OPEN ACCESS

Original research

EyeG2P: an automated variant filtering approach improves efficiency of diagnostic genomic testing for inherited ophthalmic disorders

Eva Lenassi,[1,2,3] Ana Carvalho,[4,5] Anja Thormann,[6] Liam Abrahams,[7] Gavin Arno,[8,9] Tracy Fletcher,[2] Claire Hardcastle,[2] Javier Lopez,[7] Sarah E Hunt,[6] Patrick Short,[10] Panagiotis I Sergouniotis ,[1,2,3] Michel Michaelides,[8,9] Andrew Webster,[8,9] Fiona Cunningham,[6] Simon C Ramsden,[2] Dalia Kasperaviciute,[7] David R Fitzpatrick,[4] Genomics England Research Consortium, Graeme C Black,[1,2] Jamie M Ellingford [1,7]

For numbered affiliations see end of article.

Correspondence to
Dr Jamie M Ellingford, The University of Manchester, Manchester M13 9WL, Manchester, UK; jamie.ellingford@manchester.ac.uk

Received 7 May 2022
Accepted 16 December 2022

ABSTRACT

Background Genomic variant prioritisation is one of the most significant bottlenecks to mainstream genomic testing in healthcare. Tools to improve precision while ensuring high recall are critical to successful mainstream clinical genomic testing, in particular for whole genome sequencing where millions of variants must be considered for each patient.

Methods We developed EyeG2P, a publicly available database and web application using the Ensembl Variant Effect Predictor. EyeG2P is tailored for efficient variant prioritisation for individuals with inherited ophthalmic conditions. We assessed the sensitivity of EyeG2P in 1234 individuals with a broad range of eye conditions who had previously received a confirmed molecular diagnosis through routine genomic diagnostic approaches. For a prospective cohort of 83 individuals, we assessed the precision of EyeG2P in comparison with routine diagnostic approaches. For 10 additional individuals, we assessed the utility of EyeG2P for whole genome analysis.

Results EyeG2P had 99.5% sensitivity for genomic variants previously identified as clinically relevant through routine diagnostic analysis (n=1234 individuals). Prospectively, EyeG2P enabled a significant increase in precision (35% on average) in comparison with routine testing strategies ($p<0.001$). We demonstrate that incorporation of EyeG2P into whole genome sequencing analysis strategies can reduce the number of variants for analysis to six variants, on average, while maintaining high diagnostic yield.

Conclusion Automated filtering of genomic variants through EyeG2P can increase the efficiency of diagnostic testing for individuals with a broad range of inherited ophthalmic disorders.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Genome diagnostics for ophthalmic disorders have clinical utility, but automated prioritisation of genomic variants that underpin ophthalmic disorders remains a challenge in clinical diagnostics.

⇒ This is further emphasised by the spectrum and diversity of disease-causing genomic variants that have been described to underpin these heterogeneous conditions.

WHAT THIS STUDY ADDS

⇒ We compare EyeG2P, an updated software tool available through the Ensembl Variant Effect Predictor, with other routine approaches for variant prioritisation in the diagnosis of genomic ophthalmic disorders.

⇒ We show that EyeG2P can achieve high sensitivity to disease-causing variations with improvements in the precision of variant analysis to routine approaches.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our data support the use of EyeG2P as a front-line software tool for analysis of genomic data sets for individuals with suspected genomic ophthalmic disorders.

INTRODUCTION

Inherited ophthalmic disorders are a clinically and genetically heterogeneous group of conditions that represent a major cause of blindness in children and working-age adults.[1–3] They include developmental disorders affecting the whole eye (eg, microphthalmia and anophthalmia, aniridia), childhood cataracts, inherited retinal disorders (eg, rod-cone dystrophies, cone-rod dystrophies, macular dystrophies, Leber congenital amaurosis) and inherited optic neuropathies (eg, Leber hereditary optic neuropathy, dominant optic atrophy). Notably, these conditions may present in isolation or as part of a syndrome with major extraocular features (eg, Usher syndrome, Joubert syndrome, Bardet-Biedl syndrome).

Obtaining a genetic diagnosis in individuals with inherited ophthalmic disorders has been shown to inform management and to facilitate genetic counselling.[3] As a result, genomic investigations are increasingly being used as a front-line diagnostic tool for this group of conditions.[4–9] The more widespread availability of preimplantation genetic testing and gene-directed interventions (eg, gene therapy for Leber congenital amaurosis due to

To cite: Lenassi E, Carvalho A, Thormann A, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2022-108618

pathogenic genetic variation impacting *RPE65*) has increased both the value and the risk of genetic testing, placing ever increasing demands on the delivery of genetic testing in a timely and accurate manner.[10–15]

In this study, we evaluate the diagnostic utility of EyeG2P, a publicly accessible tool for prioritised analysis of variants identified in genes known to cause inherited ophthalmic conditions. We curated disease-related genes through robust and transparent standards and assessed the sensitivity and precision of EyeG2P both retrospectively and prospectively. EyeG2P uses logical filtering of genomic variants, taking into account their predicted molecular consequence, their population frequency and prior knowledge of disease mechanisms and inheritance patterns. Overall, we show that using EyeG2P as a first-tier analytical strategy reduces the number of variants requiring analysis by clinical scientists and hence increases the precision and efficiency of diagnostic testing. We demonstrate the utility of EyeG2P both for targeted testing and whole genome-based sequencing approaches, and compare diagnostic yields and precision with those achieved by Exomiser and Genomics England tiering approaches for individuals with genomic data sets.

METHODS

Curation of known disease genes

The gene2phenotype (G2P) web portal (https://www.ebi.ac.uk/gene2phenotype/)[16] was used to develop and curate an ophthalmic disorders panel. For the purpose of this work, we defined inherited ophthalmic disorders as a group of predominantly monogenic conditions that affect the eye. This group encompasses diseases affecting the anterior segment, cornea, lens, vitreous, retina and optic nerve, but not disorders in which the main site of dysfunction is, for example, the eyelid or the visual cortex. New entries were initiated by selection of a relevant gene symbol from the list of preloaded genes (with their associated Ensembl identifiers). For each entry, a gene or locus was linked via a disease mechanism to a disease. These connections were made after inspecting MEDLINE (through the PubMed interface); search terms included the gene name (HUGO Gene Nomenclature Committee, HGNC) and the disease name (as a minimum). A disease mechanism was defined as both an allelic requirement (mode of inheritance; eg, biallelic or monoallelic) and a variant consequence (mode of pathogenicity; eg, loss-of-function). A confidence attribute (confirmed, probable or possible) was also assigned to indicate how likely it is that the gene is implicated in the cause of the disease; the rules used to assign confidence, allelic requirement and variant consequence to entries have been previously described.[16] Each *locus–genotype–mechanism–disease–evidence* link was further characterised by assigning to it a set of phenotype terms (ie, clinical signs and symptoms) from the Human Phenotype Ontology.[17] The identifiers of the relevant publications that provide evidence for a specific gene–disease thread were stored and are available through the G2P web portal.

Sequencing and variant identification

Genomic sequencing data sets were generated in a tertiary healthcare setting (North West Genomic Laboratory Hub, Manchester, UK; ISO 15189:2012, United Kingdom Accreditation Service medical reference 9865) or through the Genomics England 100,000 Genomes Project. Individuals provided written consent for genomic analysis. All clinical data were collected as part of routine clinical care.

Diagnostic gene panels

Routine diagnostic gene panel testing was performed as previously described.[4 6 18] Briefly, enrichment techniques were used to select specific genes for analysis using Agilent SureSelect (Agilent Technologies, Santa, Clara, California) target enrichment kits that were designed to capture selected intronic regions and all protein-coding exons±50 bp of flanking intronic sequences of the selected panels of known disease genes. The decision on which panel to use was made by the referring clinician (either a consultant ophthalmologist or a consultant clinical geneticist with an interest in ophthalmic genetics).

High-throughput genomic sequencing was performed using Illumina HiSeq and NextSeq platforms. Raw sequencing reads were first aligned to the GRCh37 reference genome using Burrow-Wheeler Aligner (BWA)-mem,[19] with single nucleotide variants (SNVs) and small insertions/deletions (indels) identified using the Genome Analysis Toolkit (GATK).[20] Larger and more complex indels were identified using Pindel, and CNVs were identified using DeCON.[21] Variants were filtered using quality and read depth thresholds, as well as inhouse allele frequencies. The zygosity of CNVs was estimated based on their relative read depths. Regions that are highly polymorphic and/or difficult to survey through short-read, high-throughput techniques were masked from the initial analysis, specifically *RP1L1* exon 4, *USH1C* exon 18 and *RPGR* orf15.

Whole genome sequencing

Whole genome sequencing data sets were generated as part of the Genomics England 100,000 Genomes Project.[22] Briefly, this involved complete sequencing of the accessible genome at a target coverage of 30× per nucleotide using Illumina sequencing technology. Alignment of sequencing reads and variant calling was performed through an Illumina pipeline involving Isaac aligner, Starling (for SNVs and indels), Canvas and Manta (for structural variants and CNVs). Further details of these pipelines are available at https://research-help.genomicsengland.co.uk/display/GERE/Genomic+data.

Variant analysis

Routine diagnostics for gene panel analysis

Routine genomic analysis was performed using the Congenica platform.[23] This process involves filtering variants based on gene/location depending on the gene panel applied, the population frequency and the predicted molecular consequence. A complete list of the criteria used for variant filtering is available in online supplemental tables 1 and 2. After prefiltering, variants were analysed by clinically accredited scientists and were classified in accordance with the 2015 American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG/AMP) best practice guidelines.[24 25]

EyeG2P analysis

We merged variant calls for SNVs, indels and CNVs into single Variant Call Format (VCF) files for each individual and annotated them using the G2P plugin for Ensembl Variant Effect Predictor.[16 26] This plugin requires an input file which lists genes of interest and their allelic requirements; we used the EyeG2P data set and an allele frequency cut-off of 0.001 for variants in monoallelic genes and 0.05 for variants in biallelic genes. An additional list was used as input including all ClinVar pathogenic or likely pathogenic variants, all variants predicted to have a significant impact on splicing by SpliceAI,[27] and a selection of

hypomorphic alleles that are known to be pathogenic but exceed the variant frequency thresholds specified.

Comparisons between routine diagnostic analysis and EyeG2P

Results from EyeG2P were retrospectively compared with clinically reported variants identified from routine diagnostic analysis in 1234 individuals with inherited ophthalmic conditions. All the relevant study participants had a confirmed (or a provisional) molecular diagnosis and carried pathogenic, likely pathogenic or variants of uncertain significance, in accordance with the guidelines proposed by the ACMG/AMP[21]; these changes were identified in a disease-causing state and were deemed to fully account for the patient's phenotype at the time of routine diagnostic analysis. Prospectively collected data from 83 individuals were also used for comparison. The sensitivity and precision of EyeG2P in comparison with results from routine diagnostic analysis were subsequently assessed and the 95% CI calculated through Bayesian inference using the *binom* R package. All statistical analyses were performed in R and graphics created in R and BioRender.

EyeG2P applied to whole genome data sets

Ten individuals recruited to the Genomics England 100,000 Genomes Project[22] with ophthalmic disorders were selected for analysis using EyeG2P (six cases with rod-cone dystrophies, one with a rod dysfunction disorder and three with macular dystrophies).

Genomic variants identified for each of these 10 individuals were accessed (Genomics England Research Registry ID: RR117) and variants were prioritised in four distinct phases: (1) all SNVs, indels and CNVs/structural variants impacting genes within the EyeG2P gene panel were selected; (2) variants identified as high quality, defined as those variants also present in an aggregated file of high-quality variants for the 100,000 Genomes Project cohort; (3) variants matching the user-specified criteria for EyeG2P (as described in the EyeG2P analysis section); and (4) subsequent filtering to keep only variants impacting the most important (canonical) transcripts for genes and present in fewer than 4 of the 10 individuals analysed. Variants prioritised by this process were compared with clinically reported variants available within the gmc_outcome_questionnaire in the Genomics England Research Environment from first-tier analysis of the genomic data performed by clinical scientists in UK Genome Medicine centres. Data were compared with variants prioritised by Exomiser and Genomics England tiering variant analysis approaches that were applied during the 100,000 Genomes Project.[22] We also applied an updated version of tiering to these 10 cases, which is now available through the NHS Genomic Medicine Service, and includes consideration of variants in ClinVar and triaging of CNVs (TierA and TierNull). Briefly, SNVs and indels are tiered into three discrete groups (tier 1, tier 2 and tier 3). Variants in all three groups are rare in the population (max allele frequencies are set for dominant, recessive and mitochondrial inherited disorders and variants compared with frequencies available for internal cohorts, and cohort subsets in UK10K, Genome Aggregation Database and 1kG_phase3), have a protein-altering consequence in at least one transcript (high or moderate impact) and segregate appropriately within the family (this strategy can be relaxed to incomplete penetrance). Tier 1 and 2 variants impact genes within the virtual gene panel that is applied during analysis of the genomic data, in this case the retinal disorders gene panel available through PanelApp (https://panelapp.genomicsengland.co.uk/panels/307/): Tier 1, high-impact variants (predicted loss-of-function), de novo variants in monoallelic disorders and variants in ClinVar (or have the same amino acid change as a variant in ClinVar) with at least one pathogenic or likely pathogenic assertion; Tier 2, moderate-impact variants (eg, missense variants and splice region variants); and Tier 3, includes high-impact or moderate-impact variants that are outside the virtual gene panel applied during analysis. CNVs are tiered separately into two distinct groups: TierA, high-quality CNVs that overlap with genes or regions defined in the applied virtual gene panel; TierNull, high-quality CNVs (>2 kb) that do not overlap with the applied virtual gene panel.

RESULTS

Curation of the literature identified 667 genes for inclusion in EyeG2P

Between April 2017 and June 2020, we interrogated the biomedical literature for genes associated with highly penetrant inherited ophthalmic disorders. We identified 667 unique disease-implicated genes, encompassing 564 MIM disease terms. Using evidence of gene–disease associations from 1624 scientific publications, we identified 559 as 'confirmed', 135 as 'probable' and 108 as 'possible'. Within the 559 confirmed gene–disease pairs, the associated inheritance patterns were autosomal dominant (n=155), autosomal recessive (n=341), X linked (n=31), and other patterns including both autosomal dominant and recessive (n=32). A high-level assessment of the biological disease mechanism was performed for each gene–disease combination; 405 disorders were catagorised as loss-of-function disorders, 19 as dominant negative disorders and 62 as disorders exclusively associated with specific missense or inframe indel variants.

For each gene–disease pairs, the predominant compartment of the eye was determined: retina (n=303), lens (n=120), cornea (n=65), vitreous (n=22) and optic nerve (n=39). Notably, 245 of the curated pairs (37%) were associated with multisystemic disorders. Skeletal (n=97), skin (n=49), ear (n=63), kidney (n=48) and metabolism (n=39) were the most frequently associated extraocular manifestations.

EyeG2P is highly sensitive for detection of pathogenic variants

We assessed the capability of EyeG2P to identify molecular diagnoses in 1234 individuals who had previously undergone clinical diagnostic genetic testing at the UK North West Genomic Laboratory Hub. EyeG2P was able to prioritise the causal variants in 1228 of the 1234 study participants (99.5%, 95% CI=99.1% to 99.8%). The 1267 variants prioritised by EyeG2P were identified in 166 distinct genes and had diverse predicted molecular consequences (figure 1). These variants were detected in 497 individuals with autosomal recessive, 514 individuals with autosomal dominant and 217 individuals with X linked disorders. The six variants not identified by EyeG2P included 5'untranslated region (UTR) variants and intronic variants not prioritised by SpliceAI.

EyeG2P increases precision over routine diagnostic testing

We performed a prospective comparison of the diagnostic capabilities of EyeG2P with routine diagnostic analysis in an additional cohort of 83 consecutively ascertained individuals with ophthalmic disease. A molecular diagnosis was confirmed in 33 of 83 cases (40%).

For 31 of 33 individuals (94%, 95% CI=84.0% to 99.5%), the confirmed molecular diagnosis was highlighted by EyeG2P. Importantly, EyeG2P allowed a reduction in the number of

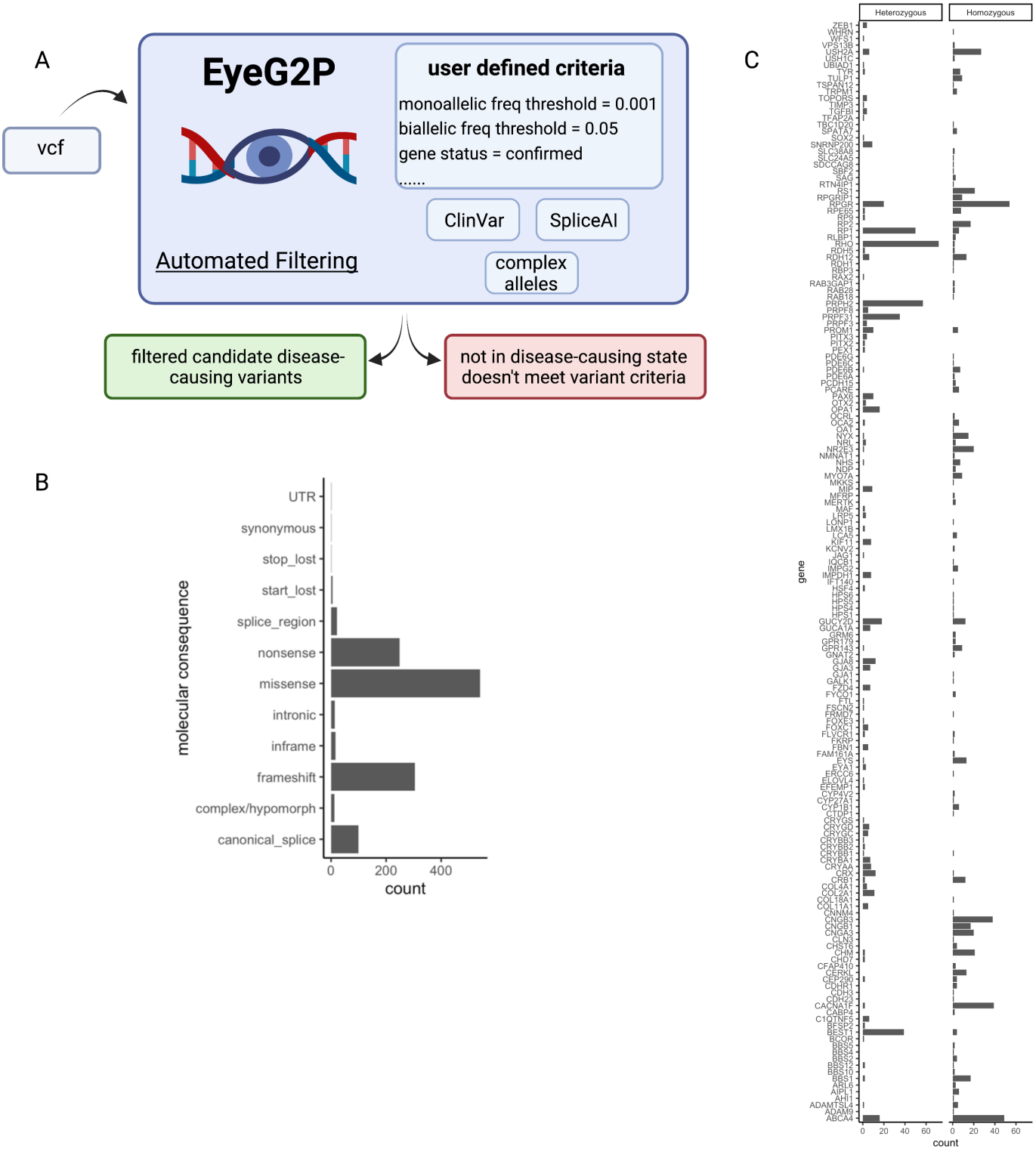


Figure 1 Molecular findings for 1228 individuals with a confirmed molecular diagnosis for inherited ophthalmic disorders. (A) Eye gene2phenotype (EyeG2P) is a plugin for the Ensembl variant effector enabling automated variant filtering and selection of variants in a disease-causing state (vcf, Variant Call Format). The specific requirements of variants to be retained can be set by the user and, through developments in the G2P software for the work described in this manuscript, can now include predefined lists of genomic variants, including pathogenic variants in ClinVar, variants predicted to impact splicing and complex alleles that comprised variants above the defined variant frequency threshold. (B) The predicted molecular consequences of 1267 variants identified as a cause of disease in 1228 individuals demonstrating the wide range of variant consequences that can be prioritised by EyeG2P (UTR, untranslated region). (C) The number of variants identified as a cause of disease in 166 genes by their proven zygosity. Hemizygous variants are included in the *Homozygous* display.

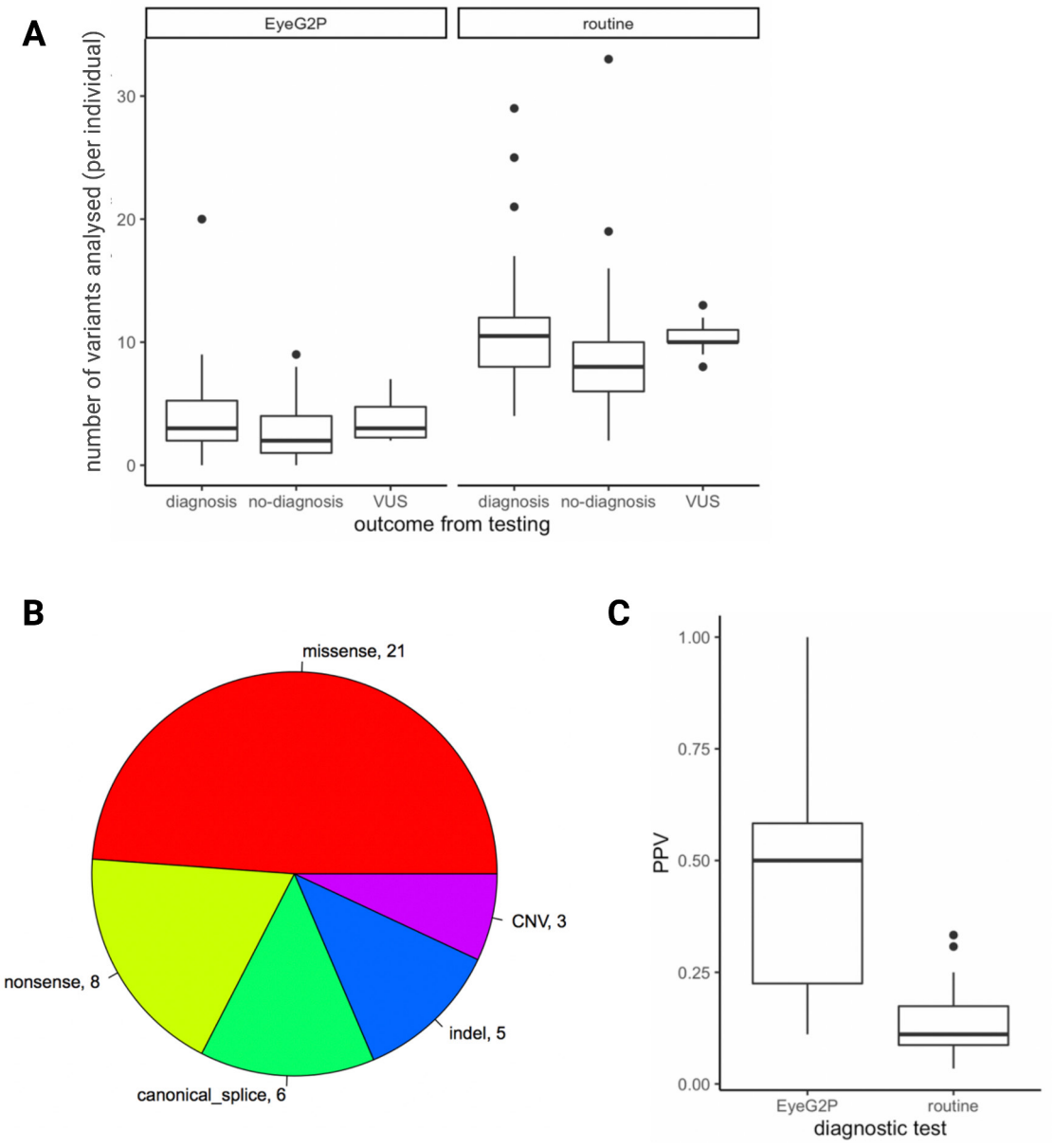
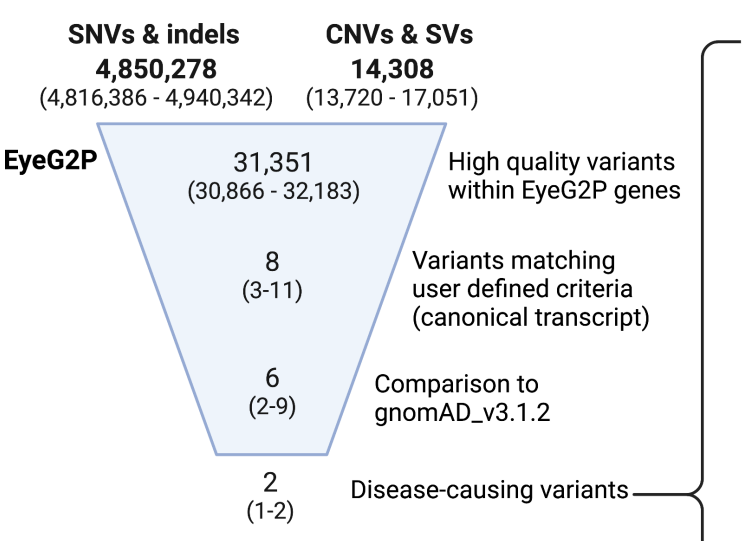


Figure 2 Direct comparison between EyeG2P and routine testing approaches identifies increased precision and efficiency of EyeG2P as a first-tier analysis approach. (A) The number of variants requiring analysis by clinical scientists for 83 individuals receiving genetic testing for inherited ophthalmic disorders (VUS, variant of uncertain significance). (B) Summary of the molecular consequences of pathogenic variants identified to underpin confirmed molecular diagnoses in 31 individuals. (C) The precision (PPV) of different testing approaches for 31 individuals receiving a diagnosis through both approaches. PPV, positive predictive value.

variants requiring analysis in each individual (mean=7.4; figure 2A). Disease-causing (pathogenic) variants were identified in 24 distinct genes (10 autosomal dominant, 3 X linked, 18 autosomal recessive disorders). The genomic variants underpinning these diagnoses included 3 CNVs (exonic deletions), 5 indels and 35 SNVs (figure 2B). Two cases had a confirmed genetic diagnosis of X linked retinitis pigmentosa as a result of variants in the orf15 region of the *RPGR* gene. These were not identified by EyeG2P as *RPGR* orf15 was excluded from EyeG2P analysis due to the limitations of short-read sequencing approaches in accurately identifying variants in this repetitive region—this is a limitation of the high-throughput short-read sequencing approaches applied, as variants in *RPGR* orf15 were not presented to EyeG2P for analysis.

In the remaining 50 cases in whom EyeG2P analysis did not identify a molecular diagnosis, we were also unable to identify a confirmed diagnosis through routine genetic testing approaches. Routine testing required an increased average analysis burden of 6.1 variants per individual (figure 2A). In 10 of 52 cases without a confirmed diagnosis, variants of uncertain significance were identified in a disease-causing state through the EyeG2P analysis.

We calculated the precision (positive predictive value) of genetic testing through EyeG2P and routine diagnostic genetic testing procedures for the 31 individuals who received a molecular diagnosis through both approaches. We observed a significant increase in precision in EyeG2P testing compared with default testing strategies (paired Wilcoxon rank-sum test, $p<0.001$), with an average increase in precision of 35% (figure 2C).



Case	Gene	Variant 1	Variant 2
1	*CNGB3*	8:86671024 CT>C NM_019098.5:c.412del NP_061971.3:p.(Arg138GlufsTer14)	8:86643780 AG>A NM_019098.5:c.1148del NP_061971.3:p.(Thr383IlefsTer13)
2	*GUCY2D*	17:8014700 C>T NM_000180.4:c.2512C>T NP_000171.1:p.(Arg838Cys)	-
3	*ABCA4*	1:93997841_del15 NM_000350.3):c.6729+5_6729+18del	1:94010911 T>A NM_000350.3:c.5603A>T NP_000341.2:p.(Asn1868Ile)
4	*ABCA4*	1:94078693 A>G NM_000350.3:c.1253T>C NP_000341.2:p.(Phe418Ser)	1:94010795 C>T NM_000350.3:c.5714+5G>A
5	*CRB1*	1:197442274 CG>C NM_201253.3:c.3988del NP_957705.1:p.(Glu1330SerfsTer11)	1:197427454 A>T NM_201253.3:c.2129A>T NP_957705.1:p.(Glu710Val)
6	*IMPG2*	3:101245876 G>A NM_016247.4:c.1469C>T NP_057331.2:p.(Thr490Ile)	-
7	*CDH3*	16:68691718 A>G NM_001793.6:c.1796-2A>G	16:68691718 A>G NM_001793.6:c.1796-2A>G
8	*USH2A*	1:215674344 C>A NM_206933.4:c.13567G>T NP_996816.3:p.(Val4523Phe)	1:216090900_del76417
9	*EYS*	6:64822650 C>T NM_001142800.2:c.3164+1G>A	6:64892927_del55368
10	*MYO7A*	11:77213858 A>G NM_000260.4:c.6439-2A>G	11:77162161 A>G NM_000260.4:c.1385A>G NP_000251.3:p.(Gln462Arg)

Figure 3 EyeG2P effectively prioritises pathogenic variation from whole genome sequencing data sets. The median and range of genomic variants identified across 10 individuals with whole genome sequencing analysis are shown at each stage of the filtering process. The complete list of clinically reported pathogenic and likely pathogenic variants prioritised by EyeG2P as a cause of inherited ophthalmic disorders is shown (genomic coordinates refer to the GRCh38 genome build). gnomAD, Genome Aggregation Database; indels, small insertion and deletion events; SNVs, single nucleotide variant; SVs, structural variants.

EyeG2P effectively filters variants from whole genome sequencing

We assessed the capability of EyeG2P to prioritise genomic variants identified in 10 individuals with visual disorders recruited to the Genomics England 100,000 Genomes Project (figure 3). EyeG2P identified a genetic diagnosis for all 10 individuals. There was 100% concordance with variants that had been clinically reported through analysis (facilitated by UK NHS Genomic Medicine centres) for these individuals. The variants prioritised further demonstrated the diversity of variant types and variant impacts that can be prioritised by EyeG2P (figure 3), including CNVs, intronic variants and hypomorphic variants. The impact of EyeG2P on the analysis of genomic variation is substantial, reducing the number of variants for analysis to six, on average, while maintaining a 100% concordance in diagnostic yield.

Findings from EyeG2P were compared with the variants triaged and the overall diagnostic outcomes available through tiering and Exomiser variant analysis strategies applied during the 100,000 Genomes Project, which identified complete molecular diagnoses for 7 of 10 individuals. These strategies have developed over time and we also compared the findings with the variants triaged by an updated tiering process at Genomics England (table 1). Updates to tiering enabled a complete diagnosis for one additional individual through identification of a known pathogenic variant which was originally excluded from tiering due to allele frequency (figure 3, case 3), and highlighted CNVs in *USH2A* and *EYS* as TierA variants that were part of the molecular diagnosis for two of the individuals analysed (figure 3, cases 8 and 9). The average precision for known disease-causing variants was slightly increased through Genomics England tiering (median=28%, range=13%–50%; table 1) in comparison with EyeG2P (median=24%, range=11%–50%; table 1), although a remaining current advantage of EyeG2P is the capability to consider CNVs and SNVs simultaneously during variant

Table 1 Comparative performance of EyeG2P, Exomiser and Genomics England tiering approaches to identify disease-causing variants from whole genome sequencing data for 10 individuals with ophthalmic disorders

Case	Number of variants (molecular diagnosis)	Number of variants (EyeG2P)	Precision (EyeG2P, %)	Number of variants (updated tiering)	Precision (updated tiering, %)	Rank (Exomiser)	Precision (Exomiser, %)
1	2	8	25	5	40	1	7
2	1	8	13	6	17	1	1
3	2	11	18	7	29	Not found	–
4	2	4	50	7	29	2	1
5	2	7	29	4	50	1	0.3
6	1	9	11	6	17	2	0.2
7	1	3	33	8	13	2	2
8	2	7	29	3	33*	Not found	–
9	2	10	20	5	20*	Not found	–
10	2	9	22	4	50	2	0.2

Updated tiering numbers presented include tier 1, tier 2 and TierA variants.
*Complete molecular diagnosis not found.

prioritisation and thereby increased diagnostic yield in this cohort.

DISCUSSION

Genomic testing for inherited ophthalmic disorders has been shown to have significant clinical utility.[4 10] The expansion from single gene-based methodologies to the routine use of large gene panels, exome and whole genome sequencing approaches requires robust and precise variant filtering strategies that take into account phenotypic information. Moreover, the continual identification of novel disease-related genes requires analytical approaches that can evolve dynamically. Here, we describe EyeG2P, a filtering approach available as a plugin for the Ensembl Variant Effect Predictor.[16] EyeG2P can be applied to any genomic variant data set in VCF format, including targeted gene panel, exome and genome data sets. We show that EyeG2P increases the precision and efficiency of genomic testing for inherited ophthalmic conditions over routine approaches for variant analysis.

The genetic basis of ophthalmic conditions is diverse and includes genes encoded by autosomes, sex chromosomes or mitochondrial DNA. This expands the initial assessments of G2P, which were largely focused on de novo variants causing dominant developmental disorders,[15] to a group of disorders that display onset in childhood and in adults, and have notable levels of incomplete penetrance. Moreover, genomic diagnostic services available through high-throughput sequencing techniques are highly mature for ophthalmic disorders and therefore provide a useful and informative benchmark for comparison with G2P software. Following curation of over 1000 biomedical publications, we identified 667 relevant genes and determined the associated modes of inheritance, high-level molecular pathological mechanisms and phenotypic features. We have released these data as a publicly available resource that can be dynamically filtered and revised to best aid users' requirements. For example, the recent elucidation of *DYNC2H1* as a cause of inherited ophthalmic conditions was not captured in our initial curation process, but this can be subsequently included in EyeG2P analysis through addition of a single data line to the released EyeG2P datafile.[28] To ensure there is long-term support for ongoing gene curation approaches and consistency in the methodology undertaken, G2P/EyeG2P has joined the Gene Curation Coalition (https://search.thegencc.org/). This initiative will enable regular updates to the curated EyeG2P panel and enable coordinated collaborative efforts across gene curation initiatives.

Our ability to detect pathogenic genomic variants from high-throughput sequencing data sets has expanded in recent years to include a wide range of mechanisms, including large structural genomic variants,[29 30] deletions and duplications within single genes ('exonic deletions),[31 32] variants deep within introns that may cause aberrant mRNA splicing,[33–36] variants in regulatory regions[37–39] and complex alleles that comprised combinations of genomic variants.[40 41] This requires a high level of specialist knowledge. We found that, in addition to characterising novel exonic variants, EyeG2P is capable of prioritising these diverse types of pathogenic variation, achieving 99.5% sensitivity in comparison with routine analytical approaches (figure 1). As our knowledge of the specific genomic variants causing inherited ophthalmic conditions inevitably expands further, it will be possible for the user to adjust the analysis settings of EyeG2P to meet these requirements and/or to provide a list of specific variants for inclusion. Such approaches can be applied at scale, as demonstrated here for the analysis of whole genome data sets (figure 3). Of note, we also demonstrate the utility of updated versions of the Genomics England tiering approach to identify variants that would have been originally excluded during the 100,000 Genomes Project, as has been shown for incomplete and complete penetrance strategies applied to trios (https://www.genomicsengland.co.uk/blog/automated-variant-interpretation-2). This further demonstrates the value of continually evaluating and evolving variant analysis software and illustrates the potential benefits of reanalysis with updated bioinformatics approaches.

In conclusion, we demonstrate that EyeG2P can be effectively integrated with clinical diagnostic testing for inherited ophthalmic conditions to increase the efficiency of variant analysis. We show that EyeG2P reduces the variant analysis workload for clinical scientists and increases the precision of diagnostic testing. Moreover, EyeG2P can identify diverse genomic variants across the spectrum of genetically and clinically heterogeneous ophthalmic genetic conditions. We propose the application of EyeG2P as a first-tier analysis strategy for the diagnosis of inherited ophthalmic conditions from high-throughput genomic data sets.

Author affiliations
[1]Division of Evolution, Infection and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom
[2]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, United Kingdom
[3]Manchester Royal Eye Hospital, Manchester University NHS Foundation Trust, Manchester, University Kingdom
[4]MRC Human Genetics Unit, MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom
[5]Medical Genetic Unit, Pediatric Hospital, Coimbra Hospital and Universitary Centre (CHUC), Coimbra, Portugal
[6]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom
[7]Genomics England Ltd, London, United Kingdom
[8]UCL Institute of Ophthalmology, University College London, London, United Kingdom
[9]Department of Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom
[10]Sano Genetics Ltd, Cambridge, United Kingdom

Acknowledgements We thank the families, clinicians, genetic counsellors, laboratory staff, clinical scientists and bioinformaticians involved in this study.

Collaborators Genomics England Research Consortium: JC Ambrose (Genomics England, London, UK), P Arumugam (Genomics England, London, UK), R Bevers (Genomics England, London, UK), M Bleda (Genomics England, London, UK), F Boardman-Pretty (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), CR Boustred (Genomics England, London, UK), H Brittain (Genomics England, London, UK), MA Brown (Genomics England, London, UK), MJ Caulfield (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), GC Chan (Genomics England, London, UK), A Giess (Genomics England, London, UK), JN Griffin (Genomics England, London, UK), A Hamblin (Genomics England, London, UK), S Henderson (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), TJP Hubbard (Genomics England, London, UK), R Jackson (Genomics England, London, UK), LJ Jones (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), D Kasperaviciute (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), M Kayikci (Genomics England, London, UK), A Kousathanas (Genomics England, London, UK), L Lahnstein (Genomics England, London, UK), A Lakey (Genomics England, London, UK), SEA Leigh (Genomics England, London, UK), IUS Leong (Genomics England, London, UK), FJ Lopez (Genomics England, London, UK), F Maleady-Crowe (Genomics England, London, UK), M McEntagart (Genomics England, London, UK), F Minneci (Genomics England, London, UK), J Mitchell (Genomics England, London, UK), L Moutsianas (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), M

Mueller (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), N Murugaesu (Genomics England, London, UK), AC Need (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), P O'Donovan (Genomics England, London, UK), CA Odhams (Genomics England, London, UK), C Patch (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), D Perez-Gil (Genomics England, London, UK), MB Pereira (Genomics England, London, UK), J Pullinger (Genomics England, London, UK), T Rahim (Genomics England, London, UK), A Rendon (Genomics England, London, UK), T Rogers (Genomics England, London, UK), K Savage (Genomics England, London, UK), K Sawant (Genomics England, London, UK), RH Scott (Genomics England, London, UK), A Siddiq (Genomics England, London, UK), A Sieghart (Genomics England, London, UK), SC Smith (Genomics England, London, UK), A Sosinsky (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), A Stuckey (Genomics England, London, UK), M Tanguy (Genomics England, London, UK), AL Taylor Tavares (Genomics England, London, UK), ERA Thomas (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), SR Thompson (Genomics England, London, UK), A Tucci (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), MJ Welland (Genomics England, London, UK), E Williams (Genomics England, London, UK), K Witkowska (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), SM Wood (Genomics England, London, UK; William Harvey Research Institute, Queen Mary University of London, London, UK), M Zarowiecki (Genomics England, London, UK).

Contributors Conceptualisation: GCB, DRF, FC, JME, SCR. Data curation: Genomics England Research Consortium, EL, AC, AT, LA, GA, TF, CH, SEH, PIS, MM, AW, JME. Formal analysis: EL, AT, TF, LA, CH, JME. Funding acquisition: GCB, DRF, FC, JME. Investigation and methodology: all authors. Software: AT, SEH, LA, JL, DK, FC, JME. Writing—original draft: EL, PIS, JME, GCB. Writing—review and editing: all authors. Guarantors: JME and GCB.

Funding Ensembl receives majority funding from Wellcome (grant numbers WT095908, WT098051, WT108749/Z/15/Z). We acknowledge funding from the Wellcome Trust Transforming Genomic Medicine Initiative (WT200990/Z/16/Z), the European Molecular Biology Laboratory, the Manchester NIHR Biomedical Research Centre (IS-BRC-1215-20007), and the NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology. JME is funded by a postdoctoral research fellowship from the Health Education England Genomics Education Programme. This work was developed in collaboration with Sano Genetics and acknowledges funding from the UK Research and Innovation Scholars Secondments: Biomedical Sciences (project ID: 75844). This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded the research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study involves human participants and was approved by the North West Research Ethics Committee (11/NW/0421 and 15/YH/0365) and East of England - Cambridge South Research Ethics Committee (14/EE/1112). All investigations were conducted in accordance with the tenets of the Declaration of Helsinki. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Shareable data is freely available through public resources, as specified in relevant sections of the manuscript, and is included in this manuscript.

ORCID iDs
Panagiotis I Sergouniotis http://orcid.org/0000-0003-0986-4123
Jamie M Ellingford http://orcid.org/0000-0003-1137-9768

REFERENCES

1 Solebo AL, Teoh L, Rahi J. Epidemiology of blindness in children. *Arch Dis Child* 2017;102:853–7.
2 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.
3 Heath Jeffery RC, Mukhtar SA, McAllister IL, Morgan WH, Mackey DA, Chen FK. Inherited retinal diseases are the most common cause of blindness in the working-age population in Australia. *Ophthalmic Genet* 2021;42:431–9.
4 Lenassi E, Clayton-Smith J, Douzgou S, Ramsden SC, Ingram S, Hall G, Hardcastle CL, Fletcher TA, Taylor RL, Ellingford JM, Newman WD, Fenerty C, Sharma V, Lloyd IC, Biswas S, Ashworth JL, Black GC, Sergouniotis PI. Clinical utility of genetic testing in 201 preschool children with inherited eye disorders. *Genet Med* 2020;22.
5 Charif M, Bris C, Goudenège D, Desquiret-Dumas V, Colin E, Ziegler A, Procaccio V, Reynier P, Bonneau D, Lenaers G, Amati-Bonneau P. Use of next-generation sequencing for the molecular diagnosis of 1,102 patients with a autosomal optic neuropathy. *Front Neurol* 2021;12:602979.
6 Ellingford JM, Barton S, Bhaskar S, O'Sullivan J, Williams SG, Lamb JA, Panda B, Sergouniotis PI, Gillespie RL, Daiger SP, Hall G, Gale T, Lloyd IC, Bishop PN, Ramsden SC, Black GCM. Molecular findings from 537 individuals with inherited retinal disease. *J Med Genet* 2016;53:761–7.
7 Consugar MB, Navarro-Gomez D, Place EM, Bujakowska KM, Sousa ME, Fonseca-Kelly ZD, Taub DG, Janessian M, Wang DY, Au ED, Sims KB, Sweetser DA, Fulton AB, Liu Q, Wiggs JL, Gai X, Pierce EA. Panel-Based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet Med* 2015;17:253–61.
8 Ellingford JM, Barton S, Bhaskar S, Williams SG, Sergouniotis PI, O'Sullivan J, Lamb JA, Perveen R, Hall G, Newman WG, Bishop PN, Roberts SA, Leach R, Tearle R, Bayliss S, Ramsden SC, Nemeth AH, Black GCM. Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. *Ophthalmology* 2016;123:1143–50.
9 Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S, Plagnol V, Penkett C, Stirrups K, Rizzo R, Wright G, Josifova D, Bitner-Glindzicz M, Scott RH, Clement E, Allen L, Armstrong R, Brady AF, Carmichael J, Chitre M, Henderson RHH, Hurst J, MacLaren RE, Murphy E, Paterson J, Rosser E, Thompson DA, Wakeling E, Ouwehand WH, Michaelides M, Moore AT, Webster AR, Raymond FL, NIHR-BioResource Rare Diseases Consortium. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am J Hum Genet* 2017;100:75–90.
10 Ellingford JM, Sergouniotis PI, Lennon R, Bhaskar S, Williams SG, Hillman KA, O'Sullivan J, Hall G, Ramsden SC, Lloyd IC, Woolf AS, Black GCM. Pinpointing clinical diagnosis through whole exome sequencing to direct patient care: a case of Senior-Loken syndrome. *Lancet* 2015;385:1916.
11 MacLaren RE, Groppe M, Barnard AR, Cottriall CL, Tolmachova T, Seymour L, Clark KR, During MJ, Cremers FPM, Black GCM, Lotery AJ, Downes SM, Webster AR, Seabra MC. Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial. *Lancet* 2014;383:1129–37.
12 Edwards TL, Jolly JK, Groppe M, Barnard AR, Cottriall CL, Tolmachova T, Black GC, Webster AR, Lotery AJ, Holder GE, Xue K, Downes SM, Simunovic MP, Seabra MC, MacLaren RE. Visual acuity after retinal gene therapy for choroideremia. *N Engl J Med* 2016;374:1996–8.
13 Tuohy GP, Megaw R. A systematic review and meta-analyses of interventional clinical trial studies for gene therapies for the inherited retinal degenerations (IRDs). *Biomolecules* 2021;11. doi:10.3390/biom11050760. [Epub ahead of print: 19 May 2021].
14 Maguire AM, Simonelli F, Pierce EA, Pugh EN, Mingozzi F, Bennicelli J, Banfi S, Marshall KA, Testa F, Surace EM, Rossi S, Lyubarsky A, Arruda VR, Konkle B, Stone E, Sun J, Jacobs J, Dell'Osso L, Hertle R, Ma J-xing, Redmond TM, Zhu X, Hauck B, Zelenaia O, Shindler KS, Maguire MG, Wright JF, Volpe NJ, McDonnell JW, Auricchio A, High KA, Bennett J. Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N Engl J Med* 2008;358:2240–8.
15 Stone EM, Andorf JL, Whitmore SS, DeLuca AP, Giacalone JC, Streb LM, Braun TA, Mullins RF, Scheetz TE, Sheffield VC, Tucker BA. Clinically focused molecular investigation of 1000 consecutive families with inherited retinal disease. *Ophthalmology* 2017;124:1314–31.
16 Thormann A, Halachev M, McLaren W, Moore DJ, Svinti V, Campbell A, Kerr SM, Tischkowitz M, Hunt SE, Dunlop MG, Hurles ME, Wright CF, Firth HV, Cunningham F, FitzPatrick DR. Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat Commun* 2019;10:2373.
17 Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, Callahan TJ, Chute CG, Est JL, Galer PD, Ganesan S, Griese M, Haimel M, Pazmandi J, Hanauer M, Harris NL, Hartnett MJ, Hastreiter M, Hauck F, He Y, Jeske T, Kearney H, Kindle G, Klein C, Knoflach K, Krause R, Lagorce D, McMurry JA, Miller JA, Munoz-Torres MC, Peters RL, Rapp CK, Rath AM, Rind SA, Rosenberg AZ, Segal MM, Seidel MG, Smedley D, Talmy T, Thomas Y, Wiafe SA, Xian J, Yüksel Z, Helbig I, Mungall CJ, Haendel MA, Robinson PN. The human phenotype ontology in 2021. *Nucleic Acids Res* 2021;49:D1207–17.

18 Gillespie RL, O'Sullivan J, Ashworth J, Bhaskar S, Williams S, Biswas S, Kehdi E, Ramsden SC, Clayton-Smith J, Black GC, Lloyd IC. Personalized diagnosis and management of congenital cataract by next-generation sequencing. *Ophthalmology* 2014;121:2124–37.
19 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv* 2013;13033997.
20 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
21 Fowler A, Mahamdallie S, Ruark E, Seal S, Ramsay E, Clarke M, Uddin I, Wylie H, Strydom A, Lunter G, Rahman N. Accurate clinical detection of exon copy number variants in a targeted NGS panel using DECoN. *Wellcome Open Res* 2016;1.
22 Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, Cipriani V, Ellingford JM, Arno G, Tucci A, Vandrovcova J, Chan G, Williams HJ, Ratnaike T, Wei W, Stirrups K, Ibanez K, Moutsianas L, Wielscher M, Need A, Barnes MR, Vestito L, Buchanan J, Wordsworth S, Ashford S, Rehmström K, Li E, Fuller G, Twiss P, Spasic-Boskovic O, Halsall S, Floto RA, Poole K, Wagner A, Mehta SG, Gurnell M, Burrows N, James R, Penkett C, Dewhurst E, Gräf S, Mapeta R, Kasanicki M, Haworth A, Savage H, Babcock M, Reese MG, Bale M, Baple E, Boustred C, Brittain H, de Burca A, Bleda M, Devereau A, Halai D, Haraldsdottir E, Hyder Z, Kasperaviciute D, Patch C, Polychronopoulos D, Matchan A, Sultana R, Ryten M, Tavares ALT, Tregidgo C, Turnbull C, Welland M, Wood S, Snow C, Williams E, Leigh S, Foulger RE, Daugherty LC, Niblock O, Leong IUS, Wright CF, Davies J, Crichton C, Welch J, Woods K, Abulhoul L, Aurora P, Bockenhauer D, Broomfield A, Cleary MA, Lam T, Dattani M, Footitt E, Ganesan V, Grunewald S, Compeyrot-Lacassagne S, Muntoni F, Pilkington C, Quinlivan R, Thapar N, Wallis C, Wedderburn LR, Worth A, Bueser T, Compton C, Deshpande C, Fassihi H, Haque E, Izatt L, Josifova D, Mohammed S, Robert L, Rose S, Ruddy D, Sarkany R, Say G, Shaw AC, Wolejko A, Habib B, Burns G, Hunter S, Grocock RJ, Humphray SJ, Robinson PN, Haendel M, Simpson MA, Banka S, Clayton-Smith J, Douzgou S, Hall G, Thomas HB, O'Keefe RT, Michaelides M, Moore AT, Malka S, Pontikos N, Browning AC, Straub V, Gorman GS, Horvath R, Quinton R, Schaefer AM, Yu-Wai-Man P, Turnbull DM, McFarland R, Taylor RW, O'Connor E, Yip J, Newland K, Morris HR, Polke J, Wood NW, Campbell C, Camps C, Gibson K, Koelling N, Lester T, Németh AH, Palles C, Patel S, Roy NBA, Sen A, Taylor J, Cacheiro P, Jacobsen JO, Seaby EG, Davison V, Chitty L, Douglas A, Naresh K, McMullan D, Ellard S, Temple IK, Mumford AD, Wilson G, Beales P, Bitner-Glindzicz M, Black G, Bradley JR, Brennan P, Burn J, Chinnery PF, Elliott P, Flinter F, Houlden H, Irving M, Newman W, Rahman S, Sayer JA, Taylor JC, Webster AR, Wilkie AOM, Ouwehand WH, Raymond FL, Chisholm J, Hill S, Bentley D, Scott RH, Fowler T, Rendon A, Caulfield M, 100,000 Genomes Project Pilot Investigators. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med* 2021;385:1868–80.
23 Furness M. Sapientia: accelerating rare disease diagnosis and treatment. *Per Med* 2016;13:415–8.
24 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. *Genet Med* 2015;17:405–24.
25 Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, Biesecker LG, ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med* 2018;20:1054–60.
26 McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl variant effect predictor. *Genome Biol* 2016;17:122.
27 Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglou S, Sanders SJ, Farh KK-H. Predicting splicing from primary sequence with deep learning. *Cell* 2019;176:535–48.
28 Vig A, Poulter JA, Ottaviani D, Tavares E, Toropova K, Tracewska AM, Mollica A, Kang J, Kehelwathugoda O, Paton T, Maynes JT, Wheway G, Arno G, Khan KN, McKibbin M, Toomes C, Ali M, Di Scipio M, Li S, Ellingford J, Black G, Webster A, Rydzanicz M, Stawiński P, Płoski R, Vincent A, Cheetham ME, Inglehearn CF, Roberts A, Heon E, Genomics England Research Consortium. DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration. *Genet Med* 2020;22:2041–51.
29 Sanchis-Juan A, Stephens J, French CE, Gleadall N, Mégy K, Penkett C, Shamardina O, Stirrups K, Delon I, Dewhurst E, Dolling H, Erwood M, Grozeva D, Stefanucci L, Arno G, Webster AR, Cole T, Austin T, Branco RG, Ouwehand WH, Raymond FL, Carss KJ. Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Med* 2018;10:95.
30 Ascari G, Rendtorff ND, De Bruyne M. Long-Read sequencing to unravel complex structural variants of. *Front Cell Dev Biol* 2021;9.
31 Ellingford JM, Horn B, Campbell C, Arno G, Barton S, Tate C, Bhaskar S, Sergouniotis PI, Taylor RL, Carss KJ, Raymond LFL, Michaelides M, Ramsden SC, Webster AR, Black GCM. Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *J Med Genet* 2018;55:114–21.
32 Zampaglione E, Kinde B, Place EM, Navarro-Gomez D, Maher M, Jamshidi F, Nassiri S, Mazzone JA, Finn C, Schlegel D, Comander J, Pierce EA, Bujakowska KM. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. *Genet Med* 2020;22:1079–87.
33 Weisschuh N, Buena-Atienza E, Wissinger B. Splicing mutations in inherited retinal diseases. *Prog Retin Eye Res* 2021;80:100874.
34 Bauwens M, Garanto A, Sangermano R, Naessens S, Weisschuh N, De Zaeytijd J, Khan M, Sadler F, Balikova I, Van Cauwenbergh C, Rosseel T, Bauwens J, De Leeneer K, De Jaegere S, Van Laethem T, De Vries M, Carss K, Arno G, Fakin A, Webster AR, de Ravel de l'Argentière TJL, Sznajer Y, Vuylsteke M, Kohl S, Wissinger B, Cherry T, Collin RWJ, Cremers FPM, Leroy BP, De Baere E. ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet Med* 2019;21:1761–71.
35 Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, van den Born LI, Khan MI, Cornelis SS, Verheij JBGM, Pott J-WR, Thiadens AAHJ, Klaver CCW, Puech B, Meunier I, Naessens S, Arno G, Fakin A, Carss KJ, Raymond FL, Webster AR, Dhaenens C-M, Stöhr H, Grassmann F, Weber BHF, Hoyng CB, De Baere E, Albert S, Collin RWJ, Cremers FPM. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. *Genet Med* 2019;21:1751–60.
36 Qian X, Wang J, Wang M, Igelman AD, Jones KD, Li Y, Wang K, Goetz KE, Birch DG, Yang P, Pennesi ME, Chen R. Identification of Deep-Intronic splice mutations in a large cohort of patients with inherited retinal diseases. *Front Genet* 2021;12:647400.
37 Small KW, DeLuca AP, Whitmore SS, Rosenberg T, Silva-Garcia R, Udar N, Puech B, Garcia CA, Rice TA, Fishman GA, Héon E, Folk JC, Streb LM, Haas CM, Wiley LA, Scheetz TE, Fingert JH, Mullins RF, Tucker BA, Stone EM. North Carolina macular dystrophy is caused by dysregulation of the retinal transcription factor PRDM13. *Ophthalmology* 2016;123:9–18.
38 Green DJ, Lenassi E, Manning CS. North Carolina macular dystrophy: phenotypic variability and computational analysis of disease-implicated non-coding variants. *medRxiv*2021.
39 Bhatia S, Bengani H, Fish M, Brown A, Divizia MT, de Marco R, Damante G, Grainger R, van Heyningen V, Kleinjan DA. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved Pax6 enhancer causes aniridia. *Am J Hum Genet* 2013;93:1126–34.
40 Norman CS, O'Gorman L, Gibson J, Pengelly RJ, Baralle D, Ratnayaka JA, Griffiths H, Rose-Zerilli M, Ranger M, Bunyan D, Lee H, Page R, Newall T, Shawkat F, Mattocks C, Ward D, Ennis S, Self JE. Identification of a functionally significant tri-allelic genotype in the tyrosinase gene (Tyr) causing hypomorphic oculocutaneous albinism (OCA1B). *Sci Rep* 2017;7:4415.
41 Lee W, Zernant J, Nagasaki T. Cis-Acting modifiers in the ABCA4 locus contribute to the penetrance of the major disease-causing variant in Stargardt disease. *Hum Mol Genet*2021.