



# Quantitative Analysis of *BRCA1* and *BRCA2* Germline Splicing Variants Using a Novel RNA-Massively Parallel Sequencing Assay

Suzette Farber-Katz<sup>1</sup>, Vickie Hsuan<sup>1</sup>, Sitao Wu<sup>2</sup>, Tyler Landrith<sup>1</sup>, Huy Vuong<sup>2</sup>, Dong Xu<sup>2</sup>, Bing Li<sup>2</sup>, Jayne Hoo<sup>3</sup>, Stephanie Lam<sup>3</sup>, Sarah Nashed<sup>4</sup>, Deborah Toppmeyer<sup>4</sup>, Phillip Gray<sup>3</sup>, Ginger Haynes<sup>1</sup>, Hsiao-Mei Lu<sup>2</sup>, Aaron Elliott<sup>3</sup>, Brigette Tippin Davis<sup>3</sup> and Rachid Karam<sup>1\*</sup>

<sup>1</sup> Translational Genomics Laboratory, Ambry Genetics, Aliso Viejo, CA, United States, <sup>2</sup> Department of Bioinformatics, Ambry Genetics, Aliso Viejo, CA, United States, <sup>3</sup> Department of Research and Development, Ambry Genetics, Aliso Viejo, CA, United States, <sup>4</sup> Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

## OPEN ACCESS

### Edited by:

Paolo Peterlongo,  
IFOM - The FIRC Institute of Molecular  
Oncology, Italy

### Reviewed by:

Eladio Andrés Velasco,  
Instituto de Biología y Genética  
Molecular (IBGM), Spain  
Paolo Radice,  
Istituto Nazionale dei Tumori (IRCCS),  
Italy

### \*Correspondence:

Rachid Karam  
rkaram@ambrygen.com

### Specialty section:

This article was submitted to  
Cancer Genetics,  
a section of the journal  
Frontiers in Oncology

**Received:** 16 March 2018

**Accepted:** 09 July 2018

**Published:** 27 July 2018

### Citation:

Farber-Katz S, Hsuan V, Wu S,  
Landrith T, Vuong H, Xu D, Li B,  
Hoo J, Lam S, Nashed S,  
Toppmeyer D, Gray P, Haynes G,  
Lu H-M, Elliott A, Tippin Davis B and  
Karam R (2018) Quantitative Analysis  
of *BRCA1* and *BRCA2* Germline  
Splicing Variants Using a Novel  
RNA-Massively Parallel Sequencing  
Assay. *Front. Oncol.* 8:286.  
doi: 10.3389/fonc.2018.00286

Clinical genetic testing for hereditary breast and ovarian cancer (HBOC) is becoming widespread. However, the interpretation of variants of unknown significance (VUS) in HBOC genes, such as the clinically actionable genes *BRCA1* and *BRCA2*, remain a challenge. Among the variants that are frequently classified as VUS are those with unclear effects on splicing. In order to address this issue we developed a high-throughput RNA-massively parallel sequencing assay—*CloneSeq*—capable to perform quantitative and qualitative analysis of transcripts in cell lines and HBOC patients. This assay is based on cloning of RT-PCR products followed by massive parallel sequencing of the cloned transcripts. To validate this assay we compared it to the RNA splicing assays recommended by members of the ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles) consortium. This comparison was performed using well-characterized lymphoblastoid cell lines (LCLs) generated from carriers of the *BRCA1* or *BRCA2* germline variants that have been previously described to be associated with splicing defects. *CloneSeq* was able to replicate the ENIGMA results, in addition to providing quantitative characterization of *BRCA1* and *BRCA2* germline splicing alterations in a high-throughput fashion. Furthermore, *CloneSeq* was used to analyze blood samples obtained from carriers of *BRCA1* or *BRCA2* germline sequence variants, including the novel uncharacterized alteration *BRCA1* c.5152+5G>T, which was identified in a HBOC family. *CloneSeq* provided a high-resolution picture of all the transcripts induced by *BRCA1* c.5152+5G>T, indicating it results in significant levels of exon skipping. This analysis proved to be important for the classification of *BRCA1* c.5152+5G>T as a clinically actionable likely pathogenic variant. Reclassifications such as these are fundamental in order to offer preventive measures, targeted treatment, and pre-symptomatic screening to the correct individuals.

**Keywords:** genetic testing, hereditary breast and ovarian cancer, HBOC, *BRCA1*, *BRCA2*, RNA, Splicing, NGS

## INTRODUCTION

Correct interpretation of genomic sequence variants, and subsequent classification of variants as benign or pathogenic, is of utmost importance to patient management, especially in clinically actionable genes such as the breast and ovarian cancer susceptibility genes *BRCA1* and *BRCA2* (OMIM 113705 and 600185, respectively). Variant interpretation is based on multiple lines of evidence (1), including molecular and functional analysis, highlighting the urgent need to develop and implement high-throughput functional assays for variant classification (2).

Genomic sequence variants in *BRCA1* and *BRCA2* have the potential to alter normal splicing of these genes (3). In fact, many alterations in *BRCA1* and *BRCA2* have been shown to be clinically significant by RNA studies and multifactorial likelihood analyses that combine bioinformatics, pathologic, and clinical data (4–6). These variants include those that affect splicing by abolishing or weakening the canonical splice sites at intron-exon boundaries, by creating a novel or activating a cryptic splice site, or by disrupting enhancer or silencer splicing regulatory sequences (7).

Recommendations for mRNA analysis best practice in clinical testing were published by the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) (8), a consortium established in 2009 with the purpose of sharing data, methods, and resources to facilitate the classification of sequence variants in hereditary breast and ovarian cancer (HBOC) genes (9). ENIGMA recommends the use of RT-PCR and digital or capillary electrophoresis to detect abnormal transcripts based on the length of the product observed, followed by cloning and Sanger sequencing to characterize the sequence of these transcripts (8). However, the consortium notes that evaluation of splicing results for variant carriers can be complicated by the detection of normal alternatively spliced transcripts (8). Both *BRCA1* and *BRCA2* genes are known to undergo alternative splicing, and clinical and functional data indicate that alternatively spliced transcripts may retain function (10, 11). Two fundamental issues in determining the functional significance of normal and abnormal spliced transcripts are whether a transcript is out-of-frame, and therefore predicted to be targeted to degradation by the nonsense-mediated RNA decay (NMD) pathway (12), and the level at which these transcripts are expressed (13). Therefore, a combination of qualitative and quantitative analysis is needed to provide proper characterization of splice variations, and to establish the clinical significance of these specific alterations.

With these in mind, we developed CloneSeq, a high-throughput RNA-based massively parallel sequencing (MPS) technique designed to perform quantitative and qualitative characterization of splicing alterations in a time-frame necessary for clinical testing. Here we describe this technique and perform a comparison of CloneSeq with the techniques recommended by the ENIGMA consortium (8). We performed this comparison using four well-characterized lymphoblastoid cell lines (LCLs) generated from carriers of the *BRCA1* or *BRCA2* germline variants *BRCA1* c.5467+5G>T, *BRCA1* c.135-1G>T, *BRCA2* c.8632+1G>A, or *BRCA2* c.9501+3A>T. These variants have been previously described and are known to be associated with

splicing defects (8). Additionally, we used a similar strategy to analyze blood samples obtained from carriers of *BRCA1* or *BRCA2* germline sequence variants (Figure 1A), including the uncharacterized alteration *BRCA1* c.5152+5G>T identified in a novel HBOC family.

## MATERIAL AND METHODS

### Samples

This study was approved and carried out in accordance with the recommendations of the Western Institutional Review Board (WIRB). All subjects gave written informed consent in accordance with the Declaration of Helsinki. Blood from normal healthy controls or patients participating in the Ambry Genetics Family Studies program was drawn in PAXgene Blood RNA Tubes and stored according to the manufacturer's recommendations (PreAnalytiX, Hombrechtikon, Switzerland). RNA was extracted using the PAXgene Blood RNA Kit according to the recommended protocol (PreAnalytiX). Informed consent was obtained from all participants. Breast RNA was purchased from Amsbio (Lake Forest, CA, USA) and BioChain (Newark, CA, USA). RNA quality was determined using the RIN number calculated by the TapeStation 2200 with RNA ScreenTape or High Sensitivity RNA ScreenTape (Agilent, Santa Clara, CA, USA).

### Cell Lines

Lymphoblastoid cell lines (LCLs) were obtained from the Kathleen Cuninghame Consortium for Research into Familial Breast Cancer (kConFab, Melbourne, Australia) from 4 carriers of *BRCA1* or *BRCA2* variants and 2 controls. Genotypes were verified by Sanger sequencing. LCLs were maintained according to the recommendations of kConFab. Inhibition of nonsense-mediated decay (NMD) was performed using puromycin (300 µg/ml) or cycloheximide (100 µg/ml) for 4 h, as previously described (8, 14).

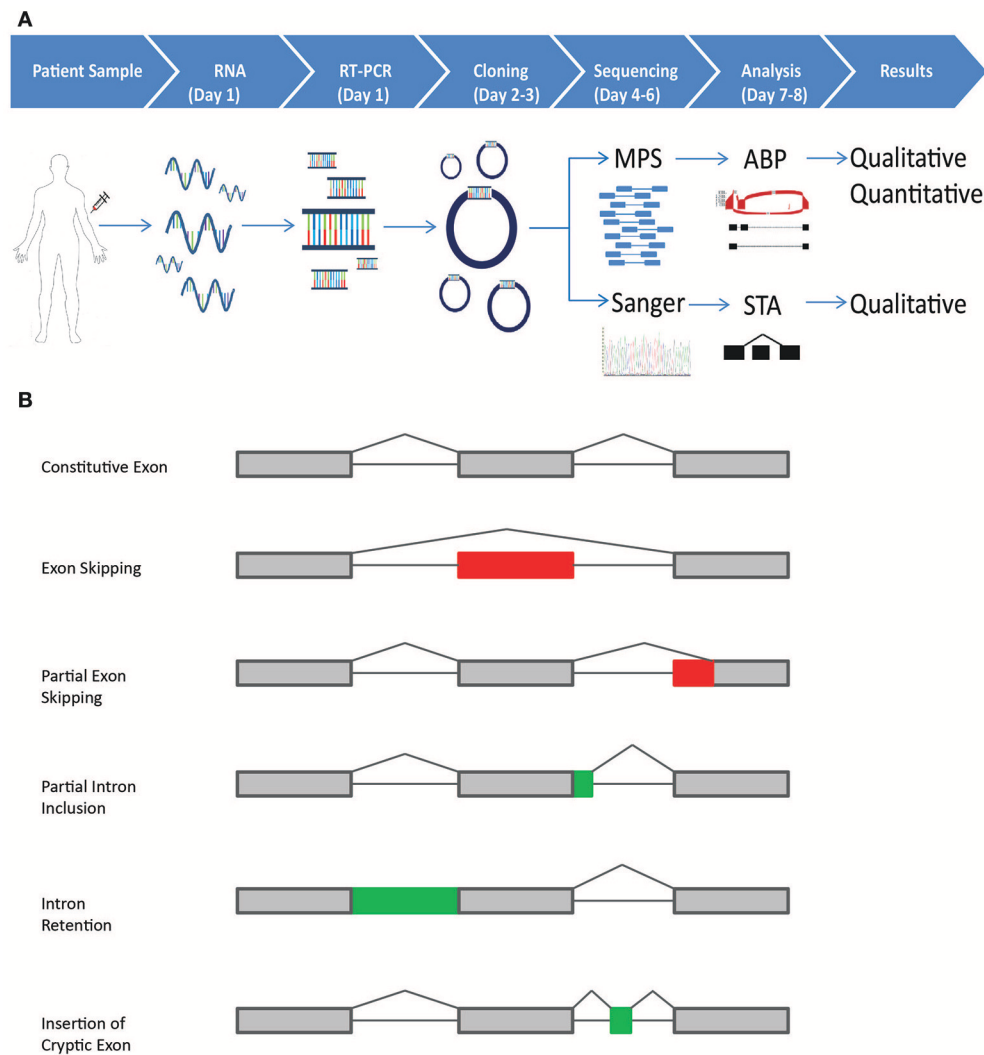
### RNA Analysis

cDNA was generated using the SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific, Chino, CA, USA). PCR was performed using either Platinum SuperFi PCR Master Mix (Thermo Fisher Scientific) or HotStarTaq Master Mix (Qiagen, Valencia, CA, USA) as previously described (8).

PCR products were analyzed using digital electrophoresis with D1000 ScreenTape and Reagents on the TapeStation 2200 (Agilent). Capillary electrophoresis (CE) was performed on an ABI 3730xl using MapMarker1000 as a standard (BioVentures, Murfreesboro, TN, USA). Primers were tagged at the 5' end with FAM or HEX for detection by CE. CE analysis was performed with GeneMapper software (Thermo Fisher Scientific).

PCR products were cloned into pGEM-T Easy and transformed into bacteria according to the manufacturer's recommended protocol (Promega, Fitchburg, WI, USA). Individual white colonies were picked, amplified by rolling-circle replication, and Sanger sequenced by Genewiz (La Jolla, CA, USA).

For CloneSeq, cDNA, PCR, and cloning were performed as described above. All colonies on a plate were scraped and



**FIGURE 1** | Schematic representation of the CloneSeq protocol and splicing events detected by the bioinformatics pipeline. **(A)** Blood from normal healthy controls and patients participating in the Ambry Genetics Family Studies program was drawn in PAXgene Blood RNA Tubes. RT-PCR was performed following ENIGMA recommendations. RT-PCR products were cloned into pGEM-T Easy and transformed into bacteria. For CloneSeq, all colonies on a plate were scraped and suspended in PBS. Plasmids were extracted, CloneSeq libraries were constructed, and Massively Parallel Sequencing (MPS) was performed, which generated  $2 \times 250$  paired-end reads. The mapped reads are then analyzed by the customized Ambry Bioinformatic Pipeline (ABP) software to generate qualitative and quantitative data for splicing events, including exon skipping, alternative 5' donor site, alternative 3' acceptor site, and intron retention. We confirmed CloneSeq results by comparing the data with ENIGMA-recommended assays, in which several individual positive colonies were picked, amplified by rolling-circle amplification, and Sanger sequenced. Single-transcript alignment (STA) was performed to characterize the transcripts' sequences. **(B)** The five types of alternative splicing events, as described by Diederichs et al. that can be detected by the ABP: (1) exon skipping; (2) partial exon skipping (as a result of the usage of alternative exonic donor or acceptor site); (3) partial intron inclusion (as a result of the usage of alternative intronic donor or acceptor site); (4) intron retention; (5) insertion of cryptic exons.

suspended in PBS. Plasmids were extracted with the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific). CloneSeq libraries were constructed according to the protocol outlined by KAPA Biosystems (Wilmington, MA, USA) using the Hyper Prep kit. Briefly, DNA was sheared to an average size of 250–350 bp using sonication (Covaris, Woburn, MA, USA). DNA fragment ends were repaired and phosphorylated. An “A” base was added to the 3' end of the blunted fragments, followed by ligation of single-indexed adapters via T-A mediated ligation. The size and concentration of the DNA library were determined

using the TapeStation 2200. Massively Parallel Sequencing (MPS) was performed on an Illumina MiSeq, which generated  $2 \times 250$  paired-end reads. Sequencing reads were aligned to the hg19 reference genome and analyzed using Ambry's Bioinformatic Pipeline (see below).

For whole transcriptome RNA-Seq, globin mRNA and ribosomal RNA were depleted using the Globin-Zero Gold rRNA removal kit (Illumina, San Diego, CA, USA). After depletion, RNA was fragmented and single-indexed cDNA libraries were generated using an RNA Hyper Prep kit

(KAPA Biosystems). Quality control was performed using the TapeStation 2200. Libraries were checked for average fragment size, concentration, and the presence of spurious peaks such as adapter dimers. Concentration was confirmed using a Qubit Fluorometer (ThermoFisher). Libraries were sequenced to a depth of  $1 \times 10^8$  paired end reads ( $2 \times 150$  bp) per sample on the Illumina NextSeq platform. Sequencing reads were aligned to the hg19 reference genome and analyzed using Ambry's Bioinformatic Pipeline for alternative splicing events and differentially expressed genes.

## Bioinformatics Analysis

Paired-end RNA-seq reads ( $2 \times 250$  bp) and Sanger sequencing reads ( $\sim 1,100$  bp) were first aligned to the hg19 human reference genome. For Sanger reads, GMAP aligner (version 2016-04-04) was used with default parameters to perform single transcript alignment (STA) of very long reads. For CloneSeq reads, STAR aligner v2.5.2a was used with default parameters except the "outSAMtype" parameter was set to "BAM SortedByCoordinate." The mapped reads were then analyzed by our customized Ambry Bioinformatics Pipeline (ABP) software to detect splicing events such as exon skipping, alternative 5' donor site, alternative 3' acceptor site, and intron retention (15). These events are detected by the pipeline, based on the alignments against the reference genome (**Figure 1B**): (1) exon skipping, if there is no reads align to one exon or several consecutive exons; (2) partial exon skipping, if there is no read alignment in one end of an exon; (3) partial intron inclusion, if there is alignment in one end of an intron; (4) intron retention, if there is alignment in a whole intron; (5) cryptic exon, if there is alignment in the middle of an intron and no alignment in the rest of the intron. Schematic representations of these splicing events are illustrated in **Figure 1B**. To quantify splicing events, we calculated the percentage of alternative splicing event against a given transcript/isoform: percent of alternative splicing event = (number of reads supporting alternative splicing event)/(number of all reads in the region covering alternative splicing event). To filter out noise caused by sequencing and alignment errors, or due to the expression of ultra-rare isoforms, the splicing events with "number of reads supporting alternative splicing event"  $< 20$ , or "number of all reads in the region covering splicing event"  $< 50$ , or "percent of splicing event"  $< 2.5\%$  were filtered out. HGVS nomenclature values were approximate for intron retention and alternative splicing site events due to differences in alignments based on NGS reads.

## RESULTS

### Quantitative and Qualitative RNA Analysis of the Variant *BRCA1* c.5467+5G>T

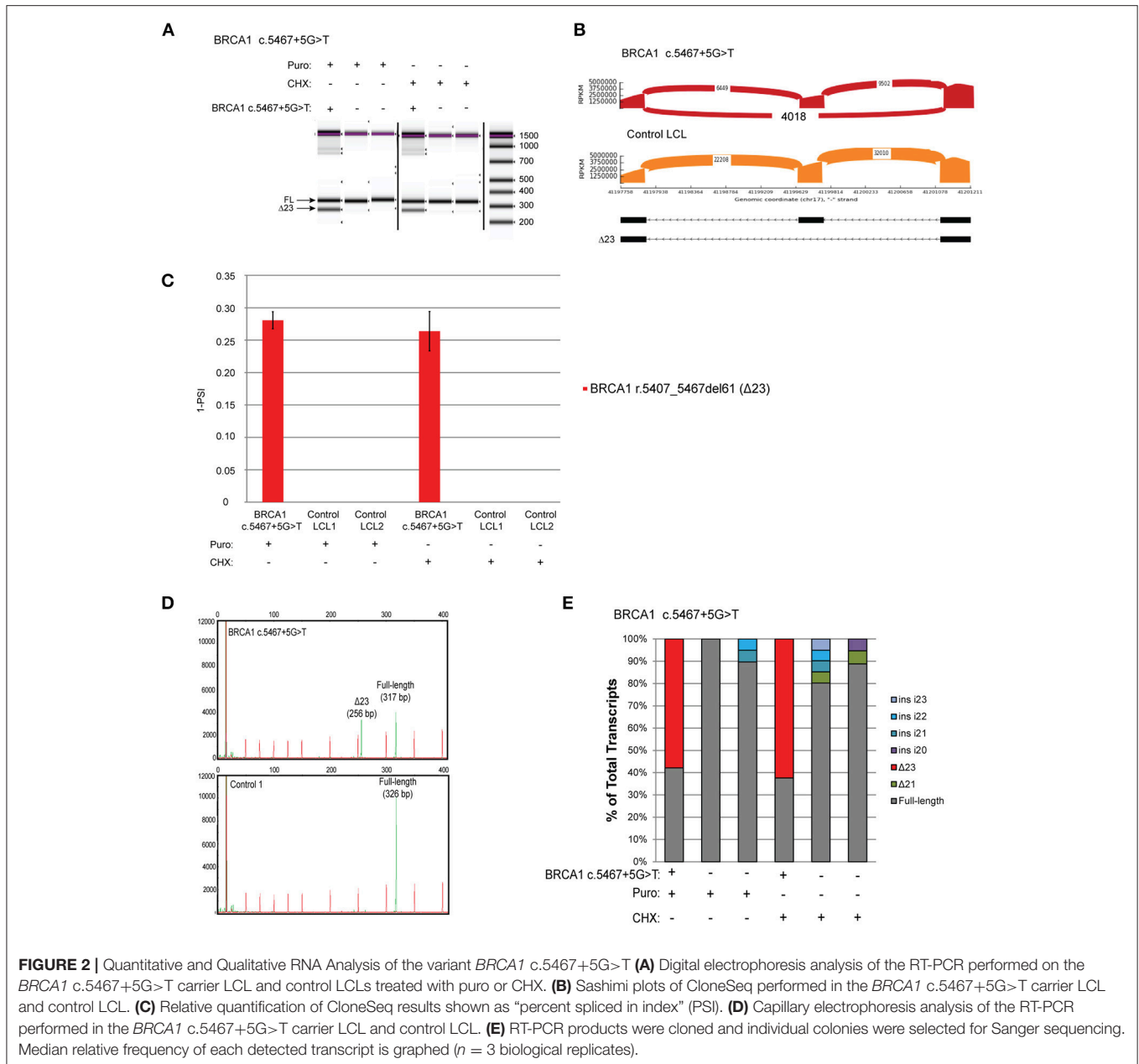
The variant *BRCA1* c.5467+5G>T, which impairs the native donor splice site of *BRCA1* exon 23, has been described to result in skipping of exon 23 ( $\Delta 23$ ) (8). This variant is currently classified as VUS by ENIGMA (class 3). LCLs were obtained from carriers of the variant and 2 controls, and reverse-transcriptase PCR (RT-PCR) was performed using the conditions

recommended by Whiley et al. including the use of the same primers, reverse transcriptase, and NMD inhibitors puromycin (puro) and cycloheximide (CHX) (8). Skipping of exon 23 ( $\Delta 23$ ) was clearly detected by digital electrophoresis of the RT-PCR products in the *BRCA1* c.5467+5G>T carrier's LCL, but not in two control LCLs treated with NMD inhibitors (**Figure 2A**). RT-PCR products were cloned and CloneSeq was performed on these samples for sequence characterization and quantification. The sequence and absolute number of reads observed in the carrier and control cell lines are shown using Sashimi plots (**Figure 2B**). Sashimi plots provide a quantitative visualization of aligned MPS reads that enables quantitative comparison of exon usage across samples (16). A total of 4,018 reads supporting exon 23 skipping (r.5407\_5467del61) were detected in the *BRCA1* c.5467+5G>T LCL, whereas none was detected in the control LCL (**Figure 2B**). Abnormal transcripts levels were then measured as a "percent spliced in index" (PSI) (**Figure 2C**). PSI demonstrates the ratio between reads including or excluding exons, indicating how efficiently sequences of interest are spliced into transcripts (17). This analysis indicated that  $\sim 25\%$  of transcripts expressed by the *BRCA1* c.5467+5G>T LCL contains skipping of exon 23, whereas skipping of exon 23 was not detected in negative control LCLs (**Figure 2C**).

To validate the CloneSeq results we performed, in the same set of samples, the mRNA splicing assays recommended by the members of the ENIGMA consortium (8), including capillary electrophoresis (**Figure 2D**), and Sanger sequencing of subcloned transcripts (**Figure 2E**). Capillary electrophoresis clearly detected  $\Delta 23$  in the *BRCA1* c.5467+5G>T LCL, in addition to the full-length WT transcript (**Figure 2D**). Sanger sequencing also detected  $\Delta 23$  exclusively in the *BRCA1* c.5467+5G>T carrier's LCL (**Figure 2E**).

### CloneSeq Characterization of the Pathogenic Alteration *BRCA1* c.135-1G>T

The variant *BRCA1* c.135-1G>T, which impairs the native acceptor site of *BRCA1* exon 5, has been associated with multiple splicing isoforms (8, 18), including an abundant transcript with skipping of exon 5 ( $\Delta 5$ ). This variant is currently classified as pathogenic by ENIGMA (class 5). RT-PCR for the *BRCA1* c.135-1G>T carrier's LCL and control LCLs was performed following ENIGMA recommendations and analyzed by digital electrophoresis, which detected a band consistent in size with  $\Delta 5$  (**Figure 3A**). RT-PCR products were cloned and CloneSeq performed, and the sequence and absolute number of reads observed in the carrier and control cell lines are shown using Sashimi plots (**Figure 3B**). A total of 10,902 reads supporting exon 5 skipping (r.135\_212del78) were detected in the *BRCA1* c.135-1G>T LCL, whereas only 584 reads were detected in the control LCL (**Figure 3B**). Quantification of splicing events indicated that the *BRCA1* c.135-1G>T LCL expresses  $\sim 50\%$  of transcripts with skipping of exon 5, whereas LCL negative controls have negligible levels of  $\Delta 5$  (**Figure 3C**). Individual colonies were selected for transcript confirmation by Sanger sequencing, and the  $\Delta 5$  transcript was the most

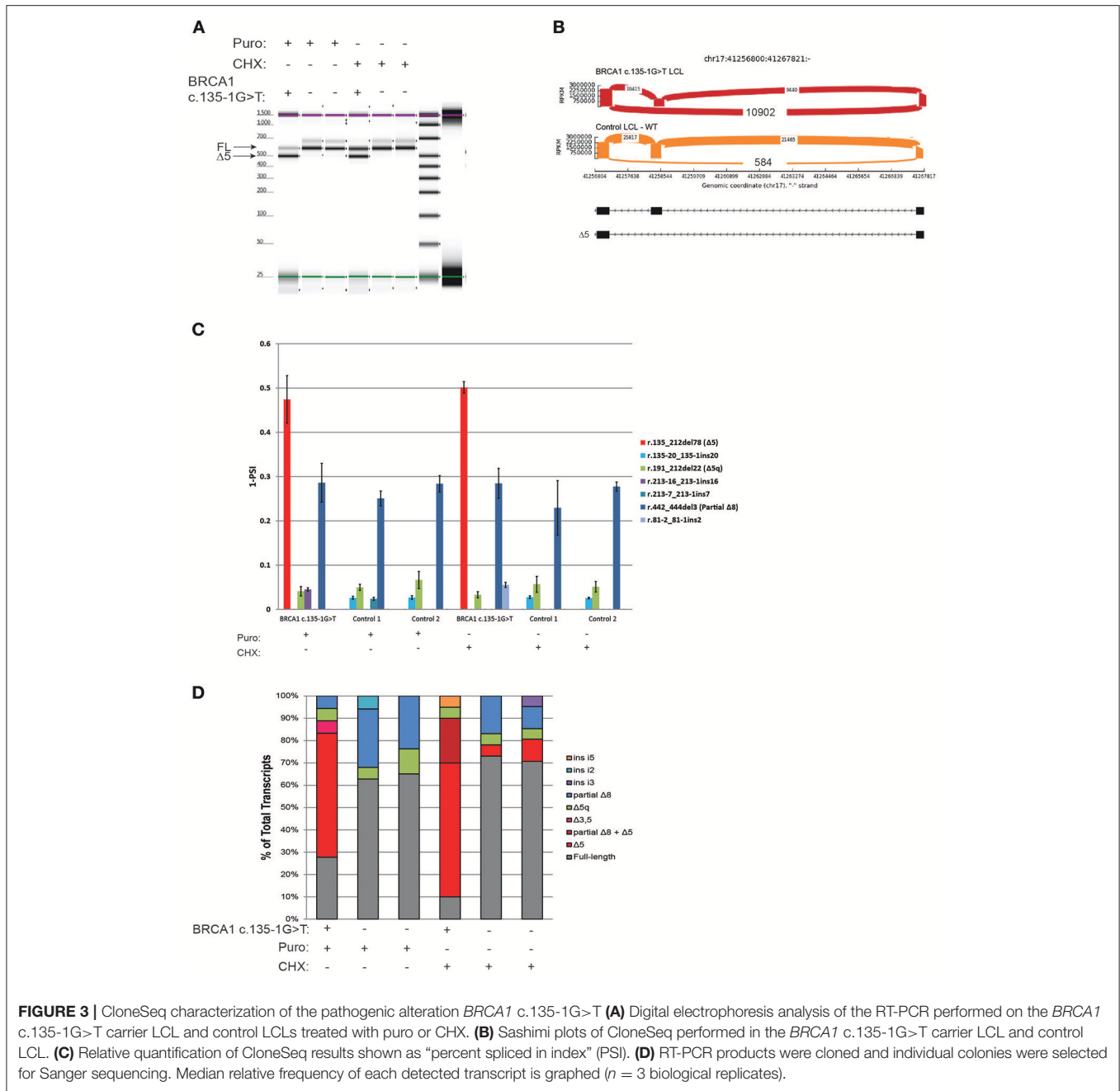


abundant abnormal transcripts in the *BRCA1* c.135-1G>T LCL (Figure 3D).

## CloneSeq Characterization of the Pathogenic Alteration *BRCA2* c.8632+1G>A

*BRCA2* c.8632+1G>A, located at intron 20 of *BRCA2*, was shown to result in skipping of exon 20 (Δ20) (8), and is currently classified by ENIGMA as a pathogenic alteration (class 5). RT-PCR of the *BRCA2* c.8632+1G>A carrier's LCL and control LCLs was performed following ENIGMA recommendations. Digital electrophoresis analysis of the RT-PCR products detected several

alternative transcripts, in addition to the full-length mRNA (Figure 4A). RT-PCR products were cloned and CloneSeq was performed. The sequence and absolute number of reads observed are shown in Sashimi plots for the carrier and control cell lines demonstrating a total of 3,002 reads supporting exon 20 skipping (r.8488\_8632del145) in the *BRCA2* c.8632+1G>A LCL, whereas only 25 reads supporting Δ20 were detected in the control LCL (Figure 4B). Quantification of splicing events indicated that ~20% of the transcripts expressed by the *BRCA2* c.8632+1G>A LCL have skipping of exon 20, whereas LCL negative controls have negligible levels of Δ20 (Figure 4C). Several alternative spliced transcripts, detected both in the *BRCA2* c.8632+1G>A LCL and controls, were also identified



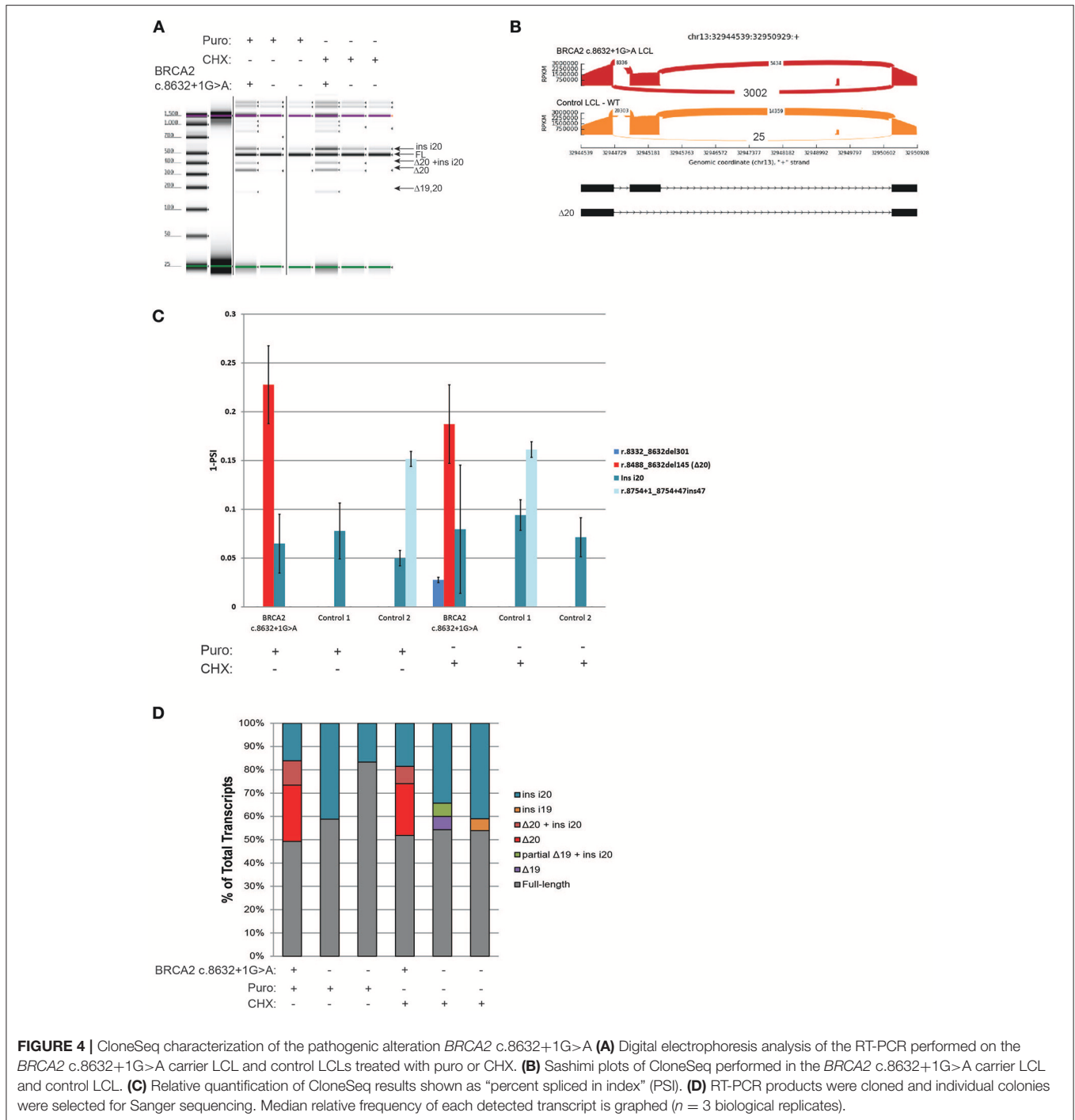
**FIGURE 3 |** CloneSeq characterization of the pathogenic alteration *BRCA1* c.135-1G>T **(A)** Digital electrophoresis analysis of the RT-PCR performed on the *BRCA1* c.135-1G>T carrier LCL and control LCLs treated with puro or CHX. **(B)** Sashimi plots of CloneSeq performed in the *BRCA1* c.135-1G>T carrier LCL and control LCL. **(C)** Relative quantification of CloneSeq results shown as “percent spliced in index” (PSI). **(D)** RT-PCR products were cloned and individual colonies were selected for Sanger sequencing. Median relative frequency of each detected transcript is graphed (*n* = 3 biological replicates).

(Figure 4C). Individual colonies were selected for transcript confirmation by Sanger sequencing, which identified the most abundant abnormal transcript Δ20 in the *BRCA2* c.8632+1G>A LCL, in addition to confirming other minor alternative isoforms detected by CloneSeq (Figure 4D).

### Characterization of the Variant *BRCA2* c.9501+3A>T in LCLs and Blood Samples

*BRCA2* c.9501+3A>T is located in the native donor site of intron 25. This variant was reported to result in low levels of skipping of exon 25 (Δ25), and it is currently classified as

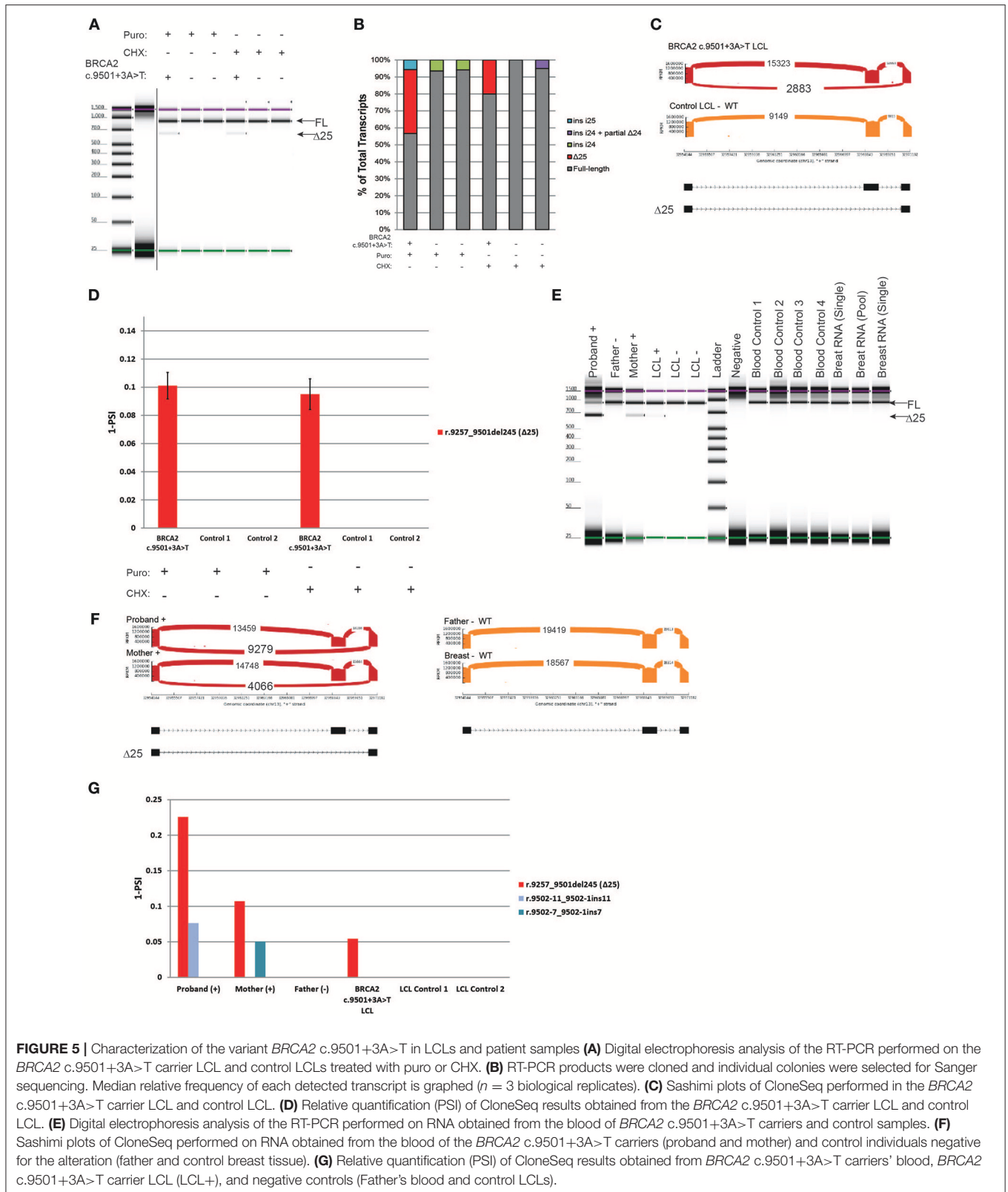
benign (class 1) by ENIGMA (8). RT-PCR was performed on the carrier’s LCL and on the control cells. Digital electrophoresis identified a minor band consistent with the size of Δ25 in the *BRCA2* c.9501+3A>T LCL that was not detected in negative controls (Figure 5A). RT-PCR products were cloned, and individual colonies were selected for Sanger sequencing, which detected Δ25 (r.9257\_9501del245) in the *BRCA2* c.9501+3A>T LCL, in addition to low levels of other alternatively spliced transcripts in controls (Figure 5B). To more accurately quantify the relative abundance of Δ25, we performed CloneSeq. The assay detected a total of 2,883 reads supporting exon 25



skipping in the *BRCA2* c.9501+3A>T LCL, whereas no reads supporting skipping of this exon were detected in the control LCL (**Figure 5C**). The reads supporting  $\Delta 25$  were  $\sim 10\%$  of the total splicing events detected in the *BRCA2* c.9501+3A>T LCL (**Figure 5D**).

Subsequently, we compared CloneSeq LCL results with an analysis of RNA isolated from the blood cells of carriers

of the *BRCA2* c.9501+3A>T alteration. RT-PCR performed on RNA from individuals that are heterozygous for *BRCA2* c.9501+3A>T (proband and mother) and negative controls (father, LCL-, normal breast RNA, and normal blood controls) detected  $\Delta 25$  only in the positive samples (**Figure 5E**). CloneSeq was performed to quantify  $\Delta 25$  in these samples, which detected 9,279 and 4,066 splicing events supporting  $\Delta 25$  in the proband



and mother's samples respectively, while none were detected in the negative controls (Figure 5F). Interestingly, the percentage of  $\Delta 25$  transcripts varies among different carriers. For the proband,

$\Delta 25$  represents  $\sim 20\%$  of splicing events, while it accounted for  $\sim 10\%$  in the mother's sample (Figure 5G),  $\sim 5\%$  in the LCL+ *BRCA2* c.9501+3A>T without NMD inhibition (Figure 5G),



and ~10% of splicing events when the LCL+ is treated with inhibitors (Figure 5D).

## Characterization of a Novel Variant, *BRCA1* c.5152+5G>T

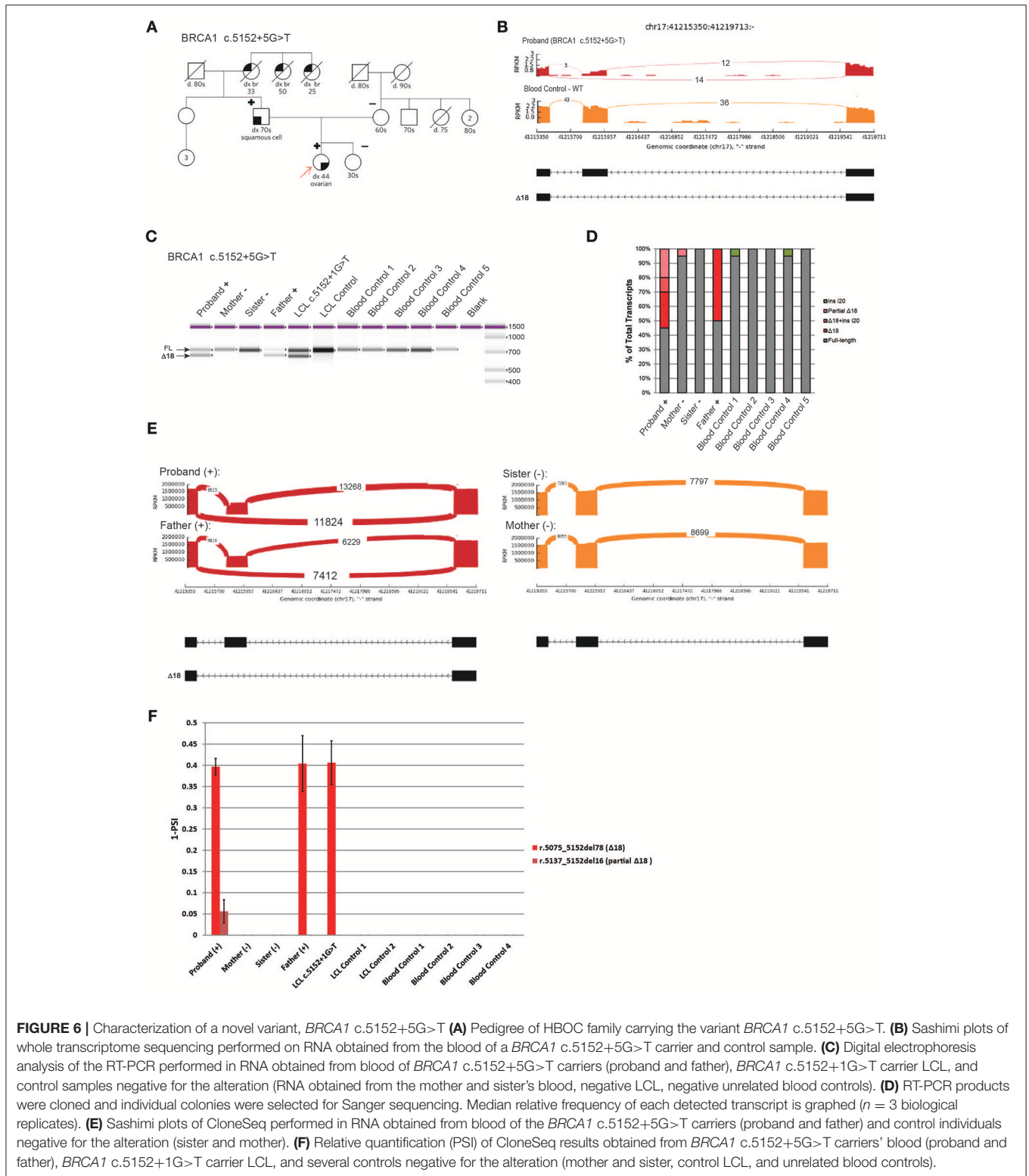
We next analyzed a novel uncharacterized VUS, *BRCA1* c.5152+5G>T, identified in a HBOC family (Figure 6A). This rare variant is located in the donor splice site of intron 18 at a highly conserved position, and was predicted by several splicing *in silico* programs to abolish the splice site (data not shown). We obtained blood samples from patients that are heterozygous for *BRCA1* c.5152+5G>T (proband and father) as well as samples from negative individuals (proband's mother and sister). To characterize the variant we performed whole transcriptome sequencing (WTS) on the proband and control blood samples. WTS detected 14 reads supporting skipping of exon 18 ( $\Delta 18$ , r.5075\_5152del78) and 12 reads supporting the WT transcript in the proband (Figure 6B, top). In the control blood sample, only WT reads ( $n = 36$ ) were detected (Figure 6B, bottom). Primers were designed in the flanking exons and RT-PCR was performed. Digital electrophoresis analysis of RT-PCR products identified a band corresponding to  $\Delta 18$ , exclusively in samples from the heterozygous carriers (Figure 6C). Sanger sequencing of subcloned transcripts was then performed, which confirmed  $\Delta 18$  sequence (Figure 6D), indicating it results in in-frame skipping of the important BRCT functional domain of *BRCA1* (19). Finally, using CloneSeq, we detected and quantified the  $\Delta 18$  transcript in heterozygous individuals, which was undetectable in non-carriers (Figure 6E). CloneSeq detected 11,824 reads supporting  $\Delta 18$  (r. c.5075\_5152del78) and 13,268 reads supporting WT transcript in the proband (Figure 6E, left top). In the proband's father, CloneSeq detected 7,412 reads supporting  $\Delta 18$ , and 6,229 reads supporting WT transcript (Figure 6E, left bottom). Quantitatively, we found heterozygous individuals to express ~40% of  $\Delta 18$  transcripts (Figure 6F). Of note, analysis of the LCL harboring the pathogenic alteration *BRCA1* c.5152+1G>T, affecting the same donor splice site as *BRCA1* c.5152+5G>T, also led to similar expression of the abnormal transcript  $\Delta 18$  (Figures 6C,F). Altogether, these data were used to reclassify *BRCA1* c.5152+5G>T from VUS to likely pathogenic, and therefore a clinically actionable alteration.

## DISCUSSION

Genetic testing for HBOC is becoming increasingly widespread in the era of precision medicine (20, 21). The implementation of next-generation sequencing has resulted in an explosion of genetic data, and germline variants with unknown function are regularly detected by clinical diagnostic laboratories (22). In particular, VUS in clinically actionable genes, such as the HBOC susceptibility genes *BRCA1* and *BRCA2*, pose a quandary to medical providers and patients (23–25). A specific challenge is the large percentage of VUS in the *BRCA1* and *BRCA2* genes predicted to affect splicing by *in silico* tools, but lack RNA evidence (26). In part, this is due to the scarcity of high-throughput assays designed to interrogate the impact of

variants on splicing. The American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines for the interpretation of germline sequence variants recommends the use of multiple types of evidence for classifying variants identified by DNA genetic testing, such as functional evidence, allele frequency data, computational and *in silico* predictions, and phenotype/family history (1). Therefore, RNA testing is critical to perform a comprehensive interpretation of sequence variants predicted to affect splicing. With this in mind, there are several RNA assays that have been used to characterize splicing alterations, each assay possessing its own advantages and drawbacks. These include the use of hybrid minigenes for the characterization of candidate splicing variants (27, 28). Given that a hybrid minigene assay analyzes the splicing outcome of a single allele, it is a great toll to evaluate allele-specific expression, i.e., the demonstration that the variant allele produces highly expressed abnormal transcripts predicted to induce NMD or to disrupt clinically important residues, an important step for the classification of splicing variants as pathogenic (10, 11). A caveat of this technique is the requirement of constructing and using artificial vectors (which are not available to most commercial diagnostic laboratories), and the need to test cell lines instead of samples derived from the patient being evaluated. The assays recommended by the ENIGMA consortium to characterize *BRCA1* and *BRCA2* transcripts include capillary electrophoresis and Sanger sequencing of subcloned transcripts, (8, 29). These assays are accessible to most laboratories, can be performed in patient samples, do not require the use of expression vectors, and can perform qualitative characterization of splicing variants; however, these assays only provide semi-quantitative data and lack the throughput required to analyze a large amount of variants (30, 31). Alternatively, there are quantitative approaches, including real-time and digital PCR, that provide robust and reliable quantitative data (29, 32), but cannot perform qualitative analysis (i.e., these assays are unable to reveal the precise transcript sequence). Here we describe a novel RNA MPS method, CloneSeq, and demonstrated that this technique is capable of performing reliable high-throughput quantitative and qualitative analysis of splicing variants, a necessary feature to obtain evidence for the large number of alterations predicted to affect splicing in HBOC genes.

Using CloneSeq coupled to our custom ABP bioinformatics analysis we were able to detect all major splicing aberrations described by the ENIGMA consortium in four well-characterized LCLs (8), with the advantage of obtaining absolute and relative quantification of the expressed transcripts (Figures 2–5). CloneSeq proficiently detected abnormal transcripts, as well as less abundant alternative splicing events, as it analyzed thousands of cloned reads. We validated the CloneSeq results using digital and capillary electrophoresis and Sanger sequencing of subcloned transcripts. Digital and capillary electrophoresis identified the abundant abnormally spliced transcripts detected by CloneSeq, but these methods were incapable of providing proper quantification of transcript levels due to their semi-quantitative nature. Additionally, sequencing was needed to



confidently identify the exact splicing event detected by digital and capillary electrophoresis. Both CloneSeq and Sanger sequencing were able to precisely determine the sequence of the transcripts, however, CloneSeq performed sequencing

of tens of thousands of transcripts (average number of *BRCA1* or *BRCA2* mapped reads *per* sample tested was 24,803). In comparison, Sanger sequencing is limited to low-throughput sequencing of colonies, each containing a

single subcloned transcript (median=56 clones sequenced *per* LCL). As examples of its high analytical sensitivity, CloneSeq was able to detect the splicing isoforms induced by the variants *BRCA1* c.135-1G>T and *BRCA2* c.8632+1G>A, and rare alternative spliced isoforms previously reported by ENIGMA (Figures 3, 4). Ultimately, CloneSeq's targeted ultra-deep *locus* sequencing of *BRCA1* or *BRCA2* proved to be a fundamental feature for the bioinformatics qualitative and quantitative characterization of splicing alterations in these genes.

In addition to LCLs, we analyzed RNA extracted from blood samples obtained from variant carriers and controls. For the previously characterized *BRCA2* c.9501+3A>T benign variant (class 1), we were able to compare LCL RNA data with RNA data from the blood of heterozygous *BRCA2* c.9501+3A>T carriers. Similar to Sanger sequencing, CloneSeq detected the major abnormal splicing event associated with this variant, skipping of exon 25, both in LCLs and in carriers' blood RNA. Because quantification of abnormally and alternatively spliced transcripts is fundamental to predict pathogenicity (10, 11, 29), we quantified the impact of this benign variant on splicing levels. The percentage of skipped exon 25 identified in different *BRCA2* c.9501+3A>T carriers ranged from ~20 to ~10% of total splicing events, suggesting that an alteration resulting in less than ~20% of abnormal splicing is clinically benign. However, it is important to note, this is an indirect measurement of each allele's expression, since we were unable to perform allele-specific expression in the individuals we tested due to the lack of informative variants in the coding sequence of *BRCA2* in the respective samples. In order to mitigate this caveat, we ran a series of normal blood and tissue controls to identify and differentiate any physiologic alternatively spliced isoform from abnormal transcripts identified in the variant carriers (Figures 5E–G). The CloneSeq results are also in agreement with a minigene single-allele analysis of this alteration that reported ~13% of  $\Delta 25$  is induced by the variant *BRCA2* c.9501+3A>T (27). By quantifying the impact that variants with benign clinical behavior have on splicing, CloneSeq could be used in the future to identify a splicing threshold that must be reached by abnormal transcripts in order to classify a *BRCA1* or *BRCA2* alteration as pathogenic.

Lastly, we analyzed blood samples obtained from a HBOC family carrying a novel uncharacterized VUS, *BRCA1* c.5152+5G>T. To characterize the VUS we performed whole transcriptome sequencing in the proband and control blood samples. WTS detected 14 reads supporting skipping of exon 18, and 12 reads supporting the WT transcript in the proband. Using CloneSeq, we were able to detect in the proband 11,824 reads supporting skipping of exon 18 and 13,268 reads supporting the WT transcript. Comparatively, the number of reads detected by WTS vs. CloneSeq highlights the higher analytical sensitivity of the later. This supports the notion that WTS can provide biased results due to low detection yields and other technical limitations (33, 34). On the other hand, CloneSeq provided the sufficient sequencing depth necessary for transcript characterization and quantification, which proved to be indispensable to reclassify the VUS *BRCA1* c.5152+5G>T as a clinically actionable likely pathogenic alteration.

Massively parallel sequencing is revolutionizing cancer genetics by enabling the detection and characterization of sequence variants at unprecedented scale and speed. For example, depending on the technology and protocol used, the number of individuals tested *per* variant, and the number of controls tested, CloneSeq can concomitantly perform analysis of multiple variants in a single MPS run (up to 96 samples). From the initial RNA extraction steps to the final bioinformatics analysis, the protocol described here can analyze multiple samples in less than 10 days (Figure 1A). Even though NGS technologies have evolved quickly over the past decade, leading to a substantial decrease in the cost per megabase (30), the cost of NGS assays may still pose challenges to laboratories with low throughput (31). However, implementation of automated steps and the development of innovative sequencing technologies could reduce the cost and time-frame of CloneSeq even further in the near future. Besides laboratory costs, one important issue to consider is the impact RNA genetic testing has on variant classification. One of the caveats of DNA MPS multi-gene testing is the high rate of VUS results (35). Since RNA genetic testing has the potential to reduce VUS rates, future research should investigate the broader impact these tests have on the overall clinical management of patients identified with germline variants in *BRCA1* and *BRCA2*.

## CONCLUSION

CloneSeq is an alternative to the current splicing assays recommended by the ENIGMA consortium. Due to its high-throughput format, quantitative and qualitative abilities, and high analytical sensitivity, CloneSeq has the potential to improve the interpretation of splicing sequence variants detected by HBOC clinical genetic tests. The enhanced classification of these germline variants as either disease-causing or neutral is fundamental to offering preventive pre-symptomatic measures or targeted treatment to the correct individuals.

## AUTHOR CONTRIBUTIONS

SF-K and RK drove the development of the intellectual concepts, performed analyses, interpreted data, and wrote the manuscript. VH, JH, SL, and TL performed experiments, generated data, and wrote the manuscript. SW, HV, DX, BL, and H-ML developed and ran the bioinformatics pipeline, and wrote the manuscript. GH, SN, and DT collected data, samples, and wrote the manuscript. PG, BT, and AE contributed to manuscript writing and to experimental design.

## FUNDING

This study was funded entirely by Ambry Genetics.

## ACKNOWLEDGMENTS

We would like to thank the patients and medical providers that contributed with samples. We thank Benjamin Paluch for his helpful comments and assistance with figures preparation.

## REFERENCES

1. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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) 17:405–24. doi: 10.1038/gim.2015.30
2. Pesaran T, Karam R, Huether R, Li S, Farber-Katz S, Chamberlin A, et al. Beyond DNA: an integrated and functional approach for classifying germline variants in breast cancer genes. *Int J Breast Cancer* (2016) 2016:2469523. doi: 10.1155/2016/2469523
3. Nielsen FC, van Overeem Hansen T, Sorensen CS. Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer* (2016) 16:599–612. doi: 10.1038/nrc.2016.72
4. Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, et al. Functional assays for classification of BRCA2 variants of uncertain significance. *Cancer Res.* (2008) 68:3523–31. doi: 10.1158/0008-5472.CAN-07-1587
5. Walker LC, Whiley PJ, Couch FJ, Farrugia DJ, Healey S, Eccles DM, et al. Detection of splicing aberrations caused by BRCA1 and BRCA2 sequence variants encoding missense substitutions: implications for prediction of pathogenicity. *Hum Mutat.* (2010) 31:E1484–1505. doi: 10.1002/humu.21267
6. Thomassen M, Blanco A, Montagna M, Hansen TV, Pedersen IS, Gutierrez-Enriquez S, et al. Characterization of BRCA1 and BRCA2 splicing variants: a collaborative report by ENIGMA consortium members. *Breast Cancer Res Treat.* (2012) 132:1009–23. doi: 10.1007/s10549-011-1674-0
7. Caminsky N, Mucaki EJ, Rogan PK. Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *F1000Res* (2014) 3:282. doi: 10.12688/f1000research.5654.1
8. Whiley PJ, de la Hoya M, Thomassen M, Becker A, Brandao R, Pedersen IS, et al. Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin Chem.* (2014) 60:341–52. doi: 10.1373/clinchem.2013.210658
9. Spurdle AB, Healey S, Devereau A, Hogervorst FB, Monteiro AN, Nathanson KL, et al. ENIGMA—evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum Mutat.* (2012) 33:2–7. doi: 10.1002/humu.21628
10. Colombo M, Blok MJ, Whiley P, Santamarina M, Gutierrez-Enriquez S, Romero A, et al. Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum Mol Genet.* (2014) 23:3666–80. doi: 10.1093/hmg/ddu075
11. Fackenthal JD, Yoshimatsu T, Zhang B, de Garibay GR, Colombo M, De Vecchi G, et al. Naturally occurring BRCA2 alternative mRNA splicing events in clinically relevant samples. *J Med Genet.* (2016) 53:548–58. doi: 10.1136/jmedgenet-2015-103570
12. Karam R, Wengrod J, Gardner LB, Wilkinson MF. Regulation of nonsense-mediated mRNA decay: implications for physiology and disease. *Biochim Biophys Acta* (2013) 1829:624–33. doi: 10.1016/j.bbagr.2013.03.002
13. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* (2008) 456:470–6. doi: 10.1038/nature07509
14. Karam R, Carvalho J, Bruno I, Graziadio C, Senz J, Huntsman D, et al. The NMD mRNA surveillance pathway downregulates aberrant E-cadherin transcripts in gastric cancer cells and in CDH1 mutation carriers. *Oncogene* (2008) 27:4255–60. doi: 10.1038/onc.2008.62
15. Diederichs S, Bartsch L, Berkmann JC, Frose K, Heitmann J, Hoppe C, et al. The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. *EMBO Mol Med.* (2016) 8:442–57. doi: 10.15252/emmm.201506055
16. Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Thorvaldsdottir H, et al. Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics* (2015) 31:2400–2. doi: 10.1093/bioinformatics/btv034
17. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). *Curr Protoc Hum Genet.* (2015) 87:11–4. doi: 10.1002/0471142905.hg1116s87
18. Tesoriero AA, Wong EM, Jenkins MA, Hopper JL, Brown MA, Chenevix-Trench G, et al. Molecular characterization and cancer risk associated with BRCA1 and BRCA2 splice site variants identified in multiple-case breast cancer families. *Hum Mutat.* (2005) 26:495. doi: 10.1002/humu.9379
19. Clark SL, Rodriguez AM, Snyder RR, Hankins GD, Boehning D. Structure-function of the tumor suppressor BRCA1. *Comput Struct Biotechnol J.* (2012) 1:e201204005. doi: 10.5936/csbj.201204005
20. Evans DG, Wisely J, Clancy T, Laloo F, Wilson M, Johnson R, et al. Longer term effects of the Angelina Jolie effect: increased risk-reducing mastectomy rates in BRCA carriers and other high-risk women. *Breast Cancer Res.* (2015) 17:143. doi: 10.1186/s13058-015-0650-8
21. Afghahi A, Kurian AW. The changing landscape of genetic testing for inherited breast cancer predisposition. *Curr Treat Options Oncol.* (2017) 18:27. doi: 10.1007/s11864-017-0468-y
22. LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, et al. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med.* (2014) 16:830–7. doi: 10.1038/gim.2014.40
23. Pal T, Cragun D, Lewis C, Doty A, Rodriguez M, Radford C, et al. A statewide survey of practitioners to assess knowledge and clinical practices regarding hereditary breast and ovarian cancer. *Genet Test Mol Biomarkers* (2013) 17:367–75. doi: 10.1089/gtmb.2012.0381
24. Garcia C, Lyon L, Littell RD, Powell CB. Comparison of risk management strategies between women testing positive for a BRCA variant of unknown significance and women with known BRCA deleterious mutations. *Genet Med.* (2014) 16:896–902. doi: 10.1038/gim.2014.48
25. Kurian AW, Li Y, Hamilton AS, Ward KC, Hawley ST, Morrow M, et al. Gaps in incorporating germline genetic testing into treatment decision-making for early-stage breast cancer. *J Clin Oncol.* (2017) 35:2232–9. doi: 10.1200/JCO.2016.71.6480
26. Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined *in silico/in vitro* studies on BRCA1 and BRCA2 variants. *Hum Mutat.* (2012) 33:1228–38. doi: 10.1002/humu.22101
27. Acedo A, Hernandez-Moro C, Curiel-Garcia A, Diez-Gomez B, Velasco EA. Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat.* (2015) 36:210–21. doi: 10.1002/humu.22725
28. Fraile-Bethencourt E, Diez-Gomez B, Velasquez-Zapata V, Acedo A, Sanz DJ, Velasco EA. Functional classification of DNA variants by hybrid minigenes: identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLoS Genet.* (2017) 13:e1006691. doi: 10.1371/journal.pgen.1006691
29. de la Hoya M, Soukarieh O, Lopez-Perolio I, Vega A, Walker LC, van Ierland Y, et al. Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum Mol Genet.* (2016) 25:2256–68. doi: 10.1093/hmg/ddw094
30. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* (2016) 17:333–51. doi: 10.1038/nrg.2016.49
31. Caspar SM, Dubacher N, Kopps AM, Meienberg J, Henggeler C, Matyas G. Clinical sequencing: from raw data to diagnosis with lifetime value. *Clin Genet.* (2018) 93:508–19. doi: 10.1111/cge.13190

32. Colombo M, Lopez-Perolio I, Meeks HD, Caleca L, Parsons MT, Li H, et al. The BRCA2 c.68-7T > A variant is not pathogenic: a model for clinical calibration of spliceogenicity. *Hum Mutat.* (2018) 39:729–41. doi: 10.1002/humu.23411
33. Costa V, Aprile M, Esposito R, Ciccodicola A. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *Eur J Hum Genet.* (2013) 21:134–42. doi: 10.1038/ejhg.2012.129
34. Park E, Pan Z, Zhang Z, Lin L, Xing Y. The expanding landscape of alternative splicing variation in human populations. *Am J Hum Genet.* (2018) 102:11–26. doi: 10.1016/j.ajhg.2017.11.002
35. Bunnell AE, Garby CA, Pearson EJ, Walker SA, Panos LE, Blum JL. The clinical utility of next generation sequencing results in a community-based hereditary cancer risk program. *J Genet Couns.* (2017) 26:105–12. doi: 10.1007/s10897-016-9985-2

**Conflict of Interest Statement:** All authors, with the exception of SN and DT, were employees of Ambry Genetics when they were engaged with this project.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Farber-Katz, Hsuan, Wu, Landrith, Vuong, Xu, Li, Hoo, Lam, Nashed, Toppmeyer, Gray, Haynes, Lu, Elliott, Tippin Davis and Karam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.