Splicing profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes

Tyler Landrith, Bing Li, Ashley A. Cass, Blair R. Conner, Holly LaDuca, Danielle B. McKenna, Kara N. Maxwell, Susan Domchek, Nichole A. Morman, Christopher Heinlen, Deborah Wham, Cathryn Koptiuch, Jennie Vagher, Rebekah Krakenberg, Meredith Seidel, Robert Pilarski, Meagan Farmer, Khateria Pyrtel, Kara Milliron, John Lee, Christopher Heinlen, Deborah Wham, Cathryn Koptiuch, Jennie Vagher, Rebekah Krakenberg, Meredith Seidel, Robert Pilarski, Meagan Farmer, Khateria Pyrtel, Kara Milliron, John Lee, Christopher Heinlen, Deborah Wham, Cathryn Koptiuch, Jennie Vagher, Rebekah Krakenberg, Meredith Seidel, Robert Pilarski, Meagan Farmer, Khateria Pyrtel, Kara Milliron, John Lee, Christopher Heinlen, Deborah Wham, Cathryn Koptiuch, Jennie Vagher, Rebekah Krakenberg, Meredith Seidel, Robert Pilarski, Meagan Farmer, Khateria Pyrtel, Kara Milliron, John Lee, Christopher Heinlen, Deborah Wham, Cathryn Koptiuch, Jennie Vagher, Rebekah Krakenberg, Meredith Seidel, Robert Pilarski, Meagan Farmer, Khateria Pyrtel, Kara Milliron, John Lee, Christopher Heinlen, Deborah Wham, Cathryn Koptiuch, Jennie Vagher, Rebekah Krakenberg, Meredith Seidel, Robert Pilarski, Meagan Farmer, Khateria Pyrtel, Kara Milliron, John Lee.

INTRODUCTION

Splicing is the removal of non-coding sequences (introns) from an RNA molecule followed by the ligation of exons, the protein coding regions of genes. DNA variants can impact this process resulting in RNA mis-splicing, such as skipping of coding sequences or inclusion of non-coding ones into the messenger RNA (mRNA), resulting in potential allele loss-of-function. Aberrant splicing data associated with a DNA variant can be used as evidence of pathogenicity, whereas normal splicing data can be used as evidence of neutrality. RNA-sequencing (RNA-seq) has shown significant potential for improving the diagnostic yield and resolution of DNA genetic testing, primarily because of the functional splicing data generated by this analysis. Importantly, RNA-seq also addresses a technical limitation of most clinically available DNA genetic tests, which typically capture only exons and short stretches of theflanking introns. Pathogenic variants (PVs) outside the captured sequence will be missed with a DNA-only approach; however, the addition of RNA-seq provides an opportunity to uncover mis-splicing caused by intronic events, leading to the identification of PVs in the non-coding region of genes. The first attempts to incorporate RNA-seq into clinical diagnostics have involved whole-transcriptome sequencing (WTS) for patients with rare Mendelian disorders who have remained without a molecular diagnosis despite receiving whole-exome or whole-genome sequencing. The addition of WTS has been shown to increase diagnostic yield by 7–36%, depending on the disease studied. Across all studies, pathogenic splicing variants were identified in regions typically captured by current DNA testing methods as well as deep-intronic regions, highlighting the utility of RNA-seq in both identification and interpretation of disease-causing splicing variants. Studies have also shown the benefits of RNA-seq for hereditary cancer predisposition genes; however, this approach has been traditionally performed as a follow-up to inconclusive DNA testing. In a recent study, RNA genetic test results facilitated classification of 88% of the cancer gene splicing variants selected for analysis as either pathogenic or benign, and was predicted to impact 1 in 43 individuals if performed simultaneously with DNA testing. Thus, a substantial proportion of patients currently receiving DNA testing are likely to benefit from the addition of RNA genetic testing. Several studies have also identified pathogenic deep-intronic variants across a range of hereditary cancer conditions, including
RESULTS

Splicing profile of TSGs in healthy controls

In this study, a splicing profile of 18 TSGs by capture RNA-seq was performed in 345 healthy individuals, 25 positive controls, and 1000 patients referred to a clinical diagnostic laboratory for suspicion of hereditary cancer (Supplementary Fig. 1a, b). All genes were found to be expressed in blood (Supplementary Fig. 2a) and had at least 85% of captured exons covered at >50 reads (Supplementary Fig. 2b), our minimum quality thresholds. We compared these expression data to the values found in the GTEx database from whole-blood WTS. All 18 genes were also detected in blood by WTS, although many at lower levels than by capture RNA-seq (Supplementary Table 1). We measured alternative splicing events defined relative to the most clinically relevant isoform for each gene (see Methods and Supplementary Fig. 1c). The relative expression of splicing events was measured by percent splicing index (PSI) (see Methods). While a minority of alternative splicing events were highly expressed, most exhibited low PSI (Supplementary Fig. 2c). Demographic distribution of the healthy donors is shown in Fig. 1a–c, the majority being Caucasians (54%), females (61%), and 30–40 years old (56%). The number of splicing events was not statistically associated with reported ethnicity (Fig. 1d). Additionally, PSI did not cluster by the observed metadata (age, gender, ethnicity, and batch), suggesting that these factors did not confound the observed PSI variation among healthy donors (Fig. 1e).

Next, we aimed to characterize alternative splicing events in TSGs common to the healthy population (i.e., PSI ≥5% in ≥5% of healthy donors). Different genes appeared to tolerate different degrees of alternative splicing. ATM, BRCA1, MUTYH, and NF1 displayed the highest median number of splicing events and MSH2, MSH6, PTEN, and TP53 the lowest (Fig. 2a, Supplementary Figs. 3–5). Interestingly, the two genes with the highest number of exons also had a high median splice event count (ATM, NF1). In order to determine if the observed data might be explained by the number of exons, we calculated the Pearson’s correlation coefficient between median splice event count and exon number for all 18 genes. Overall, we observed that there is a correlation between number of exons and median splice event count (\( \rho = 0.57, p = 0.01 \)). While 26 recurrent splicing events were identified in BRCA1 (Fig. 2b), only nine such events were observed in BRCA2 (Fig. 2c). In ATM, 12 events were identified, eight of which were detected in most of the healthy donors at low median PSI (Fig. 2d). Some alternative splicing events detected represent known alternative isoforms. For example, NF1 r.4110_4111ins4110 + 3819_4110 + 3881 corresponds to inclusion of an exon annotated in isoform NM_001042492.26,28 and exons specific to NM_001354896 and NM_001354900 were detected in APC in all controls at low levels (Fig. 2e, g). In MUTYH, 15 events were identified, most with median PSI <20% (Fig. 2f). Interestingly, 100% of controls expressed an in-frame partial skipping of exon 3 in NF1 (r.158_199del, median PSI = 38.69; interquartile range = 31.91–47.13) (Fig. 2f). Four partial exon 3 skipping events were detected in MUTYH, consistent with previous reports.24–26 Altogether, these data provide a reliable splicing landscape for these TSGs in whole blood of healthy individuals, against which putative abnormal splicing events observed in patients can be analyzed.

Splicing profile helps detect mis-splicing

Having generated a splicing reference dataset to contextualize putative pathogenic splicing events, we tested the 25 positive controls. These were blood samples from individuals heterozygous for likely PVs/PVs known to affect splicing of genes associated with HBOC, colorectal cancer predisposition (e.g., Lynch syndrome, familial adenomatous polyposis), or hereditary diffuse gastric cancer. For all tested variants, the PSI for the variant-associated splicing event was greater than the mean PSI among controls (one-sample two-sided t test, \( p < 1 \times 10^{-5} \)) (Fig. 3a and Supplementary Table 2). These results were validated using a second orthogonal methodology, CloneSeq, a highly sensitive and specific targeted RNA-seq assay based on cloning of reverse transcription-PCR (RT-PCR) products, followed by massively parallel sequencing of the cloned transcripts.27 CloneSeq results are displayed as Sashimi plots (Fig. 3b–g and Supplementary Figs. 6–8), which provide an absolute number of aligned reads that enables comparison of exon usage across probands and controls.28 Among the positive controls, we observed concordant results between the CloneSeq and capture RNA-seq for all clinically significant splicing events (Fig. 3b–g and Supplementary Figs. 6–8), demonstrating the utility of capture RNA-seq to detect mis-splicing associated with germline PVs.

Paired DNA and RNA genetic testing identifies PVs in TSG

Finally, we sought to assess in a clinical context the ability of whole-blood splicing profile by capture RNA-seq to detect variants that result in mis-splicing of these 18 TSGs. We performed simultaneous capture RNA-seq and DNA MGPT in 1000 consecutive patients referred for clinical inherited cancer predisposition testing, to evaluate the change in the diagnostic yield upon adding concurrent RNA testing to MGPT. RNA data was incorporated during variant assessment as recommended by the ACMG/AMP expert guidelines, alongside with other available lines of evidence.29–31 In total, 84 individuals received positive results (PVs/likely PVs) after performing consecutive DNA-sequencing and RNA-seq, versus 77 individuals who would have received positive results if MGPT was performed alone, resulting in a 9.1% relative increase in the diagnostic rate. All abnormal splicing events were confirmed by Sanger sequencing of cloned RT-PCR products (Supplementary Fig. 9).

Two cases exemplify the utility of splicing profile by capture RNA-seq in identifying PVs in the HBOC predisposition gene BRCA1. In the first case, the proband is a female with Caucasian/Asian descent diagnosed with breast cancer at the age of 25 and 33 years, and high-grade papillary serous ovarian cancer at the age of 44 years. She had multiple HBOC tests over the last 18 years which were inconclusive. Splicing profile identified an out-of-
frame partial retention of intron 2 (r.80_81ins81-8_81-1), which was not detected in controls (Fig. 4a, b). Analysis of the DNA identified a cytosine to guanine substitution nine nucleotides upstream from **BRCA1** intron 2 (c.81−9C > G), which is predicted in silico\(^3\) to weaken the native splice acceptor and create a cryptic acceptor, as confirmed by the RNA data. The resulting aberrant transcript is predicted to cause a glutamic acid to leucine substitution at position 29 in the amino acid chain and result in a frameshift causing a premature termination codon (p.E29Lfs*5). In the second case, the proband is a female Caucasian of Ashkenazi Jewish descent diagnosed with breast cancer at the age of 49 years. Skipping of exon 17 (r.5075_5152del) was observed in the proband but not in controls (Fig. 4a, c). A **BRCA1** variant six nucleotides downstream of exon 17 was detected by DNA MGPT (c.5152 + 6T > G), which is predicted to abolish the splice donor. This in turn is predicted to result in the deletion of amino acids at positions 1692 through 1718 and the insertion of a glycine (p. D1692_W1718delinsG). Saturation mutagenesis studies in **BRCA1** support that such a change would result in a non-functional protein.

In **BRCA2**, full skipping of exon 5 (r.426_475del) was observed at approximately twice the highest PSI observed in controls (57.7%) (Fig. 4a, d) in a Caucasian female diagnosed with breast cancer at the age of 53 years. This was associated with a rare, germline thymine deletion four nucleotides downstream from the exon (c.475 + 4delT), predicted to abolish the native donor splice site. The aberrant transcript is predicted to result in a proline to glycine substitution at position 143 and a frameshift resulting in a premature termination codon (p.P143Gfs*23). This variant was also previously detected in our laboratory cohort in a Caucasian male diagnosed with breast cancer at the age of 81 years, and prostate cancer at the age of 69 years.

In **ATM**, two splicing variants were reclassified from inconclusive to positive due to RNA evidence. The first variant was detected in a Hispanic female proband diagnosed with breast cancer at the age of 64 years. The **ATM** variant is a rare guanine to adenine substitution five nucleotides downstream from exon 56 (c.8418 +
Fig. 2 Healthy controls’ alternative splicing events detected in TSG. a Boxplot indicating median number of splicing events across controls with ≥5% PSI and coverage ≥50x for each hereditary cancer predisposition gene tested. b-g Plots indicating frequency (bar graph) and median PSI with interquartile range (boxplot) for alternative splicing events with PSI ≥5% of controls (17/345 controls). For each gene, alternative splicing events are divided into in-frame and frameshift transcripts. Splicing events are as follows: ES = exon skipping—a combination of a full and partial exon skipping event, ESP = full exon skipping—the entire length of the exon is skipped, ESP’ = partial exon skipping—some portion of the exon is skipped (i.e., alternative acceptor/donor), IC = cryptic exon—an intronic insertion (split reads on 5’ and 3’ end), IP = partial intronic insertion—an extension of the exon into the intron either upstream (5’) or downstream (3’) of the exon. b-d Hereditary breast and ovarian cancer genes (BRCA1/BRCA2/ATM), e NF1, and f, g colorectal/polyposis genes’ (APC/MUTYH) alternative splicing events.
The splicing in silico tools predicted a weakened donor splice site. The aberrant transcript identified by capture RNA-seq (Fig. 4a, e) is predicted to result in amino acid deletions at positions 2757 through 2806, which is located in a critical function domain for the protein (p.V2757_M2806del). In another ATM case (Fig. 4a, f) we identified a partial skipping of exon 19 (r.3061_3077del) in a Caucasian female diagnosed with invasive ductal carcinoma of the breast at the age of 33 years. This was associated with a thymine to guanine substitution located within the same exon (c.3065T > G). The RNA data confirmed splicing
predictions that the alteration would strengthen a cryptic donor splice site. The transcript is predicted to result in a valine to alanine substitution at position 1021 with a frameshift and resulting in a premature termination codon.

In MUTYH (Fig. 4a, g), we detected a partial intron 7 retention (r.S76_577ins577-4_S77-1) in a 31-year-old female with no personal history of cancer. This was associated with an adenine to guanine substitution five nucleotide upstream from exon 8 (c.577 – 5A > G). The transcript was predicted to result in valine to serine substitution at position 193 with a frameshift and premature termination codon. MUTYH causes an autosomal recessive polyposis syndrome, MUTYH-associated polyposis. In accordance with the expected phenotype, this variant has been previously identified in our laboratory cohort to co-occur with the founder PV MUTYH p.G396D in two patients with severe polyposis.

Finally, we detected in a Caucasian female diagnosed with breast cancer at the age of 65 years partial skipping of PMS2 exon 1 (r.11_23del), which was absent from controls (Fig. 4a, h). This splicing event was associated with a rare germline exonic variant in PMS2 (c.11C > G). Predictions in silico indicated that this cytosine to guanine substitution results in the creation of a novel donor site that is 13 base pairs upstream from the native splice donor site, and this was confirmed by the RNA data. The aberrant transcript is predicted to result in an alanine to valine substitution at position 4 with a frameshift and premature termination codon (p.A4VFS*26).

DISCUSSION

Results from our study demonstrate the feasibility and utility of capture RNA-seq in identifying patients with clinically actionable variants that would have either been missed with DNA testing alone or classified as inconclusive. This whole-blood assay also offers a clinical alternative to other less scalable methods for analysis of splicing variants, such as minigene assays. In this cohort, the 9.1% relative increase in the diagnostic yield for high-risk cancer genes is comparable if not superior to the relative increase in yield gained by the introduction of DNA copy number variation analysis. We anticipate that RNA-seq will have an even higher impact in the diagnostic yield among specific genes and phenotypic sub-groups, which will be evaluated as the testing cohort continues to expand.

The additional increase in diagnostic yield offered by RNA-seq improves a clinician’s ability to accurately apply personalized management strategies for early detection, cancer risk reduction, and treatment. In six of the seven cases with RNA-seq-related positive test results, substantial changes to medical management would be recommended based on current guidelines—not only for the probands but also for family members who test positive as well. PVS in BRCA1 and BRCA2 have implications not only for early detection and risk reduction of several associated cancer types but also have the potential to inform eligibility for PARP inhibitor therapy. Similarly, early detection and prevention options are recommended for individuals with PMS2 PVS. For individuals with pathogenic ATM variants, increased breast cancer surveillance is recommended.

Our capture RNA-seq approach required the generation of a unique control dataset to characterize the splicing landscape of TSGs in healthy individuals. Evaluating whether alternative splicing events are part of normal biological variation is an essential quality component of a clinical grade RNA-seq assay. In the absence of control data, laboratories risk over-interpreting splicing events as pathogenic. Considering the frequency of alternative splicing events in the healthy control population reported this study, failure to consider the natural splicing landscape in cancer genes could result in misclassifications.

Previous studies have leveraged WTS approaches to assess for pathogenic splicing variants across the genome. While our capture RNA-seq approach is more limited in the number of genes analyzed, it has a distinct advantage over WTS as it offers increased depth of coverage for genes that are not highly expressed in blood yet have a known clinical impact (e.g., BRCA1 and BRCA2). Along with further development of standards for the interpretation of RNA findings and the generation of publicly available databases for alternative splicing events, this approach should also be considered to accompany DNA testing panels for other hereditary diseases.

In summary, our results demonstrate the clinical utility of adding capture RNA-seq to routine genetic testing. The splicing profile approach we described can obtain a high-resolution picture of abnormal splicing events for clinically relevant genes, consequently improving the positive yield of genetic testing and supporting its adoption in a clinical context.

METHODS

Ethics

The study was approved and carried out in accordance with the recommendations of the Western Institutional Review Board (Puyallup, Washington). All participants provided written informed consent to take part in the study.

Samples

For the healthy controls and retrospective cohort, peripheral blood was drawn from 345 healthy donors and 83 patients participating in the Ambry Genetics Family Studies program, respectively. For the prospective cohort, patients were referred for paired DNA and RNA MGPT to Ambry Genetics by their physician or healthcare provider from 17 medical centers across the United States as part of routine clinical genetic cancer risk assessment (n = 1000). Harvested blood was collected and stored in Paxgene™ Blood RNA tubes (PreAnalytiX). For the retrospective cohort, RNA was extracted using either a manual method (Paxgene™ Blood RNA Kit—PreAnalytiX) or an automated one (Thermo MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit, compatible with Paxgene™ Blood RNA Tubes on Thermo King Finesse (Presto)). RNA yield and RNA integrity number were quantified using a TapeStation with RNA ScreenTape (Agilent).

DNA-sequencing and RNA-seq

For capture RNA-seq, complementary DNA (cDNA) libraries were generated using a RiboErase™ RNA Hyper Prep Kit (KAPA Biosystems) as previously described. Briefly, ribosomal RNA was captured via DNA oligonucleotide probes and digested using RNase H. The enriched RNA was heat fragmented at 94°C for 6 min and first-strand cDNA was
synthesized immediately thereafter. Second-strand synthesis and A-tailing was followed by adapter ligation, SPRI cleanup, and index PCR. The amplified and SPRI cleaned libraries were then assessed for yield and aberrant peaks using a Tapestation (Agilent) with D1000 ScreenTape. Libraries passing the quality assessment were pooled and subjected to bait capture as described in Mercer et al.18 The post-capture libraries were subjected to a second round of index PCR amplification, SPRI cleaned, and then assessed with the Tapestation. Twelve library pools (96 samples) were sequenced at 2 × 150 bp on a NextSeq 500 (Illumina). CloneSeq and Sanger sequencing of individual colonies was performed as previously described.27 DNA MGPT was performed as previously described.41
Sequencing data processing

Sequencing raw data processing was described previously. De-multiplexing was done with the RTA software 1.17.21.3 (Real Time Analysis, Illumina Inc., San Diego, CA, USA) and bc1fastq Conversion software v1.8 (Illumina Inc., San Diego, CA, USA). RNA samples that passed sequencing quality control (QC) criteria (percentage of Q30 bases >75%, mean base quality >30, and percentage of perfect index >85%) were used for downstream analysis. Paired-end RNA-seq reads (2x150 bp) were first aligned to the hg19 human reference genome using the STAR aligner v2.6.0c. Average coverage for each exon from the 18 genes (Supplementary Table 3) was calculated. A sample passed sequencing coverage QC cutoff if ≥85% of exons from the 18 genes have average coverage ≥50×. RNA samples that failed the coverage or QC cutoffs were re-prepared and re-sequenced. Data from samples that again failed the coverage and/or QC cutoffs are not included in the study. Mapped reads were then analyzed by custom in-house software based on Schafer et al. to detect splicing events relative to canonical RefSeq transcripts for the 18 hereditary cancer predisposition genes (Supplementary Fig. 2a and Supplementary Table 3). PMS2 exons 11 to 15 were excluded from analysis due to interference with >99% homologous pseudogene PMS2CL. One reference isoform was chosen for each gene. Relative to this isoform, the PSI value was defined as the number of reads supporting the alternative splicing event (i.e., the non-canonical/abnormal event) divided by the number of all reads in the region covering splicing event. This calculation differs slightly from previously published methods of splicing quantification in that it considers abnormal splicing reads as the numerator as opposed to the number of inclusion reads. To test for differential PSI between the patient and healthy controls, we used a one-sample two-sided t-test. The Holm–Sidak correction was used to adjust the p values for multiple hypothesis testing. A splicing event from a patient was considered significantly higher than healthy controls if its PSI was ≥5% and larger than the mean PSI of healthy donors, the adjusted p value was <0.05, and the number of all reads in the region covering splicing event was ≥50. Splicing events with lower PSI than controls were not considered.

Bar plots, box plots, and violin plots were generated using ggplot2 package (v3.5.3) from R v3.5.1 with default settings. Hierarchical clustering was generated using heatmap2 v3.5.2. Sashimi plots were generated using rmats2sashimiplot (https://github.com/Xinglab/rmats2sashimiplot).

Interpretation of genetic test results

Classification of germline variants identified by MGPT was performed following the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines. Multiple lines of evidence are combined, to reach one of the following classifications: PV, variant, likely pathogenic; variant of unknown significance, likely benign; or benign variant. ACMG/AMP guidelines place greater weight on experimental evidence of abnormal splicing compared with in silico predictors. In this study, RNA results were used as weighted evidence towards pathogenic or benign classification in the assessment of identified germline variants, following the ACMG/AMP guideline recommendations. Updated variant classifications are publicly available at ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/).

Diagnostic rate

Among the 1000 patients referred to Ambry for suspicion of hereditary cancer, patients with a PV within any of the 18 TSGs prior to the incorporation of RNA data were considered the denominator for diagnostic rate (n = 77). This calculation excluded moderate risk mutations (MUTYH carriers, CHEK2 1577T and APC 11307K). Physicians/healthcare providers could select up to 18 of the TSGs on the RNA panel for testing (all 18 genes were not selected in every case). The numerator for diagnostic rate was the difference between the number of patients with a PV after incorporating RNA data (n = 84) and the number of patients with a PV prior to RNA data (n = 77). This yielded a relative percent increase of 9.1%.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

RNA sequence data that support the findings of this study have been deposited in NCBI Sequence Read Archive (SRA) with the submission accession number PRJNA663342. The patient data that support the findings of this study are available on request from the corresponding author (R.K.). The patient data are not publicly available due to them containing information that could compromise research participant privacy/consent.

Received: 11 September 2019; Accepted: 30 January 2020; Published online: 24 February 2020

REFERENCES

1. Rivas, M. A. et al. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. Science 348, 666–669 (2015).
2. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. Nat. Rev. Genet. 17, 19–32 (2016).
3. Richards, 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. 17, 405–424 (2015).
4. Park, E., Pan, Z., Zhang, Z., Lin, L. & Xing, Y. The expanding landscape of alternative splicing variation in human populations. Am. J. Hum. Genet. 102, 11–26 (2018).
5. Cummings, B. B. et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci. Transl. Med. 9, https://doi.org/10.1126/scitranslmed.aal5209 (2017).
6. Fresead, L. et al. Identification of rare-disease genes using blood transcriptome sequencing and large cohort controls. Nat. Med. 25, 911–919 (2019).
7. Gonorazky, H. D. et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. Am. J. Hum. Genet. 104, 466–483 (2019).
8. Lee, H. et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. Genet. Med. https://doi.org/10.1038/s41436-019-0672-1 (2019).
9. Karam, R. et al. Assessment of diagnostic outcomes of RNA genetic testing for hereditary cancer. JAMA Netw. Open 2, e1913900, https://doi.org/10.1001/jamanetworkopen.2019.13900 (2019).
10. Cavallieri, S., Pozzi, E., Gatti, R. A. & Brusco, A. Deep-intronic ATM mutation detected by genomic resequencing and corrected in vitro by antisense morpholinoligonucleotide (AMO). Eur. J. Hum. Genet. 21, 774–778 (2013).
11. Montalban, G. et al. Screening of BRCA1/2 deep intronic regions by targeted gene sequencing identifies the first germline BRCA1 variant causing pseudoexon activation in a patient with breast/ovarian cancer. J. Med. Genet. 56, 63–74 (2019).
12. Borras, E. et al. Comprehensiv functional assessment of MLH1 variants of unknown significance. Hum. Mutat. 33, 1576–1588 (2012).
13. Clendenning, M. et al. Mutation deep within an intron of MSH2 causes Lynch syndrome. Fam. Cancer 10, 297–301 (2011).
14. Spier, I. et al. Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. Hum. Mutat. 33, 1045–1050 (2012).
15. Pros, E. et al. Nature and mRNA effect of 282 different NF1 point mutations: focus on splicing alterations. *Hum. Mutat.* **29**, E173–E193 (2008).

16. Avigad, S. et al. A novel germ line p53 mutation in intron 6 in diverse childhood malignancies. *Oncogene* **14**, 1541–1545 (1997).

17. LaDuca, H. M. et al. A clinical guide to hereditary cancer panel testing: evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. *Genet. Med.* https://doi.org/10.1038/s41436-019-0633-8 (2019).

18. Mercer, T. R. et al. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat. Protoc.* **9**, 989–1009 (2014).

19. Schafer, S. et al. Alternative splicing signatures in mRNA-seq data: percent spliced in (PSI). *Curr. Protoc. Hum. Genet.* **87**, 11 16 11–16 14 (2015).

20. Nishi, T. et al. Differential expression of two types of the neurofibromatosis type 1 (NF1) gene transcripts related to neuronal differentiation. *Oncogene* **6**, 1555–1559 (1991).

21. Trovo-Marqui, A. B. & Tajara, E. H. Neurofibromin: a general outlook. *Clin. Genet.* **70**, 1–13 (2006).

22. Horii, A., Nakatsuru, S., Ichi, S., Nagase, H. & Nakamura, Y. Multiple forms of the APC gene transcripts and their tissue-specific expression. *Hum. Mol. Genet.* **2**, 283–287 (1993).

23. Sulekova, Z., Reina-Sanchez, J. & Ballhausen, W. G. Multiple APC messenger RNA isoforms encoding exon 15 short open reading frames are expressed in the context of a novel exon 10A-derived sequence. *Int. J. Cancer* **63**, 435–441 (1995).

24. Plotz, G. et al. MUTYH gene expression and alternative splicing in controls and polyposis patients. *Hum. Mutat.* **33**, 1067–1074 (2012).

25. Out, A. A. et al. Leiden open variation database of the MUTYH gene. *Hum. Mutat.* **31**, 1205–1215 (2010).

26. Oka, S. & Nakabeppu, Y. DNA glycosylase encoded by MUTYH functions as a tumorigenesis. *Science* **326**, 677–682 (2010).

27. Farber-Katz, S. et al. Quantitative analysis of BRCA1 and BRCA2 germline splicing variants using a novel RNA-serially parallel sequencing assay. *Front. Oncol.* **8**, 286 (2018).

28. Katz, Y. et al. Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics* **31**, 2400–2402 (2015).

29. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 562–568 (2016).

30. Findlay, G. M. et al. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222 (2018).

31. Desmet, F. O. et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).

32. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).

33. Fraile-Bethencourt, E. et al. Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigen assay. *J. Pathol.* **248**, 409–420 (2019).

34. Fraile-Bethencourt, E., Valenzuela-Palomo, A., Diez-Gomez, B., Acedo, A. & Velasco, E. A. Identification of eight spliceogenic variants in BRCA2 exon 16 by minigen assay. *Front. Genet.* **9**, 188 (2018).

35. Bonnet, C. et al. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J. Med. Genet.* **45**, 438–446 (2008).

36. Stella, A. et al. A nonsense mutation in MLH1 causes exon skipping in three unrelated HNPCC families. *Cancer Res.* **61**, 7020–7024 (2001).

37. van der Klift, H. M. et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol. Genom. Genet.* **3**, 327–345 (2015).

38. Steffensen, A. Y. et al. Functional characterization of BRCA1 gene variants by mini-gene splicing assay. *Eur. J. Hum. Genet.* **22**, 1362–1368 (2014).

39. NCCN. The NCCN Clinical Practice Guidelines in Oncology™ Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic V1.2020 (2020). http://www.nccn.org/.

40. NCCN. The NCCN Clinical Practice Guidelines in Oncology™ Genetic/Familial High-Risk Assessment: Colorectal V2.2019 (2019). http://www.nccn.org/.

41. Conner, B. R. et al. RNA genetic testing identifies germline mutation MSH2.p.Lys248X in Lynch syndrome patients. *Gastroenterology* https://doi.org/10.1053/j.gastro.2019.01.248 (2019).