12 Genome Analysis for Inherited Retinal Disease: The State of the Art

Kate Oprych, Raquel Sofia Silva, Nikolas Pontikos, and Gavin Arno

Abstract

Inherited retinal disease (IRD) is a major global cause of blindness caused by mutations in a wide spectrum of genes essential to the retinal structure, maintenance and function. Current clinical diagnostic strategies in the UK are focused on targeted gene panel testing either by enrichment or virtually. Whole exome and genome sequencing (WES and WGS) have been used in rare disease genetic discovery now for a decade and are being integrated into many research pipelines and diagnostic strategies exemplified by the Genomics England 100,000 genomes project.

Here, we describe the current approaches to genetic and genomic analysis in IRD, the shortfalls and advantages of gene panel testing, WES and WGS in the context of single nucleotide, structural and copy number variants in coding, non-coding and intractable genomic regions.

Looking ahead, the missing heritability in IRD may be consequent on a number of factors: new genes, ignored or undetectable variants, new diseases for known genes, etc. Improved detection of genomic variation afforded by WGS paired with expanded variant databases, advances in variant interpretation, developing our understanding of the effect of non-coding variation using multiomics and integrating deep phenotyping and genomic data into machine learning tools will be the driving forces in better diagnosis of rare disease and discovery of novel causes of disease in the post-genomic era.

Keywords

Inherited retinal disease · Molecular genetics Whole exome sequencing · Whole genome sequencing · 100,000 genomes project Missing heritability · Multiomics · Artificial intelligence

K. Oprych
Great Ormond Street Hospital for Children, London, UK

R. S. Silva
UCL Institute of Ophthalmology, London, UK

N. Pontikos
UCL Institute of Ophthalmology, London, UK

Moorfields Eye Hospital, London, UK

G. Arno (✉)
Great Ormond Street Hospital for Children, London, UK

UCL Institute of Ophthalmology, London, UK

Moorfields Eye Hospital, London, UK
e-mail: g.arno@ucl.ac.uk

G. Prakash, T. Iwata (eds.), *Advances in Vision Research, Volume III*, Essentials in Ophthalmology,
https://doi.org/10.1007/978-981-15-9184-6_12

12.1 Background

Inherited retinal disease (IRD) defines a broad spectrum of disorders characterised by retinal cell dysfunction and/or cell death, together representing a leading cause of visual impairment and blindness worldwide [1–3]. They affect an estimated 1 in 2000 individuals and over two million people globally [4].

The IRD spectrum of disease demonstrates vast phenotypic variability across multiple clinical parameters including the age of onset, severity and progression and can be broadly classified based on the primary cell type affected, rate of degeneration and whether the retinal disease occurs in isolation or with additional systemic features [4, 5]. Fundus imaging and functional testing with electroretinography are used for deep phenotyping and classification within IRD. The most common form of IRD is rod-cone dystrophy (or retinitis pigmentosa, (RP)) which has a prevalence of 1 in 4000 worldwide [6] while the commonest single-gene recessive disease is *ABCA4*-retinopathy with a carrier frequency estimated at 1 in 25–50 [7, 8] .

Retinal disease is a common presenting feature of a number of syndromic conditions including ciliopathies, lysosomal storage diseases and metabolic disorders: Usher's syndrome, Bardet Biedl syndrome (BBS), Senior-Loken syndrome and Joubert Syndrome being a few examples, and inheritance patterns for both non-syndromic and syndromic IRD can be autosomal dominant, recessive, X-linked and mitochondrial [4].

12.2 Genetics of IRD

Since the discovery of *RHO*, as the first gene is known to cause autosomal dominant RP by linkage and Sanger sequencing of gene candidates in 1990 [9], technological advancements have enabled the discovery of over 270 genes responsible for IRD (https://sph.uth.edu/retnet/) with many thousands of pathogenic and candidate variants now reported. There is indeed vast allelic and genetic heterogeneity within IRD making a hugely complex disease model. This can be exemplified by broad phenotypic variability consequent upon single genes and indeed single variants (eg: *CRX*) [10] and conversely many variants in many genes leading to an almost indistinguishable phenotype (eg: rod-cone dystrophy or RP). Some IRD genes also demonstrate incomplete penetrance, such as *PRPF31* which encodes precursor mRNA-processing factor 31, a ubiquitously expressed protein that is required for correct splicing of pre-mRNA transcripts [11]. Haploinsufficiency for *PRPF31* leads to symptomatic adRP in an estimated 50–60% of carriers and the rescue of the phenotype is thought to be consequent upon a second genetic determinant at the same locus on the trans allele [12]. A molecular diagnosis is important for more accurate risk predictions in these cases, where the inheritance pattern will be more challenging to recognise.

12.3 Genetic Screening Approaches

A molecular diagnosis provides many benefits to patients and families with rare disease; it enables the provision of more accurate information regarding risk predictions, prognosis and investigations, improved clinical care and management, access to treatments (current or emerging) and allows families to put a name to their genetic disease with associated non-clinical benefits.

Precise genotyping will become essential with the expansion of gene therapy in research and clinical settings [13, 14]. Of note, patients value having a molecular diagnosis and the option of predictive testing and are hopeful for emerging therapies for family members [15].

Current genetic screening available for IRD is territory dependant, in the UK this includes single-gene test Sanger sequencing, targeted gene panels and unbiased testing: whole exome and whole genome sequencing (WES and WGS). Deciding on the best option is directed by cost, setting (i.e. research or clinical setting) and confidence regarding the expected genotype. Where the phenotype is highly suggestive of a specific gene, such as *BEST1* [16], *CHM* [17], *TIMP3* [18], *EFEMP1* [19], *C1QTNF5* [20], Sanger

sequencing is still likely to be the initial approach to molecular diagnosis. However, in general, the vast heterogeneity of IRD means that consecutive sequencing of genes is an expensive, time-consuming and logistically challenging approach to establishing the molecular diagnosis [21]. Consequently, a comprehensive approach is required if pathogenic variants are not to be overlooked and for new genotype-phenotype correlations to be discovered.

The development of massively parallel sequencing, also known as next-generation sequencing (NGS) has revolutionised genetic discovery and molecular diagnostic testing for IRD and other Mendelian conditions [22–24]. NGS is a powerful high-throughput technology that can perform parallel sequencing of DNA on a vast scale and sequence an entire genome in a single experiment. Such techniques are now more frequently becoming the first-line diagnostic tool [21].

The spectrum of massively parallel sequencing platforms is broad, relying on a wide range of targeting chemistries, read length, sequencing technology and data processing. For example, DNA library preparation for massively parallel sequencing can be performed in a number of ways, including enrichment through polymerase chain reaction (PCR) amplification or capturing regions with DNA probes or the relatively unbiased PCR-free whole genome sequencing (WGS). For the purpose of this review, we will break down the methods based on this into three sub-groups: targeted gene panel testing, whole exome sequencing (WES) and WGS.

12.3.1 Targeted Gene Panels

In the UK, currently the commonest approach to IRD molecular testing is using an NGS-based targeted gene panel. Selected genes that are known to be associated with IRD, and increasingly, regions known to harbour pathogenic non-coding variants, are targeted for NGS using an enrichment step. One of the first of these for IRD was a 105-gene panel first trialled in RP cases [21]. Prior to this, routine access to genetic testing for IRD was limited in the UK to direct Sanger sequencing of single genes which was predominantly aimed at autosomal dominant RP and X-linked RP patients [21] or microarray analysis for known mutations (APEX array Asper Ophthalmics, Tartu, Estonia) [25]. The introduction of NGS gene panel testing was able to improve the diagnostic rates to approximately 80% in adRP cases and from 24% to 51% for broader IRD cohorts, demonstrating the advantage of this technology and the importance of wider access to genetic testing [21, 26]. Diagnostic rates were much higher in a paediatric cohort, with almost 80% of cases having a molecular diagnosis confirmed across all paediatric IRD [27]. Clinical assessment of children can be challenging due to limited cooperation and because the full phenotype may not yet be apparent, so early molecular diagnosis can facilitate the progress of children onto specific care pathways for screening and monitoring [27].

There are however several areas where targeted gene panels will miss diagnoses. Perhaps the greatest drawback of targeted panel testing is that they may be out of date often even before they can be properly implemented and continuously require updates as novel gene discoveries are made or important intronic variants in known IRD genes are identified [27, 28]. When panels are used in a clinical setting this requires extensive validation procedures which will limit the frequency of updates.

Furthermore, the ability of targeted enrichment to identify structural variants (SV) and copy number variants (CNV) is limited to read depth analysis for detection of alterations in dosage afforded by high coverage depth of targeted genes thus enabling effective detection of copy loss/gain [29]. Structural rearrangements are emerging as an important cause of IRD, accounting for significant proportions of disease alleles in recent studies [30, 31] and include variants not detectable by dosage analysis alone, for example, inversions, translocations and complex structural rearrangements with breakpoints rarely found in coding regions and thus may evade detection [28, 32]. Alternative genetic approaches such as CGH array and SNP array platforms may be used to augment the diagnostic pipeline and can improve diagnostic rates by an estimated 7% [33].

An argument against expanded analysis afforded by WES and WGS in the clinical setting is that diagnostic reporting of genetic variants must be rigorously validated. Therefore, clinical diagnostics must be focused on identifying 'provable' disease-associated variants, those being variants in known genes with a demonstrable protein-altering pathogenic mechanism be it the loss of function, damaging missense or splice altering variants. Although variants identified outside of the protein-coding regions, deep intronic and regulatory variants and novel gene associations are of great interest for the advancement of our understanding of the rare disease, they are difficult for clinical scientists to interpret. Therefore, the choice of genetic analysis for clinical diagnostics is a complex and carefully considered balance between the cost/benefit of applying WES/WGS versus gene panels.

12.3.2 Whole Exome Sequencing

Targeting of massively parallel sequencing to the exons of protein-coding genes, WES has become the widest used method for variant discovery studies in Mendelian disease in the research setting since the first gene discovery was reported a decade ago [22]. More recently, WES has been evaluated for use in the clinical diagnostic setting for IRD [31, 34–37]. With this technique, a 'virtual gene panel' may be applied targeting the analysis only on known IRD genes. It is predicted that more than 85% of disease-causing mutations are located in the exome [38]. Advantages of WES (and WGS) over panel-based testing is that disease-causing variants in off-panel genes are captured and as discoveries are made, reanalysis of unsolved patient cohorts is possible in light of new findings without the need to perform additional and costly experiments. Additional considerations with WES (and WGS) include the potential for identification of incidental findings; the American College of Medical Genetics (ACMG) has provided a recommended list of 59 highly penetrant genes that ought to be reported when WES and WGS are undertaken [39], although in the UK there is still debate about reporting of secondary findings [40–42].

WES has been reported to successfully identify the molecular cause of IRD in approximately 50–80% of cases in selected cohorts [34–37]. However, the coverage of known IRD genes by WES has been reported to be less effective than targeted panel tests [28, 43]. Consequently, clinical services currently prefer panel testing as the first-tier choice to IRD molecular diagnosis as it is cheaper, quicker, more sensitive and limits secondary findings [28, 44].

One of the main limitations of WES is that many regions of the genome known to harbour well characterised pathogenic variants exist outside of the coverage of WES enrichment kits. Therefore, it will not identify the increasing number of pathogenic non-coding variants accounting for a significant proportion of the missing heritability in IRD [31]. In addition, like targeted capture panels, enrichment is an integral part of the library prep methodology, thus leading to an artificially distributed coverage depth across the exome and challenges with read depth interpretation for SV/CNV detection. This is compounded by the fact that breakpoints in non-coding regions cannot be identified making confident calling difficult in many cases and validation more complicated, having to rely on qPCR, aCGH or MLPA in many cases.

12.3.3 Whole Genome Sequencing

WGS is the most comprehensive short-read sequencing technique for genome analysis, enabling interrogation of over 95% of the 3 billion nucleotide human genome. Superiority over other NGS technologies is evident from head-to-head comparison of coverage, diagnostic rates and by using WGS after other NGS methods have failed to identify causative variants [45–48], with Ellingford et al. [24] extrapolating a 29% improvement of WGS over targeted gene panels and Carss et al. [31] finding a 6% improvement of WGS over WES. The reasons for the

improvement in detection rates over other NGS techniques are multiple, including identification of pathogenic variants in non-coding regions, GC rich regions and structural variations.

The introns of genes associated with Mendelian disease may harbour pathogenic variants. Identifying pathogenic non-coding variants amongst all of the benign variations is inherently challenging: a frequent analogy being the 'needle in a haystack' due to the 3–4 million variants from the reference genome harboured in the average Illumina short-read sequencing genome. Therefore, accurate prediction of the effect of non-coding variants will prove key in delineating which variants are likely pathogenic [49].

Intronic variants are an important contributor to IRD causation through disruption of flanking splice sites and altering the strength of deeply intronic cryptic splice sites, resulting in cryptic splicing and protein disruption across many genes including *CHM, ABCA4, USH2A and CEP290* [24, 31, 50–55]. Causative intronic variants do account for significant proportions of unsolved cases with the CEP290 c.2991 + 1655A > G variant alone accounting for up to 21% of cases of LCA [56].

Furthermore, exonic variants may evade enrichment; GC rich regions are highly stable and therefore resistant to the denaturation stage of PCR resulting in poor coverage of these regions in certain WES applications. Carss et al. [31] demonstrated that in a patient with Leber congenital amaurosis, WGS identified heterozygous variants in exon 1 of *GUCY2D*, which has GC content of 76%, which at the time, WES would not have captured.

CNV and SV detection and characterisation using WES have inherent difficulties in that the breakpoints are often not covered at all meaning the only mechanism of detection available is based on the read depth [57]. Although effective, the lack of coverage uniformity that WES provides limiting and complex SV, inversions, translocations, etc. will be impossible to determine without a clear loss/gain. WGS on the other-hand enables incorporation of read depth analysis and split-read data analysis into bioinformatic pipelines which can detect complex structural variants and resolve breakpoints to the nucleotide level [31, 58, 59]. For example, in a patient with typical RP, WGS identified a structural variant in *EYS* which caused a 55 kb deletion (chr6: 65,602,819–65,658,187del) that encompassed exons 15–18, with both breakpoints deeply intronic, missed by WES [31]. With an estimated 5% of IRD patients harbouring a pathogenic SV/CNV, these variants are likely to account for a significant proportion of missing heritability by WES studies [31].

Finally, WGS also allows retrospective interrogation of the data as new IRD genes and pathogenic noncoding genes are discovered. Whilst targeted gene panels do include some non-coding variants the panels quickly become out of date. For example, Elllingford et al. [32] discovered intronic variants in *ABCA4* and *GPR98,* as well as a new IRD gene (*TRPM1*), when completing WGS in cases unsolved following a targeted gene panel because they were unknown at the time of panel design.

Areas that remain intractable to all NGS methods are those with highly repetitive regions and homologous pseudogenes due to the inevitable misalignment and mapping problems associated with short-read sequencing in these regions [31, 32]. The final exon of *RPGR* (ORF15) is a key example that highlights this issue and accounts for the majority of X-linked RP and hence, additional testing with optimised Sanger protocols is required to identify pathogenic variants for these cases [31, 60]. This is an issue that single molecule, long-read sequencing (aka third-generation sequencing) should resolve with read lengths of >20 kB enabling correct alignment and read through of repetitive sequences [61].

In the UK, clinical genetics is undergoing a revolution exemplified by the completion of the sequencing of 100,000 genomes from 70,000 individuals as part of the Genomics England 100,000 genomes project (100KGP). Funded by NHS-England, this study sequenced the genomes of some 3500 NHS patients with IRD [62]. Predominantly recruiting family trios (unaffected parents with affected offspring) in the rare

disease cohort provides unprecedented power for providing individuals with molecular diagnoses and making the discovery of new pathogenic variants and disease-associated genes, since *de novo* mutations in affected individuals, compound heterozygosity and homozygosity are readily apparent. Furthermore, it has launched the development and integration of WGS within a mainstream health service with the necessary infrastructure, education, research and industrial partnerships that are fundamental for NHS patients to benefit and for management of the vast amounts of data generated [63]. The cost-effectiveness of WGS over other NGS technologies for IRD and other disorders is not yet well characterised, and interpreting the health economics of these investigations is more complicated still [64]. However, with the expansion and development of the global genomic industry [65], the price of WGS continues to fall and WGS is likely to become the most cost-effective and comprehensive molecular diagnosis in IRD and similar conditions. Genomics England Interpretation Partnerships (GeCIP) have been established in this unique project to combine the expertise of researchers and clinicians to critically analyse the data from the 100KGP and embed research in clinical care [62, 66] with great promise for novel discoveries in IRD genetics [67–69].

12.4 Missing Heritability in Retinal Disease

Currently 40–60% [23, 24, 28, 31, 35, 37] of cases remain without a molecular diagnosis following NGS, depending on cohort differences and technologies used. There are a number of explanations that may accumulatively explain the missing heritability of IRD. Undetected variants in known IRD genes: many patients who undergo NGS testing are found to harbour a single disease allele in a compatible known recessive gene [31]. In such cases, it is likely that an unidentified variant in the same gene is present on the second allele. As discussed above, intronic variant and structural rearrangements including complex SV are emerging as important disease alleles, perhaps representing as high as 10–20% of mutations. In addition, variants in regulatory regions affecting promoters, enhancers and transcription factor binding sequences are further areas that remain difficult to elucidate. Several examples of regulatory region pathogenic variants have been well characterised to date in genes including *EYS, NMNAT1 and CHM* [70–72].

As researchers and clinical scientists employ less biased genetic strategies in the search for causative variants in Mendelian disease, the spectrum of pathogenic variants in syndromic disease genes becomes broader. There are many reports now in the literature of non-syndromic IRD cases with identified pathogenic or likely pathogenic variants in syndromic disease genes. This emerging phenomenon may represent the mild end of the syndromic disease spectrum with presumed hypomorphic alleles [31, 73–77] or novel associations, thought to represent different mechanisms of disease [78]. These examples highlight the importance of incorporating syndromic disease genes with a retinal component into targeted gene panels as well as virtual panels for WES/WGS in IRD testing strategies (https://panelapp.genomicsengland.co.uk/).

12.5 Multiomics

Now that whole genome analysis is quickly becoming the preferred tool of choice for identification of disease-causing variants for inherited disorders, it allows the identification of reported or novel variants that can affect gene expression, protein function, regulatory sequences or protein level, including by SV/CNV. However, the unbiased approach of reading an entire genome comes with a liability of overloading data.

Accomplishing the task of finding the causative variant(s), and potentially novel disease-causing genes, may be facilitated by the integration of information from different *omics* approaches, as well as patient phenotypic stratification and the reference population. The next layer of informa-

tion comes from the investigation of genomic, epigenetics and cellular mechanisms that sheds light on the interface of DNA-RNA-Protein dynamics.

The DNA molecule harbours a great deal of information beyond its linear sequence. In fact, genes account for only 2% of the pool of genomic material. The remaining 98% non-coding fraction is mainly made up of repetitive sequences that have a structural function in chromosome topology, but other parts are conserved across species and have regulatory activity. Cis-acting regulatory elements account for 6% of DNA, three times the equivalent of coding genes. But how does it all connect?

The 2 metre-long DNA molecule is found in the nucleus wrapped around octamers of 4 core protein histones that have amino acids tails that can be modified. The type of post-translational modification (e.g.: acetylation and methylation) and the amino acid location within the tail, impact directly on how compacted the DNA is in that particular stretch, and the different combinations, similarly to the nucleotides, work as a histone code. The linear gene sequence is composed of the promoter region, the gene body that accounts for the transcribed sequence, and the 5′ and 3′ untranslated regions (UTRs). Actively transcribed genes and regulatory sequences have open conformations and specific histone modifications (e.g.: Histone 3 Lysine 4 tri-methyl (H3K4me3), H3K27ac), which are associated with euchromatin, while repressed genes and elements are found tightly compacted and are decorated with histone marks associated with facultative heterochromatin (H3K27me3). A third state of chromatin, constitutive heterochromatin is associated with repetitive regions, including centromeric and telomeric regions of the chromosome, mainly supporting structural functions. These are decorated by silencing marks (such as H3K9me3 and H4K20me). Another level of information is the sequence position. DNA is organised in domains that are orderly insulated in the genome and preferentially interact with specific sequence regions. These are tightly associated with chromatin state. Furthermore, within the interacting domain, we find specific short or long-range cis-interacting elements, also known as enhancers. When active, these elements are bound by transcription activators or repressors that have a critical function in gene regulation. Promoter-enhancer interactions are often mediated by DNA looping [79] and are maintained by specific architectural factors and boundary insulator elements. The perturbation of this 3D structure can lead to gene mis-expression or ectopic expression [80, 81]. Therefore, faithful genome organisation and chromatin accessibility are key to ensure precise expression patterns. Pinpointing the exact molecular mechanisms driving genetic disease can be challenging, especially in the context of tissue development. The cell type can also exert a variance since the same gene could have different expression patterns and/or be regulated by different regulatory elements [82]. Assessing enhancer function and activity is hindered by the high level of enhancer redundancy [83].

Genome-wide profiling of chromatin immunoprecipitation sites for histone modifications and transcription factors, as well as DNA interacting domains by chromosome capture conformation, are fast being employed in the context of eye disorders and development [84–87]. Additionally, DNA methylation, DNA accessibility (including ATAC-seq), RNA sequencing of the different coding and non-coding transcripts, and proteomic assays, can further highlight genomic and genetic components that are preferentially important in the context of ocular development and disease [86, 88–91]. These approaches have contributed towards the identification of novel causative non-coding variants in eye disease [92–95].

A pertinent example of the employment of these NGS techniques is the discovery of the causative variants for North Carolina macular dystrophy (NCMD) and Progressive bifocal chorioretinal atrophy (PBCRA), two rare dominantly inherited disorders that affect central vision from birth [reviewed in reference 18]. Two linked loci had been identified at 5p21 and 6q16 [96–100], and in spite of many gene-sequencing approaches, no coding defect could be recognised. Moving the approach to genome-wide scale finally unrav-

elled the nature of the causal variants. So far 5 single nucleotide variants (SNVs) and 6 independent tandem duplications were identified on both loci. All 6 SNVs were found in two clusters in 6q, 15 kb and 7 kb upstream of the PRDM13 transcription start site, where both clusters were found located in DNase hypersensitive sites [92, 95]. The 3 tandem duplications at 6q also span these sites and include *PRDM13* sequence duplication [92, 101, 102]. In 5p21, three independent structural variants were identified with a combined shared region of 39 kb. This critical region for the phenotype is located in a gene desert downstream of *IRX1* and upstream of *ADAMTS16* [92, 94]. DNase-seq from human foetal retina also identified active sites at a restricted time during retinal development. Critically it was also proven that NCMD and PBCRA may represent a spectrum of the same disorder, dependent on the extent of dysregulation of the target genes affected by these regulatory variants. The most likely pathogenic mechanism is a gain of function, although it remains to be proven, due to constrains of modelling macula development.

Advances in sequencing technologies have also allowed the mainstream use of long-read sequencing, which has opened new views on novel RNA splicing variants and DNA structural variants complexity [103]. Additionally, the decrease in cost has spiked the use of single-cell technologies. Droplet-based single-cell sequencing was initially applied in adult mouse retina [104]. More recently it was applied in developing and adult primate retinas [105–107]. In the latest study, it was particularly used to compare primate-specific cell types such as foveal cells [106]. This has particular interest towards dissecting phenotypic aspects of eye disease, such as macular disorders.

Large scale studies adapting trio-based sequencing (GE 100KGP [62]; Deciphering Developmental Disorders, [108]), have to lead the way into personalised medicine. This approach allows significant reduction of candidate variants, and additionally, allows phasing of genomes and variants; *de novo* mutations and rare chromosomic phenomenon can also be readily identified. Combined these events are responsible for genetic diversity within the population and potentially within different tissues of an individual, since they can occur as germline mutations or in somatic tissue. This genetic diversity underlies human physiology and potentially accounts for both rare and common diseases. Furthermore, mosaicism may explain certain aspects of human disease such as penetrance and severity of the disorder [95, 109–111].

On certain occasions, even after the integration of all levels of information described above, the number of variants of unknown significance (VUS) can be substantial. VUS cannot be systematically tested individually and despite major improvements of *in silico* predictions [112], there is still a high level of inaccuracy, especially for the case of non-coding variants. High-throughput techniques have emerged with the potential to answer some of these questions. Saturation genome editing (SGE) resorts to the use of CRISPR-cas9 to create a library of hundreds of mutations that are tested in vitro and assayed in a single assay for a number of genes [113]. A similar mutagenesis based method was used to recreate 210 variants in Rhodopsin to test the effect on the expression of the protein [114]. As with any technique, the limitations of both these procedures are reliant on the metric used for inferring the causal consequence on gene expression.

The systematic combination of the different molecular approaches allows the identification of specific gene regulatory networks, adding instrumental power to achieving a personalised genomic analysis. Understanding the pathomechanism is fundamental for the patient and family members' prognostic, developing novel therapeutic strategies and selecting suitable participants for clinical trials. Non-coding variants present great potential for new pharmacological targets of intervention, since they avoid risk for off-target gene sequence alterations [115].

12.6 Imaging, Genetics and Artificial Intelligence for Inherited Retinal Disease Analysis

In the era of big data and global collaborations, computational methods are an essential part of rare disease genetic diagnosis and researchers in the IRD sphere are leading these advances. Computational tools have long been used for bioinformatics analysis for processing genetic data but now are also becoming part of phenotype analysis and clinical decision support. Combining approaches in genomic data interpretation with phenotype analysis tools will lead to better understanding through improved data integration, which coupled with artificial intelligence, will yield advances in genetic diagnosis and improved efficiency in clinical practice.

Next-generation sequencing technologies such as WGS offer us the most complete view of a human genome yet. However, as described above WGS detects hundreds of thousands of rare variants per individual posing a significant challenge in the interpretation of disease causality. Furthermore, many of these mutations occur in poorly characterised regions of the human genome.

Detailed phenotyping by experienced clinicians through careful patient interrogation and clinical tests, such as electroretinograms and detailed retinal imaging, can greatly aid the process of identifying the likely disease-causing mutations by identifying similarities with previously genetically diagnosed cases thus narrowing the search space for genetic mutations.

The description and definitions of phenotypes can vary widely between clinicians. This complicates meaningful comparisons across genetic cases and makes it harder to identify genes to phenotype correlations in IRD. This also makes computational analysis of phenotypes intractable. There have been many efforts to standardise phenotypes through the introduction of controlled vocabularies of clinical terms using encoding schemes such as SNOMEDCT and the UMLS [116]. For rare diseases such as IRD, the Human Phenotype Ontology (HPO) [117] is established as the favoured option and is now adopted by large projects such as the 100,000 genomes project led by Genomics England.

HPO terms are used in computational approaches to prioritise disease-causing variants and uncover novel gene to phenotype associations. Exomiser [118] prioritises variants based on phenotype similarity with published OMIM conditions, model organisms and gene pathways. Bevimed [119] uncovers gene to phenotype relationships based on phenotype similarity regressions using Bayesian statistics.

HPO descriptions are a first step towards enabling the integration of phenotype and genetic data to match patients with similar clinical features. Nonetheless annotating genetic cases with HPO terms still requires manual input which is difficult to fit into the already busy clinical workflow of large ophthalmic hospitals. Therefore, solutions are sought which facilitate the collection of HPO terms such as, making them part of the electronic health record entry system [120], by extracting these terms automatically from patient notes using natural language processing techniques, or even directly from imaging, ERGs or visual fields, are sought. A limitation of HPO terms is that they rely on subjective clinical terms. A more objective approach is to directly analyse the primary source of these HPO terms such as the imaging data.

Retinal imaging technologies are now widely and extensively used in ophthalmology due to modern advances in the field such as Optical Coherence Tomography (OCT), which allows detailed imaging of the layers of the retina, detection of oedemas, drusen and various other features symptomatic of retinal disease. Additionally, given the accessibility of the eye, retinal imaging is both very efficient and cost-effective and is now not only part of routine care at ophthalmic hospitals but also available to community opticians [121]. However, the interpretation of these images requires expertise acquired through years of training, and for IRD in particular, which are very rare and thus hard to recognise, these images

need to be inspected by clinical experts with in-depth knowledge of genetics, of which they are very few worldwide.

Such an IRD expert may be able to recognise the pattern of retinal deterioration which are gene specific and make prognosis as to the development of the disease [122, 123]. Nonetheless, this remains a subjective process dependent on the skills and experience of the clinician. There is also a shortage of such experts worldwide and this gap is increasing with the spread and accessibility to these new technologies. The gap is unlikely to close given that this knowledge takes years of experience to acquire.

Beyond the problem of shortage of experts, human-interpretation is also limited to recognising IRD genes by looking for known features. However there are also potentially new IRD gene-specific features in images which can only be discovered once we pool sufficient data.

The promise of Artificial Intelligence (AI) is to provide a more scalable, efficient and objective solution to IRD genetic diagnosis by training a neural network on retinal images of as many confirmed genetically diagnosed cases of IRD as possible. The trained AI could then suggest a gene given a retinal image from an IRD patient. This should lead to a more objective form of clinical diagnosis and has the potential of capturing the image pattern-recognition skills of the most advanced clinicians into a neural network and making them available as a clinical decision support tool. Along with aiding in finding the most likely genetic region affected which can help identify hard to detect genetic causes such as structural variants, non-coding mutations and silent mutations, this could also guide the clinician as to which genetic test is the most appropriate.

There are reasons to be optimistic that this approach can work, as AI has recently shown good results when applied to triaging eye scan from age-related macular degeneration and diabetic patients [124]. This approach has particularly illustrated the utility of a type of neural network, known as segmentation neural networks, are able to identify features in an image such as macular holes, odemas [125], drusens, and these can aid the quantitative analysis of these and link them to disease. These segmentations were then used as input to a second type of neural networks, known as classification neural networks. It is also possible to run a classification neural network known as convolutional neural network directly on the pixel intensities of images [126, 127]. However, the challenge is then to deconvolute the features that were used in predicting the outcome in order to explain the classification process.

A concrete proof-of-concept applied to IRD has recently been published by Fujinami-Yokokawa et al. in 2017 [128]. They trained a four-class CNN classifier to distinguish foveal OCT slices between three types of IRD patients with *RP1L1*, *EYS* and *ABCA4* retinopathies and healthy patients. The IRD patients had confirmed disease-causing mutations in these genes.

Nonetheless, these approaches are limited by the amount of training data which is why international data-sharing collaborations are particularly important to augment these training data-sets, especially for rare disease. One advantage of these AI approaches when applied to IRD over common disorders such age-related macular degeneration and diabetic retinopathy, is that the training labels are more reliable as they do not depend on subjective clinical interpretation, for example, wet vs dry, but instead on objective genetic data.

12.7 Conclusions

The continuous evolution of high-throughput sequencing technologies has critically advanced our knowledge on the human genome. This genomic revolution has enabled the incorporation of WGS into clinical diagnostic pipelines and led to the generation of unprecedented volumes of data, carrying associated implications in variant interpretation. The identification and characterisation of human genes and non-coding regulatory regions have revolutionised the field of human genetics and its application in the clinical setting by providing more efficient diagnostics and potential new pharmacological targets for intervention in a personalised fashion.

As NGS technology continues to develop, we will gain further insight into the role of genetic variation in human biology and disease, which will grant us a better understanding of the mechanism by which variants affect gene expression in the dynamic context of a cell, a tissue and the integration of all systems in the single organism. It is clear that as our ability to interpret genomic variation and the effect of non-coding variants improves, the advantages of WGS in the clinical realm will far outweigh its limitations perhaps leading to the replacement of targeted gene panels or WES as a clinical tool.

Indeed, as genetics and imaging become more accessible to the general public thanks to direct-to-consumer genetic testing, eye scanners used as part of eye-tests by community opticians, portable and home devices being developed, it is important to also democratise the interpretation knowledge of these complex data to avoid the risks and dangers of misinterpretation [129].

References

1. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. BMJ Open. 2014;4:e004015.
2. Buch H, Vinding T, La Cour M, Appleyard M, Jensen GB, Nielsen NV. Prevalence and causes of visual impairment and blindness among 9980 Scandinavian adults; the Copenhagen City eye study. Ophthalmology. 2004;111:53–61.
3. Al-Merjan JI, Pandova MG, Al-Ghanim M, Al-Wayel A, Al-Mutairi S. Ophthalmic epidemiology registered blindness and low vision in Kuwait. Ophthalmic Epidemiol. 2005;12(4):251–7.
4. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. Prog Retin Eye Res. 2010;29(5):335–75.
5. Hamel CP. Gene discovery and prevalence in inherited retinal dystrophies. C R Biol. 2014 Mar;337(3):160–6.
6. Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. Lancet. 2006;368:1795–809.
7. Allikmets R, Singh N, Sun H, Shroyer NF, Hutchinson A, Chidambaram A, et al. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. Nat Genet. 1997;15(3):236–46.
8. Jaakson K, Zernant J, Külm M, Hutchinson A, Tonisson N, Glavač D, et al. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. Hum Mutat. 2003;22(5):395–403.
9. Dryja TP, McGee TL, Reichel E, Hahn LB, Cowley GS, Yandell DW, et al. A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. Nature. 1990 Jan;343:364–6.
10. Hull S, Arno G, Plagnol V, Chamney S, Russell-Eggitt I, Thompson D, et al. The phenotypic variability of retinal dystrophies associated with mutations in CRX, with report of a novel macular dystrophy phenotype. Invest Ophthalmol Vis Sci [Internet]. 2014 Jan [cited 2014 Dec 19];55(10):6934–44. Available from http://www.ncbi.nlm.nih.gov/pubmed/25270190
11. Rivolta C, Mcgee TL, Frio TR, Jensen RV, Berson EL, Dryja TP. Variation in retinitis Pigmentosa-11 (PRPF31 or RP11) gene expression between symptomatic and asymptomatic patients with dominant RP11 mutations. Hum Mutat. 2006;27(7):644–53.
12. Venturini G, Rose AM, Shah AZ, Bhattacharya SS, Rivolta C. CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. PLoS Genet. 2012 Nov;8(11):e1003040.
13. Den Hollander AI, Black A, Bennett J, Cremers FPM. Review series lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. J Clin Invest. 2010;120(9):3042–53.
14. Dalkara D, Sahel J-A. Gene therapy for inherited retinal degenerations. C R Biol. 2014;337:185–92.
15. Willis TA, Potrata B, Ahmed M, Hewison J, Gale R, Downey L, et al. Understanding of and attitudes to genetic testing for inherited retinal disease: a patient perspective. Br J Ophthalmol. 2013;97:1148–54.
16. Petrukhin K, Koisti MJ, Bakall B, Li W, Xie G, Marknell T, et al. Identification of the gene responsible for best macular dystrophy. Nat Genet. 1998;19(3):241–7.
17. Cremers FPM, Sankila E-M, Brunsmann F, Jay M, Jay B, Wright A, et al. Deletions in patients with classical choroideremia vary in size from 45 kb to several megabases. Am J Hum Genet. 1990;47(4):622–8.
18. Weber BHF, Vogt G, Pruett RC, Stohr H, Felbor U. Mutations in the tissue inhibitor of metalloproteinases-3 (TIMP3) in patients with Sorsby's fundus dystrophy. Nature. 1994;8:352–6.
19. Stone EM, Lotery AJ, Munier FL, Héon E, Piguet B, Guymer RH, et al. A single EFEMP1 mutation associated with both Malattia Leventinese and Doyne honeycomb retinal dystrophy. Nat Genet. 1999;22:199–202.
20. Hayward C, Shu X, Cideciyan A V, Lennon A, Barran P, Zareparsi S, et al. Mutation in a short-chain collagen gene, CTRP5, results in extracellular deposit formation in late-onset retinal degeneration: a genetic model for age-related macular degeneration. Hum Mol Genet [Internet]. 2003 Oct 15 [cited 2015 Jul 30];12(20):2657–67. Available from http://www.ncbi.nlm.nih.gov/pubmed/12944416

21. O'Sullivan J, Mullaney BG, Bhaskar SS, Dickerson JE, Hall G, O'Grady A, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. J Med Genet. 2012 May;49(5):322–6.
22. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010 Jan;42(1):30–5.
23. Patel A, Hayward J, Tailor V, Nyanhete R, Ahlfors H, Gabriel C, et al. The oculome panel test: next-generation sequencing to diagnose a diverse range of genetic developmental eye disorders. Ophthalmology. 2019;26(6):888–907.
24. Ellingford JM, Barton S, Bhaskar S, Williams SG, Sergouniotis PI, O'Sullivan J, et al. Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. Ophthalmology. 2016;123(5):1143–50.
25. Henderson RH, Waseem N, Searle R, van der Spuy J, Russell-Eggitt I, Bhattacharya SS, et al. An assessment of the apex microarray technology in genotyping patients with Leber congenital amaurosis and early-onset severe retinal dystrophy. Invest Ophthalmol Vis Sci. 2007 Dec;48(12):5684–9.
26. Ellingford JM, Barton S, Bhaskar S, O'Sullivan J, Williams SG, Lamb JA, et al. Molecular findings from 537 individuals with inherited retinal disease. J Med Genet [Internet]. 2016 May 11 [cited 2016 Aug 31]. Available from http://www.ncbi.nlm.nih.gov/pubmed/27208204
27. Taylor RL, Parry NRA, Barton SJ, Campbell C, Delaney CM, Ellingford JM, et al. Panel-based clinical genetic testing in 85 children with inherited retinal disease. Ophthalmology. 2017 Jul;124(7):985–91.
28. Consugar MB, Navarro-Gomez D, Place EM, Bujakowska KM, Sousa ME, Fonseca-Kelly ZD, et al. 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. 2014;17(4):253–61.
29. Ellingford JM, Horn B, Campbell C, Arno G, Barton S, Tate C, et al. Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. J Med Genet. 2017;55:114–21.
30. Stenson P, Mort M, Ball E, Evans K, Hayden M, Heywood S, et al. The human gene Mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet. 2017;136(6):665–77.
31. Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. Am J Hum Genet [Internet]. 2017 Jan 5 [cited 2017 Apr 4];100(1):75–90. Available from http://linkinghub.elsevier.com/retrieve/pii/S0002929716305274
32. Ellingford JM, Barton S, Bhaskar S, O'Sullivan J, Williams SG, Lamb JA, et al. Molecular findings from 537 individuals with inherited retinal disease. J Med Genet [Internet]. 2016 May 11 [cited 2017 Sep 21];53(11):761–7. Available from http://jmg.bmj.com/lookup/doi/10.1136/jmedgenet-2016-103837
33. Bujakowska KM, Fernandez-Godino R, Place E, Cosugar M, Navarro-Gomez D, White J, et al. Copy number variation is an important contributor to the genetic causality of inherited retinal degenerations. Genet Med. 2017;19(6):643–51.
34. Corton M, Nishiguchi KM, Avila-Fernández A, Nikopoulos K, Riveiro-Alvarez R, Tatu SD, et al. Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. PLoS One, 2013; https://doi.org/10.1371/journal.pone.0065574.
35. Xu Y, Guan L, Shen T, Zhang J, Xiao X, Jiang H, et al. Mutations of 60 known causative genes in 157 families with retinitis pigmentosa based on exome sequencing. Hum Genet. 2014;133:1255–71.
36. Tiwari A, Lemke J, Altmueller J, Thiele H, Glaus E, Fleischhauer J, et al. Identification of novel and recurrent disease-causing mutations in retinal dystrophies using whole exome sequencing (WES): benefits and limitations. PLoS One. 2016;11(7):e0158692.
37. Lee K, Berg JS, Milko L, Crooks K, Lu M, Bizon C, et al. High diagnostic yield of whole exome sequencing in participants with retinal dystrophies in a clinical ophthalmology setting. Am J Ophthalmol. 2015;160(2):354–63.
38. Broadgate S, Yu J, Downes SM, Halford S. Unravelling the genetics of inherited retinal dystrophies: past, present and future. Prog Retin Eye Res. 2017;59:53–96.
39. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med. 2017 Feb;19(2):249–55.
40. Ormondroyd E, Mackley MP, Blair E, Craft J, Knight JC, Taylor JC, et al. "Not pathogenic until proven otherwise": perspectives of UK clinical genomics professionals toward secondary findings in context of a Genomic Medicine Multidisciplinary Team and the 100,000 Genomes Project. Genet Med. 2018;20(3):320–8.
41. Mackley MP, Capps B. Expect the unexpected: screening for secondary findings in clinical genomics research. Br Med Bull. 2017;122(1):109–22.
42. Mackley MP, Fletcher B, Parker M, Watkins H, Ormondroyd E. Stakeholder views on secondary findings in whole-genome and whole-exome sequencing: a systematic review of quantitative and qualitative studies. Genet Med. 2017;19(3):283–93.
43. LaDuca H, Farwell KD, Vuong H, Lu H-M, Mu W, Shahmirzadi L, et al. Exome sequencing cov-

ers >98% of mutations identified on targeted next generation sequencing panels. PLoS One. 2017;12(2):e0170843.
44. Lee K, Garg S. Navigating the current landscape of clinical genetic testing for inherited retinal dystrophies. Genet Med. 2015;17(4):245–52.
45. Alfares A, Aloraini T, Al Subaie L, Alissa A, Al Qudsi A, Alahmad A, et al. Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. Genet Med. 2018; https://doi.org/10.1038/gim.2018.41
46. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Hum Genet. 2016;135(3):359–62.
47. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci USA. 2015;112(17):5473–8.
48. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. Hum Mutat [Internet]. 2015 Aug [cited 2016 Oct 4];36(8):815–22. Available from. http://www.ncbi.nlm.nih.gov/pubmed/25973577
49. Ellingford JM, Thomas HB, Rowlands C, Arno G, Beaman G, Gomes-Silva B, et al. Functional and in-silico interrogation of rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. bioRxiv. 2019; https://doi.org/10.1101/781088.
50. Sangermano R, Bax NM, Bauwens M, van den Born LI, De Baere E, Garanto A, et al. Photoreceptor progenitor mRNA analysis reveals exon skipping resulting from the ABCA4 c.5461-10T→C mutation in Stargardt disease. Ophthalmology. 2016 Jun;123(6):1375–85.
51. Sangermano R, Khan M, Cornelis SS, Richelle V, Albert S, Garanto A, et al. *ABCA4* midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. Genome Res [Internet]. 2018 Jan [cited 2018 Jul 24];28(1):100–10. Available from http://www.ncbi.nlm.nih.gov/pubmed/29162642
52. Sangermano R, Garanto A, Khan M, Runhart EH, Bauwens M, Bax NM, et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. Genet Med. 2019;21(8):1751–60.
53. van den Hurk JAJM, van de Pol DJR, Wissinger B, van Driel MA, Hoefsloot LH, de Wijs IJ, et al. Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon. Hum Genet [Internet]. 2003 Aug [cited 2016 Feb 2];113(3):268–75. Available from http://www.ncbi.nlm.nih.gov/pubmed/12827496
54. Vaché C, Besnard T, Le Berre P, García-García G, Baux D, Larrieu L, et al. Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy. Hum Mutat. 2011;33(1):104–8.
55. Mayer AK, Rohrschneider K, Strom TM, Glöckle N, Kohl S, Wissinger B, et al. Homozygosity mapping and whole-genome sequencing reveals a deep intronic PROM1 mutation causing cone-rod dystrophy by pseudoexon activation. Eur J Hum Genet [Internet]. 2015 Jul 8 [cited 2016 Feb 2]; Available from http://www.ncbi.nlm.nih.gov/pubmed/26153215
56. den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, Voesenek KEJ, et al. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. Am J Hum Genet. 2006;79(September):556–61.
57. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics [Internet]. 2012 Nov 1 [cited 2016 Feb 11];28(21):2747–54. Available from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3476336&tool=pmcentrez&rendertype=abstract
58. Cipriani V, Silva RS, Arno G, Pontikos N, Kalhoro A, Valeina S, et al. Duplication events downstream of IRX1 cause North Carolina macular dystrophy at the MCDR3 locus. Sci Rep. 2017;7(1):7512.
59. Sanchis-Juan A, Stephens J, French CE, Gleadall N, Mégy K, Penkett C, et al. Complex structural variants resolved by short-read and long-read whole genome sequencing in mendelian disorders. bioRxiv [Internet]. 2018 Mar 14 [cited 2018 Apr 5];281683. Available from https://www.biorxiv.org/content/early/2018/03/14/281683
60. Pusch CM, Broghammer M, Jurklies B, Besch D, Jacobi FK. Ten novel ORF15 mutations confirm mutational hot spot in the RPGR gene in European patients with X-linked retinitis pigmentosa. Hum Mutat. 2002;20(5):405.
61. Ardui S, Ameur A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res. 2018;46(5):2159–68.
62. Caulfield M, Davies J, Dennys M, Elbahy L, Fowler T, Hill S, et al. The 100,000 genomes project protocol. Genomics Engl. 2017; https://doi.org/10.6084/m9.figshare.4530893.v4.
63. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, et al The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ. 2018;361:k1687. https://doi.org/10.1136/bmj.k1687.
64. Schwarze K, Buchanan J, Taylor JC, Wordsworth S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet Med. 2018;20(10):1122–30.

65. Manolio TA, Collins FS, Cox NJ. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53.
66. Barwell JG, O'Sullivan RB, Mansbridge LK, Lowry JM, Dorkins HR. Challenges in implementing genomic medicine: the 100,000 genomes project. J Transl Genet Genomics. 2018;2:13.
67. Taylor RL, Arno G, Poulter JA, Khan KN, Morarji J, Hull S, et al. Association of steroid 5α-reductase type 3 congenital disorder of glycosylation with early-onset retinal dystrophy. JAMA Ophthalmol. 2017;135(4):339–47.
68. Fiorentino A, Fujinami K, Arno G, Robson AG, Pontikos N, Arasanz Armengol M, et al. Missense variants in the X-linked gene *PRPS1* cause retinal degeneration in females. Hum Mutat [Internet]. 2018 Jan [cited 2018 Oct 30];39(1):80–91. Available from http://doi.wiley.com/10.1002/humu.23349
69. Radziwon A, Arno GK, Wheaton D, EM MD, Baple EL, Webb-Jones K, et al. Single-base substitutions in the CHM promoter as a cause of choroideremia. Hum Mutat. 2017 Jun;38(6):704–15.
70. Eisenberger T, Neuhaus C, Khan AO, Decker C, Preising MN, Friedburg C, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. PLoS One [Internet]. 2013 Jan 12 [cited 2016 May 27];8(11):e78496. Available from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078496#pone-0078496-t001
71. Coppieters F, Todeschini AL, Fujimaki T, Baert A, De Bruyne M, Van Cauwenbergh C, et al. Hidden genetic variation in LCA9-associated congenital blindness explained by 5'UTR mutations and copy-number variations of NMNAT1. Hum Mutat [Internet]. 2015 Dec [cited 2017 Nov 24];36(12):1188–96. Available from http://doi.wiley.com/10.1002/humu.22899
72. Radziwon A, Arno GK, Wheaton D, McDonagh EM, Baple EL, Webb-Jones K, et al. Single-base substitutions in the CHM promoter as a cause of choroideremia. Hum Mutat [Internet]. 2017 Mar 8 [cited 2017 Apr 4]; Available from http://doi.wiley.com/10.1002/humu.23212
73. Hull S, Owen N, Islam F, Tracey-White D, Plagnol V, Holder GE, et al. Nonsyndromic retinal dystrophy due to bi-allelic mutations in the Ciliary transport gene *IFT140*. Investig Opthalmology Vis Sci. 2016;57(3):1053.
74. Nguyen T-M, Hull S, Roepman R, van den Born LI, Oud MM, de Vrieze E, et al. Missense mutations in the WD40 domain of AHI1 cause non-syndromic retinitis pigmentosa. J Med Genet. 2017;54(9):624–32.
75. Estrada-Cuzcano A, Koenekoop RK, Senechal A, De Baere EBW, de Ravel T, Banfi S, et al. BBS1 mutations in a wide spectrum of phenotypes ranging from nonsyndromic retinitis pigmentosa to Bardet-Biedl syndrome. Arch Ophthalmol (Chicago, Ill 1960) [Internet]. 2012 Nov [cited 2015 Dec 1];130(11):1425–32. Available from. http://www.ncbi.nlm.nih.gov/pubmed/23143442
76. Xu M, Yang L, Wang F, Li H, Wang X, Wang W, et al. Mutations in human IFT140 cause non-syndromic retinal degeneration. Hum Genet. 2015;134(10):1069–78.
77. Haer-Wigman L, Newman H, Leibu R, Bax NM, Baris HN, Rizel L, et al. Non-syndromic retinitis pigmentosa due to mutations in the mucopolysaccharidosis type IIIC gene, heparan-alpha-glucosaminide N-acetyltransferase (HGSNAT). Hum Mol Genet. 2015;24(13):3742–51.
78. Ku CA, Hull S, Arno G, Vincent A, Carss K, Kayton R, et al. Detailed clinical phenotype and molecular genetic findings in CLN3-associated isolated retinal degeneration. JAMA Ophthalmol. 2017;135(7):749–60.
79. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of chromosomal domains by loop extrusion. Cell Rep. 2016 May;15(9):2038–49.
80. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015 May;161(5):1012–25.
81. Lupiáñez DG, Spielmann M, Mundlos S. Breaking TADs: how alterations of chromatin domains result in disease. Trends Genet. 2016 Apr;32(4):225–37.
82. Kragesteen BK, Spielmann M, Paliou C, Heinrich V, Schöpflin R, Esposito A, et al. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat Genet. 2018 Oct;50(10):1463–73.
83. Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. Nature. 2018 Feb;554(7691):239–43.
84. Aldiri I, Valentine M, Shao Y, Xu B, Shirinifard A, Easton J, et al. Linking the 3D genome structure to gene expression presented in the mammalian retina, presented at ARVO 2018 meeting in Honolulu, Hi. In: Investigative Ophthalmology & Visual Science. presented at ARVO 2018 meeting in Honolulu, HI: [presented at ARVO 2018 meeting in Honolulu, Hi; Association for Research in Vision and Ophthalmology, etc.]; 2018. pp. 3495–3495.
85. Silva RS, Kraft K, Arno G, Heinrich V, Pontikos N, Cipriani V, et al. CRISPR-derived mouse model of North Carolina Macular Dystrophy reveals in trans tissue-specific upregulation of PRDM13, presented at ARVO 2018 meeting in Honolulu, Hi. In: Investigative Ophthalmology & Visual Science. presented at ARVO 2018 meeting in Honolulu, HI: [presented at ARVO 2018 meeting in Honolulu,Hi; Association for Research in Vision and Ophthalmology, etc.]; 2018. pp. 6026–6026.
86. Aldiri I, Xu B, Wang L, Chen X, Hiler D, Griffiths L, et al. The dynamic epigenetic landscape of the retina

during development, reprogramming, and tumorigenesis. Neuron [Internet]. 2017 May 3 [cited 2018 Apr 5]; 94(3):550–568.e10. Available from http://linkinghub.elsevier.com/retrieve/pii/S0896627317303483
87. Aldiri I, Valentine M, Xu B, Putnam D, Griffiths L, Lupo M, et al. The nucleome of developing murine rod photoreceptors. bioRxiv. 2018; https://doi.org/10.1101/369702.
88. Cherry TJ, Yang MG, Harmin DA, Tao P, Timms AE, Bauwens M, et al. Epigenomic profiling and single-nucleus-RNA-Seq reveal cis-regulatory elements in human retina, macula and RPE and non-coding genetic variation. bioRxiv. 2018 Sep; https://doi.org/10.1101/412361.
89. Zelinger L, Karakülah G, Chaitankar V, Kim J-W, Yang H-J, Brooks MJ, et al. Regulation of noncoding transcriptome in developing photoreceptors by rod differentiation factor NRL. Investig Opthalmology Vis Sci. 2017 Sep;58(11):4422.
90. Zelinger L, Chaitankar V, Minehart J, Aponte A, Swaroop A. Identification of putative small open reading frame peptides in the retinal transcriptome. Vol. 59, Investigative Ophthalmology & Visual Science [Association for Research in Vision and Ophthalmology, etc.]; 2018. pp. 6023–6023.
91. Zhao L, Chen Y, Bajaj AO, Eblimit A, Xu M, Soens ZT, et al. Integrative subcellular proteomic analysis allows accurate prediction of human disease-causing genes. Genome Res. 2016;26(5):660–9.
92. Small KW, DeLuca AP, Whitmore SS, Rosenberg T, Silva-Garcia R, Udar N, et al. North Carolina macular dystrophy is caused by dysregulation of the retinal transcription factor PRDM13. Ophthalmology [Internet]. 2016 Jan [cited 2017 Jan 27];123(1):9–18. Available from http://www.ncbi.nlm.nih.gov/pubmed/26507665
93. Liskova P, Dudakova L, Evans CJ, Rojas Lopez KE, Pontikos N, Athanasiou D, et al. Ectopic GRHL2 expression due to non-coding mutations promotes cell state transition and causes posterior polymorphous corneal dystrophy 4. Am J Hum Genet. 2018;102(3):447–59.
94. Cipriani V, Silva RS, Arno G, Pontikos N, Kalhoro A, Valeina S, et al. Duplication events downstream of IRX1 cause North Carolina macular dystrophy at the MCDR3 locus. Sci Rep. 2017 Dec;7(1):7512.
95. Silva RS, Arno G, Cipriani V, Pontikos N, Defoort-Dhellemmes S, Kalhoro A, et al. Unique noncoding variants upstream of *PRDM13* are associated with a spectrum of developmental retinal dystrophies including progressive bifocal chorioretinal atrophy. Hum Mutat. 2019 Feb;40(5):578–87.
96. Michaelides M, Jeffery G, Moore A. Developmental macular disorders: phenotypes and underlying molecular genetic basis. Br J Ophthalmol. 2012;96:917–24.
97. Michaelides M, Johnson S, Tekriwal AK, Holder GE, Bellmann C, Kinning E, et al. An early-onset autosomal dominant macular dystrophy (MCDR3) resembling North Carolina macular dystrophy maps to chromosome 5. Investig Ophthalmol Vis Sci. 2003;44(5):2178–83.
98. Yang Z, Tong Z, Chorich LJ, Pearson E, Yang X, Moore A, et al. Clinical characterization and genetic mapping of North Carolina macular dystrophy. Vision Research. 2008;48:470–7.
99. Small KW, Weber JL, Roses A, Lennon F, Vance JM, Pericak-Vance MA. North Carolina macular dystrophy is assigned to chromosome 6. Genomics. 1992;13(3):681–5.
100. Kelsell RE, Godley BF, Evans K, Tiffin PAC, Gregory CY, Plant C, et al. Localization of the gene for progressive bifocal chorioretinal atrophy (PBCRA) to chromosome 6q. Hum Mol Genet. 1995;4(9):1653–6.
101. Bowne SJ, Sullivan LS, Wheaton DK, Locke KG, Jones KD, Koboldt DC, et al. North Carolina macular dystrophy (MCDR1) caused by a novel tandem duplication of the PRDM13 gene. Mol Vis. 2016;22:1239–47.
102. Manes G, Joly W, Guignard T, Smirnov V, Berthemy S, Bocquet B, et al. A novel duplication of PRMD13 causes North Carolina macular dystrophy: overexpression of PRDM13 orthologue in drosophila eye reproduces the human phenotype. Hum Mol Genet. 2017 Nov;26(22):4367–74.
103. Hafford-Tear NJ, Tsai Y-C, Sadan AN, Sanchez-Pintado B, Zarouchlioti C, Maher GJ, et al. CRISPR/Cas9-targeted enrichment and long-read sequencing of the Fuchs endothelial corneal dystrophy–associated TCF4 triplet repeat. Genet Med. 2019 Feb;21(9):2092–102.
104. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using Nanoliter droplets. Cell. 2015 May;161(5):1202–14.
105. Lu Y, Yi W, Wu Q, Zhong S, Zuo Z, Zhao F, et al. Single-cell RNA-seq analysis maps the development of human fetal retina. bioRxiv. 2018 Sep;17(7):e3000365.
106. Peng Y-R, Shekhar K, Yan W, Herrmann D, Sappington A, Bryman GS, et al. Molecular classification and comparative taxonomics of foveal and peripheral cells in primate retina. Cell. 2019 Jan;176(5):1222–1237.e22.
107. Lukowski S, Lo C, Sharov A, Nguyen Q, Fang L, Hung S, et al. Generation of human neural retina transcriptome atlas by single cell RNA sequencing. bioRxiv. 2018 Sep; https://doi.org/10.1101/425223.
108. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet (London, England). 2015 Apr;385(9975):1305–14.
109. Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, et al. Modified penetrance of coding variants by cis-regulatory varia-

tion contributes to disease risk. Nat Genet. 2018 Sep;50(9):1327–34.
110. Barbosa M, Joshi RS, Garg P, Martin-Trujillo A, Patel N, Jadhav B, et al. Identification of rare de novo epigenetic variations in congenital disorders. Nat Commun. 2018 Dec;9(1):2064.
111. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, et al. Post-zygotic point mutations are an underrecognized source of De novo genomic variation. Am J Hum Genet. 2015 Jul;97(1):67–74.
112. Smedley D, Schubach M, Jacobsen JOB, Köhler S, Zemojtel T, Spielmann M, et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in mendelian disease. Am J Hum Genet [Internet]. 2016 Sep 1 [cited 2017 Jan 27];99(3):595–606. Available from http://linkinghub.elsevier.com/retrieve/pii/S0002929716302786
113. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. Nature. 2014 Sep;513(7516):120–3.
114. Wan A, Place E, Pierce EA, Comander J. Characterizing variants of unknown significance in rhodopsin: a functional genomics approach. bioRxiv. 2019 Jan;40(8):1127–44.
115. Zarouchlioti C, Sanchez-Pintado B, Hafford Tear NJ, Klein P, Liskova P, Dulla K, et al. Antisense therapy for a common corneal dystrophy ameliorates TCF4 repeat expansion-mediated toxicity. Am J Hum Genet. 2018;102(4):528–39.
116. Aronson AR. Effective mapping of biomedical text to the UMLS metathesaurus: the MetaMap program. Proc AMIA Symp. 2001;2001:17–21.
117. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res. 2017 Jan;45(D1):D865–76.
118. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the exomiser. Nat Protoc. 2015 Dec;10(12):2004–15.
119. Greene D, Richardson S, Turro E. A fast association test for identifying pathogenic variants involved in rare diseases. Am J Hum Genet. 2017 Jun;101:104–14.
120. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al. PhenoTips: patient phenotyping software for clinical and research use. Hum Mutat. 2013 Aug;34(8):1057–65.
121. Pontikos N, Wagner SK, Balaskas K, Keane PA, Faes L, Kortuem K, et al. Trends in retina specialist imaging utilization from 2012 to 2016 in the United States Medicare fee-for-service population. Am J Ophthalmol. 2019;208:12–8.
122. López-Rubio S, Chacon-Camacho OF, Matsui R, Guadarrama-Vallejo D, Astiazarán MC, Zenteno JC. Retinal phenotypic characterization of patients with ABCA4 retinopathy due to the homozygous p.Ala1773Val mutation. Mol Vis. 2018 Feb;24:105–14.
123. Lambertus S, Lindner M, Bax NM, Mauschitz MM, Nadal J, Schmid M, et al. Progression of late-onset Stargardt disease. Invest Ophthalmol Vis Sci. 2016 Oct;57(13):5186–91.
124. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018 Sep;24(9):1342–50.
125. Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. Biomed Opt Express. 2017 Jul;8(7):3440–8.
126. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018 Feb;172(5):1122–1131.e9.
127. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. Ophthalmol Retin. 2017;1(4):322–7.
128. Fujinami-Yokokawa Y, Pontikos N, Yang L, Tsunoda K, Yoshitake K, Iwata T, et al. Prediction of causative genes in inherited retinal disorders from spectral-domain optical coherence tomography utilizing deep learning techniques. J Ophthalmol. 2019;2019:1691064.
129. Topol E. The Topol Review: preparing the healthcare workforce to deliver the digital future. 2019. https://topol.hee.nhs.uk