

Multiple Alternative Splicing Markers for Ovarian Cancer

Roscoe Klinck,¹ Anne Bramard,¹ Lyna Inkel,¹ Geneviève Dufresne-Martin,¹ Julien Gervais-Bird,¹ Richard Madden,¹ Éric R. Paquet,¹ ChuShin Koh,¹ Julian P. Venables,¹ Panagiotis Prinos,¹ Manuela Jilaveanu-Pelmus,² Raymund Wellinger,³ Claudine Rancourt,³ Benoit Chabot,³ and Sherif Abou Elela³

¹Laboratoire de génomique fonctionnelle de l'Université de Sherbrooke, ²Département de pathologie, and ³Département de microbiologie et d'infectiologie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada

Abstract

Intense efforts are currently being directed toward profiling gene expression in the hope of developing better cancer markers and identifying potential drug targets. Here, we present a sensitive new approach for the identification of cancer signatures based on direct high-throughput reverse transcription-PCR validation of alternative splicing events. This layered and integrated system for splicing annotation (LISA) fills a gap between high-throughput microarray studies and high-sensitivity individual gene investigations, and was created to monitor the splicing of 600 cancer-associated genes in 25 normal and 21 serous ovarian cancer tissues. Out of >4,700 alternative splicing events screened, the LISA identified 48 events that were significantly associated with serous ovarian tumor tissues. In a further screen directed at 39 ovarian tissues containing cancer pathologies of various origins, our ovarian cancer splicing signature successfully distinguished all normal tissues from cancer. High-volume identification of cancer-associated splice forms by the LISA paves the way for the use of alternative splicing profiling to diagnose subtypes of cancer. [Cancer Res 2008;68(3):657–63]

Introduction

Epithelial ovarian cancer is the second most common gynecologic cancer and the deadliest among gynecologic pelvic malignancies. Early symptoms of ovarian cancer are often mild and unspecific, making this disease difficult to detect. In most cases, at the time of diagnosis, cancer cells have already disseminated throughout the peritoneal cavity. In fact, more than 70% of patients are diagnosed with late-stage disease and only a minority survive more than 5 years postdiagnosis (1), but early detection offers a 90% 5-year survival rate (2). Inability to detect ovarian cancer at an early stage and its propensity for peritoneal metastasis are largely responsible for these low survival rates. Currently, there are no reliable methods for detecting early-stage epithelial ovarian cancer. The CA125 serum marker combined with transvaginal ultrasonography are the usual clinical tests offered to screen for early stages of ovarian cancer in high-risk populations. These tests are used as predictors of clinical recurrence of ovarian cancers, and to monitor response to anticancer therapy. Neither of these strategies,

individually or combined, have proven reliable (3) and there is an urgent need to develop sensitive and specific screening tests for the early detection, prognosis, and tailored clinical management of epithelial ovarian cancer. Epithelial ovarian tumors are heterogeneous and include many different histopathologic subtypes: serous, endometrioid, mucinous, clear cell, undifferentiated, or mixed, the serous type being the most common and the second most lethal (4).

Cancer is in many ways a genetic disease and recent studies have focused on differences in molecular profiling of gene expression patterns to uncover diagnostic and prognostic markers as well as new therapeutic targets in a variety of cancers. Although promising results have been reported in some cancers, the genes that are differentially expressed between normal and cancer cells vary between individual microarray studies, reflecting either a variability in methods and in the choice of model systems or heterogeneity in selected tissues (5–7).

Alternative splicing of pre-mRNA, the production of distinct mRNAs from a single gene, is a posttranscriptional process that occurs in at least 70% of all genes to increase diversity of the proteome (8). Alternative splicing can also introduce or remove regulatory elements to affect mRNA translation, localization, or stability (9, 10). Alternative splicing is tightly regulated during development and between different tissues, and inherited and acquired changes in pre-mRNA splicing patterns have been associated with many human diseases (8, 11). Cancer-specific alterations in splice site selection affect genes controlling cellular proliferation (e.g., *FGFR2*, *p53*, *MDM2*, *FHIT*, and *BRCA1*), cellular adhesion, and invasion (e.g., *CD44*, *Ron*); angiogenesis (e.g., *VEGF*); apoptosis (e.g., *Fas*, *Bcl-x*, and *caspase-2*); and multidrug resistance (e.g., *MRP-1*; refs. 12–16). Some of these changes can arise from mutations at either the splice sites or within proximal splicing enhancer or silencer elements or from perturbations of *trans*-acting splicing factors (8). Examples of *trans*-acting factors that affect alternative splicing in cancer include ASF/SF2, PTB, and SRPK1 (17–20). Indeed, staining with PTB antibody was higher in invasive ovarian tumors than low malignant potential tumors and even lower in benign tumors (19). Thus, investigating cancer-specific splice variants may accelerate the discovery of novel diagnostic and prognostic cancer biomarkers, as well as identifying potentially new drug targets.

The first hints that global alternative splicing patterns were altered in cancer came from computational efforts designed to exploit collections of expressed sequence tag (EST) databases (reviewed in ref. 21) some of which were subsequently validated by reverse transcription-PCR (RT-PCR; ref. 22). Since then, there has been a concerted effort to annotate alternative splicing globally with oligonucleotide-based microarray technologies. Arrays with alternative splice junction probes have been used to detect splicing

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

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, Canada J1H 5N4. Phone: 819-564-5275; E-mail: sherif.abou.elela@usherbrooke.ca.

©2008 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-07-2580

changes in Hodgkin's lymphoma (23), breast cancer cell lines and xenografts (24), colon cancer (25), brain tumors (26), and leukemia (27). A related medium-throughput technique (Dasl) has been used to show that alternative splicing analysis can complement the power of gene expression analysis of prostate tumors (24). Microarray studies produce lists of putative alternative splicing events that must subsequently be validated by RT-PCR and usually only very few events are identified with high confidence. Indeed, the typical output of cancer-based studies is usually in the order of 10 validated alternative splicing events. Therefore, despite the low confidence in the accuracy of global gene expression signatures identified in ovarian cancer (5), classic gene expression remains the standard tool to distinguish ovarian tumors from normal tissues (28).

We report here the development of an automated high-throughput RT-PCR platform named the LISA (*l*ayered and *i*ntegrated system for *s*plicing *a*nnotation) capable of performing and analyzing 3,000 reactions per day. Using this integrated bioinformatics/robotics infrastructure, we have comprehensively analyzed alternative splicing in a set of 600 genes representing many confirmed cancer-associated genes or <4% of the human transcriptome. The sensitivity of this analysis led to the discovery of 48 highly cancer-specific splicing events. This set of potentially highly significant and biologically relevant splicing differences makes up a strong signature for epithelial ovarian cancer samples and amounts to a considerable augmentation of our knowledge of cancer-specific alternative splicing.

Materials and Methods

Reagents. All reagents were purchased from VWR International with the exception of the following: TRIzol reagent and Platinum Taq DNA polymerase (Invitrogen Canada, Inc.), Omniscript reverse transcriptase (Qiagen, Inc.), deoxynucleotide triphosphates (dNTP; GE Healthcare Bio-Sciences, Inc.), Power SYBR Green qPCR master mix (Applied Biosystems), and Porcine RNAsguard RNase inhibitor (Amersham Biosciences). Oligonucleotides were obtained from Integrated DNA Technologies. For capillary electrophoresis, RNA Nano chips and reagents were obtained from Agilent Technologies, and SE-30 DNA chips and reagents were obtained from Caliper LifeSciences.

Tissue sourcing and RNA extraction. Normal and serous epithelial ovarian cancer tissue samples were obtained as frozen specimens from the Réseau de Recherche sur le Cancer du Fonds de la recherche en santé du Québec biobank. Histopathology, grade, and stage were assigned according to the International Federation of Gynecology and Obstetrics criteria. From the 21 serous ovarian cancer tissues used as the training set, 20 were obtained from stage III and one tissue from stage I disease (Supplementary Table S1). Only chemotherapy-naïve tumor samples were used for the training set. Normal ovarian tissues were selected from women undergoing oophorectomy for reasons other than ovarian cancer, and the normality of the ovaries was confirmed by standard pathology tissue examination. Normal ovaries with benign tumors or cysts were excluded. Most of the donors were postmenopausal women of the age group when most serous tumors develop. Tumor and normal tissues were obtained from similar age group patients. For normal ovary tissues, age ranged from 39 to 84 years whereas the serous ovarian cancer tissues ranged from 42 to 79. The blinded set of 29 tumor and 10 normal tissues included 5 serous carcinomas, 1 serous benign tumor, 4 mixed carcinomas, 6 endometrioid carcinomas, 2 mucinous carcinomas, 3 serous low malignant potential tumors, 2 mucinous low malignant potential tumors, 1 undifferentiated adenocarcinoma, 1 teratoma, and 4 serous carcinomas of possible extraovarian origin (Table 1). For this set, no exclusion based on age or exposure to chemotherapy was made. RNA was extracted from 50 mg tissue samples using TRIzol reagent according to the manufacturer's protocol,

using a PowerMax homogenizing system equipped with a 10 mm saw tooth blade (VWR International). Extracted RNA was isopropanol precipitated, then resuspended in pure water and stored at -80°C . RNA concentration was quantified on an Agilent 2100 BioAnalyzer (Agilent Technologies).

Tissue quality control for the training set was established using real-time PCR amplification of known genes with known cancer or tissue type-specific expression profiles, including the epithelial cell markers *KRT7*, *KRT18*, and *CDH1*; the stromal marker *vimentin*; and the tumor cell content indicator *hTERT* (primer sequences available on request). *KRT7*, *KRT18*, and *CDH1* were shown to be up-regulated in high-grade serous ovarian cancer (29, 30). Expression data for all tissues used in the validation screen was subjected to unsupervised clustering using the R package.⁴ Four tumor tissues that failed, this quality control were considered to have low tumor tissue content and were omitted from the study. Eight normal and eight tumor RNA samples showing expression data closest to the median for each gene expression level for the normal or tumor tissue type were selected and combined in equal amounts to formulate two normal and two tumor pools of four samples each. No tissue quality control was performed on the blinded set.

Primer and PCR experiment design. For the 600 genes in this study, the AceView transcript set was mapped into the LISA database and the LISA automatically generated a splicing map (Supplementary Fig. S1B). Concurrently, the LISA applied a modified Primer-3-based (31) algorithm for the automated design of PCR primers for all exons in the transcript set. PCR experiments flanking all possible exon-exon junctions were designed. In addition, alternative splicing events were covered by at least two independent reactions. Where possible, the design was such that predicted amplicon sizes fell within the 100 to 400 bp range. The lower limit of 100 bp was set to avoid an overlap with primer and primer-dimer signals. Primers were synthesized by Integrated DNA Technologies in 96-well plates on a 10 nmol scale.

RT-PCR and capillary electrophoresis. Reverse transcription was performed on 2 μg total RNA samples in the presence of RNase inhibitor according to the manufacturers' protocols. Reactions were primed with both (dT)₂₁ and random hexamers at final concentrations of 1 and 0.9 $\mu\text{mol/L}$, respectively. The integrity of the cDNA was assessed by SYBR green-based quantitative PCR, done on three housekeeping genes: *MRPL19*, *PUM1*, and *GAPDH* (primer sequences available on request). C_t values for these genes, typically in the range of 14 to 25, depending on the gene, were used to verify the integrity of each cDNA sample. Following qPCR, the samples were analyzed by capillary electrophoresis to ensure that only one amplicon of the expected size was obtained.

End-point PCR reactions were done on 20 ng cDNA in 10 μL final volume containing 0.2 mmol/L each dNTP, 1.5 mmol/L MgCl_2 , 0.6 $\mu\text{mol/L}$ each primer, and 0.2 units of Taq DNA polymerase. An initial incubation of 2 min at 95°C was followed by 35 cycles at 94°C 30 s, 55°C 30 s, and 72°C 60 s. The amplification was completed by a 2-min incubation at 72°C . Relative quantitation using this method was verified for six alternative splicing events by comparison with isoform-specific quantitative PCR, and was found to be in strong agreement, to within 5% in all cases (data not shown).

PCR reactions are carried out using a liquid handling system linked to thermocyclers, and the amplified products were analyzed by automated chip-based microcapillary electrophoresis on Caliper LC-90 instruments (Caliper LifeSciences). Amplicon sizing and relative quantitation was performed by the manufacturer's software, before being uploaded to the LISA database.

Database, design, and analysis modules. The LISA was built around the LAMP solution stack of software programs (Linux operating system, Apache web server, Mysql database management server, and Perl and Python programming languages). In addition, several peripheral Perl and Python modules for experimental design, analysis, and display of results interact with the LISA. Statistical *t* tests and unsupervised clustering were performed using the R package.⁴ The capillary electrophoresis

⁴ <http://www.r-project.org>

Table 1. Predictions of ovarian specimens based on nearest centroid classification

Sample reference	Pathology	Distance to tumor	Distance to normal	Prediction
OVC 346	Serous carcinoma, gr 3, stage IIIC	95.43	142.66	Cancer
OVC 203	Serous carcinoma, gr 2, stage IIIC	116.88	138.60	Cancer
OVC 311	Serous carcinoma, gr 2, stage IV	145.62	165.33	Cancer
OVC 58	Serous LMP	122.81	179.12	Cancer
OVC 299	Serous LMP	168.09	229.21	Cancer
OVC 417	Serous carcinoma, gr 3, stage IIIC	118.37	135.89	Cancer
OVC 410	Endometrioid, gr 2	112.41	204.20	Cancer
OVC 37	Endometrioid, gr 2	103.95	170.50	Cancer
OVC 28	Endometrioid, gr 3	91.46	184.23	Cancer
OVC 17	Endometrioid, gr 3, stage IIIC	130.34	227.55	Cancer
OVC 98	Endometrioid, gr 1, stage IIC	114.25	187.55	Cancer
OVC 409	Mixed malignant serous-clear cell, gr 3, stage IIC	145.19	214.84	Cancer
OVC 130	Mixed malignant serous-mucinous	124.37	142.21	Cancer
OVC 385	Malignant mixed serous-clear cell, gr 1, stage IA	133.94	140.57	Cancer
OVC 126	Mucinous carcinoma, gr 2	102.06	213.80	Cancer
OVC 81	Mucinous carcinoma, stage IIB	98.11	188.01	Cancer
OVC 238	Mixed malignant serous-mucinous, gr 3, stage IIC	75.33	182.35	Cancer
OVC 2	Serous LMP	96.56	166.97	Cancer
OVC 161	Endometrioid carcinoma, gr 2, stage IIIC	129.08	193.40	Cancer
OVC 160	Teratoma	132.64	154.89	Cancer
OVC 47	Serous carcinoma, origin uncertain	117.55	127.65	Cancer
OVC 391	Undifferentiated adenocarcinoma, gr 3, stage IIIC	83.16	178.87	Cancer
OVC 78	Serous carcinoma, origin uncertain, gr 2, stage IV	149.63	135.61	Normal
OVC 121	Serous carcinoma, origin uncertain, gr 2, stage IIIA	156.58	110.55	Normal
OSC 536 BT	Serous carcinoma, origin uncertain	175.05	127.94	Normal
OVC 380	Mucinous LMP	170.45	159.97	Normal
OVC 182	Serous carcinoma, origin uncertain, gr 3, stage IIIC	211.33	101.31	Normal
OVC 374	Mixed mucinous LMP-Brenner	130.63	122.45	Normal
OVC 303	Serous benign	224.76	194.19	Normal
OVN 124	Normal	201.18	81.94	Normal
OVN 125	Normal	221.06	81.04	Normal
OVN 152	Normal	188.57	100.43	Normal
OVN 169	Normal	156.03	102.33	Normal
OVN 197	Normal	185.07	76.09	Normal
OVN 239	Normal	209.95	73.69	Normal
OVN 377	Normal	198.52	77.45	Normal
OVN 57	Normal	215.87	133.88	Normal
OVN 68	Normal	202.18	104.16	Normal
OVN 115	Normal	135.81	100.07	Normal

NOTE: Blind test of serous ovarian cancer alternative splicing markers. Thirty-nine normal and ovarian tumor tissues of the varying pathologies and origins indicated were classified based on their alternative splicing profile using the nearest centroid prediction rule (see Materials and Methods). Samples were designated normal or cancer depending on the shorter Euclidian distance (indicated in arbitrary units) to the normal or cancer centroid. Abbreviations: gr, grade; LMP, low malignant potential tumors.

instrument software provides size and concentration data for the detected peaks of each PCR reaction. These data are uploaded to the LISA database and compared with expected amplicon sizes for that experiment using a signal detection protocol and are thus stored in the LISA database with their corresponding concentrations as “assigned” or “unassigned” amplicons. In addition, gene sequence, primer sequence, single nucleotide polymorphism sites, and protein coding data are linked to each element of experimental data stored in the database. For each PCR reaction covering an alternative splicing event, the analysis module uses the concentration data from all RNA sources under consideration to determine the most prevalent assigned amplicon. For each RNA source, the ratio of the concentration of this amplicon to the total assigned amplicon concentrations measured is calculated and is expressed as a percentage, termed the percent splicing index, Ψ . Ψ values for normal

versus tumor RNA sources are used in statistical t tests, and resulting P values are used in the screening process to determine cancer specificity. Reactions showing a greater than 10 percentage point difference between the mean Ψ values for normal and tumor pools were selected for the validation screen. Following the validation screen, Ψ values were used in a t test for significant differences between normal and tumor tissue samples. Reaction sets with Bonferroni-corrected P values <0.0002 were selected.

Blinded set test. Tumor/normal class prediction was based on the “nearest-centroid prediction rule” (32). In brief, we defined two average profiles (tumor and normal) as vectors of the average Ψ values of the 48 alternative splicing events based on the clustering of the 46 samples. The sample was assigned a class label based on the smallest Euclidean distance between the sample and either one of the two average profiles.

Results

Layered and integrated system for splicing annotation. We have developed a system that provides unbiased systematic and comprehensive monitoring of splicing events, without relying on elaborate mathematical modeling for interpretation. The LISA is an automated high-throughput RT-PCR platform capable of executing 3,000 reactions per day, allowing a daily analysis of 750 alternative splicing events from four RNA sources (Supplementary Fig. S1A).

To identify cancer-associated splicing events, we first generated a list of 600 candidate genes from a keyword search for “ovarian cancer,” “breast cancer,” and “DNA damage” in the National Center for Biotechnology Information Entrez Gene database. These genes were entered into the LISA for the identification of alternative splicing events and experimental design. A transcript map for each gene was generated (Supplementary Fig. S1B) from publicly available mRNAs and ESTs, and sets of PCR primers and experiments were designed such that all putative exon-exon junctions and alternative splicing events were covered by at least two distinct PCR reactions. A total of 4,709 alternative splicing events were identified and an average of 33 primers and 30 PCR reactions were designed per gene. Most primer pairs (96%) amplified the expected products and this low failure rate was further offset by having at least two separate primer pairs overlapping each region. PCR amplification and subsequent capillary electrophoresis revealed the tissue-specific splicing patterns (Supplementary Fig. S1C). The percent splicing index (denoted Ψ) was then calculated and plotted as a heat map to show splicing changes across the tissues, and alternative splicing events showing significant changes were flagged (see Materials and Methods and Supplementary Fig. S1D). The different tools used for annotation and visualization of alternative splicing differences between different tissue samples are shown in Supplementary Fig. S2.

Comparative profiling of splicing events in normal and serous epithelial ovarian tumor tissues. As outlined in Fig. 1, we performed a screen in two stages: the first termed the “discovery screen” and the second, the “validation screen.” The discovery screen was carried out using two pools of RNA extracted from high-grade (grades 2 and 3) serous epithelial ovarian cancer specimens and two pools of RNA extracted from unmatched normal ovary sections (see Materials and Methods and Supplementary Table S1). Each pool contained an equivalent mix of four independent tissues.

Screening of the two cancer and two normal pools identified potentially cancer-associated alternative splicing events in 98 different genes. The range of shifts in Ψ values for these candidates was 10% to 70%. All results from the discovery screen are available as supplementary material.⁵ Alternative splicing events identified in the discovery screen were reexamined in a validation screen on individual RNA samples extracted from 21 serous epithelial ovarian cancer (grades 2 and 3) and 25 normal ovary sections. Overlapping PCR reactions using RNA from each tissue were carried out as in the discovery screen confirming the association of 48 alternative splicing events in 45 genes with ovarian cancer tissues. Supplementary Fig. S3 shows the distribution of Ψ values for the 48 hits and for 16 reactions that passed the discovery screen, but were not significantly cancer associated in the validation screen. Supplementary Table S2 presents the details of each of the 48 significant cancer-associated events issued from the validation screen. This

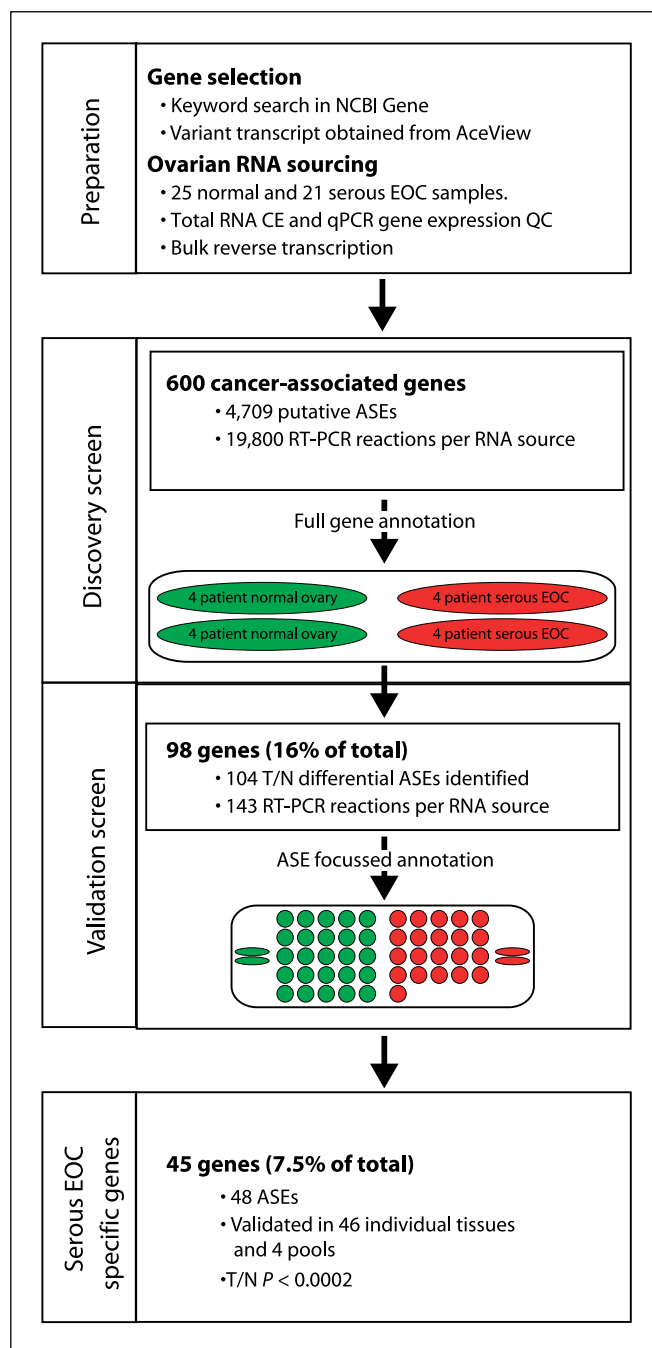


Figure 1. Four-phase pipeline for the identification of serous epithelial ovarian cancer (EOC)-associated alternative splicing events (ASE). A target gene set was identified and RNA from normal and serous epithelial ovarian cancer tissues was extracted. The discovery screen was conducted using two pools of RNA extracted from four different normal ovary tissues and two pools of RNA extracted from four different serous epithelial ovarian cancer tissues. A total of 600 cancer-associated genes were analyzed using a comprehensive set of PCR primers that span all possible splicing events. The alternative splicing events showing different profiles in normal and tumor (T/N) tissue pools were selected for the validation screen in 25 normal (green) and 21 EOC (red) samples. Forty-eight alternative splicing events with T/N Bonferroni-corrected P values < 0.0002 were identified from the initial 600-gene set.

suggests that 7.5% of our gene set harbors at least one ovarian cancer-specific splicing event.

To verify that our measurements of splicing shifts between tissue samples were not directly related to gene expression, the

⁵ <http://www.lgfus.ca/CAAS>

expression levels of a random selection of 11 genes with cancer-associated alternative splicing events were determined by quantitative PCR. As shown in Supplementary Fig. S4, expression levels varied <5-fold between the pools of the same type. In fact, for all gene measurements of expression levels in this sample, variation was within an order of magnitude, precluding the possibility of significant end point PCR measurement biases. Indeed, we did not detect any correlation between shifts in alternative splicing and expression levels in this data set, and conclude that expression and cancer-specific alternative splicing are distinctly regulated, as suggested previously for tissue-specific transcription and alternative splicing (33).

Use of alternative splicing to diagnose ovarian cancer. To assess the potential of the alternative splicing events identified by the LISA as diagnostic markers for ovarian cancer, we evaluated

their individual and collective capacity to accurately differentiate between normal and cancer tissues. The variance in the splicing pattern of each of the 48 identified ovarian cancer-associated splicing events (Supplementary Table S2), as calculated by their Ψ value, was used to classify the 46 individual and 4 pools of normal and tumor tissues based on splicing similarity (Fig. 2). Strikingly, all tumor tissues clustered together.

To evaluate the accuracy and applicability of the newly identified splicing signature, we blindly examined a set of 39 normal and ovarian tumor tissues of varying pathologies and origin, including benign tumors, low malignant potential tumors, and carcinomas of various histologic types (no prior tissue content quality control was used to select this group of tissues). The samples were processed as before using the LISA platform and the tissues were classified based on their alternative splicing profile using the nearest

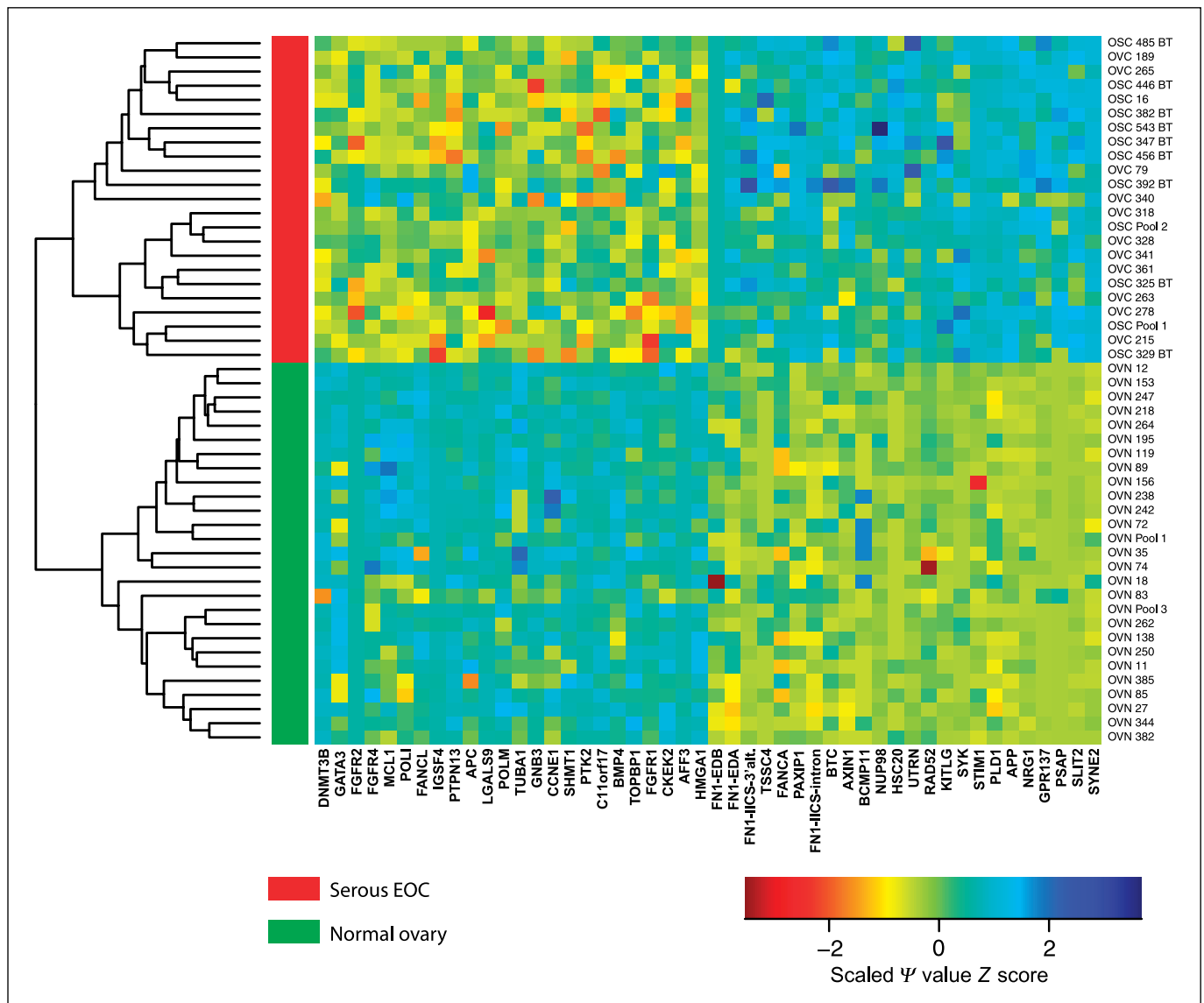


Figure 2. Summary of the serous epithelial ovarian cancer-associated alternative splicing events. Unsupervised clustering of ovarian cancer-associated variation of 48 alternative splicing events (*columns*) analyzed in 46 ovarian tissue samples (*rows*). The splicing variations are presented in scaled percent splicing index (Ψ) values shown in colors representing the number of SDs from the mean value (Z -score) for each column. Twenty-four alternative splicing events on the left side of the heat map have a shift to the longer form in cancer, whereas the 24 alternative splicing events to the right are shifted to the short isoform in cancer. Tissue hierarchical clustering is shown as a dendrogram (*left*). The column on the left shows clustered positions of normal ovary (*green*) and serous epithelial ovarian cancer tumor tissues (*red*).

centroid prediction rule (see Materials and Methods). By considering the shortest distance to either the normal or the cancer centroid, all 10 normal specimens were classified as normal. All the endometrioid (six of six), all mixed malignant cancers (four of four), and two of four mucinous cancers were correctly classified. The other two mucinous carcinomas, which were low malignant potential tumors, were classified as normal. Seven of the 11 malignant serous cancers were classified as cancer (Table 1). Interestingly, for the four malignant serous carcinoma tissues that were not classified as cancer, pathology could not define a clear ovarian origin. These tissues may have originated from an extraovarian source such as primary peritoneal carcinomatosis. The three serous low malignant potential tumors included in this study were classified as cancerous. The only serous benign tumor was classified as normal (Table 1). These data therefore confirm the utility of the alternative splicing signature as a general identifier of epithelial ovarian cancer and show the potential and sensitivity of this approach as a cancer subtype classifier.

Discussion

In the current study, we have used the LISA to provide high-quality comprehensive annotation of alternative splicing for 600 genes in 46 different tissues. Our analysis required nearly 100,000 RT-PCR reactions that were automatically carried out in a relatively short time. We based our comprehensive primer design on the AceView database as it has the most comprehensive EST coverage (34). In comparison with AceView, only 65% and 45% of our hits were represented in the smaller, curated, UCSC, and RefSeq databases, respectively. The fact that the AceView database is near complete for our purposes was tested by an extensive search through our data for novel cancer-associated splice events. One possible explanation for the lack of novel sequences recovered by this query is that we concentrated on well-known genes. It is possible that novel cancer-specific splicing events could be discovered in less well annotated genes.

In this study, we have shown that alternative splicing patterns can distinguish normal ovary from epithelial ovarian cancer. The number of cancer-specific splicing events identified exceeds the average number of events that is normally identified by splicing microarray profiling (25). The accuracy and applicability of the newly identified alternative splicing signature was shown by its capacity to identify ovarian cancer tissues (Table 1). The signature identified all normal tissues in a blinded set of 39 specimens, and 80% of the diverse cancer tissues regardless of their histologic subtypes. All cancer tissues that were identified as normal could be either of very low epithelial content, from an uncertain ovarian origin (Table 1), or possibly from an unknown tumor biology.

Our results identify some expected genes and many surprises; in ovarian tumors specifically, three genes have been reported to be regulated at the level of splicing. The splicing of the multidrug resistance protein ABCC1 (MRP1) and the *p53* regulator *MDM2* have both been shown to be disrupted in ovarian cancer and a large selection of internal deletions were found (12, 19, 35). The incapacity of the LISA to detect these previously reported alternative splicing events is possibly due to the preset stringency of the screen design, which was set to detect only events that are found in >25% of samples. Among our novel markers for ovarian cancer were several alternative splicing events that have been associated with other cancers. *Fibronectin* is known to have three

regions that are alternatively spliced. The EDB region is preferentially included in lung cancer (36), whereas the EDA region is preferentially included in liver cancer (37). The EDB and EDA exons are usually excluded in adult tissues with a major exception being ovary. Consistent with this, we found cancer-specific exclusion of the EDB exon. Interestingly, a recent splicing microarray study in colon cancer also found cancer-specific exclusion of EDB (25). Another known cancer-specific splice occurs to the α exon of *FGFR1*, which is specifically removed in glioma (38). By contrast, we found cancer-specific inclusion of this exon, as well as of the homologous exon of *FGFR2*. Alternative splicing of the *FGFR2* α exon has not been correlated with cancer before but intriguingly a recent genome-wide study associated its upstream intron with an increased risk of breast cancer (39). Alternative splicing of the *SYK* tyrosine kinase has been associated with breast cancer (40) and we found the same splicing shift in ovarian cancer. We also found alternative splicing of *DNMT3B* in ovarian cancer, in the same region, but a different isoform from that previously found in hepatocellular carcinoma (41). That short *DNMT3B* isoform (42) is lacking part of the catalytic DNA methyltransferase domain, including the TRD loop previously shown to be important for cytosine recognition (43), and is therefore inactive. The alternative splicing event in *KITLG* also affects function. In this case, the skipped exon encodes a metalloprotease cleavage site that determines whether KITLG will be membrane bound or secreted (44). The transmembrane form is more active in promoting cell-cell adhesion, cell proliferation, and survival because it induces more persistent tyrosine kinase activation than the secreted isoform (45).

Cancer-specific splicing events may drive cancer or may be a result of it (46). Another possibility is that cancer-specific splicing (and/or gene expression) reflects the tissue of origin. In an attempt to determine which applies to ovarian cancer, we studied primary cell cultures from normal ovarian surface epithelium tissues for our identified cancer-specific isoforms and found an equal representation of normal and cancer-specific events (Supplementary Table S2), implying that about half of our identified events may be truly cancer specific, whereas the other half may reflect the epithelial origin of the cancer (47). However, this interpretation remains uncertain as cells derived from normal ovarian surface epithelium tissues often adopt mesenchymal-like morphology due to their high degree of plasticity (48). Thirty-four hits (70%) are alternative splicing events that produce in-frame alterations (Supplementary Table S2), suggesting that alternative splicing of these genes contributes to ovarian tumor biology by creating proteins with novel or antagonistic functions (16). To extend this utility of alternative splicing events as markers, it will be of interest to correlate these with clinical variables in larger cohorts of patients who have suffered from serous epithelial ovarian cancer as well as other histopathologic subtypes.

Acknowledgments

Received 7/8/2007; revised 9/10/2007; accepted 10/4/2007.

Grant support: Genome Canada and Genome Quebec. B. Chabot is the Canada Research Chair in Functional Genomics. S. Abou Elela is a Chercheur-Boursier Senior of the Fonds de la recherche en santé du Québec. Tumor banking was supported by the Banque de tissus et de données of the Réseau de recherche sur le cancer of the Fonds de la recherche en santé du Québec.

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.

We thank J.P. Perreault for comments on the manuscript and Bernard Têtu for providing some of the serous ovarian cancer tissues.

References

1. Guppy AE, Nathan PD, Rustin GJ. Epithelial ovarian cancer: a review of current management. *Clin Oncol R Coll Radiol* 2005;17:399–411.
2. Bailey J, Taylor A, Naik R, et al. Risk of malignancy index for referral of ovarian cancer cases to a tertiary center: does it identify the correct cases? *Int J Gynecol Cancer* 2006;16 Suppl 1:30–4.
3. Nikolic B, Mitrovic A, Lazic J. Mean of CA 125 in making therapy decision in adnexal inflammatory tumors. *Bosn J Basic Med Sci* 2006;6:3–6.
4. Scott M, McCluggage WG. Current concepts in ovarian epithelial tumorigenesis: correlation between morphological and molecular data. *Histol Histopathol* 2006;21:81–92.
5. Agarwal R, Kaye SB. Expression profiling and individualisation of treatment for ovarian cancer. *Curr Opin Pharmacol* 2006;6:345–9.
6. Olivier RI, van Beurden M, van't Veer LJ. The role of gene expression profiling in the clinical management of ovarian cancer. *Eur J Cancer* 2006;42:2930–8.
7. 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.
8. Buratti E, Baralle M, Baralle FE. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res* 2006;34:3494–510.
9. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 2007;446:926–9.
10. Ni JZ, Grate L, Donohue JP, et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev* 2007;21:708–18.
11. Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. *Nat Biotechnol* 2004;22:535–46.
12. He X, Ee PL, Coon JS, Beck WT. Alternative splicing of the multidrug resistance protein 1/ATP binding cassette transporter subfamily gene in ovarian cancer creates functional splice variants and is associated with increased expression of the splicing factors PTB and SRp20. *Clin Cancer Res* 2004;10:4652–60.
13. 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.
14. Revil T, Shkreta L, Chabot B. [Pre-mRNA alternative splicing in cancer: functional impact, molecular mechanisms and therapeutic perspectives]. *Bull Cancer* 2006;93:909–19.
15. Skotheim RI, Nees M. Alternative splicing in cancer: Noise, functional, or systematic? *Int J Biochem Cell Biol* 2007;39:1432–49.
16. Venables JP. Unbalanced alternative splicing and its significance in cancer. *Bioessays* 2006;28:378–86.
17. 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.
18. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* 2007;14:185–93.
19. He X, Pool M, Darcy KM, et al. Knockdown of polypyrimidine tract-binding protein suppresses ovarian tumor cell growth and invasiveness *in vitro*. *Oncogene* 2007;26:4961–8.
20. Cheung HC, Corley LJ, Fuller GN, McCutcheon IE, Cote GJ. Polypyrimidine tract binding protein and Notch1 are independently re-expressed in glioma. *Mod Pathol* 2006;19:1034–41.
21. Kalnina Z, Zayakin P, Silina K, Line A. Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes Cancer* 2005;42:342–57.
22. Wang Z, Lo HS, Yang H, et al. Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. *Cancer Res* 2003;63:655–7.
23. Relogio A, Ben-Dov C, Baum M, et al. Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J Biol Chem* 2005;280:4779–84.
24. 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.
25. 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.
26. 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.
27. Milani L, Fredriksson M, Syvanen AC. Detection of alternatively spliced transcripts in leukemia cell lines by minisequencing on microarrays. *Clin Chem* 2006;52:202–11.
28. Olivier RI, Lubsen-Brandsma MA, Verhoef S, van Beurden M. CA125 and transvaginal ultrasound monitoring in high-risk women cannot prevent the diagnosis of advanced ovarian cancer. *Gynecol Oncol* 2006;100:20–6.
29. Ouellet V, Provencher DM, Maugard CM, et al. Discrimination between serous low malignant potential and invasive epithelial ovarian tumors using molecular profiling. *Oncogene* 2005;24:4672–87.
30. Sun PM, Wei LH, Luo MY, et al. The telomerase activity and expression of hTERT gene can serve as indicators in the anti-cancer treatment of human ovarian cancer. *Eur J Obstet Gynecol Reprod Biol* 2007;130:249–57.
31. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000;132:365–86.
32. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002;99:6567–72.
33. 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.
34. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* 2006;7 Suppl 1:S121–4.
35. Sigalas I, Calvert AH, Anderson JJ, Neal DE, Lunec J. Alternatively spliced mdm2 transcripts with loss of p53 binding domain sequences: transforming ability and frequent detection in human cancer. *Nat Med* 1996;2:912–7.
36. Khan ZA, Caurtero J, Barbin YP, Chan BM, Uniyal S, Chakrabarti S. ED-B fibronectin in non-small cell lung carcinoma. *Exp Lung Res* 2005;31:701–11.
37. Oyama F, Hirohashi S, Sakamoto M, Titani K, Sekiguchi K. Coordinate oncogene developmental modulation of alternative splicing of fibronectin pre-messenger RNA at ED-A, ED-B, and CS1 regions in human liver tumors. *Cancer Res* 1993;53:2005–11.
38. Jin W, Bi W, Huang ES, Cote GJ. Glioblastoma cell-specific expression of fibroblast growth factor receptor-1 β requires an intronic repressor of RNA splicing. *Cancer Res* 1999;59:316–9.
39. 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.
40. 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.
41. Saito Y, Kanai Y, Sakamoto M, Saito H, Ishii H, Hirohashi S. Overexpression of a splice variant of DNA methyltransferase 3b, DNMT3b4, associated with DNA hypomethylation on pericentromeric satellite regions during human hepatocarcinogenesis. *Proc Natl Acad Sci U S A* 2002;99:10060–5.
42. Oka M, Meacham AM, Hamazaki T, Rodic N, Chang LJ, Terada N. *De novo* DNA methyltransferases Dnmt3a and Dnmt3b primarily mediate the cytotoxic effect of 5-aza-2'-deoxycytidine. *Oncogene* 2005;24:3091–9.
43. Lappalainen I, Vihinen M. Structural basis of ICF-causing mutations in the methyltransferase domain of DNMT3B. *Protein Eng* 2002;15:1005–14.
44. Flanagan JG, Chan DC, Leder P. Transmembrane form of the kit ligand growth factor is determined by alternative splicing and is missing in the Sld mutant. *Cell* 1991;64:1025–35.
45. Miyazawa K, Williams DA, Gotoh A, Nishimaki J, Broxmeyer HE, Toyama K. Membrane-bound Steel factor induces more persistent tyrosine kinase activation and longer life span of c-kit gene-encoded protein than its soluble form. *Blood* 1995;85:641–9.
46. Venables JP. Aberrant and alternative splicing in cancer. *Cancer Res* 2004;64:7647–54.
47. Christiansen JJ, Rajasekaran AK. Reassessing epithelial to mesenchymal transition as a prerequisite for carcinoma invasion and metastasis. *Cancer Res* 2006;66:8319–26.
48. Auersperg N, Wong AS, Choi KC, Kang SK, Leung PC. Ovarian surface epithelium: biology, endocrinology, and pathology. *Endocr Rev* 2001;22:255–88.