Supplemental Table 1. Detection and Validation of 63 *BRCA1* alternative splicing events.

1Splicing events displayed in blue boxes were supported by experimental evidence of contributing laboratories. By contrast, splicing events displayed in white boxes were supported solely by public domain scientific data 2 Predicted size-fragments according to Ensemble reference transcript ENST00000357654. The actual size-fragment observed in CE might be slightly different (±2bp) 3 Splicing events displayed in blue boxes have been validated in the present study. \* Stage 2 amplicons were designed prior to the publication of these findings. 4 Exon1B terminal events have been analyzed by one contributing laboratory only. Since (E1B)+Δ2\_5 has not been sequenced, this event do not qualify for a validated event according to our own criteria (see methods). However, we have validated this imputed splicing event based on the following data. First, *BRCA1* Δ2, and Δ2,3 events have been validated in both E1B and E1A containing transcripts, suggesting that alternative splicing events at the 5' end of the BRCA1 gene is similar, regardless of the transcription star site used. Second, Δ2\_5 has been validated in E1A transcripts.

Supplemental Table 2. Stage 2 and 3 Data



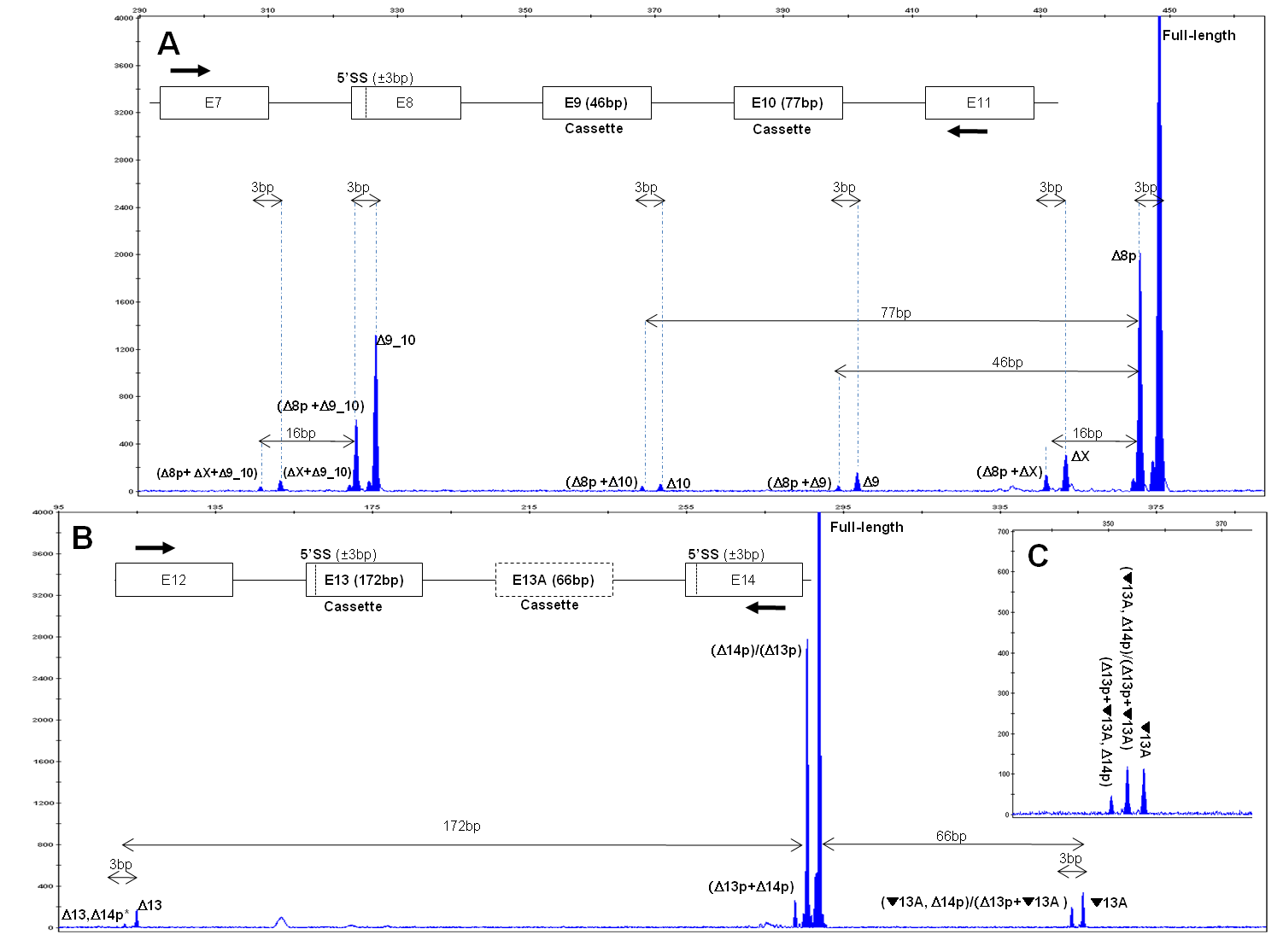
1 Qualitative abundance (QA) based on visual inspection of splicing assays (see methods for further details) 2 *Coverage* refers to data-points *per* splicing event. *Detection Rate* refers to the % of positive data-points. (See methods for further details). 3 Gray boxes indicate a minimum of one positive data-point. 4Splicing assays interrogating the corresponding event. Some events are detected by two or more different assays (see Supp Table 1 for further details).

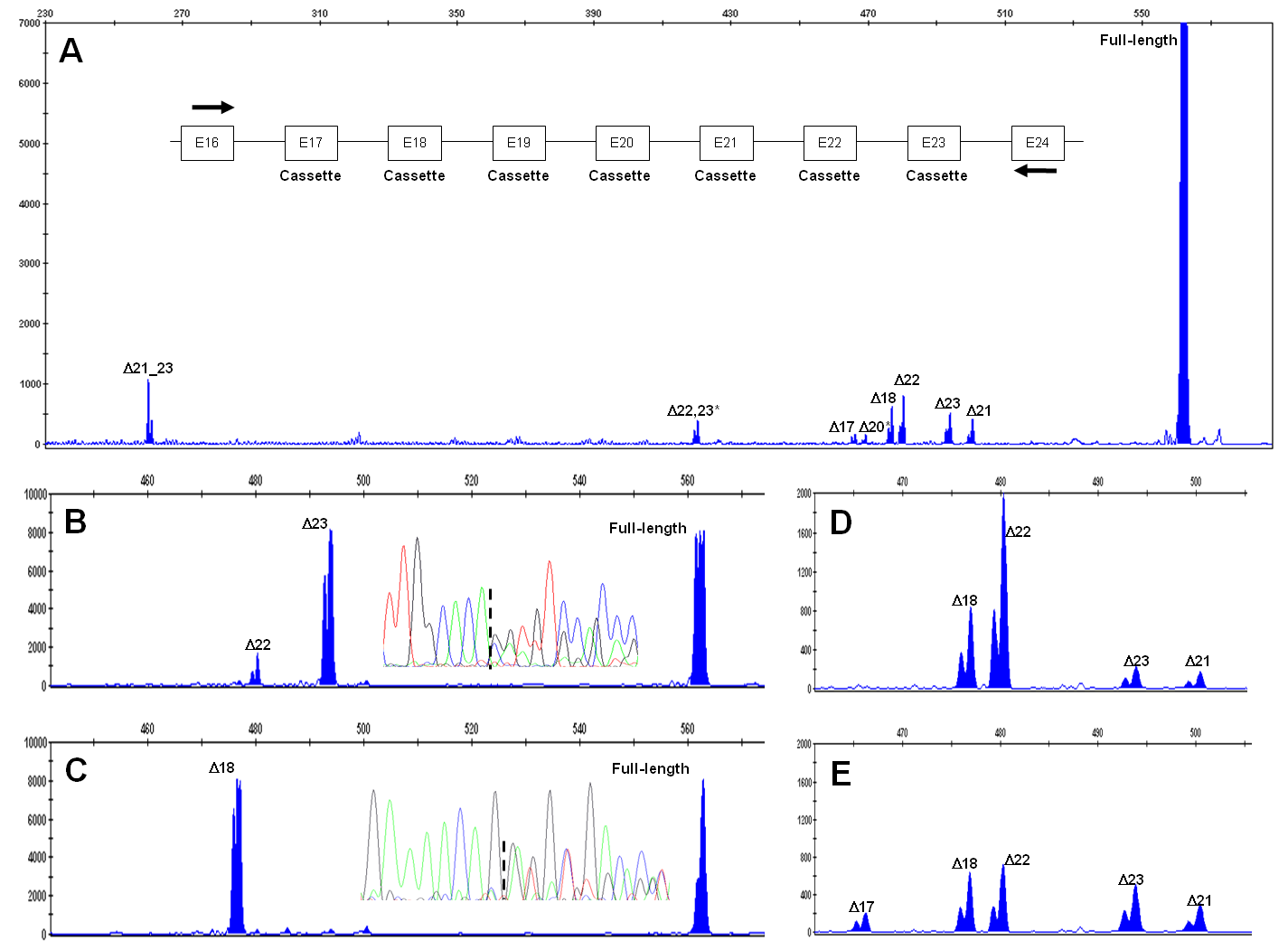
**Supplemental Table 3. *BRCA1* alternative splicing events not validated by our experimental approach.**

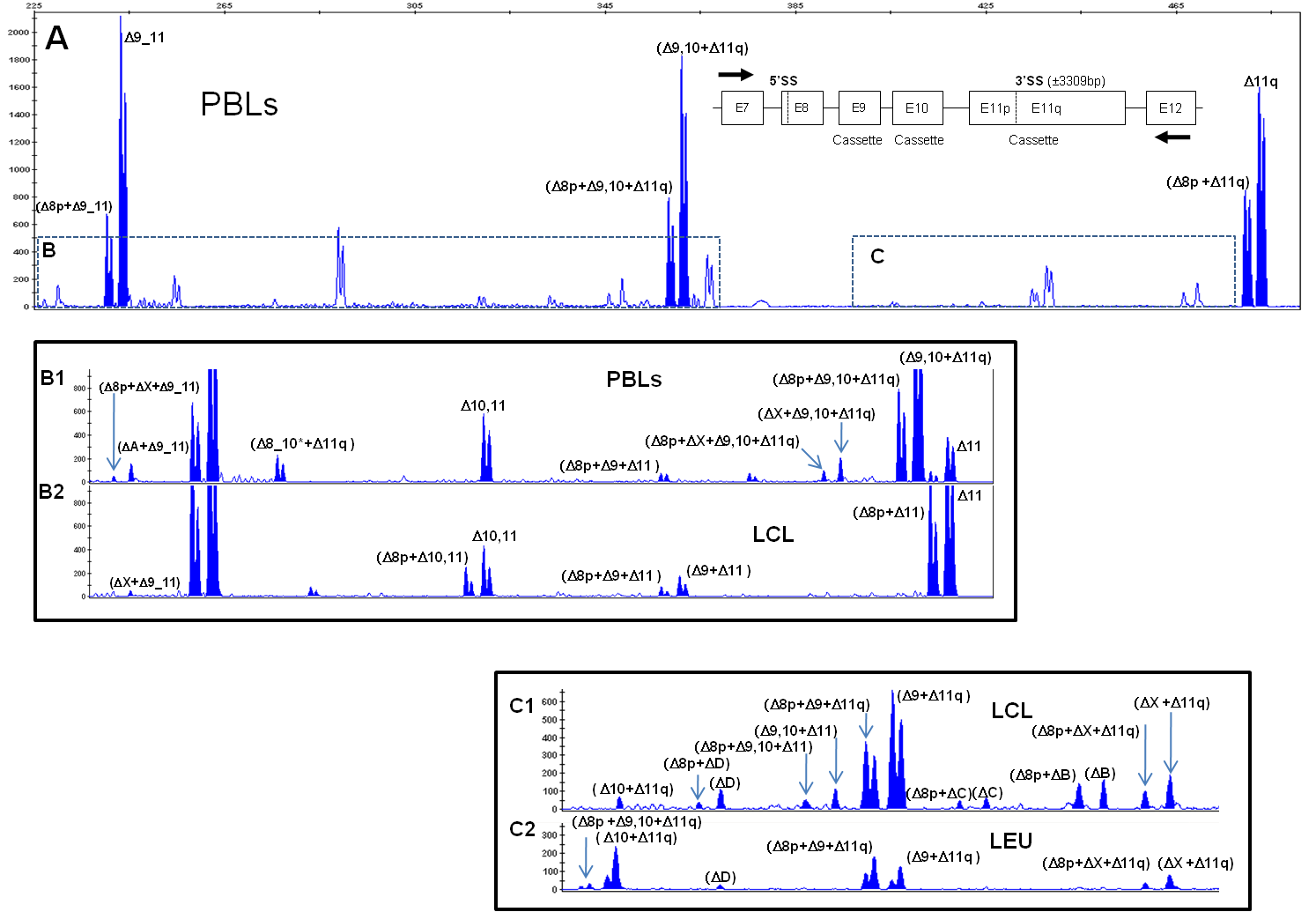
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | | | **This study** | **Previous studies** | | |
| **Designation** | **HGVS** | | **Biotype** | **Functional**  **Annotation** | **Coverage1** | **GENCODE** | **Blood** | **Others** |
| **▼1aA+Δ2\_17** | r.-20\_-19ins-20+1\_-20+89 + r.-19\_5074del | | Splice donor shift + Multi-cassette | No FS2 | 72\* | - | - | BCCLs, BCs(22) |
| **Δ3\_5** | r.81\_217del | | Multi-cassette | PTC-NMD | 61 | - | PBMCs (17) | - |
| **Δ5q,6** | r.191\_301del | | Splice donor shift + cassette | PTC-NMD | 124 | - | PBMC (48) | - |
| **Δ6,7** | r.213\_441del | | Multi-cassette | PTC-NMD | 124 | - | PBMC (48) | - |
| **Δ13\_19** | r.4186\_5193del | | Multi-cassette | No FS | 20\* | - | PBMCs (17) | - |
| **Δ17\_19** | r.4987\_5193del | | Multi-cassette | No FS | 223 | - | PBMCs (17) | - |
| **Δ18\_20** | r.5075\_5277del | | Multi-cassette | PTC-NMD | 103 | - | - | - |
| **Δ19** | r.5153\_5193del | | Cassette | PTC-NMD | 103 | - | - | - |

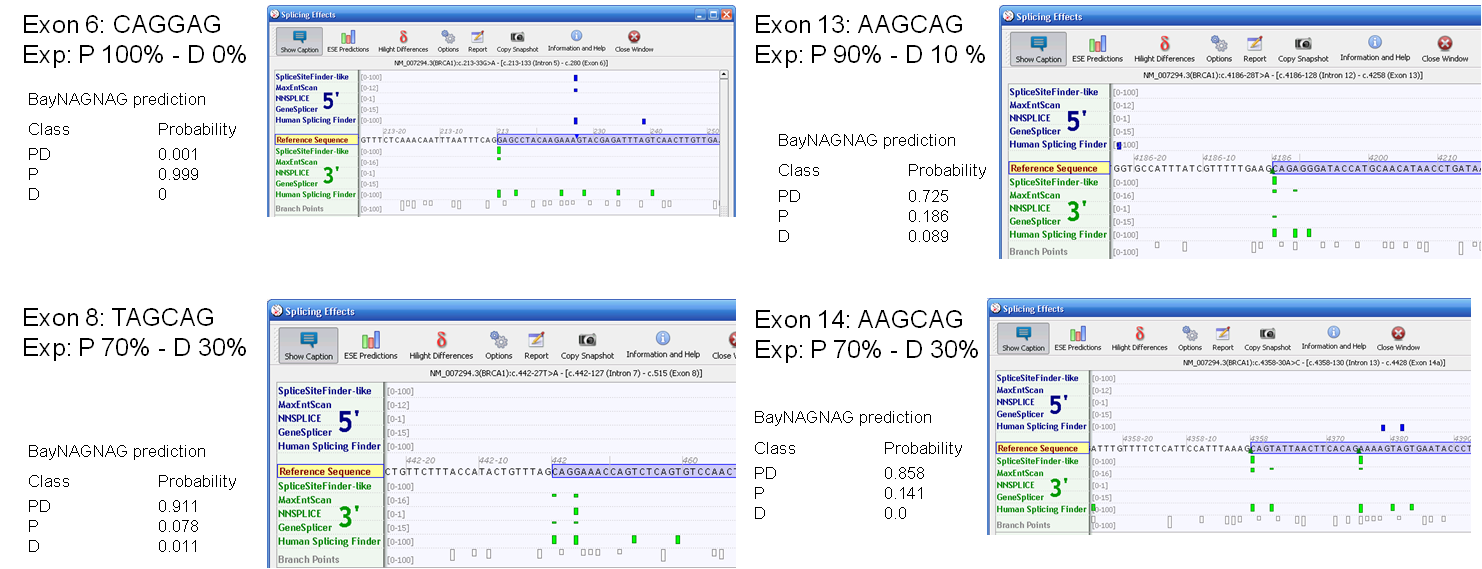
1 Coverage was calculated pooling together splicing assays performed in LEUs, PBMCs, PBLs, LCLs, and BREAST samples (independent control samples plus technical replicas). 2 Exon 1Aa is a splice donor shift event incorporating 89 nucleotides into exon 1A. The inserted nucleotides include an in-frame ATG. \* Since none of the Stage 2 splicing assays (see Supp Table 1) are able to detected ▼1aA+Δ2\_17 and Δ13\_19 events, we later performed dedicated screenings with splicing assays 1A-22 and 12-22.

Since ▼1aA+Δ2\_17 has been detected in Breast Cancer Cell Lines (BCCLs) and Breast Cancers (BCs) but not in control samples, it does not necessarily qualify for a *naturally occurring* splicing event (it might be, for instance, a by-product of the carcinogenic process). Yet, ▼1aA+Δ2\_17 combines two biotypes, one of them a donor shift (▼1aA) detected in the present study as an individual event both in blood and BREAST (Table 3). By contrast, *BRCA1* Δ(5q,6), Δ(6,7), Δ3\_5, Δ13\_19, and Δ17\_19 have been reported in blood related control samples, thus apparently qualifying for *naturally occurring* events. Yet, the authors do not inform on relevant analytical parameters such as coverage, detection rates, and/or relative expression levels (or whether the analyses were performed in the presence of NMD inhibitors). It is therefore possible that some of these events represent rather minor events. In this regard it is interesting to note that we have not detected Δ13\_19 and Δ17\_19, but we have detected other (multi)-cassette events in the same *BRCA1* region, most of them among the Stage 2 events with lower detection rates (see Supp. Table 2). We have identified in our study up to 23 PTC-NMD splicing events, suggesting that lack of NMD inhibition (at least in capillary EP based analysis) is not a major concern (37). Yet, we cannot discard that certain PTC-NMD events, including Δ(5q,6), Δ(6,7), and Δ3\_5, are particularly prone to NMD degradation. BRCA1 Δ18\_20 and Δ19 have been detected (imputed) by contributors of the present study (one contributor each). However, we do not have sequence evidence nor validation detection by a second laboratory.

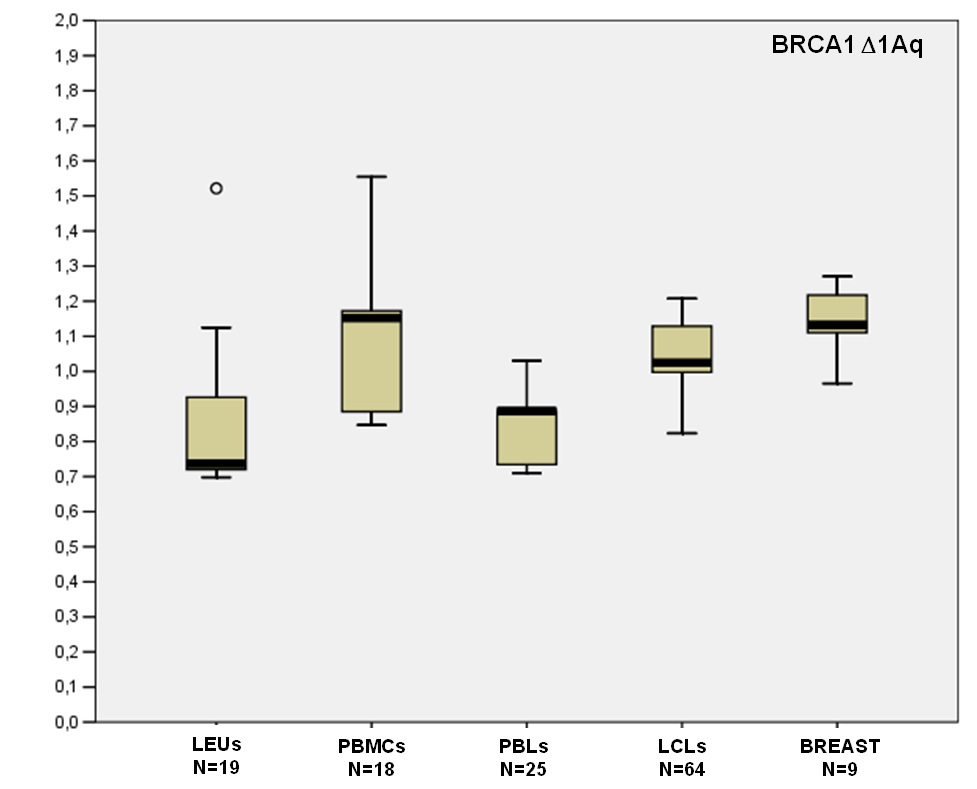
**Supplemental Figure 1** **Annotation of** ***BRCA1* alternative splicing events** Panel A shows a representative example of a 7-11q CE assay. Note that ±3bp (NAGNAG located at the 5’ end of exon 8), ±46bp (exon 9), and ±77bp (exon 10) spacing between peaks are essential for imputation and proper peak annotation. Interestingly, the assay predicts the existence of one splicing event (SE) involving ±16bp (annotated as X in the figure). However, we have not been able to impute X. Panel B shows a representative example of a 12-14 assay. In this assay, ±3bp (NAGNAG located at the 5’ ends of exon 13 and exon 14), ±172bp (exon13), and ±66bp (exon13A) spacing between peaks are essential for imputation and proper peak annotation. Panel C shows a detail of an independent 12-14 assay (a technical replica) in which one extra peak in the 350bp region is observed. Note that the extra peak shown in Panel C is essential for ▼13A,14p annotation. Peaks representing a combination of two independent splicing events are annotated as (SE1+SE2).Peaks compatible with different SE events are annotated as (SE1)/(SE2). \*Annotations not supported by sequencing data (see Tables 1 to 4 for further details). Note that, overall, the signal corresponding to transcripts containing two splicing events is consistently lower than that of the transcripts containing the corresponding individual events (compare for instance Δ8p+Δ9,10 with Δ8p and Δ9,10), as expected from a random combination of independent elements.

Supplemental Figure 2. Stochastic preferential amplification of *minor* splicing events. CE analysis of 16-24 assays allowed us to identify up to 8 different alternative splicing events. Panel A shows a representative example. Compared with the full-length, all them are rather minor events (note that the corresponding full-length peak is saturated), so direct sequencing of splice junctions is not feasible. Panels B and C show examples of stochastic preferential amplification of minor alternative splicing events (Δ23 and Δ18, respectively) which allow direct sequencing of the splicing junctions. To avoid Δ22 interference, Δ23 splice junction was sequenced with a forward primer located at exon 22. Panels D and E show two assays performed in parallel with the same sample (technical replicas). Four alternative splicing events are detected in both assays (only the 420-520bp region is shown), but Δ17 is detected only once. \*Annotations not supported by sequencing data (see Tables 1-4 for further details).

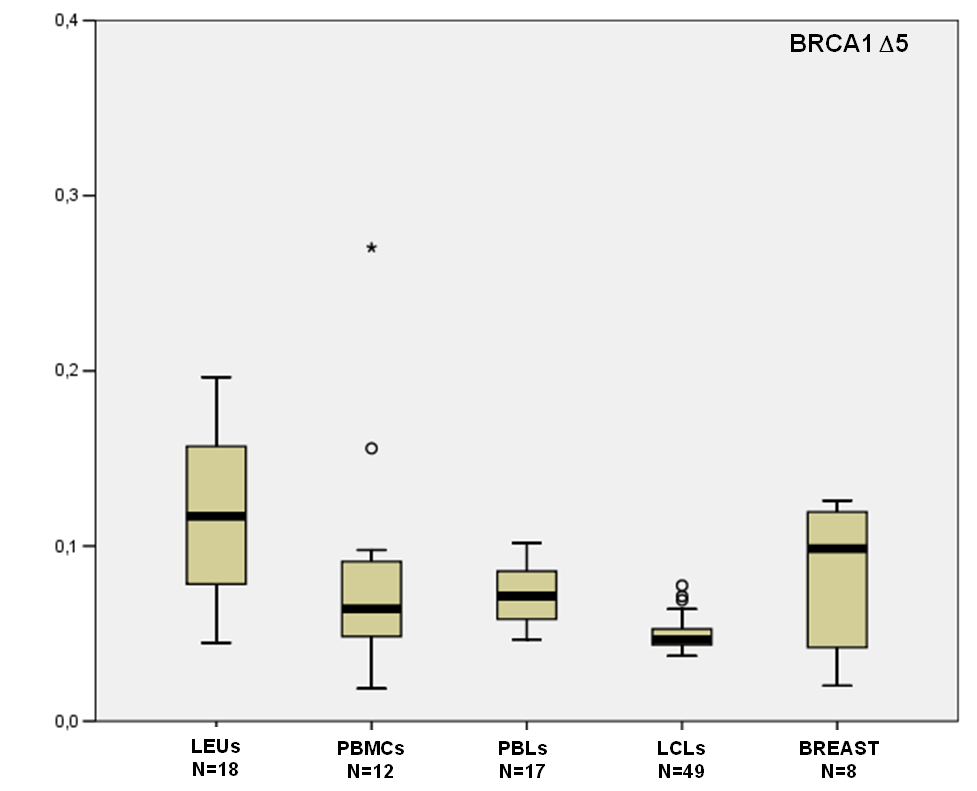
**Supplemental Figure 3. *BRCA1* alternative splicing in exon11 lacking RNAs.** Panel A shows a representative example of a 7-12 CE assay. The full-length fragment (3793bp) is out of range for CE analysis. Dashes areas are shown in greater detail in panel B and C (note that panel B1 is a magnification of panel A, but panels B2, C1 and C2 represent independent assays. Several peaks have been annotated, most of them representing combinations of independent events. Other peaks have not been annotated. Some might represent unspecific amplifications, but others (in particular ΔB, ΔC, ΔD, and ΔX) are apparently observed in combination with *BRCA1* events, supporting that they are indeed true *BRCA1* splicing events that we have not been able to annotate. Interestingly, the ΔX event (±16bp) is probably the same ΔX event detected in 7-11q CE assays (see Figure 1). Peaks representing a combination of independent splicing events (SE) are annotated as (SE1+SE2+SEn). \*Annotations not supported by sequencing data (see Tables 1-4 for further details). Note that, overall, the signal corresponding to transcripts containing two or more splicing events is consistently lower than that of the transcripts containing the corresponding individual events, as expected from a random combination of independent elements.

****

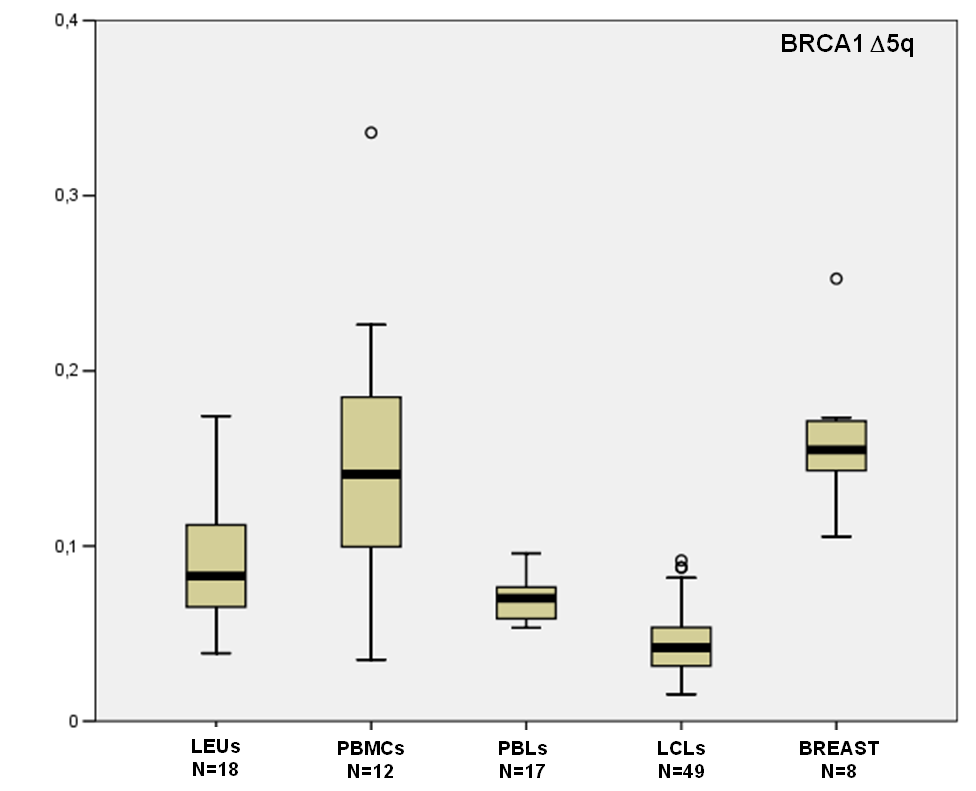
**Supplemental Figure 4. NAGNAG sequence motifs at the 5’ end of *BRCA1* exon 6, 8, 13 and 14 and local splice-site prediction.** Screenshot from Alamut software (Interactive Biosoftware) showing the local sequence at the 5’end of the exon 6, 8, 13 and 14, including the composition of the NAGNAG motif. The 3’ predictions in green represent the outcomes for predictions of the strength of a splice acceptor site as predicted by the different integrated software tools. The height of the vertical green bars is correlated with the height of the score (range of scores indicated in light gray next to the different tools). The NAGNAG-motif is displayed next to the screenshot, including the experimentally (Exp) determined ratio between the use of the proximal (P) or distal (D) acceptor site (NAGPNAGD). BayNAGNAG prediction scores (<http://www.tassdb.info/baynagnag/form.html>) are given (≥0.5 = significant result). PD = usage of both acceptors, P or D = constitutive splicing at either P or D acceptor site.

****

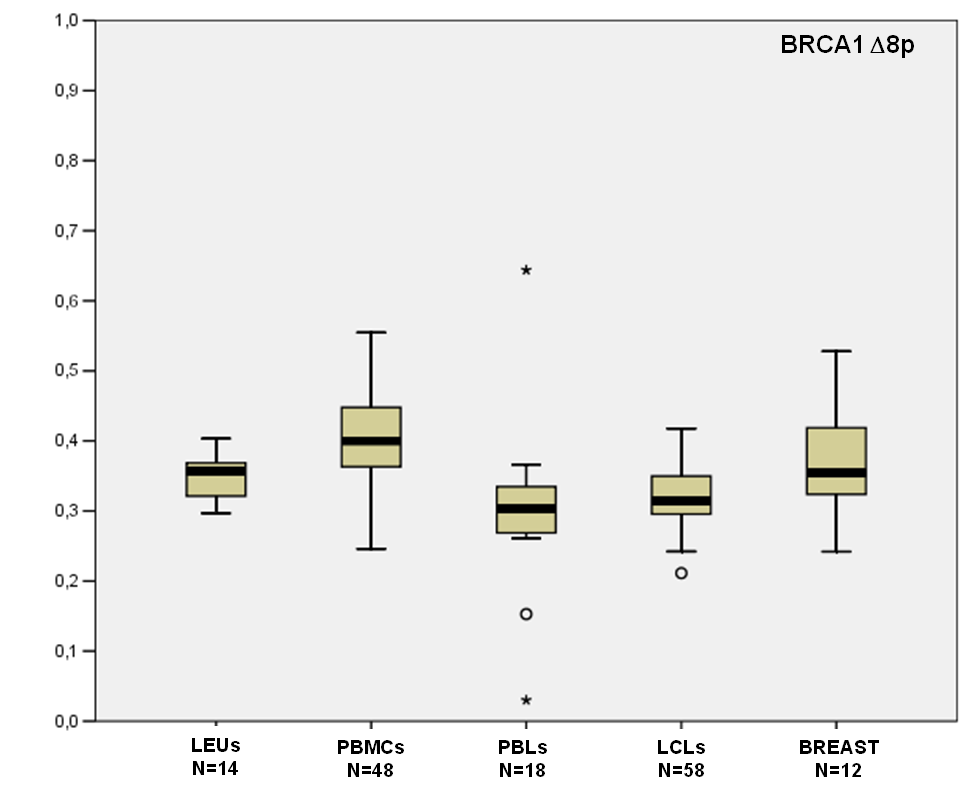
**Supplemental Figure 5.** Relative expression level of *BRCA1* Δ1Aq in 4 blood related RNA sources and one breast tissue. 33-cycle RT-PCR assays (E1A-E3) were analyzed by CE. The boxplot shows the ratios between the peak area of Δ1Aq and the peak areas of the reference full-length transcript (Low, Q1, Median, Q3, ad High values are indicated). Normal outliers (>1.5 IQR) display a small circle. N indicates the number of individual data-points (different control samples plus technical replicates). In BREAST, N indicates the number of technical replicas.



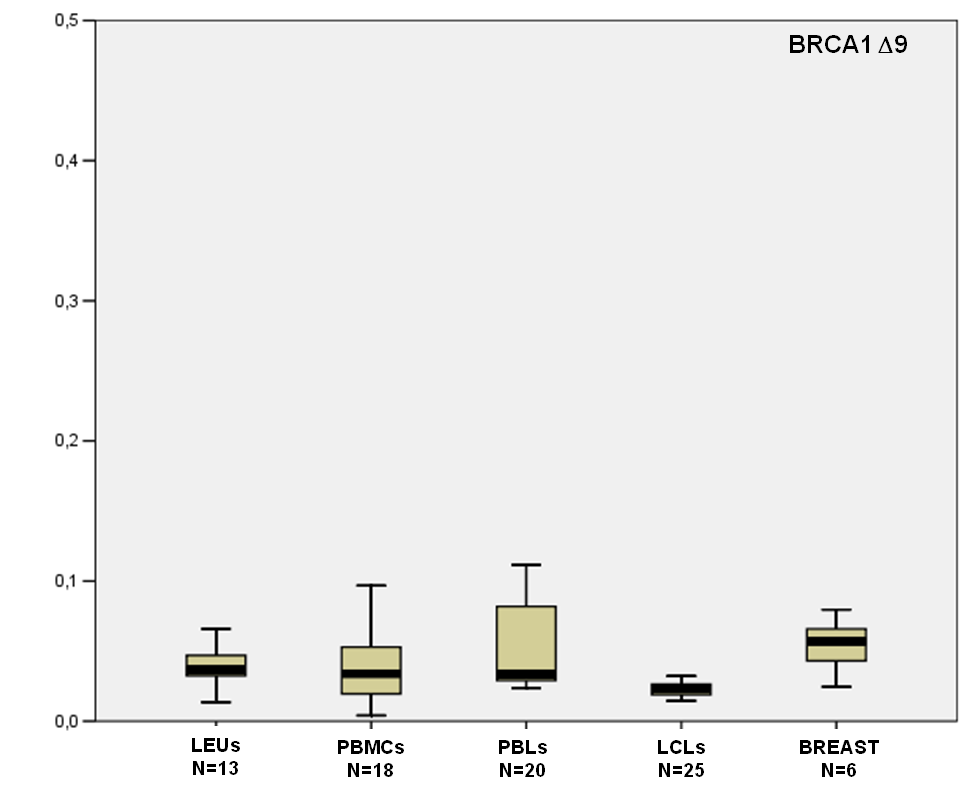
**Supplemental Figure 6.** Relative expression level of *BRCA1* Δ5 in 4 blood related RNA sources and one breast tissue. 33-cycle RT-PCR assays (E3-E8) were analyzed by CE. The boxplot shows the ratios between the peak area of Δ5 and the peak areas of the reference full-length transcript (Low, Q1, Median, Q3, ad High values are indicated). Normal outliers (>1.5 IQR) display a small circle. Extreme outliers (>3 IQR) display a star sign. N indicates the number of individual data-points (different control samples plus technical replicates). In BREAST, N indicates the number of technical replicas performed.



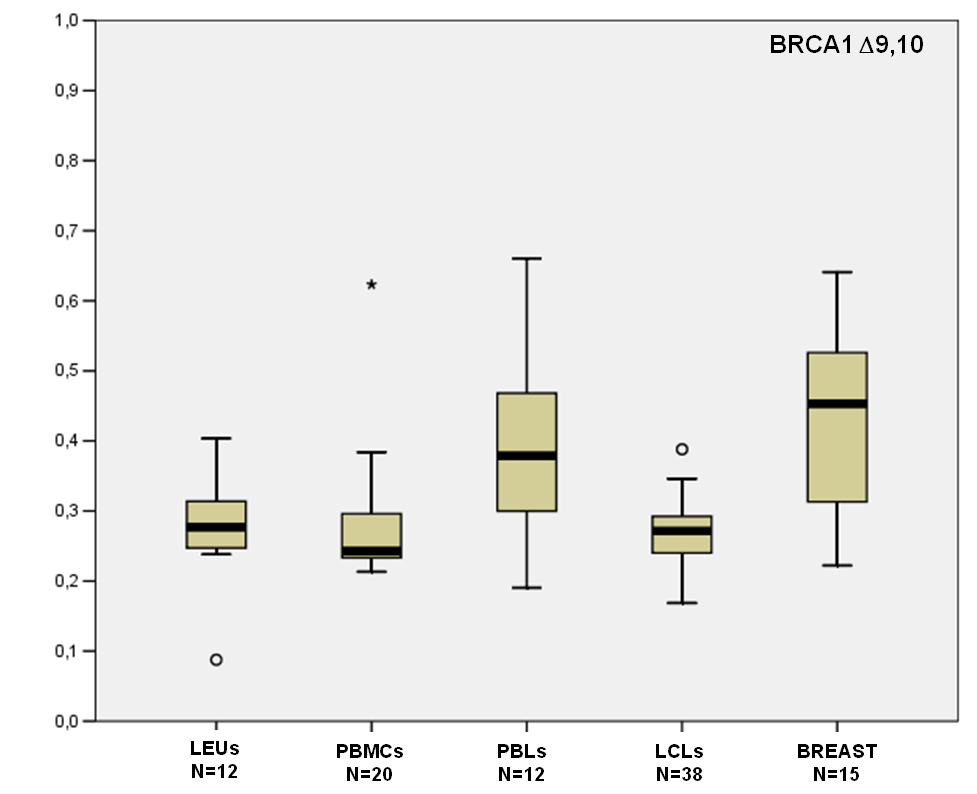
**Supplemental Figure 7.** Relative expression level of *BRCA1* Δ5q in 4 blood related RNA sources and one breast tissue. 33-cycle RT-PCR assays (E3-E8) were analyzed by CE. The boxplot shows the ratios between the peak area of Δ5q and the peak areas of the reference full-length transcript (Low, Q1, Median, Q3, ad High values are indicated). Normal outliers (>1.5 IQR) display a small circle. N indicates the number of individual data-points (different control samples plus technical replicates). In BREAST, N indicates the number of technical replicas performed.



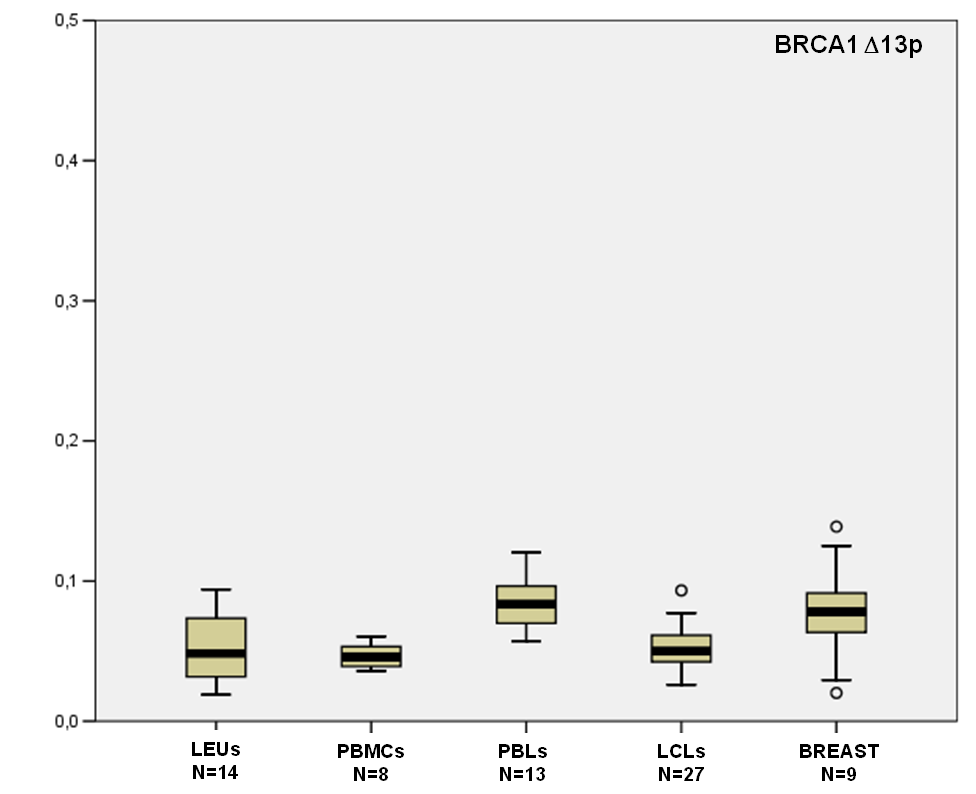
**Supplemental Figure 8.** Relative expression level of *BRCA1* Δ8p in 4 blood related RNA sources and one breast tissue. 33-cycle RT-PCR assays were analyzed by CE. The boxplot shows the ratios between the peak area of Δ8p and the peak areas of the reference full-length transcript (Low, Q1, Median, Q3, ad High values are indicated). Normal outliers (>1.5 IQR) display a small circle. Extreme outliers (>3 IQR) display a star sign. N indicates the number of individual data-points (different control samples plus technical replicates). In BREAST, N indicates the number of technical replicas performed.



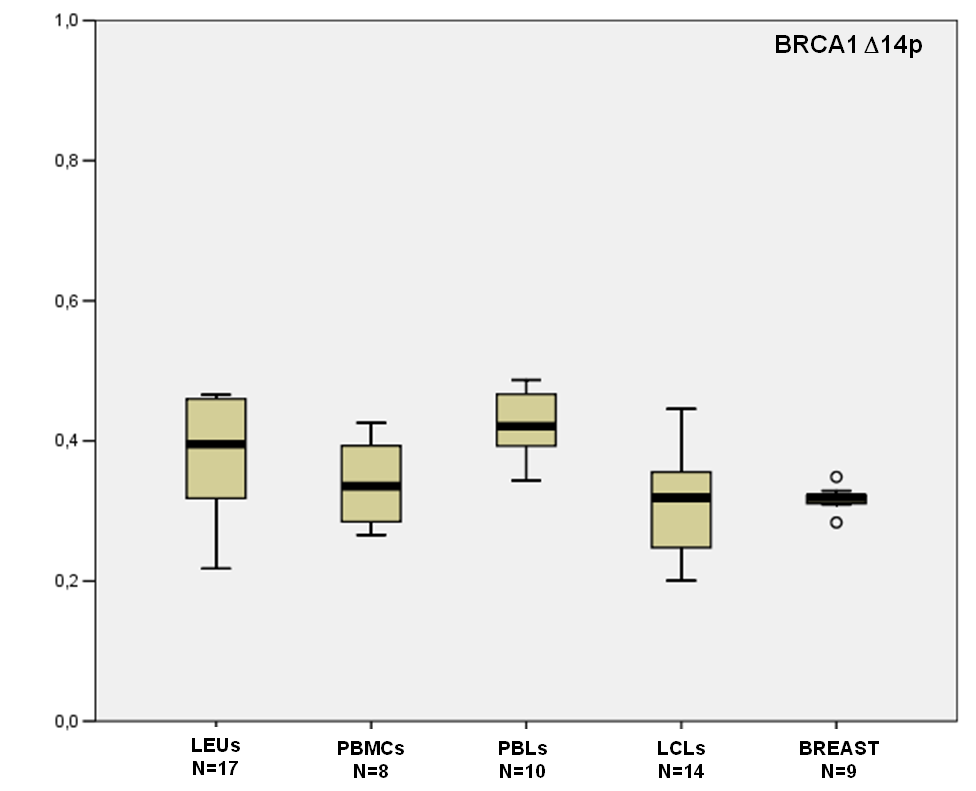
**Supplemental Figure 9.** Relative expression level of *BRCA1* Δ9 in 4 blood related RNA sources and one breast tissue. 33-cycle RT-PCR assays were analyzed by CE. The boxplot shows the ratios between the peak area of Δ9 and the peak areas of the reference full-length transcript (Low, Q1, Median, Q3, ad High values are indicated). N indicates the number of individual data-points (different control samples plus technical replicates). In BREAST, N indicates the number of technical replicas performed.



**Supplemental Figure 10.** Relative expression level of *BRCA1* Δ9,10 in 4 blood related RNA sources and one breast tissue. 33-cycle RT-PCR assays were analyzed by CE. The boxplot shows the ratios between the peak area of Δ9,10 and the peak areas of the reference full-length transcript (Low, Q1, Median, Q3, ad High values are indicated). Normal outliers (>1.5 IQR) display a small circle. Extreme outliers (>3 IQR) display a star sign. N indicates the number of individual data-points (different control samples plus technical replicates). In BREAST, N indicates the number of technical replicas performed.



**Supplemental Figure 11.** Relative expression level of *BRCA1* Δ13p in 4 blood related RNA sources and one breast tissue. 33-cycle RT-PCR assays were analyzed by CE. The boxplot shows the ratios between the peak area of Δ13p and the peak areas of the reference full-length transcript (Low, Q1, Median, Q3, ad High values are indicated). Normal outliers (>1.5 IQR) display a small circle. N indicates the number of individual data-points (different control samples plus technical replicates). In BREAST, N indicates the number of technical replicas performed.



**Supplemental Figure 12.** Relative expression level of *BRCA1* Δ14p in 4 blood related RNA sources and one breast tissue. 33-cycle RT-PCR assays were analyzed by CE. The boxplot shows the ratios between the peak area of Δ14p and the peak areas of the reference full-length transcript (Low, Q1, Median, Q3, ad High values are indicated). Normal outliers (>1.5 IQR) display a small circle. N indicates the number of individual data-points (different control samples plus technical replicates). In BREAST, N indicates the number of technical replicas performed.