Received: 31 October 2018 | Revised: 31 March 2019 | Accepted: 8 April 2019

DOI: 10.1002/humu.23762

RESEARCH ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Characterizing variants of unknown significance in rhodopsin: A functional genomics approach

Aliete Wan | Emily Place | Eric A. Pierce | Jason Comander

Department of Ophthalmology, Ocular Genomics Institute, Berman-Gund Laboratory for the Study of Retinal Degenerations, Massachusetts Eye and Ear, Harvard Medical School, Boston, Massachusetts

Correspondence
Jason Comander, Department of Ophthalmology, Ocular Genomics Institute, Berman-Gund Laboratory for the Study of Retinal Degenerations, Massachusetts Eye and Ear, 243 Charles Street, Boston, MA 02114.
Email: jason_comander_public@meei.harvard.edu

Funding information
Research to Prevent Blindness, Grant/Award Number: Career Development Award; Massachusetts Lions Eye Research Fund; National Eye Institute, Grant/Award Numbers: K12 EY016335-10, P30 EY014104; Foundation Fighting Blindness, Grant/Award Number: Enhanced Career Development Award

Abstract

Characterizing the pathogenicity of DNA sequence variants of unknown significance (VUS) is a major bottleneck in human genetics, and is increasingly important in determining which patients with inherited retinal diseases could benefit from gene therapy. A library of 210 rhodopsin (*RHO*) variants from literature and in-house genetic diagnostic testing were created to efficiently detect pathogenic *RHO* variants that fail to express on the cell surface. This study, while focused on *RHO*, demonstrates a streamlined, generalizable method for detecting pathogenic VUS. A relatively simple next-generation sequencing-based readout was developed so that a flow cytometry-based assay could be performed simultaneously on all variants in a pooled format, without the need for barcodes or viral transduction. The resulting dataset characterized the surface expression of every *RHO* library variant with a high degree of reproducibility ($r^2 = 0.92–0.95$), recategorizing 37 variants. For example, three retinitis pigmentosa pedigrees were solved by identifying VUS which showed low expression levels (p.G18D, p.G101V, and p.P180T). Results were validated across multiple assays and correlated with clinical disease severity. This study presents a parallelized, higher-throughput cell-based assay for the functional characterization of VUS in RHO, and can be applied more broadly to other inherited retinal disease genes and other disorders.

KEYWORDS
functional genomics, inherited retinal diseases, next-generation sequencing, retinitis pigmentosa, rhodopsin, variant library, variant pathogenicity, variant of unknown significance, VUS

1 | INTRODUCTION

With the increasing availability and use of human DNA sequencing, the problem of accurately characterizing newly-discovered DNA variants has become a major issue in human genetics (Lappalainen, 2015; Richards et al., 2015). Difficulties in predicting the pathogenicity of DNA "variants of unknown significance" (VUS), even with all available bioinformatic, functional, and human data, routinely produce an ambiguous final result of genetic testing (Bean, Tinker, da Silva, & Hegde, 2013; Davies et al., 2012; Richards et al., 2015). Dealing with this ambiguity is a major problem for medical geneticists and genetic counselors who have to manage this uncertainty with patients who are expecting clarity. For example, grading systems have been developed to rank each variant on a scale from one to five, representing "pathogenic", "likely pathogenic", "uncertain significance", "likely benign", and "benign". Evidence supporting pathogenicity is divided into additional categories, and counting the number of evidence points from each evidence category results in a final rank (Richards et al., 2015). Such complex

systems can attempt to manage uncertainty in pathogenicity, but clearly, it would be preferable to have more information about a variant to decrease the uncertainty level.

Current bioinformatic approaches to predicting variant pathogenicity are impressive in that they efficiently produce large numbers of predictions about very complex biological systems, but they are not sufficient to meet the increasing expectations about the fidelity of these predictions (Richards et al., 2015). For example, older algorithms for predicting the pathogenicity of missense mutations, many still in use, are only 65–80% accurate (Thusberg, Olatubosun, & Vihinen, 2011), and specificity for the predictions of pathogenicity of missense and splice variants can be low (Choi, Sims, Murphy, Miller, & Chan, 2012; Houdayer et al., 2012; Kircher et al., 2014). The latest techniques are raising the quality of the predictions, for example with ROC curve areas of 0.72–0.86 (Caminsky, Mucaki, & Rogan, 2014; Gray, Hause, Luebeck, Shendure, & Fowler, 2018), but in the end, the state-of-the-art is that "it is not recommended that these (bioinformatic) predictions be used as a sole source of evidence to make a clinical assertion" about a potentially pathogenic variant (Richards et al., 2015). In fact, "well-established *functional* studies showing a deleterious effect" is considered two levels-of-evidence higher than "multiple lines of *computational* evidence support(ing) a deleterious effect on the gene/gene product" (italics added; Richards et al., 2015). Again, while the bioinformatic predictions are becoming even more impressive, and expected to become even more accurate when trained with more empirical data (Gray et al., 2018), empirical wet-lab functional assays are still required to achieve the level of accuracy desired to, for example, decide whether a patient is eligible for gene therapy treatment (Cideciyan et al., 2018).

VUS not only cloud the interpretation and utility of clinical diagnostic testing, but also can lead to outliers and ambiguities when analyzing structure-function relationships of proteins of interest (Rakoczy, Kiel, McKeone, Stricher, & Serrano, 2011). For example, when correlating the computational prediction of misfolding propensity and the age of onset of disease among rhodopsin (RHO) mutants, some mutations are considered outliers and excluded from the regression analyses (Rakoczy et al., 2011); however, it is not clear whether the outliers could be due to imperfect computational models or to miscategorization of the mutant. Thus, improving the characterization of DNA variants is scientific importance, and has been included U.S. federal research priorities and identified as a knowledge gap in the understanding of inherited retinal diseases (Duncan et al., 2018; National Eye Institute, 2012).

This study implements an improved method to characterize potentially pathogenic DNA variants causing retinitis pigmentosa (RP). RP accounts for up to 25% of blindness or visual impairment in working-age people (21–60 years; Al-Merjan, Pandova, Al-Ghanim, Al-Wayel, & Al-Mutairi, 2005; Buch et al., 2004; Hartong, Berson, & Dryja, 2006; Hata, Yonezawa, Nakanishi, Ri, & Yanashima, 2003), and therefore it is an important cause of vision loss. Although RP is a Mendelian disease, it is genetically very heterogeneous, with mutations in over 60 different genes that can cause nonsyndromic RP (Berger, Kloeckener-Gruissem, & Neidhardt, 2010; Daiger, Rossiter, Greenberg, Christoffels, & Hide, 1998; RetNet, 2017). Genetic testing to identify the cause of disease has become increasingly important as more clinical trials for RP focus on patient populations with specific genotypes, for example, studies recruiting *MERTK*-, *MYO7A*-, *PDE6A*-, *PDE6B*-, *RPGR*-, *or RLBP1*- affected RP patients (U.S. National Institutes of Health, 2017), as well as the occasional RP patient due to RPE65 mutations who would be eligible for the first food and drug administration (FDA)-approved gene therapy, voretigene neparvovec (Luxturna; Russell et al., 2017). Patients without a genetic diagnosis are not eligible for gene-specific treatments. Despite the practical importance of obtaining a genetic diagnosis, definitive causal variant(s) can be found in only about half of patients, and slightly more using the latest next-generation sequencing (NGS) techniques (Consugar et al., 2015; Corton et al., 2013; Glockle et al., 2014; Huang, Wu, Lv, Zhang, & Jin, 2015; Neveling et al., 2012; Wang et al., 2014). For this reason, improving the characterization of DNA variants in RP is also of practical importance.

This study focuses on *RHO* mutants because *RHO* has the largest set of known pathogenic variants of any dominant RP gene, and among those genes, the structural, biochemical, and cell biological understanding of *RHO* is unmatched (Athanasiou et al., 2018; Behnen et al., 2018; Dryja, Berson, Rao, & Oprian, 1993; Krebs et al., 2010; McKeone, Wikstrom, Kiel, & Rakoczy, 2014; Mendes & Cheetham, 2008; Mendes, van der Spuy, Chapple, & Cheetham, 2005; Mendes, Zaccarini, & Cheetham, 2010; Rakoczy et al., 2011). The rich body of existing data provides a context for interpreting data on new variants, and conversely, provides an opportunity to refine existing models of *RHO* structure and function (Rakoczy et al., 2011). A recent review (Athanasiou et al., 2018) notes that ongoing questions about the pathogenicity of *RHO* variants "reinforce the need for thorough genetics, such as segregation analyses, and in-depth functional analyses to confirm pathogenicity." (Note that the term "variant" is used in this study to include any sequence change, whether that sequence change is a "mutant" i.e., known to be pathogenic, a VUS, or a benign change such as a synonymous control.)

Standard experimental methods of assaying *RHO* variants include assessment of surface expression/subcellular localization in cell-based assays (Behnen et al., 2018; Chen, Wang, Lin, & Chen, 2011; Chuang, Vega, Jun, & Sung, 2004; Davies et al., 2012; Hollingsworth & Gross, 2013; Li et al., 1998; Liu et al., 2013; McKeone et al., 2014; Sung, Davenport, & Nathans, 1993; Toledo et al., 2011; Yamasaki et al., 2014) and assessment of bulk biochemical properties (Bosch, Ramon, Del Valle, & Garriga, 2003; Bosch-Presegue, Ramon, Toledo, Cordomi, & Garriga, 2011; Dizhoor et al., 2008; Krebs et al., 2010; Opefi, South, Reynolds, Smith, & Reeves, 2013). These assays are performed on one variant at a time, and, especially for the biochemical assays, do not scale easily for larger numbers of variants.

The purpose of this study was to develop higher-throughput cell-biological methods to evaluate *RHO* variants for pathogenicity for example, a "functional genomics" approach. We hypothesized that this approach would result in the improved categorization of *RHO* DNA variants compared with using computational information alone. We also hypothesized that pedigrees with inconclusive genetic testing results and an *RHO* VUS could be solved using the functional

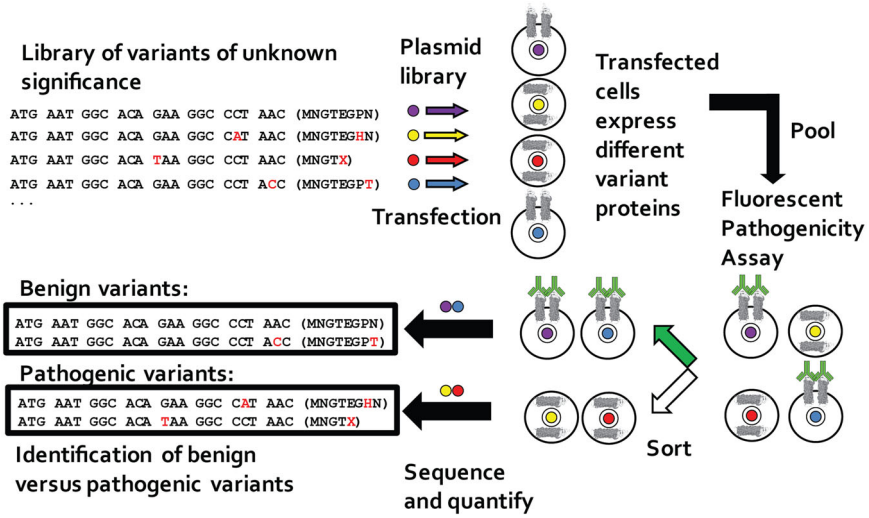


FIGURE 1 Experimental design of library screening for VUS characterization. An expression plasmid library of *RHO* variants was constructed using mutagenesis. (Four variants are shown as an example.) Each variant plasmid was individually transfected into a well of cells, and the cells were then pooled after transfection. The cells, after expressing the desired variant, were stained using a fluorescent pathogenicity assay based on RHO surface expression (fluorescent antibody). Fluorescence-activated cell sorting was used to separate high- and low-expressing cells. Each transfected cell carries with it the DNA of the variant of interest, and this "tag" can be sequenced using NGS and therefore count the relative frequency of each DNA variant in each pool, quantifying the pathogenicity of each variant in parallel. NGS: next-generation sequencing; RHO: rhodopsin; VUS: variants of unknown significance

data. This study presents a streamlined functional genomic screen (Figure 1) suitable for assaying hundreds of variants, using standard mutagenesis, transfection, flow cytometry, and NGS amplicon sequencing techniques that are broadly available. The bioinformatic analysis was custom-built, but also straightforward.

Figure 1 (see legend) describes the design of the experimental screen. Specific design considerations included: the HEK293 cell line was chosen because it has been widely used in the literature to characterize RHO mutants (Chen et al., 2011; Chuang et al., 2004; Davies et al., 2012; Hollingsworth & Gross, 2013; Li et al., 1998; Liu et al., 2013; McKeone et al., 2014; Sung et al., 1993; Toledo et al., 2011; Yamasaki et al., 2014). Also, it is a robust transfection host, and for the multiplexed assays a high transfection rate was useful to preserve library diversity. The transfection was performed in an unpooled format (one variant per well), because preliminary experiments showed that a pooled transfection at limiting dilutions of DNA, such that each cell expresses a single variant, did not produce sufficient expression (not shown). The selected pathogenicity assay was a quantification of cell surface expression of RHO using antibody staining with indirect immunofluorescence, as reported (for example, McKeone et al., 2014; any mutant protein that misfolds and fails to be detected on the cell surface was identified as pathogenic, whether that mutation is dominantly or recessively inherited in humans). We then used two different methods to implement this pathogenicity assay on the transfected cells—first, we used an unpooled format using flow cytometry analysis of RHO surface expression, which is similar to published methods (e.g. McKeone et al., 2014) and served as a validation of the system. Then, we used a method where the transfected cells from every variant were pooled (Figure 1), allowing for higher throughput, parallelized assay. Fluorescence-activated cell sorting (FACS) was used because it allowed for a large number of cells (~10^6) to be assayed and sorted into pools for NGS analysis in parallel. Amplicon sequencing of the entire RHO complementary DNA (cDNA) plasmid, rather than an implementation of plasmid barcodes for each variant, was a simpler approach given the number of variants under study (~10^2, see Section 4).

More intricate functional genomic screens have been recently developed that interrogate thousands of variants at once (Melnikov, Rogov, Wang, Gnirke, & Mikkelsen, 2014; Brenan et al., 2016; Findlay et al., 2018; Gasperini, Starita, & Shendure, 2016) but these methods require specialized library construction techniques, lentivirus packaging, and viral transduction followed by selection, which were not needed for this study of 210 variants.

Overall, a functional genomics approach was used to create a library of *RHO* variants and precisely assay their surface expression in an efficient format.

2 | MATERIALS AND METHODS

2.1 | Identification of RHO variants

A literature review (PubMed) and database search (HGMD, ClinVar) identified 211 known RHO variants. Twenty-two database variants were not created due to technical complexity (large insertions, large deletions, in an intron, or in an untranslated region), due to likely incompatibility with pathogenicity (a high-allele-frequency synonymous allele), or due to never having been reported in humans ("synthetic"). See Table S1. The remaining 189 variants were added to the variant library and of those, 175 were known as pathogenic, causing dominant RP, recessive RP, or congenital stationary night

blindness. (Table S1) 14 variants were of uncertain significance or thought to be benign polymorphisms. All of the variants were well distributed across the cDNA, except that there were no variants between bp 664 and 745, which corresponds to cytoplasmic loop 3 of RHO (Zhou, Melcher, & Xu, 2012; DNA descriptors in this study refer to NM_000539.3 and protein descriptors refer to NP_000530.1 also known as RHO_v001).

A group of 11 individual synonymous mutations, expected to have a similar expression to wildtype and to serve as controls, was designed synthetically (Table S1). The mutations were placed near the 3′-end of the gene to minimize any possible (though unexpected) changes to transcriptional or translational processes (Stoletzki & Eyre-Walker, 2007). In addition, a search for rare (detected <3 times in our internal database, ExAC allele frequency $<10^{-4}$) RHO VUS from patient DNA sequencing performed in our institution's genetic testing service (Consugar et al., 2015), identified 10 variants spanning 11 probands. All of these variants were novel (previously unreported in patients) at the beginning of this study. Altogether, 232 RHO variants were identified and 210 were ultimately constructed as a variant library: 189 variants from literature/databases, 11 synonymous controls, and 10 novel VUS.

The study protocol adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of Massachusetts Eye and Ear.

2.2 | RHO variant cDNA library construction

A Gateway destination vector was created that contained: a CMV early enhancer/chicken β actin (CAG) promoter, V5 tag, Gateway cassette (Thermo Fisher Scientific, Waltham, MA), internal ribosome entry site (IRES), mCherry fluorescent protein marker (Addgene, Watertown, MA), and ampicillin resistance gene. Plasmids were propagated in ccdb survival cells (Thermo Fisher Scientific) with chloramphenicol and ampicillin. A pDONR entry clone containing the human RHO cDNA sequence was purchased (#GC-T1321; GeneCopoeia, Rockville, MD). Each of the 210 *RHO* variants was created via site-directed mutagenesis using a one-primer modification to the QuikChange II protocol from Agilent (Braman, Papworth, & Greener, 1996; http://qb3.berkeley.edu/macrolab/quick-change-mutagenesis/). Primers were synthesized in 96-well plates by Integrated DNA Technologies (Coralville, IA). Entry clone plasmids were propagated in One Shot Top10 chemically competent *E.*coli (Thermo Fisher Scientific) with kanamycin 50 μg/ml. The 1,047 bp RHO insert and flanking at sites were Sanger sequenced in both directions to exclude second-site mutations. Reracked positive clones were resequenced to verify the identity of each clone on each plate.

Expression clones ($n = 210$) were created by recombination of destination vector and each of the mutagenized entry clones, using LR Clonase II (Thermo Fisher Scientific). The resulting expression clone consisted of pCAG-V5-*RHO* (*WT or variant*)-IRES-mCherry and was propagated in Stbl3 cells (Thermo Fisher Scientific) with ampicillin 50–100 μg/ml. Clones were sequenced, reracked, and resequenced to verify the identity of each clone on each final library plate. DNA purification and sequencing (96-well plate) were performed by the CCIB DNA Core Facility at Massachusetts General Hospital (Cambridge, MA).

2.3 | Cell culture

HEK293 cells (American Type Culture Collection) were cultured using aseptic technique and grown in 10% fetal bovine serum (FBS) in Dulbecco's modified Eagle's medium serum without antibiotics. Cells were grown at 37°C and 5% CO_2 in a standard cell culture incubator and passaged at subconfluent densities every 2–3 days as needed.

2.4 | *RHO* transfection and surface expression assays by immunofluorescence

All transfections were performed using Lipofectamine 2000 (11668019; Thermo Fisher Scientific) on HEK293 cells seeded at a density of 200 cells/mm^2 on tissue culture dishes. The day after seeding, the media was refreshed and the transfection mixture added as described below. Samples were collected 48 hr after transfection.

2.4.1 | Immunofluorescence microscopy

Glass coverslips were placed in wells before cell seeding into six-well dishes. Transfection was performed using 2 μg plasmid DNA and 3 μl Lipofectamine in 150 μl Opti-MEM I (31985062; Thermo Fisher Scientific). At collection, cells on coverslips were fixed in 4% paraformaldehyde for 20 min and blocked with 3% bovine serum albumin in phosphate-buffered saline for 10 min before applying Ret-P1 anti-RHO antibody (#O4886; Millipore, Sigma, Burlington, MA) at a final concentration of 1:1,000 in blocking buffer for 1 hr. After washing with PBS, Alexafluor-488 goat anti-mouse antibody (A-11029; Thermo Fisher Scientific) was applied at a final concentration of 1:300 in blocking buffer for 1 hr and Hoechst 33342 (#H3570; Thermo Fisher Scientific) applied for 1–5 min at a concentration of 1:5,000. Coverslips were mounted onto slides with Fluoromount G and dried overnight at room temperature. Slides were kept at 4°C before viewing on a Nikon TI Eclipse microscope or Leica TCS SP5 confocal microscope.

2.4.2 | Flow cytometry

Transfection was performed in 24-well plates using 0.5 μg plasmid and 1.5 μl Lipofectamine in 50 μl Opti-MEM I. At collection, the cells were washed briefly with PBS and then 200 μl trypsin added and allowed to incubate for 4 min at 37°C. A 300 μl aliquot full media was then added, and cells collected into 2.2-ml tubes. The cells were centrifuged at 400*g* in a floor centrifuge for 4 min and the supernatant aspirated. Cells were fixed in 4% PFA for 20 min and blocked with 3% BSA in PBS before applying the Ret-P1 primary antibody at a final concentration of 1:1,000 in blocking buffer for 30 min. After washing with PBS, Alexfluor-488 secondary antibody was applied at a final concentration of 1:300 in blocking buffer for 30 min. After a final PBS wash, cells were analyzed on a BD LSRII Flow Cytometer.

Flow cytometry results were analyzed with Flowing Software (http://www.flowingsoftware.com). The percentage of cells with high RHO surface expression was determined by setting the quadrants based on the WT and p.P23H controls in each experiment and then dividing the percent of cells in the top right quadrant (double-positives) by the sum of the top right and bottom right quadrants (all transfected cells). Resulting data were averaged over independent transfections from different weeks (n = 1–3).

2.4.3 | FACS

Cells were prepared as described for flow cytometry analysis but pooled after transfection and then fixed in zinc-based fixative (ZBF; #550523; BD Pharmingen) instead of 4% PFA (ZBF was used because PFA fixation/crosslinking prevented downstream PCR amplification of residual plasmid DNA [Wester et al., 2003], while unfixed cells did not stain with Ret-P1 antibody). All subsequent washes substituted Tris-buffered saline (TBS) for PBS to reduce salt precipitation. Cells were sorted into TBS or PBS with 1% FBS using a BD SORP 5 Laser Vantage SE DIVA. Replicate samples were derived from independent transfections on different weeks (n = 3).

Additional experiments to implement a pooled transfection method were not effective; at standard plasmid concentrations, large numbers of different variants entered each cell causing every cell to stain similarly, while at low plasmid concentrations, not enough RHO expression was achieved for robust staining and cell sorting (not shown).

2.5 | RHO cDNA OR RNA extraction and amplification from sorted cells

Immediately after sorting, cells (from 4×10^4 to 2.3×10^6) were pelleted at 400g for 4 min and resuspended in 700 μl RLT buffer (Qiagen) with β-mercaptoethanol (1:100), mixed and aliquoted into two samples of 350 μl each and stored at −80°C until extraction. Plasmid DNA and total RNA were isolated. Preliminary experiments (not shown) demonstrated that RNEasy mini columns (#74104; Qiagen) can isolate both RNA and the relatively low molecular weight plasmid DNA in one step using the manufacturer's standard RNA isolation protocol. However, as on-column DNAse treatment performed more reliably for the RNA preparation, the cells were separated into two aliquots loaded on separate columns: For total RNA extraction, the manufacturer's standard RNEasy protocol was followed, including on-column DNAse digestion (#79254; Qiagen). For plasmid DNA extraction, a separate aliquot of cells in RLT buffer was purified using RNEasy mini columns without DNAse digestion, and PCR performed without reverse transcription to avoid amplifying RNA. The final RNA or plasmid DNA/RNA mixture was eluted in 35 μl water, and 15 μl was used for the reverse transcription polymerase chain reaction (RT-PCR) or PCR reaction described below.

For DNA-extracted samples, PCR amplification was performed using forward primer 5′-GTTTGTACAAAAAAGCAGG-3′ and reverse primer 5′-GGAATTTACGTAGCGGC-3′, which were complementary to regions of the plasmid DNA flanking the RHO sequence. HotStarTaq polymerase (#203203; Qiagen) was used with the following PCR program: (a) 95°C for 15 min; (b) 36 cycles of: 94°C for 30 s, 50°C for 30 s, 72°C for 1 min; (c) 72°C for 10 min. For RNA-extracted samples, One-Step RT-PCR reagent (#210212; Qiagen) was used to create cDNA and then the cDNA was amplified by PCR. The following program was used: (a) 50°C for 30 min; (b) 95°C for 15 min; (c) 36 cycles 94°C for 30 s, 50°C for 30 s, 72°C for 1 min; (d) 72°C for 10 min. One microliter of each 100 μl reaction was run on a Tapestation (Agilent) to determine approximate size and concentration and to confirm the presence of a single band of an amplified template. Negative control reactions (minus RT for RNA, plus DNAse for DNA) showed no cross-amplification between samples that were intended to amplify RNA or DNA, respectively. PCR reactions were processed through a PCR clean-up column (#28104; Qiagen, Germantown, MD), and the concentration determined using a QuBit dsDNA HS Assay (Q32854; Thermo Fisher Scientific).

2.6 | NGS amplicon sequencing without library barcodes

Smaller scale NGS sequencing (1/96 of a MiSeq run per sample) was performed by the CCIB DNA Core Facility at Massachusetts General Hospital (Cambridge, MA). For more read depth (~1/10 of a MiSeq run per sample), NGS sequencing was performed in the Ocular Genomics Institute facility (https://ocular genomics.meei.harvard.edu/index.php/gc) using the following protocol: the PCR product was sheared on a Covaris E220 focused ultrasonicator set to a treatment time of 360 s, acoustic duty factor of 10%, peak incident power of 175 W and 200 cycles per burst. Library preparation was performed with the Truseq nano LT Kit (#FC-121–4001; Illumina, San Diego, CA) but modified to use AMpure XP beads for the clean-up steps (#A63881; Beckman Coulter, Indianapolis, IN). Briefly, steps included: clean-up, quantification, end repair, clean-up, A-tail, adapter ligation, clean-up, PCR enrichment, clean-up, quantification and normalization, denature, and run on MiSeq using 2× 121 cycles with a 6 bp index. Because of relative overrepresentation of the amplicon ends, PhiX library (10%) was added to the final mixture.

2.7 | Bioinformatic quantitation of low-frequency variants in NGS amplicon sequencing

Most variant callers are designed to work with diploid genomes and do not call variants present at less than about 50% frequency. Extensive testing with specialized low-frequency variant callers (Spencer et al., 2014) showed that these variant callers, including "lofreq" (Wilm et al., 2012), did not accurately call all variants, particularly small insertions/deletions (a known limitation of alignment algorithms; Jiang, Turinsky, & Brudno, 2015). Additional analyses with MuTect (Cibulskis et al., 2013) or MuTect2 showed low sensitivity using default settings, and settings with appropriate sensitivity and specificity were not identified (not shown).

Multinucleotide polymorphisms (e.g. 511_512delCCinsGA) were particularly problematic.

Therefore, an alignment-free approach was used to detect and quantify low-frequency variants in NGS data when the variants are known. This algorithm uses no alignment step and simply counts the number of exact-match occurrences of a 20 bp "probe" sequence string designed for each known variant. The "probes" (in analogy to Southern blot probes), were designed in silico containing each of the known 210 nucleotide changes. In most cases, the primer used for QuikChange mutagenesis was simply trimmed to 20 bp to create the "mutant probe". A corresponding "wildtype probe" was created to quantify and normalize for variations in coverage. To estimate signal-to-noise ratio in the context of NGS sequencing errors (Fox, Reid-Bayliss, Emond, & Loeb, 2014), a "noise probe" was manually created containing an alternate nucleotide change which was not present in the wildtype or mutant probes. Each occurrence of these 630 probes (wildtype, mutant, or noise × 210 variants) was counted in the text of each sample's FASTQ file using the Linux tool *grep*, similar to a strategy previously published (Bujakowska, White, Place, Consugar, & Comander, 2015). The coverage at both ends of the amplicon was overrepresented, presumably due to the sequencing of the residual unsheared amplicon.

Preliminary experiments showed that the signal-to-noise ratio was low for some variants, particularly at amplicon ends, where early MiSeq cycles had a lower base call quality value. Thus, before quantifying the frequency of each probe in each FASTQ file, a script was implemented to censor all FASTQ files on a base pair-by-base pair basis for a high-quality Q score from 36 to 41 ("E" to "J"), resulting in improved signal-to-noise ratio (not shown). For graphing, signal-to-noise ratios were capped at 1,000 when zero noise probes were detected. Run-time for quantification of the entire dataset was about 2 hr on our local computer cluster. These scripts are available at https://github.com/jcomand/VariantCounting.

The experimental design called for relative quantitation of each variant in two pools of transfected and sorted cells, in this case, "high" versus "low" surface-RHO-expressing cells. The raw read counts of each mutant probe were normalized to the read depth of the wildtype probe. that is the final "NGS ratio" used for analysis was the number of read counts: mutant probe_high/wildtype probe_high/mutant probe_low/wildtype probe_low. Three independent biological experiments (e.g. separate transfections on different weeks) were converted to log-ratios for calculation of averages and standard errors for each variant. This log transformation was sufficient to produce approximately normally distributed ratios in this dataset with relatively large read counts (not shown); however for a dataset with small read counts, a transformation of the counts would be required (Viner, Dorman, Shirley, & Rogan, 2014).

2.8 | Predicting clinical disease severity

Clinical outcomes of subjects with *RHO* mutations have previously been published, including by our institution (Berson, Rosner, Weigel-DiFranco, Dryja, & Sandberg, 2002). Electroretinography (ERG) data were analyzed from these subjects and subsequent subjects with *RHO* mutations from later years. The baseline (first visit) 30 Hz cone flicker ERG amplitude was used as the clinical outcome measure, as it has a broad dynamic range (>3 log units) across disease severities (Berson et al., 2002). Other ERG parameters or other outcomes such as visual field area and visual acuity were not used, to avoid multiple testing. To maximize sample size, data from DNA variants with the same amino acid change were pooled. Of the amino acid changes represented in the clinical data, class 2 (misfolding) mutations were the only biochemical category (Rakoczy et al., 2011) that had more than two mutations represented. The analysis was therefore limited to subjects with class 2 mutations ($n = 69$ subjects; class 2 mutations also have the best biological rationale and precedent for correlating with disease severity in this assay; Athanasiou et al., 2018; Rakoczy et al., 2011). Standard linear regression was used to predict the logged 30 Hz ERG amplitude based on the logged NGS-based final surface expression ratio. A multivariate model (analysis of variance) was also used which adjusted for age at baseline visit, as younger subjects have higher baseline ERG amplitudes. For comparison with computationally-derived datasets, ΔΔG values (Rakoczy et al., 2011) and pathogenicity predictions from the Envision dataset (Uniprot P08100; Gray et al., 2018) were tested as predictors of clinical severity as well.

3 | RESULTS

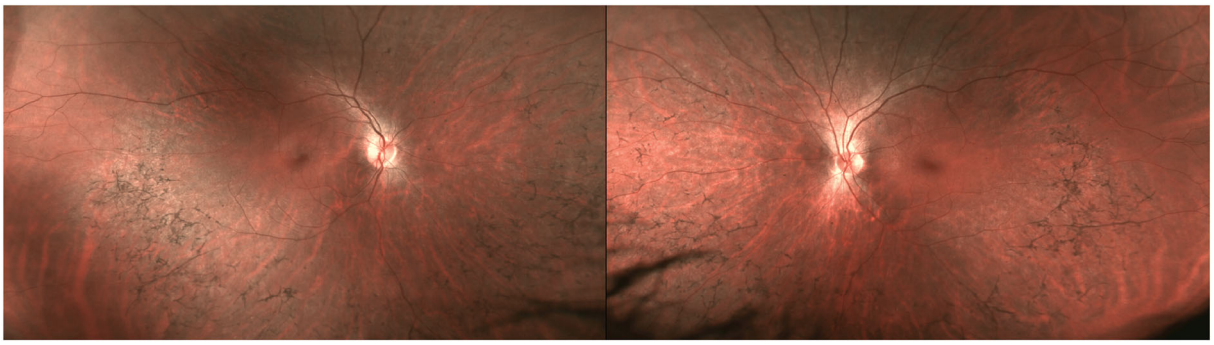
3.1 | Example of inconclusive genetic testing results including a rare *RHO* VUS

This study hypothesizes that functional studies can help interpret DNA VUS found in patient samples. Figure 2 shows an example of a clinical situation in which an inconclusive genetic testing result, including a rare RHO VUS, was obtained for a proband with RP. Of the 210 *RHO* variants identified for this study, 10 variants were from rare *RHO* VUS that were discovered in DNA samples from subjects with inherited retinal diseases (See Section 2 and Table S1). Figure 2a shows the fundus appearance of one of these 10 subjects (D00726) who was diagnosed with RP after a full clinical evaluation. In this Mendelian disorder, it is most likely that variant (s) in a single gene are truly pathogenic. Figure 2b shows the large number of VUS identified in genetic testing for this subject using panel-based sequencing of inherited retinal disease genes (GEDi test; Consugar et al., 2015). The *RHO* p.P180T variant (row 3) is predicted to be pathogenic, but so are other VUS to varying degrees, two of which are coding variants in genes that can cause dominantly-inherited RP (SNRNP200 and RP1). An expert in ocular genetics may suspect that the *RHO* variant is the most likely to be pathogenic, but it is not conclusive. Therefore, additional functional data are needed.

3.2 | Flow cytometry assay validation

RHO variant plasmids ($n = 210$) were created using site-directed mutagenesis and these were then cloned into an expression vector suitable for transfection into cultured cells. The next step (Figure 1)

(a)



(b)

chromosome	genomic position	reference allele	variant allele	type	geneID	Gene inheritance model	refseq	consequence	cDNA change	protein Change	Gnomad AF Exome	Gnomad AF Genome	frequency (of >1700 local panel seqeuncing)	SIFT	PolyPhen	CADD	GERP score
1	211652457	G	A	het	RD3	AR	NM_001164688.1	missense	c.509C>T	p.Ser170Phe	0.000021		1	deleterious	probably_damaging	28.3	4.3
2	96950323	C	T	het	SNRNP200	**AD**	NM_014014.4	missense&splice	c.4165G>A	p.Val1389Ile	0.005484	0.00236	32	tolerated *	benign *	22.6	5.6
3	129251101	C	A	het	RHO	AR;**AD**	NM_000539.3	missense	c.538C>A	p.Pro180Thr			1	deleterious	probably_damaging	24	4.2
8	55540031	A	G	het	RP1	AR;**AD**	NM_006269.1	missense	c.3589A>G	p.Ser1197Gly	0.000224	0.00026	2	deleterious	benign	14.42	4.6
10	50732202	T	G	het	ERCC6	AR	NM_000124.3	missense	c.1274A>C	p.Asp425Ala	0.001744	0.00149	10	deleterious	possibly_damaging	23.9	4.9
16	77325275	C	T	het	ADAMTS18	AR	NM_199355.2	missense	c.3290G>A	p.Arg1097His	0.000435	0.00013	1	tolerated	benign	23.3	5.1
22	29075659	G	C	het	TTC28	AR	NM_001145418.1	missense	c.53C>G	p.Ala18Gly	0	0.00013	15	tolerated_low	unknown	21.7	2.4

FIGURE 2 Example subject (ID#D00726) with multiple VUS. (a) Fundus photographs show midperipheral bone spicule pigmentation consistent with a diagnosis of RP. (b) DNA sequencing using the Gene Eye Disease panel revealed a large number of VUS, with annotations color coded (red/dark = supporting pathogenicity; green/light = supporting benign). An additional 31 rare, noncoding variants are not listed. *splice site potentially broken, Human Splice Finder −34%. RP: retinitis pigmentosa; VUS: variants of unknown significance

was to optimize a fluorescent assay to identify which variants have pathogenic surface expression levels. Based on the assay by McKeone et al. (2014), Figure 3 shows that indirect immunofluorescence staining successfully distinguishes between wildtype RHO and known mutant (p.P23H) RHO. In this case, whereas the wildtype RHO was expressed well on the cell surface, the p.P23H mutant did not express on the cell surface but rather was trapped inside the cell as expected. The results were validated using nonpermeabilized cells to ensure that only surface RHO was detected (Figures 3a,b); flow cytometry results corresponded to the immunofluorescence staining patterns (Figure 3c). By comparing flow cytometry results from wildtype and p.P23H mutant RHO as positive and negative controls, respectively, the assay to separate wildtype from p.P23H mutant RHO was highly reproducible (Z' = 0.94; Figure 3d; Z' is a measure of statistical effect size, used in high-throughput screening to judge whether the response in a particular assay is large enough to warrant further attention). While this performance in distinguishing p.P23H versus wildtype RHO does not guarantee that all mutants will be detected with the same level of fidelity, it suggests that this assay is appropriate for use in a "high-throughput screening" context.

Each of the 210 variants in the library was then tested by flow cytometry using two methods for comparison: (a) the standard unpooled flow cytometry assay in which each variant is maintained in a separate tube and analyzed separately (Figure S1); and (b) a pooled assay in which transfected cells are then pooled and FACS is used to sort RHO high-expressing versus low-expressing cells (Figure 4a). The entire cDNA within the transfected plasmid (inside the cell) serves as the "barcode" and allows deconvolution/quantitation of the level of each variant that is represented in the high and low pools. The residual plasmid is isolated from each cell pool, amplified by PCR and sequenced by NGS.

3.3 | Validation of string-based pooled variant detection and quantification

All 210 library variants were detected at a coverage-normalized frequency that was >1.5 of that observed by sequencing a wildtype *RHO* amplicon with no variants (Figure 4b). Two hundred and seven of the 209 variants were seen at >×2. One variant which was used in the positive control sample during flow cytometry (diamond) showed overrepresentation, possibly due to cross-contamination at the flow cytometry stage; this was minimized in future iterations. Next, intra-sample quality control was performed by evaluating the level of the "noise probes" as described in the Methods. Customized computational filtering of sequencing quality scores was needed to obtain high signal-to-noise data. Figure 4c shows the signal-to-noise (S-N) ratio as a function of location in the cDNA. All variants were detected with a S-N ratio >1.5×, and with 208 of 210 detected at >2× (Figure 4c).

The starred cluster of points (Figure 4c) with a lower, but acceptable, S-N ratio is from variants located near a sequencing error-prone sequence context that is GC rich with a poly-C repeat (GGCCCCGGCC). There are a large number of reported variants in this region (Table S1), and in the 3′-end of the coding region in

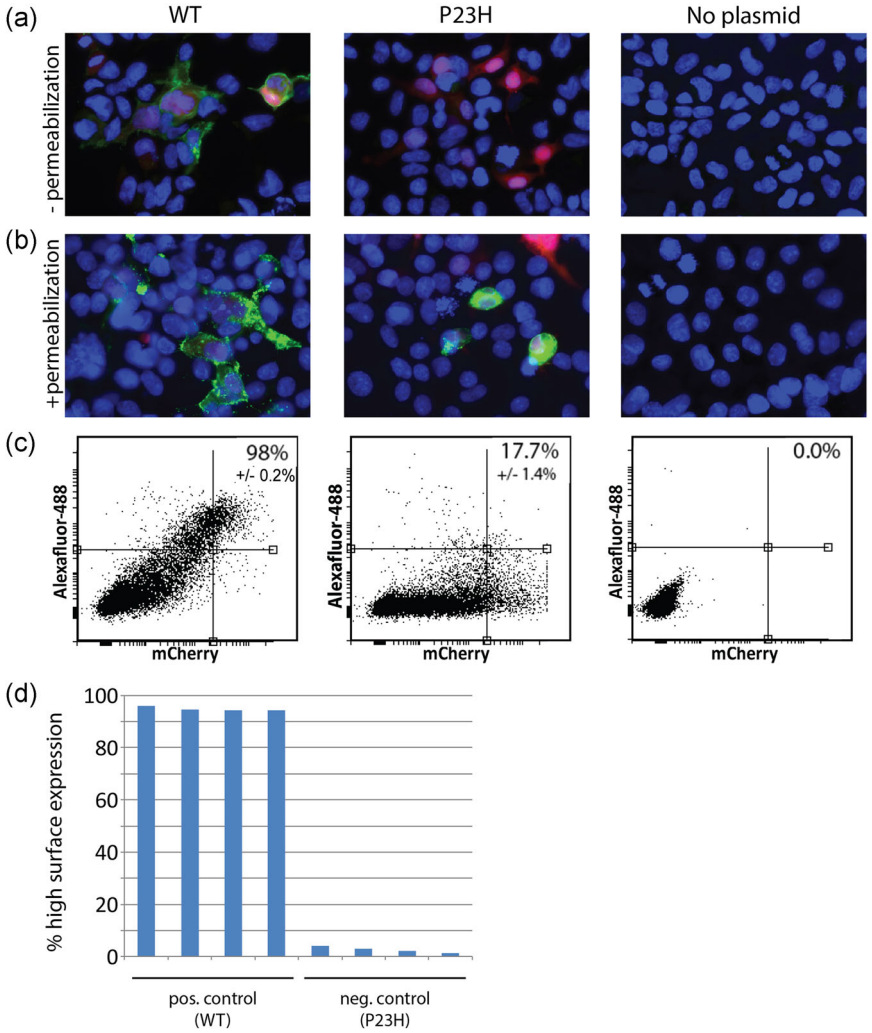


FIGURE 3 Unpooled assay validation. (a) Indirect immunofluorescence without permeabilization using Ret-P1/Alexafluor-488 antibodies shows strong RHO surface expression (green) using the wildtype plasmid (WT; left column), but not with a known mutant construct (p.P23H; middle column) or no plasmid control (right column). Red: mCherry transfection control. Blue: nuclei (Hoechst). (b) With permeabilization, mutant RHO is detectable inside the cell (middle column). (c) When the same antibodies were used in flow cytometry (without permeabilization), the percentage of RHO^+/transfection$^+$ cells (inset) correlates to the results obtained using immunofluorescence. (d) Replicate flow cytometry assays show high separation between positive and negative controls, and low noise, Z' = 0.94. RHO: rhodopsin

general, forming a "hot spot" of reported variants. In contrast, there are no reported variants between bp 665–744.

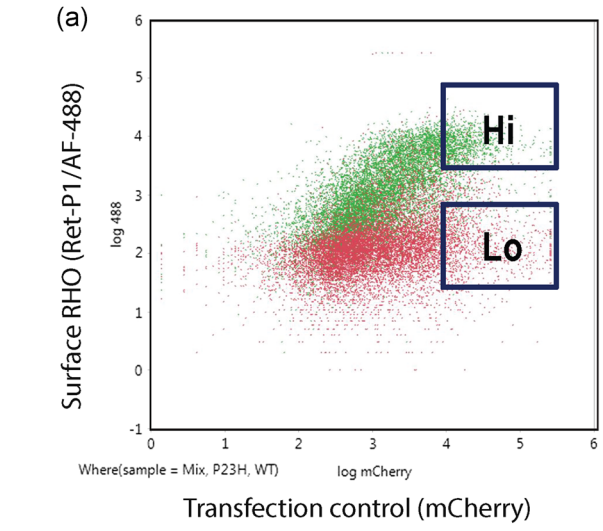
3.4 | Validation and quality control of the final NGS-based surface expression ratio

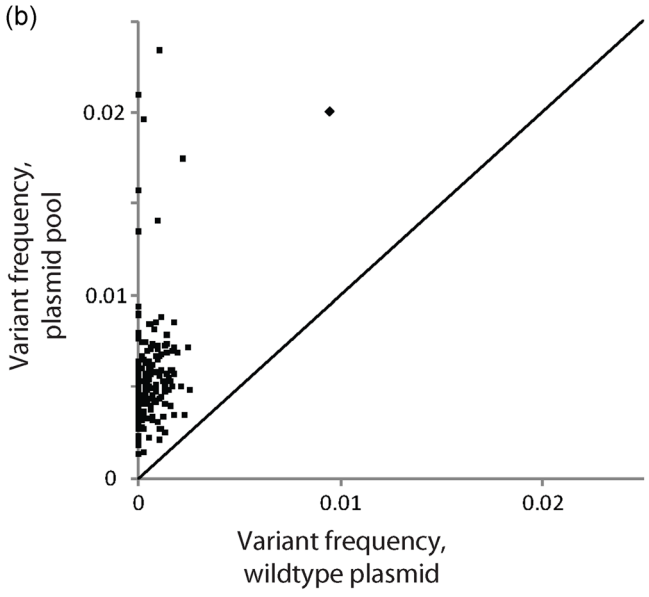
The final surface expression ratio was high when cells with a particular variant are primarily found by flow cytometry within the "Hi" gate (Figure 4a), with good surface expression, similar to wildtype RHO. The final ratio was low when cells with a particular variant are primarily found in the "Lo" gate, similar to known RHO mutants which do not express well on the cell surface. The characteristics of this ratio were evaluated both by comparing it to the standard unpooled flow cytometry assay and by its analytical and biological variability.

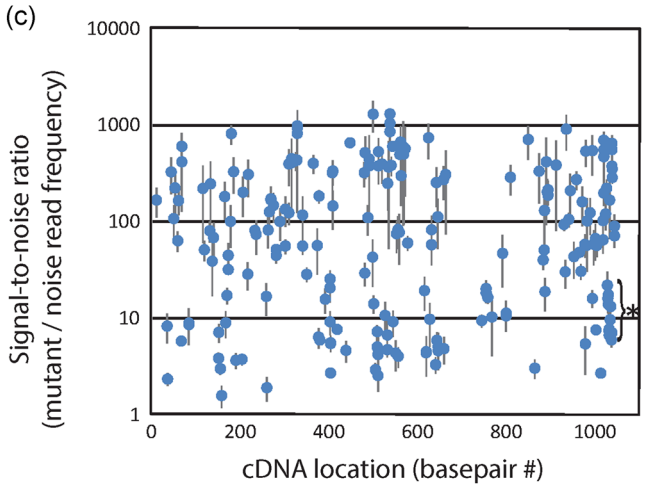
Figure 5 demonstrates that there was very good concordance between the standard unpooled assay and the NGS-based pooled assay, with an overall r^2 of 0.92. The pooled assay drastically reduced the number of samples (per biological replicate) that had to be examined by flow cytometry; unpooled: 210 samples analyzed, plus three controls. Pooled: two samples sorted, plus three controls. For the minor discrepancies/outliers in the intermediate range of the assays, it is not known which assay was more accurate. The red and pink areas contain variants with "very low" and "low" expression levels, respectively, consistent with those variants being pathogenic mutations. While hard cutoffs of such regions have no exact biological meaning, for practical purposes the regions were defined by 40% and 80% surface expression on the *x*-axis, and the corresponding pooled assay ratios can be compared with positive and negative controls (Figure 6). Of note, while the box-and-whiskers plot at the top of Figure 6 identifies the statistical range of the wildtype-like synonymous controls, it is not necessarily the case that variants falling just below these levels would be pathogenic biologically. To be conservative about making claims of pathogenicity, the above cutoffs were used instead of cutoffs based on a statistical confidence interval.

To evaluate variability, the pooled assay was performed on DNA from three separate biological replicates performed in different months. The resulting ratios showed high correlations between replicate experiments (r^2 = 0.86–0.88; see Figure S2). Of note, the data described above were derived from DNA amplification of the

residual plasmid in the transfected cell. Ratios derived from RT-PCR of RNA were much more variable, and while they showed similar trends, were less consistent (data not shown).







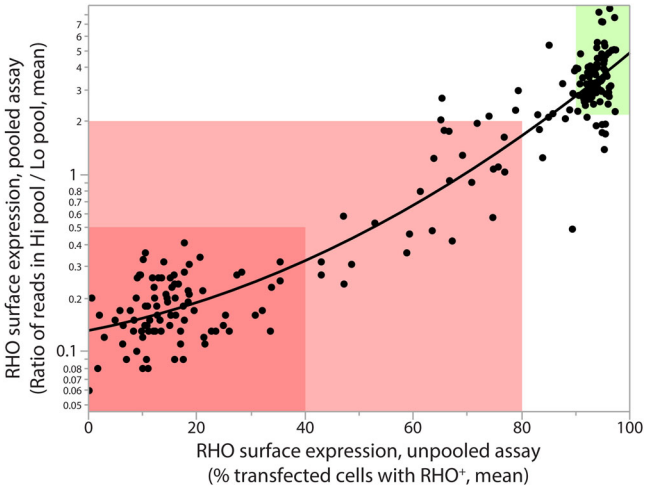


FIGURE 5 Concordance between unpooled and pooled assays. The unpooled RHO surface expression flow cytometry assay performed on individual tubes (*x*-axis) showed highly correlated results (quadratic fit; $r^2 = 0.92$) to the NGS-based, pooled assay (*y*-axis, log scale). Each point represents one variant. Red/dark = very low expression, pink/medium = low expression, and green/light = high expression. RHO: rhodopsin

Next, the pooled assay results were graphed by the variant class derived from pre-existing literature and database annotations (Figure 6). The reported biochemical classes were obtained from the literature (Athanasiou et al., 2018; Mendes et al., 2005; Rakoczy et al., 2011; additional historical categorizations have also been published; Krebs et al., 2010; Sung et al., 1993), as well as classifications that integrate pharmacologic response (Behnen et al., 2018). The synonymous controls (top row) demonstrated high surface expression, as expected. Most of the known class 2 mutants exhibited very low surface expression as expected, but a minority (19%) expressed at levels that were unexpectedly high (p.L125R, p.E150K), intermediate (p.: A164V, G51V, F52Y, F56Y, T58M, K296M, and K296N), or moderately low (p.: G51R, T58R, L88P, G109R, C167R, S186W, M207R, M216K, and K296E). Class 3 variants showed intermediate levels. Class 4 variants showed low and high levels (see Section 4). Class 5–7 variants and benign variants show intermediate to high levels. Labeled variants that were previously considered VUS (p.L47R, p.G18D, p.G101V, and p.P180T) have now been demonstrated to express at pathogenic levels. Note that the variants with high expression levels are not

FIGURE 4 Pooled assay workflow and signal-to-noise ratio. (a) Gates used for FACS sorting cells with high or low rhodopsin (RHO) surface expression, based on a positive control wildtype RHO sample (green/upper) and a negative control mutant sample (red/lower). (b) Variant detection above background is demonstrated for each of 210 library variants, comparing sequencing results from of a pool of all variants (*x*-axis) to those from a wildtype *RHO* amplicon with no variants (*y*-axis). (c) Signal-to-noise ratio of the 210 variants as a function of location in the cDNA sequence. Error bars represent standard error of three independent replicates. Star = cluster of lower signal-to-noise probes- see text. cDNA: complementary DNA

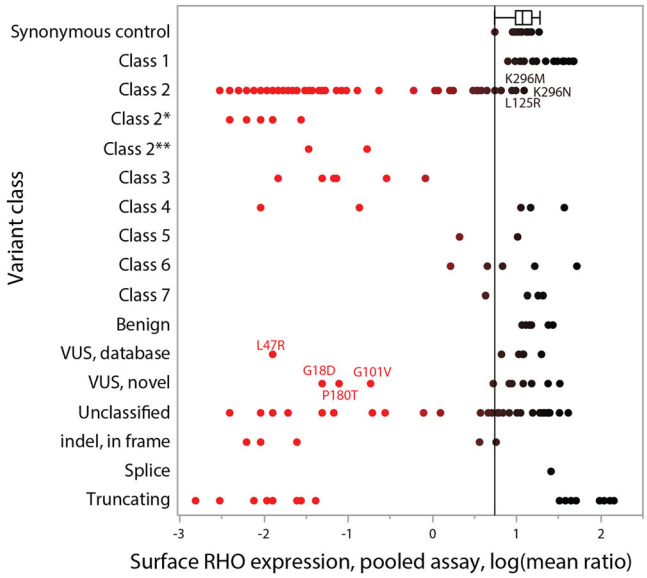


FIGURE 6 Pooled assay results by variant class. RHO surface expression results are plotted by previously-described biochemical classes. The color gradient (red/light-black) is based on the *x*-axis value, with pathogenic levels-of expression to the left, and wildtype levels to the right. A vertical line is drawn at the lower bound of the box-and-whiskers plot, which shows the interquartile range and upper and lower data point values of the synonymous controls. Class 2*, following Athanasiou et a., (20018) might behave as class 4 after the 11-cis-retinal rescue and class 2** might behave as class 2 on overexpression, but class 4 in vivo. Labeled points are examples of variants previously considered class 2 mutants that unexpectedly show intermediate or high expression levels (upper) or VUS that were previously unclassified that are now likely to be class 2, 3, or 4 mutations (lower). RHO: rhodopsin

necessarily all benign, as the assay does not detect all classes of mutations.

Next, six variants with new or unexpected findings were selected for further validation using immunofluorescence with confocal microscopy. p.A164V and p.G109R are class 2 mutants according to Athanasiou et al. (2018) and Rakoczy et al. (2011), but unexpectedly showed intermediate or relatively preserved surface RHO staining in all three assays—standard and pooled flow cytometry (Table 1) as well as confocal microscopy (Figure 7). The following variants were originally of unknown pathogenicity: p.L47R (literature) and p.G18D, p.G101V, p.P180T (internal VUS). All four of these variants showed low or very low surface expression in the standard and pooled flow cytometry assays and were further validated to have a low surface expression in the confocal immunofluorescence assay (Figure 7).

3.5 | Probands with a rare *RHO* VUS

Eleven probands that had genetic testing through our departmental genetic testing service had a novel *RHO* VUS, as listed in Table 2. These variants were submitted to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/). These represent 10 unique VUS (as G18D was found in two probands; since the start of this project, the p.R147C variant has been added to the HGMD database as a pathogenic variant causing RP). Of the remaining nine VUS, three (33%) showed pathogenic surface expression levels. The variants that were pathogenic by the functional assay were not uniformly predicted to be damaging based on computational predictions alone (Table 2). Both of the subjects with the p.G18D mutation and the one subject with the p.G101V mutation had a phenotype of pericentral RP, as described in further detail (Comander et al., 2017). The subject is shown in Figure 2 who had RP and a p.P180T *RHO* variant now has been demonstrated to have a pathogenic *RHO* mutation.

3.6 | Predicting clinical disease severity data from surface expression levels

Among known class 2 mutants, the level of surface RHO expression measured in vitro can predicted the clinical disease severity of class 2 mutants, with an observed increase of 0.67 of Ln ERG amplitude for each log of surface expression ratio ($p = 0.0008$; Figure 8). A multivariate model that takes an age-at-baseline visit into account gives similar results, with a 0.59 increase of Ln ERG amplitude for each log of surface expression ratio ($p = 0.016$). In contrast, ΔΔG values from Rakoczy et al. (2011), representing the computationally predicted misfolding propensity, did not predict ERG amplitudes alone, or in a combined model with age or age and NGS ratio (all $p > 0.05$). Similarly, computational pathogenicity predictions from the Envision dataset (Gray et al., 2018) did not correlate with clinical severity, with or without restriction of class 2 mutants ($p > 0.05$).

A more robust observation was that class 1 mutants generally had a worse severity than class 2 mutants (baseline log cone flicker ERG amplitude of −0.56 + −0.17 versus 0.38+/−0.11, $p < 0.01$), extending earlier findings (Berson et al., 2002).

4 | DISCUSSION

This study demonstrates the feasibility of a higher-throughput cell-based assay for the functional characterization of VUS in IRDs, compared with the standard approach of using a cell-based assay for each variant individually (Chen et al., 2011; Chuang et al., 2004; Davies et al., 2012; Hollingsworth & Gross, 2013; Li et al., 1998; Liu et al., 2013; McKeone et al., 2014; Sung et al., 1993; Toledo et al., 2011; Yamasaki et al., 2014). A pooled, multiplexed assay for RHO variants can efficiently identify class 2, 3, or 4a mutants. This assay was used to identify pathogenic variants within a group of VUS, provisionally solving three pedigrees of typical or pericentral retinitis pigmentosa, including the proband in Figure 2.

The barcode-free, pooled, NGS-based assay for evaluating RHO surface expression was highly reproducible and quantitative, with good agreement with the standard unpooled assay. This was achieved by optimizing several segments of the schema in Figure 1, including extensive library validation, with sequencing on both DNA

TABLE 1 New findings based on RHO surface expression assay results.

Variant number	DNA description	Protein description	Previous class, detailed (Mendes/ Rakoczy/ Athanasiou)	Previous class, simple	Unpooled RHO surface assay, mean	Pooled RHO surface assay, mean	RHO surface expression, final category	Conclusions based on new data from this study	Revised class
81	578C>T	T193M	Unclassified	Unclassified	75.7	1.10	low	Previously unclassified; now likely class 2, 3, or 4	likely class 2, 3, or 4
88	641T>A	I214N	Unclassified	Unclassified	71.8	1.94	low	Previously unclassified; now likely class 2, 3, or 4	likely class 2, 3, or 4
169	500G>A	C167Y	Unclassified	Unclassified	48.6	0.31	low	Previously unclassified; now likely class 2, 3, or 4	likely class 2, 3, or 4
182	236T>C	L79P	Unclassified	Unclassified	70.8	0.90	low	Previously unclassified; now likely class 2, 3, or 4	likely class 2, 3, or 4
19	170T>G	L57R	Unclassified	Unclassified	17.7	0.15	very low	Previously unclassified mutant; now likely class 2	likely class 2
23	233A>T	N78I	Unclassified	Unclassified	22.7	0.13	very low	Previously unclassified mutant; now likely class 2	likely class 2
66	535A>T	I179F	Unclassified	Unclassified	23.5	0.13	very low	Previously unclassified mutant; now likely class 2	likely class 2
68	538C>T	P180S	Unclassified	Unclassified	10.9	0.18	very low	Previously unclassified mutant; now likely class 2	likely class 2
135	505G>C	A169P	Unclassified	Unclassified	26.0	0.13	very low	Previously unclassified mutant; now likely class 2	likely class 2
168	509C>A	P170H	Unclassified	Unclassified	9.4	0.27	very low	Previously unclassified mutant; now likely class 2	likely class 2
159	53G>A	G18D	VUS, novel	VUS, novel	27.4	0.27	very low	newly determined as pathogenic	likely class 2, 3, or 4
197	302G>T	G101V	VUS, novel	VUS, novel	63.5	0.48	low	newly determined as pathogenic	likely class 2, 3, or 4
203	538C>A	P180T	VUS, novel	VUS, novel	10.1	0.33	very low	newly determined as pathogenic	likely class 2
13	151G>C	G51R	IIa/II/2	2	47.2	0.24	low	low level - but higher than typical class 2	was IIa/II/2, now probably unclassified
20	173C>G	T58R	IIa/II/2	2	69.1	1.28	low	low level - but higher than typical class 2	was IIa/II/2, now probably unclassified
25	263T>C	L88P	II/Unclassified	2	58.8	0.36	low	low level - but higher than typical class 2	was II/Unclassified/2, now probably unclassified
33	325G>A	G109R	IIa/Unclassified	2	76.9	1.03	low	low level - but higher than typical class 2	was IIa/Unclassified/2, now probably unclassified
56	499T>C	C167R	IIb/II/2	2	52.9	0.53	low	low level - but higher than typical class 2	was IIb/II/2, now probably unclassified
74	557C>G	S186W	IIa/Unclassified	2	63.8	1.23	low	low level - but higher than typical class 2	was IIa/Unclassified/2, now probably unclassified
83	620T>G	M207R	IIa/Unclassified	2	74.8	1.07	low	low level - but higher than typical class 2	was IIa/Unclassified/2, now probably unclassified
92	647T>A	M216K	IIb+IIc/Unclassified	2	76.8	1.62	low	low level - but higher than typical class 2	was IIb+IIc/Unclassified/2, now probably unclassified
106	886A>G	K296E	Unclassified/II/2	2	61.3	0.80	low	low level - but higher than typical class 2	was Unclassified/II/2, now probably unclassified
178	302G>A	G101E	Unclassified	Unclassified	74.7	0.57	low	literature report unclear - now suggests likely pathogenic	likely class 2, 3, or 4
186	140T>G	L47R	VUS, database/Unclassifie	VUS, database	11.2	0.15	very low	literature correct over databases- pathogenic	likely class 2
55	491C>T	A164V	IIa/II/2	2	85.0	2.10	indeterminate	intermediate level - not typical of class 2	was IIa/II/2, now unclassified
15	152G>T	G51V	IIa/II/2	2	95.5	1.69	indeterminate	intermediate level - not typical of class 2	was IIa/II/2, now unclassified
16	155T>A	F52Y	IIc+IVb/Unclassified	2	83.3	1.79	indeterminate	intermediate level - not typical of class 2	was IIc+IVb/2, now IVb
18	167T>A	F56Y	IIc+IVb/Unclassified	2	94.9	1.72	indeterminate	intermediate level - not typical of class 2	was IIc+IVb/2, now IVb
21	173C>T	T58M	Unclassified/2	2	94.9	1.91	indeterminate	intermediate level - not typical of class 2	was 2, now unclassified
107	887A>T	K296M	Unclassified/2	2	65.3	2.69	indeterminate	intermediate level - not typical of class 2	was 2, now unclassified
108	888G>T	K296N	Unclassified/2	2	79.4	2.97	indeterminate	intermediate level - not typical of class 2	was 2, now unclassified
40	374T>G	L125R	IIa/II/2	2	91.6	2.57	high	high - not consistent with class 2	was IIa/II/2, now unclassified
50	448G>A	E150K	IIa	2	97.3	2.26	high	high - not consistent with class 2	was IIa/2, now unclassified
26	266G>A	G89D	IIa/II/2	2	33.8	0.23	very low	databases correct over literature - pathogenic	likely class 2
57	501C>G	C167W	IIb/II/2	2	17.5	0.18	very low	databases correct over literature - pathogenic	likely class 2
67	538C>G	P180A	IIa/Unclassified	2	25.3	0.16	very low	databases correct over literature - pathogenic	likely class 2
101	810C>A	S270R	IIc/II/2	2	18.4	0.19	very low	databases correct over literature - pathogenic	likely class 2

Note: RHO: rhodopsin; VUS: variants of unknown significance.

Previously-reported *RHO* mutation categories ("Previous class, detailed") were revised ("Revised Class") based on RHO surface expression results. Color-coded gradients show pathogenic levels (red) and wild-type levels (green). For standardized variant descriptions (HGVS format), add prefixes NM_000539.3: for DNA and RHO_v001: or NP_000530.1: for protein descriptions

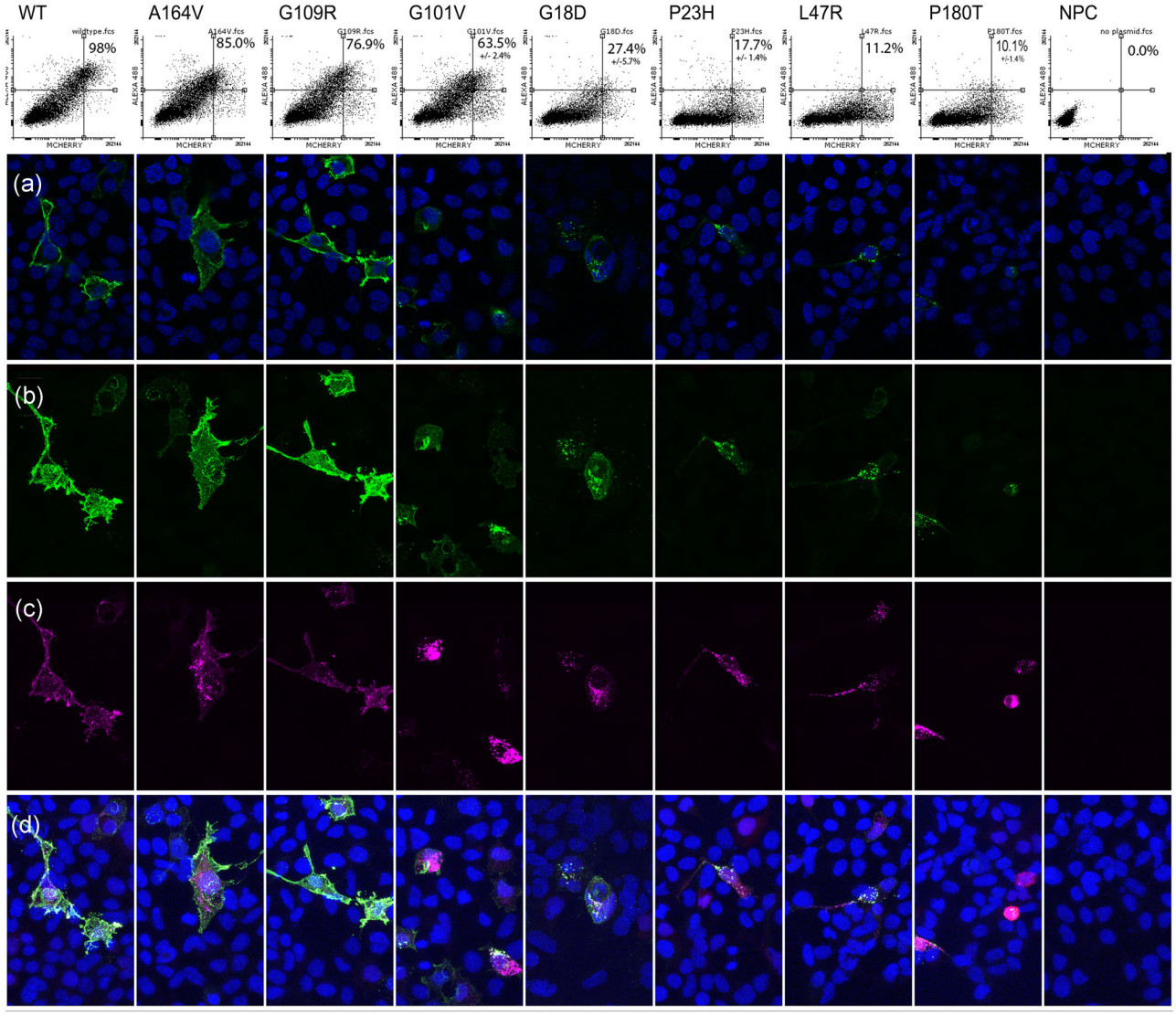


FIGURE 7 Confocal immunofluorescence of selected mutants validates flow cytometry results. (a) Each column represents a transfection with a particular labeled variant, arranged from highest (left) to lowest (right) RHO expression (WT = wildtype). Top row: flow cytometry results showing surface RHO staining (*y*-axis) and transfection marker (*x*-axis), with the percent of transfected cells with high surface RHO as an inset. Confocal staining: green = RHO cell surface staining, purple = total RHO staining after cell permeabilization, blue = nuclear stain. Row (a): single-slice confocal images demonstrate cell membrane staining for the left-most three variants but fainter, punctate perinuclear staining for the other variants. A similar pattern of decreasing RHO intensity is seen in maximum projection images of (b) surface RHO and (c) total RHO. (d) Row shows a composite of all channels. RHO: rhodopsin

strands after mutagenesis and after it was recombined into the final expression vector; optimization of fixation and staining conditions to maintain high discrimination between positive and negative controls, while also using a fixative that preserved DNA integrity for subsequent PCR; and quality filtering of NGS basecalls to maintain signal-to-noise in variant quantification.

The data in Tables 1 and S1 provide new information about previously-uncharacterized variants, identify apparently misclassified variants, and confirm pathogenicity of known mutations. Of 80 known class 2 mutants, 71 (89%) were confirmed to show pathogenically low levels of RHO surface expression. However, 9 (11%) class 2 mutants unexpectedly showed high or intermediate levels, especially those found in transmembrane helix I 0.8 of 66 (12%) mutants in other mutation classes also showed pathogenic RHO surface expression levels, especially in class 3 and 4 mutants, which tended to have an intermediate phenotype. A total of 14 of 33 (42%) unclassified pathogenic variants showed pathogenic expression levels, which are likely class 2, 3, or 4 mutants.

For example, *RHO* A292E does not cause RP but instead causes the milder disease congenital stationary night blindness. *RHO* p.A292E is known to be constitutively active in activating transduction without a chromophore (class 6; Dryja et al., 1993). Computation predictions based on misfolding propensity (ΔΔG; Rakoczy et al., 2011) led to the conclusion that this mutant should also be grouped in class 2 (IIa), but the relatively normal actual expression level of this variant indicates that it should be categorized as class 6 only.

TABLE 2 Comparison of information from computational methods and function assays to solve patient pedigrees

Variant		Allele rarity		Computational predictions					Functional assays				
DNA description	Protein Description	Gnomad allele frequency	Internal frequency (# samples)	SIFT	Polyphen	CADD	Phast Cons	gerpRS	Variant number	unpooled surface RHO assay, mean	pooled surface RHO assay, mean	RHO surface expression, final category	Conclusions based on functional data
c.53G>A	p.G18D	0.000077	2	tolerated	probably_damaging	24.7	1	5.5	159	**27.4**	**0.27**	very low	newly determined as pathogenic
c.178T>C	p.Y60H	0.000037	1	tolerated	probably_damaging	23.6	1	5.5	194	**90.4**	**3.95**	high	not informative
c.185C>A	p.T62N	0.000016	1	deleterious	probably_damaging	27.4	1	5.5	195	**94.9**	**4.54**	high	not informative
c.218A>G	p.N73S	0.000004	1	deleterious	probably_damaging	25.1	1	5.5	196	**94.8**	**2.94**	high	not informative
c.302G>T	p.G101V	0	1	deleterious	probably_damaging	25.6	1	4.4	197	**63.5**	**0.48**	low	newly determined as pathogenic
c.439C>T	p.R147C	0.000183	1	deleterious	probably_damaging	35	1	4.3	198	**87.6**	**3.25**	high	not informative
c.538C>A	p.P180T	0	1	deleterious	probably_damaging	24	1	4.2	203	**10.1**	**0.33**	very low	newly determined as pathogenic
c.755G>A	p.R252H	0.000016	1	deleterious	probably_damaging	33	1	5.5	199	**94.5**	**2.57**	high	not informative
c.895G>A	p.A299T	0.000020	1	tolerated	benign	0.01	0.01	-11.0	200	**88.1**	**2.06**	indeterminate	not informative
c.913A>G	p.I305V	0.000004	1	deleterious	probably_damaging	26.1	1	5.5	201	**96.0**	**2.49**	high	not informative

Note: RHO: rhodopsin; VUS: variants of unknown significance.

Among nine rare VUS that were identified from our genetic testing service, three variants had measured pathogenic expression levels. Color coding for computational methods is based on arbitrary cutoffs for SIFT (Ng & Henikoff, 2003), PolyPhen (Adzhubei, Jordan, & Sunyaev, 2013), CADD scores (Kircher et al., 2014), PhastCons (Siepel et al., 2005), and GerpRS (Cooper et al., 2005). Color coding for functional assays is based on the RHO surface expression final category. For standardized variant descriptions (HGVS format), add prefixes NM_000539.3:c. for DNA and RHO_v001:p. or NP_000530.1:p. for protein descriptions.

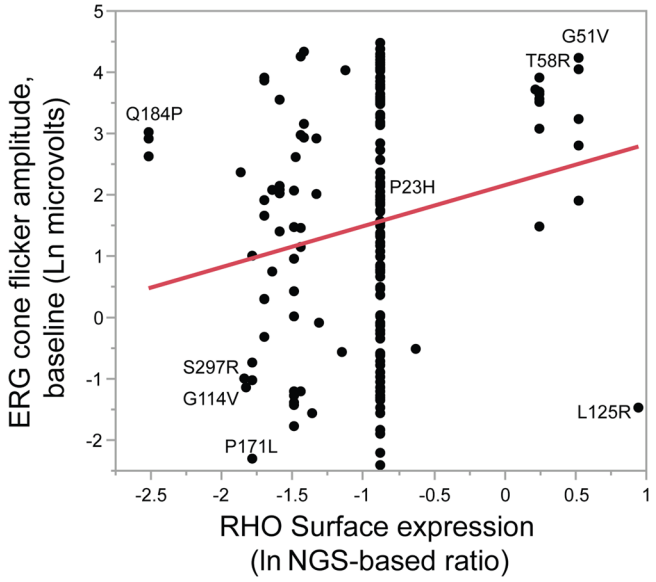


FIGURE 8 Correlating *RHO* surface expression and clinical disease severity. Among known class 2 RHO mutants, increasing amounts of RHO surface expression correlated with a milder clinical phenotype, as represented by the baseline cone flicker ERG response amplitude ($p = 0.012$). Each point represents one patient. ERG: electroretinography; RHO: rhodopsin

Conversely, p.T193M was predicted to fold correctly (Rakoczy et al., 2011), but in this study, p.T193M is not expressed well on the cell surface. It is not known if this defect is due to misfolding or a hypothesized defect in dimerization (Rakoczy et al., 2011).

This expanded dataset also provides more confidence in variant interpretation based on smaller numbers of variants or among different experimental systems. For example, D190N showed low expression levels in this study, in contrast to smaller studies showing moderate expression using a GFP fusion tag (McKeone et al., 2014) or normal localization using a bovine sequence backbone (Liu et al., 2013).

Some special variant classes deserve further discussion. Even though RHO surface expression in HEK293 cells has been widely used to demonstrate pathogenicity (Chen et al., 2011; Chuang et al., 2004; Davies et al., 2012; Hollingsworth & Gross, 2013; Li et al., 1998; Liu et al., 2013; McKeone et al., 2014; Sung et al., 1993; Toledo et al., 2011; Yamasaki et al., 2014), recessive mutations do not show any special features in this assay, and are not distinguished from the dominant mutations. The recessive mutations are not clustered in a certain segment of the gene. However, because the number of known recessive mutants is small (6.6%; 12 of 181 according to Table S1), it is much more likely that any particular low-expressed mutation is dominant. Therefore, this study provides good evidence that the "very low" and "low" expressing variants are pathogenic mutations, but cannot conclusively state that they are *dominant* pathogenic mutations (thus, the three families mentioned above are "provisionally" solved). Our initial experiments mixing together a known dominant mutation with a wildtype construct showed no effect on trafficking of the wildtype protein in this system (Figure S3). Modifications of the cell type used or of expression levels may be able to reveal such dominant negative or gain of function effects to distinguish dominant pathogenic mutations from recessive pathogenic mutations. In a different system using the SK-N-SH cell type, dominant negative effects of dominant mutants have been demonstrated (Mendes & Cheetham, 2008).

This assay was focused on identifying class 2 (misfolding) mutations, which is the most common class. Figure 6 shows that class 3 mutations (endocytosis) and some class 4 mutations (altered posttranslational modifications and reduced stability) also displayed low expression, though not quite as low as class 2, on average. Interestingly, the older Rakoczy classification separates the class 4 mutants slightly better, with the class IVa (affecting a glycosylation site; p.T4K, p.T17M, p.N15S) all showing pathogenic surface expression levels (note that three variants —p.T4K, p.P12R, and p.N15S are at or near the Ret-P1 antibody epitope at amino acids 4–10, so the results from these variants should be interpreted with caution).

Additional assays are needed to detect and distinguish other mutation classes. For example, to identify mutations that are rescued by retinal binding (Behnen et al., 2018; Krebs et al., 2010; Mendes & Cheetham, 2008; Mendes et al., 2010), the assay could be repeated in the presence of retinal (which was too toxic for use in the current transient transfection system; not shown). For context, the classical assay for analysis of RHO folding and retinal binding is ultraviolet-visible spectroscopy. However, this assay does not scale well to hundreds of variants. Therefore this study makes a compromise in that it implements an assay that is not the gold standard but, in exchange, is amenable to higher throughput. The larger dataset produced by this higher throughput has the advantage that cross-comparison within the dataset (e.g. Figure 6) produces "internal calibrators" that increase the ability the interpret the data that is collected. However, working within the constraints of assays that can be parallelized will require some creativity to implement more general pathogenicity detectors. For the purpose of evaluating pathogenicity, general toxicity- or stress-based assays would be ideal. In this cell culture system, expression of p.P23H or p.T17M mutants had no effect on cell death as measured by membrane permeability dyes, on apoptosis as measured by annexin V labeling, or on endoplasmic reticulum stress as measured by the pCAX-XBP1delDBD-venus reporter (Iwawaki, Akai, Kohno, & Miura, 2004; not shown). Other approaches to look for these more general effects might use different cell types or even in vivo. Thus, a panel of both general and specific assays is likely the best approach to identify all types of pathogenic mutations.

Truncating mutations (nonsense or frameshift) were expressed in the context of a cDNA without internal introns, after the synthetic intron in the CAG promoter. Because internal introns and their potential effects of nonsense-mediated decay (Roman-Sanchez, Wensel, & Wilson, 2016) were not included, the results for those mutants should be interpreted with caution. All truncating variants 5′ of a.a. 315 showed pathogenic expression levels, while all variants 3′ to a.a. 332, including 10 truncating variants, showed wildtype expression levels. These variants should preferably be retested in

their full genomic context (Roman-Sanchez et al., 2016), which would also allow for testing of splice variants.

4.1 | Screening small versus large variant libraries

This library construction strategy differs from larger-scale screening strategies (Brenan et al., 2016; Gasperini et al., 2016; Melnikov et al., 2014) (thousands of variants) in that this library, with 210 variants, was created with widely-available techniques at lower technical complexity. The library itself was constructed using modified Quikchange mutagenesis reactions using inexpensive short oligos ordered in 96-well plates (Synthetic DNA variant libraries may replace this step as prices continue to decrease). Each variant plasmid was introduced into cells using a standard transfection protocol, without the need for viral vectors. The disadvantage of this approach is that it does not scale well to larger sized libraries. Sequencing and reracking of individual clones, as well as the unpooled transfection, become more cumbersome with an increasing number of variants and wells. However, for biological questions such as the one addressed in this study involving about 200 variants or less, this approach is feasible and has the above advantages.

A smaller library size also has advantages in achieving a quantitative readout. In a very large, barcode-free, pooled assay, NGS sequencing errors would set a lower boundary so that the rarest variants are harder to quantify, and therefore these assays would best be used to find highly enriched "hits". In this study, because of the smaller number of variants, the quantitation of each variant could be maintained above the noise floor caused by sequencing errors. Therefore, smaller gradations in expression level were reproducibly quantified over the entire dynamic range of the assay (Figure S2), allowing for observations such as the "intermediate" expression phenotype of class 3 and 4 mutations described above. More specifically, for the quantitation of each of 210 plasmids in a mixed pool, the frequency of an individual plasmid (ideally ~1/210 = ~0.5%) has to be greater than the background noise of sequencing errors—theoretically 0.1% on the Illumina MiSeq at Q = 30, though sequence-context and cycle number dependent. With library inhomogeneity and context-dependent sequencing error rates, this ratio is not guaranteed to be maintained. Careful masking of low-quality basecalls allowed for low-frequency variant quantitation above the noise level (Section 2 and Figure 4b,c). For larger barcode-free libraries, maintaining accurate quantification at low levels in the presence of read errors would not be solved by simply increasing read depth; specialized wet-lab assays to lower the actual sequencing error rate (e.g. Schmitt et al., 2012) could be considered. Alternatively, barcoded libraries designed to avoid barcode collisions in the presence of sequencing errors could be used.

4.2 | Predicting clinical disease severity data from surface expression levels

Among known class 2 mutants, the level of RHO surface expression measured in this study was better at predicting disease severity than computational folding predictions (Figure 8 and Section 3). These approaches are not mutually exclusive, however, and structural biology can be used as a complementary approach to experimental observations. For example, folding calculations can be refined based on empirical data and then extended to new mutants. For variants that express well on the cell surface or where the mechanism is different (e.g. class 7 dimerization mutations) other classes of bioinformatic predictors are needed.

The outliers in Figure 8 that do not follow the general trend of less severe disease with increasing surface expression can be informative as well. For example, p.Q184P shows low severity despite low expression, and p.L125R shows high severity despite high expression (Figure 8). These data suggest that these variants are likely not classic class 2 mutations and influence disease course through a different mechanism than simple misfolding.

The correlation between surface expression and disease severity can produce some clinically-useful estimates. For example, a typical class 2 mutation with a low surface expression ratio of −1.5 would have a predicted baseline ERG which is 1.34 natural log units lower than a mutation with a higher expression ratio of 0.5. At an average rate of progression (0.091 Ln units per year; Berson et al., 2002), this corresponds to an extra 15 years of vision. If outliers (e.g. Q184P, L125R) in the regression had been excluded, this estimated effect would be larger. These estimates are based on the average severity of a mutant. In particular, L125R is probably not a class 2 mutation (Figure 6). However, there is a large amount of variation between individual subjects, particularly notable for the p.P23H mutation which is common in our cohort (Dryja et al., 1990; Hartong et al., 2009); additional modifying factors are yet to be identified, whether genetic, environmental, or stochastic (Chow, Kelsey, Wolfner, & Clark, 2016).

In summary, a functional genomics approach can be used to address the problem of VUS in inherited retinal diseases, which in general is currently one of the major bottlenecks in the diagnosis of human Mendelian diseases. Future studies may include generalizing these assays to more genes and more mutation types, as well as using more complex methods to screen larger numbers of variants.

ACKNOWLEDGMENTS

We would like to acknowledge Mark Consugar, Mindy Kwong, the CCIB DNA Core Facility at Massachusetts General Hospital for assistance with sequencing and colony processing, and David Dombrowski in the Massachusetts General Hospital Flow Cytometry Core for flow cytometry and flow sorting.

We are grateful for funding from NEI K12 EY016335-10 (JC), the Foundation Fighting Blindness Enhanced Career Development Award (JC), the Research for Prevention of Blindness Career Development Award (JC), the Massachusetts Lions Foundation/Lions Eye Research Foundation, and NEI/NIH P30 EY014104.

DISCLOSURE

JC serves as a consultant on a Blue Cross Blue Shield advisory board for gene therapy policy. JC is/was a consultant for AGTC, Beam Therapeutics, RBS, Editas Medicine, Gensight, and Sanofi, not directly related to this study. EP receives research funding from Casebia Therapeutics, not directly related to this study. JC and EP are also investigators for gene therapy clinical trials, not directly related to this study.

ORCID

Jason Comander http://orcid.org/0000-0002-4257-7145

REFERENCES

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*, *Chapter 7*, *76*, 20–7.20.41. https://doi.org/10.1002/0471142905.hg0720s76

Al-Merjan, J. I., Pandova, M. G., Al-Ghanim, M., Al-Wayel, A., & Al-Mutairi, S. (2005). Registered blindness and low vision in Kuwait. *Ophthalmic Epidemiology*, *12*(4), 251–257.

Athanasiou, D., Aguila, M., Bellingham, J., Li, W., McCulley, C., Reeves, P. J., & Cheetham, M. E. (2018). The molecular and cellular basis of rhodopsin retinitis pigmentosa reveals potential strategies for therapy. *Progress Retinal Eye Research*, *62*, 1–23. https://doi.org/10.1016/j.preteyeres.2017.10.002

Bean, L. J., Tinker, S. W., da Silva, C., & Hegde, M. R. (2013). Free the data: One laboratory's approach to knowledge-based genomic variant classification and preparation for EMR integration of genomic data. *Human Mutation*, *34*(9), 1183–1188. https://doi.org/10.1002/humu.22364

Behnen, P., Felline, A., Comitato, A., Di Salvo, M. T., Raimondi, F., Gulati, S., … Fanelli, F. (2018). A small chaperone improves folding and routing of rhodopsin mutants linked to inherited blindness. *Science*, *4*, 1–19. https://doi.org/10.1016/j.isci.2018.05.001

Berger, W., Kloeckener-Gruissem, B., & Neidhardt, J. (2010). The molecular basis of human retinal and vitreoretinal diseases. *Progress in Retinal and Eye Research*, *29*(5), 335–375. https://doi.org/10.1016/j.preteyeres.2010.03.004

Berson, E. L., Rosner, B., Weigel-DiFranco, C., Dryja, T. P., & Sandberg, M. A. (2002). Disease progression in patients with dominant retinitis pigmentosa and rhodopsin mutations. *Investigative Ophthalmology & Visual Science*, *43*(9), 3027–3036.

Bosch, L., Ramon, E., Del Valle, L. J., & Garriga, P. (2003). Structural and functional role of helices I and II in rhodopsin. A novel interplay evidenced by mutations at Gly-51 and Gly-89 in the transmembrane domain. *The Journal of Biological Chemistry*, *278*(22), 20203–20209. https://doi.org/10.1074/jbc.M301319200

Bosch-Presegue, L., Ramon, E., Toledo, D., Cordomi, A., & Garriga, P. (2011). Alterations in the photoactivation pathway of rhodopsin mutants associated with retinitis pigmentosa. *The FEBS journal*, *278*(9), 1493–1505. https://doi.org/10.1111/j.1742-4658.2011.08066.x

Braman, J., Papworth, C., & Greener, A. (1996). Site-directed mutagenesis using double-stranded plasmid DNA templates. *Methods in molecular biology (Clifton, N.J.)*, *57*, 31–44. https://doi.org/10.1385/0-89603-332-5:31

Brenan, L., Andreev, A., Cohen, O., Pantel, S., Kamburov, A., Cacchiarelli, D., … Johannessen, C. M. (2016). Phenotypic characterization of a comprehensive set of MAPK1/ERK2 missense mutants. *Cell Reports*, *17*(4), 1171–1183. https://doi.org/10.1016/j.celrep.2016.09.061

Buch, H., Vinding, T., La, C. M., Appleyard, M., Jensen, G. B., & Nielsen, N. V. (2004). Prevalence and causes of visual impairment and blindness among 9980 Scandinavian adults: The Copenhagen City Eye Study. *Ophthalmology*, *111*(1), 53–61.

Bujakowska, K. M., White, J., Place, E., Consugar, M., & Comander, J. (2015). Efficient in silico identification of a common insertion in the MAK gene which causes retinitis pigmentosa. *PLoS One*, *10*(11), e0142614. https://doi.org/10.1371/journal.pone.0142614

Caminsky, N., Mucaki, E. J., & Rogan, P. K. (2014). Interpretation of mRNA splicing mutations in genetic disease: Review of the literature and guidelines for information-theoretical analysis. *F1000Research*, *3*, 282. https://doi.org/10.12688/f1000research.5654.1

Chen, Y. F., Wang, I. J., Lin, L. L., & Chen, M. S. (2011). Examining rhodopsin retention in endoplasmic reticulum and intracellular localization in vitro and in vivo by using truncated rhodopsin fragments. *Journal of Cellular Biochemistry*, *112*(2), 520–530. https://doi.org/10.1002/jcb.22942

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One*, *7*(10), e46688. https://doi.org/10.1371/journal.pone.0046688

Chow, C. Y., Kelsey, K. J., Wolfner, M. F., & Clark, A. G. (2016). Candidate genetic modifiers of retinitis pigmentosa identified by exploiting natural variation in Drosophila. *Human Molecular Genetics*, *25*(4), 651–659. https://doi.org/10.1093/hmg/ddv502ddv502

Chuang, J. Z., Vega, C., Jun, W., & Sung, C. H. (2004). Structural and functional impairment of endocytic pathways by retinitis pigmentosa mutant rhodopsin-arrestin complexes. *The Journal of Clinical Investigation*, *114*(1), 131–140. https://doi.org/10.1172/JCI21136

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., … Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, *31*(3), 213–219. https://doi.org/10.1038/nbt.2514

Cideciyan, A. V., Sudharsan, R., Dufour, V. L., Massengill, M. T., Iwabe, S., Swider, M., … Beltran, W. A. (2018). Mutation-independent rhodopsin gene therapy by knockdown and replacement with a single AAV vector. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(36), E8547–E8556. https://doi.org/10.1073/pnas.1805055115

Comander, J., Weigel-DiFranco, C., Maher, M., Place, E., Wan, A., Harper, S., … Pierce, E. A. (2017). Thegenetic basis of pericentral retinitis pigmentosa-A form of mild retinitis pigmentosa. *Genes (Basel)*, *8*(10) https://doi.org/10.3390/genes8100256

Consugar, M. B., Navarro-Gomez, D., Place, E. M., Bujakowska, K. M., Sousa, M. E., Fonseca-Kelly, Z. D., … Pierce, E. A. (2015). Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *17*(4), 253–261. https://doi.org/10.1038/gim.2014.172

Cooper, G. M., Stone, E. A., Asimenos, G., Program, N. C. S., Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, *15*(7), 901–913. https://doi.org/10.1101/gr.3577405

Corton, M., Nishiguchi, K. M., Avila-Fernandez, A., Nikopoulos, K., Riveiro-Alvarez, R., Tatu, S. D., … Rivolta, C. (2013). Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. *PLoS One*, *8*(6), e65574. https://doi.org/10.1371/journal.pone.0065574

Daiger, S. P., Rossiter, B. J. F., Greenberg, J., Christoffels, A., & Hide, W. (1998). *Data services and software for identifying genes and mutations causing retinal degeneration*.

Davies, W. I., Downes, S. M., Fu, J. K., Shanks, M. E., Copley, R. R., Lise, S., … Nemeth, A. H. (2012). Next-generation sequencing in health-care delivery: Lessons from the functional analysis of rhodopsin. *Genetics in*

Medicine: Official Journal of the American College of Medical Genetics, *14*(11), 891–899. https://doi.org/10.1038/gim.2012.73

Dizhoor, A. M., Woodruff, M. L., Olshevskaya, E. V., Cilluffo, M. C., Cornwall, M. C., Sieving, P. A., & Fain, G. L. (2008). Night blindness and the mechanism of constitutive signaling of mutant G90D rhodopsin. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, *28*(45), 11662–11672. https://doi.org/10.1523/JNEUROSCI.4006-08.2008

Dryja, T. P., Berson, E. L., Rao, V. R., & Oprian, D. D. (1993). Heterozygous missense mutation in the rhodopsin gene as a cause of congenital stationary night blindness. *Nature Genetics*, *4*(3), 280–283. https://doi.org/10.1038/ng0793-280

Dryja, T. P., McGee, T. L., Reichel, E., Hahn, L. B., Cowley, G. S., Yandell, D. W., … Berson, E. L. (1990). A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature*, *343*(6256), 364–366. https://doi.org/10.1038/343364a0

Duncan, J. L., Pierce, E. A., Laster, A. M., Daiger, S. P., Birch, D. G., Ash, J. D., … Zarbin, M. A., and the Foundation Fighting Blindness Scientific Advisory, B. (2018). Inherited retinal degenerations: Current landscape and knowledge gaps. *Translational Vision Science & Technology*, *7*(4), 6. https://doi.org/10.1167/tvst.7.4.6

Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., … Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. *Nature*, *562*(7726), 217–222. https://doi.org/10.1038/s41586-018-0461-z

Fox, E. J., Reid-Bayliss, K. S., Emond, M. J., & Loeb, L. A. (2014). Accuracy of next generation sequencing platforms. *Next Gener Seq Appl*, *1*, https://doi.org/10.4172/jngsa.1000106

Gasperini, M., Starita, L., & Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. *Nature Protocols*, *11*(10), 1782–1787. https://doi.org/10.1038/nprot.2016.135

Glockle, N., Kohl, S., Mohr, J., Scheurenbrand, T., Sprecher, A., Weisschuh, N., … Neidhardt, J. (2014). Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. *European Journal Of Human Genetics: EJHG*, *22*(1), 99–104. https://doi.org/10.1038/ejhg.2013.72

Gray, V. E., Hause, R. J., Luebeck, J., Shendure, J., & Fowler, D. M. (2018). Quantitative missense variant effect prediction using large-scale mutagenesis data. *Cell Syst*, *6*(1), 116–124. https://doi.org/10.1016/j.cels.2017.11.003

Hartong, D. T., Berson, E. L., & Dryja, T. P. (2006). Retinitis pigmentosa. *Lancet*, *368*(9549), 1795–1809.

Hartong, D. T., McGee, T. L., Sandberg, M. A., Berson, E. L., Asselbergs, F. W., van der Harst, P., … Dryja, T. P. (2009). Search for a correlation between telomere length and severity of retinitis pigmentosa due to the dominant rhodopsin Pro23His mutation. *Molecular Vision*, *15*, 592–597. https://doi.org/60

Hata, H., Yonezawa, M., Nakanishi, T., Ri, T., & Yanashima, K. (2003). Causes of entering institutions for visually handicapped persons during the past fifteen years. *Japanese Journal Clinical Ophthalmology*, *57*, 259–262.

Hollingsworth, T. J., & Gross, A. K. (2013). The severe autosomal dominant retinitis pigmentosa rhodopsin mutant Ter349Glu mislocalizes and induces rapid rod cell death. *The Journal of Biological Chemistry*, *288*(40), 29047–29055. https://doi.org/10.1074/jbc.M113.495184

Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., … Stoppa-Lyonnet, D. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Human Mutation*, *33*(8), 1228–1238. https://doi.org/10.1002/humu.22101

Huang, X. F., Wu, J., Lv, J. N., Zhang, X., & Jin, Z. B. (2015). Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: A complementary approach to clinical genetic diagnostic testing. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *17*(4), 307–311. https://doi.org/10.1038/gim.2014.193

Iwawaki, T., Akai, R., Kohno, K., & Miura, M. (2004). A transgenic mouse model for monitoring endoplasmic reticulum stress. *Nature Medicine*, *10*(1), 98–102. https://doi.org/10.1038/nm970

Jiang, Y., Turinsky, A. L., & Brudno, M. (2015). The missing indels: An estimate of indel variation in a human genome and analysis of factors that impede detection. *Nucleic Acids Research*, *43*(15), 7217–7228. https://doi.org/10.1093/nar/gkv677

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. https://doi.org/10.1038/ng.2892

Krebs, M. P., Holden, D. C., Joshi, P., Clark, C. L., 3rd, Lee, A. H., & Kaushal, S. (2010). Molecular mechanisms of rhodopsin retinitis pigmentosa and the efficacy of pharmacological rescue. *Journal of Molecular Biology*, *395*(5), 1063–1078. https://doi.org/10.1016/j.jmb.2009.11.015

Lappalainen, T. (2015). Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Research*, *25*(10), 1427–1431. https://doi.org/10.1101/gr.190983.115

Li, T., Sandberg, M. A., Pawlyk, B. S., Rosner, B., Hayes, K. C., Dryja, T. P., & Berson, E. L. (1998). Effect of vitamin A supplementation on rhodopsin mutants threonine-17—> methionine and proline-347—>serine in transgenic mice and in cell cultures. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(20), 11933–11938.

Liu, M. Y., Liu, J., Mehrotra, D., Liu, Y., Guo, Y., Baldera-Aguayo, P. A., … Yan, E. C. (2013). Thermal stability of rhodopsin and progression of retinitis pigmentosa: Comparison of S186W and D190N rhodopsin mutants. *The Journal of Biological Chemistry*, *288*(24), 17698–17712. https://doi.org/10.1074/jbc.M112.397257

McKeone, R., Wikstrom, M., Kiel, C., & Rakoczy, P. E. (2014). Assessing the correlation between mutant rhodopsin stability and the severity of retinitis pigmentosa. *Molecular Vision*, *20*, 183–199.

Melnikov, A., Rogov, P., Wang, L., Gnirke, A., & Mikkelsen, T. S. (2014). Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Research*, *42*(14), e112–e112. https://doi.org/10.1093/nar/gku511

Mendes, H. F., & Cheetham, M. E. (2008). Pharmacological manipulation of gain-of-function and dominant-negative mechanisms in rhodopsin retinitis pigmentosa. *Human Molecular Genetics*, *17*(19), 3043–3054. https://doi.org/10.1093/hmg/ddn202

Mendes, H. F., Zaccarini, R., & Cheetham, M. E. (2010). Pharmacological manipulation of rhodopsin retinitis pigmentosa. *Advances in Experimental Medicine and Biology*, *664*, 317–323. https://doi.org/10.1007/978-1-4419-1399-9_36

Mendes, H. F., van der Spuy, J., Chapple, J. P., & Cheetham, M. E. (2005). Mechanisms of cell death in rhodopsin retinitis pigmentosa: Implications for therapy. *Trends in Molecular Medicine*, *11*(4), 177–185. https://doi.org/10.1016/j.molmed.2005.02.007

National Eye Institute. (2012,). Vision Research: Needs, Gaps, and Opportunities. Retrieved from https://nei.nih.gov/catalog/vision-research-needs-gaps-and-opportunities

Neveling, K., Collin, R. W., Gilissen, C., van Huet, R. A., Visser, L., Kwint, M. P., … Scheffer, H. (2012). Next-generation genetic testing for retinitis pigmentosa. *Human Mutation*, *33*(6), 963–972. https://doi.org/10.1002/humu.22045

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814.

Opefi, C. A., South, K., Reynolds, C. A., Smith, S. O., & Reeves, P. J. (2013). Retinitis pigmentosa mutants provide insight into the role of the N-terminal cap in rhodopsin folding, structure, and function. *The Journal of Biological Chemistry*, *288*(47), 33912–33926. https://doi.org/10.1074/jbc.M113.483032

Rakoczy, E. P., Kiel, C., McKeone, R., Stricher, F., & Serrano, L. (2011). Analysis of disease-linked rhodopsin mutations based on structure, function, and protein stability calculations. *Journal of Molecular Biology*, *405*(2), 584–606. https://doi.org/10.1016/j.jmb.2010.11.003

RetNet. (2017,). RetNet, the Retinal Information Network. Retrieved from http://www.sph.uth.tmc.edu/Retnet/

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Roman-Sanchez, R., Wensel, T. G., & Wilson, J. H. (2016). Nonsense mutations in the rhodopsin gene that give rise to mild phenotypes trigger mRNA degradation in human cells by nonsense-mediated decay. *Experimental Eye Research*, *145*, 444–449. https://doi.org/10.1016/j.exer.2015.09.013

Russell, S., Bennett, J., Wellman, J. A., Chung, D. C., Yu, Z. F., Tillman, A., … Maguire, A. M. (2017). Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: A randomised, controlled, open-label, phase 3 trial. *Lancet*, *390*(10097), 849–860. https://doi.org/10.1016/S0140-6736(17)31868-8

Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B., & Loeb, L. A. (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(36), 14508–14513. https://doi.org/10.1073/pnas.1208715109

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., … Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*(8), 1034–1050. https://doi.org/10.1101/gr.3715005

Spencer, D. H., Tyagi, M., Vallania, F., Bredemeyer, A. J., Pfeifer, J. D., Mitra, R. D., & Duncavage, E. J. (2014). Performance of common analysis methods for detecting low-frequency single nucleotide variants in targeted next-generation sequence data. *The Journal of Molecular Diagnostics: JMD*, *16*(1), 75–88. https://doi.org/10.1016/j.jmoldx.2013.09.003

Stoletzki, N., & Eyre-Walker, A. (2007). Synonymous codon usage in *Escherichia coli*: Selection for translational accuracy. *Molecular Biology and Evolution*, *24*(2), 374–381. https://doi.org/10.1093/molbev/msl166

Sung, C. H., Davenport, C. M., & Nathans, J. (1993). Rhodopsin mutations responsible for autosomal dominant retinitis pigmentosa. Clustering of functional classes along the polypeptide chain. *Journal of Biological Chemistry*, *268*(35), 26645–26649.

Thusberg, J., Olatubosun, A., & Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation*, *32*(4), 358–368. https://doi.org/10.1002/humu.21445

Toledo, D., Ramon, E., Aguila, M., Cordomi, A., Perez, J. J., Mendes, H. F., … Garriga, P. (2011). Molecular mechanisms of disease for mutations at Gly-90 in rhodopsin. *Journal of Biological Chemistry*, *286*(46), 39993–40001. https://doi.org/10.1074/jbc.M110.201517

U.S. National Institutes of Health. (2017). ClincalTrials.gov. Retrieved from http://clinicaltrials.gov/

Viner, C., Dorman, S. N., Shirley, B. C., & Rogan, P. K. (2014). Validation of predicted mRNA splicing mutations using high-throughput transcriptome data. *F1000Research*, *3*, 8. https://doi.org/10.12688/f1000research.3-8.v2

Wang, J., Zhang, V. W., Feng, Y., Tian, X., Li, F. Y., Truong, C., … Wong, L. J. (2014). Dependable and efficient clinical utility of target capture-based deep sequencing in molecular diagnosis of retinitis pigmentosa. *Investigative Ophthalmology & Visual Science*, *55*(10), 6213–6223. https://doi.org/10.1167/iovs.14-14936

Wester, K., Asplund, A., Backvall, H., Micke, P., Derveniece, A., Hartmane, I., … Ponten, F. (2003). Zinc-based fixative improves preservation of genomic DNA and proteins in histoprocessing of human tissues. *Laboratory Investigation*, *83*(6), 889–899.

Wilm, A., Aw, P. P., Bertrand, D., Yeo, G. H., Ong, S. H., Wong, C. H., … Nagarajan, N. (2012). LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, *40*(22), 11189–11201. https://doi.org/10.1093/nar/gks918

Yamasaki, A., Hara, T., Maejima, I., Sato, M., Sato, K., & Sato, K. (2014). Rer1p regulates the ER retention of immature rhodopsin and modulates its intracellular trafficking. *Scientific Reports*, *4*, 5973. https://doi.org/10.1038/srep05973

Zhou, X. E., Melcher, K., & Xu, H. E. (2012). Structure and activation of rhodopsin. *Acta Pharmacologica Sinica*, *33*(3), 291–299. https://doi.org/10.1038/aps.2011.171

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wan A, Place E, Pierce EA, Comander J. Characterizing variants of unknown significance in rhodopsin: A functional genomics approach. *Human Mutation*. 2019;40:1127–1144. https://doi.org/10.1002/humu.23762