The EMBO Journal (2010) 29, 3286–3300 | © 2010 European Molecular Biology Organization | All Rights Reserved 0261-4189/10
www.embojournal.org

THE EMBO JOURNAL

An ESRP-regulated splicing programme is abrogated during the epithelial–mesenchymal transition

Claude C Warzecha[1,2], Peng Jiang[3], Karine Amirikian[1], Kimberly A Dittmar[1], Hezhe Lu[4], Shihao Shen[5], Wei Guo[4], Yi Xing[3,5,6] and Russ P Carstens[1,2,7,*]

[1]Renal Division, Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA, [2]Cell and Molecular Biology Graduate Group, University of Pennsylvania School of Medicine, Philadelphia, PA, USA, [3]Department of Internal Medicine, University of Iowa, Iowa City, IA, USA, [4]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA, [5]Department of Biostatistics, University of Iowa, Iowa City, IA, USA, [6]Department of Biomedical Engineering, University of Iowa, Iowa City, IA, USA and [7]Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA.

Alternative splicing achieves coordinated changes in post-transcriptional gene expression programmes through the activities of diverse RNA-binding proteins. Epithelial splicing regulatory proteins 1 and 2 (ESRP1 and ESRP2) are cell-type-specific regulators of transcripts that switch splicing during the epithelial–mesenchymal transition (EMT). To define a comprehensive programme of alternative splicing that is regulated during the EMT, we identified an extensive ESRP-regulated splicing network of hundreds of alternative splicing events within numerous genes with functions in cell–cell adhesion, polarity, and migration. Loss of this global ESRP-regulated epithelial splicing programme induces the phenotypic changes in cell morphology that are observed during the EMT. Components of this splicing signature provide novel molecular markers that can be used to characterize the EMT. Bioinformatics and experimental approaches revealed a high-affinity ESRP-binding motif and a predictive RNA map that governs their activity. This work establishes the ESRPs as coordinators of a complex alternative splicing network that adds an important post-transcriptional layer to the changes in gene expression that underlie epithelial–mesenchymal transitions during development and disease.

The EMBO Journal (2010) **29,** 3286–3300. doi:10.1038/emboj.2010.195; Published online 13 August 2010
Subject Categories: cell & tissue architecture; RNA
Keywords: alternative splicing; epithelial–mesenchymal transition; epithelial splicing regulatory proteins

*Corresponding author. Department of Medicine, University of Pennsylvania School of Medicine, 411 Hill Pavilion, 380 S. University Avenue, Philadelphia, PA 19104, USA. Tel.: +1 215 573 1838; Fax: +1 215 898 0189; E-mail: russcars@upenn.edu

Received: 23 May 2010; accepted: 20 July 2010; published online: 13 August 2010

Introduction

Alternative splicing vastly expands transcriptomic diversity as evidenced by recent studies demonstrating that nearly all multi-exon human genes undergo alternative splicing (Pan *et al*, 2008; Wang *et al*, 2008). Regulated alternative splicing is achieved through the interplay of multiple RNA-binding proteins that associate with auxiliary *cis*-elements within pre-mRNA transcripts that can positively or negatively influence splicing of regulated exons and are designated as intronic or exonic splicing enhancers and silencers. The combinatorial assembly of RNA-binding proteins with these *cis*-elements determines whether an exon is included or skipped in the mature transcript. Combinatorial control of splicing involves contributions by the generally ubiquitous SR and hnRNP families of proteins in cooperation with tissue-specific regulators of splicing, such as Nova, nPTB, Fox, nSR100, MBNL, and CELF proteins (Chen and Manley, 2009; Hartmann and Valcarcel, 2009; Nilsen and Graveley, 2010). With the advent of novel cross-linking and high-throughput sequencing approaches, binding sites for several splicing regulators have been determined on a genome-wide scale and demonstrated position-dependent effects of binding relative to regulated exons that determine the splicing outcome (Licatalosi *et al*, 2008; Xue *et al*, 2009; Yeo *et al*, 2009). These RNA maps are thus predictive of genome-wide effects of a given splicing factor and the combined contributions of numerous regulators are likely to define a splicing code that governs diverse splicing outcomes in different cellular milieus (Barash *et al*, 2010).

The well-characterized neuron-specific Nova proteins and the neural/muscle-enriched Fox proteins co-regulate numerous splicing events as part of functionally coherent splicing regulatory networks (SRNs) (Licatalosi *et al*, 2008; Zhang *et al*, 2008; Venables *et al*, 2009). Just as sets of genes that are co-regulated at the transcriptional level have been shown to function in common processes, complexes, and pathways, increasing evidence suggests that genes co-regulated post-transcriptionally as part of SRNs also have functionally related properties. However, genes that are co-regulated transcriptionally in different cells and tissues do not overlap significantly with genes that are regulated by alternative splicing in the same cells and tissues (Pan *et al*, 2004). These observations indicate that identification of genes co-regulated within SRNs will uncover sets of functionally related gene products in addition to those uncovered by analysis of total mRNA levels alone.

The epithelial-to-mesenchymal transition (EMT) is a critical process during the development and differentiation of multiple organs and is implicated in tissue fibrosis and tumour metastasis (Thiery *et al*, 2009). The EMT is a complex process involving reorganization of the actin cytoskeleton, loss of cell–cell adhesion, and the acquisition of increased cell motility. The changes in the transcriptional profile of

epithelial cells as they undergo an EMT have been well studied, including the downregulation of E-cadherin and upregulation of vimentin and N-cadherin (LaGamba *et al*, 2005). However, the extent to which coordinated changes in splicing might contribute an additional layer of gene regulation that impacts the EMT at the post-transcriptional level has not been systematically investigated. We recently identified epithelial splicing regulatory proteins 1 and 2 (ESRP1 and ESRP2) as epithelial cell-type-specific regulators of FGFR2 splicing as well as several other transcripts with epithelial-specific splice variants (Warzecha *et al*, 2009a, b). We determined that the ESRPs are a component of an epithelial gene signature and demonstrated a downregulation of the ESRPs in cells that undergo an EMT, which corresponded with switches in splicing of several ESRP-regulated exons. We hypothesized that a global epithelial SRN regulated by the ESRPs constitutes an essential post-transcriptional gene expression programme with broad developmental and patho-physiological functions. We further predicted that the different protein isoforms encoded by these co-regulated transcript variants would have distinct functions in epithelial and mesenchymal cells that contribute to the unique features and biological functions of each cell type.

In this study, we used a next generation splicing sensitive microarray platform to characterize an ESRP-regulated SRN and identified hundreds of novel ESRP-regulated splicing events. Computational and experimental approaches defined an ESRP-binding motif that defined position-dependent functions on exon splicing, thereby deriving an RNA map for ESRP-mediated splicing regulation. Importantly, we demonstrated that a global reversion of splicing from the epithelial-to-mesenchymal patterns through sustained knockdown of ESRP1 and ESRP2 in mammary epithelial cells induced cellular changes indicative of an EMT. These results establish that the ESRPs orchestrate an epithelial cell-type-specific SRN that regulates mRNAs encoding proteins with isoform-specific functions in epithelial versus mesenchymal cells. This study also highlights the importance of coordinated changes in alternative splicing, in addition to changes in transcription, to promote dynamic changes in proteomic complexity during the EMT.

Results

Identification of an extensive ESRP-regulated epithelial SRN using splicing sensitive exon junction microarrays

Our previous observations demonstrated several splicing switches in ESRP-regulated gene transcripts during the EMT, suggesting that these splicing factors direct a large-scale splicing network of genes with functions in this process. Although no systematic analysis of global changes in alternative splicing during the EMT has previously been performed, we hypothesized that numerous ESRP-regulated transcripts undergo splicing switches during the EMT. To comprehensively determine a global ESRP-regulated alternative splicing network, we used affymetrix human exon junction arrays (HJAY) to profile splicing changes in response to ectopic expression of ESRP1 in MDA-MB-231 mesenchymal cells and knockdown of ESRP1 and ESRP2 in PNT2 epithelial cells (Figure 1A–C). To identify ESRP-regulated changes in splicing, we used the microarray analysis of differential splicing plus (MADS+) computational pipeline (Shen *et al*, 2010). The HJAY arrays consist of eight tiled probes per probeset that target 315 137 exons and 260 488 exon–exon junctions, covering most human alternative splicing events supported by mRNA and/or EST evidence. To detect splicing changes between samples, we selected events that displayed reciprocal changes in the probeset targeting the exon skipping events and at least one of three exon inclusion probesets. This approach predicted 417 and 540 changes in cassette exon splicing upon ectopic expression of ESRP1 and combined ESRP knockdown, respectively (Figure 1D; Supplementary Table S1). We noted numerous gene transcripts that contained more than one predicted ESRP-regulated exon and examples in which the ESRPs were predicted to promote either exon inclusion or exon skipping. There was a significant positive correlation of the two data sets consisting of 104 cassette exons showing opposite changes in splicing (P-value $< 4.0e{-}23$) (Figure 1D; Supplementary Table S1). A related analysis detected reciprocal changes in probesets corresponding to alternative 3′ and 5′ splice sites as well as mutually exclusive exons (Supplementary Table S1).

The ESRPs are phylogenetically conserved splicing factors and we therefore expected that many of the regulated exons and genes would be highly conserved between humans and mice, as well as among broader metazoans. Consistent with this prediction, we found a significant number of alternative conserved exons (ACEs) among the HJAY-detected cassette exons as predicted by the computational ACEScan tool (Yeo *et al*, 2005). These ACEScan exons are highly enriched for biologically relevant alternative splicing events that are conserved between humans and mice. Among the 828 non-redundant exons from both ESRP-regulated cassette exon data sets, 200 (24.2%) were ACEScan exons, whereas only 9.0% of the background set of cassette exons on the array that passed an expression filter in either experiment were ACEScan exons (P-value $< 2.2e{-}16$) (Supplementary Table S2). ESRP-regulated exons were also enriched in a data set of exons with strong evidence of more extensive phylogenetic conservation of regulated splicing. Analysis of RNA selection pressure (RSPR) across 28 vertebrate genomes recently identified 345 human exons with strong evidence of conserved alternative splicing (Lu *et al*, 2009). Of the predicted ESRP-target exons, 49 (5.9%) of these were among the 345 highly conserved alternative exons with high RSPR compared with 220 (1.9%) of the background set (P-value $< 5.1e{-}11$) (Supplementary Table S2). Finally, ESRP-regulated exons from both sets were more likely to be a multiple of three in nucleotide (nt) length thus maintaining the reading frame, further evidence of enrichment for functionally relevant splicing events (PNT2: 58 versus 47% in HJAY background; MDA-MB-231: 59 versus 46% in background) (Resch *et al*, 2004).

Validation of ESRP-regulated exons reveals numerous large switches in exon splicing

To independently validate a large set of splicing changes predicted by the HJAY analysis, we predominantly used high-throughput capillary reverse transcription PCR (HT-RT–PCR) to validate nearly all simple cassette exons with constitutive flanking exons identified by the HJAYarrays (Venables *et al*, 2009). In the ESRP1 ectopic expression system, 158 of the 173 exons with successful HT-RT–PCR reactions demonstrated the HJAY predicted changes in exon

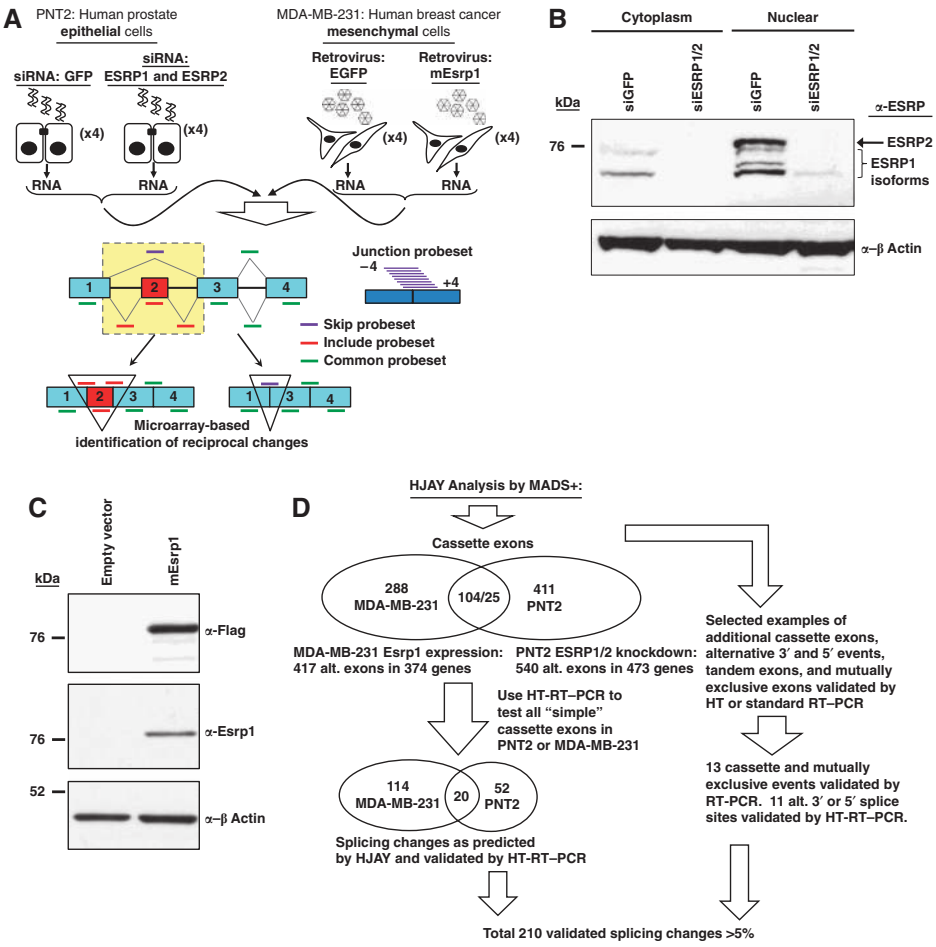


Figure 1 An exon junction microarray detects hundreds of novel ESRP-regulated alternative splicing events. (**A**) Outline of the experimental systems and array platform used to identify ESRP-regulated exons. (**B**) Western blot using a monoclonal antibody that recognizes both ESRPs to validate siRNA-mediated depletion of ESRP1 and ESRP2 in the nucleus and cytoplasm. Bracket indicates bands corresponding to isoforms encoded by ESRP1 splice variants. (**C**) Western blot demonstrating ectopic expression of Flag-tagged mouse Esrp1 protein in MDA-MB-231 cells using anti-FLAG and anti-ESRP1 antibodies. (**D**) Flowchart summarizing the results of the microarray analyses and validation assays.

inclusion of which 134 showed a higher confidence change of >5% in percent exon inclusion (Supplementary Table S3). In PNT2 knockdowns, 115 out of 183 successful HT-RT–PCR reactions yielded a lower overall validation rate, of which 71 exons had a >5% change in exon inclusion (Supplementary Table S4). We also used the HT-RT–PCR format to test 20 alternative 3′ or 5′ splice sites that were identified in the knockdown experiments and 11 validated the predicted change with a high confidence of >5% change between competing 5′ or 3′ splice sites (Supplementary Table S4). We also validated an additional 13 HJAY predicted splicing changes using traditional RT–PCR, including cases that involved predicted mutually exclusive exons (Supplementary Figure S1). In total, we validated 210 predicted changes in splicing with a >5% change in splice isoforms.

Exons showing the overall largest validated changes in exon inclusion identified in each experimental approach are shown in Table I. Among the validated events, were 22 'switch-like' changes in exon splicing in which knockdown or ectopic expression induced a switch from predominant exon inclusion (>66%) to mostly exon skipping (<34% inclusion), or *vice versa* as defined in Xing and Lee (2005). Also noteworthy were 19 examples in which one splice variant was predominant at baseline in each cell type, but where knockdown or ectopic ESRP expression induced expression of a second splice variant. Even though such changes were less complete, they nonetheless predicted expression of numerous epithelial- and mesenchymal-specific variants (e.g. *ARFGAP2*, *ARHGEF11*). *RALGPS2* and *ITGA6* demonstrated large changes in ESRP-mediated exon inclusion in both experimental systems (Figure 2A and B). In *MAGI1*, ESRP expression promoted exon skipping in both the ESRP overexpression and knockdown analysis (Figure 2C). In addition, we noted that two different exons in *MAP3K7* (also known as *TAK1*) transcripts were either enhanced or silenced by the ESRPs (Figure 2D). Of note, regulated exons in *ITGA6* and *MAP3K7* are alternative penultimate exons that generate different C-terminal domains in the resulting protein isoforms. Similar to the *ITGA6* exon, we identified splicing switches in a number of other penultimate exons containing stop codons (e.g. in *MPRIP*, *CD46*, *STX2*, and *STX3* transcripts). As most of these stop codons are <50–55 nt upstream of the terminal intron, these transcripts are predicted to elude RNA degradation through the nonsense-mediated decay pathway and this type of exon switch provides an effective means of changing the C-terminal domain. Many of

Table I Exons with the largest validated positive and negative change in splicing upon ESRP knockdown or ectopic expression

MDA-MB-231: ectopic expression of Esrp1

Gene symbol	Gene name	Exon coordinates	% Inclusion EV control	% Inclusion Esrp1	% Change
ESRP-enhanced exons					
VPS39	Vacuolar protein sorting 39	chr15:40271556–40271588	1.2	100.0	98.8
FNIP1	Folliculin-interacting protein 1	chr5:131074170–131074253	3.1	88.4	85.3
UAP1	UDP-*N*-acetylhexosamine pyrophosphorylase	chr1:160829149–160829196	4.4	86.5	82.2
KIF13A	Kinesin family member 13A	chr6:17879324–17879428	2.1	79.1	77.1
RALGPS2[a]	Ral GEF with PH domain and SH3-binding motif 2	chr1:177127988–177128065	6.4	80.9	74.5
ITGA6	Integrin α-6	chr2:173074746–173074875	30.0	96.4	66.4
KIAA0494	KIAA0494	chr1:46926612–46926803	25.8	91.5	65.7
ARHGEF12	Rho guanine nucleotide exchange factor 12	chr11:119785313–119785369	18.2	82.1	63.9
C10orf137	Chromosome 10, open reading frame 137	chr10:127407562–127407663	37.3	100.0	62.7
BTN2A1	Butyrophilin subfamily 2 member A1	chr6:26571004–26571104	17.9	76.6	58.7
ESRP-silenced exons					
MAGI1[a]	Membrane-associated guanylate kinase, WW and PDZ 1	chr3:65408737–65408772	92.5	1.9	−90.6
PLAA	Phospholipase A-2-activating protein	chr9:26907095–26907163	97.0	7.9	−89.1
LAS1L	LAS1-like protein	chrX:64670215–64670340	100.0	16.1	−83.9
LRRFIP2	Leucine-rich repeat flightless-interacting protein 2	chr3:37107962–37108033	88.2	16.1	−72.1
ELOVL1	Elongation of very long chain fatty acids protein 1	chr1:43603444–43603634	80.9	9.4	−71.4
SPINT2	Serine peptidase inhibitor, Kunitz type, 2	chr19:43466107–43466277	87.6	18.9	−68.7
ANKRD26	Ankyrin repeat domain-containing protein 26	chr10:27390079–27390177	100.0	36.1	−63.9
SOD2	Superoxide dismutase 2	chr6:160029148–160029264	85.3	25.9	−59.4
PPFIBP1	PTPRF-interacting protein-binding protein 1	chr12:27721264–27721296	62.7	3.4	−59.3
HMBS	Hydroxymethylbilane synthase	chr11:118468324–118468443	93.3	34.6	−58.7

PNT2 cells: combined knockdown of ESRP1 and ESRP2[b]

Gene symbol	Gene name	Exon coordinates	% Inclusion siGFP	% Inclusion siESRP1/2	% Change
ESRP-enhanced exons					
SLC37A2	Solute carrier family 37 member 2	chr11:124461310–124461366	80.2	3.7	−76.5
RALGPS2[a]	Ral GEF with PH domain and SH3-binding motif 2	chr1:177127988–177128065	76.5	10.7	−65.8
ITGA6	Integrin α-6	chr2:173074746–173074875	87.8	27.5	−60.3
GOLGA2	Golgin subfamily A member 2	chr9:130069294–130069374	60.8	21.5	−39.3
MYO9A	Myosin-IXa	chr15:69972360–69972572	89.6	50.9	−38.7
TCF7L2	Transcription factor 7-like 2	chr10:114714305–114714373	44.7	15.5	−29.2
MAP3K7	Mitogen-activated protein kinase kinase kinase 7 (TAK1)	chr6:91310992–91311072	31.3	5.2	−26.1
MPZL1	Myelin protein zero-like protein 1	chr1:166011925–166012027	91.9	66.2	−25.7
HISPPD1	Histidine acid phosphatase domain-containing protein 1	chr5:102546834–102547007	30.0	6.8	−23.2
INTS9	Integrator complex subunit 9	chr8:28760183–28760245	95.7	73.5	−22.2
ESRP-silenced exons					
COL16A1	Collagen α-1(XVI)	chr1:31917992–31918039	24.9	70.8	46.0
PLAA	Phospholipase A-2-activating protein	chr9:26907095–26907163	50.7	95.6	44.9
MAP3K7	Mitogen-activated protein kinase kinase kinase 7 (TAK1)	chr6:91284887–91285002	27.5	71.6	44.1
PQLC3	PQ-loop repeat-containing protein 3	chr2:11232545–11232586	49.2	91.0	41.8
JAG2	Jagged-2	chr14:104688373–104688486	33.0	72.4	39.4
MAGI1[a]	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 1	chr3:65408737–65408772	7.0	44.9	37.9
PRC1	Protein regulator of cytokinesis 1	chr15:89313313–89313354	27.3	63.3	36.1
MEST	Mesoderm-specific transcript homolog	chr7:129927866–129927967	46.6	81.6	35.0
UBE2K	Ubiquitin-conjugating enzyme E2 K	chr4:39455697–39455825	57.6	92.4	34.8
TBC1D23	TBC1 domain family member 23	chr3:101513367–101513411	31.5	64.5	33.0

The 10 cassette exons with the largest validated increase or decrease in percent change in exon inclusion in each experimental approach as determined by HT-RT–PCR. Values are shown in response to expression of a murine cDNA for Esrp1 in mesenchymal MDA-MB-231 cells or upon combined knockdown of ESRP1 and ESRP2 in epithelial PNT2 cells.
[a]Note that RALGPS2 and MAGI1 were among the top 10 enhanced and silenced exons that were validated in both experimental approaches.
[b]In PNT2 cells most values represent mean from two RT–PCRs.

these splicing switches occur in transmembrane proteins, indicating that this might be a common mechanism of altering the intracellular signalling pathways of integral membrane receptors and matrix binding proteins. To further verify that changes in splicing were the result of ESRP knockdown, we also validated a subset of splicing switches using a second combination of siRNAs directed against different sequences in ESRP1 and ESRP2. We also performed a 'rescue' using the RNAi resistant mouse Esrp1 cDNA. This analysis confirmed that these splicing switches are specifically due to ESRP1 and ESRP2 knockdown and not because of off-target effects (Figure 2E and F).

Identification of an ESRP-binding motif within ESRP-regulated exons and intronic flanking sequences

We previously demonstrated that ESRP1 binds the *FGFR2* auxiliary *cis*-element ISE/ISS-3, located in the intron between mutually exclusive exons IIIb and IIIc, which implied that the ESRPs promoted splicing of the upstream exon IIIb while silencing the downstream exon IIIc (Warzecha *et al*, 2009a). To define a consensus ESRP-binding motif, we searched for conserved hexamers in the introns flanking validated ESRP-regulated cassette exons as well as in the exons themselves. The top five motifs in each position relative to enhanced or silenced exons are presented in Figure 3A (complete motif analysis is shown in Supplementary Table S5). We noted a remarkably enriched motif downstream of ESRP-enhanced exons consisting of hexamers that contained repeats of UGG or GGU; similar to the *FGFR2*-binding site and consistent with a model in which ESRP-binding downstream of an exon promotes its inclusion. Similar motifs were enriched within ESRP-silenced exons. As UGG was a core of the most enriched motifs and to account for some degeneracy in the binding site, we mapped the 5-mer motifs UGGUG, GGUGG, and GUGGU across the set of validated ESRP-regulated exons within the windows used to derive the motifs (Supplementary Figure S2A and B). We noted that clusters of these sequences were enriched downstream of enhanced exons. In contrast, we observed enrichment of these motifs within ESRP-silenced exons, whereas these motifs were largely absent in ESRP-enhanced exons. We also noted that these 5-mers were present in greater numbers upstream of silenced exon than the enhanced exons. Therefore, the trend of these data was consistent with a 'map' in which ESRP-binding downstream of a regulated exon promotes exon inclusion and binding within the exon, and possibly the upstream intron, promotes exon skipping. We also note that although there are a number of ESRP-enhanced exons in which the motif was not identified within the 250-nt window used here, we often noted UGG-rich motifs in conserved sequence tracks located further downstream of these exons (e.g. ENAH). To determine whether the UGG-rich motifs are direct binding targets for the ESRPs, we performed RNA electrophoretic mobility shift assay (EMSAs) using recombinant ESRP1 protein. We tested whether the protein bound to RNA sequences containing repeats of UGG compared with UCG repeats. These experiments validated sequence-specific binding to the UGG repeat motifs, but not to those in which the motif is disrupted (Figure 3B). EMSA titration analysis to determine binding affinities revealed a dissociation constant (Kd) of 31.7 nM, indicating that this motif represents a bona fide high-affinity target for the ESRPs (Figure 3C).

We noted that the Fox-binding site UGCAUG was one of the top conserved hexamer motifs downstream of both ESRP-enhanced and ESRP-silenced exons. As the presence of this motif downstream of an exon predicts Fox-mediated enhanced exon inclusion, this observation suggests that the Fox proteins and the ESRPs can function both cooperatively and antagonistically in the regulation of exon splicing, although further studies are needed to investigate functional interactions between these regulators (Zhang *et al*, 2008). However, we did note a number of ESRP-regulated exons that were previously identified as Fox-regulated exons and these included examples in which these splicing regulators promoted the same as well as opposite effects on exon splicing (Supplementary Table S6).

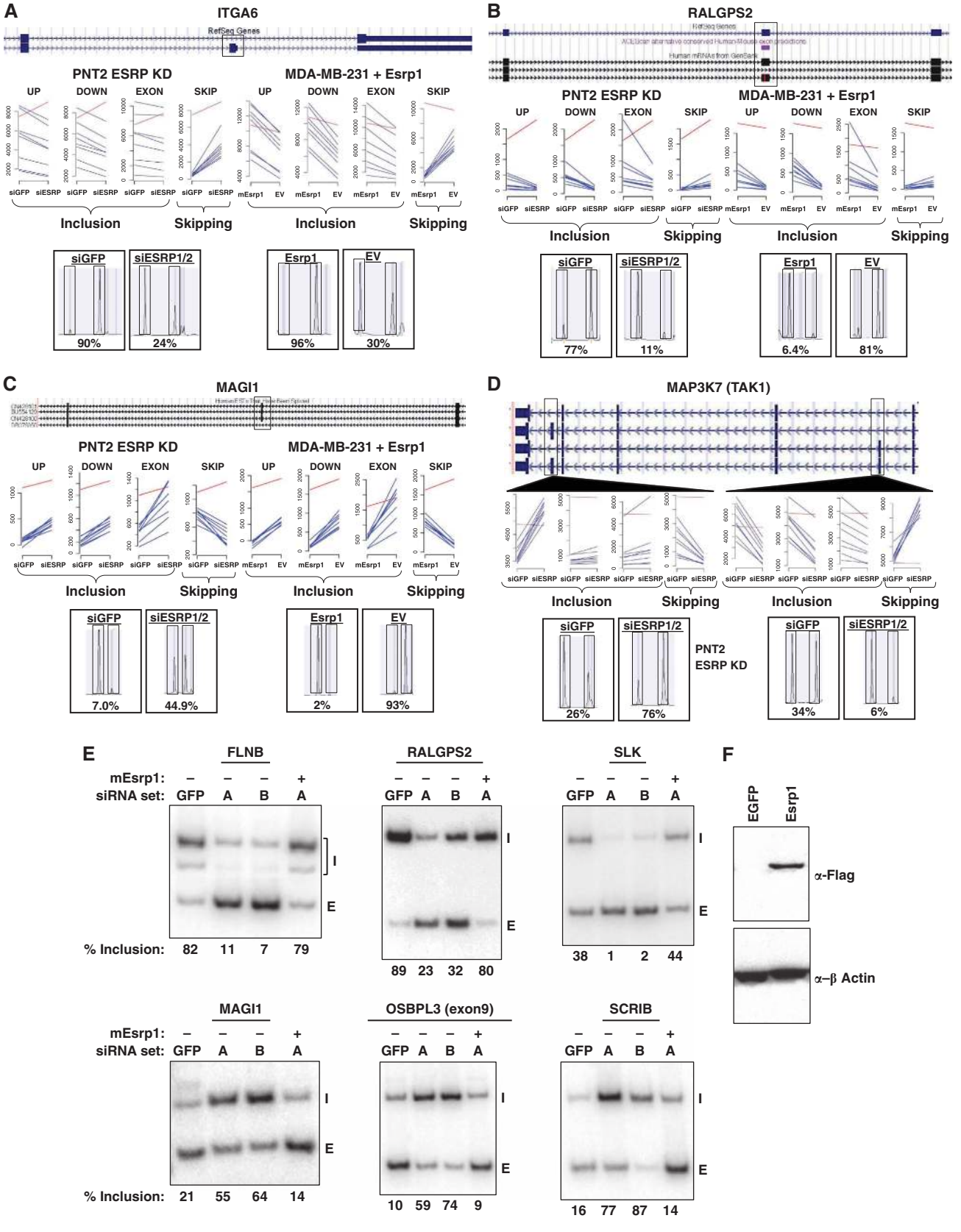
The position of ESRP-binding motifs define an epithelial splicing regulatory 'RNA map'

To validate that the ESRPs can directly regulate splicing by binding to UGG-rich motifs, we constructed minigenes corresponding to two alternative exons in the *ITGA6* and *RALGPS2* transcripts that are robustly enhanced by ESRP expression. These exons and conserved flanking intronic sequences were inserted between two adenoviral exons. Both gene transcripts contain a highly conserved sequence element downstream of the exon that contains several putative ESRP-binding sites (Figure 4A and B). We co-transfected these minigenes in 293T cells, which do not express ESRP, with a plasmid directing expression of ESRP1 or empty vector control. Exon inclusion was significantly increased in both minigenes when ESRP1 was co-expressed (Figure 4C). We introduced point mutations in both minigenes to disrupt conserved UGG motifs (Figure 4A and B). In *ITGA6* mutations in the UGG motifs abolished the ability of ESRP1 to promote exon inclusion, whereas mutations outside of these motifs did not. In the case of *RALGPS2* splicing of the exon containing the mutated sequence motifs was reduced even in the control empty vector co-transfections. This could be due either to reduced binding of other GU-rich binding proteins that may also promote exon inclusion, or through the creation of a silencing element. Nonetheless, these mutations similarly abrogated the ability of ESRP1 to promote exon inclusion (Figure 4C). We validated that these *ITGA6* and *RALGPS2* sequence elements bind ESRP1 directly by EMSA using RNA probes from the sequences presented in Figure 4A and B. Mutations in the *ITGA6* element targeting UGG motifs nearly abolished ESRP1 binding, whereas the mutations in the *RALGPS2* element disrupted binding, albeit to a lesser extent (Figure 4D).

We next investigated whether the ESRPs promote exon skipping through binding to UGG-containing motifs in the

Figure 2 Robust examples of validated ESRP-induced enhancement and silencing of microarray predicted target exons. (**A**) Validation of an ESRP-enhanced penultimate exon in *ITGA6*. The UCSC browser view of the transcripts that contain or skip the exon is shown at the top. In the centre are HJAY array data showing changes in probeset intensity in response to ectopic expression of Esrp1 (MDA-MB-231 cells) or knockdown of ESRP1 and 2 (PNT2). In each panel, the red line indicates changes in total transcript expression levels based on probesets in constitutive exons. The blue lines represent average changes detected by each of the eight probes in each probeset against each exon or exon–exon junction. UP denotes upstream junction; DOWN denotes downstream junction; SKIP denotes junction formed by exon skipping. Shown at the bottom are electropherograms from HT-RT–PCR validation assays with gel peaks corresponding to exon inclusion and skipping boxed. Exon inclusion levels are given as percentages for each condition. This format also applies to (**B**–**D**). (**B**) Validation of an ESRP-enhanced exon in *RALGPS2*. (**C**) Validation of an ESRP-silenced exon in *MAGI1*. (**D**) Validation of opposite changes in splicing of two exons in the *MAP3K7* transcript in response to ESRP knockdown. (**E**) RT–PCR was used to assay splicing of select genes in PNT2 cells treated with two different sets of siRNAs (**A**, **B**) directed against both ESRP1 and ESRP2 or cells treated with siRNA-set A and expressing an siRNA-resistant cDNA directing expression of mouse Esrp1 protein. (**F**) Western blot demonstrating expression of the Flag-tagged mouse Esrp1 protein.

upstream introns or within the exon itself as predicted by the motif analysis. We constructed a minigene corresponding to an alternative exon in the *OSBPL3* transcript that is silenced by the ESRPs. There are putative ESRP-binding sites both directly upstream of and within the alternative exon (Figure 4E). Point mutations were introduced to disrupt the



UGG motifs in the upstream intron, in the exon, or in both regions. Exon inclusion was greatly reduced when ESRP1 was co-expressed with the wild-type minigene (Figure 4F). ESRP1 was able to silence exon inclusion when the exonic motifs were mutated, but to a lesser extent compared with wild type. Interestingly, mutation of the intronic motifs alone had a negligible effect on the ability of ESRP1 to promote exon silencing, but when both sets of mutations are made in the minigene, ESRP1-mediated silencing of this exon is nearly abolished. These results suggest that silencing of *OSBPL3* exon 9 by the ESRPs is promoted by binding to both sites, although binding at the exonic sites has a predominant function. To further examine whether the ESRPs can silence splicing through binding to upstream intronic sites alone, we constructed a minigene corresponding to the alternative exon in MAGI1 that is strongly silenced by ESRP expression. Here, we noted a highly conserved UGG-containing *cis*-element upstream of the alternative exon, whereas no clear candidate motifs were present in the exon itself (Figure 4G). Co-transfections revealed that ESRP1 significantly reduced exon inclusion levels in the wild-type minigene but that the UGG mutations essentially abolished ESRP1-mediated silencing activity (Figure 4H). Collectively, these results provide evidence that the ESRPs directly regulate splicing of these exons and strongly implicate UGG-containing motifs as components of ESRP-binding sites that mediate position-dependent splicing regulation.

Sustained ESRP knockdown in mammary epithelial cells induces changes in epithelial–mesenchymal cell markers and morphology consistent with an EMT

To determine whether prolonged disruption of this broad ESRP-regulated splicing programme might induce changes in cell morphology or phenotype indicative of an EMT programme, we transduced human mammary epithelial cells (HMECs) with lentiviruses expressing shRNAs targeting *ESRP1*, *ESRP2*, or both proteins together. These shRNAs effectively reduced the levels of each protein (Figure 5A). As expected, prolonged knockdown of the ESRPs resulted in changes of exon inclusion in ESRP-target transcripts (Figure 5B). We noted little or no change in splicing upon knockdown of ESRP2 alone, whereas ESRP1 knockdown was associated with some changes in splicing and combined knockdown generally achieved the largest change, suggesting that ESRP1 is a more robust splicing regulator. To confirm that these changes in splicing were specific for the defined set

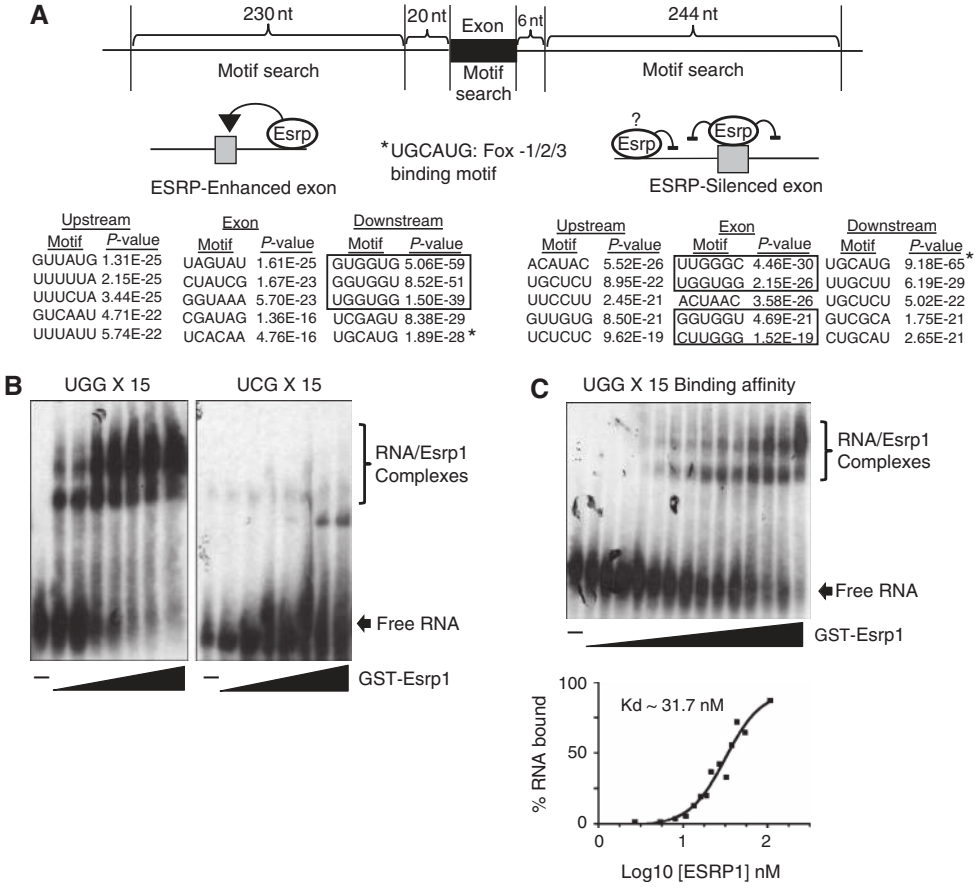


Figure 3 UGG motifs define a high-affinity ESRP-binding site with position-dependent functions. (**A**) Motif enrichment analysis in the exon and flanking 250 nt of intronic sequence, excluding the splice sites. The top five most enriched motifs ranked by Fisher *P*-value (by nucleotide) in each window for ESRP-enhanced and ESRP-silenced exons are shown. UGG-containing motifs are boxed. The Fox-binding site is indicated with an asterisk. (**B**) Esrp1 binds specifically to a UGG repeat-containing RNA. Increasing concentrations of recombinant GST-tagged Esrp1 were bound to UGG X 15 radiolabelled RNA or a mutated UCG X 15 control and resolved on a non-denaturing polyacrylamide gel. (**C**) Determination of Esrp1-binding affinity to the UGG repeats using increasing concentrations of Esrp1 (0–108 nM). The percent of bound RNA was plotted against the concentration of Esrp1 and fit to a sigmoidal curve to determine the dissociation constant (Kd).

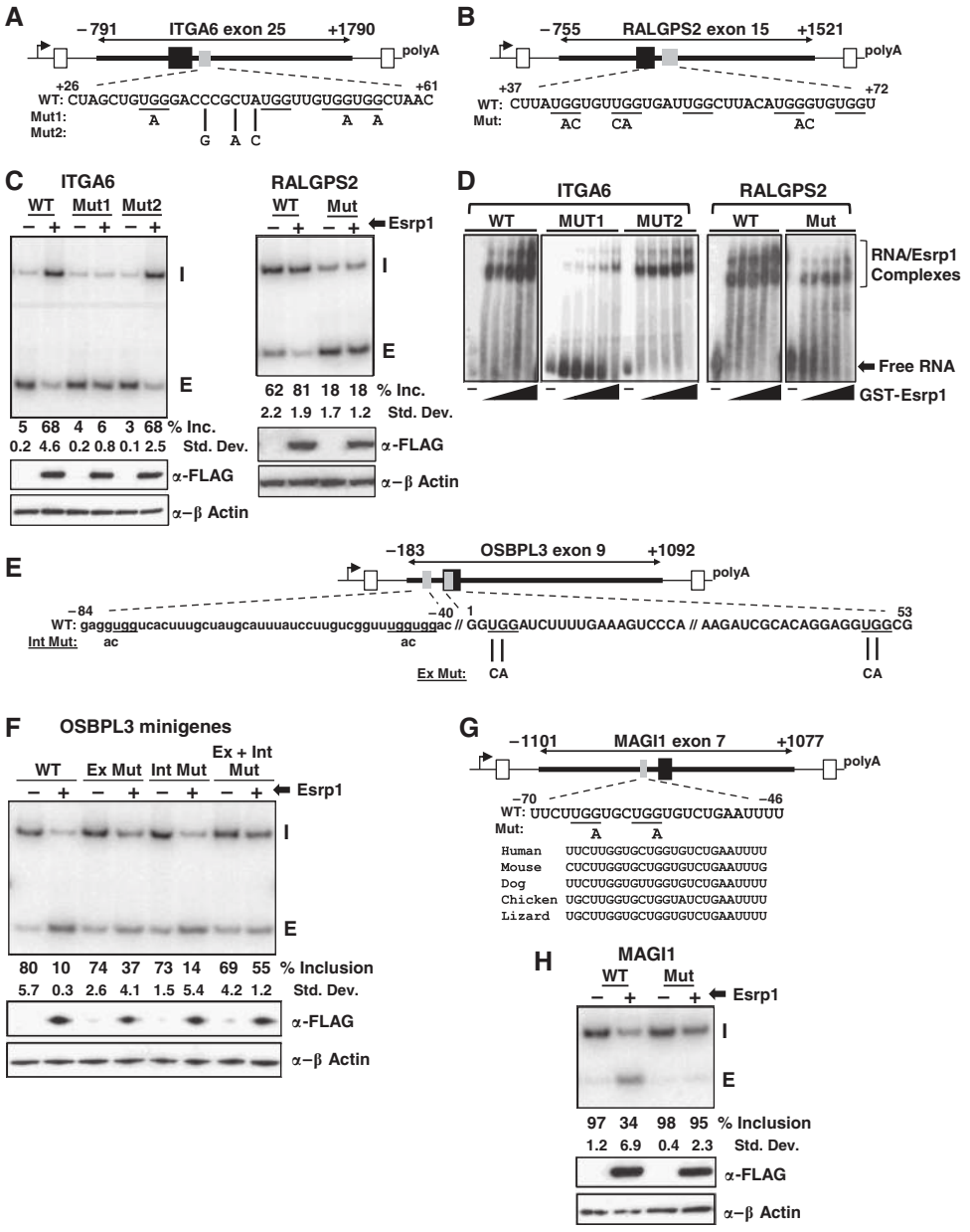


Figure 4 The ESRPs regulate splicing through direct binding to UGG-rich motifs. (**A**) *ITGA6* exon 25 and flanking intronic sequences were inserted into the intron of a minigene reporter. A conserved UGG-rich element downstream of the exon is indicated by a grey box and point mutations both within (Mut1) and outside (Mut2) of UGG motifs are indicated. Numbers above this and subsequent minigenes indicate the number of nucleotides upstream of the 3′ splice site and downstream of the 5′ splice site. (**B**) Schematic of the *RALGPS2* exon 15 minigene. (**C**) ESRP-mediated enhancement of exon inclusion requires the wild-type (WT) UGG-rich elements downstream of the *ITGA6* and *RALGPS2* exons, but does not occur when they are mutated (Mut1). Mutations outside these motifs (Mut2) do not abrogate ESRP-mediated exon splicing. RT–PCR analysis monitoring levels of exon inclusion (I) versus exclusion (E) in 293T cells co-transfected with the respective minigenes and an empty vector control (EV) or vector expressing Esrp1. Data are presented as the average exon inclusion percentage and s.d. from three independent transfections. At bottom are western blots validating expression of FLAG-tagged Esrp1 protein. (**D**) EMSA comparing binding of Esrp1 to the WT *ITGA6* and *RALGPS2* sequence elements and to the respective mutated sequences. Concentrations of Esrp1: 0, 54, 81, 108, 161, and 1075 nM. (**E**) Diagram of the *OSBPL3* exon 9 minigene. Conserved UGG-rich elements occur in both the upstream intron (lower case) and within the exon (upper case letters). Point mutations were introduced in the intron (Int mut), exon (Ex mut), and both together. (**F**) ESRP-mediated exon silencing *OSBPL3* exon 9 involves contributions by both upstream intronic and exonic UGG-rich *cis*-elements. RT–PCR analysis is as described in (**C**). (**G**) Diagram of the *MAGI1* exon 7 minigene. Sequence alignment of the element in five vertebrate species is shown below the sequence element in which UGGs are boxed. (All UGG-rich motifs presented in this figure are highly conserved, but for space reasons the data are limited to this example). (**H**) ESRP-mediated exon silencing requires UGG-rich elements in the introns upstream of the *MAGI1* exon. RT–PCR analysis is as described in (**C**).

of ESRP targets, we also tested several unrelated alternatively spliced exons that showed no change in exon inclusion levels (Supplementary Figure S3). Strikingly, HMECs in which ESRP1 alone or both ESRPs were depleted displayed altered cellular morphology, with a loss of typical epithelial cobblestone appearance and acquisition of an elongated spindle shape compared with control cells (Figure 5C). HMECs expressing the ESRP2 shRNA retained the morphology of

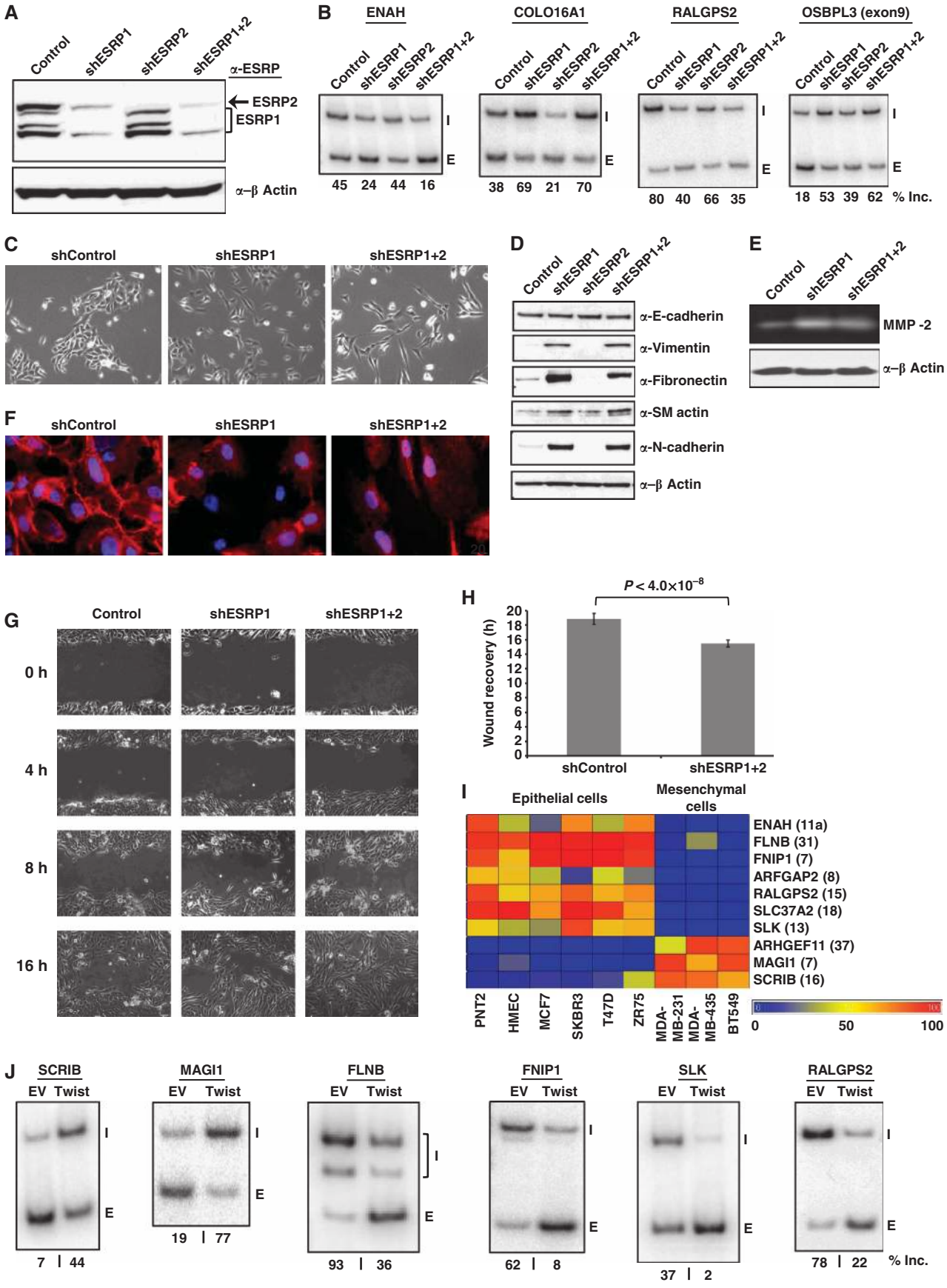
the control cells (data not shown). To assess whether these morphological changes were indicative of an EMT, we analysed for changes in expression of several prototypical EMT markers. Immunoblotting showed that there was a corresponding induction of protein expression for the mesenchymal markers vimentin, fibronectin, N-cadherin, and α-SMA with ESRP1 or combined ESRP1 and ESRP2 knockdown (Figure 5D). We also observed increased matrix metalloprotein-2 activity upon ESRP1 or combined knockdown by in gel zymography; further evidence of invasive mesenchymal cell properties (Figure 5E) (Liu *et al*, 2009). However, there was negligible if any reduction in E-cadherin protein expression or that of other epithelial cell-specific markers (Figure 5D and data not shown). Nonetheless, immunofluorescence analysis of E-cadherin protein showed that ESRP knockdown induced a generalized loss of E-cadherin localization from the plasma membrane at sites of cell–cell junctions and relocalization in a diffuse cytoplasmic pattern, consistent with changes in E-cadherin during the EMT in other model systems (Figure 5F). To investigate the phenotypic consequences of ESRP depletion, we used the scratch wound assay, which demonstrated that a loss of the ESRP-splicing programme promoted a modest, but significant increase in cell motility, consistent with the acquisition of mesenchymal cell properties (Figure 5G and H). Together, these results demonstrate that disruption of the epithelial-splicing programme, regulated predominantly by ESRP1, is sufficient to induce epithelial cells to undergo some of the changes in gene expression and morphology that are associated with the EMT. We note that although ESRP loss is associated with an increase in expression of several mesenchymal markers, there was not an overt decrease in expression of epithelial markers such as E-cadherin. These findings thus indicate that even in the absence of a complete switch in the transcription of several known EMT markers, these global changes in splicing can induce some of the transitions in cell behaviour that are observed during the EMT. These studies also support our hypothesis that the ESRP-regulated splicing programme is an essential feature of the epithelial phenotype and that many changes in cell behaviour associated with the EMT are due to alterations in the functions of proteins that undergo isoform switching during this process.

A splicing signature that can be used to define epithelial and mesenchymal cell types and the EMT

Using the validated set of ESRP-regulated exons with large changes in exon inclusion, we chose a panel of alternative exons to derive a select set of splicing assays that might distinguish epithelial from mesenchymal cell types and that might also serve as markers that change during the EMT. We used HT-RT–PCR to assay splicing of 13 ESRP-enhanced exons and seven ESRP-silenced exons across six epithelial cell lines and three mesenchymal cell lines. These cell lines have previously been characterized based on traditional epithelial and mesenchymal markers such as E-cadherin and vimentin, and we showed that ESRP1 and ESRP2 mRNA expression is limited to the epithelial cell lines (Warzecha *et al*, 2009a). A distinct epithelial 'splicing signature' of 10 exons that are specifically included or skipped in epithelial cells emerged from this analysis (Figure 5I; Supplementary Figure S4A and B). ESRP expression is strongly downregulated during the EMT induced in HMECs by the transcription factor Twist. We therefore also assayed alternative splicing following Twist-induced EMT for changes in splicing of this subset of alternative splicing events. This analysis revealed a switch from epithelial-to-mesenchymal exon inclusion for transcripts derived from this splicing signature (Figure 5J). These data, thus, clearly establish that distinct splicing switches that occur during the EMT are regulated by the ESRPs and that limited components of this splicing signature hold promise for use as specific markers that can be used to define the EMT as well as in the classification of cell types.

It is noteworthy that among the cell lines we used to profile this splicing signature were several breast cancer cell lines that have been categorized as belonging to cancer subtypes designated luminal (MCF7, SKBR3, T47D, and ZR75) or basal B (MDA-MB-231, MDA-MB-435, and BT549). In contrast to luminal or basal A cell lines, the basal B cell lines are representative of cancer cells that have undergone the EMT and have greater metastatic potential, invasive properties, and resistance to several cancer therapies (Neve *et al*, 2006; Blick *et al*, 2008; Sarrio *et al*, 2008). Consistent with a loss of an ESRP-regulated splicing signature in the cell lines analysed here, analysis of the expression data from the large set of cell

Figure 5 Identification of a discrete ESRP-regulated splicing signature that can distinguish epithelial from mesenchymal cell types. (**A**) Western blot using a monoclonal antibody that recognizes both ESRPs to validate specific shRNA-mediated depletion in the nucleus of HMEC cells (**B**) Switches in splicing of ESRP-target genes after prolonged knockdown of each ESRP alone or in tandem. RT–PCR was used to assay alternative splicing in the HMEC samples described in (**A**). Percent exon inclusion is indicated beneath each lane. (**C**) Prolonged knockdown of ESRP1 or ESRP1 and 2 induces morphological changes in HMEC cells. Phase-contrast images illustrate that cells depleted of ESRP1 and 2 or ESRP1 alone undergo EMT-like changes in morphology including a reduction in cobblestone appearance and cell–cell contacts and increased stellate-shaped cells. (**D**) Western blot demonstrating pronounced increase in expression of mesenchymal protein markers in cells depleted of ESRP1 or ESRP1 and 2 but not ESRP2 alone compared with control cells. (**E**) An increase in matrix metalloprotein-2 (MMP-2) activity accompanies the increase in expression of mesenchymal proteins. Conditioned media were collected and analysed by gelatin zymography to detect MMP-2 secretion. (**F**) Immunofluorescence staining of E-cadherin (red) shows pronounced localization near the plasma membrane in control cells and diffused cytoplasmic localization in cells depleted of ESRP1 or ESRP1 and 2. Nuclei are stained with DAPI. (**G**) Prolonged knockdown of ESRP1 alone or ESRP1 and ESRP2 results in increased cell motility. The scratch wound-healing assay was used to monitor migration of the HMEC samples shown in (**C**). (**H**) The scratch assay was repeated with nine replicates each for HMEC shRNA control and combined shRNAs against ESRP1 and ESRP2 and time to complete wound closure determined. Wound closure time was significantly shorter for cells depleted of both ESRPs compared with the control cells. Each bar represents the mean ± s.d. of replicate samples (*P*-value determined by Welch two-sample *t*-test). (**I**) Heat map illustrating exon inclusion levels from select genes across a panel of epithelial and mesenchymal cell lines as measured by HT-RT–PCR. (**J**) Switches in splicing of the 'EMT signature' exons during Twist-induced EMT in HMEC cells. RT–PCR was used to assay alternative splicing in HMEC cells transduced with empty vector (EV) or Twist expressing retroviruses. Percent exon inclusion is indicated beneath each lane.


A
Control
shESRP1
shESRP2
shESRP1+2
α-ESRP
ESRP2
ESRP1
α–β Actin
B
ENAH
COLO16A1
RALGPS2
OSBPL3 (exon9)
I
E
45 24 44 16
38 69 21 70
80 40 66 35
18 53 39 62 % Inc.
C
shControl
shESRP1
shESRP1+2
D
α-E-cadherin
α-Vimentin
α-Fibronectin
α-SM actin
α-N-cadherin
α–β Actin
E
MMP -2
F
G
0 h
4 h
8 h
16 h
H
P < 4.0×10−8
Wound recovery (h)
I
Epithelial cells
Mesenchymal cells
ENAH (11a)
FLNB (31)
FNIP1 (7)
ARFGAP2 (8)
RALGPS2 (15)
SLC37A2 (18)
SLK (13)
ARHGEF11 (37)
MAGI1 (7)
SCRIB (16)
PNT2
HMEC
MCF7
SKBR3
T47D
ZR75
MDA-MB-231
MDA-MB-435
BT549
0
50
100
J
SCRIB
MAGI1
FLNB
FNIP1
SLK
RALGPS2
EV Twist
7 44
19 77
93 36
62 8
37 2
78 22 % Inc.

lines used to define these cancer cell line subtypes revealed a very strong association of low or absent ESRP1 and ESRP2 expression with basal B-cell classification (Supplementary Figure S4C). Furthermore, low ESRP1 and ESRP2 mRNA expression was also a component of a gene signature in breast cancer specimens classified as claudin-low and metaplastic, which are also more aggressive and show EMT-like features (Hennessy *et al*, 2009). It is thus likely that this splicing signature is broadly associated with cancer cells with EMT features and might therefore have future clinical applications.

Discussion

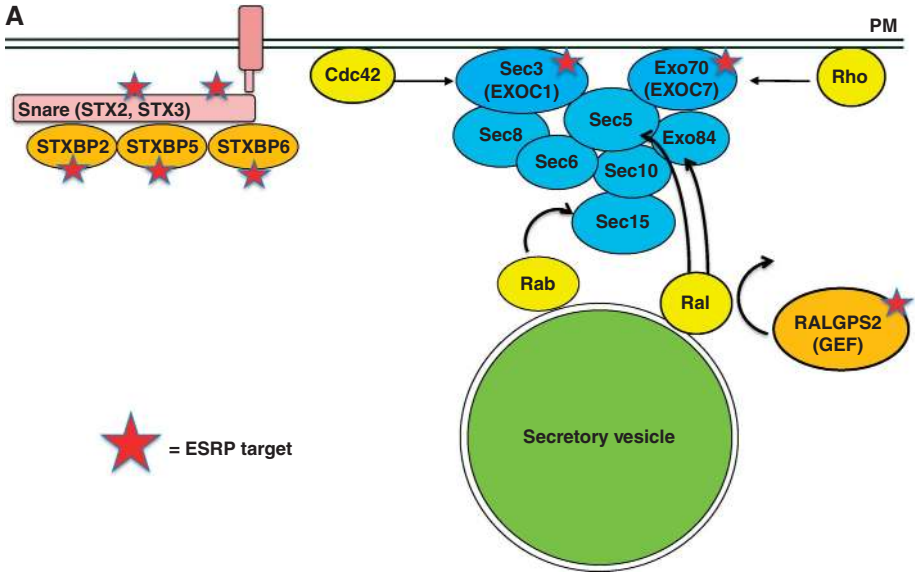
We defined an extensive ESRP-regulated splicing network and provided strong evidence that alternative splicing provides an additional layer at which gene expression is regulated during the EMT. Furthermore, our data demonstrate that this post-transcriptional gene expression programme is distinct from the transcriptional gene expression programme that is altered during this process; thereby pointing out the degree to which alternative splicing contributes to the global transcriptomic complexity that is implicated in the EMT. Our data demonstrating that the loss of this splicing programme alone can induce some of the phenotypic changes that occur during the EMT support our hypothesis that the ESRP-regulated splicing programme is an essential feature of the epithelial phenotype and that many changes in cell behaviour associated with the EMT are due to alterations in functions of proteins that undergo isoform switching during this process. We suggest that the ESRPs are among the most essential epithelial-specific genes that, like E-cadherin, must be downregulated directly or indirectly by mesenchymal transcription factors such as Twist, Snail/Slug, or Zeb1/2 in order to achieve a complete phenotypic cellular switch in the EMT.

Our observation of changes in cell morphology and behaviour through abrogation of an epithelial-splicing programme suggests that the ESRPs regulate transcripts that, while expressed in both epithelial and mesenchymal cells, encode protein isoforms specific to each cell type with isoform-specific functions that impact the EMT. A prototypical example of such a protein is p120-catenin, which has been extensively documented to have a function in the EMT and undergo isoform switches during this process. The ESRPs regulate the splicing of exons that yield distinct epithelial versus mesenchymal p120-catenin isoforms that oppositely promote cell–cell adhesion in epithelial cells and cell migration and invasion in mesenchymal cells (Yanagisawa *et al*, 2008; Warzecha *et al*, 2009a). Among the ESRP targets identified in this study are numerous examples of transcripts that encode proteins with functions in processes that underlie the EMT such as regulation of the actin cytoskeleton, cell–cell and cell–matrix adhesion, migration, and cell polarity. For example, NUMB has been shown to have an essential function in the maintenance of cell polarity and cell–cell adhesion through binding to E-cadherin through an N-terminal phosphotyrosine-binding (PTB) domain. Abrogation of this interaction during the EMT leads to loss of cell–cell adhesion and increased cell migration (Wang *et al*, 2009). The 33 nt epithelial-specific exon in *NUMB* encodes an 11 amino acid insert in this PTB domain that is likely to promote the NUMB-E-cadherin interaction. Consistent with this possibility, isoforms that contain this insert are predominantly associated with the plasma membrane, whereas isoforms that exclude it are diffusely cytoplasmic (Dho *et al*, 1999). Relocalization of β-catenin from adherens junctions to the nucleus occurs when E-cadherin is downregulated, and in association with TCF/LEF transcription family members (including TCF7L2) induces changes in gene expression that are associated with the EMT (Medici *et al*, 2008). The ESRPs promote inclusion of exon 4 in *TCF7L2* (also known as *TCF4*) transcripts, whereas this exon 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, suggesting the interesting possibility that a partial switch towards the skipped isoform during the EMT might further enhance the transcriptional activation of β-catenin-target genes upon its translocation to the nucleus (Weise *et al*, 2010). These examples thus not only highlight several of the ESRP-regulated transcripts with EMT-related functions, but also cases in which there is evidence suggesting that the resulting epithelial versus mesenchymal isoforms have distinct and potentially even opposing functions in the maintenance of cell adhesion versus cell migration.

Another notable ESRP target encoding alternative protein isoforms that have been functionally characterized with implications for the function of splicing in cancer is Bin1. Bin1 binds Myc and inhibits its oncogenic activity, but isoforms that lack the domain encoded by exon 12A lack this tumour suppressor activity (Ge *et al*, 1999). A previous study showed that overexpression of the splicing factor SRSF1 caused increased inclusion of exon 12A, one of several alterations it induced in alternative splicing suggesting that this SR protein functions as a proto-oncogene (Karni *et al*, 2007). Conversely, the ESRPs promote skipping of exon 12A, thereby promoting expression of the 'tumour suppressor' isoform and the hypothesis that ESRP1 itself is a tumour suppressor was put forth prior to our finding that it is a splicing regulator (Leontieva and Ionov, 2009). Although most of the changes in splicing of ESRP-regulated exons have not been functionally characterized, several additional examples of ESRP-regulated transcripts in which splicing can change the functions of the resulting isoforms, or affect domains predicted to change their functions in ways that might be relevant for the EMT are described in Supplementary Table S7. This study therefore sets the stage for further cell biological investigations into the functional consequences of the changes in splicing in numerous transcripts as a result of ESRP downregulation during the EMT. As more studies begin to catalog the functions of these proteins, it will likely become clear that functional annotations of gene products will need to increasingly incorporate isoform-specific properties into functional genomic databases in order to fully capture gene-function associations.

In the published literature, we also identified examples that support the hypothesis that the ESRPs co-regulate transcripts that function in common processes and pathways. One example of functionally related ESRP targets with functions in the EMT includes p38 (MAPK14), p38IP (FAM48A), and MAP4K4 (NIK). NIK has been shown to act upstream to activate p38 pathways that require p38IP to repress E-cadherin expression at the protein level during the EMT (Zohn *et al*, 2006). Thus, changes in the splicing of these transcripts may influence the dynamics of this pathway during the EMT. Another example that illustrates this concept is a coherent

set of ESRP targets encoding proteins that function in vesicle-mediated protein trafficking pathways that have essential functions in the maintenance of cell polarity and cell–cell adhesion (Mellman and Nelson, 2008). The exocyst is a hetero-octameric complex that is essential for maintenance of epithelial polarity through the targeting of secretory and membrane-bound proteins to specific sites at the plasma membrane (He and Guo, 2009). In contrast, in cells that lack epithelial E-cadherin, the exocyst is localized to the leading edge and focal complexes of migrating cells (Spiczka and Yeaman, 2008). We identified two ESRP-regulated epithelial-specific exons in two of the eight exocyst proteins, EXOC1 (Sec3) and EXOC7 (Exo70) (Figure 6). These exocyst proteins are the two membrane proximal components of the complex that are essential for vesicle targeting and Rho GTPase recruitment and are therefore at a critical interface in regulating polarized exocytosis. RalA is required upstream for exocyst assembly and function and the RalA-specific activating guanine nucleotide exchange factor (GEF) *RALGPS2* displayed nearly complete switches in splicing (Cascone *et al*, 2008). The epithelial-specific exon in *RALGPS2* is an in-frame exon located between a PxxP motif and pleckstrin homology (PH) domain, a region of the protein previously proposed to regulate its GEF activity (Ceriani *et al*, 2007). It therefore will be of interest to determine functional differences in activity of RALGPS2, EXOC1, and EXOC7 protein isoforms that may differentially impact the diverse activities of exocyst complex. We also noted several ESRP targets among SNARE protein complexes that mediate downstream vesicle fusion with the plasma membrane, including Syntaxin 2 (STX2; epimorphin) and Syntaxin 3 (STX3) as well as syntaxin-binding proteins STXBP2 (Hunc18b), STXBP5 (tomosyn), and STXBP6 (amisyn) (Kauppi *et al*, 2002; Sharma *et al*, 2006). Taken together, the multiple ESRP-



B

Gene	PNT2 siGFP	PNT2 siESRP1/2	PNT2 % Change	MDA-MB-231 EV	MDA-MB-231 Esrp1	MDA-MB-231 % Change	ESRP function
EXOC1	ND	ND	ND	6.3	42.5	36.2	Enhance
EXOC7	31.2	6.6	–24.6	2.4	42.0	39.6	Enhance
STX2	26.7	46.7	20	ND	ND	ND	Silence
STX3	97.1	87.5	–9.6	52.3	97.7	45.4	Enhance
STXBP2	ND	ND	ND	88.8	44.7	–44.1	Silence
STXBP5	12.4	18.5	6.1	9.0	4.4	–4.6	Silence

C

Rho regulators and effectors		Rab regulators and effectors
AKAP13	ECT2	ERC1
ARHGAP10	FAM13B1	RAB18
ARHGAP17	MPRIP	RAB28
ARHGEF11	MYO9B	RAB6A
ARHGEF12	PLXNB2	RABEP1
CDC42BPA	RHOT1	RABEPK
CDC42BPB	RTKN2	TBC1D13
CDC42SE1	SRGAP2	TSC2
DAAM1		

Figure 6 ESRP-regulated gene transcripts encode multiple components and regulators of the vesicular transport system. (**A**) Schematic illustrating ESRP-target transcripts that function in vesicular transport events mediated through exocyst function. Exocyst components are shown in blue. PM, plasma membrane. (**B**) Validated changes in exon inclusion among the ESRP-target transcripts involved in exocyst pathways. (**C**) RHO and RAB regulators and effectors that are predicted and/or validated to be targets of ESRP-splicing regulation.

regulated splicing events that connect upstream and downstream functions of the exocyst complex suggest that the different isoforms carry out functions that differentially impact exocytosis to either maintain cell polarity in epithelial cells or regulate focal complex turnover and cell migration in mesenchymal cells (Figure 6).

We also performed Gene Ontology analysis using the DAVID functional annotation tool to characterize processes and functions that were enriched among the ESRP-target genes (Dennis *et al*, 2003). Consistent with our hypothesis that the ESRPs regulate gene products with EMT-related functions, we observed enrichment for genes involved in maintenance of cell–cell adhesion, cell motility, and cell–matrix adhesion (Supplementary Table S8). There was also enrichment for regulators and effectors of small GTPases, including the Rho GTPases that are well characterized to have functions in the EMT as well as exocyst function (Iden and Collard, 2008; He and Guo, 2009). Another enriched category was protein serine/threonine kinases, including components of ERK/MAPK signalling pathways. Analysis of KEGG pathways showed enriched participation of these targets in regulation of the actin cytoskeleton and maintenance of adherens junctions.

Our identification of a splicing signature of ESRP-regulated exons provides a novel set of biomarkers for the EMT that can facilitate analysis of cancers in which analysis of EMT-like changes may predict prognosis or therapeutic responses. As many of the splicing-target genes we identified are expressed in both epithelial and mesenchymal cells, these splicing switches provide a set of gene expression changes distinct from those that are regulated at the level of transcription. Current markers used to characterize the EMT rely on collections of transcript or protein biomarkers that have utility in different subtypes of EMT (Zeisberg and Neilson, 2009). However, the utility of these standard markers is often compromised in tissue specimens that can often have significant protein or RNA degradation. In contrast to the use of whole transcript profiling of RNA, the determination of changes in splicing provides a tool that can define epithelial and mesenchymal cell types even when RNA quality is compromised, provided that products can be obtained by RT–PCR. We are currently testing the utility of this signature in an expanded set of cell lines and tissues in order to determine a splicing metric that can be applied to quantitatively define the EMT in diverse tissues and pathologic specimens.

Sequence enrichment analysis in ESRP-regulated exons and flanking introns was used to derive a candidate-binding motif that was validated as a direct ESRP-binding site as well as an ESRP-responsive element in minigene reporters. Mapping of this motif relative to ESRP-regulated exons revealed a trend suggestive of an RNA map of position-dependent ESRP binding in which binding sites in the downstream intron promote exon inclusion and binding within or upstream of an exon promote exon skipping. This RNA map is similar to that previously characterized for position-dependent functions of the Nova- and Fox-splicing regulators, with the possible exception that the ESRPs would appear to more commonly regulate exon silencing through direct binding within exons (Licatalosi *et al*, 2008; Zhang *et al*, 2008). However, some exceptions were observed that might be due to more degenerate UG-rich ESRP-binding sites or presence of binding sites in more distal intron regions or flanking exons. Alternatively, some of the ESRP-induced splicing changes we observed may be indirect as we identified a number of known splicing factors among the ESRP targets. It is therefore possible that some of the splicing changes may come about through modifications in the activity of these splicing regulators. Nonetheless, our results suggest that many if not most of the splicing changes we observed are likely to involve direct affects on splicing of these targets and that the identification of ESRP-binding sites can predict *de novo* targets and the splicing outcome based on the position of binding. Experiments to map genome-wide ESRP-binding sites using cross-linking and immunoprecipitation are underway and will further refine the 'RNA map' that predicts the functions of these proteins in alternative splicing, including more complex types of alternative splicing. We note, however, that an RNA-binding map for a single-splicing regulator alone is not sufficient to determine a comprehensive splicing code. Given the overlap of this ESRP-SRN with the Fox-splicing network, the predictive capabilities of an 'RNA map' derived from a single regulator is limited when applied to splicing outcomes in different cellular milieus. Such observations further confirm that global combinatorial control of alternative splicing involves large and overlapping networks of splicing regulators with different activities in different cell types that together determine the transcriptome within any given cell. However, through compilation of RNA-binding maps and splicing events regulated by expanding numbers of splicing factors we will not only unveil a more definitive splicing code, but also begin to uncover the molecular mechanisms by which these regulators function.

Materials and methods

Cell culture, transfection, and transduction

PNT2, MDA-MB-231, HMEC (HMLE), and 293T cells were maintained, transfected, and retrovirally transduced as described (Warzecha *et al*, 2009a). Permanent ESRP1 and 2 knockdown in HMLE cells was achieved with miR-30-based shRNA lentiviral plasmids with puromycin and blasticidin selection cassettes and selection in 0.5 μg/ml of puromycin and 10 μg/ml of blasticidin S.

HJAY microarray target preparation and hybridization

RNA purification, preparation of biotinylated sense-strand DNA, and hybridization was carried out as described for exon array analysis (Warzecha *et al*, 2009b). The CEL file data sets can be accessed at NCBI GEO repository through accession number GSE22937.

RT–PCR and HT-RT–PCR

Quantification of splicing by standard RT–PCR and high-throughput RT–PCR at the Université de Sherbrooke was performed as described (Warzecha *et al*, 2009b). For complex splicing events, restriction digestion and/or DNA sequencing was used to validate PCR product identity. Complete HT-RT–PCR data can be accessed at http://palace.lgfus.ca.

Antibodies and immunoblotting

Cell lysates were resolved on 4–12% NuPAGE Bis-tris gels (Invitrogen) and immunoblotted as described (Warzecha *et al*, 2009a). Antibodies used were anti-β actin (Sigma AC-15, 1:1000), anti-Flag (Stratagene, 1:1000), anti-Vimentin (Thermo Scientific Ab2 Clone V9, 1:150), anti-Fibronectin (BD Transduction, 1:10 000), anti-smooth muscle actin (Sigma, 1:300), anti-N-cadherin (BD Transduction, 1:250), and anti-E-cadherin (Cell Signaling Technology, 1:1000). ESRP1 antibodies that specifically detect ESRP1 (clone 27H12) and both ESRP1 and ESRP2 (clone 23A7) were obtained as custom monoclonal antibodies against our Glutathione-s-transferase

(GST)-Esrp1 recombinant protein as prepared by Rockland Immunochemicals, Inc. (Gilbertsville, PA). These antibodies were shown to be specific with no detectable cross-reactivity against the closest RNA-binding protein homologs GRSF1, hnRNPF, or hnRNPH1.

Recombinant protein production and RNA EMSA analysis

Production of recombinant GST-Esrp1 fusion protein from plasmid GST-Esrp1-FL was performed using standard techniques and purified using Glutathione sepharose beads (GE healthcare). Sequences for *in vitro* transcription were inserted into the *Cla*I-*Xho* restriction sites in plasmid pDP19RC-ΔEE and transcription by T7 polymerase (Ambion) after *Xho*I digestion was carried out as described (Hovhannisyan and Carstens, 2007). [^{32}P]UTP-radiolabelled RNAs were used at a specific activity of 2.4×10^8 CPM/μg and 50 000 CPM (~8–13 fmol) of radiolabelled RNA and GST-Esrp1 were incubated in a 10-μl binding reaction (16 mM Hepes pH 7.9, 95 mM KCl, 2 mM $MgCl_2$, 60 μM EDTA, 1 mM DTT, 6% glycerol, 0.1 mM heparin (Sigma #H-3393), and 0.15 mM PMSF). Binding reactions were incubated at 30°C for 30 min and loaded on 4% non-denaturing Tris-borate acrylamide gels and electrophoresed at 4°C for 1.5 h at 200 V. Quantification for Kd determination was performed using a PhosphorImager (Molecular Dynamics) and the % RNA bound was determined by calculating the ratio of [(shifted complexes)/(free RNA + shifted complexes)]. Data were analysed using the dose response with variable slope regression equation in Prism software and the Kd was determined using the EC50 value corresponding to the concentration of ESRP1 at 50% binding saturation.

In gel zymography

Cells were cultured in serum-free MEGM (Lonza) for 24 h. Cell culture media was lyophilized to dryness and rehydrated with SDS loading buffer (10% SDS, 50% glycerol, 0.4 M Tris, pH 6.8, and 0.1% bromophenol blue) and separated on 8% polyacrylamide/0.3% gelatin gels. Gels were then washed in 0.01% Triton X-100, 1 h rocking gently, and incubated in reaction buffer (50 mM Tris, pH 8.0, and 5 mM $CaCl_2$) at 37°C for 24 h. After the reaction, gels were stained with staining buffer (0.12% Coomassie blue R-250, 50% methanol, and 20% acetic acid) for 1 h and destained overnight with destaining buffer (22% methanol and 10% acetic acid). Gel loadings were normalized to total protein measured with a Bio-Rad Protein Assay (Richmond, CA).

Immunofluorescence

Transduced HMEC cells were grown on coverslips, washed with phosphate-buffered saline (PBS), fixed in 4% paraformaldehyde at room temperature for 12 min, washed, permeabilized for 5 min with PBS + 0.2% Triton X-100, and blocked for 10 min with 10% goat serum in PBS. The coverslips were incubated sequentially with anti-E-cadherin primary antibody (1:150) and Texas Red-conjugated anti-Rabbit secondary antibody (1:4000) and counterstained with DAPI blue.

In vitro scratch assay

Cells were cultured to confluence on six-well plates, and scratched with a 200-μl pipette tip. Detached cells were removed, and the remaining cells were incubated with MEGM (Lonza). Scratched areas for each sample were marked and photographed immediately or at the indicated time points. For quantification, the same analysis was performed with nine replicate scratches for the control and combined ESRP1 and ESRP2 knockdown cells. Each scratch was then visually observed until the time point at which complete closure of the wound occurred.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Don Baldwin, Yi Li, and John Hogenesch for technical support and discussion; Kristen Lynch for review of the manuscript; and the University of Iowa ICTS for computer support. This research was supported by NIH grants R01-CA093769 and R01-GM088809 (RPC), R01-HG004634 and Edward Mallinckrodt, Jr Foundation (YX), R01-GM085146 (WG), and T32-GM08216 (CCW).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ (2010) Deciphering the splicing code. *Nature* **465:** 53–59

Blick T, Widodo E, Hugo H, Waltham M, Lenburg ME, Neve RM, Thompson EW (2008) Epithelial mesenchymal transition traits in human breast cancer cell lines. *Clin Exp Metastasis* **25:** 629–642

Cascone I, Selimoglu R, Ozdemir C, Del Nery E, Yeaman C, White M, Camonis J (2008) Distinct roles of RalA and RalB in the progression of cytokinesis are supported by distinct RalGEFs. *EMBO J* **27:** 2375–2387

Ceriani M, Scandiuzzi C, Amigoni L, Tisi R, Berruti G, Martegani E (2007) Functional analysis of RalGPS2, a murine guanine nucleotide exchange factor for RalA GTPase. *Exp Cell Res* **313:** 2293–2307

Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10:** 741–754

Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4:** P3

Dho SE, French MB, Woods SA, McGlade CJ (1999) Characterization of four mammalian numb protein isoforms. Identification of cytoplasmic and membrane-associated variants of the phosphotyrosine binding domain. *J Biol Chem* **274:** 33097–33104

Ge K, DuHadaway J, Du W, Herlyn M, Rodeck U, Prendergast GC (1999) Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. *Proc Natl Acad Sci USA* **96:** 9689–9694

Hartmann B, Valcarcel J (2009) Decrypting the genome's alternative messages. *Curr Opin Cell Biol* **21:** 377–386

He B, Guo W (2009) The exocyst complex in polarized exocytosis. *Curr Opin Cell Biol* **21:** 537–542

Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee JS, Fridlyand J, Sahin A, Agarwal R, Joy C, Liu W, Stivers D, Baggerly K, Carey M, Lluch A, Monteagudo C, He X, Weigman V, Fan C, Palazzo J *et al* (2009) Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res* **69:** 4116–4124

Hovhannisyan RH, Carstens RP (2007) Heterogeneous ribonucleoprotein m is a splicing regulatory protein that can enhance or silence splicing of alternatively spliced exons. *J Biol Chem* **282:** 36265–36274

Iden S, Collard JG (2008) Crosstalk between small GTPases and polarity proteins in cell polarization. *Nat Rev Mol Cell Biol* **9:** 846–859

Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR (2007) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* **14:** 185–193

Kauppi M, Wohlfahrt G, Olkkonen VM (2002) Analysis of the Munc18b-syntaxin binding interface. Use of a mutant Munc18b to dissect the functions of syntaxins 2 and 3. *J Biol Chem* **277:** 43973–43979

LaGamba D, Nawshad A, Hay ED (2005) Microarray analysis of gene expression during epithelial-mesenchymal transformation. *Dev Dyn* **234:** 132–142

Leontieva OV, Ionov Y (2009) RNA-binding motif protein 35A is a novel tumor suppressor for colorectal cancer. *Cell Cycle* **8:** 490–497

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456:** 464–469

Liu J, Yue P, Artym VV, Mueller SC, Guo W (2009) The role of the exocyst in matrix metalloproteinase secretion and actin dynamics during tumor cell invadopodia formation. *Mol Biol Cell* **20:** 3763–3771

Lu H, Lin L, Sato S, Xing Y, Lee CJ (2009) Predicting functional alternative splicing by measuring RNA selection pressure from multigenome alignments. *PLoS Comput Biol* **5:** e1000608

Medici D, Hay ED, Olsen BR (2008) Snail and Slug promote epithelial-mesenchymal transition through beta-catenin-T-cell factor-4-dependent expression of transforming growth factor-beta3. *Mol Biol Cell* **19:** 4875–4887

Mellman I, Nelson WJ (2008) Coordinated protein sorting, targeting and distribution in polarized cells. *Nat Rev Mol Cell Biol* **9:** 833–845

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM *et al* (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10:** 515–527

Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463:** 457–463

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40:** 1413–1415

Pan Q, Shai O, Misquitta C, Zhang W, Saltzman AL, Mohammad N, Babak T, Siu H, Hughes TR, Morris QD, Frey BJ, Blencowe BJ (2004) Revealing global regulatory features of Mammalian alternative splicing using a quantitative microarray platform. *Mol Cell* **16:** 929–941

Resch A, Xing Y, Alekseyenko A, Modrek B, Lee C (2004) Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. *Nucleic Acids Res* **32:** 1261–1269

Sarrio D, Rodriguez-Pinilla SM, Hardisson D, Cano A, Moreno-Bueno G, Palacios J (2008) Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer Res* **68:** 989–997

Sharma N, Low SH, Misra S, Pallavi B, Weimbs T (2006) Apical targeting of syntaxin 3 is essential for epithelial cell polarity. *J Cell Biol* **173:** 937–948

Shen S, Warzecha CC, Carstens RP, Xing Y (2010) MADS+: discovery of differential splicing events from Affymetrix exon junction array data. *Bioinformatics* **26:** 268–269

Spiczka KS, Yeaman C (2008) Ral-regulated interaction between Sec5 and paxillin targets Exocyst to focal complexes during cell migration. *J Cell Sci* **121:** 2880–2891

Thiery JP, Acloque H, Huang RY, Nieto MA (2009) Epithelial-mesenchymal transitions in development and disease. *Cell* **139:** 871–890

Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, Durand M, Couture S, Froehlich U, Lapointe E, Lucier JF, Thibault P, Rancourt C, Tremblay K, Prinos P, Chabot B, Elela SA (2009) Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* **16:** 670–676

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476

Wang Z, Sandiford S, Wu C, Li SS (2009) Numb regulates cell-cell adhesion and polarity in response to tyrosine kinase signalling. *EMBO J* **28:** 2360–2373

Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP (2009a) ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol Cell* **33:** 591–601

Warzecha CC, Shen S, Xing Y, Carstens RP (2009b) The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol* **6:** 546–562

Weise A, Bruser K, Elfert S, Wallmen B, Wittel Y, Wohrle S, Hecht A (2010) Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/{beta}-catenin targets. *Nucleic Acids Res* **38:** 1964–1981

Xing Y, Lee CJ (2005) Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. *PLoS Genet* **1:** e34

Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H, Fu XD, Zhang Y (2009) Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* **36:** 996–1006

Yanagisawa M, Huveldt D, Kreinest P, Lohse CM, Cheville JC, Parker AS, Copland JA, Anastasiadis PZ (2008) A p120 catenin isoform switch affects Rho activity, induces tumor cell invasion, and predicts metastatic disease. *J Biol Chem* **283:** 18344–18354

Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16:** 130–137

Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB (2005) Identification and analysis of alternative splicing events conserved in human and mouse. *Proc Natl Acad Sci USA* **102:** 2850–2855

Zeisberg M, Neilson EG (2009) Biomarkers for epithelial-mesenchymal transitions. *J Clin Invest* **119:** 1429–1437

Zhang C, Zhang Z, Castle J, Sun S, Johnson J, Krainer AR, Zhang MQ (2008) Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev* **22:** 2550–2563

Zohn IE, Li Y, Skolnik EY, Anderson KV, Han J, Niswander L (2006) p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation. *Cell* **125:** 957–969

