Integrated Systems and Technologies

Exon-Based Clustering of Murine Breast Tumor Transcriptomes Reveals Alternative Exons Whose Expression Is Associated with Metastasis

Martin Dutertre1, Magali Lacroix-Triki3,4,5, Keltouma Driouch6, Pierre de la Grange2, Lise Gratadou1, Samantha Beck1,4,5, Stefania Millevoi3,4,5, Jamal Tazi7, Rosette Lidereau6, Stephan Vagner3,4,5, and Didier Auboeuf1

Abstract
In the field of bioinformatics, exon profiling is a developing area of disease-associated transcriptome analysis. In this study, we performed a microarray-based transcriptome analysis at the single exon level in mouse 4T1 primary mammary tumors with different metastatic capabilities. A novel bioinformatics platform was developed that identified 679 genes with differentially expressed exons in 4T1 tumors, many of which were involved in cell morphology and movement. Of 152 alternative exons tested by reverse transcription-PCR, 97 were validated as differentially expressed in primary tumors with different metastatic capability. This analysis revealed candidate progression genes, hinting at variations in protein functions by alternate exon usage. In a parallel effort, we developed a novel exon-based clustering analysis and identified alternative exons in tumor transcriptomes that were associated with dissemination of primary tumor cells to sites of pulmonary metastasis. This analysis also revealed that the splicing events identified by comparing primary tumors were not aberrant events. Lastly, we found that a subset of differentially spliced variant transcripts identified in the murine model was associated with poor prognosis in a large clinical cohort of patients with breast cancer. Our findings illustrate the utility of exon profiling to define novel theranostic markers for study in cancer progression and metastasis. Cancer Res; 70(3); 896–905. ©2010 AACR.

Introduction
Metastasis is the cause of >90% of deaths in breast cancer which is the most common neoplasm in women. A pressing challenge is to improve breast cancer molecular classification to better predict the risk of tumor metastasis. Great advances have been made using global gene expression profiling via DNA microarrays (1–3). However, the proposed molecular definitions have been derived by assessing only the variations in the expression level of transcripts without analyzing their exon content. However, most human genes can generate several splicing variants and misregulation of alternative splicing has been shown to occur for some genes in cancers (4–10).

In this context, genome-wide analysis of the transcriptome at the exon level is now possible thanks to a new generation of DNA microarrays that enable the profiling of virtually all human and mouse exons (11, 12). This technology has recently been used to identify differentially expressed exons in lung, colon, prostate, and brain tumor samples (12–15).

Although studies comparing the exon content of transcripts from normal versus tumor samples represent an important advance in the field, it remains to be tested whether different primary tumors expressed the same set of splicing variants. However, the comparisons of primary tumors is challenging as variations of transcript expression levels in human primary tumors reflect the huge heterogeneity that arises from many different factors, including metastatic ability, genetic alterations, age, and treatments (1–3, 16, 17).

This study aimed, first, to determine whether variations at the exon level could be identified by comparing the transcriptomes of primary tumors, and second, to search for potential splicing variants associated with different metastatic properties. For that purpose, we focused on the clinically relevant 4T1 animal model of spontaneous breast cancer metastasis. The 4T1 model comprises four syngeneic tumor cell lines isolated from a spontaneous mammary tumor in a BALB/cF13H mouse which can give rise to primary tumors with a spectrum of metastatic phenotypes when implanted into a mouse mammary fat pad (18–20). Using the Affymetrix...
Exon Array coupled with bioinformatics analysis, we found splicing variants that were associated with the ability of the murine mammary tumor cells to disseminate from primary mammary tumors into the lungs, and some of these splicing variants were linked with poor prognosis in a large cohort of patients with breast cancer.

Materials and Methods

Cell culture and mice model. 67NR, 168FARN, 4T07, and 4T1 cells were kindly provided by Dr. Fred Miller (Michigan Cancer Foundation, Detroit, MI). Cells were cultured in medium supplemented with 10% fetal bovine serum at 37°C. Three-month-old female BALB/c mice (Janvier Laboratory) were used for cell injection. 67NR, 168FARN, 4T07, and 4T1 cells (5×10⁶) were harvested, rinsed in fetal bovine serum–free medium, and injected into the fourth mammary fat pad in 100 μL of PBS. We injected 8 mice with 67NR cells, 11 mice with 168FARN cells, 15 mice with 4T07 cells, and 13 mice with 4T1 cells. The site of injection was validated by injecting a black ink solution.

Tumor samples and RNA extraction. Primary tumors were excised once the average primary tumor size in each group reached 1 to 2 cm in size. Each tumor sample used showed the presence of >70% of tumor cells. Very few stromal reactions such as fibrosis or inflammatory cell infiltrate were observed. Total RNA was extracted with Trizol (Invitrogen).

Affymetrix exon array hybridization. One microgram of total RNA from the 67NR, 168FARN, 4T07, and 4T1 primary mammary tumors was labeled with Affymetrix reagents and hybridized to Affymetrix-GeneChip Mouse Exon 1.0 ST arrays. Affymetrix Expression Console Software was used to perform quality assessment.

Array data and statistical analysis. Affymetrix exon Array data treatment was performed using FAST DB annotation and interface visualization (21, 22) and the EASANA analysis system (GenoSplice technology). Background correction was performed using antigenic probes and only probes with a detection above background of P ≤ 0.05 in at least half of the chips were considered for further statistical analysis (11–15). Only selected probes targeting exons annotated from full-length cDNA were used for analysis, as described previously (12–15). In one set of experiments, the transcriptome of each primary tumor was compared with the others (paired comparisons; see below). The experiment, from cell injection to array hybridization, was performed thrice. Paired statistical analyses were performed using Student’s paired t test on the splicing index to analyze the Exon Array data (12–15). Results were considered statistically significant for P ≤ 0.05 and fold changes ≥1.5. A statistical group analysis was performed to identify events common to different tumors. For this, only exons from genes expressed in the four samples were considered, and the means of gene-normalized exon intensity from each sample were ordered by ascending values. A Student’s unpaired t test was then performed for each possible group to select the group with the lowest P value. Hierarchical clustering was carried out to cluster the gene-normalized exon intensities and the samples using Mev4.0 software from The Institute of Genome Research. The functional analyses were generated using Ingenuity Pathways Analysis (Ingenuity Systems).

Validation by reverse transcription-PCR. One microgram of total RNA from each murine primary tumor was reverse transcribed using random primers and the Superscript II reverse transcriptase (Invitrogen). The resulting cDNAs were diluted 400×, and 5 μL of the diluted cDNAs were used for PCR amplification reactions using GoTaq DNA polymerase (Promega). Primer sequences are provided in Supplementary Table S1.

Clinical samples, quantitative PCR, and statistical analysis. This study used biopsies from primary breast tumors excised from 104 women treated at the Centre René Huguenin (Saint-Cloud, France) from 1977 to 1989 (see Supplementary Fig. S1). Each sample was normalized by the TATA box-binding protein transcript. The receiver-operating curve analysis provided the threshold expression value to balance sensitivity and specificity for detection of life-threatening cancer, and this cutpoint was used in the Kaplan-Meier analysis to estimate the metastasis-free survival distributions. The significance of differences between survival rates was ascertained using the log rank test. The linear combination of all the ratios was calculated as the sum of the KIAA1109 E+/E−, EPB41 E+/E−, CLSTN1 E−/E+, and TMEM16F E−/E+ ratios. The linear combination of all the splicing variants was calculated as the sum of weighted expression signals of all variants with their Cox’s regression coefficient as the weight.

Results

Variations of the exon content of mRNAs produced from genes involved in cellular morphology and movement in a mouse model of tumor progression. To search for variations in the transcriptome at the exon level during tumor progression, we focused on the syngeneic tumor lines 67NR, 168FARN, 4T07, and 4T1 that have differential metastatic behavior (18–20). In agreement with earlier reports (18–20), the 67NR cell line formed primary carcinomas when implanted into mouse mammary fat pads and no tumor cells were detectable in distant tissue; the 168 FARN cell line formed primary carcinomas with extensions to local lymph nodes, whereas the (4T07, 4T1) cell lines generated micrometastases and macroscopic metastasis, respectively, in the lungs (Fig. 1A; Supplementary Fig. S2).

Total RNA was purified from the primary tumors 21 days after the cell lines had been injected into the mouse mammary fat pads. The transcriptomes of the four primary mammary tumor types were analyzed using the Affymetrix GeneChip Mouse Exon 1.0 ST Array that contains multiple probes per exon, allowing to search for variations at the exon level (11–15).
We focused on probes supported by full-length mRNAs and analyzed the exon content of transcripts produced from 12,208 well-annotated mouse genes. The splicing index method (11–15) was used to identify differentially expressed exons in paired comparisons with a fold change of >1.5 and a $P \leq 0.05$. Using this strategy, 1,233 genes with at least one differentially expressed exon were identified by the paired comparisons (total events; Table 1). This corresponded to 679 unique genes because some differentially expressed exons were simultaneously identified in different paired comparisons (Supplementary Fig. S3).

Interestingly, among these 679 genes, 212 and 97 genes corresponded to genes related to cancer and reproductive system disease, respectively, whereas 94 and 113 genes were related to cellular morphology and cellular movement, respectively (Table 2; Supplementary Fig. S4). Both these two cellular functions are highly relevant to tumor progression.

**Differential expression of alternatively spliced exons in primary mouse mammary tumors.** To focus on already known alternative exons, a manual inspection was performed after uploading the Exon Array data into our FAST DB database, gathering all the known alternative exons thanks to the computational comparison of publicly available mRNA sequences with genomic sequences (21, 22). Figure 1B illustrates the annotation process using the RAI14 gene (23–25) as an example. Following FAST DB exon numeration and annotation, the mouse RAI14 gene contains one alternatively spliced exon (indicated by red lines below exon 12). Each Affymetrix probe corresponding to the RAI14 gene is represented by a column above the gray exonic structure of the gene (bottom). The color of each column is representative of the difference in probe intensities between samples (in this case, three paired comparisons of 168FARN and 4T1 samples); a green or red column indicates that the intensity of the corresponding probe was lower or greater, respectively, in the 168FARN samples than in the 4T1 samples; a black column corresponds to probes with no intensity variation between samples. Alternatively spliced RAI14 exon 12 represented by green columns was predicted to be skipped more frequently in the 168FARN tumors than in the 4T1 tumors. Screen shot is from EASANA/FAST DB.

![Figure 1. 4T1 model of tumor progression and identification of differentially expressed alternative exons in primary tumors. A, phenotypes of the 67NR, 168FARN, 4T07, and 4T1 mouse mammary tumors. Twenty-one days after injection of $5 \times 10^5$ cells into the mammary fat pads, primary tumors, and lungs were harvested and paraffin-embedded for histologic examination (H&E; magnification, x200). No lung metastases were observed in the 67NR-injected and 168FARN-injected mice, whereas lung micrometastases and macrometastases were observed in the 4T07-injected and 4T1-injected mice (arrows). B, analysis of the RAI14 gene. Following FAST DB exon numeration and annotation (top), the mouse RAI14 gene contains one alternatively spliced exon (indicated by red lines below exon 12). Each Affymetrix probe corresponding to the RAI14 gene is represented by a column above the gray exonic structure of the gene (bottom). The color of each column is representative of the difference in probe intensities between samples (in this case, three paired comparisons of 168FARN and 4T1 samples); a green or red column indicates that the intensity of the corresponding probe was lower or greater, respectively, in the 168FARN samples than in the 4T1 samples; a black column corresponds to probes with no intensity variation between samples. Alternatively spliced RAI14 exon 12 represented by green columns was predicted to be skipped more frequently in the 168FARN tumors than in the 4T1 tumors. Screen shot is from EASANA/FAST DB.](image-url)
Table 1. Number of genes with at least one differentially expressed alternative exon (“Gene with alternative exons”) in “Paired comparisons” or in “Group comparisons” as indicated

<table>
<thead>
<tr>
<th>Paired comparisons</th>
<th>Genes with Annotated alternative exons</th>
<th>Genes with alternative exons</th>
</tr>
</thead>
<tbody>
<tr>
<td>67NR vs. 168FARN</td>
<td>178</td>
<td>63</td>
</tr>
<tr>
<td>67NR vs. 4T07</td>
<td>208</td>
<td>65</td>
</tr>
<tr>
<td>67NR vs. 4T1</td>
<td>257</td>
<td>78</td>
</tr>
<tr>
<td>168FARN vs. 4T07</td>
<td>230</td>
<td>80</td>
</tr>
<tr>
<td>168FARN vs. 4T1</td>
<td>325</td>
<td>108</td>
</tr>
<tr>
<td>4T07 vs. 4T1</td>
<td>35</td>
<td>9</td>
</tr>
<tr>
<td>Total events</td>
<td>1233</td>
<td>403</td>
</tr>
<tr>
<td>Tested events</td>
<td>152</td>
<td></td>
</tr>
<tr>
<td>Validated events</td>
<td>97 (64%)</td>
<td></td>
</tr>
<tr>
<td>Total gene number</td>
<td>679</td>
<td>209</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group comparisons</th>
<th>Genes with alternative exons</th>
</tr>
</thead>
<tbody>
<tr>
<td>168FARN vs. (67NR, 4T07, 4T1)</td>
<td>148</td>
</tr>
<tr>
<td>67NR vs. (168FARN, 4T07, 4T1)</td>
<td>59</td>
</tr>
<tr>
<td>4T07 vs. (67NR, 168FARN, 4T1)</td>
<td>9</td>
</tr>
<tr>
<td>4T1 vs. (67NR, 168FARN, 4T07)</td>
<td>16</td>
</tr>
<tr>
<td>(67NR, 168FARN) vs. (4T07, 4T1)</td>
<td>78</td>
</tr>
<tr>
<td>(67NR, 4T07) vs. (168FARN, 4T1)</td>
<td>13</td>
</tr>
<tr>
<td>(67NR, 4T1) vs. (168FARN, 4T07)</td>
<td>2</td>
</tr>
</tbody>
</table>

NOTE: “Annotated Alternative exons” refers to exons that are already annotated as alternative exons in FAST DB. “Total gene number” refers to the number of nonredundant genes.

Computational analysis of the Exon Array data also predicted a set of differentially expressed exons in the 4T1 model that were not annotated in FAST DB as alternative exons. This is illustrated by exon 11 of the mouse DEFH9 gene (35–37), which was predicted to be and which was validated by RT-PCR as being differentially expressed in the 67NR and 4T1 samples (Fig. 3; Supplementary Fig. S7). This indicates that all of the differentially expressed mouse exons identified in this study are potentially alternative exons, even those that are not yet annotated in databases.

Association of a set of alternative exons with the ability of cells to disseminate from primary mammary tumors into the lungs. We noted that the (4T07, 4T1) transcriptomes were more similar to each other than to the (67NR, 168FARN) transcriptomes and that many events were common to the (4T07, 4T1) primary tumors (Table 1; Supplementary Fig. S3). Remarkably, two groups containing either the (67NR, 168FARN) samples or the (4T07, 4T1) samples were identified by performing a clustering analysis based on gene-normalized exon intensity values (Fig. 3A). These data suggested that differential expression of alternative exons was associated with the ability of primary tumors to disseminate from primary mammary tumors (4T07, 4T1) or not (67NR, 168FARN).

To identify alternative exons specific to the (67NR, 168FARN) group compared with the (4T07, 4T1) group, we performed a statistical group analysis on the splicing index values for all the exons of only genes expressed in all four tumor types. Remarkably, 78 events were identified as specific to the (67NR, 168FARN) group compared with the (4T07, 4T1) group (“Group comparisons”; Table 1; Supplementary Table S3). Much fewer events were identified as being specific to the (67NR, 4T07) group compared with the (168FARN, 4T1) group or to the (67NR, 4T1) group compared with the (168FARN, 4T07) group (Table 1).

Strikingly, among the genes with alternative exons that were differentially expressed in the (67NR, 168FARN) group compared with the (4T07, 4T1) group, two genes were known to be involved in tumor progression. The first gene was the FGFR2 gene (38–40), which encodes the fibroblast growth factor receptor 2. Exon switching of two mutually exclusive FGFR2 exons that occurred between the (67NR, 168FARN) group and the (4T07, 4T1) group (Fig. 3B) is required for epithelial cell tumor progression (38–41). The second gene was the CD44 gene, which is involved in cellular adhesion and motility (9, 42). Interestingly, the use as prognostic markers of CD44 splicing variants generated from 10 cassette exons is currently being investigated (9, 42). Remarkably, CD44 alternative splicing variants were detected only in the (4T07, 4T1) group (Fig. 3B). Likewise, several splicing variants generated from the STRN3 gene (43) were detected only in the (4T07, 4T1) group (Fig. 3B).
We also identified several single cassette exons that were differentially expressed in the two groups. For example, cassette exon 4 of the KIAA1109/FSA gene (44, 45) and exon 22 of the EPB41 gene (46, 47) were more frequently included in transcripts in the (4T07, 4T1) group compared with the (67NR, 168FARN) group (Fig. 3B). Meanwhile, cassette exon 3 of the TMEM16F gene (48, 49) and exon 11 of the CLSTN1 gene were more frequently excluded in transcripts in the (4T07, 4T1) group compared with the (67NR, 168FARN) group (Fig. 3B). Similar RT-PCR results were obtained by using RNAs extracted from other sets of primary tumors and by performing different numbers of PCR cycles (Supplementary Fig. S5 and S8). Therefore, our analysis of the transcriptome at the exon level identified a subset of alternatively spliced exons that were differentially expressed in mouse primary tumors that disseminate (4T07, 4T1) or not (67NR, 168FARN) into the lungs.

In an attempt to identify splicing factors that may explain the splicing switches observed above, we performed a statistical group analysis on gene expression levels. This led us to identify 1,526 genes whose expression differed between the (67NR, 168FARN) group and the (4T07, 4T1) group (data not shown). However, we did not find genes coding for well-characterized splicing factors such as the SR (serine-arginine-rich family of splicing factors) and hnRNP (heterogeneous nuclear ribonucleoprotein) proteins. Very interestingly, we noted that more than half of the genes (including STRN3, KIAA1109, TMEM16, and CLSTN1) with differential expression at the exon level between the (67NR, 168FARN) group and the (4T07, 4T1) group were not differentially expressed at the whole gene level. Therefore, these genes would not have been identified by using classical arrays or gene expression profiling (see Discussion).

Association of alternative transcript expression levels with metastasis-free survival prognosis in a large cohort of patients with breast cancer. By looking in FAST DB at each of the 78 genes with differentially expressed exons between tumors that disseminated (4T07, 4T1) or not (67NR, 168FARN) into the lungs, we noted that 47 exons were already annotated as alternative exons and 30 (38%) were exons differentially expressed in a collection of normal tissues (Fig. 3C; Supplementary Table S3).

Furthermore, among the 47 mouse exons that were already annotated as alternative exons in FAST DB, 18 (38%) were also annotated as alternative exons in humans in FAST DB (Fig. 3C; Supplementary Table S3). These included FGFR2, CD44, and STRN3, KIAA1109, TMEM16, and CLSTN1.

Table 2. Functions and alternative events of selected genes with differentially expressed exons in the 4T1 model

<table>
<thead>
<tr>
<th>Alternatively spliced exons</th>
<th>Symbol</th>
<th>Function(s)*</th>
<th>Alternative events†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cassette exons</td>
<td>RAII14</td>
<td>Putative actin cytoskeleton organization function</td>
<td>In-frame (29 aa); cellular localization</td>
</tr>
<tr>
<td></td>
<td>ADD3</td>
<td>Structural constituent of cytoskeleton (spectrin-actin network)</td>
<td>In-frame (32 aa)</td>
</tr>
<tr>
<td></td>
<td>FN1</td>
<td>Adhesive and migratory processes</td>
<td>In-frame (91 aa); cell adhesion</td>
</tr>
<tr>
<td></td>
<td>ECT2</td>
<td>Rho guanyl-nucleotide exchange factor; cytokinesis and epithelial cell polarity</td>
<td>In-frame (31 aa)</td>
</tr>
<tr>
<td></td>
<td>KIAA1109</td>
<td>Epithelial growth/differentiation</td>
<td>In-frame (67 aa)</td>
</tr>
<tr>
<td></td>
<td>EPB41</td>
<td>Structural constituent of cytoskeleton; membrane-associated cytoskeleton</td>
<td>In-frame (150 aa)</td>
</tr>
<tr>
<td></td>
<td>TMEM16F</td>
<td>Calcium-activated chloride channel</td>
<td>In-frame (22 aa)</td>
</tr>
<tr>
<td></td>
<td>CLSTN1</td>
<td>Transmembrane protein of the cadherin superfamily of cell adhesion molecules</td>
<td>In-frame (19 aa)</td>
</tr>
<tr>
<td>Mutually exclusive exons</td>
<td>TPM2</td>
<td>Actin binding; actin cytoskeleton functions</td>
<td>In-frame filamentous actin binding</td>
</tr>
<tr>
<td></td>
<td>CALU</td>
<td>Actin cytoskeleton functions</td>
<td>In-frame EF-hand Ca(2+) binding sites</td>
</tr>
<tr>
<td></td>
<td>FGFR2</td>
<td>Fibroblast growth factor receptor</td>
<td>In-frame differential binding of FGFs</td>
</tr>
<tr>
<td>Multiple cassette exons</td>
<td>MYO1B</td>
<td>Motor activity; actin and calmodulin binding; actin cytoskeleton organization</td>
<td>In-frame</td>
</tr>
<tr>
<td></td>
<td>HISPPD1</td>
<td>Diphosphoinositol pentakiphosphate kinase</td>
<td>In-frame IQ motifs</td>
</tr>
<tr>
<td></td>
<td>CD44</td>
<td>Cell adhesion glycoprotein</td>
<td>In-frame extracellular domain</td>
</tr>
<tr>
<td></td>
<td>STRN3</td>
<td>WD-40 repeat protein with potential scaffolding functions; cell cycle</td>
<td>In-frame</td>
</tr>
<tr>
<td>Intron retention</td>
<td>SSR3</td>
<td>Cotranslational protein targeting to membrane</td>
<td>Frame shift; C-ter truncated protein</td>
</tr>
<tr>
<td></td>
<td>SLC38A2</td>
<td>Sodium-dependent amino acid transporter</td>
<td>Frame shift; C-ter truncated protein</td>
</tr>
<tr>
<td></td>
<td>ADAM33</td>
<td>Membrane-anchored glycoprotein metalloprotease</td>
<td>Frame shift; C-ter truncated protein</td>
</tr>
</tbody>
</table>

Abbreviations: nts, nucleotides; aa, amino acids; c-ter, COOH-terminal.

*The references are provided in the Results and Discussion.

†Consequence of alternative usage of exons at the protein level as annotated in FAST DB and in the literature.
STRN3, and CLSTN1 described in Fig. 3B. However, the EPB41, KIAA1109/FSA, and TMEM16F splicing variants described in mice had not yet been described in humans. Therefore, we designed primers to test whether the corresponding splicing variants were expressed in a variety of human tissues. Remarkably, the alternative exons identified in mice were also alternatively spliced in normal human tissue samples (Supplementary Fig. S9). Collectively, these data indicated that the splicing events identified by comparing primary tumors having different abilities to disseminate were not aberrant events.

As tumor metastasis is responsible for most deaths of patients with breast cancer, we investigated whether the differentially expressed splicing variants identified in the mouse 4T1 model of tumor progression were associated with metastasis-free survival in patients with breast cancer. We focused on genes with single cassette exons, i.e., on the EPB41, KIAA1109/FSA, TMEM16F, and CLSTN1 genes (Fig. 3B), and using RT-quantitative PCR, we measured the human splicing variants containing or not the alternatively spliced exons in a set of 104 clinically annotated breast cancer tumors.

The biological outcome of expression variation of two splicing variants produced from one single gene can depend on the expression of one specific variant (independently of the other one) and/or on the relative expression of both variants (e.g., when two splicing variants produce protein isoforms with opposite activities, like proapoptotic and antiapoptotic isoforms). Therefore, we analyzed the expression of each splicing variant as well as their relative expression level (ratio). We first analyzed the KIAA1109 and EPB41 genes, which had a higher level of alternative exon inclusion in the metastatic (4T07, 4T1) primary mouse tumors than in the nonmetastatic (67NR, 168FARN) tumors (Fig. 3B). Remarkably, a high KIAA1109 alternative exon inclusion to exclusion ratio in human samples correlated with poor prognosis in patients with breast cancer (KIAA1109 +E/-E; Fig. 4A). There was a similar trend for EPB41 alternative exon 22 (EPB41 +E/-E; Supplementary Fig. S10), with higher levels of the EPB41 exclusion variant associated with better patient prognosis (EPB41 -E; Fig. 4B).

We next analyzed the TMEM16F and CLSTN1 genes, which had a lower level of alternative exon inclusion in the metastatic...
(4T07, 4T1) primary mouse tumors than in the nonmetastatic (67NR, 168FARN) tumors (Fig. 3B). Strikingly, a low TMEM16F alternative exon inclusion to exclusion ratio correlated with poor prognosis (TMEM16F +E/−E; Fig. 4A). There was a similar trend for the CLSTN1 alternative exon (CLSTN1 +E/−E; Supplementary Fig. S10), with higher levels of the CLSTN1 inclusion variant associated with better patient prognosis (CLSTN1 +E; Fig. 4B). In addition, there was a statistically significant association with metastasis-free survival when taking into account the four ratios (“All ratios”; Fig. 4A) or the combination of all splicing variants (“All variants”; Fig. 4B).

Discussion

In this report, we showed variations of the transcriptome at the exon level across mammary primary tumors with different abilities to disseminate. Interestingly, the variations at the exon level occurred often in mRNAs produced from genes with functions related to cellular morphology and movement (Table 2; Supplementary Fig. S4). In addition, transcriptome analysis at the exon level not only pointed to genes involved in cellular functions relevant to tumor progression but also hinted at specific protein features targeted by...
variations at the exon level (Table 2). For example, the RAI14 gene encodes a conserved protein that localizes in confluent cells at cell-cell adhesion sites or along the stress fibers, suggesting its involvement in the organization of the actin cytoskeleton (23–25). However, the RAI14 protein is mainly localized to nuclei in nonconfluent cells and the alternatively spliced RAI14 exon 12 (Fig. 1B) encodes for a conserved nuclear localization signal that has been experimentally validated (23–25). Therefore, skipping exon 12 may change the subcellular localization of the RAI14 protein (23–25). Likewise, protein features targeted by variations at the exon level are well characterized for the FGFR2 and CD44 genes (see Results). As described in Table 2 (and see below), many of the alternative exons identified in this study resulted in the in-frame deletion of protein domains pointing to specific protein features targeted by variations at the exon level.

Furthermore, we showed variations of the transcriptome at the exon level when comparing 67NR- and 168FARN-derived primary tumors, which did not disseminate into the lungs, as compared with 4T07- and 4T1-derived primary tumors (Fig. 3). Interestingly, many exons that were differentially expressed across primary tumors which were able to disseminate or not were found to be alternative exons in mouse normal tissues and in human (Fig. 3C; Supplementary

Figure 4. Kaplan-Meier curves for 15-y outcome in patients with breast cancer (n = 104) based on splicing variant content of KIAA1109, EPB41, TMEM16F, and CLSTN1 genes. Metastasis-free survival curves according to (A) the inclusion/exclusion ratio of the KIAA1109 (KIAA1109 E+/E−) or TMEM16F (TMEM16F E+/E−) genes or according to the combination of the inclusion/exclusion ratio of the KIAA1109, EPB41, TMEM16F, and CLSTN1 genes (All ratios); and (B) the level of exon exclusion of the EPB41 gene (EPB41 E−) or of exon inclusion of CLSTN1 gene (CLSTN1 E+) or according to the combination of the eight splicing variants generated by the KIAA1109, EPB41, TMEM16F, and CLSTN1 genes (All variants). Abscise axis (mo) and ordinate axis (percentage of metastasis-free survival).
These data indicated that the splicing events identified by comparing primary tumors were not aberrant events. The microarray analysis did not reveal changes in expression at the RNA levels of major splicing factors (data not shown) and variations of splicing variant expression observed in the 4T1 mouse model might have different origins. However, whatever the origin is, an important achievement of this study is the demonstration that some of these variations were associated with metastasis-free survival in a large cohort of patients with breast cancer.

As underlined in the Introduction, there is a pressing need for better prognostic and predictive markers in breast cancer. The analysis of the transcriptome at the exon level is likely to greatly enrich and improve molecular definitions of primary tumors. Indeed, several genes, including the KIAA1109, TMEM16F, and CLSTN1 genes identified in this study, would not have been found with classical 3′-based microarrays because the genes’ overall expression levels were not modified. In addition, we showed that a subset of alternative splicing events identified in the 4T1 model was associated with poor prognosis in a cohort of patients with breast cancer (Fig. 4; Supplementary Fig. S10 and S11). It is noteworthy that the alternative transcripts associated with poor prognosis were produced by genes (i.e., KIAA1109, EPB41, TMEM16F, and CLSTN1) that are not currently known as prognostic or predictive markers, even though their functions are highly relevant to cancer.

For example, the KIAA1109 or FSA gene is highly conserved during evolution and shares sequence similarity with the Caenorhabditis elegans lpd-3 gene, which is involved in lipid storage. The murine KIAA1109/FSA gene also plays a role in 3T3-L1 cell adipogenesis induction in vitro and is upregulated during mammary gland development (44, 45). The KIAA1109/FSA gene is expressed in many tissues (e.g., colon, ovary, and prostate) and its expression level is downregulated in tumors originating in these tissues (44, 45). In addition, the KIAA1109/FSA gene was recently identified in Chinese hamster ovary cells by positional cloning of the 1q31 fragile site, which plays an important role in regulating the amplification of the multidrug resistance gene in multidrug-resistant cells (44, 45). Altogether, these data suggest that the KIAA1109/FSA gene plays important roles in regulating mammary epithelial growth and differentiation, as well as in tumor development. Another example is the EPB41 gene that has been implicated in several diseases and is a tumor suppressor gene candidate (46, 47). The EPB41 protein, together with spectrin and actin, form the membrane-associated cytoskeleton that supports animal cell membranes (46, 47). The EPB41 gene produces several splice variants, in particular, transcripts that include exon 22 (also named exon 17b) restricted to epithelial cells (46, 47).

Interestingly, the regulation of exon 22 inclusion has been shown to correlate with cell shape: whereas the nondividing suspension human mammary epithelial cells strongly expressed transcripts that included exon 22, proliferating adherent human mammary epithelial cells mostly produced transcripts without exon 22 (46, 47). Another example is the TMEM16F gene that belongs to a new family of calcium-activated chloride channels, which are major regulators of sensory transduction, smooth muscle contraction, and epithelial secretion (48, 49). Moreover, members of this family play a role in cell adhesion and are overexpressed in some types of cancer (48, 49). Finally, the CLSTN1 gene is a transmembrane protein of the cadherin superfamily, which is involved in cell adhesion, tissue organization, and morphogenesis regulation (35–37).

In conclusion, our study showed that exon-based transcriptome profiling and clustering of tumors allows the identification of novel cancer-related alternative exons and may hint at specific protein features in biological processes. The analysis of exon expression level constitutes a valuable approach for the identification of novel prognosis markers and may help scientists better understand the molecular mechanisms underlying tumor progression and aid in the development of novel therapeutic tools.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Adrien Briaux for excellent technical assistance.

Grant Support

ANR, INCa, and European Union (NoE EURASNET). M. Dutertre was supported by INSERM; L. Gratadou and S. Beck by INCa; P. de la Grange by AFM and EURASNET. Work in the laboratory of S. Vagner was also supported by FRM (Equipe FRM, soutenue par la Fondation Recherche Médicale).

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.

Received 7/20/09; revised 12/3/09; accepted 12/3/09; published OnlineFirst 1/26/10.

References

7. Gunning P, O’Neill G, Hardeman E. Tropomyosin-based regulation...
of the actin cytoskeleton in time and space. Physiol Rev 2008;88:1–35.


17. Goldhirsh A, Glick JH, Gelber RD, et al. Meeting highlights: interna-


Exon-Based Clustering of Murine Breast Tumor Transcriptomes Reveals Alternative Exons Whose Expression Is Associated with Metastasis

Martin Dutertre, Magali Lacroix-Triki, Keltouma Driouch, et al.

Cancer Res  Published OnlineFirst January 26, 2010.

Updated version  Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-09-2703

Supplementary Material  Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2010/01/26/0008-5472.CAN-09-2703.DC1

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.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.