An ESRP-regulated splicing program is abrogated during the Epithelial Mesenchymal Transition

Claude C. Warzecha, Peng Jiang, Karine Amirikian, Kimberly A. Dittmar, Hezhe Lu, Shihao Shen, Wei Guo, Yi Xing, Russ P. Carstens

Supplemental Data

Figure S1. Additional validations of microarray predicted ESRP regulated splicing events. Semi-quantitative RT-PCR was used to assay a set of more complex candidate alternative splicing events. Identities of PCR products are labeled to the right of the gels. The percent inclusion of the relevant exon(s), shaded in red, is given below each lane. For the mutually exclusive splicing events in the FYN and MAPK14 gene transcripts, PCR products were analyzed by restriction digest in order to distinguish the different splice variants of similar size (U—uncut; B—BclI; M—MwoI). In these cases, percent Exon 7b or Exon 9a inclusion was calculated as \([\text{MwoI resistant band} / (\text{MwoI resistant band} + \text{BclI resistant band})]\). For the mutually exclusive splicing event in the RAB6A gene transcript, PCR products were digested with restriction enzymes PstI (P) or TspRI (T). Percent Exon 4a inclusion was calculated as \([\text{TspRI resistant band} / (\text{TspRI resistant band} + \text{PstI resistant band})]\). Primer sequences are listed in the Supplemental Experimental Procedures.

Figure S2. Location of ESRP binding motifs generally correlates with functions as splicing enhancers of splicing silencer. Alternative exons that showed a \(\geq 10\%\) change in exon inclusion levels upon validation were selected for motif mapping. The 5mers TGGTG, GGTGG, and GTGGT are represented by blue, green, and red bars respectively. The region mapped was restricted to the exon body and 250nt of flanking intronic sequence. A) ESRP enhanced exons demonstrating that candidate ESRP binding motifs are predominantly located in the downstream intron and largely excluded from the exons. B) ESRP silenced exons demonstrating that the motifs are most commonly seen in the exon body and the upstream intron.

Figure S3. Alternative splicing of non-ESRP targets is not altered upon sustained knockdown of the ESRPs. RT-PCR was used to assay inclusion of alternative exons in the PUF60, RFX2, and ZNF236 gene transcripts in HMLE cells after an EMT was induced via sustained knockdown of both ESRP proteins. Identities of PCR products are labeled to the right of the gels.

Figure S4. Analysis of ESRP regulated alternative splicing events and ESRP 1 and 2 expression across a panel of epithelial and mesenchymal cell lines. Complete set of splicing events profiled to derive signature events presented in Figure 5. A) Analysis of exons predicted to be spliced in epithelial cells. The bar graph presents exon inclusion levels for each event assayed across six epithelial cell lines (purple bars) and three mesenchymal cell lines (gold bars) as determined by HT RT-PCR. B) Analysis of exons predicted to be skipped in epithelial cells, presented as in A. C) Gene expression levels for ESRP1 and ESRP2 in 51 breast cancer cell lines. Median centered mRNA expression levels are shown on a log\(_2\) scale. The 51 cell lines are organized by subclass as defined in (Neve et al., 2006). Of note, these data also correlate with expression of epithelial markers and inversely with mesenchymal cell markers in the same dataset as reviewed in (Blick et al., 2008).
Figure S2A Warzecha et al.

ESRP enhanced exons

TSC2  CCC125  MPZL1  GOLGA2  EXOC1  AKAP11  NUMB  OSBP3  KIAA0494  USP4  MAP3K7  PDGFA  ARHGEF12  BBX  TTC3  LSM14A  RALGPS2  MCAT  VPS39  ZNF267  APBB2  YAP1  UAP1  CSNK1E  FNP1  KIAA1217  ZNF608  RAB18  MACF1  CEP77  GLMN  KIF13A  PARL  ASL  CCAR1  MINK1  C10orf137  ACSL3  RAPGEF1  METTL8  NBEAL2  COX4NB  PCT2  DDHD1  HISPPD1  TRERF1  ARFIP1  SLC37A2  MYO9A  DOCK9  TCFL72  PUM2  INTS9  MACF1  HMGX4  FAM62B  ITGA6  ARFGAP2  BTN2A1  KIAA1618  PXDN  ADD3  IQCB1

UGGUG
GGUGG
GUGGU
Figure S2B  Warzecha et al.

ESRP silenced exons

C6orf106
IL7R
DALRD3
POFUT1
SPINT2
JTB
FGFR1OP
ZDHHC16
TAI1
GRAMD1A
WDFY3
JAG2
UBE2K
ANKRD26
SPA9
GALE
ITSN1
RAPGEF2
STX8P2
LASS1L
MAP3K7
SEP15
WDR74
HMBS
ACA1
RIPK2
VARS
MRPL23
SLC24A6
AKR1GPF11
PARP6
MOSPD1
FAM128A
MST1R
CISD1
MAP7
PACSL2
LRBP1P2
SO2
COL26A1
STX2
PDA3
EHBP1
C10orf47
NADK
ATP6V0A1
PAR1
RBM4
ZNF207
ATP13A3
PPF1B1P1
CISL1N
MEST
ST7
POLC3
PPT2
MAGUI
ZFYVE9
MYO9B
WDR19
TBC1D23
NAC1
PRC1
ACADVL
ELOVL1
SKP2
PBX2
ATF2B4
FAM13B1
OSBPL3
REP51
MYH10
MITD1
NGLY1
PSAT1
WDR32
KIF13A
MST4
FAM13B8
MPRIP
BTN2A1
EPN2
PFKM
QTRT1D1
MAP3K9
PLAA
LAS1L
OSBPL9

UGGUG
GGUGG
GUGGU
Figure S3  Warzecha et al.

**PUF60**
- shControl
- shESRP1+2

**RFX2**
- shControl
- shESRP1+2

**ZNF236**
- shControl
- shESRP1+2
Figure S4  Warzecha et al.

A

Predicted epithelial exons

B

Predicted mesenchymal exons

C

ESRP1

ESRP2

Luminal  Basal A  Basal B
Table S1. Complete set of results from HJAY analysis in MDA-MB-231 cells with ectopic Esrp1 expression and in PNT2 cells with ESRP1 and ESRP2 knockdown with RNAi. Results include tabs to indicate predicted cassette, alternative 3’ and 5’, and mutually exclusive exons. Also shown are exons detected in both arrays, including 104 cassette exons with the “expected” pattern in which the predicted effect of ESRP expression (Control in PNT2 cells and ectopic Esrp1 expression in MDA-MB-231 cells) is the same in both experiments. We also noted 25 cassette exons in which the direction of splicing change was “unexpected” in that the direction of change was the same with ESRP ectopic expression and knockdown. In the tab for “unexpected” changes in splicing we present validations for seven of the 25 exons with the unexpected HJAY predicted changes in splicing that were tested in both cell types and 4 of these validated the apparent contradictory function of the ESRPs in the different cell types, although only 2 cases showed changes of over 5% in exon inclusion. Of the remaining 3 cases the results included three false positives from the PNT2 knockdown results, thereby showing consistent effects of the ESRPs in both cell types. Column headings include Affymetrix assigned transcript IDs and junction probeset and exon probeset IDs as well as specific exons IDs. The columns following each probeset indicate the direction of change in signal (‘-’ or ‘+’) that occurs in response to ESRP1 and 2 knockdown (PNT2 datasets) or ectopic expression of Esrp1 (MDA-MB-231 datasets). In the case of the cassette exon worksheets the predicted effect of ESRP expression is summarized as “enhance” or “silence” in a separate column. Each dataset is ranked by the lowest p-value among the probesets that predict the change in splicing. Note that several exons were identified on the basis of more than one probeset combination and hence the total numbers in the worksheets for cassette exons slightly exceed the stated number of exons identified.

Table S2. Enrichment of ACEScan and high RSPR (RNA Selection Pressure Ratio) conserved alternative exons (ACE) among HJAY predicted ESRP-regulated exons. Complete list of HJAY predicted ESRP target cassette exons that are ACEScan alternative cassette exons (ACEs) and those with a high RNA Selection Pressure Ratio (RSPR). Table include tabs to indicate ESRP predicted target genes and exons that are present in each set as well as complete statistics outlining the total exon numbers in the HJAY predicted and background exons in both the ESRP knockdown and overexpression experiments. Also shown are statistics showing the percent of ACescan and RSPR exons in the ESRP-regulated and background exons on the array that passed the same expression threshold.

Table S3. Validations of ESRP-regulated cassette exons detected in MDA-MB-231 cells upon Esrp1 ectopic expression. Full set of validated exons in addition to those shown in Figure 2 and Table 1. Complete listing of all cassette exons submitted for validation by HT-RT-PCR. Tabs representing validated are all exons with changes in splicing consistent with the HJAY prediction, whereas those that pass a higher confidence level of 5% change or listed in a separate tab. Splicing events listed as indeterminate are those in which the expected RT-PCR products were not observed or those in which the results otherwise did not confidently verify or refute the prediction. Validated events are listed based on percent change in exon splicing from most silenced to most enhanced.

Table S4. Validations of ESRP-regulated cassette exons and alternative 3’ and 5’ splice sites detected in PNT2 cells with ESRP1 and ESRP2 knockdown by RNAi. Full set of validated exons in addition to those in Figure 2 and Table 1. Results as described for Table S3, but based on PNT2 cell HJAY predictions. For these validations we performed two RT-PCR reactions for each target to ensure
increased detection of validated events and therefore ranked validated events based on the mean change in splicing where both RT-PCR reactions yielded the expected products. HT-RT-PCR validations from predicted alternative 3’ and 5’ splice sites are also shown in a separate tab.

Table S5. Candidate ESRP binding motif analysis among enriched hexamer sequences located in the upstream intron, exon, or downstream intron in ESRP-enhanced exons and ESRP-silenced exons. Motifs in each position were ranked based on Fisher p-value by nucleotide. Shown are motifs in each set selected for p-value <0.05 after Bonferroni correction for multiple comparisons (0.05/4096 possible motifs). Columns representing density of each motif in or flanking regulated exons and the reference cassette exons are shown. Data are shown separately for ESRP-enhanced and ESRP-silenced exons. Motifs in each position were ranked based on Fisher p-value by nucleotide. We also determined the mean branch length score (BLS) for each motif in the regulated and control exons to look for evidence of phylogenetic conservation of each motif in flanking intron sequences. These results demonstrated increased BLS of the candidate motifs near the regulated exons compared to the reference set.

Table S6. Overlap between ESRP and Fox regulated cassette exons that function both cooperatively and antagonistically. Based on enrichment of the binding motif for Fox family splicing regulators in the introns flanking ESRP-regulated exons as shown in Figure 3, we examined the overlap in regulation of exons by both factors. Of 123 previously validated Fox-regulated exons, 48 (39%) were identified in our HJAY analysis as ESRP regulated exons. Consistent with the Fox binding motif being the most enriched candidate downstream of silenced exons, the direction of splicing of most of these exons (27) by these factors was opposite. However, 21 exons were predicted to be cooperatively enhanced or silenced by both regulators. Exon sizes and coordinates are also indicated and HJAY predicted events that are validated are shown in blue type. Experimentally demonstrated Fox-1 or Fox-2 mediated changes in splicing were previously described (Baraniak et al., 2006; Nakahata and Kawamoto, 2005; Venables et al., 2009; Yeo et al., 2009; Zhang et al., 2008).

Table S8. GO analysis for predicted ESRP regulated genes. Gene Ontology Analysis of ESRP regulated exons as performed using the DAVID functional annotation tool. Because of the different transcriptomes in PNT2 and MDA-MB-231 cells we separately searched for enriched terms and processes associated the the HJAY predicted cassette exons in each cell type using the background of transcripts that passed the expression filter in each cell type. P-values and fold enrichment for GO terms associated with genes containing ESRP predicted cassette exons are shown. We also used the DAVID tool to identify enriched protein domains and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Table S7. Examples of ESRP-regulated exons in gene transcripts in which the resulting isoforms have known or suggestive evidence of differential functions. These examples include several with profound demonstrated functional distinctions and in processes linked to the EMT.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Functions</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMB</td>
<td>Numb has previously been shown to be required for maintenance of cell-cell adhesion through interactions with E-cadherin at adherens junctions. The interaction of Numb with E-cadherin occurs through an N-terminal phosphotyrosine binding (PTB) domain in Numb that is abrogated by a specific phosphorylation event in a cytoplasmic domain of E-cadherin. Loss of this Numb-E-cadherin interaction was thus proposed to be required for</td>
<td>(Wang et al., 2009) (Dho et al., 1999; Nie et al., 2004)</td>
</tr>
</tbody>
</table>
targeting of E-cadherin to the basolateral membrane and uncoupling of this interaction is likely to contribute to loss of cell polarity and cell-cell adhesion in early EMT events. The 33 nt epithelial-specific exon in Numb encodes an 11 amino acid insert in this PTB domain that was shown to be required for binding to the PDZ1 domain in the E3 ubiquitin ligase LNX. Thus, while not directly shown, it might be similarly proposed that loss of this insert and alterations in the PTB binding domain during the EMT may also abrogate the Numb-E-Cadherin interaction and thereby promote loss of cell-cell-adhesion and promote cell motility. Consistent with this possibility, isoforms that contain this insert in PTB1 are predominantly associated with the plasma membrane, whereas isoforms that exclude it are diffusely cytoplasmic.

**ITGA6**
Integrins are integral cell surface proteins that mediate interactions with extracellular matrix and integrin signaling pathways play major roles in cell adhesion and migration. Integrins consist of an alpha chain that is paired with a beta chain and the different heterodimers form numerous different integrins. Integrin alpha 6 typically forms heterodimers with either the integrin beta 1 or integrin beta 4 subunit to form alpha6beta1 or alpha6beta4 molecules. In addition to noting that multiple ESRP target genes function in integrin signaling pathways, we also noted a nearly switch-like change in splicing of a penultimate exon in integrin alpha 6 (ITGA6) that results in different cytoplasmic C-terminal peptides. The A isoform that is enforced by the ESRPs encodes a classical PDZ binding domain, whereas the B isoform was shown to encode an isoform with an atypical PDZ domain that was nonetheless also able to bind the PDZ domain of the cytoplasmic PDZ domain protein TIP2/GIPC as well as the A isoform. Thus, while no specific changes in binding to either isoform have been shown, it is likely that they may show differential binding to other cytoplasmic PDZ domain proteins that affect integrin alpha6-beta 1 and or alpha6-beta 4 signaling pathways.

**MAP3K7**
Transforming growth factor beta-activated kinase 1 (TAK1; official symbol MAP3K7) was first characterized in its role in TGF-beta signaling pathways, but also participates in a number of other signaling pathways as well. Our data showed that the ESRPs regulated both of the well characterized alternative cassette exons that give rise to four splice variants termed isoforms A-D. While the ESRPs promote inclusion of the in frame exon 12, they also silence the 116 nt exon 16 thereby inducing a frameshift in the 3’ terminal exon that results in protein isoforms with significantly divergent C-termini. While there is currently no direct evidence that isoforms that specifically lack exon 16 function differentially, a recent study demonstrated that peptide sequences that are present only in isoforms (A and B) that include the exon are required for interaction with upstream signaling proteins TAB2 and TAB3. Because most previous studies of TAK1 mediated signaling have used these longer isoforms, this study strongly implies that the shorter isoforms that lack TAB2/3 binding will be shown to have different cytokine-mediated signaling properties.

**MPZL1**
MPZL1, also known as Protein Zero Related (PZR), is an Ig superfamily transmembrane protein that has two well documented splice variants that include (PZR) or skip (PZRb) the 103 nt exon that is promoted by ESRP

---

(Tani and Mercurio, 2001)

(Besse et al., 2007)

(Zannettino et al., 2003; Zhao and Zhao, 2003)
expression. Isoforms that skip the exon (PZRb) have a truncated intracellular C-terminal domain that is devoid of the immunoreceptor tyrosine-based inhibitory motifs (ITIMs) that bind to SH-P. As a result, these shorter isoforms have been shown to have dominant negative effect on phosphorylation of the full length protein. Furthermore, it was proposed that these ITIMs play an important role in regulation of integrin mediated cell motility, whereas the truncated isoforms fail to regulate cell motility.

**BIN1**
Bin1 is an adapter molecule that interacts with the Myc oncoprotein and for which there is evidence that this interactions inhibits Myc oncogenic activity, thereby negatively regulating cell proliferation and malignancy. Bin1 isoforms that include alternative exon 12A, however, do not bind Myc and therefore do not have this tumor suppressor activity. Interestingly, a previous study showed that overexpression of the splicing factor caused increased inclusion of exon 12A; one of several alterations it induced in alternative splicing suggesting that SF2/ASF functions as a proto-oncogene. Conversely, we observed that the ESRPs promoted skipping of exon 12A, which suggests that this might be one piece of evidence that they may function as tumor suppressors through maintenance of splicing pathways that negatively regulate cell proliferation. It is notable in this regard that we reproducibly observe significantly slower growth rates in all of the immortalized cell lines into which we introduce cDNAs that express Esrp1 driven by the CMV promoter. Therefore, in addition to promoting epithelial cell phenotypes and differentiation, the Esrps may also retard malignancy.

**ENAH**
Enabled homolog (ENAH), also known as hMena, is an Ena/Vasp family member that has well documented roles in cytoskeletal organization and cell migration. The ESRP-regulated exon, referred to as exon 11a, is specifically downregulated in invasive cells compared the primary tumors in in vivo tumor invasion models. The exon is present within a C-terminal Eva/Vasp homology domain (EVH2) and adjacent to an F-actin binding site, leading to speculation that the peptides encoded by the exon may have negative regulatory effects on processes associated with cell migration and invasion.

**ABI1**
Abl interactor 1 (ABI1), originally identified as an Abl binding partner and substrate for the Abelson tyrosine kinase, has been shown to play roles in actin cytoskeletal organization and cell-cell adhesion through participation in a number of signaling pathways. The 87 nt alternative exon that is silenced by ESRP expression is present within a proline rich domain, itself containing strong proline enrichment. This proline rich region was shown to mediate binding to SH3 domains, including those of Ali and Eps8. Interestingly, this proline rich region was also shown to mediate binding of ABI1 to ENAH, thereby promoting Abl mediated tyrosine phosphorylation of ENAH. Therefore it can be speculated that the presence or absence of this the proline-rich amino acids encoded by this alternative exon might influence the binding affinity and/or specificity for specific interacting proteins including other SH3 domains. Very suggestive evidence from two studies indicate, for example, that an isoform that skips this exon binds to the Nck1 adapter protein, whereas isoforms that include it do not. Further studies to investigate differential binding by these isoforms may unveil specific pathways that are correspondingly regulated differently as a result of this switch in splicing.

(Ge et al., 1999) (Karni et al., 2007)
(Goswami et al., 2009)
(Biesova et al., 1997; Yamamoto et al., 2001) (Tani et al., 2003).
between epithelial and mesenchymal cells or in response to specific extracellular stimuli. For example, given the corresponding role of ENAH in cell migration, it would be of interest to determine whether binding of ABI1 to ENAH might be influenced by the inclusion of this exon.

**PALLD**

Palladin (PALLD) is an actin binding scaffold protein that plays important roles in cell adhesion and migration. Several isoforms of Palladin that arise from different alternative promoters have been well described that significantly alter the general structure of the gene, including the presence or absence of a proline region and the number of immunoglobulin (Ig) C2 type domains. Interestingly, we noted that one of the candidate ESRP regulated alternative exons corresponds to an 85-90 kD isoform that lacks the proline rich region and contains three of the five potential Ig domains. This isoform was shown to be specifically upregulated during an EMT induced by TFG-beta 1. Furthermore, palladin was shown to contribute to invasive motility of breast cancer cell lines, consistent with a role of this isoform in mesenchymal cell properties, including invasion. (Goicoechea et al., 2008) (Goicoechea et al., 2009)

**UAP1**

An interesting well defined switchlike change in one of the splicing events we identified occurs in UDP-N-acetylglucosamine pyrophosphorylase 1 (UAP1) transcripts. UAP1 encodes two isoforms with an ESRP regulated exon that encodes a 17 amino acid insert that converts the enzyme from predominantly functioning as a UDP GalNAc to a UDP-GlcNAc pyrophosphorylase. While the implications of UDP-GalNAc pyrophosphorylase activity in epithelial cells and UDP-GlcNAc pyrophosphorylase in mesenchymal cells is not implicitly clear, this finding might suggest that further investigation of the pathways these enzymes are involved in might provide further insights into cellular processes that differ in epithelial and mesenchymal cell types. (Wang-Gillam et al., 1998)

**SOS1**

A central player in numerous signaling pathways are the Sos family of Ras guanine nucleotide exchange factors. While ESRP knockdown modestly reduced splicing of a 45 nt exon in Sos1 transcripts, it is notable that isoforms containing the 15 amino acids encoded by this exon have previously been shown to have higher affinity binding to the Grb2 adapter protein and correspondingly higher Ras guanine nucleotide exchange activity. (Rojas et al., 1999)

**KITLG**

The ligand for the c-kit protooncogene, KITLG, also known as stem cell factor (SCF) exists as both soluble and membrane bound proteins due to alternative splicing of exon 6. This alternative exon encodes a site for proteolytic cleavage and therefore its inclusion leads to the soluble protein isoform, whereas skipping results in the membrane bound form. Since the ESRPs are predicted to cause exon skipping they therefore would be expected to promote expression of the membrane bound isoform. Notable, only the membrane bound form of KITLG was shown to promote cell-cell adhesion and loss of the membrane bound form was proposed to correlate with metastatic potential in melanoma cell lines (Flanagan et al., 1991; Welker et al., 2000).

**CD46**

CD46 antigen, also known as membrane cofactor protein (MCP), is one of the class of ESRP targets containing an alternative penultimate exon that contains an in frame stop codon that leads to alternative C-termini in the cytoplasmic domain of this transmembrane protein. The isoforms that result from skipping of the exon contain a tyrosine that is phosphorylated by the src (Wang et al., 2000).
kinase, Lck, thus strongly suggesting differences in signaling pathways between these isoforms. The ESRPs are predicted to silence the exon thereby promoting these isoforms containing this specific phosphorylation event.

| RAB6A | Two mutually exclusive exons in Rab6A were shown to be regulated by the ESRPs, which were predicted to promote use of the downstream exon. This form has been described as the Rab6A isoform, whereas isoforms resulting from upstream exon inclusion have been called Rab6A`. These duplicated exons encode peptide sequences that differ by only three amino acids and yet several studies have clearly demonstrated isoform-dependent differences in function in trans-Golgi vesicular trafficking. While several studies have proposed different roles for each protein, a recent study suggested that the Rab6A` isoform was the major isoform required for retrograde Golgi to ER transport. (Del Nery et al., 2006; Echard et al., 2000) |
| FYN | Splicing of mutually exclusive exons 7A and 7B in FYN transcripts previously demonstrated enrichment of exon 7B in epithelial cells and our data consistently showed that the ESRPs promoted splicing of exon 7B. FYN Protein isoforms encoded by exon 7B were shown to be less efficient in phosphorylation of Sam68, SLM-1, SLM-2, Shc, and PLC-γ. Sam68 is itself a splicing regulator and this differential phosphorylation of Sam68 was shown to affect its activity to regulate splicing of the pro- and anti-apoptotic splice forms of Bcl-x (Bcl-x₅ and Bcl-xL). (Brignatz et al., 2009) |
| TCF7L2 | TCF7L2, also known as TCF4, encodes a member of the TCF/LEF family of transcription factors that are activated in response to activation of Wnt/β-catenin signaling pathways. The ESRPs promote inclusion of exon 4, whereas it is predominantly skipped in mesenchymal cells. Isoforms that include exon 4 were recently shown to have reduced transactivation activity of several Wnt/β-catenin target genes including Axin2, CyclinD1, Siamois, and Cdx1. These findings therefore suggest that a partial switch towards the skipped isoform during the EMT might further enhance the transcriptional activation of β-catenin target genes upon it’s translocation to the nucleus. (Weise et al., 2009) |
Supplementary Materials and methods

Plasmid Construction

Retroviral and bicistronic expression vectors for mouse Esrp1 (2A) have been described (Warzecha et al., 2009). ITGA6 and RALGPS2 minigene sequences were PCR amplified from genomic DNA isolated from PNT2 cells. Primers used for amplification were ITGA6-int24-N-F, ITGA6-int25-RV-R, RALGPS2-int14-N-F, and RALGPS2-int15-RV-R. The PCR products were cloned into the Not I and EcoR V sites of the pi-11-(H3)-PL adenovirus based splicing minigene (Hovhannisyan and Carstens, 2005). The Not I to PshA I sequence of the OSBPL3 minigene with the point mutations in the exon was synthesized by Blue Heron Technology and subcloned into the pi-11-OSBPL3 plasmid, replacing the wild-type sequence. All other point mutations were introduced via QuikChange Site-Directed Mutagenesis (Stratagene). Plasmids for in vitro transcription of the TGG(x15), ITGA6 element, and RALGPS2 element and their respective mutants were generated by annealing complimentary oligos such that the resulting double-stranded DNA contained Cla I and Xho I compatible ends for cloning into the previously described pDP-19RCAEE vector (Hovhannisyan and Carstens, 2007). Creation of the FGFR2 ISE/ISS-3 5’ sequence and mutant sequences represented in Fig. 4C were also inserted into the Cla I and Xho I sites in pDP-19RCAEE. Expression of recombinant Esrp1 protein as Glutathione-S-Transferase (GST) fusion proteins was carried out using a derivative of pGEX-2TK (GE healthcare) modified to contain a tobacco etch virus (TEV) cleavage site between the GST tag and the multicloning site (MCS) and an in-frame FLAG tag downstream of the MCS. This construct, pGEX-TEV-MCS-B-FF, contains the following sequence inserted into the BamH1 and EcoR1 sites in pGEX 2TK (inclusive of the insertion sites): 5' -

GGATCCGATTATGATATTCCAACTACTGCTAGCGAGAATTTGTATTTTCAGGGTGAGCTCAGGCCTATCGATGC GGCGCCGCCGATTACAAGGATGACGACGACGATAAGGCAGATTACAAGGATGACGACGATAAGTAAATGCA TGAATTC-3'. The underlined sequence represents a modified MCS containing unique (from 5' to 3') Sac I, Stu I, Cla I, and Not I sites. A cDNA for murine Esrp1 was PCR amplified using primers Esrp-FL-Stu-F and Esrp1-RRM123-Not-R and inserted into the Stu I and Not I sites in pGEX-TEV-MCS-B-FF for expression of recombinant Esrp1 containing an N-terminal GST and C-terminal FLAG tag. This construct was named GST-Esrp1-FL and contains sequences encoding the complete N-terminus of the protein and all three RNA Recognition Motifs (RRMs), but is truncated at the C-terminus and lacking the last 126 amino acids present in the full length protein that contains two alternative cassette exons. We determined that these C-terminal sequences contain a nuclear localization signal, but are otherwise dispensable for the complete preservation of function of the protein in regulation of splicing.

Lentiviral Production and Transduction of ESRP shRNAs

A miR-30 based non-targeting shRNA and an shRNA targeting ESRP1 were expressed from the pGIPZ vector (Open Biosystems), or in the case of the ESRP2 shRNA, from a vector in which we replaced the puromycin resistance gene in pGIPZ with the BSD blasticidin resistance gene. The targeted sequences of ESRP1 and ESRP2 are 5’-ATAAGATCTTGGAAATAAT-3’ and 5’-CGCTATATTTGAGGTGTAT-3’, respectively. The shRNA non-silencing pGIPZ control vector containing a hairpin designed to avoid targeting of mammalian genes is Catalog # RHS4346 from Open Biosystems. Lentiviruses expressing the shRNAs were produced using the psPAX2 packaging vector (Addgene). 293T Cells were transfected in 10 cm dishes with 12.6 ug of the shRNA expression vector, 8.1 ug of psPAX2, and 4.5 ug
of pCMV-VSV-G using Transit 293 (Mirus). After 16–20 hr, the media was replaced with fresh DMEM with 5% FBS, and virus was harvested after an additional 24 hr. Target cells were transduced with a 50/50 mix of viral supernatant and growth media. Selection was carried out using 0.5 ug/ml puromycin and/or 10 ug/ml blasticidin.

**Analysis of Affymetrix HJAY array results using MADS Plus (MADS+)**

Each experiment was carried out in four biological replicates and total RNA was converted to labeled cDNA for array hybridization. The HJAY array format has previously been described (Yamamoto et al., 2009). HJAY data for analysis of cassette exons were normalized and filtered for expression as described (Shen et al., 2010). For all HJAY analyses a “splicing index” is calculated for each probeset, represented as the background corrected probe intensity divided by the estimated gene expression index. Two separate one-sided t-tests are performed to test if the splicing index of a probe is substantially higher in one sample group over the other group. Then, for each probeset, the p-values of individual probes are summarized using Fisher’s method to obtain a probeset-level p-value for differential splicing (Xing et al., 2008). After removing transcripts with a greater than 1.5-fold change in overall transcript level, we used a p-value cutoff of 0.001 for inverse differences in signal between the control and experimental samples as detected by probesets corresponding to at least one of the exon inclusion events and the exon skipping event. Data were then ranked by the lowest p-value for change in probeset signal for a given cassette exon. Analysis for alternative 3’ and 5’ splicing events was performed using the same filters and a p-value of less than 0.001 for reciprocal changes in probeset signal between samples using the junction probeset corresponding to the shorter resulting isoform compared to either the longer isoform junction probeset or the exon probeset (where available) in the portion of the exon unique to the longer isoform. Analysis of predicted mutually exclusive exons was performed using the same 0.001 p-value cutoff and selecting for reciprocal differences between samples of at least one exon 1 inclusion probeset (upstream junction, downstream junction, or exon) and an exon 2 inclusion probeset. These mutually exclusive exon predictions were derived from Affymetrix annotations, but among the examples identified in our analysis a limited number of the alternative exons corresponded to absolute mutually exclusive exons and therefore a very limited number were validated and analyzed further in this study.

**Identification of ESRP-regulated alternative conserved exons (ACEs) in human and mouse databases**

The alternatively spliced exons in the mouse genome (mm7) were collected from the Alternative Splicing Annotation Project 2 (ASAP2) database (Kim et al., 2007), by requiring at least 2 ESTs for both the exon inclusion form and the exon skipping form. We used liftOver (http://genome.ucsc.edu/cgi-bin/hgLiftOver) tool to identify the orthologous location of mouse exons (mm7) in human (hg18). The LiftOver transformation file (from mm7 to hg18) can be downloaded from USCS website: http://hgdownload.cse.ucsc.edu/goldenPath/mm7/liftOver/mm7ToHg18.over.chain.gz.

**Conserved motif enrichment analysis for candidate ESRP binding sites**

In order to derive a candidate ESRP binding site, we used the set of ESRP regulated cassette exons identified in this study that displayed a greater than 10% change in the level of exon inclusion in response to either ESRP knockdown in PNT2 cells or ectopic expression of ESRP1 in the MDA-MB-
231 cells. Using a control reference set of 4779 alternative cassette exons we searched for all hexamers within a window of 250 nucleotides upstream and downstream of the exons, exclusive of the last 20 nt of upstream introns and the first 6 nt of the downstream introns. These 4799 non-redundant cassette exons were collected after applying stringent filtering criteria to exons in the Alternative Splicing Annotation Project 2 (ASAP2) database. For this set of cassette exons, we selected those with at least 3 inclusion ESTs and at least 3 skipping ESTs and removed exons with length >250nt or <10nt.

We also performed the same analysis for enrichment of hexamers within the ESRP-regulated exons compared to the exons in the same reference set. We then independently analyzed enhanced exons and silenced exons for enriched hexamer motifs in the upstream intron, downstream intron, and the exon body. Enriched hexamers were then sorted by p-value from a one-sided Fisher test, by nucleotide. We also used a branch length score (BLS) to assess motif conservation using the multiple alignment and phylogenetic tree of 44 vertebrate species across ~500 million years (Stark et al., 2007). This analysis allowed movement of the sites in the assignment of the orthologous sites (e.g. TGCATG aligned with TGCAT-G is tolerated). Using this analysis, we thus determined the BLS for all 4096 possible motifs in the set of ESRP-regulated and in the reference set of alternative exons. The mean BLS for each motif was then analyzed for conservation in the ESRP-regulated set of exons compared to the reference set.

**Oligonucleotide sequences used in this study**

<table>
<thead>
<tr>
<th>RT-PCRs</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLK-F</td>
<td>5'-TTGAGCAGGAAATGATGAGGAAAA-3'</td>
</tr>
<tr>
<td>SLK-R</td>
<td>5'-CTGCTTTCTGCTGGATGA-3'</td>
</tr>
<tr>
<td>ZFAND1-F</td>
<td>5'-AGAGATGGCGGAGTTGGACA-3'</td>
</tr>
<tr>
<td>ZFAND1-R</td>
<td>5'-CATTAGGGGATTTGGATTC-3'</td>
</tr>
<tr>
<td>SCRIB-F</td>
<td>5'-GACAAGGGAGGGGCGTGGTTTCT-3'</td>
</tr>
<tr>
<td>SCRIB-R</td>
<td>5'-TATGCGCTCGTCTCCCTAT-3'</td>
</tr>
<tr>
<td>FLNB-F</td>
<td>5'-GCCGAAGAAGAGGCCCTTGG-3'</td>
</tr>
<tr>
<td>FLNB-R</td>
<td>5'-GCCGTTCTAGTCTCACTCCG-3'</td>
</tr>
<tr>
<td>FNIP1-F</td>
<td>5'-AACACAGTTATTAAATGGACTGCTTTGG-3'</td>
</tr>
<tr>
<td>FNIP1-R</td>
<td>5'-GTGCTATGCGCCTCTGTC-3'</td>
</tr>
<tr>
<td>MAGI1-F</td>
<td>5'-CCATAACACGAAAAACACATCTT-3'</td>
</tr>
<tr>
<td>MAGI1-R</td>
<td>5'-GTCTTCAATCTTTTCCAACCA-3'</td>
</tr>
<tr>
<td>SLC37A2-F</td>
<td>5'-CTGGAAGGTGTCCTCCCTAG-3'</td>
</tr>
<tr>
<td>SLC37A2-R</td>
<td>5'-TGAACAAAGCAAGTCTGAGCA-3'</td>
</tr>
<tr>
<td>OSBPL3-exon8-F</td>
<td>5'-CTCCTGCAAGACATGGAGCT-3'</td>
</tr>
<tr>
<td>OSBPL3-exon10-R</td>
<td>5'-GATGAGGTTTCCAGGACATCAGA-3'</td>
</tr>
<tr>
<td>STX3-F</td>
<td>5'-TGGAGAAGGACAGGAGTAAA-3'</td>
</tr>
<tr>
<td>STX3-R</td>
<td>5'-AAATACAGCAGCCTTGGGAGT-3'</td>
</tr>
<tr>
<td>STXB5-F</td>
<td>5'-CCTATGAGACTGTTGGTTT-3'</td>
</tr>
<tr>
<td>STXB5-R</td>
<td>5'-AATCGGCTTTAGTGAGTACAGG-3'</td>
</tr>
<tr>
<td>MARK2-F</td>
<td>5'-GAACCGTCCGGCTGGCTAG-3'</td>
</tr>
<tr>
<td>MARK2-R</td>
<td>5'-GGGCCCCCGAATCTTGGGACTT-3'</td>
</tr>
<tr>
<td>RALGPS2-F</td>
<td>5'-AGACCTCATGCGCCTTTGGAACAA-3'</td>
</tr>
<tr>
<td>RALGPS2-R</td>
<td>5'-TGTAGGCTTTTGGCCTTTTAAA-3'</td>
</tr>
<tr>
<td>ARHGEF11-F</td>
<td>5'-GGCAGCAGAGGGTTCAAAAGTT-3'</td>
</tr>
</tbody>
</table>
ARHGEF11-R  5′-TGAGTGGTCGGTGCTTGAGTC-3′
COL16A1-F   5′-AGCAAGGACAGACGGGACTCA-3′
COL16A1-R   5′-TCCGGGAAATACCCTGTTGGGAC-3′
EXO7C-F1    5′-CTTACTCCCTGCTATCCCCAACA-3′
EXO7C-R1    5′-CCCTCTCAAGCATCAGCCCATCCA-3′
ABI1-F      5′-CGGGGAACACTGGGACGGAATAC-3′
ABI1-R      5′-GAGGTGGTGCGGGCAGGAGTTGG-3′
BIN1-F      5′-CTCCAGATGGCTCCCCTGC-3′
BIN1-R      5′-CCTGGGGGCAGGTCCAAGCG-3′
VEGFA-F     5′-CGGGGGCTGCTGCAATGACGA-3′
VEGFA-R     5′-GCCCTCCGGACCCAAAGTGCTCTG-3′
FYN-exon8-F 5′-TGCAAAACCTTGACATGGTGATA-3′
FYN-exon10-R 5′-CTTGTCTGTGCTTCTCCTCAT-3′
MAPK14-exon8-F 5′-GAGTCGATAATGGCCGAGCTG-3′
MAPK14-exon10-R 5′-ATAAATACATTTGCGAAGCT-3′
pl-11-F     5′-ACTAGTAACGCCGCCAGTGTGCT-3′
pl-11-R     5′-GCATAGAGACCGCGAAGAGTTTGTCC-3′
PUF60-F     5′-GGGGAGACCGGCTGGAGTTGG-3′
PUF60-R     5′-GCAAGGGCGTTCGCCAG-3′
RFX2-F      5′-CCAGCGGTCCCCTCCCACAG-3′
RFX2-R      5′-CGCCCGCGTTTCACGCCAC-3′
ZNF236-F    5′-GCCACCGGTTCGCCAC-3′
ZNF236-R    5′-CCCACCGGCCAGGAC-3′

Minigenes
ITGA6-int24-N-F  5′-GCGGCCGCAGGAGATTTTATTTTATGGACCAG-3′
ITGA6-int25-RV-R 5′-GATATCGCTGCTTCTTCATCATGAGCTGTA-3′
RALGPS2-int14-N-F 5′-GCGGCCGCTAGACAGACATCAAATACATT-3′
RALGPS2-int15-RV-R 5′-GATATCGTTTGTTACTTTATTCAGAGCTGA-3′
MAGI1-int6-N-F2  5′-GCAGGCCCCTCTTCCTAAATGTGTTGCCCTCC-3′
MAGI1-int7-RV-R2 5′-GATATCTTCTAGCCAACCATGACTGATTAT-3′
OSBPL3-int8-N-F2 5′-GCCGCGGCGGTTCGCCAC-3′
OSBPL3-int9-RV-R2 5′-GATATCGTCTCCGTACCCCTGTTCTCTG-3′
ITGA6-mut1-qckchng 5′-ATTGCTAGCTGTGAGCCGCACATGACTAACTTT-3′
ITGA6-mut2-qckchng 5′-ATTGCTAGCTGTGAGCCGCACATGACTAUCTT-3′
RALGPS2-qckchng 5′-GAACTCCCTTATATGTGCTGTTACATGACTGTAAG-3′
MAGI1-mut1 5′-GGAGATTTCTTCTTTGATGCTGATTCT-3′
OSBPL3-int-mut 5′-GGGGATGGAGACCGGCTCCCTTACTGATTGACTAAGTAC-3′

Annealed inserts for in vitro transcription
ITGA6-ISE-WT-Top  5′-CGATCTAGCTGTGAGCCGCACATGACTAACTT TAAC-3′
ITGA6-ISE-WT-Bottom 5′-TCGAGTTAGGCCACACCATAGCCCGGTCCCACAG-3′

16
ITGA6-ISE-Mut1-Top  5’-CGATCTAGCTGTTAGGACCCGCTATGGTGTGATGACTAAC-3’
ITGA6-ISE-Mut1-Bottom  5’-TCGAGTTAGCTACAAACCACCATGCGGTCTGCTACAGCTAGAT-3’
ITGA6-ISE-Mut2-Top  5’-CGATCTAGCTGTGACGCGCTGTGGTGGATGCTAAC-3’
ITGA6-ISE-Mut2-Bottom  5’-TCGAGTTAGCCACAAACCAAGCTGCGCTCCACACAGCTAGAT-3’
RALGPS2-ISE-WT-Top  5’-CGATCAGGGGTCACAGAACACTCCCTATGTTGGTTGATGTTGGTC-3’
RALGPS2-ISE-WT-Bottom  5’-TCGAGACCAACCACCATGTAAGCCACCACCAACCAACCAATAACGATCTGACCCCTTATGATTGGTC-3’
RALGPS2-ISE-Mut-Top  5’-CGATCAGGGGTCACAGAACACTCCCTATGTTGGTTGATGTTGGTC-3’
RALGPS2-ISE-Mut-Bottom  5’-TCGAGACCAACCACCATGTAAGCCACCACCAACCAACCAATAACGATCTGACCCCTTATGATTGGTC-3’

**Primers for amplification of Esrp1 cDNA for recombinant protein expression**

Esrp-FL-Stu-F:  5’-AGGCCCTGCCACCAGATGACGGGCGTCTCCGGGATTAG-3’
Esrp1-RRM123-Not-R:  5’-GGCCGCGCCGCGGAGGACAGGACAGGCGATGG-3’
Supplementary References


