Weisschuh N, Schimpf-Linzenbold S, Mazzola P, *et al.* Mutation spectrum of the OPA1 gene in a large cohort of patients with suspected dominant optic atrophy: identification and classification of 48 novel variants. *PLoS One* 2021;16:e0253987.
- 30 Weisschuh N, Feldhaus B, Khan MI, *et al.* Molecular and clinical analysis of 27 German patients with Leber congenital amaurosis. *PLoS One* 2018;13:e0205380.
- 31 Sturm M, Schroeder C, Bauer P. Seqpurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics* 2016;17:208.
- 32 Li H, Durbin R. Fast and accurate short read alignment with burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- 33 Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 2014;30:2503–5.
- 34 Mose LE, Perou CM, Parker JS. Improved Indel detection in DNA and RNA via realignment with AbRA2. *Bioinformatics* 2019;35:2966–73.
- 35 Erik G. Haplotype-based variant detection from short-read sequencing. *arXiv* 2012:1207.3907.
- 36 McLaren W, Gil L, Hunt SE, *et al.* The Ensembl variant effect predictor. *Genome Biol* 2016;17:122.
- 37 Chen X, Schulz-Trieglaff O, Shaw R, *et al.* Manta: rapid detection of structural variants and Indels for Germline and cancer sequencing applications. *Bioinformatics* 2016;32:1220–2.
- 38 Demidov G, Ossowski S. ClinCNV: novel method for allele-specific somatic copy-number alterations detection. *Bioinformatics* [Preprint] 2019.
- 39 Landrum MJ, Lee JM, Benson M, *et al.* Clinvar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46:D1062–7.
- 40 Stenson PD, Mort M, Ball EV, *et al.* The human gene mutation database (HGMD((R))) : optimizing its use in a clinical diagnostic or research setting. *Hum Genet* 2020;139:1197–207.
- 41 Fokkema IFAC, Taschner PEM, Schaafsma GCP, *et al.* LOVD V2.0: the next generation in gene variant databases. *Hum Mutat* 2011;32:557–63.
- 42 The 1000 Genomes Project Consortium, Corresponding authors, Auton A, *et al.* A global reference for human genetic variation. *Nature* 2015;526:68–74.
- 43 Sherry ST, Ward MH, Kholodov M, *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
- 44 Karczewski KJ, Weisburd B, Thomas B, *et al.* The Exac browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res* 2017;45:D840–5.
- 45 Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
- 46 Richards S, Aziz N, Bale S, *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–24.
- 47 Otto EA, Hurd TW, Airik R, *et al.* Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat Genet* 2010;42:840–50.
- 48 Weisschuh N, Mazzola P, Heinrich T, *et al.* First Submicroscopic inversion of the OPA1 gene identified in dominant optic atrophy - a case report. *BMC Med Genet* 2020;21:236.
- 49 Neveling K, Collin RWJ, Gilissen C, *et al.* Next-generation genetic testing for retinitis pigmentosa. *Hum Mutat* 2012;33:963–72.
- 50 Branham K, Othman M, Brumm M, *et al.* Mutations in RPGR and RP2 account for 15% of males with simplex retinal degenerative disease. *Invest Ophthalmol Vis Sci* 2012;53:8232–7.
- 51 Vervoort R, Lennon A, Bird AC, *et al.* Mutational hot spot within a new RPGR Exon in X-linked retinitis pigmentosa. *Nat Genet* 2000;25:462–6.
- 52 Cideciyan AV, Hufnagel RB, Carroll J, *et al.* Human cone visual pigment deletions spare sufficient photoreceptors to warrant gene therapy. *Hum Gene Ther* 2013;24:993–1006.
- 53 den Hollander AL, Koenekoop RK, Yzer S, *et al.* Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet* 2006;79:556–61.
- 54 Sangermano R, Garanto A, Khan M, *et al.* Deep-Intronic ABCA4 variants explain missing heritability in stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* 2019;21:1751–60.
- 55 Weisschuh N, Mazzola P, Bertrand M, *et al.* Clinical characteristics of POC1B-associated retinopathy and assignment of pathogenicity to novel deep Intronic and non-canonical splice site variants. *Int J Mol Sci* 2021;22:5396.
- 56 Reith M, Zeltner L, Schäferhoff K, *et al.* A novel, apparently silent variant in MFSD8 causes neuronal ceroid lipofuscinosis with marked Intrafamilial variability. *Int J Mol Sci* 2022;23:2271.
- 57 Dekker J, Schot R, Bongaerts M, *et al.* Web-accessible application for identifying pathogenic transcripts with RNA-Seq: increased sensitivity in diagnosis of neurodevelopmental disorders. *Am J Hum Genet* 2023;110:251–72.
- 58 Yépez VA, Gusic M, Kopajtich R, *et al.* Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med* 2022;14:38.
- 59 Shanks ME, Downes SM, Copley RR, *et al.* Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease. *Eur J Hum Genet* 2013;21:274–80.
- 60 Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
- 61 Meynert AM, Ansari M, FitzPatrick DR, *et al.* Variant detection sensitivity and biases in whole genome and Exome sequencing. *BMC Bioinformatics* 2014;15:247.
- 62 Sun Y, Liu F, Fan C, *et al.* Characterizing sensitivity and coverage of clinical WGS as a diagnostic test for genetic disorders. *BMC Med Genomics* 2021;14:102.
- 63 Farnaes L, Hildreth A, Sweeney NM, *et al.* Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. *NPJ Genom Med* 2018;3:10.
- 64 Stavropoulos DJ, Merico D, Jobling R, *et al.* Whole genome sequencing expands diagnostic utility and improves clinical management in pediatric medicine. *NPJ Genom Med* 2016;1:15012:15012.
- 65 Turnbull C, Scott RH, Thomas E, *et al.* The 100 000 Genomes project: bringing whole genome sequencing to the NHS. *BMJ* 2018;361:k1687.
- 66 Reurink J, Weisschuh N, Garanto A, *et al.* Whole genome sequencing for Ush2A-associated disease reveals several pathogenic deep-intronic variants that are amenable to splice correction. *HGG Adv* 2023;4:100181.