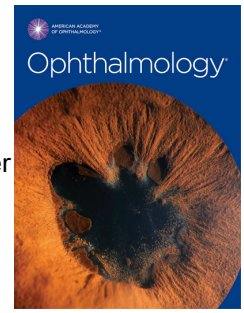
Journal Pre-proof



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

Please cite this article as: Pontikos N, Arno G, Jurkute N, Schiff E, Ba-Abbad R, Malka S, Gimenez A, Georgiou M, Wright G, Armengol M, Knight H, Katz M, Moosajee M, Yu-Wai-Man P, Moore AT, Michaelides M, Webster AR, Mahroo OA, Genetic basis of inherited retinal disease in a molecularly characterised cohort of over 3000 families from the United Kingdom, *Ophthalmology* (2020), doi: https://doi.org/10.1016/j.ophtha.2020.04.008.

Genetic basis of inherited retinal disease in a molecularly characterised 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, FRCOphth[1,2,7,8]

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%); *CNGA3* (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.[1-3] Pathogenic variants in over 250 genes can give rise to inherited retinal disease (IRD), with multiple modes of inheritance.[4] For the majority of these diseases, there are no medical or surgical treatments, but a large number of therapeutic trials are underway.[5] There is now a commercially available licensed gene-replacement treatment for a particular genetic cause, namely IRD due to bi-allelic variants in *RPE65*.[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.[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 *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 *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)[8] and later as part of the "100,000 Genomes" project.[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,[9] 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%); *CNGB3* (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 *RHO*, *RP1* and *PRPF31*; for X-linked and autosomal recessive forms of RP, the most frequently associated genes were *RPGR* and *USH2A* respectively. For macular dystrophies, the most common gene by far was *ABCA4* (autosomal recessive), whilst *PRPH2* and *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 *ABCA4* and *BEST1*), 20.7% in X-linked genes (most commonly *RS1* and *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 *RPGR* (associated with RP or cone-rod dystrophy), *RS1* (associated with X-linked retinoschisis), *CHM* (choroideremia), *CACNA1F* (incomplete congenital stationary night blindness (CSNB)), *RP2* (associated with RP), *NYX* (complete CSNB) and *NDP* (associated with Norrie disease or X-linked FEVR). Of these, some affected females were seen in the overall cohort with disease associated with *RPGR*, *CHM* and *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 *RPGR*).

Notably also, *PRPH2* and *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: *CACNA1F*, associated with incomplete congenital stationary night blindness; *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.[8,10-27] 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.[11] Variants in *EYS* were a more frequent cause of disease than *USH2A* in Korean patients;[12] this has also been reported in a large cohort of Japanese retinitis pigmentosa patients.[27] In a large Chinese retinitis pigmentosa cohort, *CYP4V2* was the second most implicated gene after *USH2A*.[28] 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)[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 *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, *RS1* (X-linked) and *BEST1* (usually associated with autosomal dominant disease) also featured in a number of families.[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 *KCNV2* retinopathy, which has a pathognomonic electroretinography phenotype,[29] in the UAE study).[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.

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.[5] 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 under-estimated. In addition, some of the earlier results pre-date current guidelines[9] 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 mulitidisciplinary 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,[8] 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

1. 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 Feb 12;4(2):e004015.
2. Solebo AL, Teoh L, Rahi J. Epidemiology of blindness in children. Arch Dis Child. 2017 Sep;102(9):853-857.
3. Solebo AL, Rahi J. Epidemiology, aetiology and management of visual impairment in children. Arch Dis Child. 2014 Apr;99(4):375-9.
4. Farrar GJ, Carrigan M, Dockery A, Millington-Ward S, Palfi A, Chadderton N, Humphries M, Kiang AS, Kenna PF, Humphries P. Toward an elucidation of the molecular genetics of inherited retinal degenerations. Hum Mol Genet. 2017 Aug 1;26(R1):R2-R11.
5. Vázquez-Domínguez I, Garanto A, Collin RWJ. Molecular Therapies for Inherited Retinal Diseases-Current Standing, Opportunities and Challenges. Genes (Basel). 2019 Aug 28;10(9). pii: E654.
6. Russell S, Bennett J, Wellman JA, Chung DC, Yu ZF, Tillman A, Wittes J, Pappas J, Elci O, McCague S, Cross D, Marshall KA, Walshire J, Kehoe TL, Reichert H, Davis M, Raffini L, George LA, Hudson FP, Dingfield L, Zhu X, Haller JA, Sohn EH, Mahajan VB, Pfeifer W, Weckmann M, Johnson C, Gewaily D, Drack A, Stone E, Wachtel K, Simonelli F, Leroy BP, Wright JF, High KA, Maguire AM. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. Lancet. 2017 Aug 26;390(10097):849-860.
7. Turnbull C, Scott RH, Thomas E, et al. The 100,000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ 2018;361:k1687.
8. Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S, Plagnol V, Penkett C, Stirrups K, Rizzo R, Wright G, Josifova D, Bitner-

Glindzicz M, Scott RH, Clement E, Allen L, Armstrong R, Brady AF, Carmichael J, Chitre M, Henderson RHH, Hurst J, MacLaren RE, Murphy E, Paterson J, Rosser E, Thompson DA, Wakeling E, Ouwehand WH, Michaelides M, Moore AT; NIHR-BioResource Rare Diseases Consortium, Webster AR, Raymond FL. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. Am J Hum Genet. 2017 Jan 5;100(1):75-90.

9. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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 May;17(5):405-24.
10. Khan AO. PHENOTYPE-GUIDED GENETIC TESTING OF PEDIATRIC INHERITED RETINAL DISEASE IN THE UNITED ARAB EMIRATES. Retina. 2019 Nov 8. doi: 10.1097/IAE.0000000000002675. [Epub ahead of print]
11. Sharon D, Ben-Yosef T, Goldenberg-Cohen N, Pras E, Gradstein L, Soudry S, Mezer E, Zur D, Abbasi AH, Zeitz C, Cremers FPM, Khan MI, Levy J, Rotenstreich Y, Birk OS, Ehrenberg M, Leibu R, Newman H, Shomron N, Banin E, Perlman I. A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). Hum Mutat. 2020 Jan;41(1):140-149.
12. Holtan JP, Selmer KK, Heimdal KR, Bragadóttir R. Inherited retinal disease in Norway - a characterization of current clinical and genetic knowledge. Acta Ophthalmol. 2019 Aug 19. doi: 10.1111/aos.14218. [Epub ahead of print]
13. Avela K, Salonen-Kajander R, Laitinen A, Ramsden S, Barton S, Rudanko SL. The genetic aetiology of retinal degeneration in children in Finland - new founder mutations identified. Acta Ophthalmol. 2019 Dec;97(8):805-814.

14. Kim MS, Joo K, Seong MW, Kim MJ, Park KH, Park SS, Woo SJ. Genetic Mutation Profiles in Korean Patients with Inherited Retinal Diseases. J Korean Med Sci. 2019 Jun 2;34(21):e161.
15. Tayebi N, Akinrinade O, Khan MI, Hejazifar A, Dehghani A, Cremers FPM, Akhlaghi M. Targeted next generation sequencing reveals genetic defects underlying inherited retinal disease in Iranian families. Mol Vis. 2019 Feb 8;25:106-117. eCollection 2019.
16. Motta FL, Martin RP, Filippelli-Silva R, Salles MV, Sallum JMF. Relative frequency of inherited retinal dystrophies in Brazil. Sci Rep. 2018 Oct 29;8(1):15939.
17. Wang L, Zhang J, Chen N, Wang L, Zhang F, Ma Z, Li G, Yang L. Application of Whole Exome and Targeted Panel Sequencing in the Clinical Molecular Diagnosis of 319 Chinese Families with Inherited Retinal Dystrophy and Comparison Study. Genes (Basel). 2018 Jul 19;9(7). pii: E360.
18. Stone EM, Andorf JL, Whitmore SS, DeLuca AP, Giacalone JC, Streb LM, Braun TA, Mullins RF, Scheetz TE, Sheffield VC, Tucker BA. Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease. Ophthalmology. 2017 Sep;124(9):1314-1331.
19. Dockery A, Stephenson K, Keegan D, Wynne N, Silvestri G, Humphries P, Kenna PF, Carrigan M, Farrar GJ. Target 5000: Target Capture Sequencing for Inherited Retinal Degenerations. Genes (Basel). 2017 Nov 3;8(11). pii: E304.
20. Ellingford JM, Barton S, Bhaskar S, O'Sullivan J, Williams SG, Lamb JA, Panda B, Sergouniotis PI, Gillespie RL, Daiger SP, Hall G, Gale T, Lloyd IC, Bishop PN, Ramsden SC, Black GCM. Molecular findings from 537 individuals with inherited retinal disease. J Med Genet. 2016 Nov;53(11):761-767.
21. Haer-Wigman L, van Zelst-Stams WA, Pfundt R, van den Born LI, Klaver CC, Verheij JB, Hoyng CB, Breuning MH, Boon CJ, Kievit AJ, Verhoeven VJ, Pott JW, Sallevelt SC, van Hagen JM, Plomp AS, Kroes HY, Lelieveld SH, Hehir-Kwa JY, Castelein S, Nelen M, Scheffer H, Lugtenberg

D, Cremers FP, Hoefsloot L, Yntema HG. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. Eur J Hum Genet. 2017 May;25(5):591-599.

22. Riera M, Navarro R, Ruiz-Nogales S, Méndez P, Burés-Jelstrup A, Corcóstegui B, Pomares E. Whole exome sequencing using Ion Proton system enables reliable genetic diagnosis of inherited retinal dystrophies. Sci Rep. 2017 Feb 9;7:42078.
23. Tiwari A, Bahr A, Bähr L, Fleischhauer J, Zinkernagel MS, Winkler N, Barthelmes D, Berger L, Gerth-Kahlert C, Neidhardt J, Berger W. Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. Sci Rep. 2016 Jun 29;6:28755.
24. Bernardis I, Chiesi L, Tenedini E, Artuso L, Percesepe A, Artusi V, Simone ML, Manfredini R, Camparini M, Rinaldi C, Ciardella A, Graziano C, Balducci N, Tranchina A, Cavallini GM, Pietrangelo A, Marigo V, Tagliafico E. Unravelling the Complexity of Inherited Retinal Dystrophies Molecular Testing: Added Value of Targeted Next-Generation Sequencing. Biomed Res Int. 2016;2016:6341870.
25. Farrar GJ, Carrigan M, Dockery A, Millington-Ward S, Palfi A, Chadderton N, Humphries M, Kiang AS, Kenna PF, Humphries P. Toward an elucidation of the molecular genetics of inherited retinal degenerations. Hum Mol Genet. 2017 Aug 1;26(R1):R2-R11.
26. Martin-Merida I, Avila-Fernandez A, Del Pozo-Valero M, Blanco-Kelly F, Zurita O, Perez-Carro R, Aguilera-Garcia D, Riveiro-Alvarez R, Arteche A, Trujillo-Tiebas MJ, Tahsin-Swafiri S, Rodriguez-Pinilla E, Lorda-Sanchez I, Garcia-Sandoval B, Corton M, Ayuso C. Genomic Landscape of Sporadic Retinitis Pigmentosa: Findings from 877 Spanish Cases. Ophthalmology. 2019 Aug;126(8):1181-1188.
27. Koyanagi Y, Akiyama M, Nishiguchi KM, Momozawa Y, Kamatani Y, Takata S, Inai C, Iwasaki Y, Kumano M, Murakami Y, Omodaka K, Abe T, Komori S, Gao D, Hirakata T, Kurata K, Hosono K, Ueno S, Hotta Y, Murakami A, Terasaki H, Wada Y, Nakazawa T, Ishibashi T, Ikeda Y, Kubo

549 M, Sonoda KH. Genetic characteristics of retinitis pigmentosa in 1204 Japanese patients. J
550 Med Genet. 2019 Oct;56(10):662-670.
551 28. Gao FJ, Li JK, Chen H, Hu FY, Zhang SH, Qi YH, Xu P, Wang DD, Wang LS, Chang Q, Zhang YJ,
552 Liu W, Li W, Wang M, Chen F, Xu GZ, Wu JH. Genetic and Clinical Findings in a Large Cohort
553 of Chinese Patients with Suspected Retinitis Pigmentosa. Ophthalmology. 2019
554 Nov;126(11):1549-1556.
555 29. Vincent A, Robson AG, Holder GE. Pathognomonic (diagnostic) ERGs. A review and update.
556 Retina. 2013 Jan;33(1):5-12.

557

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.

579 **Table Legends**

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

583

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

593

594

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.

619

620 **Supplementary Table 3. Genes in adult, but not pediatric cohort.**

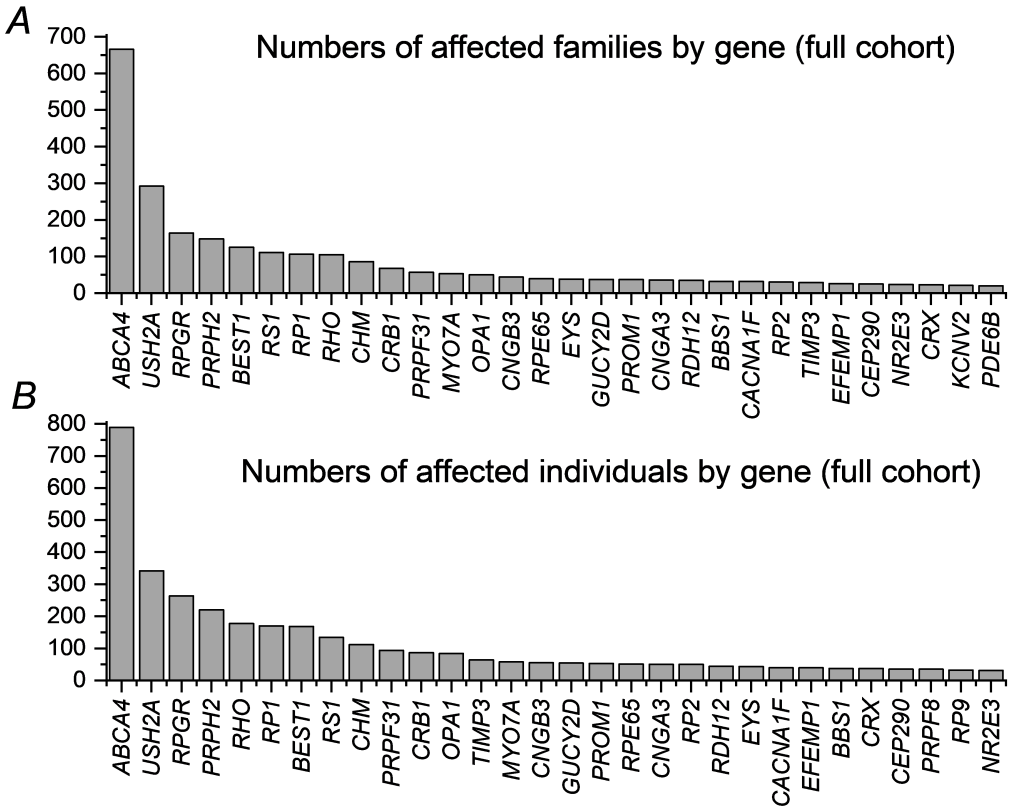
621

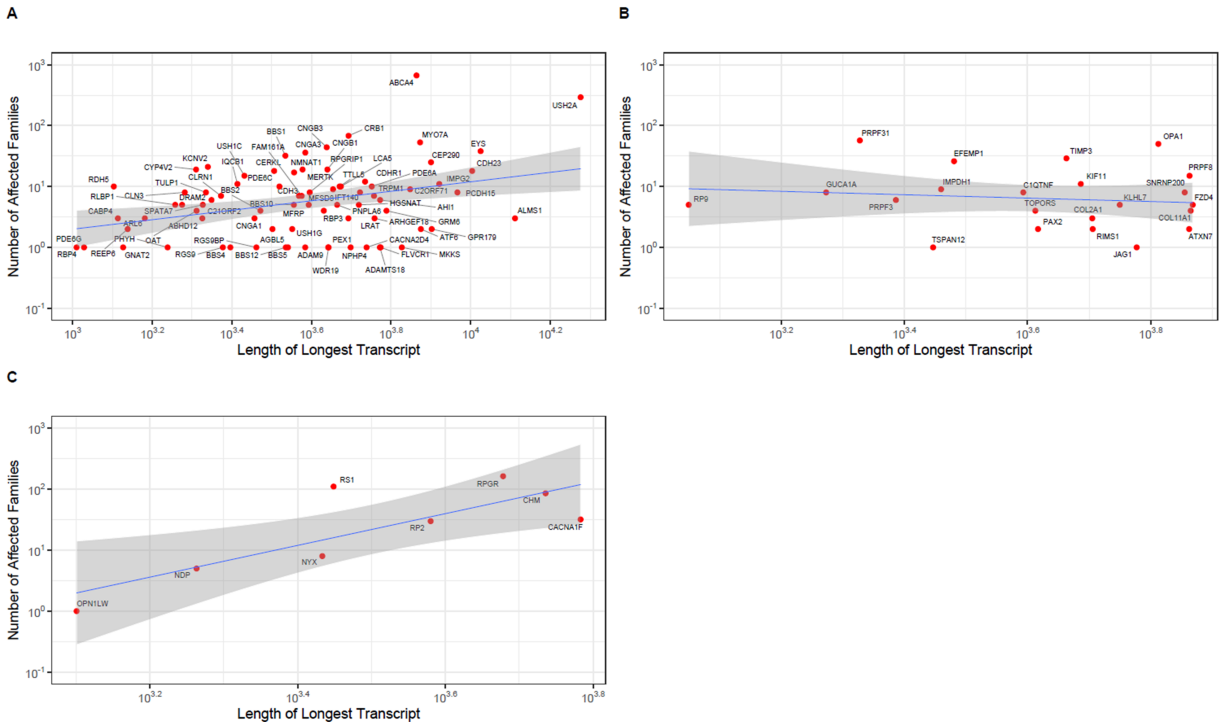
622 **Supplementary Table 4. Results (genes, numbers of families, numbers of affected individuals,**

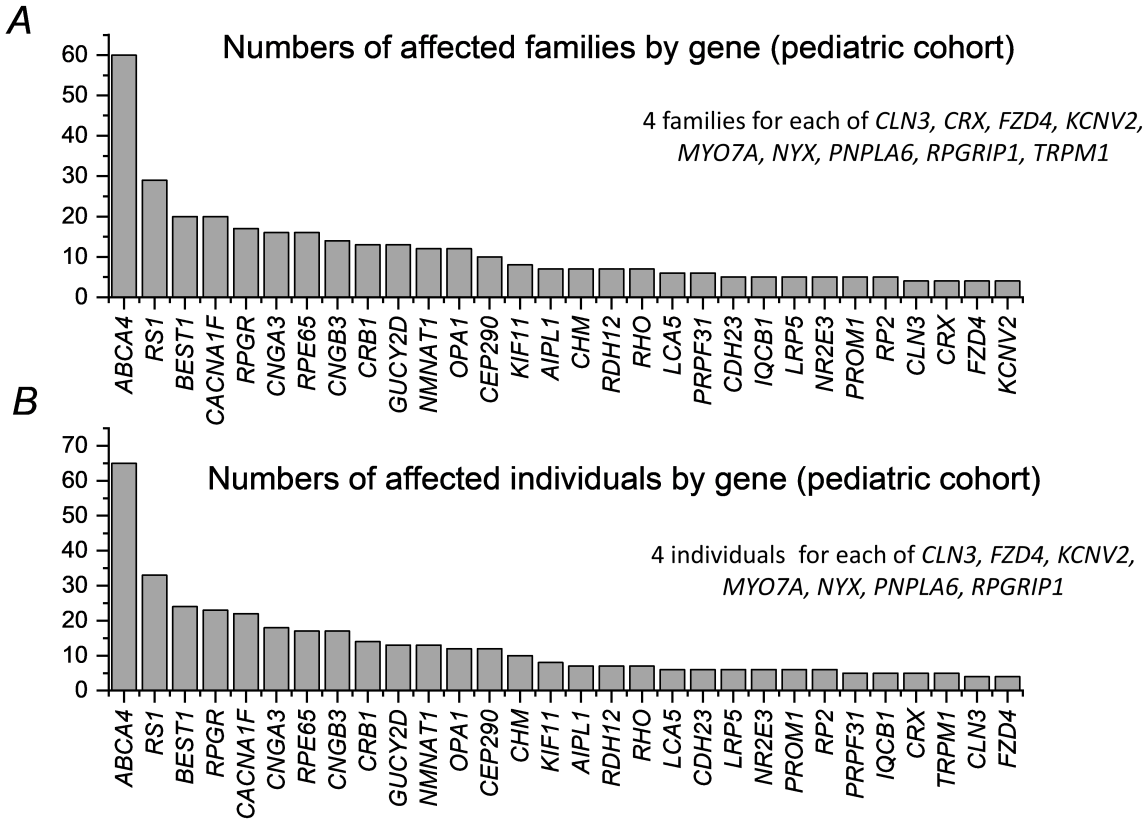
623 **published modes of inheritance) for the "current" cohort.**

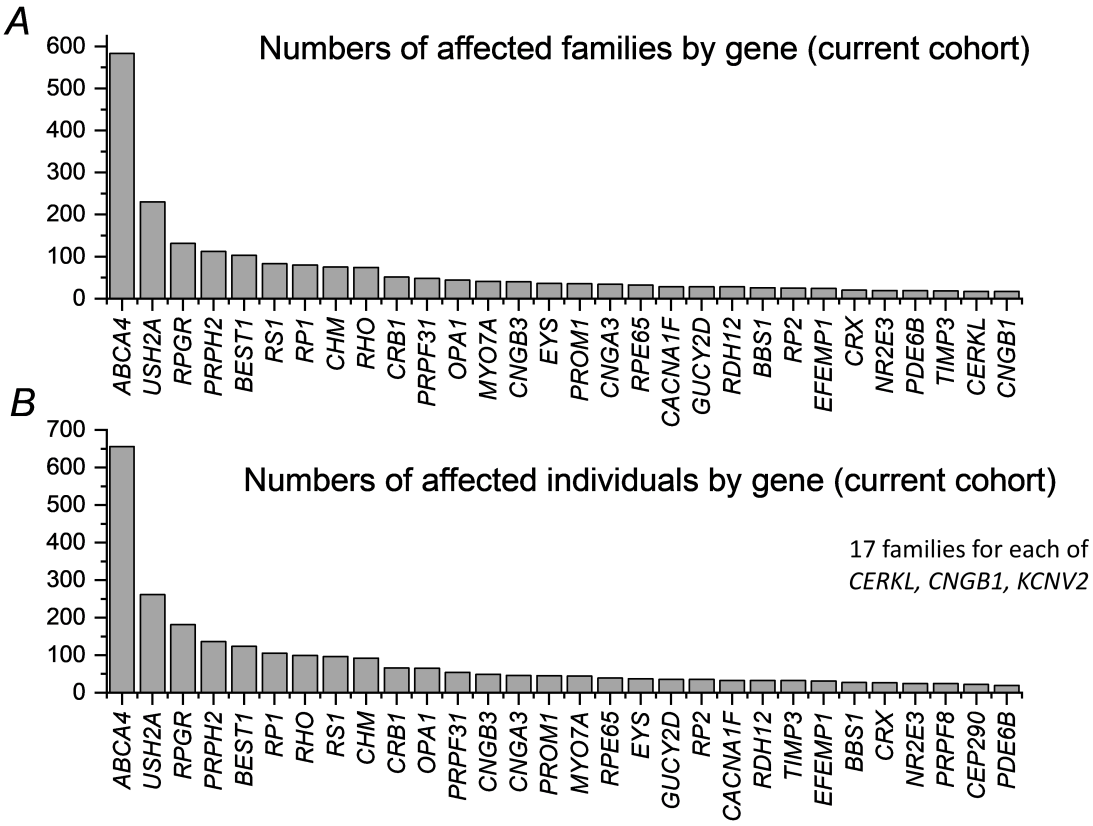
624

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 (full cohort)
700
600
500
400
300
200
100
0
ABCA4
USH2A
RPGR
PRPH2
BEST1
RS1
RP1
RHO
CHM
CRB1
PRPF31
MYO7A
OPA1
CNGB3
RPE65
EYS
GUCY2D
PROM1
CNGA3
RDH12
BBS1
CACNA1F
RP2
TIMP3
EFEMP1
CEP290
NR2E3
CRX
KCNV2
PDE6B
B
Numbers of affected individuals by gene (full cohort)
800
700
600
500
400
300
200
100
0
ABCA4
USH2A
RPGR
PRPH2
RHO
RP1
BEST1
RS1
CHM
PRPF31
CRB1
OPA1
TIMP3
MYO7A
CNGB3
GUCY2D
PROM1
RPE65
CNGA3
RP2
RDH12
EYS
CACNA1F
EFEMP1
BBS1
CRX
CEP290
PRPF8
RP9
NR2E3




A
Numbers of affected families by gene (pediatric cohort)
4 families for each of CLN3, CRX, FZD4, KCNV2, MYO7A, NYX, PNPLA6, RPGRIP1, TRPM1
60
50
40
30
20
10
0
ABCA4
RS1
BEST1
CACNA1F
RPGR
CNGA3
RPE65
CNGB3
CRB1
GUCY2D
NMNAT1
OPA1
CEP290
KIF11
AIPL1
CHM
RDH12
RHO
LCA5
PRPF31
CDH23
IQCB1
LRP5
NR2E3
PROM1
RP2
CLN3
CRX
FZD4
KCNV2
B
Numbers of affected individuals by gene (pediatric cohort)
4 individuals for each of CLN3, FZD4, KCNV2, MYO7A, NYX, PNPLA6, RPGRIP1
70
60
50
40
30
20
10
0
ABCA4
RS1
BEST1
RPGR
CACNA1F
CNGA3
RPE65
CNGB3
CRB1
GUCY2D
NMNAT1
OPA1
CEP290
CHM
KIF11
AIPL1
RDH12
RHO
LCA5
CDH23
LRP5
NR2E3
PROM1
RP2
PRPF31
IQCB1
CRX
TRPM1
CLN3
FZD4


A
Numbers of affected families by gene (current cohort)
600
500
400
300
200
100
0
ABCA4
USH2A
RPGR
PRPH2
BEST1
RS1
RP1
CHM
RHO
CRB1
PRPF31
OPA1
MYO7A
CNGB3
EYS
PROM1
CNGA3
RPE65
CACNA1F
GUCY2D
RDH12
BBS1
RP2
EFEMP1
CRX
NR2E3
PDE6B
TIMP3
CERKL
CNGB1
B
Numbers of affected individuals by gene (current cohort)
17 families for each of
CERKL, CNGB1, KCNV2
700
600
500
400
300
200
100
0
ABCA4
USH2A
RPGR
PRPH2
BEST1
RP1
RHO
RS1
CHM
CRB1
OPA1
PRPF31
CNGB3
CNGA3
PROM1
MYO7A
RPE65
EYS
GUCY2D
RP2
CACNA1F
RDH12
TIMP3
EFEMP1
BBS1
CRX
NR2E3
PRPF8
CEP290
PDE6B

Tables

Gene	Chromosomal location	Number of families affected (%)	Number of individuals affected (%)	Modes of inheritance	Range of phenotypes in literature
ABCA4	1p22.1	666 (20.8)	789 (18.6)	Recessive	Stargardt macular dystrophy; cone-rod dystrophy
USH2A	1q41	292 (9.1)	342 (8.1)	Recessive	RP; Type 2 Usher Syndrome
RPGR	Xp11.4	164 (5.1)	263 (6.2)	X-linked	RP; Cone or cone-rod dystrophy
PRPH2	6p21.1	148 (4.6)	220 (5.2)	Dominant and Recessive	Pattern dystrophy; RP
BEST1	11q12.3	125 (3.9)	168 (4.0)	Dominant and Recessive	Best disease; autosomal recessive bestrophinopathy
RS1	Xp22.13	111 (3.5)	134 (3.2)	X-linked	X-linked retinoschisis
RP1	8q12.1	106 (3.3)	170 (4.0)	Dominant and Recessive	RP
RHO	3q22.1	105 (3.3)	177 (4.2)	Dominant and Recessive	RP; stationary night blindness
CHM	Xq21.2	86 (2.7)	112 (2.6)	X-linked	Choroideremia
CRB1	1q31.3	68 (2.1)	86 (2.0)	Recessive	LCA; RP; macular dystrophy
PRPF31	19q13.42	57 (1.8)	94 (2.2)	Dominant	RP
MYO7A	11q13.5	53 (1.7)	58 (1.4)	Recessive	Type 1 Usher syndrome
OPA1	3q29	50 (1.6)	84 (2.0)	Dominant	Optic atrophy; optic atrophy with sensorineural hearing loss
CNGB3	8q21.3	44 (1.4)	55 (1.3)	Recessive	Achromatopsia; cone dystrophy
RPE65	1p31.2	39 (1.2)	51 (1.2)	Recessive and Dominant	LCA; RP
EYS	6q12	38 (1.2)	43 (1.0)	Recessive	RP
GUCY2D	17p13.1	37 (1.2)	54 (1.3)	Recessive and Dominant	LCA; RP; Cone or cone-rod dystrophy
PROM1	4p15.32	37 (1.2)	53 (1.2)	Recessive and Dominant	Macular dystrophy; cone-rod dystrophy; RP
CNGA3	2q11.2	36 (1.1)	50 (1.2)	Recessive	Achromatopsia; cone dystrophy
RDH12	14q24.1	35 (1.1)	44 (1.0)	Recessive and Dominant	LCA; RP

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.

Tables

Year published	Authors	Study cohort/ country	Number molecularly diagnosed (number of genes)	Most frequently implicated genes
Current study		United Kingdom	4241 individuals from 3197 families (135 genes)	By family: *ABCA4, USH2A, RPGR, PRPH2, BEST1* By individuals: *ABCA4, USH2A, RPGR, PRPH2, RHO*
2019	Khan AO[10]	United Arab Emirates (children)	71 individuals (26 genes)	*ABCA4, KCNV2, CRB1, CNGA3*
	Sharon et al.[11]	Israel	1369 families (129 genes)	*ABCA4, USH2A, FAM161A, CNGA3, EYS*
	Holtan et al.[12]	Norway	207 patients (56 genes)	*ABCA4, USH2A, BEST1, RHO, RS1*
	Avela et al.[13]	Finland (children)	41 families (17 genes)	*RS1, GUCY2D, RPGR*
	Kim et al.[14]	Korea	38 individuals (24 genes)	*ABCA4, EYS, PDE6B, USH2A, PDE6A, GUCY2D*
	Tayebi et al.[15]	Iran	36 families (19 genes)	*ABCA4, RPE65, CERKL, RPGRIP1*
2018	Motta et al.[16]	Brazil	400 individuals (66 genes)	*ABCA4, CEP290, USH2A, CRB1, RPGR*
	Wang et al.[17]	China	132 families (47 genes)	*USH2A, RPGR, CYP4V2, ABCA4, CRB1, RHO*
2017	Stone et al.[18]	United States	760 families (104 genes)	*ABCA4, USH2A, RPGR, RHO, PRPH2*
	Carss et al.[8]	United Kingdom	404 individuals (94 genes)	*ABCA4, USH2A, EYS, RP1, CACNA1F, RPGR*
	*Dockery et al.[19]	Ireland (adults)	357 families (59 genes)	*ABCA4, USH2A, BBS1, RHO, RP1*
	Ellingford et al.[20]	UK, genomic laboratory	271 individuals (62 genes)	*USH2A, CRB1, ABCA4, CERKL, CEP290*
	Haer-Wigman et al.[21]	Netherlands	136 individuals (56 genes)	*USH2A, EYS, ABCA4, RPGR, GUCY2D, PDE6B*
	Riera et al.[22]	Spain	42 individuals (29 genes)	*ABCA4, USH2A, PDE6A, CRB1, EYS, GUCY2D, PDE6B*
2016	Tiwari et al.[23]	Switzerland	58 individuals (18 genes)	*ABCA4, C2orf71, RP1, CEP290, FLVCR1, CRB1*
	Bernardis et al.[24]	Italy	52 individuals (16 genes)	*ABCA4, USH2A, RPGR, CNGB1, BEST1*

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]