

SCIENTIFIC REPORTS



OPEN

Molecular genetic analysis using targeted NGS analysis of 677 individuals with retinal dystrophy

Cathrine Jespersgaard¹, Mingyan Fang^{2,3,4}, Mette Bertelsen^{1,5}, Xiao Dang^{2,3}, Hanne Jensen⁵, Yulan Chen^{2,3}, Niels Bech⁵, Lanlan Dai^{2,3}, Thomas Rosenberg⁵, Jianguo Zhang^{2,3}, Lisbeth Birk Møller¹, Zeynep Tümer^{1,6}, Karen Brøndum-Nielsen¹ & Karen Grønskov¹

Inherited retinal diseases (IRDs) are a common cause of visual impairment. IRD covers a set of genetically highly heterogeneous disorders with more than 150 genes associated with one or more clinical forms of IRD. Molecular genetic diagnosis has become increasingly important especially due to expanding number of gene therapy strategies under development. Next generation sequencing (NGS) of gene panels has proven a valuable diagnostic tool in IRD. We present the molecular findings of 677 individuals, residing in Denmark, with IRD and report 806 variants of which 187 are novel. We found that deletions and duplications spanning one or more exons can explain 3% of the cases, and thus copy number variation (CNV) analysis is important in molecular genetic diagnostics of IRD. Seven percent of the individuals have variants classified as pathogenic or likely-pathogenic in more than one gene. Possible Danish founder variants in *EYS* and *RP1* are reported. A significant number of variants were classified as variants with unknown significance; reporting of these will hopefully contribute to the elucidation of the actual clinical consequence making the classification less troublesome in the future. In conclusion, this study underlines the relevance of performing targeted sequencing of IRD including CNV analysis as well as the importance of interaction with clinical diagnoses.

The introduction of next generation sequencing (NGS) has improved molecular genetic diagnosis of genetically heterogeneous conditions substantially. This is also true for inherited retinal diseases (IRDs) which are associated with sequence variations in more than 150 genes (<https://sph.uth.edu/retnet/>). IRDs encompass a heterogeneous group of retinal disorders with degeneration of photoreceptors, and are considered one of the leading causes of visual impairment in children and young individuals¹. Retinitis pigmentosa (RP) is the most frequent clinical diagnosis of the IRD specific diagnostic subgroups with a worldwide prevalence of 1:4000². Diagnoses include both syndromic and non-syndromic conditions. IRD can be inherited as autosomal recessive (AR), autosomal dominant (AD), X-linked (XL) and mitochondrial traits; furthermore, digenic inheritance has been reported in rare cases³.

A genetic diagnosis is not only important to improve the genetic counselling of the families and for prognostic reasons but also for identifying patients that might benefit from new candidate treatments such as gene therapy, where promising results are being reported in numerous clinical trials worldwide⁴.

In this study, we present genetic investigation of a cohort of 677 individuals residing in Denmark and clinically diagnosed or suspected with IRD. Coding regions together with 20 bp flanking intronic sequence and untranslated regions (UTRs) of 125 genes previously shown to be associated with IRD, were analyzed using panel-based NGS.

¹Kennedy Center, Department of Clinical Genetics, Rigshospitalet, University of Copenhagen, Glostrup, DK2600, Denmark. ²BGI-Shenzhen, Shenzhen, 518083, China. ³China National GeneBank, BGI-Shenzhen, Shenzhen, 518120, China. ⁴Department of Laboratory Medicine, Division of Clinical Immunology and Transfusion Medicine, Karolinska University Hospital Huddinge, Stockholm, S141 86, Sweden. ⁵Department of Ophthalmology, Rigshospitalet-Glostrup, University of Copenhagen, Glostrup, DK2600, Denmark. ⁶Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. Cathrine Jespersgaard and Mingyan Fang contributed equally. Correspondence and requests for materials should be addressed to K.G. (email: karen.groenskov@regionh.dk)

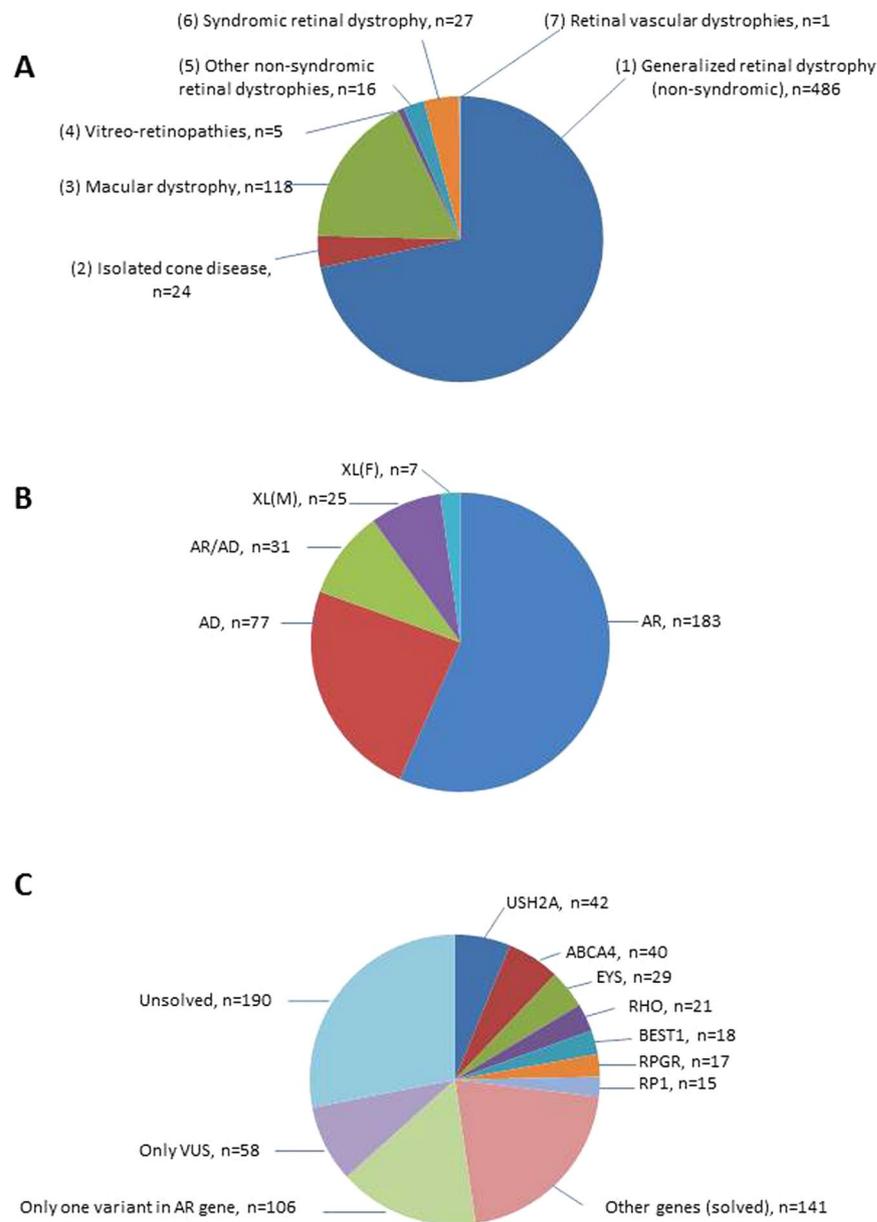


Figure 1. (A) Distribution of 677 individuals in seven groups based on clinical diagnosis. The number in brackets refer to the clinical group, and n = number of individuals. (B) Inheritance based on genetic findings in 323 individuals with a molecular genetic diagnosis; n = number of individuals. (C) Mutational spectrum of variants identified in 677 individuals with IRDs; n = number of individuals.

We report all variants that could be potentially disease causing; it is noteworthy that the number of rare variants per individual requires meticulous scrutiny combining several tools; and a close collaboration between clinical and laboratory medicine is mandatory.

Results

Patient characteristics. Patients clinically suspected of having a retinal dystrophy were selected for the targeted analysis based on evaluation of available clinical records (clinical diagnosis, clinical history, fundus changes, OCT imaging, ERG, visual acuity, visual fields, family history). A total of 677 individuals were investigated by a targeted NGS panel consisting of 125 genes. Samples have been collected to a clinical biobank over many years. The clinical diagnoses were divided into seven groups (Table S1, Fig. 1A). The group “Generalized retinal dystrophies” which includes RP and cone-rod dystrophies were by far the most common clinical diagnosis (72%), followed by “Macular dystrophy” which includes Stargardt disease (17%)⁵. The distribution of clinical diagnoses is biased due to several previous research projects of specific diagnostic subgroups and diagnostic efforts during the years. Individuals with one of following clinical diagnoses: Stargardt disease⁶, achromatopsia⁷, Usher syndrome⁸, Bardet-Biedl syndrome⁹, Best vitelliform macular dystrophy¹⁰, choroideremia¹¹, Leber congenital amaurosis¹²,

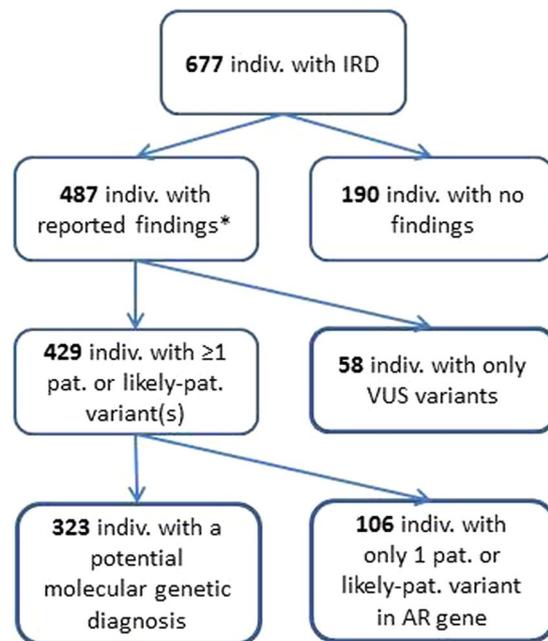


Figure 2. Flow of individuals. A schematic representation showing the outcome of the 677 individuals with IRD participating in the study. *Supplementary Table 4. Indiv: individuals.

XL and AD retinitis pigmentosa¹³, congenital stationary night blindness¹⁴, cone dystrophy and Åland eye disease¹⁵ were included in previous research projects, and are therefore represented by lower numbers than might be expected.

NGS panel. The 125 genes in the panel are known to be associated with: retinitis pigmentosa (41%), Bardet-Biedl syndrome (14%), cone-rod dystrophy (12%), Leber congenital amaurosis (10%), Usher syndrome (7%) and congenital stationary night blindness (3%) (some genes cause more than one eye disorder); Other genetic retinal diseases account for a small proportion in the panel (Table S2). 70% of the genes are associated with an AR inheritance pattern, 20% with an AD inheritance pattern and 5% with a XL inheritance pattern; furthermore, two genes (*PRPH2* and *ROM1*) have been reported to be involved in digenic inheritance (Table S2).

We obtained on average 1156.22 Mb per individual that mapped to the target region. The mean depth of the target regions of all samples was 533X (range from 181X to 1587X) and the average coverage of at least 4X and 20X was 98.27% and 97.29%, respectively (Table S3).

Variants were filtered based on quality, frequency and function. A total of 7440 rare variants (4030 SNVs and 3410 indels) remained with a mean number of 9.24 rare variants per individual (range from 2 to 29). The reference sequence is GRCh37/hg19 and the exons are numbered according to reference sequences as listed in Table S2.

Sequence variations. We report all variants considered to be potentially disease causing, including variants classified as VUS (variant of unknown clinical significance) when it was estimated that they could affect the phenotype (Table S4). Variants were classified according to the American College of Medical Genetics and Genomics (ACMG) classification. A total of 806 variants, 532 unique variants, are reported; of these 187 were novel at the time of manuscript submission. Potential disease causing variants were found in 487 individuals while in 190 individuals we found no variants to report (flow of individuals are shown in Fig. 2). Pathogenic or likely-pathogenic variants was identified in 429 out of 677 individuals (63%), while 58 individuals out of 677 (9%) had either one variant classified as VUS in a gene with AD or XL inheritance, or two variants classified as VUS in a gene with AR inheritance. In 323 individuals (48%) a plausible molecular genetic diagnosis could be established while in 106 individuals (16%) only one pathogenic or likely-pathogenic variant in an AR gene was identified. The study revealed unprecedented complexity in the interpretation of the sequence variants as 7% (47/677) of individuals harbored pathogenic or likely-pathogenic variants in more than one gene.

The sequence variations are distributed across 81 genes; 59 genes with AR inheritance ($n = 183$), 14 genes with AD inheritance ($n = 77$), four with XL inheritance ($n = 32$) and four genes known with both AR and AD inheritance ($n = 31$) (Fig. 1B). Numbers in brackets indicate the number of individuals. Of the 323 individuals with a plausible molecular genetic diagnosis, the most frequently mutated genes were *USH2A* ($n = 42$, 6%), *ABCA4* ($n = 40$, 6%), *EYS* ($n = 29$, 4%), *RHO* ($n = 21$, 3%), *BEST1* ($n = 18$, 3%), *RPGR* ($n = 17$, 2%) and *RPI* ($n = 15$, 2%). Variants in these seven genes could explain more than half of the solved cases (182 out of 323, 56%) (Fig. 1C). Variants in 22 genes solved only one or two cases each, demonstrating the necessity of a large gene panel for IRD analysis. Furthermore, in 47 individuals, sequence variations classified as either pathogenic or likely-pathogenic were detected in more than one gene (Table S4).

CNV analysis was performed using NGS data, and indications of a deletion or duplication were confirmed with another method (MLPA (Multiplex Ligation-dependent Probe Amplification), qPCR (quantitative PCR analysis) or chromosome microarray). Furthermore, all samples heterozygous for a pathogenic or likely-pathogenic sequence variant in *ABCA4*, *USH2A* or *EYS* were analyzed by MLPA to search for deletions and duplications which revealed duplications in *EYS* (case 126) and *USH2A* (case 297). In total, this strategy detected 17 deletions and seven duplications resulting in a molecular genetic diagnosis in 17 cases (Table S5).

Danish disease associated variants and other disease associated founder variants. A novel disease associated variant, c.232del in *EYS*, was detected in 11 unrelated individuals. Using the surrounding SNPs in *EYS* we detected a common haplotype spanning 1.9 Mb (chr6:64,431,148-66,415,690, GRCh37/hg19). Likewise, a novel disease associated variant, c.2690_2695del in *RPI1*, was detected in five unrelated individuals. Analysis of the surrounding SNPs demonstrated a common haplotype spanning 14 kb (chr8:55,528,953-55,543,161, GRCh37/hg19) (Table S6).

Discussion

Targeted NGS analysis is a valuable method for molecular genetic diagnostics of IRDs as also supported by several previous studies^{16–18}. Our detection rate of 48% is relatively low compared to other studies; however, this may be biased as our study population was derived from clinical biobank samples that had been extensively investigated over the years in several research projects and diagnostics efforts, resulting in many solved cases not included in the present study.

A substantial number of individuals were found with only one pathogenic or likely-pathogenic sequence variant in a gene with AR inheritance and clinical symptoms that support the genetic diagnosis. This type of incomplete genetic diagnosis (missing heritability) is a well-known phenomenon. The missing genetic alterations might be found in regulatory regions or deep intron sequences, but also frequent hypomorph variants which are filtered out during data analysis or combinations of such in a haplotype could be an explanation of the missing heritability. Furthermore, given the high frequency of AR inherited retinal dystrophy, it is of course possible that a yet unknown cause is responsible for the eye disorder. Digenic inheritance has also been observed for retinal dystrophies, however, we did not observe any obvious signs of this. Whole genome sequencing combined with segregation analysis and functional studies could solve some of these cases in future studies.

In 9% of individuals, variants classified as VUS (either one in a gene with AD or XL inheritance, or two in a gene with AR inheritance) were identified where the clinical symptoms were supportive. These variants are also reported in this study since the classification of variants may change over time, and ideally, variants should be classified as either pathogenic (disease associated) or benign; but currently many variants are classified as VUS and the only way forward is to report the variants and combine studies.

The most genetically heterogeneous group was the RP group in which causative variations were found in 32 genes, all of which are known to be associated with RP. The diagnosis Stargardt disease was almost exclusively associated with variants in *ABCA4*. However, we also found one individual with p.(Arg373Cys) missense variant in *PROM1* which has previously been reported to be associated with Stargardt disease. Notably, we found two individuals with truncating variants in *CRX*. Recently, a nonsense variant in *CRX* was found in an individual with adult onset macular dystrophy¹⁹.

CNV analysis revealed a total of 24 deletions and duplications. In some cases their pathogenicity was questionable, while in others they were plausible explanations. It is obvious that CNV analysis is important in molecular genetic diagnosis of IRD, both shown in the present study and by others in previous studies²⁰.

Several individuals had variants in more than one gene. Six individuals (cases 53, 135, 138, 180, 194 and 202) had sequence variations in two genes that potentially both could be the cause of IRD (Table S4). In case 53 (a male), a pathogenic change in *BEST1* explains the Best vitelliform macular dystrophy phenotype, however, a truncating variant in *RPGR* would also be expected to be causative. Case 135 was found to be compound heterozygote for variants in both *EYS* and *CDH23*; he was diagnosed clinically with RP and had a mild hearing impairment. It is noteworthy though, that analysis restricted to genes known to be associated with Usher syndrome based on the clinical symptoms would have revealed only the *CDH23* variants while the *EYS* variants would have remained undetected. Case 138 with a clinical diagnosis of congenital stationary night blindness was homozygous for likely-pathogenic variants in both *GRK1* and *TRPM1*. Both genes are associated with congenital stationary night blindness with AR inheritance. Likewise, case 180 had variants in both *PRPF31* and *ELOVL4*; both genes are associated with RP with AD inheritance. Case 194 had deletions in two genes (*PRPF31* and *IMPDH1*); both genes show AD inheritance, and both genes are associated with RP. Case 202 had a clinical RP diagnosis and we detected a *PRPF31* splice variant and two *ABCA4* variants. Segregation analysis was unfortunately not possible in these cases.

In conclusion, this study shows that targeted NGS is an effective method for establishing a molecular genetic diagnosis of IRDs. CNV analysis should be part of the strategy. Using a large gene panel is a prerequisite since a similar phenotype can be caused by many genes, and variants can be present in more than one gene, which is important knowledge if the result is to be used in treatment purposes and for genetic counseling. We report sequence variants both from individuals who received what is considered to be a molecular genetic diagnosis with results consequently reported to the clinician, as well as from those receiving a partly molecular genetic diagnosis. This information is valuable in the further classification of variants, and in phenotype-genotype correlation studies.

Methods

Patients and clinical evaluation. The investigated cohort comprises 677 unrelated individuals residing in Denmark with a clinical diagnosis of IRD. Ages of the individuals ranged from 4 to 100 years with a mean age of 53.3 years and a median age of 55 years. 493 individuals without a previous molecular genetic diagnosis and with a DNA sample in the clinical biobank at Kennedy Center, Rigshospitalet, were identified from the Retinitis Pigmentosa Registry and from assessment of clinical records based on diagnosis codes related to any form of retinal dystrophy at the Department of Ophthalmology, Rigshospitalet. The Retinitis Pigmentosa Registry is a nation-wide register in which all individuals residing in Denmark with a diagnosis of a generalized retinal dystrophy born after 1850 has been registered^{5,21}. Furthermore, 184 individuals were enrolled during the study. Most participants were of Danish origin. Unfortunately, due to the fact that samples have been collected over many years, segregation analysis was in most cases not possible.

The project was approved by The National Committee on Health Research Ethics (Denmark). The project was performed according to the Declaration of Helsinki and approved by the Regional Ethics Committee. Written informed consent was obtained from all the participants and if under the age of 18 from a parent or legal guardian, before the molecular genetic testing.

Gene selection and target enrichment. We selected 125 genes that were reported to be associated with inherited retinal disorders according to RetNet database (<https://sph.uth.edu/retnet/>), OMIM (the Online Mendelian Inheritance in Man, <http://www.omim.org/>) and literature searching (Table S2). The custom NimbleGen SeqCap Target Enrichment System (NimbleGen, Madison, WI, USA) was designed to capture and enrich all coding exons, 5-/3-UTR regions and 20 bp flanking intronic regions. The size of the target region was 2.07 Mb and the coverage rate was 95%.

Individuals who were heterozygous for an *ABCA4* or *USH2A* pathogenic or likely-pathogenic variant were analyzed for deep intron variants reported to be associated with disease. For investigation of *ABCA4* intronic regions, chr1:94,483,922-94,484,082 and chr1:94,493,000-94,492,973 were amplified using primer set *ABCA4-V1-3-FH* acccactgcttactggcttaccGGATCATTATGACATCAACCCC and *ABCA4-V1-3-RH* gaggggcaacaacagatggcCTCCATAGGCTCAGAGATCCC and primer set *ABCA4-V4-5-FH* acccactgcttactggcttaccACACCATGTAGGTAGGCTTGG and *ABCA4-V4-5-RH* gaggggcaacaacagatggcAGGGATCCCAAAA-GAAGGAC²² respectively, followed by Sanger sequencing. For *USH2A*, primer set *USH2A-intron40-FH* accctgcttactggcttaccAGCTTCCTCTCCAGAATCACA and *USH2A-intron40-RH* gaggggcaacaacagatggcG-GTTTTTCATCTGGGTCTTGCA were used for PCR amplification, followed by Sanger sequencing. Sequences in lower case letters were used as tags for Sanger sequencing. In addition all individuals with a clinical diagnosis of Leber congenital amaurosis were investigated for the *CEP290* c.2991 + 1655A > G variant using primer set *CEP290-intron26-FH* acccactgcttactggcttaccGGTTCAGGCCGTTCTCCT and *CEP290-intron26-RH* gaggggcaacaacagatggcCACATGGGAGTCACAGGGTA for PCR amplification, followed by Sanger sequencing.

Next-generation sequencing and bioinformatics. Genomic DNA was extracted from peripheral blood using standard protocols. The enriched DNA libraries were sequenced using the Illumina HiSeq 2000 (494 patients) or HiSeq 4000 (184 patients) platforms (Illumina, San Diego, CA, USA). To optimize cost-efficient sequencing of the custom target panels, a pilot study was performed, pooling libraries from 1, 2, 5 and 8 samples. After comparing the sequencing data quality, coverage rate, genotyping concordance, sequencing depth, mismatch rate, capture specificity and total detected variants, pooling of five libraries were the optimal strategy.

Raw sequencing image files and base-calling were processed with the Illumine Pipeline and raw paired-end low quality reads and adapter sequences were removed using the SOAPnuke software (<http://soap.genomics.org.cn/>). The remaining high-quality reads were aligned to the human reference genome (GRCh37/hg19) using Burrows–Wheeler Algorithm (BWA) version 0.7.10²³, with the MEM algorithm. The SAMtools (version 0.1.19)²⁴ was used to sort and index SAM/BAM files and the Picard (version 1.117, <http://broadinstitute.github.io/picard/>) was used to mark PCR-duplicates. Local realignment and base recalibration were performed using GATK (version 3.3-0)²⁵ and single nucleotide variants (SNVs) and insertions/deletions (InDels) were called with GATK HaplotypeCaller.

All identified variants were annotated using Ensembl's VEP (Variant Effect Predictor)²⁶ and variants were prioritized based on the following criteria: (1) variants not present in BGI (Beijing Genome Institute) in-house databases and with a minor allele frequency less than 1% using three public variant databases, including 1000 Genomes Project (KG, <http://www.1000genomes.org/>), Exome Variant Server (ESP, <http://evs.gs.washington.edu/EVS/>) and Exome Aggregation Consortium (ExAC; <http://exac.broadinstitute.org/>); there are known disease associated variants causing IRD which have a frequency above 1%, the data were screened for these separately; (2) kept all non-synonymous, small indels, frameshift, nonsense, or affect canonical splice-site donor/acceptor sites variants. Variants were verified by Sanger sequencing prior to return of results to the clinician.

Interpretation and classification of variants. Sequence variations were classified into five categories: pathogenic (class 5), likely-pathogenic (class 4), VUS (class 3), likely benign (class 2) and benign (class 1), according to the ACMG guidelines²⁷. Alamut Visual (Interactive Biosoftware, Rouen, France) was used to evaluate missense and splice variants. Alamut Visual includes the in silico tools Align GVD^{28,29}, SIFT³⁰, MutationTaster³¹ and PolyPhen2³² for missense variants and SpliceSiteFinder-like^{33,34}, MaxEntScan³⁵, GeneSplicer³⁶, NNSPLICE³⁷ and Human Splicing Finder³⁸ for splice variants.

Deletion and duplication analysis. For targeted NGS samples, copy number analysis was performed either using the R software package ExomeDepth (v1.0.7)³⁹ (494 samples) or using the VarSeq Copy number variation (CNV) software (184 samples) (Golden Helix, Montana, USA). The ExomeDepth uses read depth data

to call CNVs. For each tested individual, the ExomeDepth algorithm builds the most appropriate reference set from the BAM files of a group of samples and ranks the CNV calls by their confidence level. The Varsseq software generates a set of matched reference controls and the sample is compared to this set. A ratio and z-score is computed for each target region defined in the BED file. The z-scores measure the number of standard deviations that a sample's coverage is from the mean coverage of the reference set. CNVs were validated either by MLPA, qPCR or chromosome microarray. MLPA was performed using kits P151 and P152 (*ABCA4*), P361 and P362 (*USH2A*), P328 (*EYS*) and P235 (*PRPF31*), following the manufacturer's instructions (MRC-Holland, Amsterdam, the Netherlands). If no MLPA kit was available we performed qPCR using SYBR Power Green (Applied Biosystems, Foster City, CA, USA), 50 ng DNA and 12,5 pmol of each primer. Three primer sets were used to verify the aberrations and minimum two normal controls were included in the analysis. Real time PCR was run on a 7500 ABI SDS system (Applied Biosystems). Chromosome microarray was carried out using CytoScan HD array (Affymetrix, Santa Clara, CA, USA). Data analysis was performed using ChAS software (Affymetrix).

References

- Rahi, J. S. *et al.* Severe visual impairment and blindness in children in the uk. *The Lancet* **362**, 1359–1365 (2003).
- Hartong, D. T., Berson, E. L. & Dryja, T. P. Retinitis pigmentosa. *The Lancet* **368**, 1795–1809 (2006).
- Kajiwara, K., Berson, E. L. & Dryja, T. P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/rds and rom1 loci. *Science* **264**, 1604–1608 (1994).
- MacLaren, R. E. Gene therapy for retinal disease: what lies ahead. *Ophthalmologica* **234**, 1–5 (2015).
- Bertelsen, M., Jensen, H., Bregnhøj, J. F. & Rosenberg, T. Prevalence of generalized retinal dystrophy in denmark. *Ophthalmic epidemiology* **21**, 217–223 (2014).
- Bertelsen, M. *et al.* Generalized choriocapillaris dystrophy, a distinct phenotype in the spectrum of abca4-associated retinopathies. *Investigative ophthalmology & visual science* **55**, 2766–2776 (2014).
- Kohl, S. *et al.* Cnbg3 mutations account for 50% of all cases with autosomal recessive achromatopsia. *European Journal of Human Genetics* **13**, 302 (2005).
- Dad, S. *et al.* Usher syndrome in denmark: mutation spectrum and some clinical observations. *Molecular genetics & genomic medicine* **4**, 527–539 (2016).
- Hjortshøj, T. D. *et al.* Bardet-biedl syndrome in denmark—report of 13 novel sequence variations in six genes. *Human mutation* **31**, 429–436 (2010).
- Bitner, H., Schatz, P., Mizrahi-Meissonnier, L., Sharon, D. & Rosenberg, T. Frequency, genotype, and clinical spectrum of best vitelliform macular dystrophy: data from a national center in denmark. *American journal of ophthalmology* **154**, 403–412 (2012).
- Schwartz, M., Rosenberg, T., van den Hurk, J. A., van den Pol, D. J. & Cremers, F. P. Identification of mutations in danish choroideremia families. *Human mutation* **2**, 43–47 (1993).
- Astuti, G. D. *et al.* Comprehensive genotyping reveals rpe65 as the most frequently mutated gene in leber congenital amaurosis in denmark. *European Journal of Human Genetics* **24**, 1071 (2016).
- Neidhardt, J. *et al.* Identification of novel mutations in x-linked retinitis pigmentosa families and implications for diagnostic testing. *Molecular vision* **14**, 1081 (2008).
- Szabo, V., Kreienkamp, H.-J., Rosenberg, T. & Gal, A. P. gln200 glu, a putative constitutively active mutant of rod α -transducin (gnat1) in autosomal dominant congenital stationary night blindness. *Human mutation* **28**, 741–742 (2007).
- Hove, M. N. *et al.* Clinical characteristics, mutation spectrum, and prevalence of åland eye disease/incomplete congenital stationary night blindness in denmark. *Investigative ophthalmology & visual science* **57**, 6861–6869 (2016).
- Ellingford, J. M. *et al.* Assessment of the incorporation of cnv surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *Journal of Medical Genetics* **55**, 114–121 (2017).
- Bernardis, I. *et al.* Unravelling the complexity of inherited retinal dystrophies molecular testing: Added value of targeted next-generation sequencing. *Biomed Res Int* **2016**, 6341870 (2016).
- Khan, K. N. *et al.* Advanced diagnostic genetic testing in inherited retinal disease: experience from a single tertiary referral centre in the uk national health service. *Clin Genet* **91**, 38–45, <https://doi.org/10.1111/cge.12798> (2017).
- Griffith, J. F., DeBenedictis, M. J. & Traboulsi, E. I. A novel dominant crx mutation causes adult-onset macular dystrophy. *Ophthalmic Genet* **39**, 120–124, <https://doi.org/10.1080/13816810.2017.1373831> (2018).
- Eisenberger, T. *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* **8**, e78496 (2013).
- Haim, M. The epidemiology of retinitis pigmentosa in denmark. *Acta Ophthalmologica Scandinavica* **80**, 1–34 (2002).
- Bauwens, M. *et al.* An augmented abca4 screen targeting noncoding regions reveals a deep intronic founder variant in belgian stargardt patients. *Hum Mutat* **36**, 39–42, <https://doi.org/10.1002/humu.22716> (2015).
- Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Li, H. *et al.* The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–2079 (2009).
- McKenna, A. *et al.* The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome research* (2010).
- McLaren, W. *et al.* Deriving the consequences of genomic variants with the ensembl api and snp effect predictor. *Bioinformatics* **26**, 2069–2070 (2010).
- Richards, C. S. *et al.* Acmg recommendations for standards for interpretation and reporting of sequence variations: Revisions2007. *Genetics in Medicine* **10**, 294 (2008).
- Tavtigian, S. V. *et al.* Comprehensive statistical study of 452 brca1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* **43**, 295–305, <https://doi.org/10.1136/jmg.2005.033878> (2006).
- Mathe, E. *et al.* Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res* **34**, 1317–25, <https://doi.org/10.1093/nar/gkj518> (2006).
- Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. *Nature protocols* **4**, 1073 (2009).
- Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. Mutationtaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **11**, 361–2, <https://doi.org/10.1038/nmeth.2890> (2014).
- Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248–9, <https://doi.org/10.1038/nmeth0410-248> (2010).
- Zhang, M. Q. Statistical features of human exons and their flanking regions. *Hum Mol Genet* **7**, 919–32 (1998).
- Shapiro, M. B. & Senapathy, P. Rna splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15**, 7155–74 (1987).
- Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to rna splicing signals. *J Comput Biol* **11**, 377–94, <https://doi.org/10.1089/1066527041410418> (2004).

36. Pertea, M., Lin, X. & Salzberg, S. L. Genesplicer: a new computational method for splice site prediction. *Nucleic Acids Res* **29**, 1185–90 (2001).
37. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in genie. *J Comput Biol* **4**, 311–23, <https://doi.org/10.1089/cmb.1997.4.311> (1997).
38. Desmet, F. O. *et al.* Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**, e67, <https://doi.org/10.1093/nar/gkp215> (2009).
39. Pagnol, V. *et al.* A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **28**, 2747–2754 (2012).

Acknowledgements

We would like to thank the patients and their families who participated in this study. We thank Pia Skovgaard, Bodil Olsen, Judy Rasmussen and Anne Obling Madsen for technical assistance. We also thank all the ophthalmologists who have sent samples to the study and especially the project follow committee: Torben Hansen, Toke Bek, Torben Sørensen, Michael Larsen, Else Gade and Lotte Welinder. This work was funded by the Velux foundation grant 32700.

Author Contributions

K.G., L.B.M., Z.T. and K.B.N. conceived the experiments, M.F., X.D., Y.C., L.D. and J.Z. conducted the NGS sequencing and bioinformatic analysis, C.J. and K.G. performed additional bioinformatic analysis, interpretation and validation of data, M.B., H.J., T.R. and N.B. carried out the clinical investigation of the patients and collected samples, K.G. and C.J. drafted the manuscript and all authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-38007-2>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019