HMG Advance Access published February 25, 2014

Comprehensive annotation of splice junctions supports pervasive alternative splicing at the *BRCA1* locus: a report from the ENIGMA consortium

Mara Colombo[1,#], Marinus J. Blok[2,#], Phillip Whiley[3,4], Marta Santamariña[5], Sara Gutiérrez-Enríquez[6], Atocha Romero[7], Pilar Garre[7], Alexandra Becker[8], Lindsay Denise Smith[9], Giovanna De Vecchi[1], Rita D. Brandão[2], Demis Tserpelis[2], Melissa Brown[4], Ana Blanco[10], Sandra Bonache[6,11], Mireia Menéndez[12], Claude Houdayer[13], Claudia Foglia[1], James D. Fackenthal[14], Diana Baralle[9], Barbara Wappenschmidt[8], kConFaB[15], Eduardo Díaz-Rubio[7,16], Trinidad Caldés[7], Logan Walker[17], Orland Díez[6,11,18], Ana Vega[10], Amanda B. Spurdle[3], Paolo Radice[1], Miguel de la Hoya[7,*]

[1]Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

[2]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands

[3]Molecular Cancer Epidemiology Laboratory, Genetics and Computational Division, QIMR Berghofer Medical Research Institute, Brisbane, Australia

[4]School of Chemistry and Molecular Biosciences,The University of Queensland, Brisbane, Australia

[5]Grupo de Medicina Xenómica-USC, Universidad de Santiago de Compostela, CIBERER, IDIS, Santiago de Compostela, Spain

[6]Oncogenetics Group, Vall d'Hebron Institute of Oncology (VHIO), Universitat Autonoma de Barcelona, Barcelona, Spain

[7]Laboratorio de Oncología Molecular. Instituto de Investigación Sanitaria San Carlos (IdISSC). Hospital Clínico San Carlos. Madrid. Spain

[8]Center of Familial Breast and Ovarian Cancer, University Hospital Cologne, Cologne Germany. Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany

[9]Human Development and Health Academic Unit, Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton, UK

[10]Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica-USC, CIBERER, IDIS, Santiago de Compostela, Spain

[11]Oncogenetics Group, Vall d'Hebron Research Institute (VHIR) and Universitat Autonoma de Barcelona, Barcelona, Spain

[12]Genetic Diagnosis Unit, Hereditary Cancer Program, Institut Català d'Oncologia, Barcelona, Spain

[13]Service de Génétique and INSERM U830, Institut Curie and Université Paris Descartes, Sorbonne Paris Cité, Paris, France

[14]Department of Medicine, The University of Chicago Medical Center, Chicago, Illinois, USA

[15]Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria, Australia

[16]Servicio de Oncología Médica. Hospital Clínico San Carlos. Madrid. Spain

[17]Department of Pathology, University of Otago, Christchurch, New Zealand

[18]Oncogenetics Group, University Hospital of Vall d'Hebron, Barcelona, Spain

[*]Correspondence to: Miguel de la Hoya, Laboratorio de Oncología Molecular. Instituto de Investigación Sanitaria San Carlos (IdISSC). Hospital Clínico San Carlos. c/Martín Lagos s/n, Madrid 28040, Spain. Telephone: +34 913303348. Fax: +34913303544. mhoya@hotmail.com

[#]co-first authors.

Abstract

Loss-of-function germ-line mutations in *BRCA1* (MIM #113705) confer markedly increased risk of breast and ovarian cancer. The full-length transcript codifies for a protein involved in DNA repair pathways and cell-cycle checkpoints. Several *BRCA1* splicing isoforms have been described in public domain databases, but the physiological role (if any) of *BRCA1* alternative splicing remains to be established. An accurate description of *naturally occurring* alternative splicing at this locus is a prerequisite to understand its biological significance. However, a systematic analysis of alternative splicing at the *BRCA1* locus is yet to be conducted.

Here, the Evidence-based Network for the Interpretation of Germ-line Mutant Alleles (ENIGMA) consortium combines RT-PCR, exon scanning, cloning, sequencing, and relative semi-quantification to describe naturally occurring *BRCA1* alternative splicing with unprecedented resolution. The study has been conducted in blood related RNA sources, commonly used for clinical splicing assays, as well as in one healthy breast tissue. We have characterized a total of 63 *BRCA1* alternative splicing events, including 35 novel findings. A minimum of 10 splicing events (Δ1Aq, Δ5, Δ5q, Δ8p, Δ9, Δ(9,10), Δ9_11, Δ11q, Δ13p, and Δ14p) represent a substantial fraction of the full-length expression level (ranging from 5% to 100%). Remarkably, our data indicates that *BRCA1* alternative splicing is similar in blood and breast, a finding supporting the clinical relevance of blood-based in vitro splicing assays.

Overall, our data suggests an alternative splicing model in which most non-mutually exclusive alternative splicing events are randomly combined into individual mRNA molecules to produce hundreds of different *BRCA1* isoforms.

Introduction

Virtually all human multi-exon *loci* are subject to alternative splicing (1–4), a biological process that can produce multiple mature RNA transcripts (RNA isoforms) from a single locus (5, 2). Alternative splicing is believed to occur in all metazoan organisms, but it is more prevalent in vertebrates (in particular birds and mammals), thus suggesting a link with phenotypic complexity (6, 7). However, the adaptive role of this mechanism remains elusive (8), in part because the function of many splicing isoforms is unclear (4). Indeed, many of them lack annotated coding sequences (CDSs) (9), or introduce premature termination codons (PTCs) that are predicted to induce the nonsense-mediated mRNA decay (NMD) pathway (10). In this regard, it has been suggested that alternative splicing not only increases the complexity of transcriptomes and proteomes, but also plays a significant role in gene regulation (8, 11). However, it has also been proposed that many alternative splicing events do not have functional significance at all, but rather represent stochastic noise in the splicing process (12).

The breast cancer predisposing gene *BRCA1* (MIM# 113705) was identified in 1994 by positional cloning in families with multiple cases of breast and ovarian cancer (13). The full-length *BRCA1* transcript includes 23 exons (22 coding exons) that encode a 1863 amino acid protein involved in multiple DNA repair pathways and cell-cycle checkpoint regulation (14). In addition to mRNA aberrations arising as a consequence of pathogenic germline mutations associated with high risk of cancer (15), several *BRCA1* alternative splicing isoforms have been described in the literature (13, 16–27). Some of them, in particular those detected solely in tumor samples and/or cell lines (20, 16, 22), probably represent aberrant by-products of (or somatic alteration contributing to) tumorigenesis. However, others certainly represent *naturally occurring* splicing isoforms, here defined as alternative splicing isoforms produced by wild-type alleles in non-malignant tissues. For instance, Δ(9,10), Δ11q and Δ(9,10,11q) have been described (together with the full-length) as *predominant* isoforms expressed in a wide variety of tissues (16, 28). More recently, six BRCA1 transcripts (Δ5, Δ5q, Δ8p, Δ(9,10), Δ13p, and Δ14p) have been reported to be "consistently found in control samples" (29). The most comprehensive review published so far describes up to 21 *BRCA1* alternative splicing isoforms, albeit not all formally validated as *naturally occurring* isoforms (16).

According to several reports, the relative expression levels of Δ(9,10), Δ11q and Δ(9,10,11q) are tissue specific, cell-cycle regulated, and markedly altered in tumor samples, albeit conflicting results have been published (16). These observations suggest that *BRCA1* alternative splicing could play a role in certain cellular functions and might be involved in carcinogenesis (16). The Δ11q and Δ(9,10,11q) isoforms are remarkable in that they lack more than 50% of the full-length coding sequence, suggesting that the encoded proteins greatly differ from the full-length in their biological activities. Apparently, engineered mice models supported specific roles for Δ11 (mice do not express Δ11q) during early embryogenesis (30–33). However, the later discovery of BRCA1-IRIS (18), a *BRCA1* locus product containing an open reading frame that extends from full-length start codon in exon 2 to the end of exon 11, continuing for 34 more triplets into intron 11 where it terminates, complicated the interpretation of previous data derived from mice models (28). Interestingly, mounting evidence indicates that BRCA1-IRIS, contrary to BRCA1 full-length, has oncogenic-like activity (34). Despite these and other efforts, the physiological and pathological roles (if any) of *BRCA1* alternative splicing remain to be established.

An accurate description of *naturally occurring* alternative splicing at the *BRCA1* locus is a prerequisite to understand its biological significance. In addition, it will become a valuable resource for the design and interpretation of *in vitro* splicing assays conducted to investigate the pathogenicity of germ-line genetic variants (35, 26, 36, 15, 37). To date, a systematic analysis of alternative splicing at the *BRCA1* locus is yet to be conducted. Here we report a collaborative effort of the ENIGMA consortium (38) conducted to comprehensively analyze *BRCA1* alternative splicing events occurring in four blood related RNA sources, commonly used for clinical splicing assays, and one healthy breast tissue.

Results

After conducting a 4-stage project (Fig. 1), we have been able to annotate 63 independent *BRCA1* alternative splicing events (Tables 1 to 4, and Supp. Table 1 online), including 46 fully characterized by sequencing of the spliced junctions and 17 imputed from capillary electrophoresis (CE) alternative splicing models (Supp. Fig. 1 to 3 online). Out of 63 events, 61 were observed in lymphoblastoid cell lines (**LCLs)**, 53 in primary cultures of stimulated peripheral blood leukocytes (**PBLs),** 51 in whole blood leukocytes (**LEUs)**, and 46 in ficoll-isolated peripheral blood mononuclear cell (**PBMCs)** (Supp. Table 2 online). The overlap was significant, with 39 events (62%) detected in all four RNA sources, and 51 events (81%) detected in at least three of them. Most discrepancies were best explained by low *coverage* in low *detection rate* events (see Methods and Supp. Table 2 online for further details).

We have classified these alternative splicing events into six basic structural biotypes: cassettes, multi-cassettes, splice donor shifts, splice acceptor shifts, terminal modifications, and intronizations (Fig. 2). Splice donor shifts include the alternative use of proximal and distal sites at all *BRCA1* tandem acceptor (NAGNAG) sites (39), with the single notable exception of exon 6. The latter is probably explained by the local sequence context, as suggested by dedicated *in silico* analysis (see Supp. Fig. 4 online for further details). Some splicing events are best described as mixed biotypes (Fig. 2 and Table 4). Most annotated events classify into (multi)-cassette biotypes (N=37), including two inclusion events that introduce intron 3 and intron 13 genomic sequences into mature transcripts (Table 1). The former corresponds to the genomic sequence originally reported as *BRCA1* exon 4 (13), and later considered an intronic Alu element (27).

Functional annotation classifies *BRCA1* splicing events into: 23 PTC-NMDs (splicing events introducing Premature Termination Codons predicted to induce the Nonsense-Mediated RNA Decay pathway), 15 in-frames (ranging from subtle effects at NAGNAG sites to large deletions removing more than 50% of the reference coding sequence), 13 Non-Coding (eliminating the full-length start codon), 5 UTRs (splicing events modifying UnTranslated Regions), 4 frame-shifts generating PTCs not predicted to induce NMD (PTCs located in exon 24), and 1 internal PTC with polyadenylation (IRIS) (Tables 1-4).

Splicing *assays* were developed to detect *BRCA1* splicing events, not to address quantitative aspects. Yet, visual inspection allowed us to identify 10 *predominant BRCA1* splicing events: Δ1Aq, Δ5, Δ5q, Δ8p, Δ9, Δ(9,10), Δ9_11, Δ11q, Δ13p, and Δ14p, eight of which were later analyzed (Stage 4) by semi-quantitative CE (not feasible in the case of Δ9_11 and Δ11q, see methods for further details).

We performed a comprehensive screening of *BRCA1* alternative splicing events in one healthy breast tissue sample (**BREAST**). Remarkably, most splicing events previously identified in blood samples (43 out of 63) were detected, despite the lower *coverage*. Those not detected tended to be low *detection rate* events in blood derived samples (see Supp. Table 2 online for further details). Equally relevant, our analysis did not identify any splicing events that had not been detected previously in blood. Visual inspection of saturating PCR assays detected the very same 10 *predominant* splicing events previously identified in blood related samples.

Semi-quantitative CE indicated obvious differences among *predominant* events both in blood and breast tissue (Fig. 3). While some events represented roughly 5% of the full-length signal (Δ5, Δ5q, Δ9, Δ13p), others represented up to 30% (Δ8p, Δ(9,10), Δ14p). Finally, we observed similar levels of Δ1Aq and full-length transcripts. Very similar splicing patterns were observed when analyzing **LEU**, **PBMC**, **PBL**, and **LCL** samples separately (see Supp. Fig. 5 to 12 online). Although semi-quantitative profiling of all *BRCA1* splicing events described here is beyond the scope of the present study, there are some notable observations in relation with Stage 2 *detection rate* (Supp. Table 2 online). The average *detection rate* of Stage 2 splicing events is 46%, but there is a clear distinction between the average *detection rates* of *predominant* versus *minor* events (86% vs. 39%). Furthermore, the *detection* rate reaches 100% in four *predominant* events representing ≥30% of the full-length signal (Δ1Aq, Δ8p, Δ(9,10), Δ14p), but decreases to 72% (range 62% to 87%) for those representing only 5% of the full-length signal (Δ5, Δ5q, Δ9, Δ13p). Taken together, these observations suggest that, in our experimental setting, the *detection rate* of an individual splicing event (Supp. Table 2 online) is related with the actual expression level of that particular event.

Interestingly, CE analysis allowed us to identify peaks imputed to transcripts combining two or more independent splicing events. For instance, 7-11q assays (Supp. Fig. 1A online) demonstrated the existence of RNA species combining Δ8p with Δ9, Δ10, and Δ(9,10) events. Similarly, 12-14 assays (Supp. Fig. 1B online) demonstrated the existence of RNA species containing Δ13, ▼13A, Δ13p, and Δ14p events in almost all possible combinations, with the only exception being that RNA species combining Δ13 with ▼13A were not observed. Further supporting this scenario, the analysis of 7-12 assays revealed a high diversity of transcripts combining Δ8p, Δ9, Δ10, Δ(9,10), Δ11, and Δ11q splicing events (Supp. Fig. 3 online), including the detection of transcripts combining Δ(9,10) with Δ11q. The latter, for the sake of consistence annotated as (Δ9,10+Δ11q) in Supp. Fig. 3 online, but often referred to as Δ(9,10,11q) in the literature, is one of few *BRCA1* splicing isoforms described previously as predominant (16). *Detection rate* of (Δ9,10+Δ11q) reached 100% in Stage 2 (data not shown), further supporting a link between *detection rate* and actual expression level.

Note that, overall, the CE signal corresponding to transcripts containing multiple splicing events is consistently lower than that of the transcripts containing the corresponding individual events, as expected from a random combination of independent elements (see several examples in Supp. Fig. 1 and 3 online).

Discussion

We have combined RT-PCR, CE, cloning and conventional sequencing to describe naturally occurring *BRCA1* alternative splicing with unprecedented resolution. To our knowledge, 34 out of the 63 splicing events reported here are novel findings; while only 22 events have been described previously in blood samples (see Tables 1-4 for further details). However, we have not been able to validate up to 8 *BRCA1* splicing events previously reported by others (including 2 Stage 1 events reported by contributors of the present manuscript). While it is likely that some of these events do not qualify for *naturally occurring* events, the data suggests that characterizing the full complexity of *BRCA1* splicing will require further studies (see supplemental Table 3 online for further details). This is also suggested by the fact that we have identified several signals compatible with additional splicing events that, nonetheless, we have not been able to annotate (see Supp. Fig. 1 and 3 online).

Overall, our data indicates that most naturally occurring *BRCA1* splicing events are rather minor if compared with the full-length signal. However, we have identified 10 *predominant* splicing events that appear to represent a substantial fraction of the full-length expression using semi-quantitative measures. Not surprisingly, all 10 *predominant* events have been described previously. Indeed, six of them (Δ5, Δ5q, Δ8p, Δ(9,10), Δ13p, and Δ14p) have been described recently as *BRCA1* splicing events "consistently found in control samples" (29).

Genome-wide analyses suggest that cassette events (30-50% of all splicing events) are the commonest alternative splicing structural biotypes observed in mammals (5, 4, 10). In this regard, the human *BRCA1* gene can be described as typical, since 37 out of the 63 (58%) splicing events here reported are cassette-like. Remarkably, all *BRCA1* internal exons are involved in one or more cassette events so that, formally speaking, *BRCA1* lacks constitutive exons. Yet, with few exceptions (Δ5, Δ9, Δ(9,10), Δ9_11) cassette events are rather *minor events*, so that most internal exons are best described as "quasi-constitutive exons". The number of splice site shifts (N=9) is much lower, but 6 out of 10 *predominant* events correspond to this biotype. Remarkably, we have not identified structural biotypes such as mutually exclusive cassette exons, or retained introns (5, 10). Yet, we have identified two intronization events (Fig. 2). This would appear to be a rare structural biotype, since it is not included in a comprehensive catalogue of structural biotypes (68 different biotypes) identified in human, mouse and several non-mammal vertebrates (4). Perhaps, intronization events occurring in vertebrates are associated with exceptionally long exons, such as human *BRCA1* exon 11 (3426bp vs. an average exon length of approximately 180bp in the human genome).

The spectrum of possible splicing events occurring at a single locus is so wide that any attempt to catalogue it will inevitably be biased by the analytical approach employed. In our experience, CE analysis of RT-PCR products is very sensitive for detecting minor events, subtle size-effects, and multi-cassette events (37). However, minor events involving long (≥1000bp) intron retentions (if any) will usually escape detection, as expected product sizes are out of range of CE. Furthermore, our analytical approach does not allow discovery of novel terminal events which are not formally splicing events, but are nonetheless reported as major contributors to exon variability in mRNAs (5, 3). In this regard, we have limited our study to analyze the expression of two previously reported terminal events (Exon1B and IRIS) by dedicated assays.

We have shown previously that semi-quantitative CE is able to detect splicing quantitative trait loci (sQTLs) such as rs1799965, a rare SNP (MAF <0.001) that is associated with an increase in expression of *BRCA1* Δ9,10 (26), and rs9534262, a common SNP (MAF>0.40) associated with an increase in *BRCA2* Δ17,18 (40). In the present study we have not been able to detect inter-individual variability, suggesting that this is below technical replica variability. Since we have analyzed alternative splicing in a relative large number of control samples (N=48), our data suggests that common sQTLs at the *BRCA1* gene (if any) are likely to induce more subtle effects than those reported for rs1799965 (c.591C>T) and rs9534262 (c.7806-14T>C).

It is important to point out that we have produced a catalogue of alternative splicing events, which may not represent a catalogue of RNA isoforms (we have not cloned individual mRNAs). At present, we cannot rule out the possibility that certain splicing events tend to occur together (linked splicing model), so that the actual number of *BRCA1* mRNA isoforms might be lower than the number of splicing events reported here. Yet, overall our data favors an unlinked splicing model in which most, if not all, non-mutually exclusive alternative splicing events are randomly combined into individual mRNA molecules to potentially produce hundreds of different *BRCA1* isoforms. According to GENCODE v7, human protein coding loci express on average 6.31 alternatively spliced transcripts (10), and those loci with more than 20 annotated isoforms are very rare (3). Therefore, our analysis raises the interesting possibility of *BRCA1* being a locus with particularly high levels of alternative splicing. However, global estimations of alternative splicing levels are based on genome-wide RNA-seq efforts that may underestimate the true level of alternative splicing. At least, this is supported by targeted RNA-seq experiments that identify hundreds of previously un-annotated isoforms in even extensively studied protein-coding loci such as *TP53* and *HOX* (41). If proven true, the finding that *BRCA1* is a locus with a high level of alternative splicing would be consistent with recent genome-wide analyses connecting high level alternative splicing loci with intrinsically disordered proteins/domains (IDPs/IDDs), IDPs/IDDs with Hub proteins, and Hub proteins with disease (42–45). *TP53* represents a paradigm of this association (43, 41, 44). Similar to *TP53*, *BRCA1* is a disease associated genetic locus coding for a IDP/IDD protein with Hub properties (43). Accordingly, high level alternative splicing would indeed be an expected feature of the *BRCA1* locus. Remarkably, CE splicing analysis at the *BRCA2* locus (a gene fairly similar to *BRCA1* in terms of overall size and exon/intron structure, but coding for a protein that lacks IDDs and/or Hub features) reveals a much lower extent of alternative splicing (ENIGMA manuscript in preparation).

Regardless of its biological significance, we believe that the comprehensive description of *BRCA1* alternative splicing reported here will be highly relevant for diagnosis, in particular when assessing the impact of *BRCA1* germ-line variants on splicing. Recently, the ENIGMA consortium conducted a multicentre investigation aimed at comparing *in vitro* splicing assay protocols and elaborating best practice guidelines (37). The study addressed analytical aspects such as primers design, reverse transcriptase protocols, NMD inhibition, and detection methods, and identified primers design (positioning primers) as a major source of variability across laboratories. The study concluded that *a prior* knowledge of the expected transcripts (naturally occurring alternative splicing isoforms) was a key factor for proper primers design and clinical assessment (37). Previous studies have identified as well alternative splicing as a critical aspect to be considered in the design and analysis of *BRCA1 in vitro* splicing assays (26). In this regard, the catalogue of splicing events here identified will be a valuable tool to improve the design (primers can be strategically positioned to include or exclude specific splicing events in function of the position of the variants under scrutiny) and analysis (at both qualitative and quantitative level) of future *BRCA1 in vitro* splicing assays, thus improving the clinical interpretation of the outcomes. In turn, this will facilitate the integration of *BRCA1 in vitro* splicing assays into the multifactorial likelihood models that are developed by the ENIGMA consortium to assess the clinical relevance of genetic variants (38).

Despite its comprehensiveness, the above mentioned ENIGMA study comparing *in vitro* splicing assays protocols was conducted in RNAs isolated from LCLs, so that did not evaluate the impact of using other blood related RNA sources. Yet, LEUs or PBLs are common RNA sources for *in vitro* splicing assays in genetic testing laboratories worldwide (15). In the present study we have shown that *BRCA1* alternative splicing is similar in four different blood RNA sources (LEUs, PBLs, PBMCs, and LCLs), suggesting that the actual blood related RNA source used for assessing the role of *BRCA1* germ-line variants on splicing is unlikely to represent a major contributor to variability of results. Further on, our data suggests that *BRCA1* alternative splicing is similar in blood and breast tissues, supporting that *in vitro* splicing assays performed in blood are relevant for diagnosis.

Although the biological relevance of *BRCA1* alternative splicing is largely unknown, the precise knowledge of the different splicing events will be instrumental for the definition of its functional (and clinical) relevance. Individual *BRCA1* mRNA isoforms can be monitored more closely in future splicing assays (preferably including accurate quantification), and the functional relevance of their putatively encoded proteins can be further evaluated by *in vitro* transfection of the corresponding cDNA constructs to rescue gene expression, as recently shown for *BRCA1* missense variants (53), and two *BRCA1* alternative splicing isoforms (17, 54).

Finally, we believe that CE scanning, as here conducted for *BRCA1* analysis, is a feasible approach to develop accurate catalogues of locus specific alternative splicing events that can assist the analysis and validation of data from targeted RNAseq experiments.

Materials and Methods

Samples

We have analyzed *BRCA1* alternative splicing in RNA samples from healthy control individuals. RNA was isolated from whole blood leukocytes (**LEUs**), ficoll-isolated peripheral blood mononuclear cells (**PBMCs**), primary cultures of stimulated peripheral blood leukocytes (**PBLs**), and lymphoblastoid cell lines (**LCLs**). In addition, RNA was isolated from an epithelial enriched area of one healthy breast tissue obtained after cosmetic surgery (**BREAST**). In Stage 1 (see workflow in Fig. 1) different contributing laboratories used different isolation protocols and/or cDNA synthesis strategies, as described in a recent ENIGMA paper (37). A full description of RNA isolation and cDNA synthesis protocols used in Stages 2 and 3 (see workflow in Fig. 1), is provided in supplemental methods online. The study was approved by the Institutional Review Board of each participating center.

Identification, Validation, and Relative Quantification of *naturally occurring BRCA1* alternative splicing events For the purpose of this study, we define alternative splicing events as those incorporating splice junctions not present in the reference transcript Ensemble ENST00000357654 (hereafter referred as full-length transcript). The only exception is BRCA1-IRIS (see introduction), a locus product for which no specific splice junction exists (18). Multiple combinations of forward and reverse primers located at exonic regions (as defined by the full-length transcript) were used to amplify cDNAs. A PCR performed with a particular combination of primers will be referred throughout the text as a *BRCA1* splicing *assay*. We conducted a 4-stage project as follows (see workflow in Fig. 1):

In Stage 1, contributing centers used their own control samples (blood related) and *assays* to identify alternative splicing events at the *BRCA1* gene. All Stage 1 primers are available upon request. At this stage, splicing *assays* were analyzed by Et-Br stained agarose gel electrophoresis, capillary electrophoresis (CE) and/or direct sequencing, depending on the contributing center. Both confirmed (sequenced) and predicted (size-matching) events were considered. In addition, we performed a comprehensive review of the literature in order to identify all *BRCA1* splicing events previously described, including *naturally occurring* events, but also splicing events not formally validated as such (like those solely detected in tumor samples and/or cell lines). Stage 1 experimental and review data were pooled together to elaborate a working list of 42 *BRCA1* alternative splicing events (see Supp. Table 1 online).

In Stage 2, stage 1 information was used to develop a panel of 18 overlapping *assays* (all primer sequences are provided as Supplemental Methods) that allowed a comprehensive scanning of *BRCA1* splicing events by CE (see Supp. Fig. 1 to 3, and Supp. Table 1 online). Thermal cycling consisted of an initial 10-min hold at 95ºC, followed by 30-second hold at 95ºC, 30-second hold at 58 ºC, and 30-second hold at 72ºC (increased to 2 min for exon11 containing assays) for 45 cycles to maximize sensitivity. Stage 2 screening was performed in 48 healthy control samples of European ancestry, including 10 **LEUs**, 8 **PBMCs**, 20 **PBLs**, and 10 **LCLs**. Several centers contributed samples at this stage, but actual screening was centralized in one laboratory. CE analyses were performed in a 3130 Genetic Analyzer (Applied Biosystems) with GeneScan 500/1200 Size Standards (Applied Biosystems) as internal markers. Size-calling was performed with GeneMapper v4.0 Software (Applied Biosystems). Some splicing events were captured by one *assay*, while others were captured by two or more overlapping *assays* (Supp. Table 1 online). Stage 2 centralized screening involved a total of 4281 CE data-points (one data-point defined as each technical replica of an individual splicing event assayed in one sample). *Coverage* (defined here as data-points *per* splicing event) ranged from 18X to 163X (67X on average). *Detection Rate* (% of positive data-points) ranged from 3% to 100%. By far, the highest *coverage* was obtained in **PBLs** samples, with 2055 data-points (see Supp. Table 2 online for further details). None of the *assays* listed in Supp. Table 1 online allowed BRCA1-IRIS detection. For that purpose, we developed a dedicated assay that does not rely on CE analysis (see supplemental methods online). Stage 2 allowed us to validate 34 out of 42 Stage 1 events in a cohort of control samples, but also to validate 29 additional splicing events. For the purpose of this study, we have validated findings only if sequenced, or imputed by two or more contributing centers with different primer sets.

Visual inspection of CE assays (or Ethidium Bromide agarose stained gels in the case of exon 11 containing assays) revealed that most splicing events were easily classified into two categories according to their signal relative to the full-length transcript, hereafter refereed as *predominant* and *minor* events. BRCA1-IRIS and exon1B transcripts were not classifiable because the full-length reference transcript was not co-amplified in the corresponding assays. Later, splicing events classified as *predominant* were further characterized by semi-quantitative CE assays (see below).

In Stage 3, screening of *BRCA1* splicing events was performed in one normal breast sample (**BREAST**). *Assays* and CE protocols were as in Stage 2, although *coverage* was much lower (see Supp. Table 2 online).

Finally, in Stage 4 we investigated the expression level relative to the full-length in 8 alternative splicing events previously annotated as *predominant*. With this aim, **LEUs**, **PBMCs**, **PBLs**, **LCLs**, and **BREAST** samples were reanalyzed with four splicing *assays* (E1-E6, E7-E11q, E12-E13, and E12-E14) performed in semi-quantitative (33 PCR cycles) conditions (semi-quantitative CE). Relative quantification of individual splicing events was expressed as the average ratio between the peak area of that particular event and the peak area of the full-length signal (Fig. 3 and Supp. Fig. 5 to 12 online). Semi-quantitative CE analysis of two *predominant* events (Δ9_11, and Δ11q) was not feasible because of the large size difference (>3300bp) between spliced and full-length products.

Splicing Events Designation We have designated *BRCA1* exons following the Breast Core Informative (BIC) database nomenclature (46), so that the 22 coding exons of the reference full-length transcript are numbered from 2 to 24 with no exon 4 defined (13). We have designated splicing events combining the following symbols: Δ (skipping), ▼ (retention), p (proximal), and q (distal). In addition, we have also used non-systematic designations previously established in the scientific literature, including IRIS, exon1A, exon1aA, exon1B, exon4, and exon13A (13, 18, 19, 22, 23).

Splice junction sequencing Depending on the particular splicing event investigated (and/or contributing center), different approaches were followed. Direct sequencing of individual *assays* (sometimes with internal primers at selected locations) allowed us to sequence 25 events; including *predominant* and *minor* events (see Tables 1-4). The latter was possible thanks to stochastic preferential amplification of *minor* events observed in 45-cycle RT-PCR assays (an illustrative example is shown in

Supp. Fig. 2 online). Alternatively, agarose or polyacrylamide gel excised splicing assay products were cloned into the pGEM-T vector (Promega) and sequenced. Cloning allowed us to sequence 32 events (see Tables 1-4). All sequence reactions were performed using the ABI PRISM® BigDye™ Terminator Cycle Sequencing kit (Applied Biosystems) and examined with an ABI 3130 Genetic Analyzer (Applied Biosystems), using the Sequencing Analysis software (Applied Biosystems).

Imputation of splicing events For this purpose, we elaborated alternative splicing models that best explained the peak pattern observed in CE analyses. Imputations were performed combining CE size-calling data, sequencing findings, and GENCODE annotations retrieved through the Ensemble Genome Browser (http://www.ensembl.org). As a rule, we imputed splicing events only if compatible with the use of canonical splice sites (GT-AG) present in the *BRCA1* reference genomic sequence GRCh37:17:41195712:41322890:-1 (http://www.ensembl.org). The approach allowed us to annotate *BRCA1* splicing events not supported by direct sequencing evidence (referred throughout the text as imputed events). Imputation was also used to deduce the existence of transcripts combining multiple splicing events. Two representative examples of *BRCA1* splicing models are shown in Supp. Fig. 1 online.

Structural and Functional Annotation of Alternative Splicing Events Structural and functional annotation has been performed as in Mudge and co-workers (4), although we incorporated an additional structural biotype referred throughout the text as intronization. First described in *Caenorhabditis* species (47), intronization refers to the conversion of a single exon into two exons and one intervening intron (see Fig. 2). Functional annotation of *BRCA1* splicing events includes *Non-Coding* (splicing events eliminating the full-length start codon), *PTC-NMDs* (splicing events introducing Premature Termination Codons predicted to induce the Nonsense-Mediated RNA Decay pathway), *No-FS* (in-frame splicing events), *FS-alternative STOP* (frame-shift events generating PTCs not predicted to induce NMD as they are located in the most downstream *BRCA1* exon), *UTRs* (splicing events modifying UnTranslated Regions), and one internal PTC with polyadenylation (IRIS).

Identification of *BRCA1* alternative splicing events in public domain databases Studies published up to June 2013 that contained data from *BRCA1* splicing assays were identified by carrying out literature searches using the LOVD database (http://chromium.liacs.nl/LOVD2/cancer/home), PubMed (http://www.ncbi.nlm.nih.gov/pubmed), and Google Scholar (http://scholar.google.com), using the following keywords: BRCA1, BRCA2, splicing. Each report was reviewed in detail to extract the following data: splicing events detected (excluding those directly attributed to germ-line pathogenic mutations), and RNA source. The data, together with information retrieved from Ensembl (BRCA1 transcripts) has been incorporated into Tables 1-4 and Supp. Tables 1 and 3 online.

Acknowledgements

This work was supported by the Spanish Instituto de Salud Carlos III research grants [grant numbers PI10/01422, PI13/00285, CP10/00617 to SGE, PI12/02585 to OD, PI12/00539 to MH]; the Spanish Instituto de Salud Carlos III/Red Tematica de Investigación Cooperativa en Cáncer [research grants RD06/0020/1051 and RD12/0036/008]; the Italian Association for Cancer Research [grant number 11897 to PR]; the United States Department of Defense Idea Award [grant number BC061352 to JDF]; the Australian National Health and Medical Research Council [grant number 1010719 to ABS]; the National Breast Cancer Foundation and Cancer Australia [grant number 628333 to KConFab] and funds from The University of Queensland. DB and LD were both funded by CRUK. DB is hefce senior fellow. PW has an honorary apartment at UQ. LCW is funded by a HRC Sir Charles Hercus Health Research Fellowship. SGE is funded by a Miguel Servet contract (CP10/00617). Spanish Instituto de Salud Carlos III research grants and Red Tematica de Investigación Cooperativa en Cáncer are initiatives of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds.

We wish to thank the following personnel of the Fondazione IRCCS Istituto Nazionale dei Tumori of Milano: Ferdando Ravagnani for providing biological samples, Donata Penso and Maria Teresa Radice for technical assistance. We also wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up

Study for their contributions to this resource, and the many families who contribute to kConFab. We wish to thank all the member of the ICO Hereditary Cancer Program. We thank Anna Tenés and Paula Diaque for technical support.

Conflict of Interest Statement

The authors have declared no conflicts of interest.

References

1. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.

2. Chen, M. and Manley, J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.*, **10**, 741–754.

3. Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin,W., Schlesinger, F., *et al*. (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.

4. Mudge, J.M., Frankish, A., Fernandez-Banet, J., Alioto, T., Derrien, T., Howald, C., Reymond, A., Guigó, R., Hubbard, T. and Harrow, J. (2011) The origins, evolution, and functional potential of alternative splicing in vertebrates. *Mol. Biol. Evol.*, **28**, 2949–2959.

5. Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell*, **126**, 37–47.

6. Kim, H., Klein, R., Majewski, J. and Ott, J. (2004) Estimating rates of alternative splicing in mammals and invertebrates. *Nat. Genet.*, **36**, 915–916.

7. Kim, E., Magen, A. and Ast, G. (2007) Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.*, **35**, 125–131.

8. Skandalis, A., Frampton, M., Seger, J. and Richards, M.H. (2010) The adaptive significance of unproductive alternative splicing in primates. *RNA*, **16**, 2014–2022.

9. Tress, M.L., Martelli, P.L., Frankish, A., Reeves, G.A., Wesselink, J.J., Yeats, C., Olason, P.I., Albrecht, M., Hegyi, H., Giorgetti, A., *et al*. (2007) The implications of alternative splicing in the ENCODE protein complement. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 5495–5500.

10. Frankish, A., Mudge, J.M., Thomas, M. and Harrow, J. (2012) The importance of identifying alternative splicing in vertebrate genome annotation. *Database (Oxford)*, **2012**, bas014.

11. Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 189–192.

12. Melamud, E. and Moult, J. (2009) Stochastic noise in splicing machinery. *Nucl. Acids Res.*, **37**, 4873–4886.

13. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M. and Ding, W. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, **266**, 66–71.

14. Roy, R., Chun, J. and Powell, S.N. (2012) BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer*, **12**, 68–78.

15. Walker, L.C., Whiley, P.J., Houdayer, C., Hansen, T.V.O., Vega, A., Santamarina, M., Blanco, A., Fachal, L., Southey, M.C., Lafferty, A., *et al*. (2013) Evaluation of a 5-Tier Scheme Proposed for Classification of Sequence Variants Using Bioinformatic and Splicing Assay Data: Inter-Reviewer Variability and Promotion of Minimum Reporting Guidelines. *Hum. Mutat.*, 34, 1424-1431.

16. Orban, T.I. and Olah, E. (2003) Emerging roles of BRCA1 alternative splicing. *MP, Mol. Pathol.*, **56**, 191–197.

17. Sevcik, J., Falk, M., Kleiblova, P., Lhota, F., Stefancikova, L., Janatova, M., Weiterova, L., Lukasova, E., Kozubek, S., Pohlreich, P., *et al*. (2012) The BRCA1 alternative splicing variant Δ14-15 with an in-frame deletion of part of the regulatory serine-containing domain (SCD) impairs the DNA repair capacity in MCF-7 cells. *Cell. Signal.*, **24**, 1023–1030.

18. ElShamy, W.M. and Livingston, D.M. (2004) Identification of BRCA1-IRIS, a BRCA1 locus product. *Nat. Cell Biol.*, **6**, 954–967.

19. Fortin, J., Moisan, A.M., Dumont, M., Leblanc, G., Labrie, Y., Durocher, F., Bessette, P., Bridge, P., Chiquette, J., Laframboise, R., *et al*. (2005) A new alternative splice variant of BRCA1 containing an additional in-frame exon. *Biochim. Biophys. Acta*, **1731**, 57–65.

20. Lixia, M., Zhijian, C., Chao, S., Chaojiang, G. and Congyi, Z. (2007) Alternative splicing of breast cancer associated gene BRCA1 from breast cancer cell line. *J. Biochem. Mol. Biol.*, **40**, 15–21.

21. Fetzer, S., Tworek, H.A., Piver, M.S. and Dicioccio, R.A. (1998) An alternative splice site junction in exon 1a of the BRCA1 gene. *Cancer Genet. Cytogenet.*, **105**, 90–92.

22. Gu, M., Li, H., Shen, C., Wu, L., Liu, W., Miao, L. and Zheng, C. (2010) Cloning and characterization of a new BRCA1 variant: A role for BRCT domains in apoptosis. *Cancer Letters*, **295**, 205–213.

23. Xu, C.F., Brown, M.A., Chambers, J.A., Griffiths, B., Nicolai, H. and Solomon, E. (1995) Distinct transcription start sites generate two forms of BRCA1 mRNA. *Hum. Mol. Genet.*, **4**, 2259–2264.

24. Xu, C.F., Chambers, J.A., Nicolai, H., Brown, M.A., Hujeirat, Y., Mohammed, S., Hodgson, S., Kelsell, D.P., Spurr, N.K., Bishop, D.T., *et al*. (1997) Mutations and alternative splicing of the BRCA1 gene in UK breast/ovarian cancer families. *Genes Chromosomes Cancer*, **18**, 102–110.

25. Claes, K., Vandesompele, J., Poppe, B., Dahan, K., Coene, I., De Paepe, A. and Messiaen, L. (2002) Pathological splice mutations outside the invariant AG/GT splice sites of BRCA1 exon 5 increase alternative transcript levels in the 5' end of the BRCA1 gene. *Oncogene*, **21**, 4171–4175.

26. Dosil, V., Tosar, A., Cañadas, C., Pérez-Segura, P., Díaz-Rubio, E., Caldés, T. and de la Hoya, M. (2010) Alternative splicing and molecular characterization of splice site variants: BRCA1 c.591C>T as a case study. *Clin. Chem.*, **56**, 53–61.

27. Friedman, L.S., Szabo, C.I., Ostermeyer, E.A., Dowd, P., Butler, L., Park, T., Lee, M.K., Goode, E.L., Rowell, S.E. and King, M.C. (1995) Novel inherited mutations and variable expressivity of BRCA1 alleles, including the founder mutation 185delAG in Ashkenazi Jewish families. *Am. J. Hum. Genet.*, **57**, 1284–1297.

28. Tammaro, C., Raponi, M., Wilson, D.I. and Baralle, D. (2012) BRCA1 exon 11 alternative splicing, multiple functions and the association with cancer. *Biochemical Society Transactions*, **40**, 768–772.

29. Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., Bronner, M., Buisson, M., Coulet, F., Gaildrat, P., *et al*. (2012) Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum. Mutat.*, **33**, 1228–1238.

30. Gowen, L.C., Johnson, B.L., Latour, A.M., Sulik, K.K. and Koller, B.H. (1996) Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. *Nat. Genet.*, **12**, 191–194.

31. Hakem, R., de la Pompa, J.L., Sirard, C., Mo, R., Woo, M., Hakem, A., Wakeham, A., Potter, J., Reitmair, A., Billia, F., *et al*. (1996) The tumor suppressor gene Brca1 is required for embryonic cellular proliferation in the mouse. *Cell*, **85**, 1009–1023.

32. Ludwig, T., Chapman, D.L., Papaioannou, V.E. and Efstratiadis, A. (1997) Targeted mutations of breast cancer susceptibility gene homologs in mice: lethal phenotypes of Brca1, Brca2, Brca1/Brca2, Brca1/p53, and Brca2/p53 nullizygous embryos. *Genes Dev.*, **11**, 1226–1241.

33. Hakem, R., de la Pompa, J.L., Elia, A., Potter, J. and Mak, T.W. (1997) Partial rescue of Brca1 (5-6) early embryonic lethality by p53 or p21 null mutation. *Nat. Genet.*, **16**, 298–302.

34. Shimizu, Y., Luk, H., Horio, D., Miron, P., Griswold, M., Iglehart, D., Hernandez, B., Killeen, J. and ElShamy, W.M. (2012) BRCA1-IRIS overexpression promotes formation of aggressive breast cancers. *PLoS ONE*, **7**, e34102.

35. Claes, K., Poppe, B., Machackova, E., Coene, I., Foretova, L., De Paepe, A. and Messiaen, L. (2003) Differentiating pathogenic mutations from polymorphic alterations in the splice sites of BRCA1 and BRCA2. *Genes Chromosomes Cancer*, **37**, 314–320.

36. Spurdle, A.B., Couch, F.J., Hogervorst, F.B.L., Radice, P., Sinilnikova, O.M. and IARC Unclassified Genetic Variants Working Group (2008) Prediction and assessment of splicing alterations: implications for clinical testing. *Hum. Mutat.*, **29**, 1304–1313.

37. Whiley, P., de la Hoya, M., Thomassen, M., Becker, A., Brandão, R., Pedersen, I.S., Montagna, M., Menéndez, M., Quiles, F., Enríquez, S.G., et al. (2014) Comparison of mRNA Splicing Assay Protocols across Multiple Laboratories: Recommendations for Best Practice in Standardized Clinical Testing. *Clin. Chem.*, **60**, 341-352.

38. Spurdle, A.B., Healey, S., Devereau, A., Hogervorst, F.B.L., Monteiro, A.N.A., Nathanson, K.L., Radice, P., Stoppa-Lyonnet, D., Tavtigian, S., Wappenschmidt, B., *et al*. (2012) 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.*, **33**, 2–7.

39. Hiller, M., Huse, K., Szafranski, K., Jahn, N., Hampe, J., Schreiber, S., Backofen, R. and Platzer, M. (2004) Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat. Genet.*, **36**, 1255–1257.

40. De Garibay, G.R., Acedo, A., García-Casado, Z., Gutiérrez-Enríquez, S., Tosar, A., Romero, A., Garre, P., Llort, G., Thomassen, M., Díez, O., *et al*. (2014) Capillary Electrophoresis Analysis of Conventional Splicing Assays: IARC Analytical and Clinical Classification of 31 BRCA2 Genetic Variants. *Hum. Mutat.*, **35**, 53-57.

41. Mercer, T.R., Gerhardt, D.J., Dinger, M.E., Crawford, J., Trapnell, C., Jeddeloh, J.A., Mattick, J.S. and Rinn, J.L. (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.*, **30**, 99–104.

42. Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A. and Babu, M.M. (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell*, **46**, 871–883.

43. Cortese, M.S., Uversky, V.N. and Dunker, A.K. (2008) Intrinsic disorder in scaffold proteins: getting more from less. *Prog. Biophys. Mol. Biol.*, **98**, 85–106.

44. Uversky, V.N., Oldfield, C.J., Midic, U., Xie, H., Xue, B., Vucetic, S., Iakoucheva, L.M., Obradovic, Z. and Dunker, A.K. (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*, **10 Suppl 1**, S7.

45. Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E. and Muñoz, M.J. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.*, **14**, 153–165.

46. Szabo, C., Masiello, A., Ryan, J.F. and Brody, L.C. (2000) The breast cancer information core: database design, structure, and scope. *Hum. Mutat.*, **16**, 123–131.

47. Irimia, M., Rukov, J.L., Penny, D., Vinther, J., Garcia-Fernandez,J . and Roy, S.W. (2008) Origin of introns by 'intronization' of exonic sequences. *Trends Genet.*, **24**, 378–381.

48. Munnes, M., Zuther, I., Schmitz, B. and Doerfler, W. (2000) A Novel Insertional Mutation and Differentially Spliced mRNAs in the Human BRCA1 Gene. *Gene Function & Disease*, **1**, 38–47.

49. Bonnet, C., Krieger, S., Vezain, M., Rousselin, A., Tournier, I., Martins, A., Berthet, P., Chevrier, A., Dugast, C., Layet, V., *et al*. (2008) 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.

50. Brandão, R.D., van Roozendaal, K., Tserpelis, D., Gómez García, E. and Blok, M.J. (2011) Characterisation of unclassified variants in the BRCA1/2 genes with a putative effect on splicing. *Breast Cancer Res. Treat.*, **129**, 971–982.

52. Pyne, M.T., Pruss, D., Ward, B.E. and Scholl, T. (1999) A characterization of genetic variants in BRCA1 intron 8 identifies a mutation and a polymorphism. *Mutat. Res.*, **406**, 101–107.

53. Bouwman, P., van der Gulden, H., van der Heijden, I., Drost, R., Klijn, C.N., Prasetyanti, P., Pieterse, M., Wientjens, E., Seibler, J., Hogervorst, F.B., *et al*. (2013) A High-Throughput Functional Complementation Assay for Classification of BRCA1 Missense Variants. *Cancer Discov*., **3**, 42-55.

54. Sevcik, J., Falk, M., Macurek, L., Kleiblova, P., Lhota, F., Hojny, J., Stefancikova, L., Janatova, M., Bartek, J., Stribrna, J., *et al*. (2013) Expression of human BRCA1Δ17-19 alternative splicing variant with a truncated BRCT domain in MCF-7 cells results in impaired assembly of DNA repair complexes and aberrant DNA damage response. *Cell. Signal*., **25**, 1186-1193.

Legends to Figures

Figure 1 Workflow We display the workflow of the 4-stage study conducted by ENIGMA investigators in order to elucidate the complexity of alternative splicing at the *BRCA1* locus. Key findings have been incorporated into the figure. Note that two novel splicing events detected by ENIGMA contributors in Stage 1 were not validated in Stage 2 and have not been incorporated into the final list of 63 *BRCA1* alternative splicing events. CE (capillary electrophoresis).

Figure 2 Splicing structural biotypes identified in the present study For the sake of clarity, the figure represents conceptual schemes of splicing structural biotypes, not a description of the *BRCA1* locus itself. Exonic sequences are indicated by black boxes. Intronic sequences by gray lines (or boxes if exonic in the reference splicing pattern).

Figure 3 *BRCA1* splicing events in blood and breast tissues. Panel A. The boxplots (Low, Q1, Median, Q3, and High values are displayed) show the expression level of 8 *predominant BRCA1* alternative splicing events relative to the full-length. Relative expression level was measured by semi-quantitative CE (see methods). BLOOD displays LEUs, PBMCs, PBMCs and LCLs data pooled together (different control samples plus technical replicates). N indicates the number of individual data-points (different control samples plus technical replicates). BREAST displays data from one BREAST sample. In this case, N equals the number of technical replicates performed. Normal outliers (>1.5 Inter Quartile Range) display a small circle. Extreme outliers (>3 IQR) display an asterisk. Panel B shows representative examples of 7-11q CE assays performed with four different RNA sources. Overall, the analysis suggests that *BRCA1* alternative splicing is similar, regardless of the RNA source analyzed. Differences are restricted to the presence/absence of *minor* events. More important, replica experiments show that differences are not RNA source (or sample) specific, but rather the results of stochastic preferential amplification of *minor* isoforms. Peaks representing a combination of two independent splicing events (SE) are annotated as (SE1+SE2).

Table 1. *BRCA1* alternative splicing events (cassette biotype)

Designation	RNA[1]	Status[2]	Functional Annotation[3]	CDS[1]	Breast[4]	QA[5]	Previously described?[6]		
							GENCODE[7]	Blood[8]	Others[9]
Δ2	r.-19_80del	cloned	Non-Coding	-	yes	Minor	-	-	BCCL (20)
Δ3	r.81_134del	cloned	PTC-NMD[1]	-	yes	Minor	006	LEU(17), PBMC(17, 48, 20)	BCCL (20)
▼4	r.134_135ins135-4047_135-3932	dir seq	PTC-NMD	-	yes	Minor	002	-	NP (13)
Δ5	r.135 _212del	cloned/dir seq	No FS	p.Phe46_Arg71del	yes	Predominant	004	LEU(13, 17, 25, 49), LCLs(29)	NB (13)
Δ9	r.548_593del	dir seq	PTC-NMD	-	yes	Minor	-	LEU(17), PBMC(22)	-
Δ10	r.594_670del	cloned	PTC-NMD	-	yes	Minor	-	LEU(17)	BCCL (20)
Δ11	r.671_4096del	cloned	No FS	p.Ala224_Leu1365del	yes	Minor	204	PBLs(50)	-
Δ13	r.4186_4357del	cloned	PTC-NMD	-	-	Minor	-	-	-
▼13A	r.4357_4358ins4358-2785_4358-2719	cloned	No FS	p.Lys1452_Ala1453ins22	yes	Minor	005	LCLs(35, 19)	NB(19)
Δ14	r.4358_c.4484del	imputed	PTC-NMD	-	-	Minor	-	-	-
Δ15	r.4485_4675del	cloned	PTC-NMD	-	-	Minor	-	LEU(17)	-
Δ17	r.4987_5074del	cloned	PTC-NMD	-	-	Minor	-	LEU(17)	-
Δ18	r.5075_5152del	dir seq	No FS	p.Asp1692_Trp1718delinsGly	-	Minor	-	-	-
Δ20	r.5194_5277del	imputed	No FS	p.His1732_Lys1759del	yes	Minor	-	-	-
Δ21	r.5278_5332del	cloned	PTC-NMD	-	yes	Minor	-	LEU(17)	-
Δ22	r.5333_5406del	cloned	FS-alternative STOP	p.Asp1778_Thr1802fsX32	yes	Minor	007	-	-
Δ23	r.5407_5467del	cloned/dir seq	FS-alternative STOP	p.Gly1803_Ala1823delfsX11	yes	Minor	-	-	-

[1]According to HGVS guide-lines (http://www.hgvs.org/mutnomen). Nucleotide +1 corresponding to the A of the AUG translation initiation codon in the Ensemble reference transcript ENST00000357654. Ensemble reference protein ENSP00000350283. [2]Events have been cloned and sequenced (cloned), directly sequenced from splicing assays (dir seq.), or imputed (see methods). [3]According to Mudge and cols (4). See methods for further details. [4] Detected in normal breast tissue. [5] Qualitative Abundance (QA) based on visual inspection of splicing assays (see methods for further details) [6]We have excluded splicing events described as the outcome of germ-line mutation (i.e. for instance, Δ18 has been described previously as the outcome of various germ-line pathogenic mutations). [7]GENECODE transcript IDs retrieved from Ensemble (if the corresponding splicing events is present in more than one transcript, the lowest ID number is shown). [8]Leukocytes (LEU), Peripheral blood mononuclear cell (PBMCs), primary cultures (PBLs), Lymphoplastoid cell lines (LCLs). [9] Breast Cancer Cell Lines (BCCLs), Non-Malignant Placenta (NP). Non-Malignant Breast (NB). [10] In-frame event generating a PTC at the splice junction.

Table 2. BRCA1 alternative splicing events (multi-cassette biotype).

Designation	RNA[1]	Status[2]	Functional Annotation[3]	CDS[1]	Breast[4]	QA[5]	Previously described?[6]		
							GENCODE[7]	Blood[8]	Others[9]
Δ2,3	r.-19_134del	impute	Non-Coding	-	yes	Minor	-	-	
Δ2_5	r.-19_217del	impute	Non-Coding	-	yes	Minor	-	-	
Δ2_10	r.-19_670del	dir seq.	Non-Coding	-	-	Minor	003	LCLs (24)	NB(24)
Δ8,9	r.442_593del	impute	PTC-NMD	-	-	Minor	-	-	
Δ8_10	r.442_670del	impute	PTC-NMD	-	yes	Minor	-	-	
Δ9,10	r.548_670del	cloned/dir seq	No FS	p.Gly183_Lys223del	yes	Predominant	015	LEU(13, 26),LCLs (29, 52)	NB,NO, (13) BCCL(20)
Δ9_11	r.548_4096del	cloned/dir seq	No FS	p.Ser184_Gly1366del	yes	Predominant	203	LEU(17), LCL(24)	NB(13)
Δ9_12	r.548_4185del	cloned	PTC-NMD	-	yes	Minor	-	-	
Δ10,11	r.594_4096del	dir seq.	PTC-NMD	-	yes	Minor	-	-	
Δ10_12	r.594_4185del	cloned	PTC-NMD	-	yes	Minor			
Δ11,12	r.671_4185del	impute	PTC-NMD	-	yes	Minor	-	-	
Δ14_15	r.4358_4675del	cloned	No FS	p.Ala1453_Leu1558del	-	Minor	-	-	
Δ14_17	r.4358_5074del	cloned	No FS	p.Ala1453_Thr1691del	-	Minor	202	LCLs (24)	NB(24)
Δ14_18	r.4358_5152del	cloned	No FS	p.Ala1453_Trp1718delinsGly	-	Minor	205	LCLs (24)	NB(24)
Δ14_19	r.4358_5196del	cloned	PTC-NMD	-	-	Minor	-	-	
Δ15_17	r.4485_5074del	cloned	PTC-NMD	-	yes	Minor	-	LEU(17), LCL(24)	NB(24)
Δ15_19	r.4485_5193del	cloned	PTC-NMD	-	yes	Minor	-	-	
Δ21,22	r.5278_5406del	impute	No FS	p.Ile1760_Thr1802del	yes	Minor	-	-	
Δ21_23	r.5278_5467del	cloned	FS-alternative STOP	p.Ile1760_Ala1823delfsX11	yes	Minor	-	-	
Δ22,23	r.5333_5467del	impute	FS-alternative STOP	p.Asp1778_Ala1823delfsX11	yes	Minor	-	-	

[1]According to HGVS guide-lines (http://www.hgvs.org/mutnomen). Nucleotide +1 corresponding to the A of the AUG translation initiation codon in the Ensemble reference transcript ENST00000357654. Ensemble reference protein ENSP00000350283. [2]Events have been cloned and sequenced (cloned), directly sequenced from splicing assays (dir seq.), or imputed (see methods). [3]According to Mudge and cols (4). See methods for further details. [4] Detected in normal breast tissue. [5] Qualitative Abundance (QA) based on visual inspection of splicing assays (see methods for further details) [6]We have excluded splicing events described as the outcome of germ-line mutation (i.e. for instance, Δ18 has been described previously as the outcome of various germ-line pathogenic mutations). [7]GENECODE transcript IDs retrieved from Ensemble (if the corresponding splicing events is present in more than one transcript, the lowest ID number is shown). [8] Leukocytes (LEU), Peripheral blood mononuclear cell (PBMCs), primary cultures (PBLs), Lymphoplastoid cell lines (LCLs). [9] Breast Cancer Cell Lines (BCCLs), Non-Malignant Ovarian (NO). Non-Malignant Breast (NB).

Table 3. ***BRCA1*** **alternative splicing events (miscellaneous biotypes).**

Designation	RNA[1]	Status[2]	Functional Annotation[3]	CDS[1]	Breast[4]	QA[5]	Previously described?[6]		
							GENCODE[7]	Blood[8]	Others[9]
Splice acceptor shift									
Δ2p	r.-19_-7del	dir seq.	UTR	-	yes	Minor	-	-	
Δ8p	r.442_444del	cloned/dir seq	No FS	p.Gln148del	yes	Predominant	009	PBMC (13)(20)(29)	NB, NO (13)
Δ13p	r.4186_4188del	cloned/dir seq	No FS	p.Gln1396del	yes	Predominant	-	LCLs (29)	
Δ14p	r.4358_4360del	cloned/dir seq	No FS	p.Ala1453del	yes	Predominant	005	PBMC (13), LCLs(29)	NB (13)
Splice donor shifts									
Δ1Aq	r.-25_-20del	cloned/dir seq	UTR	-	yes	Predominant	006	LCLs (21)	BCCL (20), BT(21)
▼1aA	r.-20_-19ins-20+1_-20+89	cloned	UTR	-	yes	Minor	010	-	
Δ5q	r.191_212del	cloned/dir seq	PTC-NMD	-	yes	Predominant	010	PBMC(48), LEU (25), LCL (29)	
Δ11q	r.788_4096del	cloned/dir seq	No FS	p.Ser264_Gly1366del	yes	Predominant	007	LCLs (24)	BCCL (20), NB(24)
Intronization									
11Δ3110	r.788_3897del	dir seq.	PTC-NMD	-	-	Minor	-	-	
11Δ3240	r.788_4027del	dir seq.	No FS	p.Gly263_Ser1342del	-	Minor	-	-	
Terminal modification									
(1B)		dir seq	UTR	-	yes	-	-	-	BCCL,BT,OT,NB(23)
(IRIS)		dir seq	IntronicSTOP+polyA		not tested	-	012	LCLs(18)	BCCL, BT(18)

[1]According to HGVS guide-lines (http://www.hgvs.org/mutnomen). Nucleotide +1 corresponding to the A of the AUG translation initiation codon in the Ensemble reference transcript ENST00000357654. Ensemble reference protein ENSP00000350283. [2]Events have been cloned and sequenced (cloned), directly sequenced from splicing assays (dir seq.), or imputed (see methods). [3]According to Mudge and cols (4). See methods for further details. [4] Detected in normal breast tissue. [5] Qualitative Abundance (QA) based on visual inspection of splicing assays (see methods for further details) [6]We have excluded splicing events described as the outcome of germ-line mutation (i.e. for instance, Δ18 has been described previously as the outcome of various germ-line pathogenic mutations). [7]GENECODE transcript IDs retrieved from Ensemble (if the corresponding splicing events is present in more than one transcript, the lowest ID number is shown). [8]Leukocytes (LEU), Peripheral blood mononuclear cell (PBMCs), primary cultures (PBLs), Lymphoplastoid cell lines (LCLs). [9]Breast Cancer Cell Lines (BCCLs), Non-Malignant Breast (NB). Non-Malignant Ovarian (NO). Breast Tumor (BT). Ovarian Tumor (OT).

Table 4. *BRCA1* alternative splicing events (mixed biotypes)

Designation	RNA[1]	Status[2]	Functional Annotation[3]	CDS[1]	Breast[4]	QA[5]	Previously described?[6]		
							Gencode[7]	Blood	Others
Splice donor shift + (Multi)-cassette									
Δ1Aq,2	r.-25_80del	dir seq	Non-Coding	-	yes	minor	-	-	-
Δ1Aq_3	r.-25_134del	imputed	Non-Coding	-	yes	minor	-	-	-
Δ1Aq_5	r.-25_217del	imputed	Non-Coding	-	yes	minor	-	-	-
Δ1Aq_10	r.-25_670del	dir seq	Non-Coding	-	yes	minor	-	-	-
(Multi)-cassette + Splice acceptor shift									
Δ10_13p	r.594_4188del	imputed	PTC-NMD	-	yes	minor	-	-	-
Δ11_13p	r.671_4188del	imputed	PTC-NMD	-	yes	minor	-	-	
Δ13_14p	r.4186_4360del	imputed	PTC-NMD	-	-	minor	-	-	-
▼13A,Δ14p	r.4357_4358ins4358-2785_4358-2719+r.4358_4360del	imputed	No FS	p.Ala1453delins22	yes	minor	005	-	-
Terminal Modification + (Multi)-cassette									
(1B),Δ2	r.-19_80del	dir seq	Non-Coding	-	yes	-	206	-	-
(1B),Δ2,3	r.-19_134del	dir seq	Non-Coding	-	yes	-	-	-	-
(1B),Δ2_5	r.-19_217del	imputed	Non-Coding	-	-	-	-	-	-
Multi-cassette + Cassette									
Δ2,3,▼4	r.-19_134del+r.134_135ins135-4047_135-3932	imputed	Non-Coding	-	-	minor	-	-	-
Splice donor shift + Splice acceptor shift									
Δ1Aq,Δ2p	r.-25_-7del	cloned	UTR	-	yes	minor	-	-	-
Splice donor shift + Multi-cassette + Cassette									
Δ1Aq_3,▼4	r.-25_134del+r.134_135ins135-4047_135-3932	cloned	Non-Coding	-	-	minor	-	-	-

[1]According to HGVS guide-lines (http://www.hgvs.org/mutnomen). Nucleotide +1 corresponding to the A of the AUG translation initiation codon in the Ensemble reference transcript ENST00000357654. Ensemble reference protein ENSP00000350283. [2]Events have been cloned and sequenced (cloned), directly sequenced from splicing assays (dir seq.), or imputed (see methods). [3]According to Mudge and cols **(4)**. See methods for further details. [4] Detected in normal breast tissue. [5] Qualitative Abundance (QA) based on visual inspection of splicing assays (see methods for further details) [6]We have excluded splicing events described as the outcome of germ-line mutation (i.e. for instance, Δ18 has been described previously as the outcome of various germ-line pathogenic mutations). [7]GENECODE transcript IDs retrieved from Ensemble (if the corresponding splicing events is present in more than one transcript, the lowest ID number is shown).

Abbreviations

bp: base pairs.

BCCLs: Breast Cancer Cell Lines.

BT: Breast Tumor.

CDS: Coding Sequence.

CE: Capillary Electrophoresis**.**

dir seq: Direct Sequencing.

Et-Br: Ethidium Bromide.

FS: Frame-Shift.

HGVS: Human Genome Variation Society.

IDDs: Intrinsically Disordered Domains.

IDPs: Intrinsically Disordered Proteins.

IQR: Inter Quartile Range.

IRIS: In-frame Reading of BRCA1 Intron 11 Splice Variant.

LCLs: Lymphoblastoid Cell Lines.

LEUs: Leukocytes.

NMD: Nonsense-Mediated mRNA Decay.

MAF: Minor Allele Frequency.

NB: Non Malignant Breast.

NO: Non Malignant Ovarian.

NP: Non Malignant Placenta.

OT: Ovarian Tumor.

PBLs: Peripheral Blood Leukocytes.

PBMCs: Peripheral Blood Mononuclear Cells.

PTC: Premature Termination Codon.

Q1: First Quartile.

Q3: Third Quartile.

QA: Qualitative Abundance.

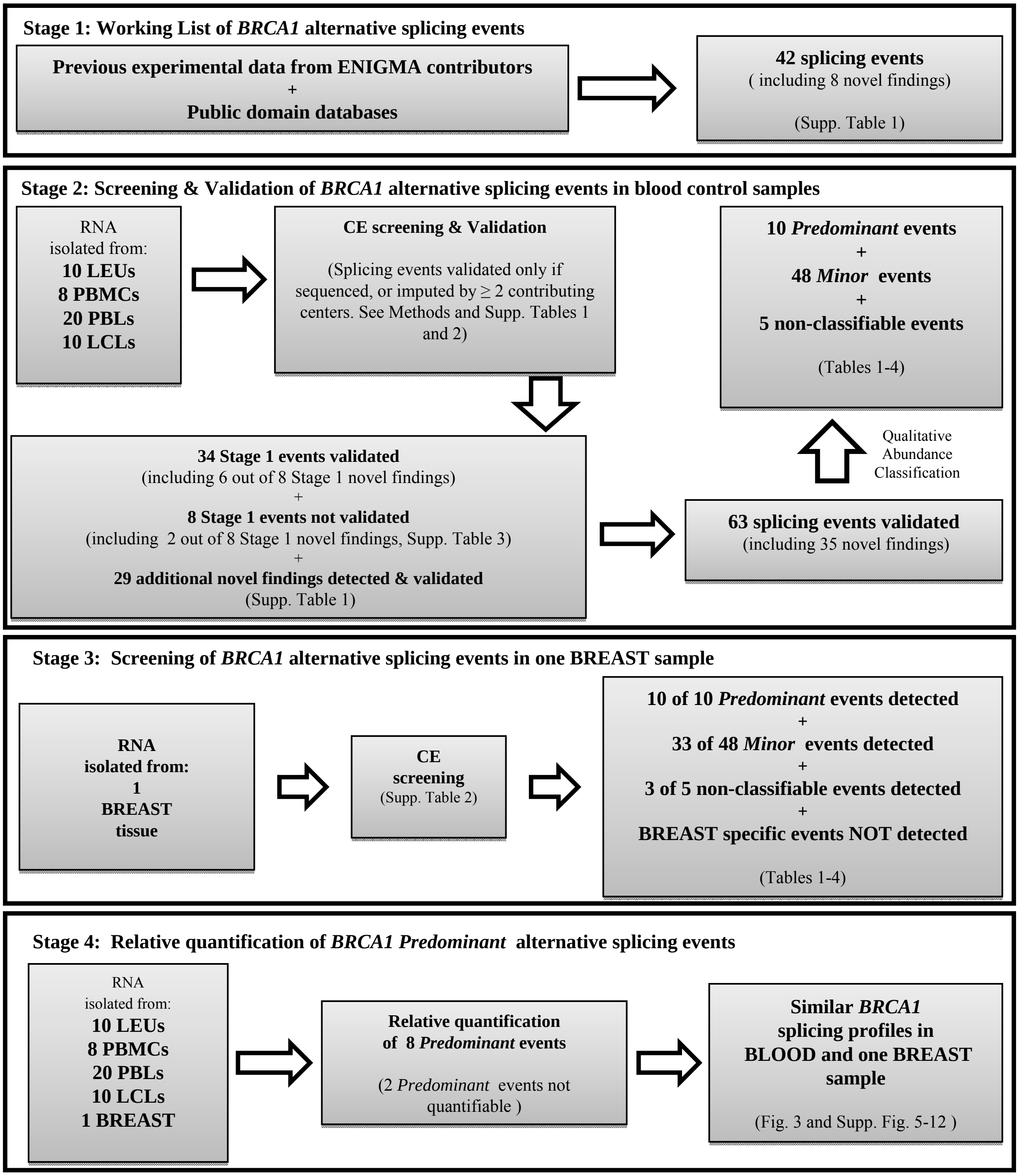
RT-PCR: Reverse Transcription-Polymerase Chain Reaction.

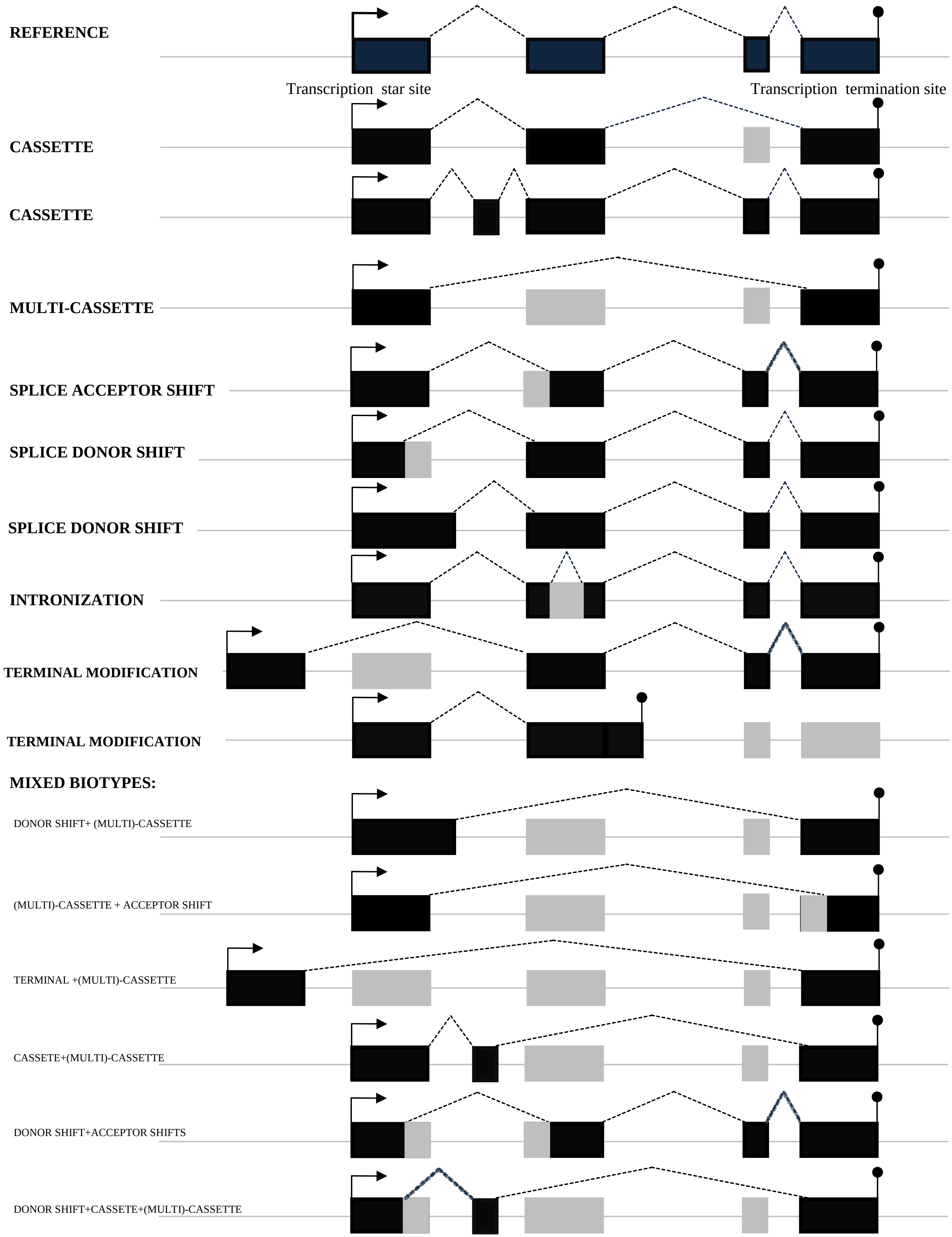
SE: Splicing Event.

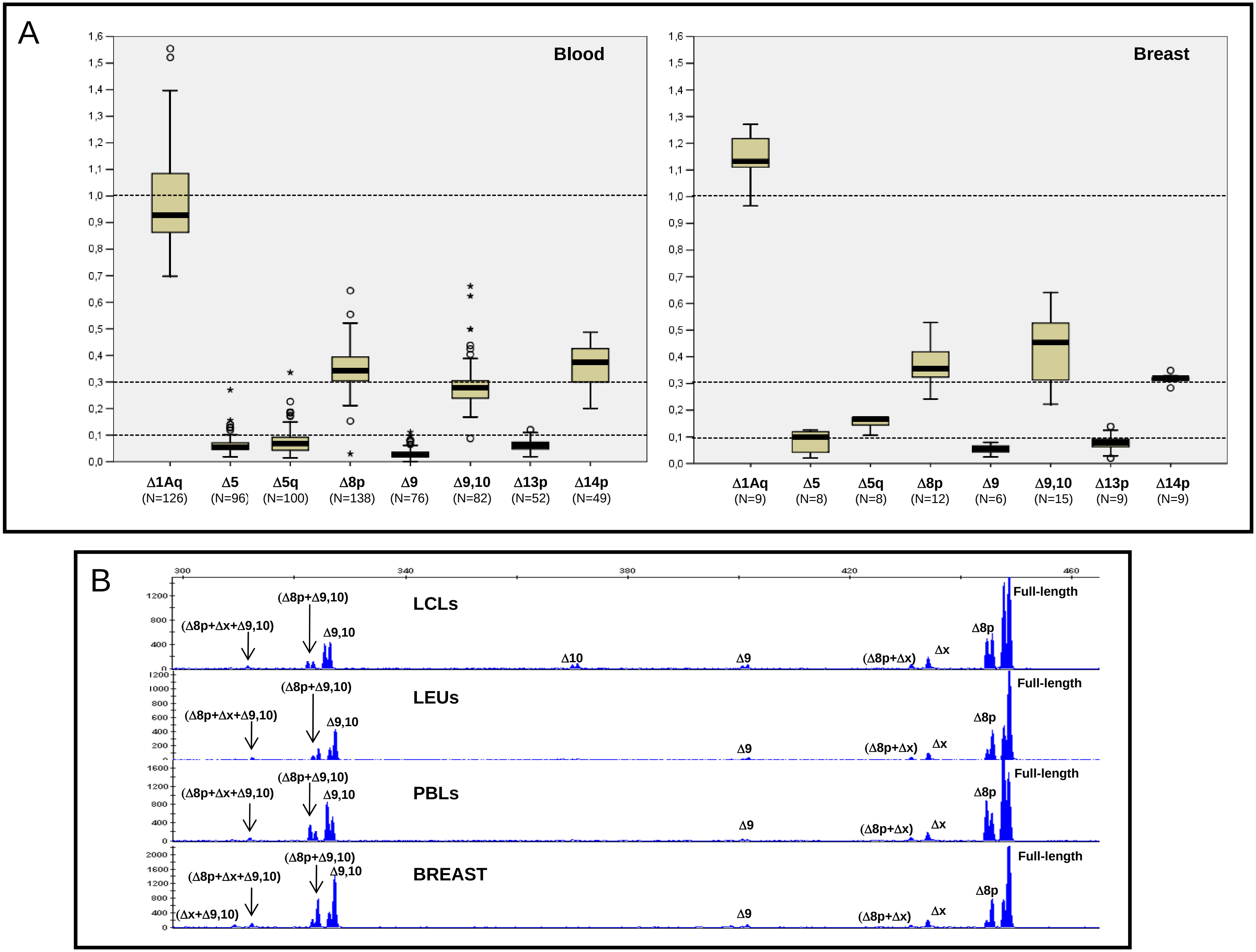
SNP: Single Nucleotide Polymorphism.

sQTL: splicing Quantitative Trait Loci.

UTR: Untranslated Region.


Stage 1: Working List of BRCA1 alternative splicing events
Previous experimental data from ENIGMA contributors
+
Public domain databases
42 splicing events
(including 8 novel findings)
(Supp. Table 1)
Stage 2: Screening & Validation of BRCA1 alternative splicing events in blood control samples
RNA
isolated from:
10 LEUs
8 PBMCs
20 PBLs
10 LCLs
CE screening & Validation
(Splicing events validated only if sequenced, or imputed by ≥ 2 contributing centers. See Methods and Supp. Tables 1 and 2)
10 Predominant events
+
48 Minor events
+
5 non-classifiable events
(Tables 1-4)
Qualitative Abundance Classification
34 Stage 1 events validated
(including 6 out of 8 Stage 1 novel findings)
+
8 Stage 1 events not validated
(including 2 out of 8 Stage 1 novel findings, Supp. Table 3)
+
29 additional novel findings detected & validated
(Supp. Table 1)
63 splicing events validated
(including 35 novel findings)
Stage 3: Screening of BRCA1 alternative splicing events in one BREAST sample
RNA
isolated from:
1
BREAST
tissue
CE
screening
(Supp. Table 2)
10 of 10 Predominant events detected
+
33 of 48 Minor events detected
+
3 of 5 non-classifiable events detected
+
BREAST specific events NOT detected
(Tables 1-4)
Stage 4: Relative quantification of BRCA1 Predominant alternative splicing events
RNA
isolated from:
10 LEUs
8 PBMCs
20 PBLs
10 LCLs
1 BREAST
Relative quantification
of 8 Predominant events
(2 Predominant events not quantifiable)
Similar BRCA1
splicing profiles in
BLOOD and one BREAST
sample
(Fig. 3 and Supp. Fig. 5-12)


REFERENCE
Transcription star site
Transcription termination site
CASSETTE
CASSETTE
MULTI-CASSETTE
SPLICE ACCEPTOR SHIFT
SPLICE DONOR SHIFT
SPLICE DONOR SHIFT
INTRONIZATION
TERMINAL MODIFICATION
TERMINAL MODIFICATION
MIXED BIOTYPES:
DONOR SHIFT+ (MULTI)-CASSETTE
(MULTI)-CASSETTE + ACCEPTOR SHIFT
TERMINAL +(MULTI)-CASSETTE
CASSETE+(MULTI)-CASSETTE
DONOR SHIFT+ACCEPTOR SHIFTS
DONOR SHIFT+CASSETE+(MULTI)-CASSETTE


A
Blood
Breast
Δ1Aq
(N=126)
Δ5
(N=96)
Δ5q
(N=100)
Δ8p
(N=138)
Δ9
(N=76)
Δ9,10
(N=82)
Δ13p
(N=52)
Δ14p
(N=49)
Δ1Aq
(N=9)
Δ5
(N=8)
Δ5q
(N=8)
Δ8p
(N=12)
Δ9
(N=6)
Δ9,10
(N=15)
Δ13p
(N=9)
Δ14p
(N=9)
B
LCLs
LEUs
PBLs
BREAST
Full-length
Δ8p
Δx
(Δ8p+Δx)
Δ9
Δ10
Δ9,10
(Δ8p+Δ9,10)
(Δ8p+Δx+Δ9,10)
(Δx+Δ9,10)