Identification of Alternative Splicing Markers for Breast Cancer


Abstract

Breast cancer is the most common cause of cancer death among women under age 50 years, so it is imperative to identify molecular markers to improve diagnosis and prognosis of this disease. Here, we present a new approach for the identification of breast cancer markers that does not measure gene expression but instead uses the ratio of alternatively spliced mRNAs as its indicator. Using a high-throughput reverse transcription-PCR–based system for splicing annotation, we monitored the alternative splicing profiles of 600 cancer-associated genes in a panel of 21 normal and 26 cancerous breast tissues. We validated 41 alternative splicing events that significantly differed in breast tumors relative to normal breast tissues. Most cancer–specific changes in splicing that disrupt known protein domains support an increase in cell proliferation or survival consistent with a functional role for alternative splicing in cancer. In a blind screen, a classifier based on the 12 best cancer-associated splicing events correctly identified cancer tissues with 96% accuracy. Moreover, over a subset of these alternative splicing events could order tissues according to histopathologic grade, and 5 markers were validated in a further blind set of 19 grade 1 and 19 grade 3 tumor samples. These results provide a simple alternative for the classification of normal and cancerous breast tumor tissues and underscore the putative role of alternative splicing in the biology of cancer.

Introduction

Over half a million women die of breast cancer each year and the numbers are expected to increase as the world’s population ages. This fact has stimulated concerted efforts to identify gene expression patterns that could be used for early detection and better prediction of prognosis for breast cancer (1). However, gene expression levels alone cannot fully explain cellular phenotype or function. Moreover, a catalogue of markers for diagnostic and prognostic purposes cannot be considered comprehensive without consideration of alternative pre-mRNA splicing (2). Alternative splicing provides an additional layer of genomic complexity by producing multiple mRNAs and protein variants from any given gene (3). Indeed, 99% of genes contain introns and changes in splicing pattern profoundly affect cell function, independently of changes in the genes’ expression levels (4). Splice variants have been identified for a large variety of cancer genes, suggesting that widespread aberrant and alternative splicing may be a consequence or even a cause of cancer (5). Nevertheless, the biological activity of the majority of alternatively spliced isoforms, and in particular, their contribution to cancer biology, has yet to be elucidated.

There have been many reports of alternative splicing events (ASEs) specific to breast cancer. An early example was Tenascin, which contains a large central alternatively spliced region that can induce focal adhesion of cultured cells and facilitate cell migration (6). In addition, it is known that many of the proteins that influence splicing decisions are up-regulated in breast cancer (7–9). A study of 64 ASEs showed changes in splicing between each of two highly invasive breast cell lines and cultured mammary epithelial cells; ~10 differences in alternative splicing were found between each cancer cell line (10).

Here, we have used a recently developed high-throughput reverse transcription-PCR (RT-PCR)-based platform for splicing annotation, termed the LISA, to examine 600 cancer-associated genes (11). The LISA platform was used to identify breast cancer-associated ASEs with the goal of developing a tissue classifier that reflects the biology of breast cancer. Previously, the LISA approach showed its high sensitivity and fidelity by identifying 48 ovarian cancer–specific ASEs (11). Here, we present the first high-throughput survey of alternative splicing in breast cancer. This screen identified a set of 41 validated markers for breast cancer. The newly identified breast cancer markers partially overlap with the previously described ovarian cancer–specific splicing events, demonstrating the richness of alternative splicing as a source of markers for cancer and suggesting that a subset of splicing events may be general markers for cancer.

Materials and Methods

Tissue selection. Ductal epithelial breast tumors and normal breast samples were obtained as frozen specimens from the Réseau de Recherche sur le Cancer du Fonds de la Recherche en Santé du Québec Biobank. Only chemotherapy-naïve tumor samples were used for the training set. Normal breast samples were obtained either from mammary reductions of healthy individuals or through mastectomies from patients with matched tumor samples (Supplementary Table S1). In addition, 3 samples were obtained 3 to 8 h postmortem after accidental death or death due to cardiovascular disease. Histopathology, grade, and stage were assigned according to the American Joint Commission on Cancer criteria. Tumor and normal tissues were obtained from similar age group patients. The ages ranged from 31 to 83 y for both groups.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/). All original data is available in searchable format at http://palace.lgfus.ca.

Requests for reprints: Sherif Abou Elela, Département de microbiologie et d’infectiologie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada.

J.P. Venables and R. Klinck contributed equally to this work.

Data deposition. Supplementary Table S1 is available on Cancer Research Online (http://cancerres.aacrjournals.org/).


DOI: 10.1158/0008-5472.CAN-08-1769

Downloaded from cancerres.aacrjournals.org on April 13, 2017. © 2008 American Association for Cancer Research.
RNA extraction, RNA quality control, RT-PCR, and capillary electrophoresis were done as previously described for ovary (11) except some normal breast RNA samples had an additional cleaning step (RNeasy; Qiagen). Tissue selection for the discovery screen was established using qPCR of the epithelial cell markers CDH1, the stromal marker Vimentin, and the tumor cell content indicator hTERT (primer sequences available on request). Expression levels of these genes were put through a logistics regression classifier to give a score from 0 (normal) to 1 (cancer; Supplementary Table S1) using the “glm” (Generalised log-linear model package).3 There were no such selection criteria for the blind set.

PCR screen design. The same gene selection, primer design, and analysis were used as in our previous study (11). Reaction design differed from our previous study on ovarian cancer in that only ASEs, as opposed to all splicing events, indicated in the AceView database were targeted. Thus, an average of five PCR reactions were performed per gene.

Quantitative PCR. The gene expression levels of the 41 cancer-specific ASEs were measured in the 4 pooled cDNA samples (2 normal, 2 ductal tumor, 4 patients per pool) used in the discovery screen using SYBR green referenced to 3 validated housekeeping genes (B2M, RPL13A, and PUM1), and data were processed using qBASE (12). This data processing framework allows interrun calibration to correct for run-to-run variations. Reaction efficiencies were used for the cycle-to-quantity transformation. Initial SEs of technical replicates and of standard curve linear regressions were propagated through all calculations. Primers were designed using a script based on Primer3.4 qPCR was also performed to investigate global gene expression of the 7 grade-associated alternative splicing markers in the unknown sample was assigned a class label based on the smallest Euclidean distance from our previous study on ovarian cancer in that only ASEs, as opposed to all splicing events, indicated in the AceView database were targeted. Thus, an average of five PCR reactions were performed per gene.

The same gene selection, primer design, and analysis were used as in our previous study (11). Reaction design differed from our previous study on ovarian cancer in that only ASEs, as opposed to all splicing events, indicated in the AceView database were targeted. Thus, an average of five PCR reactions were performed per gene.

quantitative PCR. The gene expression levels of the 41 cancer-specific ASEs were measured in the 4 pooled cDNA samples (2 normal, 2 ductal tumor, 4 patients per pool) used in the discovery screen using SYBR green referenced to 3 validated housekeeping genes (B2M, RPL13A, and PUM1), and data were processed using qBASE (12). This data processing framework allows interrun calibration to correct for run-to-run variations. Reaction efficiencies were used for the cycle-to-quantity transformation. Initial SEs of technical replicates and of standard curve linear regressions were propagated through all calculations. Primers were designed using a script based on Primer3.4 qPCR was also performed to investigate global gene expression of the 7 grade-associated alternative splicing markers in the same tissues [43 normal and 12 grade 1, 17 grade 2, and 6 grade 3 estrogen receptor (ER) positive (ER+)]. Relative expression was calculated as described above.

Nearest centroid ASE classifier. Tumor/normal class prediction was based on the “nearest-centroid prediction rule” described previously (11). In brief, we defined two average profiles (tumor and normal) as vectors of the average Psi values of the 12 best hits (P < 10−7) based on 37 samples. An unknown sample was assigned a class label based on the smallest Euclidean distance between the sample and either one of the two average profiles.

Results

A comprehensive screen of breast cancer–associated alternative splicing. To explore the potential of ASEs as markers for breast cancer, we examined the splicing pattern of 600 cancer-associated genes in normal and cancerous breast tissues. The 600 candidate genes were identified by a keyword search for “ovarian cancer,” “breast cancer,” and “DNA damage” in the National Center for Biotechnology Information Entrez Gene database. Transcript maps for these genes were downloaded from the April 2007 release of the AceView database (13), and predicted ASEs were identified and catalogued by the LISA. Only these alternative events were considered in this study. PCR primers were designed such that each putative ASE was flanked by at least two independent primer pairs, as described previously (11).

The purpose of the screen was to identify new splicing differences associated with breast cancer. Therefore, to maximize the difference between our normal and tumor samples, we used only high-grade tumors in the discovery stages of the screen. Tissue selection was based on histologic profiling and the expression pattern of tissue markers (Supplementary Table S1). Twenty six high-grade (G2–G3) ductal breast tumor tissues were selected as the training set. At the time of the tumor resection, all patients were chemo-naı¨ve with ages varying between ages 31 and 83 years. Thirteen patients were ER+ and 13 were ER−. No selection for node status or disease stage was done.

After total RNA extraction of each tissue specimen, two consecutive high-throughput RT-PCR screens were performed. The first, the discovery screen, aimed to identify potential breast cancer-associated ASEs, and the second, the validation screen, was used to confirm these events. The discovery screen was conducted using two pools of four normal breast tissues and two pools of four ductal breast cancer tissues. The ASEs that clearly discriminated between the cancer and normal tissue pools were tested in the validation screen of 21 individual normal and 26 individual tumors.

RT-PCR data analysis involved assigning ampiclons to predicted splice variants. The concentrations of each variant were annotated, and the relative abundance of variants was expressed as a percent splicing index (psi or Ψ) calculated as the percentage of the ampiclon concentration of the longest variant relative to the total ampiclon concentration of both long and short variants. Out of 3,327 PCR reactions performed on each of the 4 pools, the discovery screen identified 233 reactions covering 140 potential breast cancer–specific ASEs in 136 genes. The subsequent validation of these events in 47 individual tissues found 41 ASEs in 40 genes, which showed statistically significant differences (P < 0.001) between normal and malignant samples (Supplementary Table S2; Fig. 1). Interestingly, 13 of these events had also been previously identified as ovarian cancer–specific changes (Supplementary Table S2; ref. 11). The direction of the splicing shift between normal and tumor was the same for all 13 events common to breast and ovarian cancers. The validation data for the 12 best markers (P < 10−7) are shown in Supplementary Fig. S1.

It was previously proposed that tissue-specific gene expression and alternative splicing in mouse are largely independent and form two distinct classes of regulatory mechanisms (14). To verify whether this notion also applies to cancer-specific genes in humans, we measured the expression levels of the genes harboring the breast-cancer-specific ASEs in four pools of normal and tumor tissues using quantitative PCR (qPCR). Although expression levels between normal and tumor tissues of these genes varied between extremes of 50-fold up-regulation and 11-fold down-regulation (Supplementary Table S2), no association between the cancer-specific splicing pattern and expression level was observed, indicating that there is no strict link between the regulation of transcription and of splicing in cancer-specific genes.

Validation of breast cancer–associated splicing pattern as a tissue classifier. To test the predictive power of the alternative splicing pattern for breast cancer, the 12 most significant splicing markers from the training set were used to create a classifier based on a nearest-centroid prediction rule (see Materials and Methods). The classifier was then tested on a blind set of 35 malignant breast tumors and 20 normal breast tissue samples (Supplementary Table S2). This set included low-grade tumors as well as tumors of the lobular subtype. All normal samples and 33 of the 35 tumor samples were correctly classified with our alternative splicing classifier.

In parallel, the blind set samples were also tested by qPCR with a three-gene epithelial content classifier used to define the training set (see Materials and Methods). This expression classifier correctly classified all normal samples but classified 11 tumor samples. Of the 35 blind set tumors, 29 had defined histopathologic grade. The alternative splicing classifier correctly assigned 28 of the 29 samples for which grade information was available. The qPCR expression classifier correctly assigned all grade 3 tumors (8 of 8)

3 http://probability.ca/cran/src/contrib/Descriptions/glm.html
4 http://primer3.sourceforge.net/
but was less effective with grade 2 (3 of 10 misclassified) and grade 1 tumors (6 of 11 misclassified; Supplementary Table S1).

Association of ASEs with prognostic variables. To investigate the significance of our markers for tumor biology and their prognostic potential, we interrogated our splicing data from the combined training and blind sets of which 37 tumor samples were ER+ and 19 tumor samples were ER−/C0. Despite the fact that the discovery screen aimed to validate markers that distinguish between normal and cancer tissues, 13 of our markers nevertheless showed a significant difference between ER+ and ER− tumors (P < 0.05). These are represented as a box plot in Supplementary Fig. S2. The five most significant markers, FANCA, PLD1, POLB, GPR137, and RUNX2, showed a distinct trend toward exon loss (lower Ψ) in ER− tissues.

We also found 12 significant markers that individually correlate with histopathologic grade (Supplementary Fig. S3A). However, as expected (15), there was a strong correlation in our samples between ER negativity and higher histopathologic grade (Supplementary Table S1). To determine whether our cancer-specific ASEs could predict tumor grade, independently of ER status, we looked within our ER+ tumors of known grades (12, 17, and 6 tumor specimensof grades 1, 2, and 3, respectively). Seven ASEs showed a statistically significant (P < 0.05) shift between the 12 grade 1 and 6 grade 3 tumors (Supplementary Fig. S3B). Interestingly, qPCR in the same tissues (43 normal, 12 grade 1, 17 grade 2, and 6 grade 3) showed no correlation between the expression level of the genes (exhibiting grade-specific splicing differences) and tumor grade (Supplementary Fig. S3C).

As the training set and the blind set of ER+ tumors combined unexpectedly identified the putative grade markers, we studied an extra blind set of 19 more grade 1 and 19 more grade 3 tumors.

![Figure 1. Heat map showing the best 41 ASEs differentiate the normal and cancer samples of the training set. Forty-one alternative splice events are shown on the X-axis ordered from left to right by increasing Ψ shift in cancer. Y-axis shows the normal and cancer tissues ordered by unsupervised clustering. N, normal; C, cancer. The normal and cancer samples segregate. Scaled Ψ values are shown for each measurement, color coded according to the scale (blue, high Ψ, e.g., high exon incorporation relative to the mean inclusion for all the samples). White squares, major detected peaks were below 5 nmol/L as measured by capillary electrophoresis.](image-url)
(detailed in Supplementary Table S1) to ascertain if we had discovered genuine markers for tumor de-differentiation. Of the seven putative markers, five were confirmed in this further large set (Supplementary Fig. S3D). These five markers, POLB, GPR137, RUNX2, PCSK6, and BCAS1 all differentiated between the combined 31 grade 1 and 25 grade 3 ductal ER+ tumors with \( P \) values of <0.001 (Fig. 2).

We examined additional potential correlations between splicing and other tumor features. There were relatively weak correlations between splicing and progesterone receptor and Her2 status as well as with tumor size and type (Supplementary Figs. S4–7). In contrast, we found a strong correlation between the alternative splicing pattern of the insulin receptor (INSR) mRNA and node status (Supplementary Fig. S8). This splice has been reported to be predictive of metastasis and poor outcome (16).

**Discussion**

In this study, we present the first large-scale screen of breast cancer-associated ASEs. Our results show that ASEs can act as independent markers of breast cancer and suggest that consideration of alternative splicing will greatly increase the number of markers that can currently be identified by standard expression profiling alone. Analysis of the alternative splicing of only 600 cancer-associated genes revealed 41 ductal breast cancer-specific markers. Five of the ASEs that associate with cancer are capable of differentiating between different cancer grades. Thus, the analysis of alternative splicing provides information about the biology of the tumor that complements global expression profiling.

**Biological significance of cancer-associated splicing patterns.** Previously, many ASEs had been serendipitously observed in cancer and the function of these splicing events was invariably consistent with alternative splicing playing an active role in cancer (5, 17). To discuss the likely effects of alternative splicing in breast cancer, the encoded protein isoforms of the cancer and normal splice forms are shown in Supplementary Fig. S9. Thirty-nine of these 41 events are predicted to alter the coding region and the protein products. Of these, 19 disrupt known functionally critical protein domains (Table 1). Counter-intuitively, in 12 of these cases, the disruption by splicing seems to occur in normal cells, not in cancer cells (first 12 ASEs; Table 1), consistent with earlier observations of high rates of novel splice forms in normal tissues (18). Surprisingly, most of these genes have roles in promoting cell growth, whereas for the other seven cases where alternative splicing causes disruption in cancer cells, most of these genes are tumor suppressors or DNA damage response genes (see discussion below). Therefore, these combined observations are strongly suggestive of functional selection in cancer for oncogenic splice variants, and against tumor suppressing and checkpoint control splice variants.

As stated above, of our 41 splicing events, 12 likely produce nonfunctional molecules by loss of known protein domains in normal breast tissues (first 12 ASEs in Table 1). For example, we found the full-length form of MCL1, which is an antiapoptotic member of the Bcl family of cell death regulators, is present in cancer, whereas a COOH-terminally truncated, apoptotic form (19) predominates in normal tissues. Similarly, we found the oncogenic (20) and antiapoptotic (21) protein DNMT3b with its methylase domain removed, predominantly in normal cells. The adaptor protein SHC1 promotes cell proliferation (22) and it, too, is more fully spliced in cancer; in normal cells, we found more of a splice variant with a deletion in the phosphotyrosine-binding domain. DBF4B (also known as ASKL1) drives cell cycle progression (23), and it is truncated by alternative splicing in normal breast but it is fully spliced in breast cancer. The chemokine CCL4 (MIP1b), which is involved in invasion of cancer cells (24), is severely disrupted by alternative splicing in normal tissues only, as is the cell surface receptor CD40, which is thought to promote neoplastic growth (25). Tissue factor (F3) promotes tumor invasion (26), and it loses its transmembrane and intracellular domain by alternative splicing (27) preferentially in normal breast but not in tumors.

HMGA1 is a small DNA binding protein causally related to neoplastic transformation that enhances the invasiveness of cancer cells (28), and we also found that all its exons were incorporated in breast cancer, whereas omission of exon 7 removes an AT-hook DNA-binding domain in normal breast. FGFR2 has been suggested to be a transforming oncogene, and the longer form is found in cancer, whereas it more often lacks the first extracellular domain in normal breast, consistent with a possible role of this region in breast cancer (29). PTPRB is a tyrosine phosphatase that also lacks an extracellular region in normal cells. The mitotic Tubulin A4A is shorter in normal than in cancer cells. Finally, cancer cells express more of the fibronectin form containing IIICS, a region known to be important for cell recognition (30).

In contrast to the above examples, genes that are fully functional in normal breast and compromised in cancer form a different functional group with growth regulating or tumor suppressor properties (bottom 7 ASEs in Table 1). For example, PAXIP1 (PTIP) binds to p53 and is involved in the maintenance of genome stability. It is totally disrupted by alternative splicing in breast cancer but not in normal breast. DNA polymerase β (POLB), which is also required for DNA maintenance, is also inactivated by alternative splicing in breast cancer. DNA ligase (LIG3) is involved in excision repair, and the NH₂ terminus is absent due to alternative splicing in cancerous but not normal breast. DSC3 is down-regulated in breast cancer and this is thought to increase...
and alternative gene expression events. For example, one report detected methods used or to a lack of distinction between ASEs. Discrepancies may be due to differences in tumor subtypes or in either our tumors or normal tissues. The main reasons for these cases involve cancer-specific exon loss that we did not detect. CLIP1 down-regulation also promotes invasion of breast cancer cells (32), and breast cancer splicing causes removal of most of its central chromosome segregation domain. Overall, the directionality of splicing changes strongly suggest that alternative splicing can provide breast tumors with protein function necessary for their survival and expansion.

Even where alternative splicing does not affect known protein domains, the splicing differences are likely to affect tumor biology. For example, we found that the short form of the hyaluronan-mediated motility receptor, HMMR (RHAMM), preferentially accumulates in breast tumors. Although the altered function of this isoform has not been elucidated, this splicing event predicts the DDR1 tyrosine kinase in breast cancer tissues. The incorporated) in whole tissues (42), whereas another study used boundary-spanning primers rather than competitive PCR (43). In other cases, short isoforms were identified in breast tumor samples using antibodies (39, 40), which may reflect altered protein stability rather than alternative splicing. This underscores the need for establishing clear guidelines for the annotation of cancer-specific splicing events and shows the utility of standardized and validated systems like the LISA for detection.

We were able to confirm two ASEs previously associated with breast cancer. The first involves the up-regulation of the INSR isoform lacking exon 11 in tumor tissues (44). This isoform is found in normal fetal tissue and binds insulin-like growth factor as well as insulin and has been proposed to contribute to altered signaling in cancer (45). Interestingly, the expression of this short isoform correlates with node status affecting metastasis (Supplementary Fig. S8). Indeed, this splicing isoform has been reported to mediate cell migration/invasion effects of insulin-like growth factor II on cancer cells (45). In addition, we confirmed the cancer-specific splicing pattern of the mRNA encoding the very low density lipoprotein receptor, VLDLR. The VLDLR alternative splicing pattern may be due to the epithelial origin of cancer cells (46).

Overall, the main distinction between previous reports of cancer-specific exon exclusion and the information generated by our screen is that many of the splicing events identified by the LISA (18 of 41) were specific exon inclusions in breast cancer tissues (Supplementary Table S2). This indicates that breast cancer-specific changes in alternative splicing are not restricted to splicing defects resulting in loss of protein functions but may also include modifications that generate proteins with new functions.

### Table 1. Functional consequences of alternative splicing in breast cancer

<table>
<thead>
<tr>
<th>Gene/protein</th>
<th>Cancer-relevant gene functional tendency</th>
<th>Type of change to reading frame</th>
<th>AS removes % of protein</th>
<th>Domain(s) removed by alternative splicing</th>
<th>Fully spliced (all exons incorporated) in</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD40</td>
<td>Pro-proliferative</td>
<td>Frameshift</td>
<td>95</td>
<td>Cytokine receptor domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>TUBA4A</td>
<td>Mitosis</td>
<td>Frameshift</td>
<td>80</td>
<td>Tubulin domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>CCL4</td>
<td>Proinvasive</td>
<td>Frameshift</td>
<td>70</td>
<td>Cytokine domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>DBF4B</td>
<td>Pro-proliferative</td>
<td>Frameshift</td>
<td>60</td>
<td>Zinc finger domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>MCL1</td>
<td>Antiapoptotic</td>
<td>Frameshift</td>
<td>30</td>
<td>Bcl2 domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>F3</td>
<td>Proinvasive</td>
<td>Frameshift</td>
<td>30</td>
<td>Cytokine receptor domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>HMGA1</td>
<td>Proneoplastic</td>
<td>Ins/del</td>
<td>20</td>
<td>DNA-binding domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>DNMT3B</td>
<td>Antiapoptotic oncogenic</td>
<td>Ins/del</td>
<td>10</td>
<td>Methylase domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>FGR2</td>
<td>Proinvasive mitosis and differentiation</td>
<td>Ins/del</td>
<td>10</td>
<td>Ig-like domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>SHC1</td>
<td>Pro proliferative</td>
<td>Ins/del</td>
<td>5</td>
<td>Phosphotyrosine-binding domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>PTPRB</td>
<td>Cell signaling</td>
<td>Ins/del</td>
<td>5</td>
<td>Fibronectin-like domain</td>
<td>Cancer</td>
</tr>
<tr>
<td>FN1</td>
<td>Migration and metastasis</td>
<td>Ins/del</td>
<td>4</td>
<td>IVICS</td>
<td>Cancer</td>
</tr>
<tr>
<td>POBL</td>
<td>DNA maintenance</td>
<td>Frameshift</td>
<td>95</td>
<td>Polymerase domain</td>
<td>Normal</td>
</tr>
<tr>
<td>PAXIP1</td>
<td>Genome stability</td>
<td>Frameshift</td>
<td>90</td>
<td>BRCA1 C terminal domain</td>
<td>Normal</td>
</tr>
<tr>
<td>BTC</td>
<td>Growth factor</td>
<td>Ins/Del</td>
<td>30</td>
<td>EGF-like domain</td>
<td>Normal</td>
</tr>
<tr>
<td>DSC3</td>
<td>Tumor suppressor cell adhesion</td>
<td>Frameshift</td>
<td>10</td>
<td>Cadherin domain</td>
<td>Normal</td>
</tr>
<tr>
<td>CLIP1</td>
<td>Tumor suppressor cytoskeleton</td>
<td>Ins/del</td>
<td>10</td>
<td>Chromosome segregation domain</td>
<td>Normal</td>
</tr>
<tr>
<td>PCSK6</td>
<td>Proinvasive</td>
<td>Ins/del</td>
<td>5</td>
<td>Protease-associated domain</td>
<td>Normal</td>
</tr>
<tr>
<td>NOTCH3</td>
<td>Pro-proliferative</td>
<td>Ins/del</td>
<td>2</td>
<td>EGF-like domain</td>
<td>Normal</td>
</tr>
</tbody>
</table>

**Column 1.** 19 markers for which the alternative splice event alters a known functional domain. **Column 2.** The cancer-relevant function of the corresponding protein. **Columns 3 to 5** list the effect of alternative splicing on the reading frame: (3) whether it is truncated by a frameshift or undergoes an internal insertion or deletion (ins/del); (4) the approximate proportion of the protein removed in the short form; and (5) the known protein domains affected. The final column indicates whether breast tumors or normal breast has an increased amount of the full-length (likely the fully functional) form.
Identification of potential generic cancer-specific splicing events. Thirteen of the 41 breast cancer–specific ASEs found in this study were previously linked to serous ovarian cancer tissues (Supplementary Table S2; ref. 11) raising the possibility that a subset of splicing events may be common to a wide variety of cancer types. Indeed, 6 of the 12 most significant breast cancer–associated splicing events are also among the most discriminating ovarian cancer–associated events (11). Five of these marker ASEs (STIM1, SYNE2, APP, PLD1, and PAXIP1) are in the top six ovarian cancer–specific alternative splicing markers. Interestingly, several of these genes are known effectors of well-established tumorigenic pathways. For example, STIM1 has been implicated in breast cancer and rhabdomyosarcoma (47), and has recently emerged as a critical regulator of intracellular calcium (48). Another of the common markers Betacellulin (BTC) is one of the epidermal growth factor (EGF) family of ligands, which bind and activate the ERBB/EGF receptor family of receptors (49).

Although important functions can be inferred for all the common splicing events found in both ovarian and breast cancers, it is clear that establishing these splicing events as general markers of cancer will require further studies. They may only be common to a subset of cancer types or may simply reflect the epithelial origin of ovarian and breast cancers. However, the results presented here provide a defined set of splicing events that could be linked functionally to the general mechanism of tumorigenesis. Screening in a wide variety of cancer types will be required to confirm the status of one or more splicing events as general cancer markers.

Alternative splicing and ER status. Of particular relevance to breast cancer biology, we identified seven ASEs whose changes in splicing correlate with the ER status of the tumors. These genes differ from reported ER markers identified by expression profiling (50). The best ER-sensitive splicing event was found in FANCA, which is regulated by steroid hormone (51). Two other ER-sensitive ASEs were found in MCL-1 and CCL4 that are tamoxifen and estrogen target genes, respectively (52, 53). Notably, it was recently shown that estrogen, ER, and its coregulators may affect the alternative splicing of their target genes (54). Thus, it is tempting to propose that estrogen may directly or indirectly affect the splicing of several breast-specific ASEs.

Alternative splicing and tumor grade. We identified five novel ASEs exhibiting significant differences between tumor grades 1 and 3 (Fig. 2). These splicing events did not occur in genes whose expression level is linked to tumor grade, confirming the apparent lack of intersection between cancer-specific alternative splicing and expression markers (14). This specific association of splicing changes with grade suggests that multiple changes in alternative splicing may be involved in the de-differentiation of tumors. This is consistent with the fact that the splicing factor kinase SRPK1 is up-regulated in breast cancer and in higher tumor grades (55). Because tumor grade is a prognostic indicator for breast cancer, the results presented here show the potential of alternative splicing to guide therapeutic strategies.

The sensitivity and specificity of our LISA approach has revealed a vast yet largely unexplored domain of cancer biology. The present study suggests that >7% of cancer-associated genes harbor disease-specific ASEs, which could be important novel biomarkers and therapeutic targets. In addition, subsets of these events, initially selected for their ability to distinguish cancer from normal tissue, also reflect ER, node status, and tumor grade. Further studies specifically designed to identify alternative splicing markers that reflect distinct breast cancer biology with relation to clinical outcomes and prognoses show promise to improve our understanding of breast cancer at the molecular level.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.


## Identification of Alternative Splicing Markers for Breast Cancer

Julian P. Venables, Roscoe Klinck, Anne Bramard, et al.


<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: <a href="http://cancerres.aacrjournals.org/content/68/22/9525">http://cancerres.aacrjournals.org/content/68/22/9525</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://cancerres.aacrjournals.org/content/suppl/2008/11/12/68.22.9525.DC1">http://cancerres.aacrjournals.org/content/suppl/2008/11/12/68.22.9525.DC1</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 55 articles, 19 of which you can access for free at: <a href="http://cancerres.aacrjournals.org/content/68/22/9525.full.html#ref-list-1">http://cancerres.aacrjournals.org/content/68/22/9525.full.html#ref-list-1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citing articles</td>
<td>This article has been cited by 31 HighWire-hosted articles. Access the articles at: <a href="http://cancerres.aacrjournals.org/content/68/22/9525.full.html#related-urls">http://cancerres.aacrjournals.org/content/68/22/9525.full.html#related-urls</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-mail alerts</th>
<th>Sign up to receive free email-alerts related to this article or journal.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reprints and Subscriptions</td>
<td>To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a>.</td>
</tr>
<tr>
<td>Permissions</td>
<td>To request permission to re-use all or part of this article, contact the AACR Publications Department at <a href="mailto:permissions@aacr.org">permissions@aacr.org</a>.</td>
</tr>
</tbody>
</table>