A high throughput splicing assay to investigate the effect of variants of unknown significance on exon inclusion

Hilary A. Scott[1], Emily M. Place[1], Evelyn Harper[1], Sudeep Mehrotra[1], Rachel Huckfeldt[1], Jason Comander[1], Eric A. Pierce[1], Kinga M. Bujakowska[1*]

[1]Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA.

Correspondence:

Kinga M. Bujakowska, Ph.D.

Massachusetts Eye and Ear Infirmary, Harvard Medical School

243 Charles Street, Boston, MA 02114

Phone: 617-391-5933, Email: kinga_bujakowska@meei.harvard.edu

Abstract

Purpose: Inconclusive interpretation of pathogenicity of variants is a common problem in Mendelian disease diagnostics. We hypothesized that some variants of unknown significance (VUS) may lead to aberrant pre-mRNA splicing. To address this we have developed a high throughput splicing assay (HTSA) than can be utilized to test the effects of 1000s of variants on exon recognition.

Methods: 2296 reference, control and variant sequences from 380 exons of 89 genes associated with inherited retinal degenerations (IRDs) were cloned as a pool into a split-GFP HTSA construct and expressed in landing pad RCA7 HEK293T cells. Exon inclusion led to disruption of GFP and exon skipping led to GFP reconstitution, enabling to separate GFP+ve and GFP-ve cells by fluorescence activated cell sorting. After deep sequencing-based quantification of studied sequences in each cell pool, exon inclusion index (EII) was determined, where EII = GFP-ve oligo count/total oligo count.

Results: HTSA showed high reproducibility when compared between different biological replicates (tetrachoric correlation coefficient r = 0.91). Reference exon sequences showed a high level of exon recognition (median EII = 0.85) which was significantly reduced by mutations to the essential splice sites (donor site variants: median EII=0.06; acceptor site variants: median EII=0.39). Of the 1099 studied VUSs, 103 variants led to decreased exon inclusion (ΔEII ≤ -0.3) with 23 variants showing a strong effect (ΔEII ≤ -0,6). Using the HTSA data we were able to provide a likely genetic diagnosis to ten IRD cases.

Conclusion: HTSA offers a robust method to study the effects of VUSs on exon recognition allowing to provide new diagnoses for patients with Mendelian disorders.

Key Words: High throughput splicing assay, Mendelian trait, inherited retinal degeneration, intronic variant

INTRODUCTION

Despite substantial advances in next-generation sequencing technologies and their decreasing costs, genetic diagnoses are often hindered by our limited understanding of gene function, gene-disease relationship and uncertain pathogenicity of the uncovered variants[1–4]. For example, in the case of inherited retinal degenerations (IRDs), a Mendelian group of diseases affecting 2 million people worldwide[5], the cause of disease in ~30% of cases is unknown, despite substantial progress in genetic methodology and ~280 known IRD genes[6–13]. The remaining missing diagnoses are in part due to the currently unknown IRD genes, however their contribution will be likely small as shown by the limited number of genes discovered within the last decade[14–34]. This suggests that the missing genetic causality largely lies in elusive pathogenic variation in the already known IRD genes, including structural and non-coding changes of the genome[1,35–45]. Since next generation sequencing (NGS) produces a vast quantity of variants, solutions to some of the unsolved cases may already be available in the existing data. We therefore need robust functional assays to test large numbers of variants uncovered in the remaining unsolved cases.

One of the ways that a variant can lead to disease is by causing aberrant pre-mRNA splicing, resulting in transcript degradation due to premature stop codons or deletion of a crucial protein domain[46]. These variants are not always changes of the canonical splice sites, but can be non-canonical splice site mutations in the proximity of the exon[47], deep intronic variants[48–52], synonymous[51,53] or missense changes[54]. The pathogenicity of these variants is often difficult to predict even when using the recently developed machine learning algorithms designed for this purpose[55,56]. Due to the lack of certainty about the variants' functional consequence, they are labeled as variants of unknown significance (VUS) or ignored altogether in the clinical diagnostic reports and research. Apart from having an effect on splicing, a VUS may affect transcription or transcript stability if they fall within the enhancer, promoter or the untranslated regions (UTRs)[57–

[59]. Also, despite the existence of a variety of pathogenicity prediction tools[60–64], we lack sufficient data to interpret with confidence many of the missense variants, which can alter protein structure, enzymatic activity or interaction with other proteins or molecules. Our recent study of the genetics of IRDs showed that research "solutions" in ~23% of cases are based on one or more VUSs[13,65]. All of these classes of VUSs require sophisticated variant class or protein specific assays[54,58,66,67], which cannot be undertaken by a single laboratory. However, given the recent advances in the gene therapy field, an effort to study these variants is particularly important as it can finalize genetic diagnoses for patients, some of whom may be eligible for the emerging genetic therapies[68–74].

In this study we wanted to address the contribution of VUSs to aberrant splicing[75–78]. To address this challenge, we developed a high throughput splicing assay (HTSA) to test the effect of hundreds of VUSs on aberrant splicing simultaneously. Our assay is an adaptation of a previously published methodology by Cheung and colleagues[54] with an important alteration that allows for studying out-of-frame exons. We have tested our assay on 1131 variants in IRD genes and found likely genetic diagnoses for ten patients.

MATERIAL AND METHODS

Ethics declaration. The study was approved by the Institutional Review Board at the Massachusetts Eye and Ear (Human Studies Committee MEE, Mass General Brigham, USA) and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from all individuals on whom genetic testing and further molecular evaluations were performed.

Patient sequencing and annotation. Targeted panel sequencing was performed at MEE genomics core and exome sequencing was performed at the Center for Mendelian Genomics at

the Broad Institute of MIT and Harvard using methodology described previously[79]. Sequencing data was aligned to human genome 38 and variants were called using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3[80] (https://software.broadinstitute.org/gatk/), and annotated using VEP[81] and VCFAnno[82]. CNV predictions were produced using gCNV[80].

Exon selection for splicing assay and oligo design. Internal exons in known IRD genes that were 200 base pairs (bp) or less were selected as possible candidates for the splicing assay, which allowed for inclusion of at least 40bp of 5' intron (acceptor site), 30bp of the 3' intron (donor site) and 15 bp flanking adapter sequences containing *AgeI* and *NheI* restriction sites. After selecting exons that were not longer than 200bp and did not include the *AgeI* and *NheI* restriction sites, we used an in-house software, Mendelian analysis toolkit (MATK)[79] to prioritize rare variants in these exons (MAF <0.003) found in the unsolved IRD subjects. Additional rare variants in the *EYS* gene were taken from the gnomAD database[83]. In total 1131 variants in 380 IRD exons were selected (Tables S1 and S2). A pool of ssDNA oligos synthesized on silicone array-based method was purchased from a commercial vendor (TWIST Biosciences, USA).

High throughput splicing assay (HTSA) design. We have adapted a published approach from Cheung and colleagues[54], who kindly shared with us the *SMN1* intron backbone plasmid with split GFP-T2A-mCherry , the *Bxb1* integrase plasmid and the HEK2937-RCA7 cell line, that provide a chromosomal landing pad site which allows stable expression of splicing reporter library at the adeno-associated virus integration site 1 (AAVS1) safe harbor locus[54,84]. In our design, we altered the *SMN1* split GFP-T2A-mCherry plasmid, where we decoupled the mCherry and GFP expression to be able to study out-of-frame exons. For the uniform expression of both fluorophores, we introduced EF1alpha promoters in front of the split GFP and mCherry genes. We will refer to this altered vector as the HTSA vector. The vector also contains attB sites for integration into the *AAVS1* locus of the HEK293T-RCA7 landing pad cell line, as well as a promoter-less puromycin resistance cassette. HEK293T-RCA7 contain an attP

site and the promoter that allows for the site-specific recombination of the HTSA insert and coupling of the promoter and the puromycin resistance gene.

HTSA library design and cloning. We designed two oligo libraries: 1) control library containing 380 study exons with reference sequences, acceptor and donor site changes (1140 oligos in total); and 2) variant library containing 380 study exons with reference sequences, acceptor site changes, donor site changes and study variants (2296 sequences in total). Sequences that included the *AgeI* and *NheI* restriction sites were excluded. The library was amplified with Fwd-AgeI primer: 5'-ACGGCCAGTACCGGT-3' and NheI-Rev primer 5'-GCTATGACCGCTAGC-3' with a high fidelity polymerase (KAPA HiFi HotStart ReadyMix, Roche) for a maximum of 20 cycles and cloned to the HTSA vector by the *AgeI* and *NheI* restriction cloning.

Cell culture and FACS. After NGS confirmation of the desired diversity (Novaseq 6000, Illumina), the libraries were co-transfected with a vector containing *Bxb1* integrase into the HEK293T-RCA7 landing pad cell line which allowed for the integration of the split GFP exons and mCherry reporter at the AAVS1 locus[54]. Since the integrated constructs enable expression of puromycin resistance gene in the landing pad cell line, the cells were selected with puromycin and maintained in culture and passaged for 20 days to ensure elimination of non-integrated plasmids. Thus the fluorescence signal observed in a single cell should come only from the integrated construct. After 20 days in culture, the cells were fluorescence activated cell sorted (FACS) (Sony MA900 Multi-Application Cell Sorter, Broad Institute Cell Sorting Facility) into GFP positive and GFP negative bins. All collected cells were mCherry positive. To ensure library diversity, at least 1000 cells per oligo in the library were sorted.

Genomic DNA extraction and deep sequencing. gDNA extraction was performed with a commercial kit (Qiagen All-Prep Kit) and the integrated fragments were PCR amplified with

HTSA-NGS-F primer 5'-CTATATATAGCTATCTATGTCTACCGGT-3' and a commercial kit (Qiagen All-Prep Kit) and the integrated fragments were PCR amplified with HTSA-NGS-R primer 5'- GGCTGGAACTCTTGCGCTAG-3' with a high-fidelity polymerase (Primestar GXL Polymerase, Takara) for a maximum of 20 cycles. The amplicons from each fluorescent bin were multiplexed and deep sequenced with a targeted depth of 500 reads per oligo (250x250 pair-end reads, Novaseq 6000, Illumina).

Data analysis and statistics*.* Test exon sequence reads in each fluorescence bin were determined by NGS and matched to the oligo sequences in the corresponding library, counted and normalized by coverage (reads per million in each pool) and the FACS fraction of the fluorescence bin. Sequences with coverage of less than 50 reads across both pools were excluded. Exon inclusion index (EII) was calculated for each sequence, where EII = GFP-ve read count/(GFP negative + GFP positive read counts) (Supplemental Figure 1). The impact of the variant was calculated by the difference of exon inclusion of the variant and the reference sequence: $\Delta EII = EII_{control} - EII_{reference}$. The experiments were performed in duplicate, and an average of ΔEII values were calculated.

Reproducibility of the assay was calculated from two biological replicates using the EII and ΔEII values with EII=0.05 and ΔEII = -0.3 at cutoff points, which allowed us to construct contingency tables and estimate tetrachoric correlation coefficients (Supplementary table 3).

All figures were prepared with Graphpad Prism 8 and Biorender.

RESULTS

Optimization of the high throughput splicing assay (HTSA)

As the starting point of the assay design we used constructs developed by Cheung and colleagues for the multiplexed functional assay of splicing using Sort-seq (MFASS)[54]. In the original MFASS construct a GFP gene is split by 1000bp-long sequence derived from introns 7 and 8 of the *SNM1* transcript (NM_000344), into which test sequences can be cloned in a multiplexed fashion using the *AgeI* and *NheI* restriction sites (Figure 1A). In the MFASS design the split GFP and the control fluorophore mCherry were separated by a self-cleavage 2A peptide, which allowed for the study of in-frame exons only, which was an important limitation of the assay. To increase the number of possible exons to characterize, we have changed the construct by separating expression of the two fluorophores, which are now both under the ubiquitous *EF1*alpha promoter, allowing for testing out-of-frame exons and cryptic exons without compromising mCherry expression (Figure 1A). For every mini-gene transcript there are two outcomes: 1) an inserted test sequence is recognized as an exon, leading to GFP disruption and no green fluorescence; 2) an inserted sequence is not recognized as an exon, leading to full reconstitution of GFP and presence of green fluorescence (Figure 1B). The control mCherry fluorescence is present regardless of the splicing outcome.

Other elements of the HTSA are analogous to MFASS[54]. The multiplexed libraries are integrated into the *AAVS1* safe harbor locus of the RCA7-HEK293T landing pad human cell line by co-transfection with a plasmid carrying the *Bxb1* integrase (Figure 1C)[54,84]. The integrated plasmid carries a puromycin resistance gene, which can only be expressed after integration of the HTSA construct in the *AAVS1* locus. Puromycin selection and multiple passages over 20 days of cell culture ensure that a single copy of the reporter construct is expressed per cell and episomic HTSA plasmids are degraded. This enables fluorescence activated cell sorting (FACS)

separation of GFP negative (-ve) and GFP positive (+ve) cells (Figure 1D). The sequences in the GFP-ve and GFP+ve cell pools are subsequently analyzed by deep next generation sequencing, and the matching sequences counted and normalized by the read depth and by the fraction of cells sorted into the GFP+ve or GFP-ve bins. To determine which variants lead predominantly to exon inclusion or exclusion, the exon inclusion index (EII) is calculated for each sequence using the normalized oligo sequence count, where EII= GFP-ve read count/(GFP-ve + GFP+ve read count) (Figure 1E). EII of 1 signifies complete exon inclusion and EII of 0 indicates complete exon skipping.

Evaluation of HTSA based on 380 IRD gene exons and essential splice site controls

In the initial experiment we tested 380 exons from 89 IRD genes that were chosen based on rare unclassified mutations found in unsolved IRD patient cohort from the MEE clinic (MAF < 0.003) and additional rare variants found in the gnomAD database (MAF < 0.001). The control library consisted of the reference exon sequences and two control sequences for each exon with mutated acceptor site (AS) or donor site (DS) (Supplementary Table 1). Owing to the improvement of the oligo synthesis methodology we were able to investigate exons up to 200 bp in length (median exon length 132 bp, Figure 2A) compared to 100 bp long exons studied previously[54].

Of the 380 exons, seventeen were not sufficiently represented in the library (< 50 reads per sequence) and were dropped from the subsequent analyses. For the remaining 363 exons, we measured reproducibility of the assay across biological replicates using the EII values. The data showed a bimodal distribution and a high level or reproducibility between the replicates (tetrachoric correlation r = 0.97) (Figure 2B). Most of the reference exons showed a high level of exon recognition (median EII = 0.85) with 282 reference sequences recognized as exons at least half of the time (EII ≥ 0.5) (Figure 3C). Sixty-three reference sequences showed

preferential exon skipping (EII<0.2) and 18 showed lower than expected exon recognition (0.2 ≤ EII < 0.5). There was no correlation between the size of a studied exons and their performance in the splicing assay (Fig. 2D).

Introducing mutations to the essential splice sites reduced exon recognition, where DS controls showed stronger effect (median EII=0.06) than AS controls (median EII=0.39) (Figure 2C). To evaluate how essential splice site controls performed in comparison to their corresponding references, we used a ΔEII metric ($\Delta EII = EII_{control} - EII_{reference}$). The ΔEII values clustered in two major groups: around ΔEII = -1 and around ΔEII = 0, where ΔEII = -1 indicates a complete exon skipping due to the mutation in a well performing reference exon, and ΔEII = 0 indicates no effect of variant on exon inclusion (Figure 2 E, F). In a few extreme cases ΔEII showed positive values, which indicated a stronger exon inclusion in sequences with an essential splice site variant than in the corresponding reference sequence (Figure 2 E,F). DS controls showed on average higher effect on exon recognition (median ΔEII = -0.27), compared to AS controls (median ΔEII = -0.09). Based on the bimodal distribution of the ΔEII values, we have determined that ΔEII ≤ -0.3 is a reliable cutoff for variants that show effect on exon inclusion and ΔEII ≤ -0.6 is a reliable cutoff for a strong effect on exon inclusion (Figure 3F).

Evaluation of the effect on splicing of 1131 variants in IRD genes

Next we tested variants of unknown significance in IRD genes. The variant library consisted of 380 exons, 785 essential splice site controls and 1131 test variants (2296 sequences in total) (Supplementary Table 2). Nineteen exons were excluded due to a low coverage of a reference and/or the variant sequence, leaving 1099 variants that met coverage thresholds to be tested. The reproducibility of the assay for the variant library across biological replicates was high (tetrachoric correlation r = 0.91) (Figure 3A). Since the control library and the variant library contained the same reference and essential splice site control sequences, we were able to

compare exon inclusion of both libraries across four experiments, demonstrating high reproducibility (Figure 3B). Pairwise comparison of the average values of EII for the control library and EII for the variant library were highly correlated, demonstrating robustness and reproducibility of the assay (tetrachoric correlation r = 0.98, Supplementary table 3). As in the control library before, the DS controls exerted on average a higher impact on exon recognition (median ΔEII = -0.27) compared to AS controls (median ΔEII = -0.07). The overall effect of tested variants on exon inclusion was smaller than that of the positive controls (median ΔEII = -0.04) (Figure 3C).

Of the 1099 tested variants, 103 changes (9.4%) reduced exon recognition by at least 30% (ΔEII ≤ -0.3) and forty of those variants showed stronger effect than the essential splice site controls (Fig 3D, Supplementary Table 2). Twenty-three variants (2.1%) showed a strong effect on exon inclusion, reducing exon recognition by 60% or more (ΔEII ≤ -0.6) (Table 1, Supplementary Table 2). These included six synonymous variants, eight missense changes and nine intronic variants. Only six of the variants showing a strong exon skipping effect were predicted to have a significant effect on splicing by the Splice AI algorithm (ΔSplice AI score >0.2)[55] (Table 1).

New genetic solutions in IRD patients

When assessing pathogenicity of genetic variants that lead to exon skipping, we have to consider multiple criteria: 1) the frame of the exon that is eliminated; 2) protein domains that are encoded by the in-frame exon; 3) mode of inheritance of a particular trait and the pathogenicity of the allele in *trans* for the recessive genes; 4) gene dosage sensitivity[85]; 5) extent of exon skipping (full or partial). These criteria will allow us to determine if a particular variant leads to a hypomorphic or to a null allele. For example, partial skipping of an out-of-frame exon will lead to a hypomorphic allele, which can be pathogenic in a recessive gene, when coupled with a

pathogenic or likely pathogenic variant in *trans* [86,87]. Skipping of an in-frame exon, will not lead to transcript degradation but may lead to a non-functional protein if an important domain is deleted. Weighing these criteria, we evaluated variants that showed an effect on exon skipping in a cohort of ~3000 IRD cases analyzed by targeted sequencing or exome sequencing[13,38]. We found ten cases for whom the splicing results likely lead to new genetic diagnoses, including six cases with autosomal recessive gene solutions (*ABCA4*[88,89], *EYS*[90,91], *FLVCR1*[92], *USH2A*[93]) three with X-linked disease, (*CHM*[94], *RPGR*[95]) and one dominant haploinsufficient gene (*JAG1*[96,97]) (Table 2).

Three patients carried mutations in *ABCA4*, a gene associated with a number of retinal disorders including an inherited macular degeneration (OMIM #248200), cone-rod (OMIM #604116) and rod-cone dystrophy (OMIM #601718)[88,89]. Subject OGI2271_003379 with a clinical diagnosis of macular disease was heterozygous for c.1928T>G; p.Val643Gly, which led to skipping of an in-frame exon 13 in 2/3rd transcripts (ΔEII = -0.69) coupled with a known pathogenic/likely pathogenic splicing variant c.5714+5G>A[98]. The deleted exon 13 encodes 59 residues, which form a part of the first exocytoplasmic domain (ECD1) of unknown function[99]. Exon 13 carries 61 pathogenic or likely pathogenic missense variants (https://databases.lovd.nl/shared/variants/ABCA4/) suggesting that ECD1 plays an important role in ABCA4 function and therefore deletion of this exon is likely deleterious. No other variants in known IRD genes explaining the IRD disease were uncovered in this patient. Two additional cases with cone-rod dystrophy (OGI3781_0052218) and macular degeneration (OGI2923_004508) had heterozygous variants leading to an out-of-frame exon 40 skipping, which is thought to result in nonsense-mediated decay[46]. Variants c.5693G>A, p.Arg1898His (ΔEII = -0.45) and c.5680C>T, p.Leu1894= variant (ΔEII = -0.33), both lead to a partial exon 40 skipping. In both cases they are coupled with pathogenic/likely pathogenic variants c.4139C>T, p.Pro1380Leu and c.223T>C, p.Cys75Arg respectively (Table 2)[98,100,101] . All three alleles that

showed an effect on splicing can be considered as hypomorphic alleles, which have been shown to lead to *ABCA4*-associated disease[86].

A male subject OGI3782_0052219 with a diagnosis of an X-linked condition of choroideremia carried a hemizygous variant c.940G>A, p.(Gly314Arg) in *CHM*, a gene associated with the subject's condition (OMIM #303100)[94]. The c.940G>A variant leads to skipping of an out-of-frame exon 7 in 2/3rd of the transcripts (ΔEII = -0.63), is absent in gnomAD and segregated with the same phenotype in a maternal male cousin and thus is likely to lead to disease. At the time of the study the variant was classified as VUS, however recently the classification has changed to likely pathogenic, which corroborates our findings[98].

A rod-cone dystrophy case OGI2026_003452 carried two variants in *EYS*, associated with a recessive form of IRD (OMIM # 602772) [90,91]. The subject carried a heterozygous deletion of an out-of-frame exon 31 and a c.1056+3A>C change leading to skipping of an out-of-frame exon 6 in over 2/3rd of the transcripts (ΔEII = -0.71). These two changes are a very likely cause of disease in this subject (Table 2).

Subject OGI1487_002682 carried two heterozygous variants in *FLVCR1,* associated with posterior column ataxia with IRD (OMIM # 609033)[92]. OGI1487_002682 has a clinical diagnosis of with rod-cone degeneration though no clinical information regarding extraocular features was available. The c.806T>C, p.Leu269Pro variant did not affect splicing (ΔEII = -0.12), however the change is extremely rare (absent in gnomAD), it resides in a conserved protein domain (Major facilitator superfamily (MFS) domain[102,103]) and is predicted to be protein damaging by two algorithms (Polyphen-2[61] and FATHMM[104]). The second variant is a synonymous c.1026T>C, p.Gly342= change, that showed skipping of an out-of-frame exon 4 in the majority of transcripts (ΔEII = -0.75). Together, these two alleles are very likely causal in the subject (Table 2).

An early onset rod-cone dystrophy case OGI658_001334 was heterozygous for a c.2237G>A, (p.Ser746Asn) variant in *JAG1*, a gene associated with multiple autosomal dominant diseases including Alagille syndrome (OMIM 118450 #)[96] and Charcot-Marie-Tooth disease (OMIM # 619574)[97]. The c.2237G>A, p.Ser746Asn variant lead to exon 18 skipping in a third of transcripts (ΔEII = -0.35). Exon 18 is in-frame and encodes 13th EGF domain[102,103]. It is unclear if this change is pathogenic but given that *JAG1* is missense and loss of function intolerant and it is sensitive to dosage[105,106], this change may be the cause of disease in the studied case and warrants further functional evaluation.

Two simplex male patients with rod-cone degeneration each were found to carry hemizygous variants in *RPGR,* a gene associated with an X-linked IRD (OMIM # 312610)[95]. Both variants affected splicing of an in-frame exon 4, leading to exon skipping in up to 2/3rd of transcripts: c.310G>A; p.Glu104Lys (ΔEII = -0.65) and c.310+7T>G (ΔEII = -0.62). Exon 4 codes for 21 residues that are involved in forming of the second regulator of chromosome condensation (RCC1) domain[102,103] (Table 2).

A non-syndromic rod-cone dystrophy case OGI2102_003536 was heterozygous for a splice region variant c.5298+3A>G and a frameshift variant c.10318del (p.Ile3440LeufsTer17) in *USH2A*, a gene associated with a recessive form of retinal degeneration and hearing loss (Usher syndrome type II, OMIM # 276901) and non-syndromic IRD (OMIM # 613809)[93]. The splice affecting variant leads to skipping of an out-of-frame exon 26 in about half of the transcripts (ΔEII = -0.45). Given that the second mutation is a null allele, the patient is predicted to have about 1/4th of the normal dose of the USH2A protein, which is likely the explanation of his retinal phenotype but no defect in hearing, as hypomorphic mutations in *USH2A* have been reported in nonsyndromic IRD cases[107] (Table 2).

DISCUSSION

In this study we have developed an improved version of a high-throughput splicing assay (HTSA), which was based on a previously published method[54]. We used HTSA to study 1131 variants of unknown significance in 380 exons from 89 IRD genes. We observed that 103 (9.4%) variants lead to decreased exon inclusion. Of these, 23 variants showed a strong effect, with at least 60% of transcripts leading to exon skipping compared to the reference sequences. Using the splicing assay data we were able to provide a likely genetic diagnosis to ten IRD cases.

HTSA showed high reproducibility when compared between different biological replicates. Most of the reference exons showed a high level of exon recognition (median EII = 0.85), however in 81 reference sequences the exon was recognized less than 50% of the times and 63 exons were preferentially skipped (EII<0.2). One of the reasons for the poor performance of these exons could have been the lack of the crucial intronic context, necessary for the exon recognition. However we have not observed a correlation between the size of a studied exon (the smaller the exon the larger the flanking exonic sequences in the 270bp oligos) and its performance in the splicing assay.

We have detected 23 variants with strong effects on splicing, however not all of these variants led to genetic solutions in our patient cohort, which can be due to a number of reasons. In some autosomal recessive cases the second variant in the same gene was determined not to be pathogenic, a phenotype didn't match or a another more likely cause of disease was found. Nevertheless in five recessive cases, our HTSA was able to provide additional functional evidence to reconsider variants of unknown significance that were coupled with likely pathogenic variants in the same gene. In three X-linked male cases, exon skipping led to the deletion of a crucial domain (RCC1 in *RPGR*) or creation of a frameshift mutation (*CHM*), which

were likely causal variants. Variants in the remaining two cases are not as straightforward and will require further investigation and additional IRD cases with the same mutations will strengthen their likelihood of pathogenicity. In subject OGI1487_002682 with two VUSs in *FLVCR1* (c.806T>C, p.Leu269Pro and c.1026T>C, p.Gly342=), we were able to find that only the c.1026T>C led to skipping of an out-of-frame exon 4 in the majority of transcripts (ΔEII = -0.75), whereas the c.806T>C, p.Leu269Pro change had no effect on splicing. Despite the fact that c.806T>C, p.Leu269Pro is absent in gnomAD, it resides in a conserved protein domain[102,103] and is predicted to be protein damaging by two algorithms[61,104], it remains classified as VUS in ClinVar. One subject carried a hypomorphic c.2237G>A, p.Ser746Asn allele in *JAG1* leading to an in-frame exon 18 skipping in 1/3rd of the transcripts. Given that JAG1 is mutation intolerant and dosage sensitive[83,85,105], we judged it important to report our findings of this allele.

In HTSA we measured the disruption of GFP, which we labeled as exon recognition. However, disruption of GFP may also occur when part of an exon is included, or an intron retained, which can only be identified by the analysis of HTSA minigene transcript from the GFP-ve pool of cells. Therefore it is possible that variants that seemingly showed no effect on exon inclusion, actually led to aberrant splicing by one of the above mechanisms. Further studies will be undertaken to test this hypothesis. Of the 23 variants that showed a strong effect on exon inclusion, only six were predicted to have a significant effect on splicing by the Splice AI algorithm (ΔSplice AI score >0.2)[55]. This illustrates the need for improved splicing prediction programs especially in regions not typically thought of as splice altering. Amercian College of Medical Genetics (ACMG) guidelines often require functional validation to assess the pathogenicity of unclassified variants[65]. However, the cost and effort to individually interrogate VUSs is often prohibitive which leaves many variants that may provide the genetic solution as unclassified. We present an elegant and robust method for interrogating hundreds to thousands

of variants and their impact on exon recognition leading to splicing dysfunction. In the current study HTSA was employed to investigates the effects of variants in or near exons, however it can also be used to study deep intronic variants leading to cryptic exon inclusion[48–52,77].

ACKNOWLEDGMENTS

This work was supported by grants from the National Eye Institute R01EY026904 (KMB/EAP), [R01EY012910 (EAP) and P30EY014104 (MEEI core support)], the Foundation Fighting Blindness [EGI-GE-1218-0753-UCSD, (KMB/EAP)] and the Research to Prevent Blindness International Research Collaborators Award (KMB). Exome sequencing and analysis were provided by the Broad Institute of MIT and Harvard Center for Mendelian Genomics (Broad CMG) and was funded by the National Human Genome Research Institute, the National Eye Institute, and the National Heart, Lung and Blood Institute grant UM1HG008900 and in part by National Human Genome Research Institute grant R01 HG009141. The authors thank all subjects for their participation in this study and the OGI Genomics Core members for their experimental assistance.

References

1. Jespersgaard C, Fang M, Bertelsen M, et al. Molecular genetic analysis using targeted NGS analysis of 677 individuals with retinal dystrophy. *Sci Rep*. 2019;9(1219):1-7. doi:10.1038/s41598-018-38007-2
2. Chakravorty S, Hegde M. Gene and Variant Annotation for Mendelian Disorders in the Era of Advanced Sequencing Technologies. *Annu Rev Genomics Hum Genet*. 2017;18(1):229-256. doi:10.1146/annurev-genom-083115-022545
3. ClinVar [http://www.ncbi.nlm.nih.gov/clinvar/].
4. Petersen BS, Fredrich B, Hoeppner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet*. 2017;18(1):1-13. doi:10.1186/s12863-017-0479-5
5. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res*. 2010;29(5):335-375. doi:10.1016/j.preteyeres.2010.03.004
6. Daiger SP, Sullivan LS, Sara J. Bowne. Retinal Information Network (RetNet). https://sph.uth.edu/retnet/home.htm. Published 2020. Accessed October 15, 2021.
7. Consugar MB, Navarro-Gomez D, Place EM, et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genet Med*. 2015;17(4):253-261. doi:10.1038/gim.2014.172
8. Bujakowska KM, Consugar MB, Place E, et al. Targeted exon sequencing in Usher syndrome type I. *Invest Ophthalmol Vis Sci*. 2014;55(12):8488-8496. doi:10.1167/iovs.14-15169
9. Weisschuh N, Obermaier CD, Battke F, et al. Genetic architecture of inherited retinal degeneration in Germany: A large cohort study from a single diagnostic center over a 9-year period. *Hum Mutat*. 2020;(January):1-14. doi:10.1002/humu.24064
10. Perez-Carro R, Corton M, Sánchez-Navarro I, et al. Panel-based NGS Reveals Novel Pathogenic Mutations in Autosomal Recessive Retinitis Pigmentosa. *Sci Rep*. 2016;6(December 2015):1-10. doi:10.1038/srep19531
11. Bernardis I, Chiesi L, Tenedini E, et al. Unravelling the Complexity of Inherited Retinal Dystrophies Molecular Testing : Added Value of Targeted Next-Generation Sequencing. *Biomed Res Int*. 2016;2016:6341870. doi:10.1155/2016/6341870
12. Costa KA, Salles MV, Whitebirch C, Chiang J, Sallum JMF. Gene panel sequencing in Brazilian patients with retinitis pigmentosa. *Int J Retin Vitr*. 2017;3(1):33. doi:10.1186/s40942-017-0087-6
13. Zampaglione E, Maher M, Place EM, et al. The Importance of Automation in Genetic Diagnosis: Lessons from Analyzing an Inherited Retinal Degeneration Cohort with the Mendelian Analysis Toolkit (MATK). *medRxiv (Accepted to GiM)*. January 2021:2021.04.09.21255188. doi:10.1101/2021.04.09.21255188
14. Audo I, Bujakowska K, Orhan E, et al. The familial dementia gene revisited: A missense mutation revealed by whole-exome sequencing identifies ITM2B as a candidate gene underlying a novel autosomal dominant retinal dystrophy in a large family. *Hum Mol Genet*. 2014;23(2). doi:10.1093/hmg/ddt439
15. El Shamieh S, Neuillé M, Terray A, et al. Whole-exome sequencing identifies KIZ as a ciliary gene associated with autosomal-recessive rod-cone dystrophy. *Am J Hum Genet*. 2014;94(4):625-633. doi:10.1016/j.ajhg.2014.03.005
16. Arno G, Holder GE, Chakarova C, et al. Recessive Retinopathy Consequent on Mutant G-Protein β Subunit 3 (GNB3). *JAMA Ophthalmol*. 2016;134(8):924-927. doi:10.1001/jamaophthalmol.2016.1543
17. Biswas P, Ramana V, Chavali M, et al. A missense mutation in ASRGL1 is involved in causing autosomal recessive retinal degeneration. *Hum Mol Genet*. 2016;25(12):2483-

2497. doi:10.1093/hmg/ddw113
18. Coppieters F, Ascari G, Dannhausen K, et al. Isolated and Syndromic Retinal Dystrophy Caused by Biallelic Mutations in RCBTB1, a Gene Implicated in Ubiquitination. *Am J Hum Genet*. 2016;99(2):470-480. doi:10.1016/j.ajhg.2016.06.017
19. Wu JH, Liu JH, Ko YC, et al. Haploinsufficiency of RCBTB1 is associated with Coats disease and familial exudative vitreoretinopathy. *Hum Mol Genet*. 2016;25(8):1637-1647. doi:10.1093/hmg/ddw041
20. Xu M, Eblimit A, Wang J, et al. ADIPOR1 Is Mutated in Syndromic Retinitis Pigmentosa. *Hum Mutat*. 2016;37(3):246-249. doi:10.1002/humu.22940
21. Zhang J, Wang C, Shen Y, et al. A mutation in ADIPOR1 causes nonsyndromic autosomal dominant retinitis pigmentosa. *Hum Genet*. 2016;135(12):1375-1387. doi:10.1007/s00439-016-1730-2
22. Saksens NTM, Krebs MP, Schoenmaker-Koller FE, et al. Mutations in CTNNA1 cause butterfly-shaped pigment dystrophy and perturbed retinal pigment epithelium integrity. *Nat Genet*. 2016;48(2):144-151. doi:10.1038/ng.3474
23. Vincent A, Audo I, Tavares E, et al. Biallelic Mutations in GNB3 Cause a Unique Form of Autosomal-Recessive Congenital Stationary Night Blindness. *Am J Hum Genet*. 2016;98(5):1011-1019. doi:10.1016/j.ajhg.2016.03.021
24. Branham K, Matsui H, Biswas P, et al. Establishing the involvement of the novel gene *AGBL5* in retinitis pigmentosa by whole genome sequencing. *Physiol Genomics*. 2016;48(12):922-927. doi:10.1152/physiolgenomics.00101.2016
25. Xu M, Xie Y (Angela), Abouzeid H, et al. Mutations in the Spliceosome Component CWC27 Cause Retinal Degeneration with or without Additional Developmental Anomalies. *Am J Hum Genet*. 2017;100(4):592-604. doi:10.1016/j.ajhg.2017.02.008
26. Neuille M, El Shamieh S, Orhan E, et al. Lrit3 deficient mouse (nob6): A novel model of complete congenital stationary night blindness (cCSNB). *PLoS One*. 2014;9(3):e90342. doi:10.1371/journal.pone.0090342
27. Agrawal SA, Burgoyne T, Eblimit A, et al. REEP6 deficiency leads to retinal degeneration through disruption of ER homeostasis and protein trafficking. *Hum Mol Genet*. 2017;26(14):2667-2677. doi:10.1093/hmg/ddx149
28. Ma X, Guan L, Wu W, et al. Whole-exome sequencing identifies OR2W3 mutation as a cause of autosomal dominant retinitis pigmentosa. *Sci Rep*. 2015;5:9236. doi:10.1038/srep09236
29. El-Asrag ME, Sergouniotis PI, McKibbin M, et al. Biallelic mutations in the autophagy regulator DRAM2 cause retinal dystrophy with early macular involvement. *Am J Hum Genet*. 2015;96(6):948-954. doi:10.1016/j.ajhg.2015.04.006
30. Goyal S, Jäger M, Robinson PN, Vanita V. Confirmation of TTC8 as a disease gene for nonsyndromic autosomal recessive retinitis pigmentosa (RP51). *Clin Genet*. July 2015. doi:10.1111/cge.12644
31. Avila-Fernandez A, Perez-Carro R, Corton M, et al. Whole-exome sequencing reveals ZNF408 as a new gene associated with autosomal recessive retinitis pigmentosa with vitreal alterations. *Hum Mol Genet*. 2015;24(14):4037-4048. doi:10.1093/hmg/ddv140
32. Segarra NG, Ballhausen D, Crawford H, et al. NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver, immune system, and retina. *Am J Med Genet Part A*. 2015;167(12):2902-2912. doi:10.1002/ajmg.a.37338
33. Bujakowska KM, Zhang Q, Siemiatkowska AM, et al. Mutations in IFT172 cause isolated retinal degeneration and Bardet-Biedl syndrome. *Hum Mol Genet*. 2015;24(1):230-242. doi:10.1093/hmg/ddu441
34. Arno G, Agrawal SA, Eblimit A, et al. Mutations in REEP6 Cause Autosomal-Recessive Retinitis Pigmentosa. *Am J Hum Genet*. 2016;99:1-11. doi:10.1016/j.ajhg.2016.10.008
35. Pieras JI, Barragán I, Borrego S, et al. Copy-number variations in EYS: A significant

event in the appearance of arRP. *Investig Ophthalmol Vis Sci*. 2011;52(8):5625-5631. doi:10.1167/iovs.11-7292
36. Bowne SJ, Sullivan LS, Wheaton DK, et al. North Carolina macular dystrophy (MCDR1) caused by a novel tandem duplication of the PRDM13 gene. *Mol Vis*. 2016;22:1239-1247. http://www.ncbi.nlm.nih.gov/pubmed/27777503.
37. de Bruijn SE, Fiorentino A, Ottaviani D, et al. Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. *Am J Hum Genet*. 2020;107(5):802-814. doi:10.1016/j.ajhg.2020.09.002
38. Zampaglione E, Kinde B, Place EM, et al. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. *Genet Med*. 2020;22(6). doi:10.1101/742106
39. Van Cauwenbergh C, Van Schil K, Cannoodt R, et al. ArrEYE: A customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. *Genet Med*. 2017;19(4):457-466. doi:10.1038/gim.2016.119
40. Ellingford JM, Horn B, Campbell C, et al. Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *J Med Genet*. 2018;55:114-121. doi:10.1136/jmedgenet-2017-104791
41. Eisenberger T, Neuhaus C, Khan AO, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PLoS One*. 2013;8(11):e78496. doi:10.1371/journal.pone.0078496
42. Neuhaus C, Eisenberger T, Decker C, et al. Next-generation sequencing reveals the mutational landscape of clinically diagnosed Usher syndrome: copy number variations, phenocopies, a predominant target for translational read-through, and PEX26 mutated in Heimler syndrome. *Mol Genet Genomic Med*. 2017;5(5):531-552. doi:10.1002/mgg3.312
43. Austin-Tse CA, Mandelker DL, Oza AM, Mason-Suares H, Rehm HL, Amr SS. Analysis of intragenic USH2A copy number variation unveils broad spectrum of unique and recurrent variants. *Eur J Med Genet*. 2018;61(10):621-626. doi:10.1016/j.ejmg.2018.04.006
44. Khateb S, Hanany M, Khalaileh A, et al. Identification of genomic deletions causing inherited retinal degenerations by coverage analysis of whole exome sequencing data. *J Med Genet*. 2016;53(9):600-607. doi:10.1136/jmedgenet-2016-103825
45. Small KW, DeLuca AP, Whitmore SS, et al. North Carolina Macular Dystrophy Is Caused by Dysregulation of the Retinal Transcription Factor PRDM13. *Ophthalmology*. 2016;123(1):9-18. doi:10.1016/j.ophtha.2015.10.006
46. Chang Y-FF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem*. 2007;76:51-74. doi:10.1146/annurev.biochem.76.050106.093909
47. Jamshidi F, Place EM, Mehrotra S, et al. Contribution of non-coding mutations to RPGRIP1-mediated inherited retinal degeneration. *Genet Med*. 2019;3(March 21):694-704. doi:10.1038/s41436-018-0104-7
48. den Hollander AI, Koenekoop RK, Yzer S, et al. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet*. 2006;79(3):556-561. doi:10.1086/507318
49. Rio Frio T, McGee TL, Wade NM, et al. A single-base substitution within an intronic repetitive element causes dominant retinitis pigmentosa with reduced penetrance. *Hum Mutat*. 2009;30(9):1340-1347. doi:10.1002/humu.21071
50. Vaché C, Besnard T, le Berre P, et al. Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy. *Hum Mutat*. 2012;33(1):104-108. doi:10.1002/humu.21634
51. Braun TA, Mullins RF, Wagner AH, et al. Non-exomic and synonymous variants in

ABCA4 are an important cause of Stargardt disease. *Hum Mol Genet*. 2013;22(25):5136-5145. doi:10.1093/hmg/ddt367
52. Webb TR, Parfitt DA, Gardner JC, et al. Deep intronic mutation in OFD1, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23). *Hum Mol Genet*. 2012;21(16):3647-3654. doi:10.1093/hmg/dds194
53. Bitner-Glindzicz M, Lindley KJ, Rutland P, et al. A recessive contiguous gene deletion causing infantile hyperinsulinism, enteropathy and deafness identifies the Usher type 1C gene. *Nat Genet*. 2000;26:56-60. doi:10.1038/79178
54. Cheung R, Insigne KD, Yao D, et al. A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. *Mol Cell*. 2019:1-12. doi:10.1016/j.molcel.2018.10.037
55. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019;176(3):535-548.e24. doi:10.1016/j.cell.2018.12.015
56. Riepe T V., Khan M, Roosing S, Cremers FPM, 't Hoen PAC. Benchmarking deep learning splice prediction tools using functional splice assays. *Hum Mutat*. 2021;42(7):799-810. doi:10.1002/humu.24212
57. Coppieters F, Todeschini AL, Fujimaki T, et al. Hidden Genetic Variation in LCA9-Associated Congenital Blindness Explained by 5'UTR Mutations and Copy Number Variations Of NMNAT1. *Hum Mutat*. 2015;12:1188-1196. doi:10.1002/humu.22899
58. Jin Y, Chen Z, Liu X, Zhou X. Evaluating the MicroRNA targeting sites by luciferase reporter gene assay. *Methods Mol Biol*. 2013;936(4):117-127. doi:10.1007/978-1-62703-083-0
59. Conte I, Hadfield KD, Barbato S, et al. MiR-204 is responsible for inherited retinal dystrophy associated with ocular coloboma. *Proc Natl Acad Sci U S A*. 2015;112(25):E3236-E3245. doi:10.1073/pnas.1401464112
60. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12824425.
61. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013. doi:10.1002/0471142905.hg0720s76
62. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019. doi:10.1093/nar/gky1016
63. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet*. 2016;99(4):877-885. doi:10.1016/j.ajhg.2016.08.016
64. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315. doi:10.1038/ng.2892
65. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-423. doi:10.1038/gim.2015.30
66. Wan A, Place E, Pierce EA, Comander J. Characterizing variants of unknown significance in rhodopsin: A functional genomics approach. *Hum Mutat*. 2019;40(8):1127-1144. doi:10.1002/humu.23762
67. Li G, Martínez-Bonet M, Wu D, et al. High-throughput identification of noncoding functional SNPs via type IIS enzyme restriction. *Nat Genet*. 2018. doi:10.1038/s41588-

018-0159-z
68. Maguire AM, Simonelli F, Pierce EA, et al. Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N Engl J Med*. 2008;358:2240-2248. doi:10.1056/NEJMoa0802315
69. Barnard AR, Groppe M, MacLaren RE. Gene Therapy for Choroideremia Using an Adeno-Associated Viral (AAV) Vector. *Cold Spring Harb Perspect Med*. 2014;Oct 30;5(3). doi:10.1101/cshperspect.a017293
70. Hashimoto T, Gibbs D, Lillo C, et al. Lentiviral gene replacement therapy of retinas in a mouse model for Usher syndrome type 1B. *Gene Ther*. 2007;14(7):584-594. doi:10.1038/sj.gt.3302897
71. Han Z, Conley SM, Naash MI. Gene therapy for Stargardt disease associated with ABCA4 gene. *Adv Exp Med Biol*. 2014;801:719-724. doi:10.1007/978-1-4614-3209-8_90
72. Jacobson SG, Cideciyan A V., Ratnakaram R, et al. Gene Therapy for Leber Congenital Amaurosis Caused by RPE65 Mutations: Safety and Efficacy in 15 Children and Adults Followed Up to 3 Years. *Arch Ophthalmol*. 2012;130:9-24. doi:10.1001/archophthalmol.2011.298
73. Ali RR, Sarra GM, Stephens C, et al. Restoration of photoreceptor ultrastructure and function in retinal degeneration slow mice by gene therapy. *Nat Genet*. 2000;25(3):306-310. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10888879.
74. Russell S, Bennett J, Wellman JA, et al. 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*. 2017;390(10097):849-860. doi:10.1016/S0140-6736(17)31868-8
75. Weil D, El-Amraoui A, Masmoudi S, et al. Usher syndrome type I G (USH1G) is caused by mutations in the gene encoding SANS, a protein that associates with the USH1C protein, harmonin. *Hum Mol Genet*. 2003;12(5):463-471. doi:10.1093/hmg/ddg051
76. Jamshidi F, Place EM, Mehrotra S, et al. Contribution of noncoding pathogenic variants to RPGRIP1-mediated inherited retinal degeneration. *Genet Med*. 2019;21(3). doi:10.1038/s41436-018-0104-7
77. Khan M, Cornelis SS, Pozo-Valero M Del, et al. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet Med*. 2020;22(7):1235-1246. doi:10.1038/s41436-020-0787-4
78. Sangermano R, Khan M, Cornelis SS, et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res*. 2018;28(1):100-110. doi:10.1101/gr.226621.117
79. Zampaglione E, Maher M, Place EM, et al. The importance of automation in genetic diagnosis: Lessons from analyzing an inherited retinal degeneration cohort with the Mendelian Analysis Toolkit (MATK). *Genet Med*. 2022;24(2):332-343. doi:10.1016/j.gim.2021.09.015
80. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;9(9):1297-1303. doi:10.1101/gr.107524.110.20
81. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1). doi:10.1186/s13059-016-0974-4
82. Pedersen BS, Layer RM, Quinlan AR. Vcfanno: Fast, flexible annotation of genetic variants. *Genome Biol*. 2016;17(1):1-9. doi:10.1186/s13059-016-0973-5
83. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7
84. Duportet X, Wroblewska L, Guye P, et al. A platform for rapid prototyping of synthetic

gene networks in mammalian cells. *Nucleic Acids Res*. 2014;42(21):13440-13451. doi:10.1093/nar/gku1082
85. Collins RL, Glessner JT, Porcu E, et al. A cross-disorder dosage sensitivity map of the human genome. *Cell*. 2022;185(16):3041-3055.e25. doi:10.1016/j.cell.2022.06.036
86. Zernant J, Lee W, Collison FT, et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. *J Med Genet*. 2017;54:404-412. doi:10.1136/jmedgenet-2017-104540
87. Bauwens M, Garanto A, Sangermano R, et al. ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet Med*. 2019;21(8):1761-1771. doi:10.1038/s41436-018-0420-y
88. Allikmets R, Singh N, Sun H, et al. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet*. 1997;15(3):236-246. doi:10.1038/ng0397-236
89. Cremers FP, van de Pol DJ, van Driel M, et al. Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum Mol Genet*. 1998;7(3):355-362. http://www.ncbi.nlm.nih.gov/pubmed/9466990. Accessed January 17, 2016.
90. Collin RWJ, Littink KW, Klevering BJ, et al. Identification of a 2 Mb human ortholog of Drosophila eyes shut/spacemaker that is mutated in patients with retinitis pigmentosa. *Am J Hum Genet*. 2008;83(5):594-603. doi:10.1016/j.ajhg.2008.10.014
91. Abd El-Aziz M, Barragan I, O'Driscoll C, et al. EYS, encoding an ortholog of Drosophila spacemaker, is mutated in autosomal recessive retinitis pigmentosa. *Nat Genet*. 2008;40(11):1285-1287. doi:10.1038/NG.241
92. Rajadhyaksha AM, Elemento O, Puffenberger EG, et al. Mutations in FLVCR1 cause posterior column ataxia and retinitis pigmentosa. *Am J Hum Genet*. 2010;87(5):643-654. doi:10.1016/j.ajhg.2010.10.013
93. Rivolta C, Sweklo EA, Berson EL, Dryja TP. Missense mutation in the USH2A gene: association with recessive retinitis pigmentosa without hearing loss. *Am J Hum Genet*. 2000;66:1975-1978. doi:10.1086/302926
94. Cremers FPM, Van De Pol DJR, Van Kerkhoff LPM, Wieringa B, Ropers HH. Cloning of a gene that is rearranged in patients with choroideraemia. *Nature*. 1990;347(6294):674-677. doi:10.1038/347674A0
95. Meindl A, Dry K, Herrmann K, et al. A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nat Genet*. 1996;13(1):35-42. doi:10.1038/ng0596-35
96. Oda T, Elkahloun AG, Pike BL, et al. Mutations in the human Jagged1 gene are responsible for Alagille syndrome. *Nat Genet*. 1997;16(July):235-242.
97. Sullivan JM, Motley WW, Johnson JO, et al. Dominant mutations of the Notch ligand Jagged1 cause peripheral neuropathy. *J Clin Invest*. 2020;130(3):1506-1512. doi:10.1172/JCI128152
98. Landrum MJ, Lee JM, Benson M, et al. ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067. doi:10.1093/nar/gkx1153
99. Liu F, Lee J, Chen J. Molecular structures of the eukaryotic retinal importer abca4. *Elife*. 2021;10:1-30. doi:10.7554/eLife.63524
100. Fingert JH, Eliason DA, Phillips NC, Lotery AJ, Sheffield VC, Stone EM. Case of Stargardt Disease Caused by Uniparental Isodisomy. *Arch Ophthalmol*. 2006;124(5):744-745. doi:10.1001/ARCHOPHT.124.5.744
101. Lewis RA, Shroyer NF, Singh N, et al. Genotype/phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. *Am J Hum*

Genet. 1999;64(2):422-434. doi:10.1086/302251
102. PROSITE: https://prosite.expasy.org/. https://prosite.expasy.org/. Published 2015.
103. Sigrist CJA, De Castro E, Cerutti L, et al. New and continuing developments at PROSITE. *Nucleic Acids Res*. 2013;41(D1):344-347. doi:10.1093/nar/gks1067
104. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: Accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*. 2018;34(3):511-513. doi:10.1093/bioinformatics/btx536
105. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet*. 2014;advance on(9):944-950. doi:10.1038/ng.3050
106. Collins RL, Brand H, Karczewski KJ, et al. A structural variation reference for medical and population genetics. *Nature*. 2020. doi:10.1038/s41586-020-2287-8
107. Hufnagel RB, Liang W, Duncan JL, et al. Tissue-specific genotype-phenotype correlations among USH2A-related disorders in the RUSH2A study. *Hum Mutat*. 2022;43(5):613-624. doi:10.1002/HUMU.24365

TABLES

Table 1. Variants showing a strong effect on exon inclusion

Gene name	Transcript ID	Exon	cDNA change	Protein change	gnomAD	Average ΔEII	SpliceAI	Exon frame
ABCA4	ENST00000370225.3	13	c.1928T>G	p.(Val643Gly)	0.001684	-0.69	0.00	In-frame
ABCA4	ENST00000370225.3	47	c.6438C>T	p.(=)	0.000039	-0.61	0.03	In-frame
CEP164	ENST00000278935.3	10	c.1233+5G>A	intronic variant	0.000007	-0.87	**0.95**	In-frame
CHM	ENST00000357749.2	7	c.940G>A	p.(Gly314Arg)	Absent	-0.63	**0.84**	Out-of-frame
CPLANE 1	ENST00000651892.2	27	c.5520T>G	p.(=)	Absent	-0.62	0.06	Out-of-frame
CRX	ENST00000221996.7	3	c.158A>G	p.(Glu53Gly)	Absent	-0.80	0.00	Out-of-frame
EYS	ENST00000370621.7	6	c.1056+3A>C	intronic variant	Absent	-0.71	**0.94**	Out-of-frame
EYS	ENST00000370621.7	6	c.863-15T>C	intronic variant	0.000004	-0.64	0.00	Out-of-frame
EYS	ENST00000370621.7	6	c.863-5T>G	intronic variant	0.000029	-0.68	0.12	Out-of-frame
EYS	ENST00000370621.7	6	c.969A>G	p.(=)	0.000007	-0.61	0.00	Out-of-frame
EYS	ENST00000370621.7	7	c.1184+39A>G	intronic variant	0.0000082	-0.60	0.05	Out-of-frame
EYS	ENST00000370621.7	15	c.2260-63A>G	intronic variant	0.00005	-0.64	0.00	Out-of-frame
EYS	ENST00000370621.7	28	c.5836-11T>G	intronic variant	0.0000069	-0.61	**0.35**	Out-of-frame
EYS	ENST00000370621.7	33	c.6572-7C>T	intronic variant	0.0000059	-0.66	0.00	Out-of-frame
EYS	ENST00000370621.7	33	c.6624A>C	p.(=)	0.000007	-0.63	0.00	Out-of-frame
EYS	ENST00000370621.7	37	c.7392T>C	p.(=)	Absent	-0.70	0.00	In-frame
FLVCR1	ENST00000366971.4	4	c.1026T>C	p.(=)	Absent	-0.75	0.01	Out-of-frame
JAG1	ENST00000254958.5	14	c.1826C>T	p.(Ser609Leu)	0.000079	-0.69	0.00	In-frame
NPHP3	ENST00000337331.5	5	c.944A>T	p.(Asp315Val)	0.000171	-0.61	0.04	Out-of-frame
OFD1	ENST00000340096.6	19	c.2524G>A	p.(Gly842Arg)	0.000286*	-0.63	0.02	In-frame
RPGR	ENST00000645032.1	4	c.310+7T>G	intronic variant	Absent	-0.65	**0.37**	In-frame
RPGR	ENST00000645032.1	4	c.310G>A	p.(Glu104Lys)	Absent	-0.62	**0.94**	In-frame
USH2A	ENST00000307340.3	30	c.5945G>A	p.(Gly1982Asp)	0.000013	-0.78	0.00	In-frame

Notes: *13 hemizygotes are present in gnomAD, indicating that partial defect in exon inclusion is not pathogenic or leads to a mild later onset disease, overlooked in the gnomAD cohort.

Table 2. New likely genetic solutions

Patient number	Gene	gDNA Hg38	HGVSc	HGVSp	gnomAD v3	CADD	ClinVar	Average ΔEII
OGI2271_003379	*ABCA4*	chr1:94062586A>C	c.1928T>G	p.Val643Gly	0.00168	24.5	Conflicting interpretations	-0.69
		chr1:94010795C>T	c.5714+5G>A		0.00034	19.12	P/LP	N/A
OGI3781_0052218	*ABCA4*	chr1:94010821C>T	c.5693G>A	p.Arg1898His	0.00147	19.52	Conflicting interpretations	-0.45
		chr1:94031110G>A	c.4139C>T	p.Pro1380Leu	0.0002431	25.3	P/LP	N/A
OGI2923_004508	*ABCA4*	chr1:94111517A>G	c.223T>C	p.Cys75Arg	absent	29.8	LP	N/A
		chr1:94010834G>A	c.5680C>T	p.(=)	absent	5.195	B*	-0.34
OGI3782_0052219	*CHM*	chrX:85957855C>T	c.940G>A	p.(Gly314Arg)	Absent	33	VUS/LP**	-0.63
OGI2026_003452	*EYS*	chr6:64230343-64231073:1	Exon 31 heterozygous deletion		Absent (gnomADv2-SV)	N/A		N/A
		chr6:65405171T>G	c.1056+3A>C		absent	22.7		-0.71
OGI1487_002682	*FLVCR1*	chr1:212863792T>C	c.806T>C	p.Leu269Pro	absent	20.6	VUS	-0.12***
		chr1:212883372T>C	c.1026T>C	p.(=)	absent	15.92	VUS	-0.75
OGI658_001334	*JAG1*	chr20:10644970C>T	c.2237G>A	p.Ser746Asn	absent	23.4	VUS	-0.35
OGI2075_003501 (male)	*RPGR*	chrX:38321027C>T	c.310G>A	p.Glu104Lys	absent	34	VUS	-0.65
OGI1347_002520 (male)	*RPGR*	chrX:38321020A>C	c.310+7T>G		absent	10.65	VUS	-0.62
OGI2102_003536	*USH2A*	chr1:216083453T>C	c.5298+3A>G		absent	22	VUS	-0.45
		chr1:215786738AT>A	c.10318del	p.Ile3440LeufsTer17	absent	N/A	P/LP	N/A

*Benign by one submitter; ** classification if this variant has recently been changed from VUS to LP in CLinVar. ***this variant is not thought to act via exon skipping but it is predicted to be likely damaging by PolyPhen, fathmm and has an elevated CADD and Revel (0.5) scores

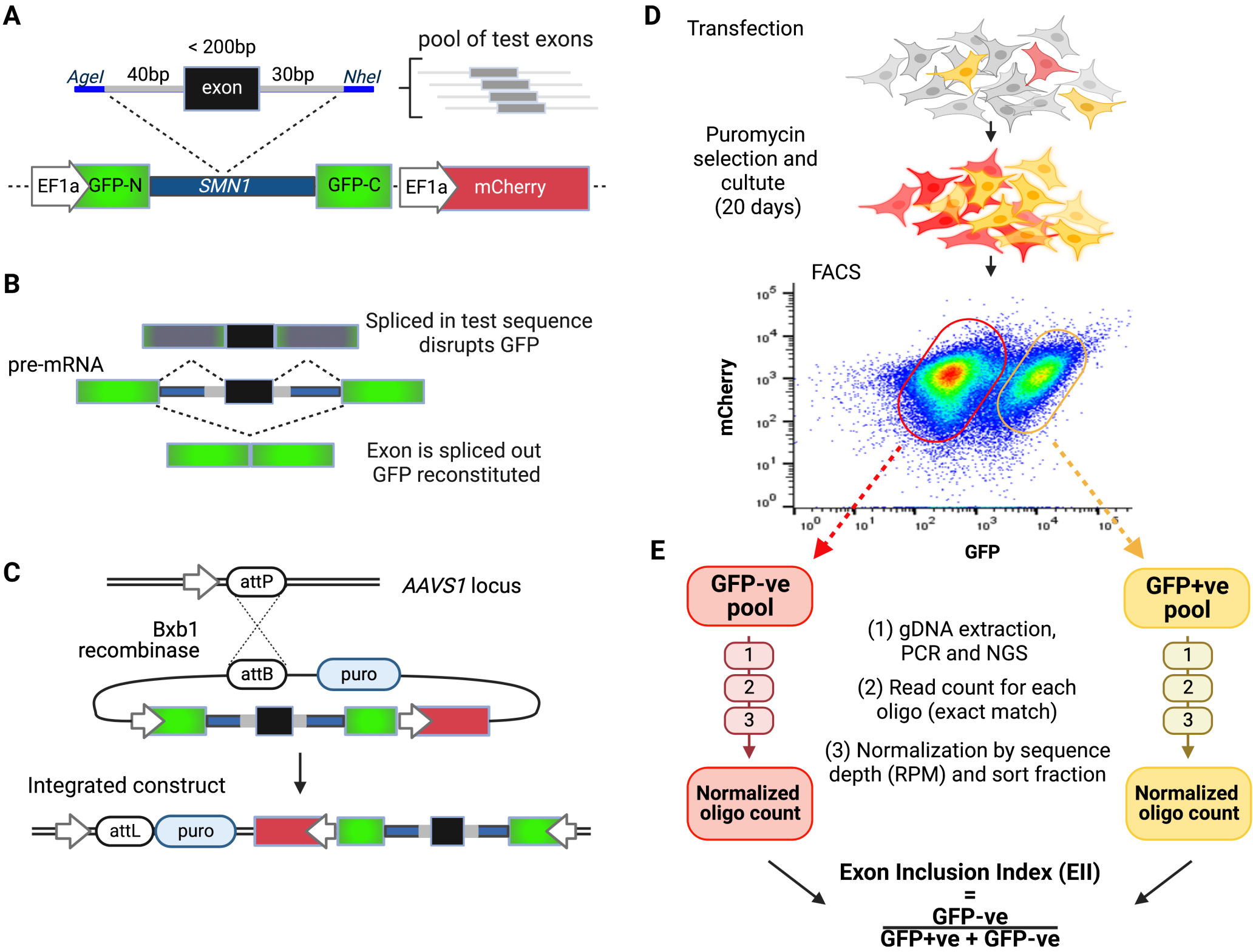
FIGURE LEGENDS

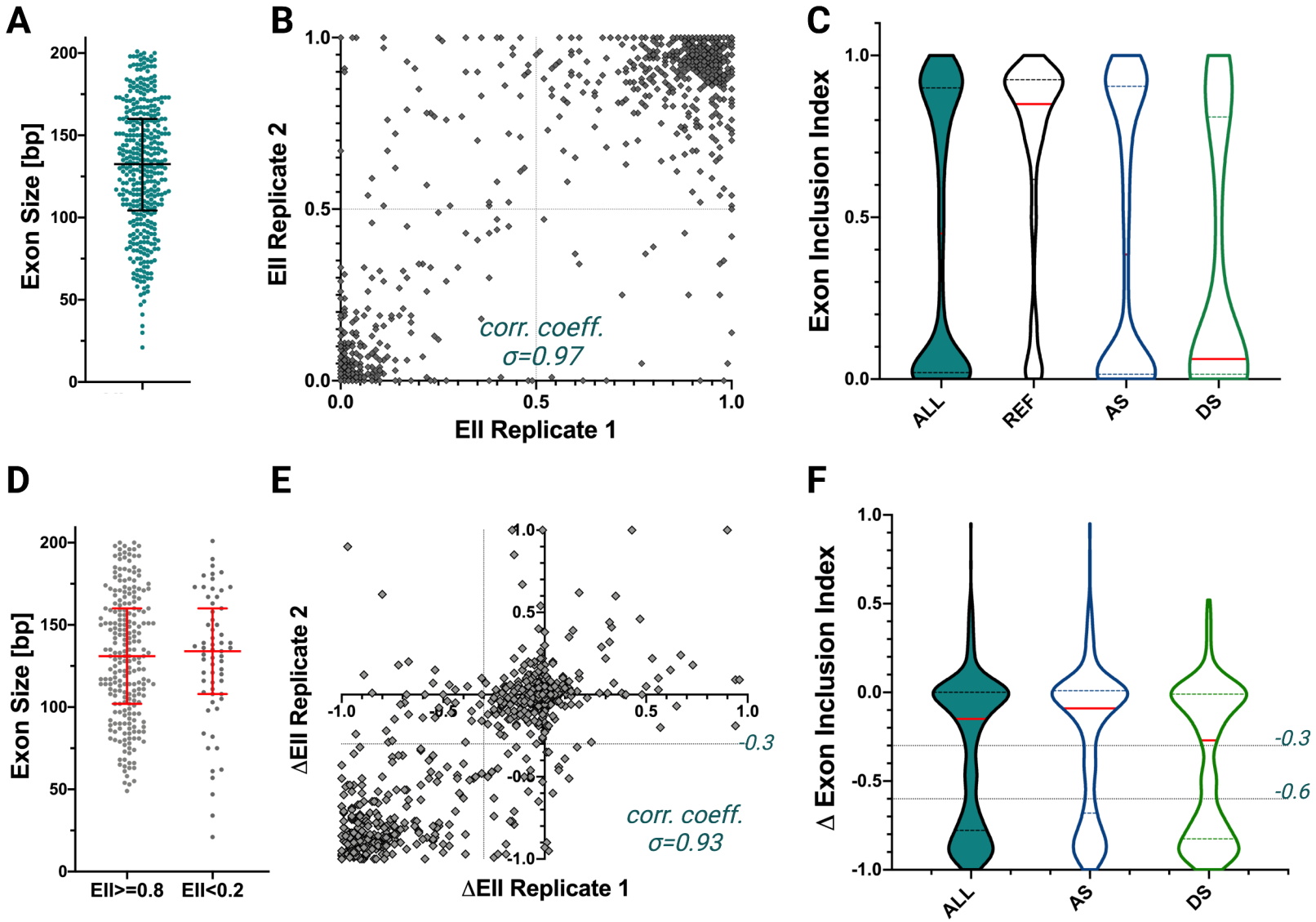
Figure 1. High throughput splicing assay (HTSA) experimental design and analysis. A) HTSA construct contains *GFP* split by an *SMN1* intron, into which a test exon is cloned by *AgeI* and *NheI* restriction cloning. mCherry is expressed independently as a background fluorescence control. A pool of exons (up to 200 bp in size) and flanking intronic sequences are cloned into a sequence derived from *SMN1* intron 7 and 8. **B)** Major splicing outcomes with exon inclusion leading to GFP disruption and exon exclusion leading to GFP reconstitution. Other outcomes are possible, e.g. partial exon inclusion or intron retention wich will also lead to the GFP disruption. **C)** RCA7 HEK293T landing pad locus contains an attP site and a gene-less promoter and the HTSA construct contains split GFP-SMN1 splicing minigene, mCherry and a promotor-less puromycin resistance gene. Co-transfection of the HTSA construct and a plasmid containing the Bxb1 integrase facilitates integration of one HTSA construct into the landing pad via the attB | attP pairing. Successful integration enables expression of puromycin resistance gene. **D)** RCA7 HEK293T landing pad cells are transfected with the HTSA plasmid pool and Bxb1 integrase. Selection with puromycin ensures elimination of cells without the integrated HTSA construct and prolonged cell growth (20 days) ensures elimination of episomal HTSA plasmids. Fluorescence activated cell sorting (FACS) leads to a clear separation of GFP+ve and GFP-ve cells. **E)** After isolation of gDNA from the cells, the integrated constructs are PCR-amplified and deep sequenced. The sequence reads are matched to the library sequences and read of each integrated sequence is counted in each cell population separately. The oligo counts are subsequently normalized by the sequence depth of the sample (reads per million, RPM) and by the FACS sorting cell fraction of the sample. The resulting normalized oligo cell counts are then used to calculate exon inclusion index as shown.

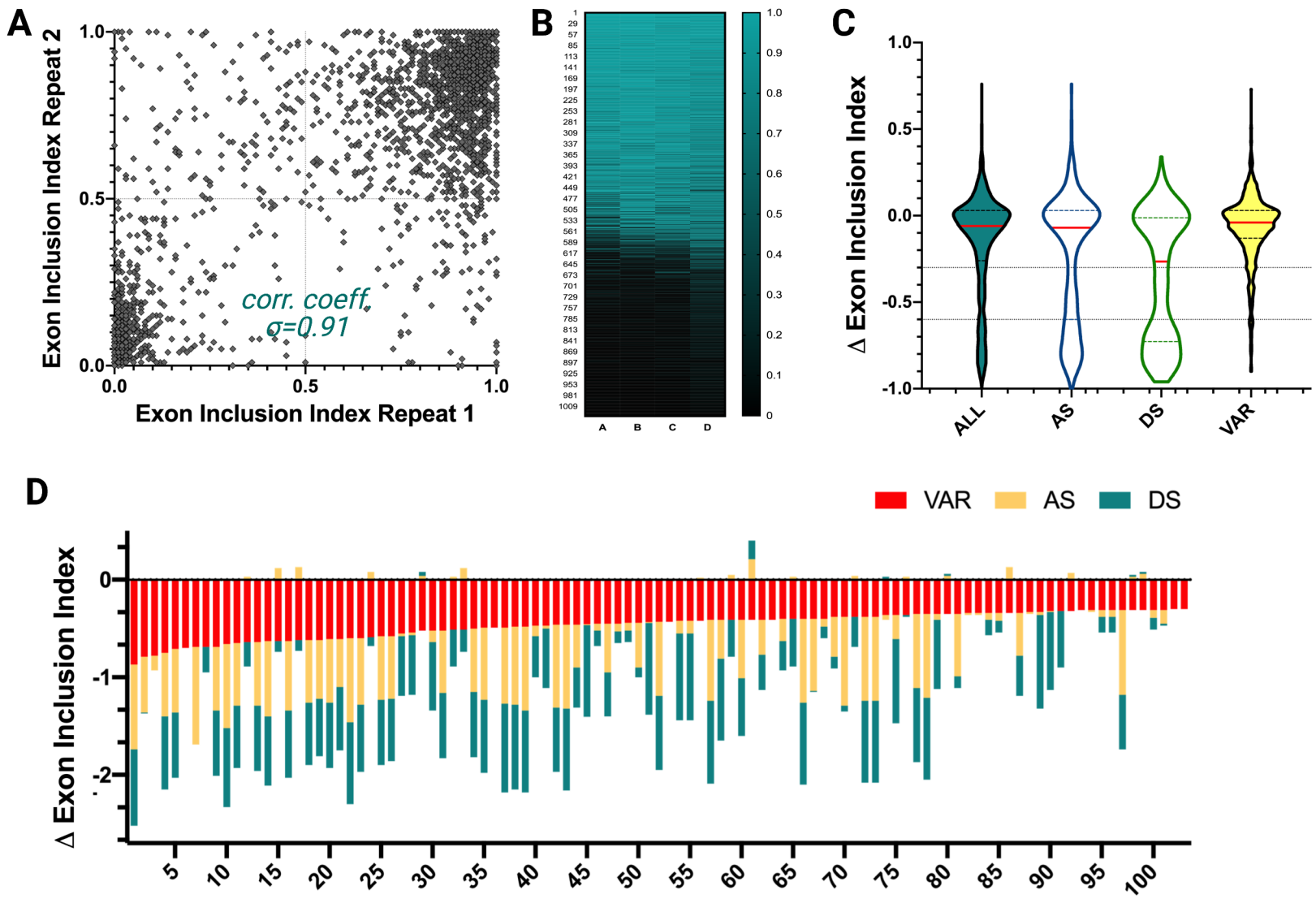
Figure 2. HTSA performance on a control library. A) Distribution of exon lengths in base pairs (bp) included in the study (Median length – 132 bp). **B)** Reproducibility of exon inclusion

indexes (EII) from replicate 1 and 2 (tetrachoric correlation coefficient, rho=0.97). **C)** Exon inclusion indexes for all control library sequences and separately for reference, acceptor (AS) and donor (DS) splice site mutations. **D)** Distribution of exon lengths for reference exon sequences that showed high rate of exon recognition (EII ≥ 0.8) and for reference exon sequences that showed preferential exons skipping (EII < 0.2). **E)** Reproducibility of delta exon inclusion indexes (ΔEII) from replicate 1 and 2 (tetrachoric correlation coefficient, rho=0.93). **F)** ΔEII values for the essential splice site controls, demonstrating a bimodal distribution of variants around ΔEII = 0 and ΔEII = -1. The distribution of data suggests two cutoff point to estimate splicing effects: ΔEII≤ -0.3 for moderate effect on exon inclusion and ΔEII ≤ -0.6 for strong effect on exon inclusion.

Figure 3. Measuring the effects of variants of unknown significance on exon recognition. **A)** Reproducibility of exon inclusion indexes (EII) from replicate 1 and 2 from a library containing 2166 differenct sequences of references, essential splice site controls and VUSs (tetrachoric correlation coefficient, rho=0.97). **B)** A heatmap showing reproducibility of EII in control sequences (reference and essential splice site controls) from four biological experiments. **C)** Delta exon inclusion indexes (ΔEII) for variant library sequences (essential splice site controls and tested variants) and separately for acceptor (AS), donor (DS) and tested variants (VAR). **D)** Individual ΔEII values for 103 variants that showed a significant effect on exon recognition, presented alongside the ΔEII values for the essential splice site controls in the same exon (VAR, red, AS yellow and DA, green).


A
< 200bp
AgeI
40bp
exon
30bp
NheI
pool of test exons
EF1a
GFP-N
SMN1
GFP-C
EF1a
mCherry
B
Spliced in test sequence disrupts GFP
pre-mRNA
Exon is spliced out GFP reconstituted
C
attP
AAVS1 locus
Bxb1 recombinase
attB
puro
Integrated construct
attL
puro
D
Transfection
Puromycin selection and cultute (20 days)
FACS
mCherry
GFP
E
GFP-ve pool
GFP+ve pool
1
2
3
(1) gDNA extraction, PCR and NGS
(2) Read count for each oligo (exact match)
(3) Normalization by sequence depth (RPM) and sort fraction
Normalized oligo count
Normalized oligo count
Exon Inclusion Index (EII)
=
GFP-ve
GFP+ve + GFP-ve


A
Exon Size [bp]
200
150
100
50
0
B
EII Replicate 2
EII Replicate 1
1.0
0.5
0.0
corr. coeff.
σ=0.97
C
Exon Inclusion Index
ALL
REF
AS
DS
D
Exon Size [bp]
EII>=0.8
EII<0.2
E
ΔEII Replicate 2
ΔEII Replicate 1
-1.0
-0.5
0.5
1.0
-0.3
corr. coeff.
σ=0.93
F
Δ Exon Inclusion Index
ALL
AS
DS
-0.3
-0.6


A
Exon Inclusion Index Repeat 2
Exon Inclusion Index Repeat 1
corr. coeff.
σ=0.91
B
A
B
C
D
C
Δ Exon Inclusion Index
ALL
AS
DS
VAR
D
VAR
AS
DS
Δ Exon Inclusion Index