Journal Pre-proof

Genetic basis of inherited retinal disease in a molecularly characterised cohort of over 3000 families from the United Kingdom

Nikolas Pontikos, PhD, Gavin Arno, PhD, Neringa Jurkute, MD, FEBO, Elena Schiff, PhD, Rola Ba-Abbad, PhD, FRCS(Glasg), Samantha Malka, Ainoa Gimenez, Michalis Georgiou, MD, Genevieve Wright, MSc, Monica Armengol, Hannah Knight, Menachem Katz, Mariya Moosajee, PhD, FRCOphth, Patrick Yu-Wai-Man, Anthony T. Moore, MD(Res), FRCOphth, Michel Michaelides, MD(Res), FRCOphth, Andrew R. Webster, MD(Res), FRCOphth, Omar A. Mahroo, PhD, FRCOphth

PII: S0161-6420(20)30332-8
DOI: https://doi.org/10.1016/j.ophtha.2020.04.008
Reference: OPHTHA 11200

To appear in: Ophthalmology

Received Date: 31 December 2019
Revised Date: 27 March 2020
Accepted Date: 6 April 2020


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Inc. on behalf of the American Academy of Ophthalmology
Genetic basis of inherited retinal disease in a molecularly
ccharcterised cohort of over 3000 families from the United
Kingdom

Nikolas Pontikos PhD,1,2 Gavin Arno PhD,1,3 Neringa Jurkute MD, FEBO,1,2 Elena Schiff PhD,1,2 Rola Ba-Abbad PhD, FRCS(Glasg),1,2 Samantha Malka,1,2 Ainoa Gimenez,1,2 Michalis Georgiou MD,1,2 Genevieve Wright MSc,2 Monica Armengol,2 Hannah Knight,2 Menachem Katz,2 Mariya Moosajee PhD, FRCOphth,1,3 Patrick Yu-Wai-Man,1,2,4,5 Anthony T Moore MD(Res), FRCOphth,1,2,6 Michel Michaelides MD(Res), FRCOphth,1,2 Andrew R Webster MD(Res), FRCOphth,1,2 Omar A. Mahroo PhD, FRCOphth1,2,7,8

1. UCL Institute of Ophthalmology, University College London, 11-43 Bath Street, London UK
2. Genetics Service, Moorfields Eye Hospital, 162 City Road, London UK
3. North East Thames Regional Genetics Service, Great Ormond Street Institute of Child Health, London, United Kingdom
4. Cambridge Centre for Brain Repair and MRC Mitochondrial Biology Unit, Department of Clinical Neurosciences, University of Cambridge, Cambridge UK.
5. Cambridge Eye Unit, Addenbrooke’s Hospital, Cambridge University Hospitals, Cambridge UK.
6. Department of Ophthalmology, University of California San Francisco, California USA
7. Section of Ophthalmology, King’s College London, St Thomas’ Hospital Campus, London UK
8. Physiology, Development and Neuroscience, University of Cambridge, Cambridge UK

Corresponding author: Omar Mahroo at address 1 above. Email: o.mahroo@ucl.ac.uk

Running title: Genetic diagnosis in a large UK inherited retinal disease cohort

Funding Sources: Wellcome Trust (including grants 206619/Z/17/Z and 205174/Z/16/Z), Fight for Sight UK, NIHR Biomedical Research Centre at Moorfields Eye Hospital and the UCL Institute of
Ophthalmology, Moorfields Eye Charity. GA is supported by a Fight For Sight (UK) Early Career Investigator Award and the Biomedical Research Centre at Great Ormond Street Hospital Institute of Child Health. NP is funded by a Moorfields Eye Charity Career Development Award (R190031A) and a donation from the Archer family. PYWM is supported by a Clinician Scientist Fellowship Award (G1002570) from the Medical Research Council (UK) and receives funding from Fight for Sight (UK), the Isaac Newton Trust, the UK National Institute of Health Research (NIHR) as part of the Rare Diseases Translational Research Collaboration. The funding organisations had no role in the design or conduct of this research. The views expressed are those of the authors and not the funding organisations.

Conflict of interest: No conflicting relationship exists for any author.
Abstract

**Purpose:** In our cohort of >3000 molecularly characterised inherited retinal disease (IRD) families, we investigated proportions with disease attributable to causative variants in each gene.

**Design:** Retrospective study of electronic patient record

**Participants:** Patients and relatives managed in the Genetics Service of Moorfields Eye Hospital in whom a molecular diagnosis had been identified.

**Methods:** Genetic screening used a combination of single gene testing, gene panel testing, whole exome sequencing, and, more recently, whole genome sequencing. For this study, genes listed in the Retinal Information Network online resource (https://sph.uth.edu/retnet/) were included. Transcript length was extracted for each gene (Ensembl, release 94).

**Main Outcome Measures:** We calculated proportions of families with IRD attributable to variants in each gene in the whole cohort, a cohort <18 years, and a “current” cohort (at least one patient encounter between 1 Jan 2017 and 2 Aug 2019). Additionally, we explored correlation between numbers of families and gene transcript length.

**Results:** We identified 3195 families with a molecular diagnosis (variants in 135 genes), including 4236 affected individuals. The pediatric cohort comprised 452 individuals from 411 families (66 genes). The current cohort comprised 2614 families (131 genes; 3130 affected individuals). The pediatric cohort showed some differences, including a higher proportion of families with X-linked disease. The 20 most frequently implicated genes overall were as follows: *ABCA4* (20.8% of families); *USH2A* (9.1%); *RPGR* (5.1%); *PRPH2* (4.6%); *BEST1* (3.9%); *RS1* (3.5%); *RP1* (3.3%); *RHO* (3.3%); *CHM* (2.7%); *CRB1* (2.1%); *PRPF31* (1.8%); *MYO7A* (1.7%); *OPA1* (1.6%); *CNGB3* (1.4%); *RPE65* (1.2%); *EYS* (1.2%); *GUCY2D* (1.2%); *PROM1* (1.2%); *CNAG3* (1.1%); *RDH12* (1.1%). These accounted for 71.8% of all molecularly diagnosed families. Spearman coefficients for correlation between numbers of
families and transcript length were 0.20 ($p=0.025$) overall, and 0.27 ($p=0.017$), -0.17 ($p=0.46$) and 0.71 ($p=0.047$) for genes in which variants exclusively cause recessive, dominant or X-linked disease respectively.

Conclusions: Our findings help quantify the burden of inherited retinal disease attributable to each gene. Over 70% of families had pathogenic variants in one of 20 genes. Transcript length (relevant to gene delivery strategies) correlated significantly with numbers of affected families (but not for dominant disease).
Monogenic retinal diseases are a major cause of blindness in the pediatric and working age population in many countries.\textsuperscript{1-3} Pathogenic variants in over 250 genes can give rise to inherited retinal disease (IRD), with multiple modes of inheritance.\textsuperscript{4} For the majority of these diseases, there are no medical or surgical treatments, but a large number of therapeutic trials are underway.\textsuperscript{5} There is now a commercially available licensed gene-replacement treatment for a particular genetic cause, namely IRD due to bi-allelic variants in \textit{RPE65}.\textsuperscript{6} As more therapies are likely to become available in the future, with many likely to be specific to a particular genetic cause, it is of increasing relevance to understand the burden of disease attributable to variants in particular genes.

The Genetics Service of Moorfields Eye Hospital oversees the care of the largest number of IRD patients of any one site in the United Kingdom. A significant proportion of these families have a molecular diagnosis, more recently with the advent of parallel nucleotide sequencing and the availability of whole genome sequencing.\textsuperscript{7} When a positive genetic diagnosis is made, regarded by the specialist physician to be in keeping with the patients’ clinical phenotype and mode of inheritance, this is recorded with the pedigree in the electronic record. In this study, we interrogated the database to quantify the number of families with pathogenic variants in different genes, to build a picture of the most prevalent causes of IRD, within the limitations of such a retrospective analysis. We performed a similar analysis exclusively in patients under the age of 18, to explore the burden of disease in the pediatric cohort. We also investigated relationships between gene transcript length (of relevance when considering development of gene replacement therapies) and number of families affected. We highlight in particular the 20 most frequently implicated genes, which accounted for over 70% of the cohort.
Methods

Genetic database search

Specialist clinics at Moorfields Eye Hospital receive secondary and tertiary referrals for patients with suspected inherited retinal disease from throughout the UK. Probands, and in many cases family members, are examined by experienced retinal specialists. Once a family is considered solved by the physician (AW, MM, ATM, PYWM, OM), the causative gene is recorded within a “Genetics Module” within the hospital electronic patient record (OpenEyes Electronic Medical Record; information available at https://openeyes.org.uk/). Each pedigree has a unique identifier. In this study, we interrogated the back-end database retrospectively to identify all families with inherited retinal disease in whom a positive molecular diagnosis had been made. The search date was 2 Aug 2019, and identified all families in whom a patient encounter had occurred since 2003.

Genetic testing pathway at Moorfields Eye Hospital

Patients are referred to the retinal genetics service when their primary care physician, optometrist or ophthalmologist suspects an inherited retinal disease. A detailed clinical history is taken from the patient (and/or parents or guardians in the case of children), which includes the presence, age, and order of onset, of symptoms, including night vision problems, central vision disturbances, photophobia or hemeralopia, as well as a full medical history and family history (including construction of a pedigree). Patients undergo ophthalmic examination including visual acuity and intraocular pressure measurement, slit lamp biomicroscopy, and retinal imaging, comprising spectral domain optical coherence tomography and short-wavelength fundus autofluorescence (not always possible in children). Some patients also undergo electroretinography. If patients are suspected by the inherited retinal disease physician of having an inherited retinal disease, genetic testing is discussed. In the past, screening was performed most commonly by Sanger sequencing of single
genes or small panels. The decision to go ahead with genetic testing in the past was based on a number of factors including the patient’s eagerness to be tested (to help inform prognosis and likelihood of transmission to future generations), the likelihood of a positive result, and the possibility of a particular genetic cause that might enable eligibility to treatment trials (early examples were RPE65 and CHM).

In the last decade, next generation sequencing of large gene panels has become more accessible, and testing in our service has been more widely offered and relatively less prone to the above biases. Over the last 5-7 years, including the time period covering the “current cohort” of the present study, our service has sought to offer the opportunity for investigating the molecular diagnosis in all patients suspected by the specialist physician of having an inherited retinal disease. The costs are not borne directly by the patients themselves, but are covered by bodies including the National Health Service (NHS) or its research arm, the National Institute of Health Research (NIHR).

Patients with retinitis pigmentosa, other monogenic chorioretinal degenerations, macular dystrophies, cone and cone-rod dystrophies, stationary conditions (including stationary night blindness and achromatopsia), and suspected syndromic retinal dystrophies, all undergo genetic testing. Some patients (including those late in life, who may have no children) might decline genetic testing, but the majority choose to undergo testing. In some cases, including conditions with very mild changes evident on retinal imaging and minimal symptoms, or adult vitelliform maculopathies, where the chances of a positive genetic diagnosis are lower, genetic testing has not been uniformly considered. Supplementary Figure 1 broadly illustrates the methods and sequence for genetic testing.

During the past 5 years, the following strategy was adopted for genetic testing. For patients with a retinal dystrophy affecting generalised retinal function (with abnormal full-field scotopic or photopic
electroretinograms), a gene panel test was offered covering >150 genes known to be implicated in retinal dystrophies (usually performed by the Manchester Centre for Genomic Medicine). In the presence of known autosomal dominant or X-linked inheritance, a restricted panel was requested covering the relevant genes; for X-linked retinal degeneration, this included a request for specific sequencing of the ORF15 exon of \textit{RPGR} as pathogenic variants in this region can be easily missed. For macular dystrophies, restricted panels were requested, frequently using the Stargardt/Macular Dystrophy Panel of the Molecular Vision Laboratory (Hillsboro, Oregon). Single-gene testing was performed in very few cases in recent years, usually only where a recognisable phenotype implicated a single gene (for example testing for \textit{RS1} in a male with retinoschisis, a pedigree suggestive of X-linked inheritance and a negative electroretinogram). Where results of gene panels were negative, but a monogenic disorder was still strongly suspected, further sequencing was initiated if available (either as part of a clinical or research test), including whole genome sequencing.

Whole genome sequencing was available as part of a number of national research projects from 2013 onwards. Initially this was via the NIHR Bioresource project (described in a previous publication)\textsuperscript{8} and later as part of the “100,000 Genomes” project.\textsuperscript{7} For the latter study, patients were recruited to a pilot study from 2014, with the main study recruiting from 2015 until September 2018. Initial recruitment to the 100,000 Genomes project was for patients who had previously tested negative in initial gene panel screening and for whom DNA samples from additional family members were available. Later, criteria were relaxed, and patients with suspected monogenic disease and no prior testing were eligible, even if samples from family members were not available. The largest number of retinal disease patients were recruited to this study via the retinal genetics service of Moorfields Eye Hospital. When possible, results for patients from our institution are reviewed by a multidisciplinary panel including molecular biologists, clinical geneticists, as well as the retinal specialist managing the family, and consensus is reached, taking into account prior reports of
pathogenicity of the variant, prevalence in publicly available genome databases, the clinical phenotype and mode of inheritance, before the molecular diagnosis is established. Approximately 600 probands in the cohort of the present study achieved a molecular diagnosis by whole genome sequencing.

Finally, prior to access to whole genome sequencing, whole exome sequencing was performed for a number of families (some, but not all, of whom had tested negative previously with single gene or limited gene panel screening). This testing was performed largely at the Institute of Ophthalmology, University College London, and achieved a molecular diagnosis in approximately 160 families of the cohort reported in this paper.

Inclusion of genes and transcript lengths

For the purposes of the present study, only genes listed on the Retinal Information Network online resource (https://sph.uth.edu/retnet/ accessed 10 Oct 2019) were included. Transcript lengths for each gene were extracted from online resources (Ensembl, release 94; longest transcript chosen in case of multiple transcripts). We calculated the correlation between numbers of families affected by variants in each gene and the gene’s transcript size. As the data were not normally distributed, Spearman correlation coefficients were used.

Consent and ethical approval

Patients and relatives gave written informed consent for genetic testing. The study had relevant local research ethics committee approval (Moorfields Eye Hospital and the Northwest London Research Ethics Committee), and conformed to the tenets of the Declaration of Helsinki.
Results

Full cohort

Our study identified 4236 individuals from 3195 families with a molecular diagnosis for their disease. Pathogenic variants were found in 135 distinct genes. The full dataset is given in Supplementary Table 1. The 20 most frequently implicated genes (by number of affected families) were as follows: ABCA4 (20.8% of families); USH2A (9.1%); RPGR (5.1%); PRPH2 (4.6%); BEST1 (3.9%); RS1 (3.5%); RP1 (3.3%); RHO (3.3%); CHM (2.7%); CRB1 (2.1%); PRPF31 (1.8%); MYO7A (1.7%); OPA1 (1.6%); CNG83 (1.4%); RPE65 (1.2%); EYS (1.2%); GUCY2D (1.2%); PROM1 (1.2%); CNGA3 (1.1%); RDH12 (1.1%). These accounted for 71.8% of all molecularly characterised families. Table 1 summarises key features of these genes and Figure 1 schematically demonstrates expression by cellular subtype. Figure 2 illustrates numbers affected by the 30 most frequently implicated genes (by number of affected families and numbers of affected individuals, upper and lower panels respectively). When genes are ranked by numbers of individuals affected, rather than families, autosomal dominant genes, as expected, move upwards in rank (e.g. RHO, TIMP3, PRPF8).

85.3% of families had causative variants in autosomal genes (most frequently ABCA4, USH2A, PRPH2 and BEST1), 13.7% in X-linked genes (most commonly RPGR, RS1 and CHM), and 1.0% in mitochondrial genes (including those implicated in Leber hereditary optic neuropathy (LHON) and maternally inherited diabetes and deafness (MIDD)). Of the autosomal genes, the majority were genes in which variants acted exclusively recessively (52.6% of all families); 8.2% of families had variants in genes in which disease-causing variants are solely dominant; 24.5% had variants in genes which can contain dominant or recessively acting pathogenic variants.
For patients with autosomal dominant RP, the most frequently associated genes were \textit{RHO}, \textit{RP1} and \textit{PRPF31}; for X-linked and autosomal recessive forms of RP, the most frequently associated genes were \textit{RPGR} and \textit{USH2A} respectively. For macular dystrophies, the most common gene by far was \textit{ABCA4} (autosomal recessive), whilst \textit{PRPH2} and \textit{BEST1} were frequently implicated in autosomal dominantly inherited macular dystrophies.

For all genes (excluding mitochondrial) the Spearman coefficient of correlation between number of families and transcript size was 0.20 ($p=0.025$). Figure 3 separately plots numbers of families against transcript size for autosomal genes in which variants act solely recessively (top left panel A), autosomal genes in which variants are solely dominant (right panel B), and for X-linked genes (bottom left panel C). A significant positive correlation was observed in each case with the exception of autosomal dominant genes.

Ethnicity and phenotypic subtypes

Data on ethnicity were not uniformly recorded for all of the patients in the genetics service. However, these data were available almost completely for all of the families recruited from our service for whole genome sequencing via the “100,000 Genomes” project. This distribution of ethnicity is representative of our cohort. Of 1287 IRD probands recruited to the main project, 62.2% were white, 17.9% Asian (largely South Asian), 7.0% were black (African, Caribbean and “other black background”), 1.8% were mixed. These reflect a combination of the demographics of London and the wider United Kingdom (given that many patients seen in the genetics service are referred from outside London, sometimes from outside England). The full ethnic distribution is presented in Supplementary Figure 2.
We were also not able to readily extract phenotypic subgroups in an automated way for from our electronic data record due to variability in data entry and diagnostic labelling. However, this information was available for the above 1287 probands. Of this group, the largest diagnostic category was rod-cone dystrophy (49%), followed by macular dystrophy (35%), inherited optic neuropathy (4.9%), Leber congenital amaurosis or early onset severe retinal dystrophy (3%), rod dysfunction syndrome (3%), cone dysfunction syndrome (2%) and familial exudative vitreoretinopathy (2%). Entries were limited to these categories; patients with cone dystrophies were entered within the macular dystrophy or cone dysfunction categories. The proportions were similar within each of the major ethnic categories (white, Asian, black) and pairwise comparisons did not reveal significant differences, except for a smaller proportion of Asian patients (29.4%) diagnosed with macular dystrophy than the corresponding proportions of white (36.6%) or black patients (43.3%), and more Asian patients diagnosed with familial exudative vitreoretinopathy (FEVR, 4.3%) than black patients (0%). However, the differences were no longer significant following correction for multiple testing.

**Pediatric Cohort**

In order to explore burden of disease in a pediatric population, an additional analysis was performed separately for patients under 18 years. Our search yielded 452 individuals from 411 molecularly diagnosed families with variants in 66 genes. This dataset is given in Supplementary Table 2. The 69 genes implicated in the overall dataset that were not present in those under 18 in our cohort are listed separately in Supplementary Table 3. Figure 4 illustrates the 30 most frequently encountered genes by number of affected families or affected individuals (for comparison with Figure 2). In this cohort, the top 20 genes accounted for 73% of the cohort (by number of affected families).
In the pediatric cohort, 78.8% of families had causative variants in autosomal genes (most frequently \textit{ABCA4} and \textit{BEST1}), 20.7% in X-linked genes (most commonly \textit{RS1} and \textit{RPGR}), and 0.5% in mitochondrial genes (associated with LHON). Of the autosomal genes, the majority were genes in which variants acted exclusively recessively (47.7% of all families); 8.8% of families had variants in genes in which disease-causing variants are solely dominant; 22.4% had variants in genes which can contain dominant or recessively acting pathogenic variants. In comparison to the overall cohort, the proportion of families with causative variants in X-linked genes was significantly greater in the pediatric cohort ($p < 0.001$).

The X-linked genes in both cohorts included \textit{RPGR} (associated with RP or cone-rod dystrophy), \textit{RS1} (associated with X-linked retinoschisis), \textit{CHM} (choroideremia), \textit{CACNA1F} (incomplete congenital stationary night blindness (CSNB)), \textit{RP2} (associated with RP), \textit{NYX} (complete CSNB) and \textit{NDP} (associated with Norrie disease or X-linked FEVR). Of these, some affected females were seen in the overall cohort with disease associated with \textit{RPGR}, \textit{CHM} and \textit{RP2}, consistent with the possibility of females developing symptoms. These tend to be more mild and usually appear later in life than in males. Thus, in the pediatric cohort, very few affected females were seen for the X-linked genes (only 2 females recorded as being affected by variants in \textit{RPGR}).

Notably also, \textit{PRPH2} and \textit{USH2A} were not amongst the most frequently implicated genes, in contrast to the overall cohort, consistent with variants in these genes more frequently leading to visual impairment later in life, relative to some of the other commonly associated genes. On the other hand, some genes associated with congenital stable, or very early onset progressive, visual impairment were amongst the top 10 genes in the pediatric cohort, but not in the overall cohort, as follows: \textit{CACNA1F}, associated with incomplete congenital stationary night blindness; \textit{CNGA3} and
*CNGB3*, associated with achromatopsia; *RPE65* and *CRB1*, associated with Leber congenital amaurosis or early onset severe retinal dystrophy.

**Current cohort**

To reduce the bias inherent in the inclusion of all molecularly characterised families, some of whom will not have accessed the clinical services for many years, but appear due to a historic and specific interest in their disorder, or ease of genetic testing for a specific gene, we conducted a third data search. This was limited to families in which patients had undergone an encounter with our service within the last 2-3 years (specifically between 1 Jan 2017 and the search date, 2 Aug 2019). This may include both clinical examination or a “virtual clinic” consisting of correspondence with patients informing them of their genetic results if these have only recently come to light.

This “current cohort” yielded 3130 individuals from 2614 distinct, molecularly characterised families. Causative variants were in 131 genes. The full dataset is given in Supplementary Table 4. The 20 most frequently implicated genes accounted for 71.2% of the total number of families. Figure 5 illustrates the 30 most common genes by number of families affected and by numbers of individuals, in the same format as Figures 2 and 4. The order of genes was very similar to the overall cohort. The proportions of families with causative variants in X-linked genes (13.5%), in autosomal genes in which pathogenic variants act exclusively dominantly (8.3%), in autosomal genes in which variants act exclusively recessively (53.7%), in autosomal genes in which pathogenic variants can be dominant or recessive (23.5%), and in mitochondrial genes (0.9%), were not significantly different from the corresponding proportions in the overall cohort.
Discussion

In this study, we investigated the burden of inherited retinal disease attributable to different genes in a large United Kingdom cohort of 3197 molecularly diagnosed families (over 4000 affected individuals). This is the largest published molecularly solved IRD cohort to date, as far as the authors are aware. Our families had variants in 135 genes that are associated with IRD on the Retinal Information Network online resource. We found the 20 most frequently involved genes accounted for more than 70% of the cohort. Of these 20 genes, one (RPE65) is the subject of licensed commercially available gene therapy, and a further 7 (ABCA4, CHM, CNGA3, CNGB3, MYO7A, RPGR, RS1) are subject of experimental gene-replacement trials.5

The most frequently encountered gene was ABCA4 (causing Stargardt macular dystrophy, or cone-rod dystrophy). The most frequent gene accounting for autosomal recessive retinitis pigmentosa was USH2A. For autosomal dominant retinitis pigmentosa, the most commonly encountered genes were RHO, RP1 and PRPF31. A significant proportion (nearly 40%) of X-linked retinopathy was due to variants in RPGR. Whilst, as expected, the vast majority of affected individuals with pathogenic variants in X-linked genes were male, some affected females were recorded in the RPGR, CHM and RP2 gene groups, consistent with the known possibility of females being affected; no affected females were seen associated with the other X-linked genes in the overall cohort (and no affected CHM or RP2 females in the pediatric cohort).

We additionally analysed genes implicated in our pediatric cohort, and a more current subsection of the full cohort in an attempt to partially mitigate the effect of historical bias in the overall cohort. Whilst the “current” cohort was very similar to the overall cohort, there were some important differences noted in the pediatric cohort. The proportion of families affected by variants in X-linked
genes was significantly higher in the pediatric cohort. This might reflect the earlier onset and severity of some of the X-linked diseases, and also the likelihood of earlier diagnosis in individuals in whom parents and clinicians are alerted by a positive family history (which is often absent in autosomal recessive conditions, these forming the largest proportion of both cohorts). A number of genes noted in the overall cohort were absent in the pediatric cohort, which might reflect rarity of these variants (and hence their absence in a cohort of smaller size) or that many genotypes lead to later onset visual impairment. PRPH2 and USH2A were amongst the 5 most frequently implicated genes in the overall cohort, but not the pediatric cohort, consistent with older ages of diagnosis (or significant visual impairment) in many cases. Conversely, a number of genes with congenital or early onset visual impairment appeared more frequently in the pediatric cohort (detailed in the Results).

Findings in other cohorts

There have been a number of prior studies of inherited retinal disease cohorts. Table 2 presents the most frequently involved genes in many of these published studies over the last 3 to 4 years. There are obvious similarities in terms of genes affected, across diverse geographic regions. However, there are also interesting differences. Variants in FAM161A account for a substantial proportion of disease in a large Israeli cohort. Variants in EYS were a more frequent cause of disease than USH2A in Korean patients; this has also been reported in a large cohort of Japanese retinitis pigmentosa patients. In a large Chinese retinitis pigmentosa cohort, CYP4V2 was the second most implicated gene after USH2A. Differences between populations can reflect founder effects, and are important in guiding genetic testing and future interpretation of results of whole genome sequencing. In addition, a relative paucity of studies of inherited retinal disease cohorts in other large regions, such as Africa, is apparent, and worthy of addressing in future investigations.
Rates of consanguinity also differ between population groups. When consanguinity or endogamy is more common, autosomal recessive diseases associated with homozygous variants will be more likely. Recently published findings from the United Arab Emirates (UAE)\textsuperscript{10} showed that the most frequently implicated genes in a pediatric cohort were those in which pathogenic variants are recessively inherited, with many associated with homozygous variants. In contrast, in our pediatric cohort, after \textit{ABCA4}, the next 4 most frequently inherited genes were associated with X-linked or predominantly autosomal dominant disease (although recessive disease did feature in a number of the top 20 genes). Eliciting a history of consanguinity can be helpful not just in selecting genes for screening, but also in interpreting results of whole genome sequencing, where preliminary focus can concentrate on regions of homozygosity. Other modes of inheritance are of course possible even in consanguineous cohorts: in the pediatric study from the UAE, \textit{RS1} (X-linked) and \textit{BEST1} (usually associated with autosomal dominant disease) also featured in a number of families.\textsuperscript{10}

Genetic testing strategies and their accessibility also differ between countries: those in which targeted restricted gene panels are employed selectively in patients with recognisable phenotypes could lead to a greater reported prevalence of those genes (for example, possibly contributing to the higher prevalence of \textit{KCNV2} retinopathy, which has a pathognomonic electroretinography phenotype,\textsuperscript{29} in the UAE study).\textsuperscript{10} The availability of whole genome sequencing to a proportion of our cohort, as part of a national research project, and access to particular gene panels with testing paid for by the National Health Service or its research arm, might not be applicable to other countries with different accessibility to clinical and research tests, and different arrangements for reimbursement.

\textit{Correlations with transcript length}
The length of the transcript is of relevance in the context of gene replacement therapy; there is a limit to the size of cDNA that can be delivered by different virus vectors. Adeno-associated viruses have been a vector of choice for ocular gene therapy trials, targeting retinal cells with relatively low immunogenicity, but their capacity is limited. We explored transcript lengths and relationships with numbers of families affected. We found a weak, but statistically significant, correlation in the overall cohort. For autosomal genes in which pathogenic variants act recessively, the correlation remained significant, whilst there was no apparent correlation for dominant genes.

Longer transcripts might be expected, by virtue of their length, to contain more sites in which a variant can potentially bring about premature termination or loss of function, which is the usual mode of action in recessive disease. Thus a greater prevalence of pathogenic variants in longer genes might be anticipated. For many dominant diseases, however, loss of function variants in many cases do not cause disease. Pathogenicity is frequently consequent upon a gain of function or specific effects of mutations (for some genes, only a few dominantly acting variants have been identified), and so prevalence of disease might not be expected to correlate in the same way with transcript length. In contrast, X-linked disease is often a result of loss of the single functioning allele in males (and again longer genes might have more sites at which mutation can lead to premature termination or loss of function); this might explain the significant correlation between number of affected families and transcript size observed for X-linked genes (although the number of genes here is relatively small).

Limitations

Our findings should be taken in the context of a number of important limitations inherent in such a retrospective study. The study relies on prior data entry, which might be incomplete or inconsistent.
or in some cases contain errors, although efforts are made to correct these when they come to light. It is likely that a number of genes are over-represented, including those discovered earlier, those more amenable to sequencing by earlier methods, or those in which there was historic or current interest particularly in light of potential gene-specific therapies. This effect will lessen over time, as more patients have undergone whole genome sequencing, permitting unbiased analysis of data. Detection of structural variants and variants in non-coding regions can still be challenging, as can detection of pathogenic variants in the repetitive ORF15 exon of RPGR. The latter can be easily missed in whole genome sequencing and so the burden of disease due to RPGR might be underestimated. In addition, some of the earlier results pre-date current guidelines and the availability of large databases of common variants, and so variants previously classified as pathogenic may no longer be regarded so.

We sought to partly mitigate the effect of historical biases by performing a time-limited analysis of more current patient data, which comprised a large proportion of the overall cohort. During the time period pertaining to the “current” cohort, clinical or research genetic testing was routinely offered to all patients who were reviewed in clinic and were suspected by the specialist physician of having an inherited retinal disease, with no direct cost borne by the patient.

A further source of potential ascertainment bias relates to the types of patients managed in our service. Although Moorfields Eye Hospital cares for both children and adults, some of the more severe syndromic conditions tend to be managed in other specialist centres, with multidisciplinary medical input. Thus these disorders are likely to be under-represented in our cohort.
Given the retrospective nature of the study, we were unable to ascertain a number of other potentially useful data. The total number of patients enrolled for genetic testing was not available, thus precluding calculation of a “molecularly solved rate” for the entire cohort. Whole genome sequencing, when available, was initially offered to patients who had tested negative with prior gene panels, but was later offered to all patients, thus making this a mixed group. In a prior study partly from our service, 63% of patients with no prior testing achieved a molecular diagnosis from whole exome/genome sequencing, compared with 54% of those who had previously tested negative on prior gene panels.

Also, date of first symptoms or first clinical diagnosis was not available. Due to variability in data entry, we could not readily extract the proportions attributable to particular variants of each gene, or ethnicity data by genotype, or the specific frequency of phenotypic subgroups for the whole cohort, but these would be useful subjects for further exploration. Our findings thus give a sense of relative burdens of disease attributable to different genes in a large multi-ethnic United Kingdom-based cohort, but might not apply precisely to other populations with different ethnic compositions (as discussed in relation to Table 2) or with different availabilities or strategies for genetic testing.

With parallel developments in genomic testing and novel therapies, we envisage that it will become a standard of care to seek the molecular diagnosis in the majority of inherited retinal disease patients. Quantification of burden of disease attributable to particular genes, and particular genetic variants, in diverse populations will be important in guiding both individual patient management as well as planning within healthcare systems to address this important cause of blindness.
References


Figure Legends

Figure 1. Schematic of retina, showing site of expression of proteins encoded by the 20 most frequently implicated genes in the cohort.

Figure 2. The 30 most frequently involved genes in the full cohort. A, Genes ranked by numbers of affected families. B, Genes ranked by numbers of affected individuals.

Figure 3. Numbers of affected families plotted against transcript length. A, Autosomal genes in which pathogenic variants act exclusively recessively (Spearman correlation coefficient 0.27, \( p=0.017 \)). B, Autosomal genes in which pathogenic variants act exclusively dominantly (correlation coefficient -0.17, \( p=0.459 \)). C, X-linked genes (correlation coefficient 0.71, \( p=0.047 \)).

Figure 4. The 30 most frequently involved genes in the cohort under 18 years of age. A, Genes ranked by numbers of affected families. B, Genes ranked by numbers of affected individuals.

Figure 5. The 30 most frequently involved genes in the “current” cohort (in whom a patient encounter had occurred within the preceding 2.5 years). A, Genes ranked by numbers of affected families. B, Genes ranked by numbers of affected individuals.
Table Legends

**Table 1.** The 20 most frequently implicated genes in the full cohort (by number of families). Modes of inheritance and range of possible phenotypes are given. RP, retinitis pigmentosa; LCA, Leber Congenital Amaurosis.

**Table 2.** Selected previous studies in inherited retinal disease cohorts. Some authors report results from gene panel or whole exome/genome testing, leading to likely under-representation of disorders diagnosed with single-gene testing. Studies restricted to specific phenotypes (e.g. retinitis pigmentosa) are not shown. Some smaller cohorts are included to allow wider geographical representation. The right-hand common gives the most frequently implicated genes. (In most cases these are the top 5, but where multiple genes contributed the same proportion, additional genes might be included.) For some 2019 studies, year published relates to year of online publication (print publication in some cases was in 2020). *Some data relating to this study was taken from the publication Farrar et al.*
Supplementary Data

Supplementary Figure 1. Strategy for genetic testing in our cohort. Testing was performed via a combination of single gene tests, gene panels and whole genome sequencing. The majority of patients in this molecularly characterised cohort achieved a diagnosis via gene panel testing (middle box, associated with large solid arrow). Historically, single-gene tests were used (left box). If patients were negative for single gene testing, gene panel testing was employed (left hand dashed arrow). More recently, for a period of 5 years (2013 to 2018) patients were recruited for whole genome sequencing (initially whole exome, then whole genome sequencing, for the NIHR Bioresource study), represented by the right-hand box. Some of these patients had previously tested negative for single gene or gene panel tests (indicated by horizontal dashed arrows) and some were recruited directly (right solid arrow). For a limited period prior to this, whole exome sequencing (not shown) was performed as part of research studies (largely based at the Institute of Ophthalmology, University College London), which led the molecular diagnosis in 162 families.

Supplementary Figure 2. Ethnic distribution of probands. Pie chart shows proportions of each ethnicity for all 1287 IRD probands recruited from the genetics service to the “100,000 Genomes” project. The category “South Asian” includes those from Indian, Pakistani and Bangladeshi backgrounds.

Supplementary Table 1. Results (genes, numbers of families, numbers of affected individuals, published modes of inheritance) for the full cohort. LHON mutations have been combined.

Supplementary Table 2. Results (genes, numbers of families, numbers of affected individuals, published modes of inheritance) for individuals under 18 years.
Supplementary Table 3. Genes in adult, but not pediatric cohort.

Supplementary Table 4. Results (genes, numbers of families, numbers of affected individuals, published modes of inheritance) for the "current" cohort.
This study explored burden of disease attributable to different genes in >3000 genetically characterised families. Over 70% of families had variants in the top 20 genes. Numbers correlated with transcript length (except in dominant disease).
A

Numbers of affected families by gene (pediatric cohort)

4 families for each of CLN3, CRX, FZD4, KCNV2, MYO7A, NYX, PNPLA6, RPRIP1, TRPM1

B

Numbers of affected individuals by gene (pediatric cohort)

4 individuals for each of CLN3, FZD4, KCNV2, MYO7A, NYX, PNPLA6, RPRIP1
A

Numbers of affected families by gene (current cohort)

B

Numbers of affected individuals by gene (current cohort)

17 families for each of CERKL, CNGB1, KCNV2
### Table 1. The 20 most frequently implicated genes in the full cohort (by number of families). Modes of inheritance and range of possible phenotypes are given. RP, retinitis pigmentosa; LCA, Leber Congenital Amaurosis.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosomal location</th>
<th>Number of families affected (%)</th>
<th>Number of individuals affected (%)</th>
<th>Modes of inheritance</th>
<th>Range of phenotypes in literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCA4</td>
<td>1p22.1</td>
<td>666 (20.8)</td>
<td>789 (18.6)</td>
<td>Recessive</td>
<td>Stargardt macular dystrophy; cone-rod dystrophy</td>
</tr>
<tr>
<td>USH2A</td>
<td>1q41</td>
<td>292 (9.1)</td>
<td>342 (8.1)</td>
<td>Recessive</td>
<td>RP; Type 2 Usher Syndrome</td>
</tr>
<tr>
<td>RPGR</td>
<td>Xp11.4</td>
<td>164 (5.1)</td>
<td>263 (6.2)</td>
<td>X-linked</td>
<td>RP; Cone or cone-rod dystrophy</td>
</tr>
<tr>
<td>PRPH2</td>
<td>6p21.1</td>
<td>148 (4.6)</td>
<td>220 (5.2)</td>
<td>Dominant and Recessive</td>
<td>Pattern dystrophy; RP</td>
</tr>
<tr>
<td>BEST1</td>
<td>11q12.3</td>
<td>125 (3.9)</td>
<td>168 (4.0)</td>
<td>Dominant and Recessive</td>
<td>Best disease; autosomal recessive bestrophinopathy</td>
</tr>
<tr>
<td>RS1</td>
<td>Xp22.13</td>
<td>111 (3.5)</td>
<td>134 (3.2)</td>
<td>X-linked</td>
<td>X-linked retinoschisis</td>
</tr>
<tr>
<td>RP1</td>
<td>8q12.1</td>
<td>106 (3.3)</td>
<td>170 (4.0)</td>
<td>Dominant and Recessive</td>
<td>RP</td>
</tr>
<tr>
<td>RHO</td>
<td>3q22.1</td>
<td>105 (3.3)</td>
<td>177 (4.2)</td>
<td>Dominant and Recessive</td>
<td>RP; stationary night blindness</td>
</tr>
<tr>
<td>CHM</td>
<td>Xq21.2</td>
<td>86 (2.7)</td>
<td>112 (2.6)</td>
<td>X-linked</td>
<td>Choroideremia</td>
</tr>
<tr>
<td>CRB1</td>
<td>1q31.3</td>
<td>68 (2.1)</td>
<td>86 (2.0)</td>
<td>Recessive</td>
<td>LCA; RP; macular dystrophy</td>
</tr>
<tr>
<td>PRPF31</td>
<td>19q13.42</td>
<td>57 (1.8)</td>
<td>94 (2.2)</td>
<td>Dominant</td>
<td>RP</td>
</tr>
<tr>
<td>MYO7A</td>
<td>11q13.5</td>
<td>53 (1.7)</td>
<td>58 (1.4)</td>
<td>Recessive</td>
<td>Type 1 Usher syndrome</td>
</tr>
<tr>
<td>OPA1</td>
<td>3q29</td>
<td>50 (1.6)</td>
<td>84 (2.0)</td>
<td>Dominant</td>
<td>Optic atrophy; optic atrophy with sensorineural hearing loss</td>
</tr>
<tr>
<td>CNGB3</td>
<td>8q21.3</td>
<td>44 (1.4)</td>
<td>55 (1.3)</td>
<td>Recessive</td>
<td>Achromatopsia; cone dystrophy</td>
</tr>
<tr>
<td>RPE65</td>
<td>1p31.2</td>
<td>39 (1.2)</td>
<td>51 (1.2)</td>
<td>Recessive and Dominant</td>
<td>LCA; RP</td>
</tr>
<tr>
<td>EYS</td>
<td>6q12</td>
<td>38 (1.2)</td>
<td>43 (1.0)</td>
<td>Recessive</td>
<td>RP</td>
</tr>
<tr>
<td>GUCY2D</td>
<td>17p13.1</td>
<td>37 (1.2)</td>
<td>54 (1.3)</td>
<td>Recessive and Dominant</td>
<td>LCA; RP; Cone or cone-rod dystrophy</td>
</tr>
<tr>
<td>PROM1</td>
<td>4p15.32</td>
<td>37 (1.2)</td>
<td>53 (1.2)</td>
<td>Recessive and Dominant</td>
<td>Macular dystrophy; cone-rod dystrophy; RP</td>
</tr>
<tr>
<td>CNGA3</td>
<td>2q11.2</td>
<td>36 (1.1)</td>
<td>50 (1.2)</td>
<td>Recessive</td>
<td>Achromatopsia; cone dystrophy</td>
</tr>
<tr>
<td>RDH12</td>
<td>14q24.1</td>
<td>35 (1.1)</td>
<td>44 (1.0)</td>
<td>Recessive and Dominant</td>
<td>LCA; RP</td>
</tr>
<tr>
<td>Year published</td>
<td>Authors</td>
<td>Study cohort/country</td>
<td>Number molecularly diagnosed (number of genes)</td>
<td>Most frequently implicated genes</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>---------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>Current study</td>
<td>Khan AO</td>
<td>United Arab Emirates (children)</td>
<td>4241 individuals from 3197 families (135 genes)</td>
<td>By family: ABCA4, USH2A, RPGR, PRPH2, BEST1 By individuals: ABCA4, USH2A, RPGR, PRPH2, RHO</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Sharon et al.</td>
<td>Israel</td>
<td>71 individuals (26 genes)</td>
<td>ABCA4, KCNV2, CRB1, CNGA3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Holtan et al.</td>
<td>Norway</td>
<td>1369 families (129 genes)</td>
<td>ABCA4, USH2A, FAM161A, CNGA3, EYS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avela et al.</td>
<td>Finland (children)</td>
<td>207 patients (56 genes)</td>
<td>ABCA4, USH2A, BEST1, RHO, RS1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kim et al.</td>
<td>Korea</td>
<td>41 families (17 genes)</td>
<td>RS1, GUCY2D, RPGR</td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>Motta et al.</td>
<td>Brazil</td>
<td>38 individuals (24 genes)</td>
<td>ABCA4, RPE65, CERKL, RPGRIP1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wang et al.</td>
<td>China</td>
<td>36 families (19 genes)</td>
<td>ABCA4, EPAC1, CRB1, CNGA3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stone et al.</td>
<td>United States</td>
<td>400 individuals (66 genes)</td>
<td>ABCA4, CEPP90, USH2A, CRB1, RPGR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carss et al.</td>
<td>United Kingdom</td>
<td>760 families (104 genes)</td>
<td>ABCA4, USH2A, CRB1, CEP290</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>*Dockery et al.</td>
<td>Ireland (adults)</td>
<td>404 individuals (94 genes)</td>
<td>ABCA4, USH2A, EYS, RP1, CACNA1F, RPGR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ellingford et al.</td>
<td>UK, genomic laboratory</td>
<td>357 families (59 genes)</td>
<td>ABCA4, USH2A, BBS1, RHO, RP1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haer-Wigman et al.</td>
<td>Netherlands</td>
<td>132 families (47 genes)</td>
<td>USH2A, CRB1, ABCA4, CERKL, CEPP90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Riera et al.</td>
<td>Spain</td>
<td>271 individuals (62 genes)</td>
<td>ABCA4, USH2A, RPE65, CRB1, EYS, GUCY2D, PDE6B</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>Tiwari et al.</td>
<td>Switzerland</td>
<td>136 individuals (56 genes)</td>
<td>ABCA4, USH2A, C2orf71, RP1, CEP290, FLVCR1, CRB1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bernardis et al.</td>
<td>Italy</td>
<td>42 individuals (29 genes)</td>
<td>ABCA4, USH2A, RPE65, CRB1, EYS, GUCY2D, PDE6B</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Selected previous studies in inherited retinal disease cohorts. Some authors report results from gene panel or whole exome/genome testing, leading to likely under-representation of disorders diagnosed with single-gene testing. Studies restricted to specific phenotypes (e.g. retinitis pigmentosa) are not shown. Some smaller cohorts are included to allow wider geographical representation. The right-hand common gives the most frequently implicated genes. (In most cases these are the top 5, but where multiple genes contributed the same proportion, additional genes might be included.) For some 2019 studies, year published relates to year of online publication (print publication in some cases was in 2020). *Some data relating to this study was taken from the publication Farrar et al.25