Identification of Alternative Splicing Events Regulated by the Oncogenic Factor SRSF1 in Lung Cancer

Fernando J. de Miguel¹, Ravi D. Sharma⁴, María J. Pajares¹,², Luis M. Montuenga¹,², Angel Rubio⁴, and Ruben Pio¹,³

Abstract
Abnormal alternative splicing has been associated with cancer. Genome-wide microarrays can be used to detect differential splicing events. In this study, we have developed ExonPointer, an algorithm that uses data from exon and junction probes to identify annotated cassette exons. We used the algorithm to profile differential splicing events in lung adenocarcinoma A549 cells after downregulation of the oncogenic serine/arginine-rich splicing factor 1 (SRSF1). Data were generated using two different microarray platforms. The PCR-based validation rate of the top 20 ranked genes was 60% and 100%. Functional enrichment analyses found a substantial number of splicing events in genes related to RNA metabolism. These analyses also identified genes associated with cancer and developmental and hereditary disorders, as well as biologic processes such as cell division, apoptosis, and proliferation. Most of the top 20 ranked genes were validated in other adenocarcinoma and squamous cell lung cancer cells, with validation rates of 80% to 95% and 70% to 75%, respectively. Moreover, the analysis allowed us to identify four genes, ATP11C, IQCB1, TUBD1, and proline-rich coiled-coil 2C (PRRC2C), with a significantly different pattern of alternative splicing in primary non–small cell lung tumors compared with normal lung tissue. In the case of PRRC2C, SRSF1 downregulation led to the skipping of an exon overexpressed in primary lung tumors. Specific siRNA downregulation of the exon-containing variant significantly reduced cell growth. In conclusion, using a novel analytical tool, we have identified new splicing events regulated by the oncogenic splicing factor SRSF1 in lung cancer. Cancer Res; 74(4); 1–11. © 2013 AACR.

Introduction
Alternative splicing is the process by which different mRNA isoforms are processed from a single primary transcript. In humans, estimates indicate that as many as 95% of multiexonic genes are alternatively spliced (1). Cassette exon, in which an exon is either retained or spliced out of the transcript, is the most common alternative splicing event in mammals, and accounts for approximately half of all events (2–4). Alternative splicing has an essential role in the regulation of cell homeostasis. When this process becomes deregulated it can lead to pathologic conditions such as cancer (5–8). Deregression of splicing is often associated with an abnormal expression pattern of splicing factors (9). One of these factors is serine/arginine-rich splicing factor 1 (SRSF1). This molecule is a member of the serine/arginine-rich protein family, which is involved in constitutive and alternative splicing (10). The SRSF1 protein contains 2 RNA recognition motives (RRM) at the N-terminal and one arginine/serine rich domain at the C-terminal. The arginine/serine rich domain mediates protein–protein interactions in the spliceosome assembly (11). The SRSF1 protein is also implicated in other functions, such as nonsense-mediated RNA decay, translation and RNA transport, genome stability, and senescence (12, 13). The expression of SRSF1 is upregulated in several human neoplasias and participates in the establishment and maintenance of a transformed phenotype (14, 15). Consequently, the study of the alternative splicing events regulated by this factor is of particular interest.

A number of genome-wide array approaches have been developed to study changes in alternative splicing (16). These arrays contain probes in exons (exon arrays), probes along the whole genome (tiling arrays), or probes in exons and junctions (junction arrays). The characteristics of the splicing arrays developed to date are summarized in Supplementary Table S1. The analytical process for the evaluation of differential splicing events is a key aspect to consider when using these arrays. A serious limitation of exon arrays is the exclusion of an exon by the absence of a signal. In contrast, a microarray design that includes junction probes indicates the exclusion of an exon by the hybridization of probes to the
skipping junction. Consequently, the current trend is to work with arrays that include probes in annotated junctions. Unfortunately, there are very few algorithms that take advantage of the redundancy provided by junction probes (17–25).

In this study we have developed an algorithm, named ExonPointer, which combines data from exon and junction probes and was optimized to identify annotated cassette exons. We used the algorithm to identify differential alternative cassette events in lung adenocarcinoma A549 cells after the downregulation of SRSF1. Expression data were generated using 2 platforms: Oryzon custom arrays (Agilent technology) and GeneSplice arrays (Affymetrix technology). Validation rates were 60% and 100% for Agilent and Affymetrix platforms, respectively, which demonstrate the effectiveness of this procedure in finding alternative splicing events. In addition, most of these events were validated in other lung cancer cell lines and, more importantly, allowed us to identify genes with significant differences in splicing between primary lung tumors and normal lung tissue. Finally, functional studies demonstrated the implication of one splicing isoform of proline-rich coiled-coil 2C (PRRC2C), one of the SRSF1-regulated genes, in cell proliferation. In conclusion, we have developed a useful tool for the analysis of alternative splicing and have identified new splicing events regulated by the oncogenic splicing factor SRSF1.

Materials and Methods

Biological material

Primary lung tumors were obtained from patients with nonsmall cell lung cancer who had been treated with resectional surgery at the Clínica Universidad de Navarra (Pamplona, Spain). Clinopathologic characteristics of the patients are shown in Supplementary Table S2. Nontumor lung specimens were sampled at a distance from the tumor to guarantee that the tissues were free from cancerous cells. Tissue samples were immediately frozen in liquid nitrogen and stored at −80°C until extraction of RNA. Samples were included in the study if greater than 70% of their cells were tumor cells. The study protocol was approved by the Institutional Ethical Committee and all patients gave written informed consent. RNA extraction was performed as previously described (26).

Lung adenocarcinoma cell lines A549, HCC827, and H2087 were obtained from the American Type Culture Collection. Squamous cell carcinoma cell lines HCC95 and LUDLU-1 were obtained from the Korean Cell Line Bank and the European Collection of Cell Cultures, respectively. Cell lines were authenticated by analysis of their genetic alterations. Cells were grown in RPMI supplemented with 2 mmol/L glutamine, 10% FetalClone (Thermo), 100 U/mL of penicillin and 100 μg/mL of streptomycin (Invitrogen).

Downregulation of SRSF1 and expression analysis

Cells were transfected with a 30 nmol/L solution of a nontargeting scramble siRNA (5’-AGGACAGGAGTTCGCTT-3’) or an SRSF1 siRNA (5’-TGAGGAGGTGATGTATGT-3’) using lipofectamine as described by the manufacturer (Invitrogen). Cells were incubated for 48 hours after transfection. Total RNA was extracted using the RNA/DNA Mini Kit (Qiagen). The efficiency of the SRSF1 siRNA knockdown was determined by real-time PCR, using SYBR Green PCR Master Mix and the following primers: 5’-GGAAACACG- GATTGGCGCATCTA-3’ (forward); 5’-CTTGAAGTGATGTAG- TGCGGGATA-3’ (reverse).

Microarray hybridization

Samples from 3 independent experiments using lipofectamine alone, scramble siRNA, or SRSF1 siRNA (i.e., a total of nine samples) were labeled and hybridized in 2 different platforms: Agilent custom-built microarrays from Oryzon Genomics and Affymetrix GeneSplice arrays. The former were developed by Oryzon Genomics in collaboration with our group, and have been previously described (27). These arrays contain exon probes and thermodynamically balanced junction probes. The noncommercial Human GeneSplice Arrays were obtained through a Technology Access program with Affymetrix. These high-density arrays contain an average of 8 probes per probe-set. RNA labeling and hybridization on the Agilent microarrays were performed by Oryzon Genomics, as previously described (27). Labeling and hybridization on the Affymetrix GeneSplice microarrays were performed by the Proteomics, Genomics and Bioinformatics Core Facility of the Center for Applied Medical Research (CIMA) following manufacturer’s instructions. The microarray data from this study have been submitted to Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE53410.

ExonPointer algorithm

The ExonPointer algorithm was used to process and analyze the microarray data. The aim in the design of ExonPointer was to identify differential alternative splicing, i.e., genes that show different relative concentrations of their isoforms, under different conditions. Fig. 1 illustrates the summarized steps of ExonPointer. The mathematical definition of differential alternative splicing and the analytical steps are described in detail in Supplementary Methods. Probes that did not indicate expression levels greater than the background noise were excluded from the analysis. Short genes with no more than 2 exons (i.e., no more than 3 probe sets) were also excluded. Supplementary Fig. S1 shows a representative example of the report generated by the algorithm, which points to the differential splicing event in a given gene.

Downregulation, expression analysis, and sequencing of PRR2C

Cells were cultured and transfected as described in section “Down-regulation of SRSF1 and expression analysis.” The siRNA sequences were as follows: PRR2C 5’-CGAACACGCA-GUCCGCAA[da][da]-3’; PRR2C2-L (long) 5’-GGAAGCAG- CUCAGCCCAT[da][da]-3’; PRR2C2-S (short) 5’-CGACCCCA- GGCAAGAC[da][da]-3’. Primers used to quantify and sequence the PRR2C2-L variant were: 5’-GTCCAGCAAAAT- GAACAGCA-3’ (forward) and 5’-GAGTCTCTTCTCCACAGC- TCCT-3’ (reverse). Primers used to quantify and sequence the PRR2C2-S variant were: 5’-ACCACGCAAGCAGAG-3’ (forward) and 5’-GATCGGCTAGTTGTATG-3’ (reverse).
PCR products were sequenced in the Proteomics, Genomics, and Bioinformatics Core Facility of CIMA.

**MTT and clonogenic assays**

Cells were cultured and transfected as described in section "Downregulation of SRSF1 and expression analysis." Six hours after transfection, cells were trypsinized and seeded in 96-well plates at 1,500 cells/well for MTT assays. Thiazolyl Blue Tetrazolium Bromide (MTT; Sigma) was added at 0, 24, 48, 72, 96, and 120 hours, at a final concentration of 0.5 mg/mL. After 4 hours of incubation, 100 μL of solubilization solution (10% sodium dodecyl sulfate, 50% N,N-dimethylformamide, pH 4.7) was added. Absorbance was measured at 540 nm using 690 nm as a reference. For clonogenic assays, 6 hours after transfection, cells were trypsinized and seeded in 6-well plates at 300 cells/well. Culture medium was replaced after 1 week of incubation and removed after 2 weeks. Clones were fixed with 1 mL of formaldehyde/well for 30 minutes, dyed with few drops of crystal violet for 10 minutes, and counted.

**Cell-cycle and apoptosis analyses**

Cells were cultured and transfected as described in section "Downregulation of SRSF1 and expression analysis," and the medium was replaced after 6 hours. At the indicated times, suspended and attached cells were harvested. For apoptosis analysis, cells were washed and stained with 10 μg/mL of propidium iodide (Sigma) and 5 μL of Annexin V (BD Biosciences). For cell-cycle analysis, cells were fixed with 70% ethanol for at least 1 hour at 4°C and treated with 0.2 mg/mL RNase A (Sigma) for 1 hour at 37°C. Cells were then stained with 10 μg/mL of propidium iodide (Sigma). All samples were analyzed on a FACSCalibur Flow cytometer (Becton Dickinson). Percentages of cells in sub-G0–G1, G0–G1, S, and G2–M, for cell-cycle analysis, or in quadrants, for apoptosis analysis, were determined with FlowJo 9.3 software (Tree Star).

**Enrichment analyses**

The DAVID Functional Annotation Clustering application was used to create clusters in accordance with annotations from several databases (28). The DAVID Functional
Annotation tool was run to obtain \( P \) values for enriched annotation terms. Ingenuity Pathway Analysis from Ingenuity Systems (http://www.ingenuity.com/products/ipa) was used to perform enrichment analyses of networks and functions. Whole genome databases of the respective tools were used as a reference for the analyses.

**Statistical analyses**

The statistical analysis for the selection of significant alternative splicing events is detailed in the description of the ExonPointer algorithm (Supplementary Methods). Differences in the expression of splice variants in cell lines were assessed by native splicing events is detailed in the description of the Wilk test. The Wilcoxon signed rank test was used to determine statistical significance of the differences between these two groups. For MTT assays, the natural logarithm was calculated for each growth curve to obtain a straight line. Linear regression was performed for each condition and the slopes were compared with the Student \( t \) test. Clonogenic assays were analyzed with the Student \( t \) test. All tests were performed in STATA/IC 12.1 software (StataCorp).

**Results**

We developed the ExonPointer algorithm to identify genome-wide differential alternative splicing events in A549 cells in which the levels of the oncogenic SRSF1 were downregulated with a specific siRNA. The efficiency of downregulation was assessed by real-time PCR (Supplementary Fig. S2A). The splicing pattern of caspase-9 (CASP9) was used as a positive control to confirm the validity of our model. Caspase-9 has two splice variants that differ in four consecutive exons, proapoptotic caspase-9a, and anti-apoptotic caspase-9b. It has been reported that, in A549 cells, SRSF1 regulates the inclusion (CASP9a) or exclusion (CASP9b) of this exon cassette (29). In accordance with the literature, downregulation of SRSF1 induced a marked increase of the CASP9b variant (Supplementary Fig. S2B).

Microarray data were obtained from three independent experiments using two different platforms: Agilent custom-built microarrays from Oryzon Genomics and Affymetrix GeneSplice arrays. The ExonPointer algorithm was used to process and analyze the results. The number of cassette events, after the application of the selection criteria, was 24,205 for Oryzon and GeneSplice arrays. The total number of genes was 4,062 and 7,326 for Oryzon and GeneSplice, respectively. Of these, 2,300 were common to both platforms. For each platform, events were ranked in accordance with their \( P \) values. For validation, we selected the top 20 events ranked by ExonPointer in the two microarray platforms. Only one gene, PPP5K2, was common to both shortlists. For each splicing event, a PCR was run with primers designed in the exons that flank the exon of interest, and the ratio between the higher (exon inclusion) and lower band (exon exclusion) was calculated. Twelve of the 20 genes (60% validation rate) selected by ExonPointer from the Oryzon platform data showed differential alternative splicing (Supplementary Fig. S3 and Table S3). The other 8 genes showed differences in expression or no observable changes (data not shown). Results were consistent in all genes, that is the validated events were found in all the triplicates, and, conversely, the events that were not validated did not appear in any of the samples. Among the validated genes, SRSF1 downregulation resulted in exon exclusion in 9 genes (NPR3L, PBP1, DHPS, CHD3, MYOB1, RAB2B, TMPO, MATR3, and Cxorf26) and exon inclusion in 3 genes (FN1, WNK1, and PPP5K2). These changes were confirmed using samples from two independent experiments with A549 cells transfected under the same experimental conditions as those used in the microarray studies (data not shown).

The same validation strategy was followed for the genes selected by ExonPointer from the GeneSplice platform data. Remarkably, 100% of the top 20 events were validated (Table 1 and Fig. 2) and the results were fully consistent among triplicates. Exclusion was again the predominant effect of SRSF1 downregulation, with 15 genes that showed an increase in exon exclusion (ATP11C, SHH, ABCD4, FIP1L1, TFDPI, IQCBI, MVP17L, NAGK, RAD51C, PARBPB, ASAP1, EWSR1, TUBD1, PRRC2C, and USP8), and only 5 genes that showed an increase in exon retention (MYCIP2, MORF4L2, PPP5K2, MAPT, and SETD5), These changes were confirmed using samples from the 2 independent experiments (data not shown). To establish a reliable limit for the validity of the ranked list generated by ExonPointer using GeneSplice microarray data, we validated 20 genes placed at lower positions in the list. We chose the genes listed at positions 96–100, 146–150, 196–200, and 246–250. We were able to validate 12 of these 20 genes (Supplementary Fig. S4). Exon exclusion was again the most frequent event caused by SRSF1 downregulation (11 of the 12 validated events). The top 250 events selected by ExonPointer were used to perform an enrichment analysis of the cellular networks and functions affected by the downregulation of SRSF1 in A549 cells. In the DAVID Functional Annotation Chart, the 2 most common and enriched annotations were "alternative splicing" and "splice variant" (Fig. 3). Within the top annotations, "RNA binding," "mRNA processing," "Nucleotide recognition," and "RNA recognition motives" were present. These results indicated that SRSF1 regulates splicing in pathways that are intrinsically implicated in RNA metabolism. "Phosphoprotein" was also on the top of the enriched annotations. The ingenuity analysis of the differentially spliced genes regulated by SRSF1 built a total of 11 networks. The first contained 22 genes and included the following annotations: "Cell Morphology, Developmental Disorder, and Hereditary Disorder" (Supplementary Fig. S5). The next two networks contained 23 and 21 genes highly involved in tumorigenesis and cancer development. The annotations included in these two networks were: "Cell Cycle, Cellular Assembly, and Organization and Cellular Movement" for the second network (Supplementary Fig. S6); and "Organismal Development, Cancer, Cell Death, and Survival" for the third network (Supplementary Fig. S7). Many genes involved in these networks have functions related to RNA metabolism (Table 2). Functional annotations related to infection were also at the top of the rank, although its biologic relevance needs to be established.

We further assessed the reliability of the list generated from the GeneSplice data by searching for events described in the
Table 1. List of the top 20 ranked genes selected by ExonPointer from the expression data obtained using GeneSplice arrays

<table>
<thead>
<tr>
<th>Rank</th>
<th>Gene</th>
<th>P value</th>
<th>Exon</th>
<th>Ensembl transcript ID</th>
<th>Event</th>
<th>A549</th>
<th>HCC827</th>
<th>H2087</th>
<th>HCC95</th>
<th>LUDLU-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MYCBP2</td>
<td>1.78e–91</td>
<td>57</td>
<td>201</td>
<td>Inclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>ATP11C</td>
<td>4.78e–87</td>
<td>20</td>
<td>007</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>SSH1</td>
<td>2.10e–82</td>
<td>11</td>
<td>001</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>ABCD4</td>
<td>9.01e–79</td>
<td>8</td>
<td>001</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>FIP1L1</td>
<td>9.35e–73</td>
<td>2</td>
<td>005</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>TFDPI</td>
<td>7.90e–72</td>
<td>10</td>
<td>202</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>MORF4L2</td>
<td>1.07e–71</td>
<td>3</td>
<td>006</td>
<td>Inclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>PPIPSK2</td>
<td>1.73e–71</td>
<td>24</td>
<td>201</td>
<td>Inclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>IQCB1</td>
<td>2.70e–70</td>
<td>12</td>
<td>001</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>MPV17L</td>
<td>3.20e–70</td>
<td>2</td>
<td>002</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>NAGK</td>
<td>1.38e–69</td>
<td>3</td>
<td>201</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>RAD51C</td>
<td>2.56e–69</td>
<td>4</td>
<td>001</td>
<td>Exclusion</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>MAPT</td>
<td>1.55e–66</td>
<td>8</td>
<td>201</td>
<td>Inclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>PARPB</td>
<td>1.40e–65</td>
<td>5</td>
<td>201</td>
<td>Inclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>ASAP1</td>
<td>1.11e–60</td>
<td>2</td>
<td>002</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>16</td>
<td>EWSR1</td>
<td>5.74e–60</td>
<td>9</td>
<td>002</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>17</td>
<td>TUBD1</td>
<td>1.48e–59</td>
<td>4</td>
<td>201</td>
<td>Inclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>SETD5</td>
<td>1.48e–58</td>
<td>4</td>
<td>020</td>
<td>Inclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>19</td>
<td>PRRC2C</td>
<td>6.32e–58</td>
<td>34</td>
<td>201</td>
<td>Inclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>USP8</td>
<td>1.40e–55</td>
<td>12</td>
<td>003</td>
<td>Exclusion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

literature as targets of SRSF1. We first analyzed the genes reported by Karni and colleagues in their seminal paper describing the oncogenic potential of SRSF1 (14). Interestingly, even though the cell models were different (A549 vs. IMR90 and HeLa cells), most of the events described in their study were correctly identified by ExonPointer in our model (Supplementary Table S4). Other genes described in the literature as targets of SRSF1 were also detected in our analysis: MAPT, P = 1.55e–66 (30); CD44, P = 2.04e–22 (31); EPP4, P = 2.22e–22 (32); SRSF3, P = 1.91e–21 (33); CASP9, P = 4.79e–21 (29); BCL2L1, P = 5.03e–03 (34); MCL1, P = 1.13e–17 (34); and FNI, P = 3.32e–15 (35). These data reveal the notable capacity of our procedure to identify alternative splicing events regulated by SRSF1.

To confirm the relevance of the information obtained from lung adenocarcinoma A549 cells, we validated the results in four additional lung cancer cell lines: two adenocarcinomas (HCC827 and H2087) and two squamous cell lung carcinomas (HCC95 and LUDLU-1). Results of the validation after SRSF1 downregulation are shown in Table 1. Most of the genes were validated in both histologic subtypes, although, as expected, validation rates were higher in adenocarcinoma cell lines (95% and 80%) than in squamous cell lines (75% and 70%).

Human cell lines retain many properties of the cells of origin, but they also show clear differences and are not fully representative of the in vivo tumors (36). Therefore, we evaluated the clinical relevance of our findings by determining the presence of the identified splicing events in a cohort of primary lung tumors and their corresponding normal lung tissues. After confirming the overexpression of SRSF1 in these tumors (Fig. 4A), we found a significant deregulation in the splicing of four genes: ATP11C, IQCB1, TUBD1, and PRRC2C (Fig. 4B–E).
PRRC2C-L (Fig. 5F). An increase in the apoptotic sub-G₀–G₁ peak was also detected (Fig. 5F). This increase in the rate of apoptosis was confirmed using Annexin V (Fig. 5G). Comparable results were obtained with lung adenocarcinoma HCC827 and H2087 cells (Supplementary Fig. S8A–S8F). In lung squamous carcinoma HCC95 cells, downregulation of PRRC2C resulted in a reduction of in vitro tumor growth, mostly detected in clonogenic assays, although no differences were observed between the 2 splice variants (Supplementary Fig. S8G and S8H). No differences were observed in lung squamous carcinoma LUDLU-1 cells (data not shown). Taken together, these data suggest that the splicing of PRRC2C is involved in the oncogenic activity of SRSF1.

**Discussion**

We have developed the ExonPointer algorithm to identify differentially spliced genes in data obtained from genome-wide
exon-junction microarray technologies. We applied this algorithm to analyze differential alternative cassette events regulated by SRSF1 in the context of lung cancer. The experimental validation rates using data from both Oryzon (Agilent) and GeneSplice (Affymetrix) arrays demonstrated the efficacy of ExonPointer, although GeneSplice performed noticeably better. Agilent probes are longer and more sensitive than Affymetrix probes (37). However, the number of probes per array, and consequently the number of probes used to interrogate each splicing event, was considerably smaller in the Oryzon arrays than in the GeneSplice arrays (40,000 vs. 6,000,000, respectively). This characteristic counterbalances the drawback of shorter and less-specific probes. During the implementation of ExonPointer we also found that the use of junction probes (particularly those that skip the exon) was fundamental for the success of the algorithm. Other aspects of ExonPointer deserve to be mentioned. The main source of false positives was weakly expressed sequences (i.e., probes with low signal) and the filtering of these probes drastically improved the results. The combination of $P$ values of several probes (as opposed to the combination of signals of the probes) was proven to be an effective approach to the quantification of splicing events. The algorithm exploits the redundancy of the junctions, and effectively combines them with the exon probes. ExonPointer is an extension of a linear model and, therefore, can be applied to any experiment where a linear model holds (most of the present algorithms are focused on case–control studies). Finally, the characteristics of ExonPointer allow for a straightforward implementation to other microarray platforms, provided that they contain exon and junction probes (e.g., the GeneChip Human Transcriptome 2.0 array, recently launched by Affymetrix).

RNA sequencing (RNA-seq) is an alternative genome-wide approach used to characterize alternative splicing. The RNA-seq technology is able to generate millions of reads that can be mapped to the whole genome using specific RNA aligners (38). This process requires a laborious tuning of numerous parameters involved in each of the steps. Using this approach, approximately 300 million reads would be required to accurately quantify 90% of the transcriptome in a single sample (39). With the decrease in cost and the improvement in computing resources, read length and algorithms, RNA-seq will become a major player in the identification of alternative splicing events, but, at present, storage and computational requirements can be unaffordable if the number of samples is high. Alternatively, if the focus is on the detection of annotated splicing events in a reliable and cost effective manner, a microarray platform with an adequate design is a very viable option (16). Despite the limitations of microarrays (e.g., the results are tied to the probes that are included in the array design, based on the existing annotations of gene structures), we have demonstrated that this technology performs well, requires only modest computing power and is simple to interpret. In less than 1 hour of analysis, ExonPointer obtained a remarkably reliable list of differential splicing events.

Figure 3. DAVID Functional Annotation Chart derived from the analysis of differential alternative splicing events after SRSF1 downregulation. Annotations are sorted by significance. A, each column depicts $-\log_{10}$ of $P$ values for the corresponding annotation. B, number of analyzed genes present in each annotation. The database for each annotation is shown in parenthesis: SP, SwissProt; UP, UniProt; GO, Gene Ontology; SM, SMART; IP, InterPro.
The use of ExonPointer allowed us to study the alterations in cassette exons caused by the downregulation of SRSF1 in lung cancer cells. Upregulation of SRSF1 is frequent in cancer and contributes to the oncogenic potential of H-Ras and Myc (14, 40). Overexpression of SRSF1 in human non–small cell lung cancers leads to a more invasive phenotype (15, 41). Through the control of caspase-9 alternative splicing, SRSF1 also regulates the sensitivity of lung cancer cells to chemotherapy (42). The serine/arginine-rich proteins are best known for their ability to promote exon inclusion (43). In accordance with this observation, our results show that SRSF1 mainly functions as a splicing repressor in A549 cells. This is also consistent with the exon-skipping phenotypes of disease-associated mutations that affect the SRSF1-specific splicing motifs (44). The splicing analysis described here also provides new information on genes and functions regulated by SRSF1, which helps to understand the transforming capacity of SRSF1. In our study we identified a substantial number of splicing events in genes related to RNA metabolism. This supports the data reported by Sanford and colleagues, which showed an enrichment for genes that encode RNA binding proteins within a group of alternative exons targeted by SRSF1 (45). In agreement with our functional analysis, the authors also found enrichment for genes associated with biologic processes such as cell division, apoptosis, and proliferation. To our knowledge, the implication of SRSF1 in the regulation of alternative splicing in genes related to developmental and hereditary disorders is a novel observation. The association with functional annotations related to phosphoproteins and infection also merit further attention.

In support of the validity of our approach, the ExonPointer algorithm found statistically significant differences in many

### Table 2. Ingenuity function enrichment analysis

<table>
<thead>
<tr>
<th>Rank</th>
<th>Function annotation</th>
<th>Category</th>
<th>P value</th>
<th>No. genes*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Accumulation of RNA</td>
<td>Molecular transport</td>
<td>1.16e–07</td>
<td>5/16</td>
</tr>
<tr>
<td>2</td>
<td>Infection of embryonic cell lines</td>
<td>Infection of embryonic cell lines</td>
<td>1.01e–04</td>
<td>12/325</td>
</tr>
<tr>
<td>3</td>
<td>Infection of epithelial cell lines</td>
<td>Dermatological diseases and conditions</td>
<td>1.01e–04</td>
<td>12/325</td>
</tr>
<tr>
<td>4</td>
<td>Infection of kidney cell lines</td>
<td>Infectious disease</td>
<td>1.42e–04</td>
<td>12/325</td>
</tr>
<tr>
<td>5</td>
<td>Accumulation of mRNA</td>
<td>Molecular transport</td>
<td>5.94e–04</td>
<td>2/5</td>
</tr>
<tr>
<td>6</td>
<td>Processing of RNA</td>
<td>RNA trafficking</td>
<td>8.27e–04</td>
<td>10/310</td>
</tr>
<tr>
<td>7</td>
<td>Tetramerization of protein</td>
<td>RNA trafficking</td>
<td>8.86e–04</td>
<td>5/78</td>
</tr>
<tr>
<td>8</td>
<td>Formation of cervical cancer cell lines</td>
<td>Cellular growth and proliferation</td>
<td>9.84e–04</td>
<td>2/6</td>
</tr>
<tr>
<td>9</td>
<td>Formation of nuclear pores</td>
<td>Cellular assembly and organization</td>
<td>9.84e–04</td>
<td>2/5</td>
</tr>
<tr>
<td>10</td>
<td>Transport of axons</td>
<td>Cellular assembly and organization</td>
<td>9.84e–04</td>
<td>2/6</td>
</tr>
<tr>
<td>11</td>
<td>Delay in initiation of prometaphase</td>
<td>Cell cycle</td>
<td>1.47e–03</td>
<td>2/6</td>
</tr>
</tbody>
</table>

*Number of genes in the list that belong to the functional annotation/total number of genes in the functional annotation.

#### Figure 4. Differential alternative splicing events in a cohort of primary lung tumors and their corresponding normal lung tissues (n = 6). A, SRSF1 expression measured by real-time PCR using IPO8 as housekeeping gene. B–E, PCR validation of differential alternative splicing events in ATP11C, IQBC1, TUBD1, and PRRC2C, respectively. The ratio of the densitometry values is shown. F, ratios of PRRC2C-L/PRRC2C-S expression, measured by real-time PCR, in an extended cohort of patients (n = 16). *P < 0.05.

The use of ExonPointer allowed us to study the alterations in cassette exons caused by the downregulation of SRSF1 in lung cancer cells. Upregulation of SRSF1 is frequent in cancer and contributes to the oncogenic potential of H-Ras and Myc (14, 40). Overexpression of SRSF1 in human non–small cell lung cancers leads to a more invasive phenotype (15, 41). Through the control of caspase-9 alternative splicing, SRSF1 also regulates the sensitivity of lung cancer cells to chemotherapy (42). The serine/arginine-rich proteins are best known for their ability to promote exon inclusion (43). In accordance with this observation, our results show that SRSF1 mainly functions as a splicing repressor in A549 cells. This is also consistent with the exon-skipping phenotypes of disease-associated mutations that affect the SRSF1-specific splicing motifs (44). The splicing analysis described here also provides new information on genes and functions regulated by SRSF1, which helps to understand the transforming capacity of SRSF1. In our study we identified a substantial number of splicing events in genes related to RNA metabolism. This supports the data reported by Sanford and colleagues, which showed an enrichment for genes that encode RNA binding proteins within a group of alternative exons targeted by SRSF1 (45). In agreement with our functional analysis, the authors also found enrichment for genes associated with biologic processes such as cell division, apoptosis, and proliferation. To our knowledge, the implication of SRSF1 in the regulation of alternative splicing in genes related to developmental and hereditary disorders is a novel observation. The association with functional annotations related to phosphoproteins and infection also merit further attention.

In support of the validity of our approach, the ExonPointer algorithm found statistically significant differences in many
genes previously reported to be regulated by SRSF1 (14, 29–35). Nevertheless, most of the splicing events that we found were novel. Moreover, the analysis allowed us to identify 4 genes, ATP11C, IQCB1, TUBD1, and PRRC2C, with a significantly different pattern of splicing in primary non–small cell lung tumors. The association of these molecules with lung cancer had not been previously reported, although, in the case of PRRC2C, there was some information suggesting its role in cancer. PRRC2C is amplified and overexpressed in bladder cancer (46), and it has been related to proliferation and cell-cycle regulation through an interaction with histone H4 transcription factor (47). We have now been able to identify a new splice variant of PRRC2C, which is regulated by SRSF1 and is overexpressed in primary lung tumors. We have also shown that PRRC2C is able to regulate lung cancer growth in vitro. Interestingly, in lung adenocarcinoma cells this activity is mostly restricted to the splice variant overexpressed in primary lung tumors. The SRSF1 protein has been reported to module...
p53 by interfering with its murine double minute 2 (MDM2)-dependent proteasomal degradation (13), and there is evidence for an interaction between PRRC2C and MDM2 (48). The role of the PRRC2C splice variants in the regulation of p53 by MDM2 merits further investigation.

In conclusion, exon–junction microarrays, together with the appropriate algorithm, provide a reliable analytical tool for genome-wide examination of differential alternative splicing in a time and cost-effective manner. This technology has allowed us to identify new splicing events regulated by the oncogenic genome-wide examination of differential alternative splicing in normal and pathologic human biology.

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

Authors’ Contributions
Conception and design: F.J. de Miguel, R.D. Sharma, L.M. Montuenga, A. Rubio, R. Pio

Acknowledgments
The authors thank A. Lavin and C. Sainz for technical support, and Dr. L. Ortiz for performing the Affymetrix array hybridizations.

Grant Support
This work has been supported by “UTE project CIMA,” the Spanish Government (ISCIII-RTICC RD12/0036/0040, PI10/00166, and PI1100618), and the European Community’s Seventh Framework Programme (HEALTH-F2-2010-258677-CURELUNG). F.J. de Miguel is supported by a predoctoral fellowship from the Instituto de Salud Carlos III and Ministerio de Economía y Competitividad.

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 May 22, 2013; revised December 4, 2013; accepted December 9, 2013; published OnlineFirst December 26, 2013.

References


Identification of Alternative Splicing Events Regulated by the Oncogenic Factor SRSF1 in Lung Cancer

Fernando J. de Miguel, Ravi D. Sharma, María J. Pajares, et al.

Cancer Res Published OnlineFirst December 26, 2013.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-13-1481

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2013/12/30/0008-5472.CAN-13-1481.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.