

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

Martin Dutertre¹, Magali Lacroix-Triki^{3,4,5}, Keltouma Driouch⁶, Pierre de la Grange², Lise Gratadou¹, Samantha Beck^{3,4,5}, Stefania Millevoi^{3,4,5}, Jamal Tazi⁷, Rosette Lidereau⁶, Stephan Vagner^{3,4,5}, 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/c/cf3H 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

Authors' Affiliations: ¹INSERM UMRS 940, ²GenoSplice Technology, Hôpital Saint-Louis, Paris, France; ³Institut National de la Santé et de la Recherche Médicale, ⁴Institut Claudius Regaud, ⁵Université Toulouse III Paul Sabatier, Toulouse, France; ⁶INSERM U735, Centre René Huguenin, Saint-Cloud, France; and ⁷Université Montpellier 2, Université Montpellier 1, CNRS, Institut de Génétique Moléculaire de Montpellier, Montpellier, France

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

M. Dutertre, M. Lacroix-Triki, and K. Driouch contributed equally to this work.

Corresponding Authors: Stephan Vagner, INSERM U563, 20-24 Rue du Pont St-Pierre, 31052 Toulouse, France. Phone: 33-56769-6311; E-mail: stephan.vagner@inserm.fr and Didier Auboeuf, INSERM UMRS 940, 1 Avenue Claude Vellefaux, 75010 Paris, France. Phone: 33-15727-6727; Fax: 33-14240-9557; E-mail: didier.auboeuf@inserm.fr.

doi: 10.1158/0008-5472.CAN-09-2703

©2010 American Association for Cancer Research.

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 (5×10^5) were harvested, rinsed in fetal bovine serum-free medium, and injected into the fourth mammary fat pad in 100 μ L 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.⁸ 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).⁹ Background correction was performed using antigenomic probes and only probes with a detection above background of $P \leq 0.05$ 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 $P \leq 0.05$ 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).¹⁰

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 \times , and 5 μ L 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 168FARN 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).

⁸ <http://www.affymetrix.com/>

⁹ <http://www.genosplice.com/>

¹⁰ <http://www.ingenuity.com/>

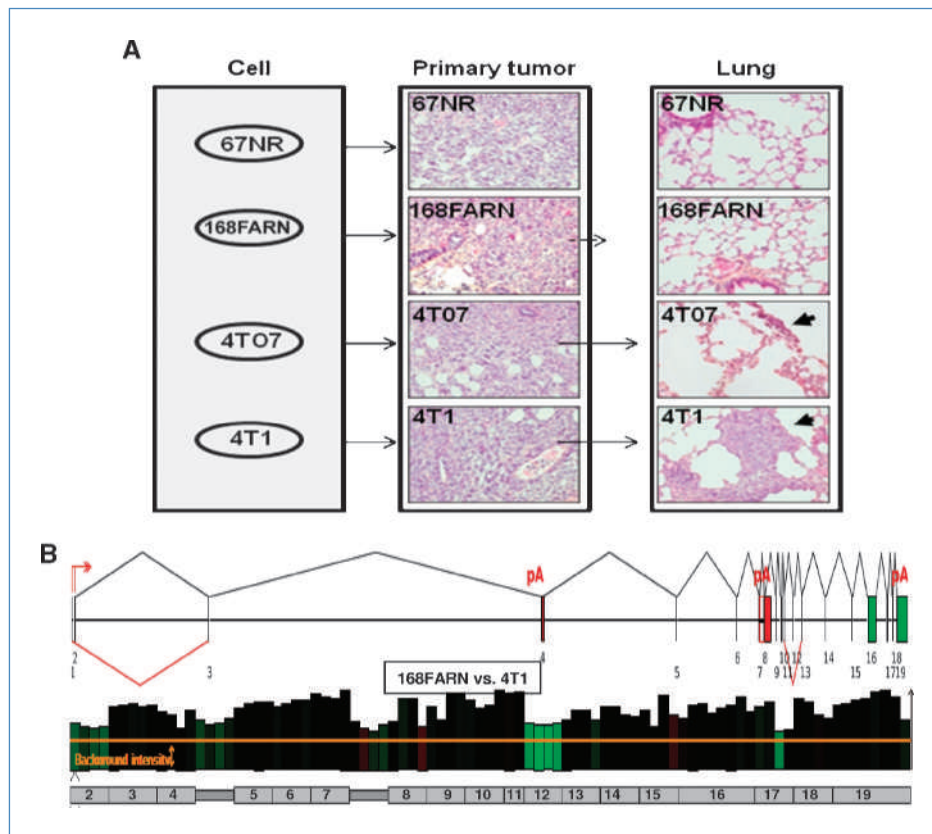


Figure 1. 4T1 model of tumor progression and identification of differentially expressed alternative exons in primary tumors. A, phenotypes of the 67NR, 168FARN, 4T07, and 4T1 mouse mammary tumors. Twenty-one days after injection of 5×10^5 cells into the mammary fat pads, primary tumors, and lungs were harvested and paraffin-embedded for histologic examination (H&E; magnification, $\times 200$). 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.

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* exon 12 is alternatively spliced, as indicated by red lines below exon 12. The computational analysis and visualization of the Exon Array data predicted that the *RAI14* exon 12 was differentially expressed when comparing the 168FARN to the 4T1 tumors (Fig. 1B). Reverse transcription-PCR analysis showed that indeed the *RAI14* exon 12 was included more frequently in the 4T1 than in the 168FARN samples (Fig. 2A).

In addition to other cases of cassette exons illustrated with the *ADD3* (26), *FNI* (27), and *ECT2* (28) genes (Fig. 2A), several genes that used mutually exclusive exons (i.e., exons that

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

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

Group comparisons	Genes with alternative exons	
	alternative exons	alternative exons
168FARN vs. (67NR, 4T07, 4T1)	148	
67NR vs. (168FARN, 4T07, 4T1)	59	
4T07 vs. (67NR, 168FARN, 4T1)	9	
4T1 vs. (67NR, 168FARN, 4T07)	16	
(67NR, 168FARN) vs. (4T07, 4T1)	78	
(67NR, 4T07) vs. (168FARN, 4T1)	13	
(67NR, 4T1) vs. (168FARN, 4T07)	2	

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.

are not included in the same transcript) were identified as illustrated for the *TPM2* (7) and *CALU* (29) genes (Fig. 2B). Furthermore, splicing events that simultaneously affected several exons were also identified as illustrated for the *MYO1B* (30) and the *HISPPDI* (31) genes (Fig. 2C). Finally, several cases of intron retention were identified as illustrated for the *SSR3* (32), *SLC38A2* (33), and *ADAM33* (34) genes (Fig. 2D). Similar RT-PCR results were obtained using RNAs extracted from other sets of primary tumors (Supplementary Fig. S5). Furthermore, similar results were obtained by using RNAs from the cell lines instead of from the primary tumors, which indicates that splicing variations appeared in the cell lines from which the tumors were derived (Supplementary Fig. S6).

Going through the 1,233 alternative events that were identified by paired comparisons, 403 events corresponded to annotated alternative exons in FAST DB (Table 1); 152 of these events were analyzed by RT-PCR and 97 (64%) were validated. In several cases that were not validated, only one PCR product was obtained, suggesting that one alternative transcript was predominantly expressed (data

not shown). As mentioned previously, a subset of differentially expressed exons was predicted simultaneously in several paired comparisons. Taking into account this redundancy, 209 genes with at least one annotated alternative exon that were differentially expressed in the 4T1 breast cancer model were identified (Table 1; Supplementary Table S2).

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 *CLSTN1* 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*, *TMEM16F*, 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*,

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

Alternatively spliced exons	Symbol	Function(s)*	Alternative events [†]
Cassette exons	<i>RAI14</i>	Putative actin cytoskeleton organization function	In-frame (29 aa); cellular localization
	<i>ADD3</i>	Structural constituent of cytoskeleton (spectrin-actin network)	In-frame (32 aa)
	<i>FN1</i>	Adhesive and migratory processes	In-frame (91 aa); cell adhesion
	<i>ECT2</i>	Rho guanyl-nucleotide exchange factor; cytokinesis and epithelial cell polarity	In-frame (31 aa)
	<i>KIAA1109</i>	Epithelial growth/differentiation	In-frame (67 aa)
	<i>EPB41</i>	Structural constituent of cytoskeleton; membrane-associated cytoskeleton	In-frame (150 aa)
	<i>TMEM16F</i>	Calcium-activated chloride channel	In-frame (22 aa)
Mutually exclusive exons	<i>CLSTN1</i>	Transmembrane protein of the cadherin superfamily of cell adhesion molecules	In-frame (19 aa)
	<i>TPM2</i>	Actin binding; actin cytoskeleton functions	In-frame filamentous actin binding
	<i>CALU</i>	Actin cytoskeleton functions	In-frame EF-hand Ca(2+) binding sites
	<i>FGFR2</i>	Fibroblast growth factor receptor	In-frame differential binding of FGFRs
Multiple cassette exons	<i>MYO1B</i>	Motor activity; actin and calmodulin binding; actin cytoskeleton organization	In-frame
	<i>HISPPD1</i>	Diphosphoinositol pentakisphosphate kinase	In-frame IQ motifs
	<i>CD44</i>	Cell adhesion glycoprotein	In-frame extracellular domain
Intron retention	<i>STRN3</i>	WD-40 repeat protein with potential scaffolding functions; cell cycle	In-frame
	<i>SSR3</i>	Cotranslational protein targeting to membrane	Frame shift; C-ter truncated protein
	<i>SLC38A2</i>	Sodium-dependent amino acid transporter	Frame shift; C-ter truncated protein
	<i>ADAM33</i>	Membrane-anchored glycoprotein metalloprotease	Frame shift; C-ter truncated protein

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.

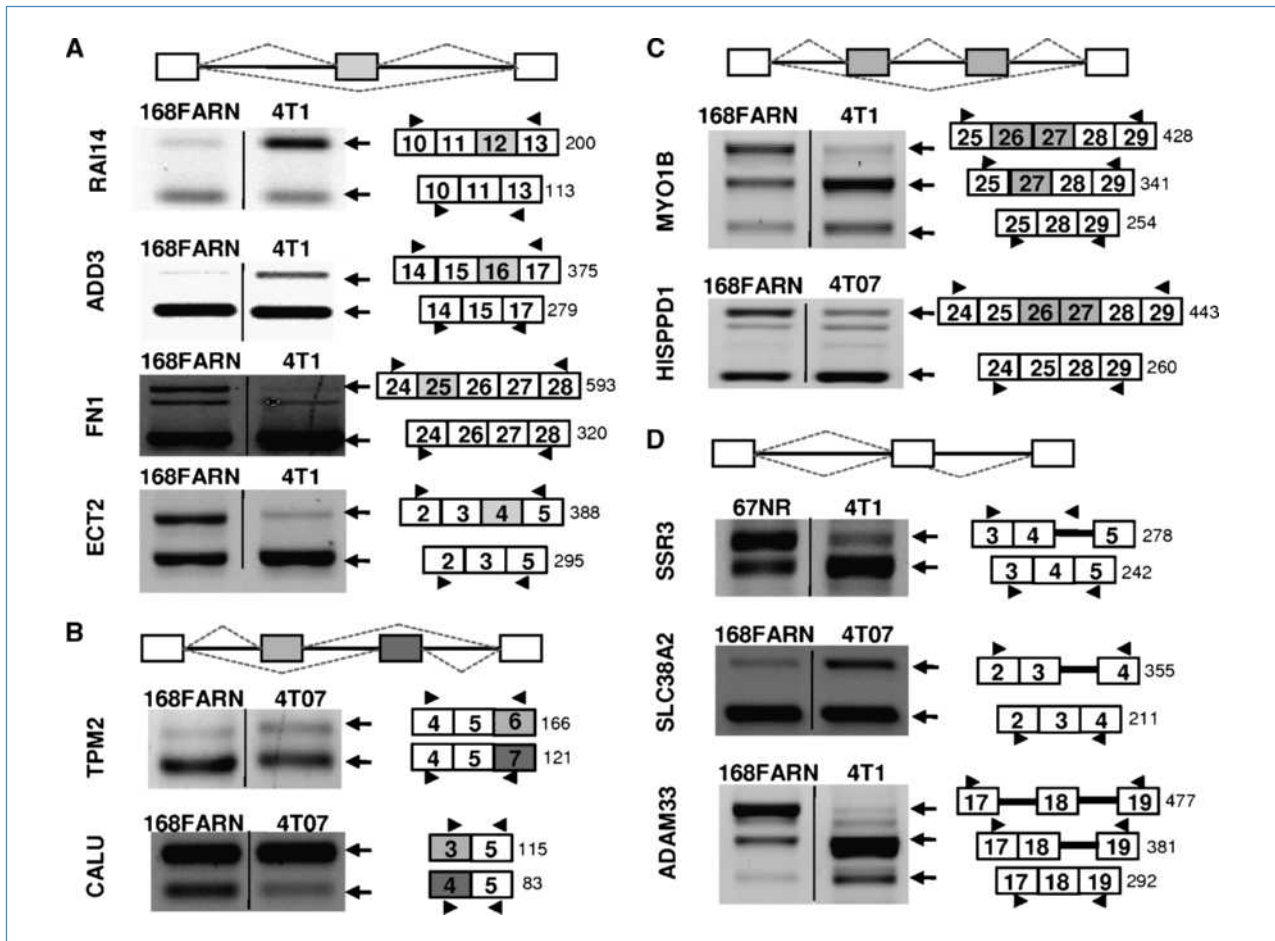


Figure 2. Validation of differentially expressed exons in primary tumors. Definition and RT-PCR analysis of alternative exons for (A) *RAI14*, *ADD3*, *FN1*, and *ECT2* genes; (B) *TPM2* and *CALU* genes; (C) *MYO1B* and *HISPPD1* genes; and (D) *SSR3*, *SLC38A2*, and *ADAM33* genes. The RT-PCR results are representative of three independent experiments.

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

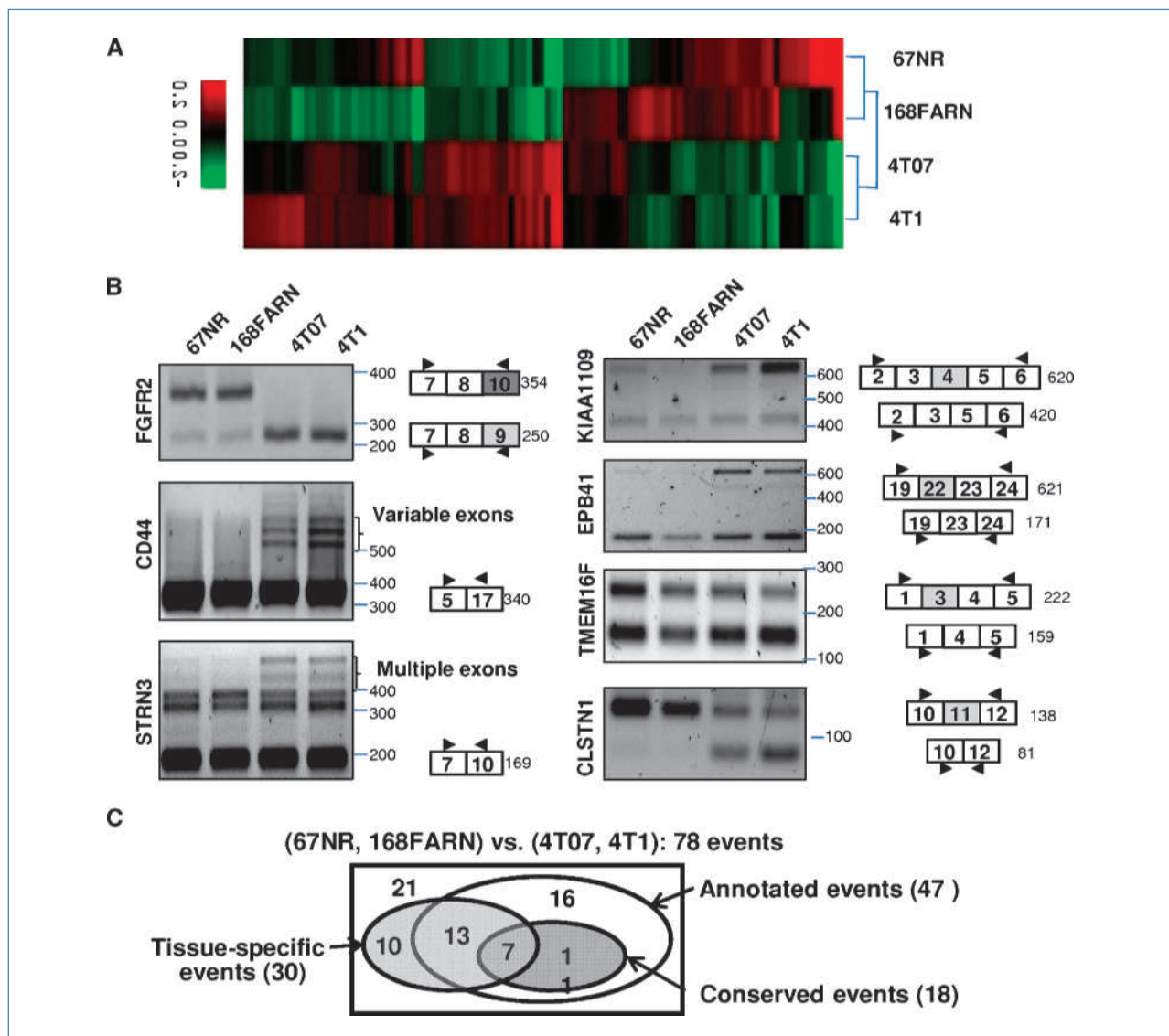


Figure 3. Validation and characterization of alternatively spliced exons associated with metastatic capability. A, hierarchical clustering of samples was performed with Mev4.0 software from The Institute of Genome Research using the gene-normalized exon intensities corresponding to the exons with the greatest intensity variations. B, RT-PCR analysis of alternatively spliced exons that were predicted to be differentially expressed in the (67NR, 168FARN) group compared with the (4T07, 4T1) group. The RT-PCR results are representative of three independent experiments. C, number of exons that were differentially expressed in the (67NR, 168FARN) group compared with the (4T07, 4T1) group and that were annotated in FAST DB as mouse alternative exons (Annotated events) or as human alternative exons (Conserved events) or as expressed in a tissue-specific manner (Tissue-specific events).

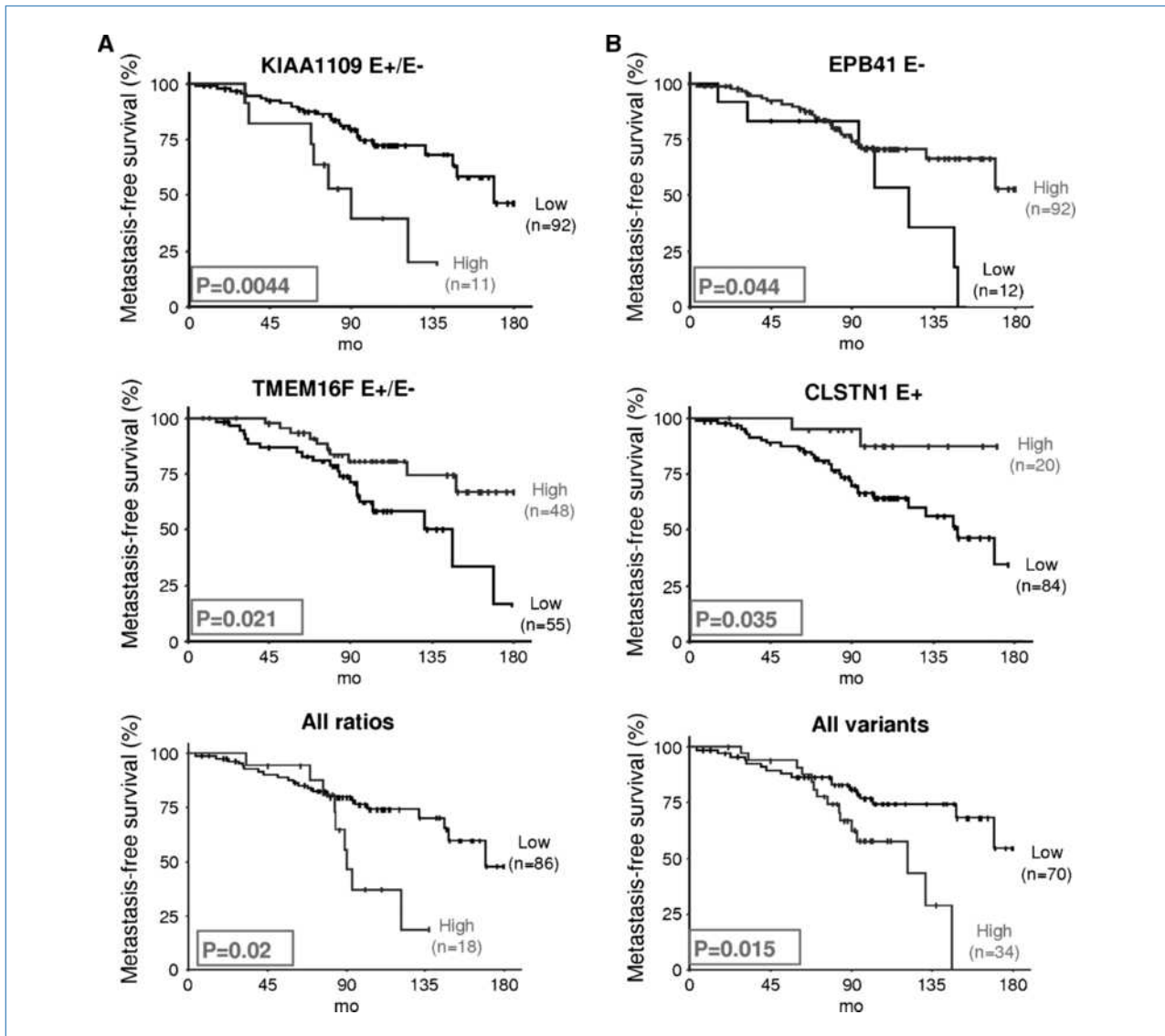


Figure 4. Kaplan-Meier curves for 15-y outcome in patients with breast cancer ($n = 104$) based on splicing variant content of *KIAA1109*, *EPB41*, *TMEM16F*, and *CLSTN1* genes. Metastasis-free survival curves according to (A) the inclusion/exclusion ratio of the *KIAA1109* (*KIAA1109* E+/E-) or *TMEM16F* (*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).

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

Table S3). 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 mammalian 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

- Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer* 2007;7:545–53.
- Bertucci F, Birnbaum D. Reasons for breast cancer heterogeneity. *J Biol* 2008;7:6.
- 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.
- Stamm S, Ben-Ari S, Rafalska I, et al. Function of alternative splicing. *Gene* 2005;344:1–20.
- 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.
- Fackenthal JD, Godley LA. Aberrant RNA splicing and its functional consequences in cancer cells. *Dis Model Mech* 2008;1:37–42.
- Gunning P, O'Neill G, Hardeman E. Tropomyosin-based regulation

- of the actin cytoskeleton in time and space. *Physiol Rev* 2008;88:1–35.
8. Paulson KE, Rieger-Christ K, McDevitt MA, et al. Alterations of the HBP1 transcriptional repressor are associated with invasive breast cancer. *Cancer Res* 2007;67:6136–45.
 9. Heider KH, Kuthan H, Stehle G, Munzert G. CD44v6: a target for antibody-based cancer therapy. *Cancer Immunol Immunother* 2004;53:567–79.
 10. Venables JP, Klinck R, Bramard A, et al. Identification of alternative splicing markers for breast cancer. *Cancer Res* 2009;68:9525–31.
 11. Clark TA, Schweitzer AC, Chen TX, et al. Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol* 2007;8:R64.
 12. Gardina PJ, Clark TA, Shimada B, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* 2006;7:325.
 13. French PJ, Peeters J, Horsman S, et al. Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays. *Cancer Res* 2007;67:5635–42.
 14. Thorsen K, Sorensen KD, Brems-Eskildsen AS, et al. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol Cell Proteomics* 2008;7:1214–24.
 15. Xi L, Feber A, Gupta V, et al. Whole genome exon arrays identify differential expression of alternatively spliced, cancer-related genes in lung cancer. *Nucleic Acids Res* 2008;36:6535–47.
 16. Fitzgibbons PL, Page DL, Weaver D, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med* 2000;124:966–78.
 17. Goldhirsch A, Glick JH, Gelber RD, et al. Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol* 2005;16:1159–83.
 18. Aslakson CJ, Miller FR. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res* 1992;52:1399–405.
 19. Eckhardt BL, Parker BS, van Laar RK, et al. Genomic analysis of a spontaneous model of breast cancer metastasis to bone reveals a role for the extracellular matrix. *Mol Cancer Res* 2005;3:1–13.
 20. Mani SA, Yang J, Brooks M, et al. Mesenchyme forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. *Proc Natl Acad Sci U S A* 2007;104:10069–74.
 21. de la Grange P, Dutertre M, Correa M, Auboeuf D. A new advance in alternative splicing databases: from catalogue to detailed analysis of regulation of expression and function of human alternative splicing variants. *BMC Bioinformatics* 2007;8:180.
 22. de la Grange P, Dutertre M, Martin N, Auboeuf D. FAST DB: a website resource for the study of the expression regulation of human gene products. *Nucleic Acids Res* 2005;33:4276–84.
 23. Peng YF, Mandai K, Sakisaka T, et al. Ankyrborin: a novel actin cytoskeleton-associated protein. *Genes Cells* 2000;5:1001–8.
 24. Kutty RK, Chen S, Samuel W, et al. Cell density-dependent nuclear/cytoplasmic localization of NORPEG (RAI14) protein. *Biochem Biophys Res Commun* 2006;345:1333–41.
 25. Yuan W, Zheng Y, Huo R, et al. Expression of a novel alternative transcript of the novel retinal pigment epithelial cell gene NORPEG in human testes. *Asian J Androl* 2005;7:277–88.
 26. Matsuoka Y, Li X, Bennett V. Adducin: structure, function and regulation. *Cell Mol Life Sci* 2000;57:884–95.
 27. Hashimoto-Uoshima M, Yan YZ, Schneider G, Aukhil I. The alternatively spliced domains EIIIB and EIIIA of human fibronectin affect cell adhesion and spreading. *J Cell Sci* 1997;110:2271–80.
 28. Liu XF, Ohno S, Miki T. Nucleotide exchange factor ECT2 regulates epithelial cell polarity. *Cell Signal* 2006;18:1604–15.
 29. Jung DH, Kim DH. Characterization of isoforms and genomic organization of mouse calumenin. *Gene* 2004;327:185–94.
 30. Lin T, Tang N, Ostap EM. Biochemical and motile properties of Myo1b splice isoforms. *J Biol Chem* 2005;280:41562–7.
 31. Fridy PC, Otto JC, Dollins DE, York JD. Cloning and characterization of two human VIP1-like inositol hexakisphosphate and diphosphoinositol pentakisphosphate kinases. *J Biol Chem* 2007;282:30754–62.
 32. Hartmann E, Gorlich D, Kostka S, et al. A tetrameric complex of membrane proteins in the endoplasmic reticulum. *Eur J Biochem* 1993;214:375–81.
 33. Franchi-Gazzola R, Dall'Asta V, Sala R, et al. The role of the neutral amino acid transporter SNAT2 in cell volume regulation. *Acta Physiol (Oxf)* 2006;187:273–83.
 34. Powell RM, Wicks J, Holloway JW, Holgate ST, Davies DE. The splicing and fate of ADAM33 transcripts in primary human airways fibroblasts. *Am J Respir Cell Mol Biol* 2004;31:13–21.
 35. Konecna A, Frischknecht R, Kinter J, et al. Calsyntenin-1 docks vesicular cargo to kinesin-1. *Mol Biol Cell* 2006;17:3651–63.
 36. Hulpiau P, van Roy F. Molecular evolution of the cadherin superfamily. *Int J Biochem Cell Biol* 2009;41:349–69.
 37. Vogt L, Schrimpf SP, Meskenaite V, et al. Calsyntenin-1, a proteolytically processed postsynaptic membrane protein with a cytoplasmic calcium-binding domain. *Mol Cell Neurosci* 2001;17:151–66.
 38. Moffa AB, Tannheimer SL, Ethier SP. Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. *Mol Cancer Res* 2004;2:643–52.
 39. Oltean S, Sorg BS, Albrecht T, et al. Alternative inclusion of fibroblast growth factor receptor 2 exon IIIc in Dunning prostate tumors reveals unexpected epithelial mesenchymal plasticity. *Proc Natl Acad Sci U S A* 2006;103:14116–21.
 40. Cha JY, Lambert QT, Reuther GW, Der CJ. Involvement of fibroblast growth factor receptor 2 isoform switching in mammary oncogenesis. *Mol Cancer Res* 2008;6:435–45.
 41. Chaffer CL, Dopheide B, Savagner P, Thompson EW, Williams ED. Aberrant fibroblast growth factor receptor signaling in bladder and other cancers. *Differentiation* 2007;75:831–42.
 42. Herrera-Gayol A, Jothy S. Adhesion proteins in the biology of breast cancer: contribution of CD44. *Exp Mol Pathol* 1999;66:149–56.
 43. Sanghamitra M, Talukder I, Singarapu N, Sindhu KV, Kateriya S, Goswami SK. WD-40 repeat protein SG2NA has multiple splice variants with tissue restricted and growth responsive properties. *Gene* 2008;420:48–56.
 44. Wei Y, Lin-Lee YC, Yang X, et al. Molecular cloning of Chinese hamster 1q31 chromosomal fragile site DNA that is important to mdr1 gene amplification reveals a novel gene whose expression is associated with spermatocyte and adipocyte differentiation. *Gene* 2006;372:44–52.
 45. Kuo MT, Wei Y, Yang X, et al. Association of fragile site-associated (FSA) gene expression with epithelial differentiation and tumor development. *Biochem Biophys Res Commun* 2006;340:887–93.
 46. Sun CX, Robb VA, Gutmann DH. Protein 4.1 tumor suppressors: getting a FERM grip on growth regulation. *J Cell Sci* 2002;115:3991–4000.
 47. Schischmanoff PO, Yaswen P, Parra MK, et al. Cell shape-dependent regulation of protein 4.1 alternative pre-mRNA splicing in mammary epithelial cells. *J Biol Chem* 1997;272:10254–9.
 48. Galindo BE, Vacquier VD. Phylogeny of the TMEM16 protein family: some members are overexpressed in cancer. *Int J Mol Med* 2005;16:919–24.
 49. Caputo A, Caci E, Ferrera L, et al. TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity. *Science* 2008;322:590–4.

Cancer Research

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

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 2010;70:896-905. Published OnlineFirst January 26, 2010.

Updated version Access the most recent version of this article at:
doi:[10.1158/0008-5472.CAN-09-2703](https://doi.org/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>

Cited articles This article cites 49 articles, 20 of which you can access for free at:
<http://cancerres.aacrjournals.org/content/70/3/896.full#ref-list-1>

Citing articles This article has been cited by 11 HighWire-hosted articles. Access the articles at:
<http://cancerres.aacrjournals.org/content/70/3/896.full#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.

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