

# Identification of Alternative Splicing Markers for Breast Cancer

Julian P. Venables,<sup>1</sup> Roscoe Klinck,<sup>1</sup> Anne Bramard,<sup>1</sup> Lyna Inkel,<sup>1</sup> Geneviève Dufresne-Martin,<sup>1</sup> ChuShin Koh,<sup>1</sup> Julien Gervais-Bird,<sup>1</sup> Elvy Lapointe,<sup>1</sup> Ulrike Froehlich,<sup>1</sup> Mathieu Durand,<sup>1</sup> Daniel Gendron,<sup>1</sup> Jean-Philippe Brosseau,<sup>1</sup> Philippe Thibault,<sup>1</sup> Jean-Francois Lucier,<sup>1</sup> Karine Tremblay,<sup>1</sup> Panagiotis Prinios,<sup>1</sup> Raymund J. Wellinger,<sup>2</sup> Benoit Chabot,<sup>2</sup> Claudine Rancourt,<sup>2</sup> and Sherif Abou Elela<sup>2</sup>

<sup>1</sup>Laboratoire de génomique fonctionnelle de l'Université de Sherbrooke and <sup>2</sup>Département de microbiologie et d'infectiologie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada

## 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, 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. [Cancer Res 2008;68(22):9525–31]**

## 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 gene 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 mastectomy 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>.

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

**Requests for reprints:** Sherif Abou Elela, Département de microbiologie et d'infectiologie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, 3001, 12e Avenue Nord, Sherbrooke, Québec J1H 5N4, Canada. Phone: 819-564-5275; Fax: 819-564-5392; E-mail: [sherif.abou.elela@usherbrooke.ca](mailto:sherif.abou.elela@usherbrooke.ca).

©2008 American Association for Cancer Research.

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

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).<sup>3</sup> 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 (*RPL13A*, *B2M*, 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.<sup>4</sup> 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 ( $P < 10^{-5}$ ) 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-naïve 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 amplicons 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  $\Psi$ ) calculated as the percentage of the amplicon concentration of the longest variant relative to the total amplicon 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 ( $P < 10^{-5}$ ) 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 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.

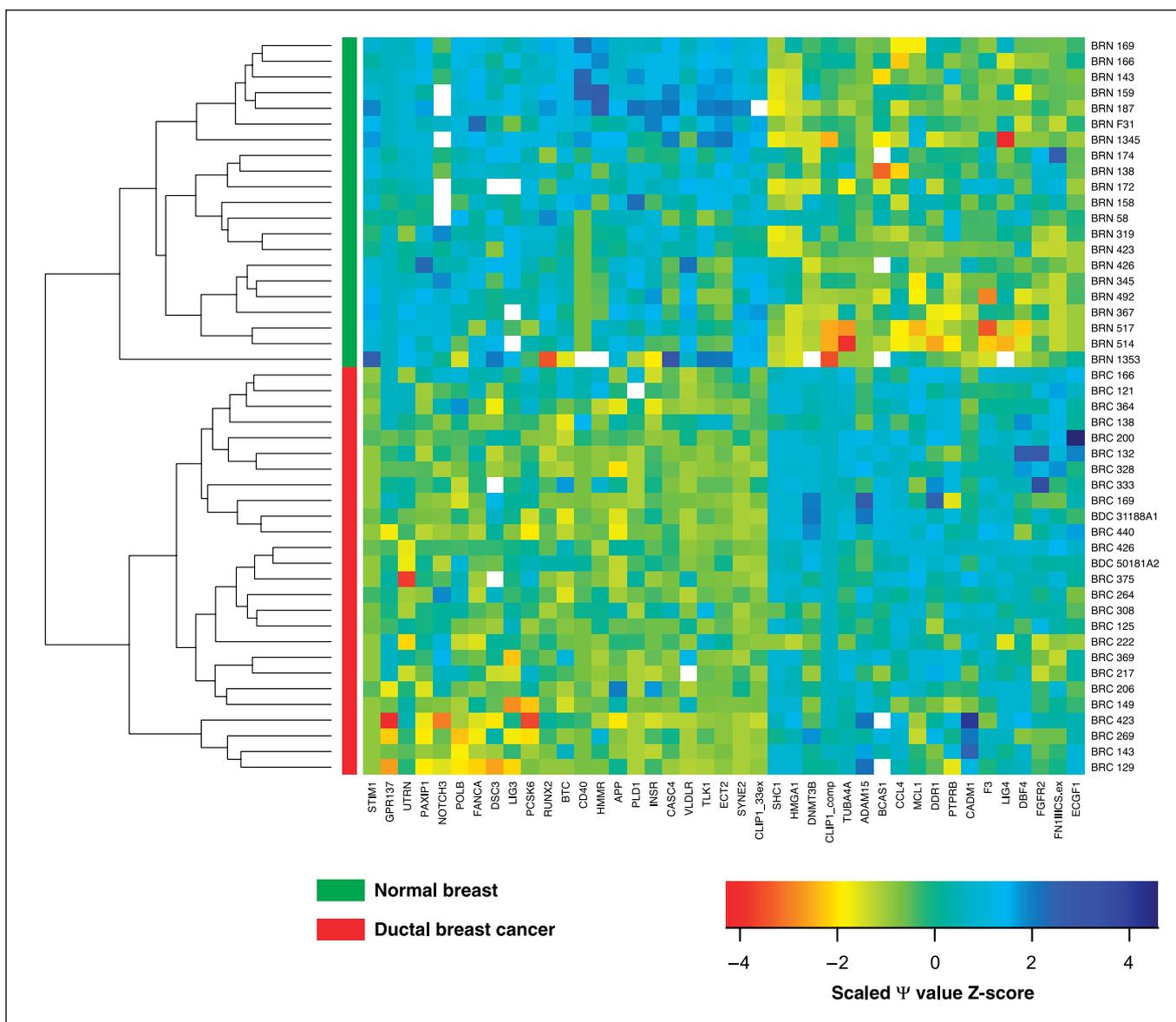
**Validation of breast cancer-associated splicing pattern as a tissue classifier.** To test the predictive power of the alternative splicing pattern for breast cancer, the 12 most significant splicing markers from the training set were used to create a classifier based on a nearest-centroid prediction 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 misclassified 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)

<sup>3</sup> <http://probability.ca/cran/src/contrib/Descriptions/glm.html>

<sup>4</sup> <http://primer3.sourceforge.net/>



**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  $\Psi$  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  $\Psi$  values are shown for each measurement, color coded according to the scale (*blue*, high  $\Psi$ , 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.

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-. 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  $\Psi$ ) 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 specimens of 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

(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.

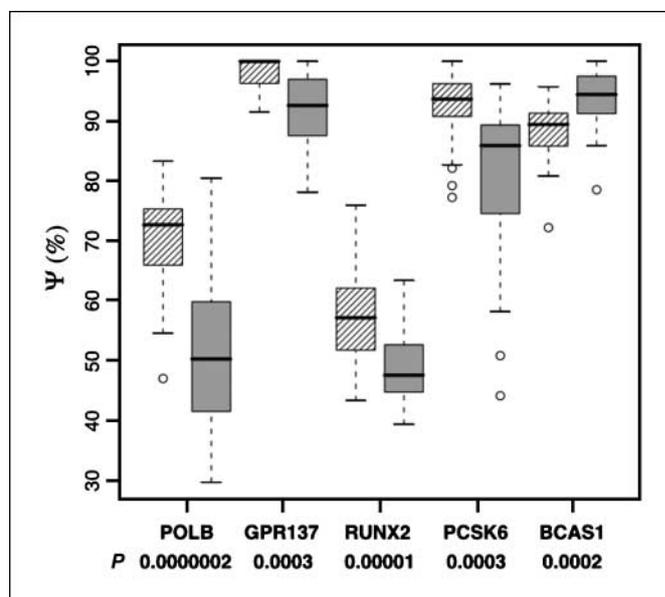
**Biological significance of cancer-associated splicing patterns.** Previously, many ASEs had been serendipitously observed in cancer and the function of these splicing events was invariably consistent with alternative splicing playing an active role in cancer (5, 17). To discuss the likely effects of 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.

HMGAI 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 IIIICS, 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  $\beta$  (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  $\text{NH}_2$  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



**Figure 2.** Five validated grade markers. Boxplot showing the  $\Psi$  distribution of the 5 grade markers from a total of 31 grade 1 and 25 grade 3 ER+ tumors from the combined data from the validation set, the blind set, and the extra blind set. Diagonally lined boxes, grade 1 ER+ tumors; gray boxes, grade 3 tumors. The strength of the markers are indicated as  $P$  values.

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

Gene/protein	Cancer-relevant gene functional tendency	Type of change to reading frame	AS removes % of protein	Domain(s) removed by alternative splicing	Fully spliced (all exons incorporated) in
CD40	Pro-proliferative	Frameshift	95	Cytokine receptor domain	Cancer
TUBA4A	Mitosis	Frameshift	80	Tubulin domain	Cancer
CCL4	Proinvasive	Frameshift	70	Cytokine domain	Cancer
DBF4B	Pro-proliferative	Frameshift	60	Zinc finger domain	Cancer
MCL1	Antiapoptotic	Frameshift	30	Bcl2 domain	Cancer
F3	Proinvasive	Frameshift	30	Cytokine receptor domain	Cancer
HMGA1	Proneoplastic	Ins/del	20	DNA-binding domain	Cancer
DNMT3B	Antiapoptotic oncogenic	Ins/del	10	Methylase domain	Cancer
FGFR2	Proinvasive mitosis and differentiation	Ins/del	10	Ig-like domain	Cancer
SHC1	Proproliferative	Ins/del	5	Phosphotyrosine-binding domain	Cancer
PTPRB	Cell signaling	Ins/del	5	Fibronectin-like domain	Cancer
FN1	Migration and metastasis	Ins/del	4	IIICS	Cancer
POLB	DNA maintenance	Frameshift	95	Polymerase domain	Normal
PAXIP1	Genome stability	Frameshift	90	BRCA1 C terminal domain	Normal
BTC	Growth factor	Ins/Del	30	EGF-like domain	Normal
DSC3	Tumor suppressor cell adhesion	Frameshift	10	Cadherin domain	Normal
CLIP1	Tumor suppressor cytoskeleton	Ins/del	10	Chromosome segregation domain	Normal
PCSK6	Proinvasive	Ins/del	5	Protease-associated domain	Normal
NOTCH3	Pro-proliferative	Ins/del	2	EGF-like domain	Normal

*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.

tumor cell migration (31). Consistent with this model, a frameshift removes the COOH-terminal cadherin domain from DSC3 and potentially limits its function in breast cancer. 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 poor prognosis in multiple myeloma (33). Likewise, we also identified the preferential inclusion of a 111 nucleotide exon of the DDR1 tyrosine kinase in breast cancer tissues. The incorporation of 37 amino acids in this long isoform alters the autophosphorylation and glycosylation activities of kinase (34). Together, these observations suggest that many cancer-associated splicing isoforms may be functionally relevant to tumor biology.

**Observation of reported breast cancer-associated splicing events.** Alternative splicing of at least 15 genes was previously reported to occur in breast cancer (e.g., refs. 35–45). The majority of these cases involve cancer-specific exon loss that we did not detect in either our tumors or normal tissues. The main reasons for these discrepancies may be due to differences in tumor subtypes or detection methods used or to a lack of distinction between ASEs and alternative gene expression events. For example, one report

used purified epithelial cells from mammoplasties rather than 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.

**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 (EGFR) 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 can 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.

## Acknowledgments

Received 5/12/2008; revised 8/28/2008; accepted 9/11/2008.

**Grant support:** Genome Canada and Genome Quebec. B. Chabot is the Canada Research Chair in Functional Genomics. S.A. Elela is a Chercheur-Boursier Senior of the Fonds de la Recherche en Santé du Québec (FRSQ). Tumor banking was supported by the Banque de tissus et de données of the Réseau de Recherche sur le Cancer of the FRSQ.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

## References

- Cheang MC, van de Rijn M, Nielsen TO. Gene expression profiling of breast cancer. *Annu Rev Pathol* 2007;3:67–97.
- Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM. Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol* 2007;8:349–57.
- Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell* 2006;126:37–47.
- Moore MJ, Silver PA. Global analysis of mRNA splicing. *RNA* 2008;14:197–203.
- Venables JP. Aberrant and alternative splicing in cancer. *Cancer Res* 2004;64:7647–54.
- Borsi L, Carnemolla B, Nicolo G, Spina B, Tanara G, Zardi L. Expression of different tenascin isoforms in normal, hyperplastic and neoplastic human breast tissues. *Int J Cancer* 1992;52:688–92.
- Huang CS, Shen CY, Wang HW, Wu PE, Cheng CW. Increased expression of SRp40 affecting CD44 splicing is associated with the clinical outcome of lymph node metastasis in human breast cancer. *Clin Chim Acta* 2007;384:69–74.
- Skotheim RI, Nees M. Alternative splicing in cancer: noise, functional, or systematic? *Int J Biochem Cell Biol* 2007;39:1432–49.
- Watermann DO, Tang Y, Zur Hausen A, Jager M, Stamm S, Stickeler E. Splicing factor Tra2-1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene. *Cancer Res* 2006;66:4774–80.
- Li C, Kato M, Shiue L, Shively JE, Ares M, Jr., Lin RJ. Cell type and culture condition-dependent alternative splicing in human breast cancer cells revealed by splicing-sensitive microarrays. *Cancer Res* 2006;66:1990–9.
- Klinck R, Bramard A, Inkel L, et al. Multiple alternative splicing markers for ovarian cancer. *Cancer Res* 2008;68:657–63.
- Hellems J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 2007;8:R19.
- Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* 2006;7 Suppl 1:S12 1–4.
- Pan Q, Shai O, Misquitta C, et al. Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol Cell* 2004;16:929–41.
- Fisher ER, Osborne CK, McGuire WL, et al. Correlation of primary breast cancer histopathology and estrogen receptor content. *Breast Cancer Res Treat* 1981;1:37–41.
- Belfiore A. The role of insulin receptor isoforms and hybrid insulin/IGF-1 receptors in human cancer. *Curr Pharm Des* 2007;13:671–86.
- Venables JP. Unbalanced alternative splicing and its significance in cancer. *Bioessays* 2006;28:378–86.
- Xu Q, Lee C. Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. *Nucleic Acids Res* 2003;31:5635–43.
- Bae J, Leo CP, Hsu SY, Hsueh AJ. MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. *J Biol Chem* 2000;275:25255–61.
- Soejima K, Fang W, Rollins BJ. DNA methyltransferase 3b contributes to oncogenic transformation induced by SV40T antigen and activated Ras. *Oncogene* 2003;22:4723–33.
- Beaulieu N, Morin S, Chute IC, Robert MF, Nguyen H, MacLeod AR. An essential role for DNA methyltransferase DNMT3B in cancer cell survival. *J Biol Chem* 2002;277:28176–81.

22. Veeramani S, Igawa T, Yuan TC, et al. Expression of p66(Shc) protein correlates with proliferation of human prostate cancer cells. *Oncogene* 2005;24:7203-12.
23. Yoshizawa-Sugata N, Ishii A, Taniyama C, Matsui E, Arai K, Masai H. A second human Dbf4/ASK-related protein, Drf1/ASKL1, is required for efficient progression of S and M phases. *J Biol Chem* 2005;280:13062-70.
24. Saito S, Kitayama J, Jin ZX, et al.  $\beta$ -chemokine, macrophage inflammatory protein-1 $\beta$  (MIP-1 $\beta$ ), is highly expressed in diffuse type human gastric cancers. *J Exp Clin Cancer Res* 2003;22:453-9.
25. Baxendale AJ, Dawson CW, Stewart SE, et al. Constitutive activation of the CD40 pathway promotes cell transformation and neoplastic growth. *Oncogene* 2005;24:7913-23.
26. Nitori N, Ino Y, Nakanishi Y, et al. Prognostic significance of tissue factor in pancreatic ductal adenocarcinoma. *Clin Cancer Res* 2005;11:2531-9.
27. Bogdanov VY, Balasubramanian V, Hathcock J, Vele O, Lieb M, Nemerson Y. Alternatively spliced human tissue factor: a circulating, soluble, thrombogenic protein. *Nat Med* 2003;9:458-62.
28. Liau SS, Jazag A, Whang EE. HMGA1 is a determinant of cellular invasiveness and *in vivo* metastatic potential in pancreatic adenocarcinoma. *Cancer Res* 2006;66:11613-22.
29. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007;39:870-4.
30. White ES, Baralle FE, Muro AF. New insights into form and function of fibronectin splice variants. *J Pathol* 2008;216:1-14.
31. Oshiro MM, Kim CJ, Wozniak RJ, et al. Epigenetic silencing of DSC3 is a common event in human breast cancer. *Breast Cancer Res* 2005;7:R669-80.
32. Suzuki K, Takahashi K. Regulation of lamellipodia formation and cell invasion by CLIP-170 in invasive human breast cancer cells. *Biochem Biophys Res Commun* 2008;368:199-204.
33. Maxwell CA, Rasmussen E, Zhan F, et al. RHAMM expression and isoform balance predict aggressive disease and poor survival in multiple myeloma. *Blood* 2004;104:1151-8.
34. Alves F, Vogel W, Mossie K, Millauer B, Hofler H, Ullrich A. Distinct structural characteristics of discoidin I subfamily receptor tyrosine kinases and complementary expression in human cancer. *Oncogene* 1995;10:609-18.
35. Debily MA, Camarca A, Ciullo M, et al. Expression and molecular characterization of alternative transcripts of the ARHGEF5/TIM oncogene specific for human breast cancer. *Hum Mol Genet* 2004;13:323-34.
36. Radisky DC, Levy DD, Littlepage LE, et al. Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. *Nature* 2005;436:123-7.
37. Datta D, Flaxenburg JA, Laxmanan S, et al. Ras-induced modulation of CXCL10 and its receptor splice variant CXCR3-B in MDA-MB-435 and MCF-7 cells: relevance for the development of human breast cancer. *Cancer Res* 2006;66:9509-18.
38. Hisatomi H, Kohno N, Wakita K, et al. Novel alternatively spliced variant with a deletion of 52 BP in exon 6 of the progesterone receptor gene is observed frequently in breast cancer tissues. *Int J Cancer* 2003;105:182-5.
39. Wang L, Duke L, Zhang PS, et al. Alternative splicing disrupts a nuclear localization signal in spleen tyrosine kinase that is required for invasion suppression in breast cancer. *Cancer Res* 2003;63:4724-30.
40. Kaczmarek J, Castellani P, Nicolo G, Spina B, Allemanni G, Zardi L. Distribution of oncofetal fibronectin isoforms in normal, hyperplastic and neoplastic human breast tissues. *Int J Cancer* 1994;59:11-6.
41. Luqmani YA, Mortimer C, Yiangou C, et al. Expression of 2 variant forms of fibroblast growth factor receptor 1 in human breast. *Int J Cancer* 1995;64:274-9.
42. Poola I, Speirs V. Expression of alternatively spliced estrogen receptor  $\alpha$  mRNAs is increased in breast cancer tissues. *J Steroid Biochem Mol Biol* 2001;78:459-69.
43. Erenburg I, Schachter B, Mira y Lopez R, Ossowski L. Loss of an estrogen receptor isoform (ER  $\alpha$   $\delta$  3) in breast cancer and the consequences of its reexpression: interference with estrogen-stimulated properties of malignant transformation. *Mol Endocrinol* 1997;11:2004-15.
44. Sciacca L, Costantino A, Pandini G, et al. Insulin receptor activation by IGF-II in breast cancers: evidence for a new autocrine/paracrine mechanism. *Oncogene* 1999;18:2471-9.
45. Sciacca L, Mineo R, Pandini G, Murabito A, Vigneri R, Belfiore A. In IGF-I receptor-deficient leiomyosarcoma cells autocrine IGF-II induces cell invasion and protection from apoptosis via the insulin receptor isoform A. *Oncogene* 2002;21:8240-50.
46. Martensen PM, Oka K, Christensen L, et al. Breast carcinoma epithelial cells express a very low-density lipoprotein receptor variant lacking the O-linked glycosylation domain encoded by exon 16, but with full binding activity for serine proteinase/serpin complexes and Mr-40,000 receptor-associated protein. *Eur J Biochem* 1997;248:583-91.
47. Sabbioni S, Barbanti-Brodano G, Croce CM, Negrini M. GOK: a gene at 11p15 involved in rhabdomyosarcoma and rhabdoid tumor development. *Cancer Res* 1997;57:4493-7.
48. Taylor CW. Store-operated Ca<sup>2+</sup> entry: A Stimulating stOrai. *Trends Biochem Sci* 2006;31:597-601.
49. Dunbar AJ, Goddard C. Structure-function and biological role of  $\beta$ cellulin. *Int J Biochem Cell Biol* 2000;32:805-15.
50. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530-6.
51. Larder R, Chang L, Clinton M, Brown P. Gonadotropin-releasing hormone regulates expression of the DNA damage repair gene, Fanconi anemia A, in pituitary gonadotroph cells. *Biol Reprod* 2004;71:828-36.
52. Fanti P, Nazareth M, Bucelli R, et al. Estrogen decreases chemokine levels in murine mammary tissue: implications for the regulatory role of MIP-1  $\alpha$  and MCP-1/JE in mammary tumor formation. *Endocrine* 2003;22:161-8.
53. Gauduchon J, Gouilleux F, Maillard S, Marsaud V, Renoir JM, Sola B. 4-Hydroxytamoxifen inhibits proliferation of multiple myeloma cells *in vitro* through down-regulation of c-Myc, up-regulation of p27Kip1, and modulation of Bcl-2 family members. *Clin Cancer Res* 2005;11:2345-54.
54. Auboeuf D, Batsche E, Dutertre M, Muchardt C, O'Malley BW. Coregulators: transducing signal from transcription to alternative splicing. *Trends Endocrinol Metab* 2007;18:122-9.
55. Hayes GM, Carrigan PE, Miller LJ. Serine-arginine protein kinase 1 overexpression is associated with tumorigenic imbalance in mitogen-activated protein kinase pathways in breast, colonic, and pancreatic carcinomas. *Cancer Res* 2007;67:2072-80.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Identification of Alternative Splicing Markers for Breast Cancer

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

*Cancer Res* 2008;68:9525-9531.

**Updated version** Access the most recent version of this article at:  
<http://cancerres.aacrjournals.org/content/68/22/9525>

**Supplementary Material** Access the most recent supplemental material at:  
<http://cancerres.aacrjournals.org/content/suppl/2008/11/12/68.22.9525.DC1>

**Cited articles** This article cites 55 articles, 19 of which you can access for free at:  
<http://cancerres.aacrjournals.org/content/68/22/9525.full.html#ref-list-1>

**Citing articles** This article has been cited by 31 HighWire-hosted articles. Access the articles at:  
</content/68/22/9525.full.html#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, contact the AACR Publications Department at [permissions@aacr.org](mailto:permissions@aacr.org).