BMJ Open Ophthalmology

Exploring the mutational landscape of genes associated with inherited retinal disease using large genomic datasets: identifying loss of function intolerance and outlying propensities for missense changes

Alexander Tanner ,[1,2] Hwei Wuen Chan ,[3,4] Elena Schiff,[3,5] Omar A Mahroo ,[3,6] Jose S Pulido[3,7]

To cite: Tanner A, Chan HW, Schiff E, *et al*. Exploring the mutational landscape of genes associated with inherited retinal disease using large genomic datasets: identifying loss of function intolerance and outlying propensities for missense changes. *BMJ Open Ophthalmology* 2022;**7**:e001079. doi:10.1136/bmjophth-2022-001079

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjophth-2022-001079).

Received 28 May 2022
Accepted 18 August 2022

For numbered affiliations see end of article.

Correspondence to
Dr Jose S Pulido; jpulido@willseye.org

ABSTRACT

Background Large databases permit quantitative description of genes in terms of intolerance to loss of function ('haploinsufficiency') and prevalence of missense variants. We explored these parameters in inherited retinal disease (IRD) genes.

Methods IRD genes (from the 'RetNet' resource) were classified by probability of loss of function intolerance (pLI) using online Genome Aggregation Database (gnomAD) and DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER) databases. Genes were identified having pLI ≥0.9 together with one or both of the following: upper bound of CI <0.35 for observed to expected (o/e) ratio of loss of function variants in the gnomAD resource; haploinsufficiency score <10 in the DECIPHER resource. IRD genes in which missense variants appeared under-represented or over-represented (Z score for o/e ratio of <−2.99 or >2.99, respectively) were also identified. The genes were evaluated in the gene ontology Protein Analysis THrough Evolutionary Relationships (PANTHER) resource.

Results Of 280 analysed genes, 39 (13.9%) were predicted loss of function intolerant. A greater proportion of X-linked than autosomal IRD genes fulfilled these criteria, as expected. Most autosomal genes were associated with dominant disease. PANTHER analysis showed >100 fold enrichment of spliceosome tri-snRNP complex assembly. Most encoded proteins were longer than the median length in the UniProt database. Fourteen genes (11 of which were in the 'haploinsufficient' group) showed under-representation of missense variants. Six genes (*SAMD11*, *ALMS1*, *WFS1*, *RP1L1*, *KCNV2*, *ADAMTS18*) showed over-representation of missense variants.

Conclusion A minority of IRD-associated genes appear to be 'haploinsufficient'. Over-representation of spliceosome pathways was observed. When interpreting genetic tests, variants found in genes with over-representation of missense variants should be interpreted with caution.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Large genomic datasets provide metrics for individual genes relating to under-representation of predicted loss of function variants and over-representation or under-representation of missense variants.

WHAT THIS STUDY ADDS

⇒ This study explores the above metrics for genes associated with inherited retinal disease: 39 were predicted 'loss of function intolerant' according to such metrics; 14 showed under-representation and 6 showed over-representation, of missense variants.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The findings can be compared with future studies of genes associated with other disorders. Also, when interpreting genetic tests, variants found in genes with over-representation of missense variants should be interpreted with caution.

INTRODUCTION

Inherited retinal diseases (IRDs) are a leading cause of blindness in children and the working age in many countries.[1–4] Variants in over 250 genes are implicated. There are a number of unresolved questions relating to the spectrum of variants and mechanisms of disease.[2] Some associated genes are ubiquitously expressed, yet pathogenic variants appear to give rise only to IRD.[5] A number of genes show mutational hotspots, while other regions exist that rarely harbour disease-causing variants, either because the regions are highly conserved or because polymorphisms rarely cause disease. Identifying genes, or genetic regions, with particular characteristics might shed light on particular selection pressures, and also help in future interpretation of novel variants.[6 7] The

range of genes and variants involved in IRDs has been recently reviewed comprehensively by Schneider *et al*,[8] who discussed, among other things, the prevalence of different types of variant, as well as their amenability to various gene-based therapeutic approaches.

Metrics are available from large genomic datasets which can identify those genes in which loss of function variants appear to be under-represented (conventionally termed 'haploinsufficient' genes).[9 10] These metrics are an indication of those genes in which heterozygosity for loss of function variants is selected against, presumably due to a survival or molecular disadvantage.[11] Genes can also be interrogated as to whether missense changes are significantly under-represented or over-represented. It is possible that variants that result in effects on vision, particularly if these are mild, or manifest late in life, will not have a strong effect on survival or reproductive success and so these metrics might not be affected. However, exploring these metrics for IRD genes might still yield insights into aspects of those genes in particular, potentially highlighting particularly conserved pathways, and could improve our understanding of the mutational landscape of IRD-associated genes more generally.

For this study, we curated a list of IRD genes (from the Retinal Information Network online resource, https://sph.uth.edu/retnet/), and investigated the above metrics in two large genomic databases, namely Genome Aggregation Database (gnomAD)[10] and DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER).[12] Both databases were used to identify genes with predicted 'loss of function intolerance', and the gnomAD resource was used additionally to identify those in which missense mutations were over-represented or under-represented. Genes of interest were evaluated in terms of associated pathways using the online gene ontology resource Protein Analysis THrough Evolutionary Relationships (PANTHER).[13]

The parameters investigated have been computed for each gene as a whole (based on the range of variants observed in the large datasets), rather than for any specific variants within the genes. Such parameters have been used, with some success, to identify candidate genes in whole genome data from patients with no molecular cause yet identified.[14] In the present study, we took a converse approach: we took genes already known to be associated with retinal disease, and interrogated which of these were, in the general population, found to have an under-representation of loss of function variants, and also which had an under-representation or over-representation of missense variants.

We were interested to observe any particular patterns that emerged, estimating the proportion of IRD-associated genes classified as having an under-representation of loss of function variants and whether particular modes of inheritance were more commonly seen in this group. Similar investigations have been performed for loss of function intolerant genes in general,[15] but our study focused in particular on IRD genes. We also explored whether such genes were more associated with syndromic disease, and whether certain pathways were over-represented. Identifying those genes with outlying propensities for missense variants could also be potentially useful: those IRD genes in which missense variants are over-represented may constitute 'noisy genes' such that missense variants in these genes, when found in patients, should be interpreted with caution.

MATERIALS AND METHODS

Databases and metrics

The gnomAD (https://gnomad.broadinstitute.org/) has over 141 456 individuals sequenced with 125 748 exomes and 15 708 genomes aligned against the Genome Reference Consortium Human genome build 37.[10] Constraint variables are computed for most genes. Probability of loss of function intolerance (pLI) refers to the probability that loss of function mutations are selected against. The ratio of observed variants to the number expected (by random chance) (o/e) is also computed along with a CI. A pLI of 0.9 or greater suggests a high level of intolerance to loss of function, and this is confirmed when the CI of o/e for loss of function variants is 0.35 or lower. The o/e for missense variants can also be explored: for this study Z scores of 2.99 or greater, or −2.99 or less, were taken to indicate a significant over-representation or under-representation of missense variants (a Z value of −2.99 means that the chance of variants occurring randomly with such low frequency in the population is only 0.14% (0.0014)).

DECIPHER (https://decipher.sanger.ac.uk/)[12] comprises genomic data from 36 000 children with rare diseases from over 270 specialist centres. Previously, a pLI separate to that from gnomAD was computed, but the pLI currently used is the gnomAD pLI. A haploinsufficiency score (HI) is also given where an index of less than 10% is taken to indicate that loss of function is significantly selected against.

Gene classification

Genes listed in the Retinal Information Network online resource (https://sph.uth.edu/retnet/) were included in this study. Those which met both of the following criteria were identified: (1) a pLI in gnomAD of ≥0.9, and (2) an upper CI o/e limit for loss of function variants in gnomAD of <0.35 and/or an HI in DECIPHER of <10. These genes were taken as likely to be intolerant to loss of function. These genes were then evaluated in the PANTHER[13] resource (http://pantherdb.org/) to identify common pathways in which the encoded proteins were involved, exploring which biological processes might be particularly over-represented in these gene groups (using the over-representation analysis).[16] Also, genes with gnomAD missense (non-synonymous variants) o/e Z scores of −2.99 or less, or of 2.99 or greater, were identified and analysed similarly. The gene list curation from RetNet and investigation of metrics in gnomAD and DECIPHER were performed in September 2021;

the evaluation in PANTHER was performed in February 2022.

RESULTS

Intolerance of loss of function analysis

Of 309 genes and loci listed in the Retinal Information Network online resource, 29 were excluded (owing to one of the following: a mitochondrial location, no specific gene yet identified for the locus or lack of relevant data available on gnomAD and DECIPHER), leaving 280 available for inclusion (including 262 autosomal and 18 X-linked genes; online supplemental table 1). Of these, 39 genes (13.9%) met the specified criteria for loss of function intolerance. Of the IRD genes with pLI ≥0.9, there were no additional genes identified in the DECIPHER resource with an HI <10 that did not also have a gnomAD upper CI o/e limit for loss of function variants of <0.35. The 39 genes are listed in table 1. Of note, 8 of these genes are X-linked. Thus, the proportion of X-linked IRD genes fulfilling criteria for intolerance to loss of function (8/18) was 44.4% while the corresponding proportion of autosomal IRD genes (31/262) was significantly lower at 11.8% (p<0.0001 for difference in proportions). The majority of the 31 autosomal genes listed in table 1 are associated with dominantly inherited disorders. Only three genes are associated exclusively with recessively inherited disease: of note, implication of each of these genes with retinal disease has been established only by a single report, suggesting the association may not be secure. The final column of table 1 contains comments on the strength of evidence for association with monogenic retinal disease; these will be mentioned in the Discussion section.

The 39 genes identified as showing intolerance to loss of function were then evaluated for over-representation in biological processes using the PANTHER database. The analysis showed that the following process was enriched more than 100-fold: spliceosome tri-snRNP complex assembly ($p=3.44\times10^{-6}$, false discovery rate (FDR) of 0.0049). Other processes showing significant over-representation included visual perception, eye morphogenesis, cellular protein localisation and nervous system development (online supplemental table 2) give these results in detail).

Exploration of size of encoded proteins

It has been noted that haploinsufficient genes are significantly longer than haplosufficient genes.[17] From 20 394 reviewed genes on the Uniprot database (https://www.uniprot.org/),[18] we found the median protein length was 415 amino acids. Only five of the 39 genes in table 1 encode a protein with fewer than 415 amino acids, and four of these (*RS1*, *PRPS1*, *RP2*, *OPN1LW*) are X-linked. Only one gene (*OTX2*) from table 1 was both autosomal and associated with a protein with fewer than 415 amino acids.

Analysis by frequencies of missense mutations

The 280 genes were also classified by observed/expected frequencies of non-synonymous missense variants. Fourteen genes (5.0%) were identified in which such variants appeared to be negatively selected (under-represented). These are given in table 2. Eleven of the 14 genes had already been classified as intolerant to loss of function (table 1); only *KLHL7*, *PNPLA6* and *PRPF6* are not present in table 1. As in table 1, the majority of genes in table 2 are associated with dominantly inherited disorders. PANTHER analysis revealed >100 fold enrichment of spliceosome tri-snRNP complex assembly process ($p=3.00\times10^{-10}$, FDR of 4.76×10^{-6}).

Finally, those genes showing significantly more missense variants than expected were identified. Six genes (2.1%) met this criterion, listed in table 3. Most have long exons, and *SAMD11* has relatively poor coverage on exome sequencing analyses. All of these genes were associated with recessively inherited diseases (exclusively autosomal recessive inheritance for 5 of the 6 genes). The PANTHER pathway analysis did not find particular biological processes enriched in this small number of genes.

DISCUSSION

In this study, we explored a novel classification of retinal disease-associated genes according to metrics relating to predicted tolerance to loss of function (as defined in two large genomic databases) and to under-representation or over-representation of missense variants (as computed in the gnomAD resource). We also sought to identify any broad biological pathways that were enriched in any of these groups.

We found that approximately 14% of IRD genes (as listed in the RETnet resource) overall were predicted to be intolerant to loss of function. The proportion for X-linked genes was significantly higher than that for autosomal genes. This might be expected as the mechanism of disease for many X-linked conditions, including X-linked retinal disease is frequently loss of function.[19] For the autosomal genes, almost all were associated with dominantly inherited conditions (and those where exclusively recessive inheritance has been described have less strong evidence for association with retinal disease). Our finding of higher proportions of autosomal dominant and X-linked (and low proportion of autosomal recessive) Mendelian diseases associated with genes with high pLI is consistent with a previous study (not focusing on retinal disease) where these genes were compared with a random sample of other genes.[15]

A number of genes encoding proteins involved in splicing complexes (*PRPF3*, *SNRNP200*, *PRPF4*, *PRPF8*, *PRPF31*)[20] were found to fulfil criteria for intolerance to loss of function and are all associated with dominantly inherited disease. Of these, *PRPF31* is known to be associated with disease resulting from haploinsufficiency. Why pathogenic variants in such ubiquitously expressed genes give only a rod-cone dystrophy is still not clear: rod photoreceptors appear uniquely vulnerable to loss of function in one *PRPF31* allele. In contrast to haploinsufficiency, gain of function, often from specific missense variants, is

Table 1 The 39 genes which fulfilled criteria of intolerance to loss of function in one or both databases

Gene	Location	Mode of inheritance of disorders	Reported phenotypes	Comments on strength of association with retinal disease
COL11A1	1p21.1	Dominant	Dominant Stickler syndrome type II; dominant Marshall syndrome	
MFN2	1p36.22	Dominant	Dominant optic atrophy with neuropathy and myopathy; dominant Charcot-Marie-Tooth disease	
PRPF3	1q21.2	Dominant	Dominant retinitis pigmentosa	
EFEMP1	2p16.1	Dominant	Dominant drusen	
SNRNP200	2q11.2	Dominant	Dominant retinitis pigmentosa	
ATXN7	3p14.1	Dominant	Dominant spinocerebellar atrophy with macular dystrophy or retinal degeneration	
OPA1	3q29	Dominant	Dominant optic atrophy; dominant optic atrophy with sensorineural hearing loss	
VCAN	5q14.3	Dominant	Dominant Wagner disease and erosive vitreoretinopathy	
NR2F1	5q15	Dominant	Dominant optic atrophy with intellectual disability and developmental delay (Bosch-Boonstra optic atrophy)	
CTNNA1	5q31.2	Dominant	Dominant macular pattern dystrophy (butterfly-shaped pigment dystrophy)	
RIMS1	6q13	Dominant	Dominant cone-rod dystrophy	One family reported in detail and a single patient in a second report with a different phenotype. (Other variants reported, but lacking detailed information)
AHR	7p21.1	Recessive	Recessive retinitis pigmentosa	One family reported
KIAA1549	7q34	Recessive	Recessive retinitis pigmentosa	One report of two families
GDF6	8q22.1	Dominant and recessive	Recessive Leber congenital amaurosis; dominant Klippel-Feil syndrome; dominant microphthalmia	Single report of Leber congenital amaurosis
TOPORS	9q21.1	Dominant	Dominant retinitis pigmentosa	
PRPF4	9q32	Dominant	Dominant retinitis pigmentosa	
HK1	10q22.1	Dominant and recessive	Dominant retinitis pigmentosa; recessive nonspherocytic haemolytic anaemia; recessive hereditary neuropathy	
KIF11	10q23.33	Dominant	Dominant microcephaly, lymphedema and chorioretinopathy	
TEAD1	11p15.3	Dominant	Dominant atrophia areata/Sveinsson peripapillary degeneration	
FZD4	11q14.2	Dominant	Dominant FEVR	
COL2A1	12q13.11	Dominant	Dominant Stickler syndrome, type I; dominant bone dysplasias, developmental disorders, osteoarthritic diseases, syndromic disorders	
CCT2	12q15	Recessive	Recessive Leber congenital amaurosis	One report
RB1	13q14.2	Dominant	Dominant (or somatic) retinoblastoma; pinealoma; osteogenic sarcoma	
OTX2	14q22.3	Dominant	Dominant syndromic microphthalmia; combined pituitary deficiency 6; early onset retinal dystrophy and pattern dystrophy	

Continued

Table 1 Continued

Gene	Location	Mode of inheritance of disorders	Reported phenotypes	Comments on strength of association with retinal disease
FBLN5	14q32.12	Dominant	Dominant familial age-related macular degeneration; hereditary neuropathy with or without age-related macular degeneration	
ZNF423	16q12.1	Dominant and Recessive	Recessive nephronophthisis; dominant Joubert syndrome	One report including two families with Joubert syndrome
PITPNM3	17p13.2	Dominant	Dominant cone-rod dystrophy	One report of two families
PRPF8	17p13.3	Dominant	Dominant retinitis pigmentosa	
C3	19p13.3	Dominant and recessive	Dominant susceptibility to atypical haemolytic-uraemic syndrome 5; recessive CS deficiency; polymorphisms confer risk for AMD	Association with AMD, but not proven to cause monogenic retinal disease
PRPF31	19q13.42	Dominant	Dominant retinitis pigmentosa	
JAG1	20p12.2	Dominant	Dominant Alagille syndrome	
RP2	Xp11.23	X-linked	Retinitis pigmentosa	
RPGR	Xp11.4	X-linked	Retinitis pigmentosa; cone-rod dystrophy; macular dystrophy	
DMD	Xp21.2-p21.1	X-linked	Duchenne muscular dystrophy	Electroretinogram may be abnormal
RS1	Xp22.13	X-linked	X-linked retinoschisis	
OFD1	Xp22.2	X-linked	Joubert syndrome; orofaciodigital syndrome 1, Simpson-Golabi-Behmel syndrome 2	
CHM	Xq21.2	X-linked	Choroideremia	
PRPS1	Xq22.3	X-linked	Retinitis pigmentosa, neuropathy, optic atrophy, deafness	
OPN1LW	Xq28	X-linked	Deuteranopia; blue-cone monochromacy	

A number are associated also with syndromic or non-retinal disorders. The rightmost column contains comments, for some of the genes, on the strength of association with retinal disease (including highlighting those genes where there have been only single reports). These will be considered further in the Discussion section.

an important mechanism in much of dominant disease. The appearance of genes in which gain of function causes retinal disease in table 1 might suggest that loss of function adversely affects survival in a way other than by affecting the retina; heterozygosity for loss of function might have severe consequences for other systems. Many, but not all, of the genes in table 1 are also associated with syndromic or non-retinal disease. We also found that the majority of IRD genes identified as intolerant to loss of function encoded proteins above the median length in terms of amino acids.

Eight of the 31 autosomal genes have associations with monogenic retinal disease that arise from reports of relatively few families, suggesting less strong evidence for causative association. The first implication of *RIMS1* in retinal disease came from a report of a single large family.[21 22] A later report described a patient with the same *RIMS1* variant, but a different phenotype (retinitis pigmentosa),[23] and there have further variants reported, but without detailed evidence to secure a causative relationship.[24–26] Implication of *AHR* in retinal disease comes from a report of a single family.[27] For *KIAA1549*, there is a single report of two families.[28] For *GDF6*, there is a single report of a patient with Leber Congenital Amaurosis, and in this case, both parents showed electroretinography (ERG) abnormalities.[29] Similarly, for *CCT2*, there is a single report of Leber Congenital Amaurosis.[30] With respect to *ZNF423*, one study reported a patient with recessive nephronophthisis (no ocular data) and two families with Joubert syndrome.[31] For *PITPNM3*, one study reported two families[32]; a subsequent study showed the carrier frequency of the variant in the general population was high for a dominant disease.[33] Other reports of disease associated with this gene are either isolated cases or report variants that also have high carrier frequencies.[34–37] A critical analysis of reported variants in autosomal dominant retinal dystrophies has questioned the evidence for disease association with *PITPNM3* (and also with *RIMS1*), among other genes.[38] Finally, while the *C3* variants are associated with AMD, evidence of specific

Table 2 Genes which fulfilled criteria of negative selection for missense variants

Gene	Location	Mode of inheritance of disorders	Phenotypes
PRPF3	1q21.2	Dominant	Dominant retinitis pigmentosa
SNRNP200	2q11.2	Dominant	Dominant retinitis pigmentosa
NR2F1	5q15	Dominant	Dominant optic atrophy with intellectual disability and developmental delay (Bosch-Boonstra optic atrophy)
CTNNA1	5q31.2	Dominant	Dominant macular pattern dystrophy (butterfly-shaped pigment dystrophy)
KLHL7	7p15.3	Dominant	Dominant retinitis pigmentosa
HK1	10q22.1	Dominant and recessive	Dominant retinitis pigmentosa; recessive nonspherocytic haemolytic anaemia; recessive hereditary neuropathy
KIF11	10q23.33	Dominant	Dominant microcephaly, lymphedema and chorioretinopathy
COL2A1	12q13.11	Dominant	Dominant Stickler syndrome, type I; dominant bone dysplasias, developmental disorders, osteoarthritic diseases, syndromic disorders
PRPF8	17p13.3	Dominant	Dominant retinitis pigmentosa
PNPLA6	19p13.2	Recessive	Boucher-Neuhauser syndrome with chorioretinal dystrophy
PRPF31	19q13.42	Dominant	Dominant retinitis pigmentosa
JAG1	20p12.2	Dominant	Dominant Alagille syndrome
PRPF6	20q13.33	Dominant	Dominant retinitis pigmentosa
PRPS1	Xq22.3	X-linked	Retinitis pigmentosa, neuropathy, optic atrophy, deafness

association with monogenic retinal disease is lacking. If these eight genes are excluded from the IRD list, the proportion of autosomal IRD genes meeting loss of function intolerance criteria is then 9.1%.

Interestingly, the only genes in table 1 associated with exclusively recessive inheritance are among those with evidence only from single reports. This might suggest that in cases where a novel genetic cause of recessive IRD is reported, with bi-allelic loss of function proposed as the disease mechanism, if that gene shows loss of function intolerance according to the criteria of the present study, such a report should be interpreted with caution.

We found that 5% of IRD genes showed significant under-representation of non-synonymous missense variants. The majority of these were also intolerant to loss of function, highlighting their importance in both terms. They were again mostly associated with dominantly inherited disorders, and a further gene encoding a splicing factor (*PRPF6*) emerged. Given that missense variants are relatively rare in these genes, any such novel variants found in patients with a consistent phenotype might be regarded as more likely to be pathogenic rather than incidental.

Only 2% of IRD genes met the criterion of significant over-representation of missense variants. These were all associated with autosomal recessive inheritance. This can make identifying the pathogenicity of novel missense variants in these genes more challenging. Fortunately, in *KCNV2*-associated retinopathy, electroretinography is pathognomonic for disease associated with this gene, facilitating judgements of pathogenicity of rare variants.[39] On the other hand, variants in *RP1L1* give rise to a dominantly inherited occult maculopathy or a recessively inherited rod-cone dystrophy. The phenotypic

Table 3 Genes which fulfilled criteria of over-representation of missense variants

Gene	Location	Mode of inheritance of disorders	Phenotypes
SAMD11	1p36.33	Recessive	Recessive retinitis pigmentosa
ALMS1	2p13.1	Recessive	Alstrom syndrome
WFS1	4p16.1	Recessive	Recessive Wolfram syndrome (also autosomal dominant low frequency sensorineural hearing loss)
RP1L1	8p23.1	Dominant and Recessive	Dominant occult macular dystrophy; recessive retinitis pigmentosa
KCNV2	9p24.2	Recessive	Cone dystrophy with supernormal rod response
ADAMTS18	16q23.1	Recessive	Knobloch syndrome; recessive early onset retinal dystrophy

features are not unique to this gene; the identification of this gene as one of those in which missense variants are over-represented is consistent with the reported highly polymorphic nature of *RP1L1*, and further supports the notion that novel variants, particularly those outside the two known hotspots for pathogenic variants, should be interpreted with great caution in terms of potential pathogenicity.[40]

We found that biological pathways relating to spliceosome complex assembly were enriched (>100 fold) in the group of retinal disease genes predicted to be loss of function intolerant. The spliceosome complex assembly pathway was also enriched in the group of genes with significant under-representation of missense variants. The importance of the splicing pathway is thus emphasised, together with the above-mentioned questions as to why only rod photoreceptors appear affected by heterozygous pathogenic variants.

A number of limitations of our study deserve mention. There can be pitfalls of relying on pLI indicators as described by other authors[41]; genes that do not meet the pLI criteria frequently may encode essential proteins where loss of function in one allele may cause dominant disease. The thresholds (criteria) used are somewhat arbitrary, but there was significant overlap between predictions from the gnomAD and DECIPHER databases. The reliance on only the PANTHER resource to investigate which pathways were over-represented might also represent a limitation. We therefore checked that similar results would be obtained from other resources. Entering the list of 'haploinsufficient' genes (from table 1) into the Reactome[42] resource (https://reactome.org/ accessed 30 July 2022) similarly showed that mRNA splicing was most significantly over-represented (p=0.0002). We also entered these genes into the Molecular Signatures Database[43 44] resource (V.7.5.1 available at http://www.gsea-msigdb.org/gsea/msigdb/index.jsp accessed 30 July 2022) to compute overlaps with other gene sets (selecting the Gene Ontology sets): the most significant overlap (after sensory perception gene sets, not unexpected for a group of IRD genes) was the 'SPLICEOSOMAL_TRI_SNRNP_COMPLEX' gene set ($p=5.74\times10^{-10}$). These findings support the validity of our results from the PANTHER analysis.

Our study is mainly exploratory, and many of the conclusions tentative. These metrics have been proposed to help identify candidate genes for unsolved diseases, whereas we have, conversely, applied these metrics to known disease-associated genes, with a view to further exploring the variant landscape of these genes.

Author affiliations
[1]Ophthalmology, University College London Institute of Ophthalmology, London, UK
[2]Ophthalmology, Moorfields Eye Hospital City Road Campus, London, UK
[3]University College London Institute of Ophthalmology, London, UK
[4]Department of Ophthalmology, University of Singapore, Singapore
[5]Moorfields Eye Hospital City Road Campus, London, UK
[6]Medical Retina Service, Moorfields Eye Hospital NHS Foundation Trust, London, UK
[7]Wills Eye Hospital, Philadelphia, Pennsylvania, USA

Correction notice This article has been corrected since it first published. Author name 'Omar A Mahroo' has been updated.

Acknowledgements This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk.

Contributors All authors contributed to data collection, analysis and interpretation. AT, OAM, JSP contributed to the drafting of the manuscript. All authors contributed to the revision of manuscript and approval of final version. JSP, OAM contributed to supervision. Guarantor - JSP.

Funding OAM is funded by the Wellcome Trust (206619/Z/17/Z).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Alexander Tanner http://orcid.org/0000-0002-7194-0978
Hwei Wuen Chan http://orcid.org/0000-0001-8079-8398
Omar A Mahroo http://orcid.org/0000-0003-1254-0832

REFERENCES

1 Cremers FPM, Boon CJF, Bujakowska K, *et al*. Special issue introduction: inherited retinal disease: novel candidate genes, genotype-phenotype correlations, and inheritance models. *Genes* 2018;9:215.
2 Duncan JL, Pierce EA, Laster AM, *et al*. Inherited retinal degenerations: current landscape and knowledge gaps. *Transl Vis Sci Technol* 2018;7:6.
3 Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. *BMJ Open* 2014;4:e004015.
4 Solebo AL, Teoh L, Rahi J. Epidemiology of blindness in children. *Arch Dis Child* 2017;102:853–7.
5 Wright AF, Chakarova CF, Abd El-Aziz MM, *et al*. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat Rev Genet* 2010;11:273–84.
6 Farrar GJ, Carrigan M, Dockery A, *et al*. Toward an elucidation of the molecular genetics of inherited retinal degenerations. *Hum Mol Genet* 2017;26:R2–11.
7 Ellingford JM, Barton S, Bhaskar S, *et al*. Molecular findings from 537 individuals with inherited retinal disease. *J Med Genet* 2016;53:761–7.
8 Schneider N, Sundaresan Y, Gopalakrishnan P, *et al*. Inherited retinal diseases: linking genes, disease-causing variants, and relevant therapeutic modalities. *Prog Retin Eye Res* 2022;89:101029.
9 Lek M, Karczewski KJ, Minikel EV, *et al*. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
10 Karczewski KJ, Francioli LC, Tiao G. *Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes*, 2019.

11 Fuller ZL, Berg JJ, Mostafavi H, *et al*. Measuring intolerance to mutation in human genetics. *Nat Genet* 2019;51:772–6.
12 Firth HV, Richards SM, Bevan AP, *et al*. Decipher: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am J Hum Genet* 2009;84:524–33.
13 Thomas PD, Campbell MJ, Kejariwal A, *et al*. Panther: a library of protein families and subfamilies indexed by function. *Genome Res* 2003;13:2129–41.
14 Seaby EG, Smedley D, Taylor Tavares AL, *et al*. Genomics England research Consortium. A gene-to-patient approach uplifts novel disease gene discovery and identifies 18 putative novel disease genes. *Genet Med* 2022;S1098-3600:00748–1.
15 Fabre A, Mancini J. No preferential mode of inheritance for highly constrained genes. *Intractable Rare Dis Res* 2022;11:25–8.
16 Mi H, Muruganujan A, Huang X, *et al*. Protocol update for large-scale genome and gene function analysis with the panther classification system (v.14.0). *Nat Protoc* 2019;14:703–21.
17 Huang N, Lee I, Marcotte EM, *et al*. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 2010;6:e1001154.
18 UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47:D506–15.
19 De Silva SR, Arno G, Robson AG, *et al*. The X-linked retinopathies: physiological insights, pathogenic mechanisms, phenotypic features and novel therapies. *Prog Retin Eye Res* 2021;82:100898.
20 Tanackovic G, Ransijn A, Thibault P, *et al*. PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Hum Mol Genet* 2011;20:2116–30.
21 Johnson S, Halford S, Morris AG, *et al*. Genomic organisation and alternative splicing of human Rim1, a gene implicated in autosomal dominant cone-rod dystrophy (CORD7). *Genomics* 2003;81:304–14.
22 Michaelides M, Holder GE, Hunt DM, *et al*. A detailed study of the phenotype of an autosomal dominant cone-rod dystrophy (CORD7) associated with mutation in the gene for Rim1. *Br J Ophthalmol* 2005;89:198–206.
23 Warwick AN, Shawkat F, Lotery AJ. Retinitis pigmentosa and bilateral cystoid macular oedema in a patient heterozygous for the Rim1 mutation previously associated with cone-rod dystrophy 7. *Ophthalmic Genet* 2017;38:178–82.
24 Glöckle N, Kohl S, Mohr J, *et al*. Panel-Based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *Eur J Hum Genet* 2014;22:99–104.
25 Seong M-W, Seo SH, Yu YS, *et al*. Diagnostic application of an extensive gene panel for Leber congenital amaurosis with severe genetic heterogeneity. *J Mol Diagn* 2015;17:100–5.
26 Wang X, Feng Y, Li J, *et al*. Retinal diseases caused by mutations in genes not specifically associated with the clinical diagnosis. *PLoS One* 2016;11:e0165405.
27 Zhou Y, Li S, Huang L, *et al*. A splicing mutation in aryl hydrocarbon receptor associated with retinitis pigmentosa. *Hum Mol Genet* 2018;27:2563–72.
28 de Bruijn SE, Verbakel SK, de Vrieze E, *et al*. Homozygous variants in *KIAA1549*, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa. *J Med Genet* 2018;55:705–12.
29 Asai-Coakwell M, March L, Dai XH, *et al*. Contribution of growth differentiation factor 6-dependent cell survival to early-onset retinal dystrophies. *Hum Mol Genet* 2013;22:1432–42.
30 Minegishi Y, Sheng X, Yoshitake K, *et al*. CCT2 mutations evoke Leber congenital amaurosis due to chaperone complex instability. *Sci Rep* 2016;6:33742.
31 Chaki M, Airik R, Ghosh AK, *et al*. Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling. *Cell* 2012;150:533–48.
32 Köhn L, Kadzhaev K, Burstedt MSI, *et al*. Mutation in the PYK2-binding domain of PITPNM3 causes autosomal dominant cone dystrophy (CORD5) in two Swedish families. *Eur J Hum Genet* 2007;15:664–71.
33 Berg JS, Adams M, Nassar N, *et al*. An informatics approach to analyzing the incidentalome. *Genet Med* 2013;15:36–44.
34 Köhn L, Kohl S, Bowne SJ, *et al*. Pitpnm3 is an uncommon cause of cone and cone-rod dystrophies. *Ophthalmic Genet* 2010;31:139–40.
35 Huang X-F, Huang F, Wu K-C, *et al*. Genotype-Phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing. *Genet Med* 2015;17:271–8.
36 Neveling K, Feenstra I, Gilissen C, *et al*. A post-hoc comparison of the utility of Sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat* 2013;34:1721–6.
37 Bakhoum MF, Sengillo JD, Cui X, *et al*. Autoimmune retinopathy in a patient with a missense mutation in PITPNM3. *Retin Cases Brief Rep* 2018;12 Suppl 1:S72–5.
38 Hanany M, Sharon D. Allele frequency analysis of variants reported to cause autosomal dominant inherited retinal diseases question the involvement of 19% of genes and 10% of reported pathogenic variants. *J Med Genet* 2019;56:536–42.
39 Vincent A, Robson AG, Holder GE. Pathognomonic (diagnostic) ERGs a review and update. *Retina* 2013;33:5–12.
40 Bowne SJ, Daiger SP, Malone KA, *et al*. Characterization of RP1L1, a highly polymorphic paralog of the retinitis pigmentosa 1 (RP1) gene. *Mol Vis* 2003;9:129–37.
41 Ziegler A, Colin E, Goudenège D, *et al*. A snapshot of some pLI score pitfalls. *Hum Mutat* 2019;40:839–41.
42 Gillespie M, Jassal B, Stephan R, *et al*. The reactome pathway knowledgebase 2022. *Nucleic Acids Res* 2022;50:D687–92.
43 Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
44 Liberzon A. A description of the molecular signatures database (MSigDB) web site. *Methods Mol Biol* 2014;1150:153–60.

Supplementary Table 1. Parameters obtained from online gnomAD and DECIPHER databases relating to IRD-associated genes listed in the RetNet online resource.

Gene	Chromosomal location	gnomAD pLI	gnomAD plof o/e	gnomAD plof o/e CI	DECIPHER HI	gnomAD missense Z score
ABCA4	1p22.1	0	0.76	0.64-0.91	18.63	-0.66
ABCC6	16p13.11	0	0.83	0.67-1.03	72.89	-1.09
ABHD12	20p11.21	0	0.47	0.3-0.76	52.8	0.19
ACBD5	10p12.1	0.01	0.31	0.19-0.53	47.98	0.87
ACO2	22q13.2	0.21	0.24	0.14-0.42	9.79	2.92
ADAM9	8q11.23	0.03	0.26	0.17-0.42	58.29	1.9
ADAMTS18	16q23.1	0	0.74	0.6-0.93	48.74	-3.51
ADGRV1	5q14.3	0	0.44	0.38-0.52	25.57	0.07
ADIPOR1	1q32.1	0.66	0.18	0.08-0.47	11.69	2.82
AFG3L2	18p11.21	0	0.49	0.34-0.7	30.83	1.99
AGBL5	2p23.3	0	0.47	0.33-0.68	29.25	1.29
AHI1	6q23.3	0	0.75	0.59-0.95	11.12	-0.03
AHR	7p21.1	1	0.05	0.02-0.15	33.82	0.73
AIPL1	17q13.2	0	0.63	0.39-1.07	58.68	0.08
ALMS1	2p13.1	0	0.75	0.64-0.87	78.65	-2.99
ARHGEF18	19p13.3	0	0.3	0.2-0.46	73.49	2.25
ARL2BP	16q13.3	0	0.75	0.44-1.35	36.61	0.83
ARL3	10q24.32	0	0.56	0.3-1.1	12.85	1.16
ARL6	3q11.2	0.01	0.44	0.23-0.92	10.16	0.56
ARMS2	10q26.13	0	1.66	0.68-1.94	97.47	0.68
ARSG	17q24.2	0	0.76	0.52-1.15	66.13	0.89
ASRGL1	11q12.3	0	0.77	0.45-1.39	74.89	0.2
ATF6	1q23.3	0	0.47	0.32-0.7	26.13	0.56
ATXN7	3p14.1	0.96	0.16	0.08-0.33	17.71	-0.62
BBIP1	10q25.2	0	0.72	0.35-1.57	22.73	0.52
BBS1	11q13.5	0	0.69	0.5-0.98	40.7	0.16
BBS10	12q21.2	0	0.93	0.64-1.36	59.88	0.09
BBS12	4q27	0	0.62	0.41-0.96	80.48	0.34
BBS2	16q13	0	0.79	0.59-1.07	15.37	0.87
BBS4	15q24.1	0	0.86	0.63-1.18	25.15	-0.96
BBS5	2q31.1	0	0.55	0.35-0.89	9.96	0.85
BBS7	4q27	0	0.51	0.35-0.73	20.8	1.23
BBS9	7p14.3	0	0.63	0.48-0.84	17.13	0.4
BEST1	11q12.3	0	1.01	0.69-1.49	51.6	0.62
C12ORF65	12q24.31	0.21	0.29	0.12-0.91	75.97	-0.43
C1QTNF5	11q23.3	0.84	0	0-0.5	NA	1.16
C2	6p21.32	0	0.76	0.56-1.03	52.58	1.12
C21ORF2	21q22.3	0	0.5	0.53-1.4	73.17	-0.16
C2ORF71	2p23.2	0	0.82	0.61-1.11	84.17	-2.14
C3	19p13.3	0.9	0.21	0.14-0.3	66.19	2.75
C8orf37	8q22.1	0	0.79	0.46-1.42	54.96	-0.07
CA4	17q23.2	0	0.68	0.42-1.15	82.9	-0.12
CABP4	11q13.1	0	1	0.67-1.54	71.95	-0.52
CACNA1F	Xp11.23	0	0.32	0.23-0.45	39.96	2.6
CACNA2D4	12p13.33	0	0.82	0.66-1.01	65.4	0.75
CAPN5	11q13.5	0	0.53	0.36-0.78	56.23	0.53
CC2D2A	4p15.33	0	0.63	0.51-0.78	79.39	0.65

CCT2	12q15	1	0.07	0.03-0.23	4.82	1.47
CDH23	10q22.1	0	0.38	0.26-0.57	15.82	0.71
CDH3	16q22.1	0	0.59	0.42-0.85	56.91	0.08
CDHR1	10q23.1	0	0.98	0.74-1.31	54.73	-0.7
CEP164	11q23.3	0	0.67	0.54-0.84	29.41	0.35
CEP19	3q29	0.02	0.48	0.23-1.09	39.64	-0.14
CEP250	20q11.22	0	0.54	0.44-0.65	50.61	1.33
CEP290	12q21.32	0	0.84	0.71-0.98	13.34	0.47
CEP78	9q21.2	0	0.71	0.51-1.02	49.15	-0.58
CERKL	2q31.3	0	1.16	0.88-1.54	41.25	-0.27
CFB	6p21.32	0	0.33	0.22-0.52	56.27	1.48
CFH	1q31.3	0.86	0.2	0.13-0.32	73.44	1
CHM	Xq21.2	1	0.04	0.01-0.19	25.27	0.79
CIB2	15q25.1	0	0.55	0.3-1.09	53.95	0.52
CLCC1	1p13.3	0.01	0.33	0.2-0.57	69.69	1.1
CLN1/ PPT1	1p34.2	0	0.52	0.32-0.89	21.97	0.03
CLN3	16p11.2	0	0.59	0.39-0.91	55	-0.15
CLRN1	3q25.1	0	1.17	0.73-1.8	42.64	-0.59
CLUAP1	16p13.3	0	0.4	0.25-0.68	44.51	0.04
CNGA1	4p12	0	0.65	0.46-0.93	52.11	0.44
CNGA3	2q11.2	0	1.08	0.8-1.49	61.65	-0.04
CNGB1	16q21	0	0.93	0.75-1.15	68.86	-0.83
CNGB3	8q21.3	0	0.76	0.58-1.02	57.91	1.18
CNNM4	2q11.2	0	0.6	0.4-0.94	64.14	2.37
COD2	Xq27	0	0.75	0.56-1.02	27.36	-0.18
COL11A1	1p21.1	1	0.14	0.1-0.22	8.56	1.02
COL2A1	12q13.11	1	0.07	0.04-0.13	2.04	3.29
COL9A1	6q13	0	0.63	0.48-0.81	23.88	0.19
CORD1	18q21.1-q21.3	NA	NA	NA	NA	NA
CORD4	17q	NA	NA	NA	NA	NA
CORD8	1q23.1-q23.3	NA	NA	NA	NA	NA
CRB1	1q31.3	0	0.62	0.47-0.83	64.17	-1.28
CRX	19q13.32	0.51	0.19	0.08-0.6	56.39	0.33
CSPP1	8q13.1-q13.2	0	0.74	0.6-0.93	45.84	0.65
CTNNA1	5q31.2	0.97	0.17	0.1-0.31	2.28	3.66
CWC27	5q12.3	0	0.57	0.38-0.9	17.87	0.76
CYP4V2	4q35.2	0	0.7	0.49-1.03	70.45	-0.12
DHDDS	1p36.11	0.25	0.24	0.13-0.51	5.63	1.09
DHX38	16q22.2	0	0.45	0.34-0.61	31.05	2.67
DMD	Xp21.2-p21.1	1	0.1	0.07-0.15	0.26	-2.43
DRAM2	1p13.3	0	0.6	0.35-1.08	29.96	0.45
DTHD1	4p14	0	0.54	0.36-0.83	67.05	0.99
DYNC2I2	9q34.11	0	0.64	0.42-0.99	NA	-0.12
DYNC2H1	11q22.3	0	0.49	0.42-0.58	29.65	0.91
EFEMP1	2p16.1	1	0.03	0.01-0.15	11.29	1.82
ELOVL1	1p34.2	0.68	0.18	0.08-0.46	24.31	1.81
ELOVL4	6q14.1	0.83	0.16	0.07-0.41	28.93	1.19
EMC1	1p36.13	0	0.79	0.62-1.01	33.43	1.35
ERCC6	10q11.23	0	0.63	0.49-0.8	56.19	0.1
ESPN	1p36.31	0	0.63	0.43-0.93	46.08	-0.08
EXOSC2	9q34.12	0	0.66	0.42-1.1	18.74	0.39

EYS	6q12	0	0.69	0.58-0.83	25.63	0.32
FAM161A	2p15	0	0.72	0.52-1.01	75.75	-0.49
FBLN5	14q32.12	1	0.04	0.01-0.17	21.01	1.56
FLVCR1	1q32.3	0	0.38	0.23-0.67	50.23	0.85
FSCN2	17q25.3	0	1.01	0.7-1.48	55.98	-0.11
FZD4	11q14.2	0.97	0.07	0.02-0.31	22.13	0.73
GDF6	8q22.1	0.99	0	0-0.22	17.17	0.93
GNAT1	3p21.31	0	1.05	0.72-1.55	17.63	0.72
GNAT2	1p13.3	0	0.59	0.37-0.98	23.08	0.66
GNB3	12p13.31	0	0.92	0.63-1.38	28.89	1.05
GNPTG	16p13.3	0	0.79	0.53-1.22	80.99	-2.06
GPR125/ADGRA3	4p15.2	0.29	0.23	0.15-0.36	42.05	0.45
GPR179	17q12	0	0.66	0.53-0.82	79.16	0.84
GRK1	13q34	0	0.49	0.31-0.82	53.1	0.43
GRM6	5q35.3	0	0.85	0.63-1.17	55.29	-0.51
GUCA1A	6p21.1	0	0.72	0.4-1.42	40.08	0.37
GUCA1B	6p21.1	0	0.66	0.38-1.24	50.31	0.22
GUCY2D	17p13.1	0	0.52	0.37-0.75	64.67	0.78
HARS	5q31.3	0	0.49	0.32-0.76	19.43	1.26
HGSNAT	8p11.21-p11.1	0	0.49	0.34-0.73	68.49	0.71
HK1	10q22.1	0.91	0.19	0.11-0.33	49.12	3.23
HMCN1	q25.3-q31.1	0	0.41	0.35-0.48	29.46	0.28
HMX1	4p16.1	0.73	0	0-0.7	66.38	-0.1
HTRA1	10q26.13	0	0.43	0.26-0.75	19.96	1.05
IDH3B	20p13	0	0.85	0.59-1.26	20.66	0.42
IFT140	16p13.3	0	0.64	0.51-0.82	69.23	-0.81
IFT172	2p33.3	0	0.62	0.51-0.76	16.01	1.19
IFT27	22q12.3	0	0.81	0.49-1.41	55.96	0.22
IFT81	12q24.11	0	0.57	0.41-0.8	18.76	0.83
IMPDH1	7q32.1	0	0.45	0.3-0.7	17.38	1.69
IMPG1	6q14.1	0	1.05	0.82-1.35	68.61	-0.94
IMPG2	3q12.3	0	0.53	0.4-0.72	48.42	-0.21
INPP5E	9q34.3	0	0.37	0.22-0.68	77.32	0.51
INVS	9q31.1	0	0.69	0.53-0.92	32.04	1.07
IQCB1	3q13.33	0	0.64	0.47-0.9	22.44	0.31
ITM2B	13q14.2	0.63	0.17	0.07-0.54	20.15	1.04
JAG1	20p12.2	1	0.06	0.03-0.15	1.06	3.25
KCNJ13	2q37.1	0.01	0.41	0.21-0.86	17.84	1.93
KCNV2	9p24.2	0	1.88	1.34-1.97	44.67	-4.48
KIAA1549	7q34	0.99	0.18	0.12-0.29	82.97	-0.09
KIF11	10q23.33	1	0.04	0.01-0.12	9.01	3.27
KIZ	20p11.23	NA	NA	NA	NA	NA
KLHL7	7p15.3	0	0.4	0.26-0.64	12.47	3.91
KSS	Mitochondrial	NA	NA	NA	NA	NA
LAMA1	18p11.31-p11.23	0	0.51	0.43-0.62	60.72	-0.14
LCA5	6q14.1	0	0.41	0.27-0.65	45.14	-0.62
LHON	Mitochondrial	NA	NA	NA	NA	NA
LRAT	4q32.1	0.03	0.52	0.23-1.32	31.38	-0.02
LRIT3	4q25	0	0.95	0.65-1.42	77.16	0.31
LRP5	11q13.2	0.51	0.22	0.15-0.34	8.86	1.67

LZTFL1	3p21.31	0.06	0.29	0.16-0.57	50.24	1.65
MAK	6p24.2	0	0.89	0.66-1.21	58.36	0.4
MAPKAPK3	3p21.2	0	0.73	0.5-1.11	32.41	1.44
MCDR3	5p15.33-p13.1	NA	NA	NA	NA	NA
MCDR4	14q11.2	NA	NA	NA	NA	NA
MCDR5	19q13.31-q13.32	NA	NA	NA	NA	NA
MDDC	7p21-p15	NA	NA	NA	NA	NA
MERTK	2q13	0	0.54	0.39-0.75	61.44	0.59
MFN2	1p36.22	0.99	0.13	0.07-0.28	12.07	1.66
MFRP	11q23.3	0	0.88	0.64-1.22	61.74	-1.47
MFSD8	4q28.2	0	0.75	0.53-1.07	58.38	0.11
MIR204	9q21.12	NA	NA	NA	NA	NA
MKKS	20p12.2	0	0.83	0.56-1.28	38.6	-0.05
MKS1	17q22	0	0.76	0.57-1.04	34.89	0.49
MT-ATP6	Mitochondrial	NA	NA	NA	NA	NA
MT-TH	Mitochondrial	NA	NA	NA	NA	NA
MT-TL1	Mitochondrial	NA	NA	NA	NA	NA
MTTP	4q23	0	0.39	0.26-0.59	40.96	1.03
MT-TP	Mitochondrial	NA	NA	NA	40.96	NA
MT-TS2	Mitochondrial	NA	NA	NA	NA	NA
MVK	12q24.11	0.17	0.26	0.13-0.55	64.64	0.94
MYO7A	11q13.5	0	0.7	0.58-0.85	15.89	1.07
NBAS	2p24.3	0	0.65	0.55-0.78	58.53	-0.87
NDP	Xp11.3	0.65	0	0-0.88	3.8	0.97
NEK2	1q32.3	0	0.52	0.34-0.82	25.91	1.3
NEUROD1	2q31.3	0.77	0.11	0.04-0.51	0.94	0.23
NMNAT1	1p36.22	0.04	0.38	0.19-0.88	62.8	0.63
NPHP1	2q13	0	0.73	0.55-0.97	60.81	0.23
NPHP3	3q22.1	0	0.5	0.38-0.65	40.01	0.86
NPHP4	1p36.31	0	0.78	0.63-0.98	68.33	-0.24
NR2E3	15q23	NA	NA	NA	NA	NA
NR2F1	5q15	0.99	0	0-0.19	2.69	4.17
NRL	14q11.2	0.05	0.43	0.19-1.1	34.98	0.61
NYX	Xp11.4	0.13	0.36	0.15-1.13	58.32	2.11
OAT	10q26.13	0	0.61	0.39-0.96	31.64	0.88
OFD1	Xp22.2	0.96	0.17	0.1-0.32	70.24	0.32
OPA1	3q29	0.99	0.18	0.12-0.29	6.93	1.97
OPA2	Xp11.4-p11.2	NA	NA	NA	NA	NA
OPA3	19q13.32	0.57	0	0-1.13	71.9	0.05
OPA4	18q12.2-q12.3	NA	NA	NA	NA	NA
OPA5	22q12.1-q13.1	NA	NA	NA	NA	NA
OPA6	8q21-q22	NA	NA	NA	NA	NA
OPA8	16q21-q22.3	NA	NA	NA	NA	NA
OPN1LW	Xq28	0.98	0	0-0.26	67.35	0.47
OPN1MW	Xq28	0.04	0.73	0.29-1.78	71.79	0.75
OPN1SW	7q32.1	0.02	0.37	0.19-0.79	40.77	0.01
OR2W3	1q44	0	0.99	0.53-1.77	77.38	-0.66
OTX2	14q22.3	0.92	0.08	0.03-0.38	0.71	1.05
PANK2	20p13	0	0.61	0.4-0.97	28.84	0.18
PAX2	10q24.31	0.67	0.19	0.09-0.43	0.23	1.49
PCDH15	10q21.1	0	0.63	0.5-0.79	23.23	-1.68

PCYT1A	3q29	0	0.52	0.33-0.86	18.95	1.64
PDE6A	5q33.1	0	0.86	0.67-1.12	20.89	-0.43
PDE6B	4p16.3	0	0.92	0.72-1.19	31.25	-0.8
PDE6C	10q23.33	0	0.65	0.49-0.87	40.69	1.14
PDE6G	17q25.3	0.01	0.7	0.32-1.66	34.65	-0.03
PDE6H	12p12.3	0.01	0.88	0.39-1.81	39.72	0.19
PDZD7	10q24.31	0	0.76	0.52-1.13	40.76	-0.37
PEX1	7p21.2	0	0.5	0.38-0.66	50.89	1.14
PEX2	8q21.13	0	0.64	0.36-1.2	51.39	0.05
PEX7	6q23.3	0	0.97	0.67-1.42	31.6	0.5
PGK1	Xq21.1	0.77	0.15	0.06-0.47	2.45	0.34
PHYH	10q13	0	0.68	0.43-1.12	73.66	0.03
PITPNM3	17p13.2	1	0.09	0.04-0.21	47.8	2.01
PLA2G5	1p36.13-p36.12	0	0.85	0.48-1.57	78.31	0.27
PLK4	4q28.2	0	0.38	0.26-0.57	23.23	0.8
PNPLA6	19p13.2	0	0.48	0.36-0.64	32.35	4.35
POC1B	12q21.33	0	0.61	0.41-0.93	43.82	-0.09
POC5	5q13.3	0	0.59	0.4-0.92	70.73	0.39
POMGNT1	1p34.1	0	0.79	0.6-1.04	12.19	0.9
PRCD	17q25.1	0.02	0.65	0.29-1.58	64.9	0.3
PRD	Xp11.3-p11.23	NA	NA	NA	NA	NA
PRDM13	6q16.2	0.56	0.2	0.1-0.46	56.2	0.42
PROM1	4p15.32	0	0.8	0.62-1.05	42.13	-0.61
PROS1	3q11.1	0	0.35	0.23-0.57	42.01	0.58
PRPF3	1q21.2	1	0	0-0.07	11.61	3.85
PRPF31	19q13.42	0.98	0.12	0.05-0.3	27.79	3.05
PRPF4	9q32	1	0.09	0.04-0.23	5.52	1.99
PRPF6	20q13.33	0	0.29	0.19-0.44	34.68	4.85
PRPF8	17p13.3	1	0.11	0.07-0.18	4.91	8.28
PRPH2	6p21.1	0.12	0.29	0.14-0.67	36.43	0.1
PRPS1	Xq22.3	0.92	0	0-0.38	26.11	3.73
RAB28	4p15.33	0.02	0.36	0.19-0.76	12.5	0.38
RAX2	19p13.3	0.01	0.88	0.39-1.81	84.16	0.17
RB1	13q14.2	1	0.05	0.02-0.13	0.53	2.67
RBP3	10q11.22	0	0.4	0.26-0.65	56.69	-0.42
RBP4	10q23.33	0.52	0.19	0.08-0.59	32.94	0.78
RCBTB1	13q14.2	0	0.51	0.34-0.79	29.73	1.03
RCD1	6q25-q26	NA	NA	NA	34.93	NA
RD3	1q32.3	0	0.98	0.55-1.73	59.23	0.61
RDH11	14q24.1	0	0.83	0.53-1.34	47.73	0.55
RDH12	14q24.1	0	0.92	0.61-1.44	31.99	-0.14
RDH5	14q24.1	0	1.06	0.7-1.61	38.88	0.39
REEP6	19p13.3	0	0.84	0.49-1.5	83.06	-0.16
RGR	10q23.1	0	1.11	0.73-1.69	68.24	-0.28
RGS9	17q24.1	0	0.62	0.47-0.85	55.96	0.04
RGS9BP	19q13.12	0	0.87	0.42-1.76	74.65	0.67
RHO	3q22.1	0	0.6	0.35-1.08	2.43	0.24
RIMS1	6q13	0.99	0.19	0.14-0.28	9.51	2.01
RLBP1	15q26.1	0	0.64	0.39-1.08	33.98	-0.31
ROM1	11q12.3	0	0.59	0.33-1.1	15.19	-0.91
RP1	8q12.1	0	0.52	0.39-0.7	72.98	-0.89

RP17	17q23.2	0	0.68	0.42-1.15	82.9	-0.12
RP1L1	8p23.1	0	1.72	1.03-1.95	96.43	-10.07
RP2	Xp11.23	0.96	0	0-0.31	33.65	0.63
RP8	NA	NA	NA	NA	NA	NA
RP22	16p12.3-p12.1	NA	NA	NA	NA	NA
RP29	4q32-q34	NA	NA	NA	NA	NA
RP6	Xp21.3-p21.2	NA	NA	NA	NA	NA
RP63	6q23	NA	NA	NA	NA	NA
RP89/ KIF3B	20q11.21	0.08	0.26	0.15-0.45	22.85	3
RP9	7p14.3	0.02	0.38	0.2-0.8	65.7	0.69
RPE65	1p31.2	0	0.79	0.57-1.11	17.77	-0.24
RPGR	Xp11.4	1	0.04	0.01-0.21	70.26	1.25
RPGRIP1	14q11.2	0	0.69	0.54-0.88	59.78	0.25
RPGRIP1L	16q12.2	0	0.77	0.62-0.96	9.18	-0.11
RS1	Xp22.13	0.96	0	0.0-0.3	28.71	0.97
RTN4IP1	6q21	0	0.58	0.37-0.94	36.98	0.73
SAG	2q37.1	0	1.02	0.74-1.44	39.84	0.68
SAMD11	1p36.33	0	0.9	0.64-1.28	75.37	-3.44
SDCCAG8	1q43	0	0.56	0.41-0.78	19.88	-0.1
SEMA4A	1q22	0	0.41	0.27-0.64	53.05	0.22
SLC24A1	15q22.31	0	0.36	0.24-0.57	79.17	1.8
SLC25A46	5q22.1	0	0.55	0.34-0.91	15.47	0.16
SLC7A14	3q26.2	0.02	0.3	0.17-0.54	25.72	0.78
SNRNP200	2q11.2	1	0.04	0.02-0.08	19.63	5.94
SPATA7	14q31.3	0	0.8	0.57-1.15	75.54	0.04
SPP2	2q37.1	0	1.41	0.97-1.89	87.97	-0.39
TEAD1	11p15.3	1	0	0-0.12	3.27	1.66
TIMM8A	Xq22.1	0.65	0	0-0.87	16.92	1.15
TIMP3	22q12.3	0.63	0.17	0.07-0.53	3.47	1.83
TLR3	4q35.1	0	0.53	0.36-0.8	29.89	0.51
TLR4	9q33.1	0	0.66	0.45-0.99	7.75	0.65
TMEM126A	11q14.1	0	0.63	0.33-1.33	44.01	-0.57
TMEM216	11q12.2	0	0.66	0.34-1.37	24.22	0.49
TMEM237	2q33.1	0	0.86	0.6-1.25	49.63	0.09
TOPORS	9q21.1	1	0.11	0.05-0.24	18.11	1.05
TREX1	3p21.31	0.58	0.15	0.05-0.69	77.67	-0.82
TRIM32	9q33.1	0	0.45	0.26-0.85	26.93	0.83
TRNT1	3p26.2	0	0.5	0.3-0.88	70.25	-1.15
TRPM1	15q13.3	0	0.88	0.72-1.07	57.81	-0.13
TSPAN12	7q31.31	0.69	0.18	0.08-0.46	16.62	0.77
TTC8	14q32.11	0	0.48	0.32-0.74	23.33	-0.08
TTLL5	14q24.3	0	0.6	0.47-0.76	12.66	-0.04
TTPA	8q12.3	0	0.55	0.3-1.09	46.94	0.3
TUB	11p15.4	0	0.33	0.2-0.56	13.93	0.47
TUBGCP4	15q15.3	0	0.5	0.35-0.72	13.88	2.58
TUBGCP6	22q13.33	0	0.74	0.6-0.93	77.41	-0.94
TULP1	6p21.31	0	0.37	0.23-0.6	57.11	0.65
UNC119	17q11.2	0	0.71	0.4-1.33	26.92	0.93
USH1C	11p15.1	0	0.67	0.51-0.89	22.31	-0.87
USH1G	17q25.1	0	0.62	0.37-1.09	45.66	0.63
USH2A	1q41	0	0.44	0.38-0.52	4.17	0.07

VCAN	5q14.3	1	0.13	0.08-0.2	13.43	0.14
WDPCP	2p15	0	0.55	0.4-0.78	24.68	0.9
WDR19	4p14	0	0.42	0.31-0.56	40.57	1.5
WFS1	4p16.1	0	1.62	1.27-1.93	33.79	-4.71
WHRN	9q32	0	0.41	0.26-0.67	38.98	-0.32
ZNF408	11p11.2	0	0.52	0.34-0.82	80.49	0.71
ZNF423	16q12.1	1	0.07	0.03-0.19	4.5	2.49
ZNF513	2p23.3	0.82	0.16	0.07-0.41	23.47	0.53

Supplementary Table 2. Results of evaluation of 39 genes in Table 1 for overrepresentation in biological processes using the PANTHER database.

Analysis Type:	PANTHER Overrepresentation Test (Released 20220202)
Annotation Version and Release Date:	GO Ontology database DOI: 10.5281/zenodo.5725227 Released 2021-11-16
Analyzed List:	Client Text Box Input (Homo sapiens)
Reference List:	Homo sapiens (all genes in database)
Test Type:	FISHER
Correction:	FDR

GO biological process complete	Homo sapiens - REFLIST (20595)	Client Text Box Input (39)	Client Text Box Input (expected)	Client Text Box Input (over/under)	Client Text Box Input (fold Enrichment)	Client Text Box Input (raw P-value)	Client Text Box Input (FDR)
spliceosomal tri-snRNP complex assembly (GO:0000244)	13	3	0.02	+	> 100	3.44E-06	4.90E-03
spliceosomal snRNP assembly (GO:0000387)	37	3	0.07	+	42.82	5.88E-05	3.29E-02
spliceosomal snRNP assembly (GO:0000387)	218	11	0.41	+	26.65	3.15E-13	4.94E-09
sensory perception of light stimulus (GO:0050953)	221	11	0.42	+	26.28	3.64E-13	2.85E-09
eye morphogenesis (GO:0048592)	153	5	0.29	+	17.26	1.15E-05	8.60E-03
camera-type eye morphogenesis (GO:0048593)	125	4	0.24	+	16.9	1.01E-04	4.67E-02
sensory organ morphogenesis (GO:0090596)	271	7	0.51	+	13.64	7.95E-07	1.56E-03
mRNA splicing, via spliceosome (GO:0000398)	234	5	0.44	+	11.28	8.35E-05	4.22E-02

RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	234	5	0.44	+	11.28	8.35E-05	4.09E-02
RNA splicing, via transesterification reactions (GO:0000375)	238	5	0.45	+	11.09	9.03E-05	4.29E-02
visual system development (GO:0150063)	375	7	0.71	+	9.86	6.51E-06	6.80E-03
camera-type eye development (GO:0043010)	324	6	0.61	+	9.78	3.34E-05	2.02E-02
sensory system development (GO:0048880)	381	7	0.72	+	9.7	7.21E-06	7.06E-03
eye development (GO:0001654)	371	6	0.7	+	8.54	7.01E-05	3.66E-02
sensory organ development (GO:0007423)	569	8	1.08	+	7.42	1.02E-05	7.99E-03
sensory perception (GO:0007600)	979	12	1.85	+	6.47	1.65E-07	4.32E-04
protein-containing complex assembly (GO:0065003)	1201	11	2.27	+	4.84	1.00E-05	8.25E-03
nervous system process (GO:0050877)	1434	12	2.72	+	4.42	8.85E-06	8.16E-03
cellular protein localization (GO:0034613)	1335	11	2.53	+	4.35	2.67E-05	1.82E-02
cellular component assembly (GO:0022607)	2320	19	4.39	+	4.32	7.33E-09	3.83E-05
cellular macromolecule localization (GO:0070727)	1345	11	2.55	+	4.32	2.87E-05	1.80E-02
protein-containing complex subunit organization (GO:0043933)	1372	11	2.6	+	4.23	3.44E-05	2.00E-02
cellular component biogenesis (GO:0044085)	2552	19	4.83	+	3.93	3.50E-08	1.37E-04
macromolecule localization (GO:0033036)	2289	15	4.33	+	3.46	9.16E-06	7.98E-03
anatomical structure morphogenesis (GO:0009653)	2180	14	4.13	+	3.39	2.60E-05	1.85E-02
nervous system development (GO:0007399)	2195	14	4.16	+	3.37	2.81E-05	1.83E-02
system process (GO:0003008)	2056	13	3.89	+	3.34	6.62E-05	3.58E-02

system development (GO:0048731)	4222	21	8	+	2.63	4.57E-06	5.52E-03
multicellular organism development (GO:0007275)	4564	22	8.64	+	2.55	3.74E-06	4.88E-03
cellular component organization (GO:0016043)	5314	25	10.06	+	2.48	5.71E-07	1.28E-03
anatomical structure development (GO:0048856)	5062	23	9.59	+	2.4	5.17E-06	5.79E-03
cellular component organization or biogenesis (GO:0071840)	5517	25	10.45	+	2.39	1.22E-06	2.12E-03
developmental process (GO:0032502)	5613	25	10.63	+	2.35	1.72E-06	2.70E-03
multicellular organismal process (GO:0032501)	6635	29	12.56	+	2.31	8.62E-08	2.70E-04