

# 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 <sup>1,2</sup> Hwei Wuen Chan <sup>3,4</sup> Elena Schiff,<sup>3,5</sup>  
Omar A Mahroo <sup>3,6</sup> Jose S Pulido<sup>3,7</sup>

**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



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Jose S Pulido; [jpulido@willseye.org](mailto: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  $\geq 0.9$  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.<sup>1–4</sup> Variants in over 250 genes are implicated. There are a number of unresolved questions relating to the spectrum of variants and mechanisms of disease.<sup>2</sup> Some associated genes are ubiquitously expressed, yet pathogenic variants appear to give rise only to IRD.<sup>5</sup> 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.<sup>6,7</sup> The



range of genes and variants involved in IRDs has been recently reviewed comprehensively by Schneider *et al*,<sup>8</sup> 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).<sup>9 10</sup> 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.<sup>11</sup> 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)<sup>10</sup> and DatabasE of genomic Variation and Phenotype in Humans using Ensembl Resources (DECIPHER).<sup>12</sup> 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).<sup>13</sup>

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.<sup>14</sup> 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,<sup>15</sup> 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.<sup>10</sup> 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/>)<sup>12</sup> 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  $\geq 0.9$ , 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<sup>13</sup> 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).<sup>16</sup> 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  $\geq 0.9$ , 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 ( $p = 3.44 \times 10^{-6}$ , 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.<sup>17</sup> From 20 394 reviewed genes on the Uniprot database (<https://www.uniprot.org/>),<sup>18</sup> 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*, *OPNILW*) 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 ( $p = 3.00 \times 10^{-10}$ , FDR of  $4.76 \times 10^{-6}$ ).

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.<sup>19</sup> 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.<sup>15</sup>

A number of genes encoding proteins involved in splicing complexes (*PRPF3*, *SNRNP200*, *PRPF4*, *PRPF8*, *PRPF31*)<sup>20</sup> 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
<i>COL11A1</i>	1p21.1	Dominant	Dominant Stickler syndrome type II; dominant Marshall syndrome	
<i>MFN2</i>	1p36.22	Dominant	Dominant optic atrophy with neuropathy and myopathy; dominant Charcot-Marie-Tooth disease	
<i>PRPF3</i>	1q21.2	Dominant	Dominant retinitis pigmentosa	
<i>EFEMP1</i>	2p16.1	Dominant	Dominant drusen	
<i>SNRNP200</i>	2q11.2	Dominant	Dominant retinitis pigmentosa	
<i>ATXN7</i>	3p14.1	Dominant	Dominant spinocerebellar atrophy with macular dystrophy or retinal degeneration	
<i>OPA1</i>	3q29	Dominant	Dominant optic atrophy; dominant optic atrophy with sensorineural hearing loss	
<i>VCAN</i>	5q14.3	Dominant	Dominant Wagner disease and erosive vitreoretinopathy	
<i>NR2F1</i>	5q15	Dominant	Dominant optic atrophy with intellectual disability and developmental delay (Bosch-Boonstra optic atrophy)	
<i>CTNNA1</i>	5q31.2	Dominant	Dominant macular pattern dystrophy (butterfly-shaped pigment dystrophy)	
<i>RIMS1</i>	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)
<i>AHR</i>	7p21.1	Recessive	Recessive retinitis pigmentosa	One family reported
<i>KIAA1549</i>	7q34	Recessive	Recessive retinitis pigmentosa	One report of two families
<i>GDF6</i>	8q22.1	Dominant and recessive	Recessive Leber congenital amaurosis; dominant Klippel-Feil syndrome; dominant microphthalmia	Single report of Leber congenital amaurosis
<i>TOPORS</i>	9q21.1	Dominant	Dominant retinitis pigmentosa	
<i>PRPF4</i>	9q32	Dominant	Dominant retinitis pigmentosa	
<i>HK1</i>	10q22.1	Dominant and recessive	Dominant retinitis pigmentosa; recessive nonspherocytic haemolytic anaemia; recessive hereditary neuropathy	
<i>KIF11</i>	10q23.33	Dominant	Dominant microcephaly, lymphedema and chorioretinopathy	
<i>TEAD1</i>	11p15.3	Dominant	Dominant atrophy areata/Sveinsson peripapillary degeneration	
<i>FZD4</i>	11q14.2	Dominant	Dominant FEVR	
<i>COL2A1</i>	12q13.11	Dominant	Dominant Stickler syndrome, type I; dominant bone dysplasias, developmental disorders, osteoarthritic diseases, syndromic disorders	
<i>CCT2</i>	12q15	Recessive	Recessive Leber congenital amaurosis	One report
<i>RB1</i>	13q14.2	Dominant	Dominant (or somatic) retinoblastoma; pinealoma; osteogenic sarcoma	
<i>OTX2</i>	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
<i>FBLN5</i>	14q32.12	Dominant	Dominant familial age-related macular degeneration; hereditary neuropathy with or without age-related macular degeneration	
<i>ZNF423</i>	16q12.1	Dominant and Recessive	Recessive nephronophthisis; dominant Joubert syndrome	One report including two families with Joubert syndrome
<i>PITPNM3</i>	17p13.2	Dominant	Dominant cone-rod dystrophy	One report of two families
<i>PRPF8</i>	17p13.3	Dominant	Dominant retinitis pigmentosa	
<i>C3</i>	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
<i>PRPF31</i>	19q13.42	Dominant	Dominant retinitis pigmentosa	
<i>JAG1</i>	20p12.2	Dominant	Dominant Alagille syndrome	
<i>RP2</i>	Xp11.23	X-linked	Retinitis pigmentosa	
<i>RPGR</i>	Xp11.4	X-linked	Retinitis pigmentosa; cone-rod dystrophy; macular dystrophy	
<i>DMD</i>	Xp21.2-p21.1	X-linked	Duchenne muscular dystrophy	Electroretinogram may be abnormal
<i>RS1</i>	Xp22.13	X-linked	X-linked retinoschisis	
<i>OFD1</i>	Xp22.2	X-linked	Joubert syndrome; orofaciocigital syndrome 1, Simpson-Golabi-Behmel syndrome 2	
<i>CHM</i>	Xq21.2	X-linked	Choroideremia	
<i>PRPS1</i>	Xq22.3	X-linked	Retinitis pigmentosa, neuropathy, optic atrophy, deafness	
<i>OPN1LW</i>	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.<sup>21 22</sup> A later report described a patient with the same *RIMS1* variant, but a different phenotype (retinitis pigmentosa),<sup>23</sup> and there have further variants reported, but without detailed evidence to secure a causative

relationship.<sup>24–26</sup> Implication of *AHR* in retinal disease comes from a report of a single family.<sup>27</sup> For *KIAA1549*, there is a single report of two families.<sup>28</sup> For *GDF6*, there is a single report of a patient with Leber Congenital Amaurosis, and in this case, both parents showed electroretinography (ERG) abnormalities.<sup>29</sup> Similarly, for *CCT2*, there is a single report of Leber Congenital Amaurosis.<sup>30</sup> With respect to *ZNF423*, one study reported a patient with recessive nephronophthisis (no ocular data) and two families with Joubert syndrome.<sup>31</sup> For *PITPNM3*, one study reported two families<sup>32</sup>; a subsequent study showed the carrier frequency of the variant in the general population was high for a dominant disease.<sup>33</sup> Other reports of disease associated with this gene are either isolated cases or report variants that also have high carrier frequencies.<sup>34–37</sup> 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.<sup>38</sup> 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
<i>PRPF3</i>	1q21.2	Dominant	Dominant retinitis pigmentosa
<i>SNRNP200</i>	2q11.2	Dominant	Dominant retinitis pigmentosa
<i>NR2F1</i>	5q15	Dominant	Dominant optic atrophy with intellectual disability and developmental delay (Bosch-Boonstra optic atrophy)
<i>CTNNA1</i>	5q31.2	Dominant	Dominant macular pattern dystrophy (butterfly-shaped pigment dystrophy)
<i>KLHL7</i>	7p15.3	Dominant	Dominant retinitis pigmentosa
<i>HK1</i>	10q22.1	Dominant and recessive	Dominant retinitis pigmentosa; recessive nonspherocytic haemolytic anaemia; recessive hereditary neuropathy
<i>KIF11</i>	10q23.33	Dominant	Dominant microcephaly, lymphedema and chorioretinopathy
<i>COL2A1</i>	12q13.11	Dominant	Dominant Stickler syndrome, type I; dominant bone dysplasias, developmental disorders, osteoarthritic diseases, syndromic disorders
<i>PRPF8</i>	17p13.3	Dominant	Dominant retinitis pigmentosa
<i>PNPLA6</i>	19p13.2	Recessive	Boucher-Neuhauser syndrome with chorioretinal dystrophy
<i>PRPF31</i>	19q13.42	Dominant	Dominant retinitis pigmentosa
<i>JAG1</i>	20p12.2	Dominant	Dominant Alagille syndrome
<i>PRPF6</i>	20q13.33	Dominant	Dominant retinitis pigmentosa
<i>PRPS1</i>	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.<sup>39</sup> 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
<i>SAMD11</i>	1p36.33	Recessive	Recessive retinitis pigmentosa
<i>ALMS1</i>	2p13.1	Recessive	Alstrom syndrome
<i>WFS1</i>	4p16.1	Recessive	Recessive Wolfram syndrome (also autosomal dominant low frequency sensorineural hearing loss)
<i>RP1L1</i>	8p23.1	Dominant and Recessive	Dominant occult macular dystrophy; recessive retinitis pigmentosa
<i>KCNV2</i>	9p24.2	Recessive	Cone dystrophy with supernormal rod response
<i>ADAMTS18</i>	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 *RP11*, 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.<sup>40</sup>

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<sup>41</sup>; 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<sup>42</sup> 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<sup>43 44</sup> 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=5.74×10<sup>-10</sup>). 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

<sup>1</sup>Ophthalmology, University College London Institute of Ophthalmology, London, UK

<sup>2</sup>Ophthalmology, Moorfields Eye Hospital City Road Campus, London, UK

<sup>3</sup>University College London Institute of Ophthalmology, London, UK

<sup>4</sup>Department of Ophthalmology, University of Singapore, Singapore

<sup>5</sup>Moorfields Eye Hospital City Road Campus, London, UK

<sup>6</sup>Medical Retina Service, Moorfields Eye Hospital NHS Foundation Trust, London, UK

<sup>7</sup>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](mailto: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.

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

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### 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 Bakhom 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.