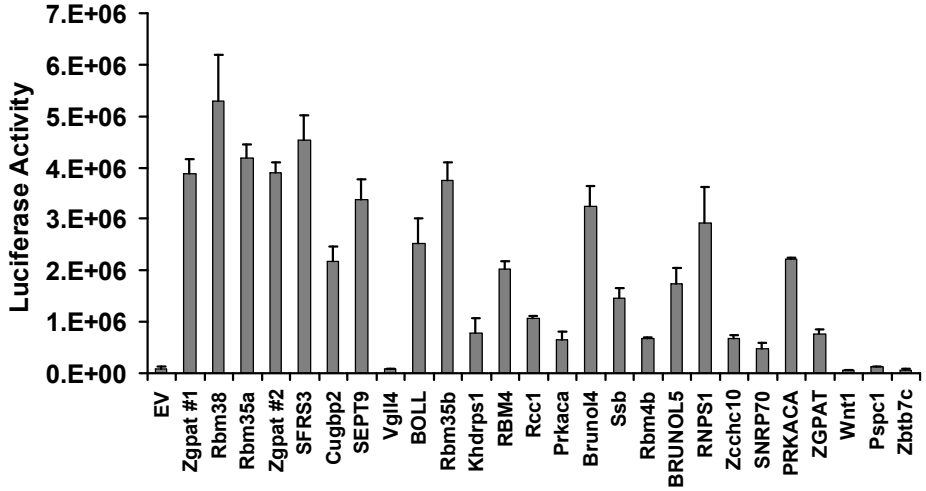
ESRP1 and ESRP2 are epithelial cell type-specific regulators of FGFR2 splicing

Claude C. Warzecha, Trey K. Sato, Behnam Nabet, John B. Hogenesch, and Russ P. Carstens

Supplemental Data

A



B

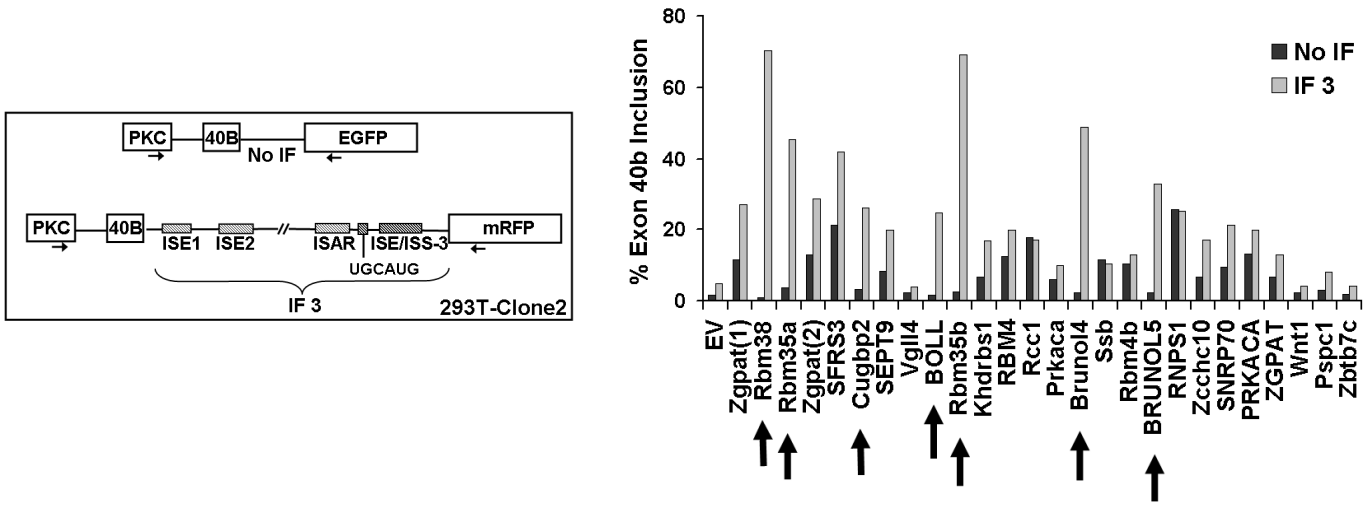


Figure S1. Validation of MGC screen hits in luciferase- and fluorescence-based minigene expressing 293T cell lines.

(A) 293T cells stably expressing the luciferase reporter minigene were transiently transfected with an empty pCMV-Sport vector control, Fox1, or MGC cDNAs from Table 1. Luciferase activity was assayed and the mean values for three independent transfections are shown in the graph. Error bars denote standard deviation.

(B) To determine whether enhanced inclusion of exon 40B by each cDNA was dependent on the FGFR2 cis-elements we used a 293T clone (Clone 2) stably co-expressing separate EGFP and mRFP-based minigenes (Newman et al., 2006). As shown schematically at the left, the control EGFP minigene does not contain any FGFR2 sequences whereas the experimental mRFP minigene with IF3 is similar to the luciferase-based minigene used in the primary screen. 293T-clone2 cells were transiently transfected with the MGC clones from

Table 1 or empty vector (EV) and exon inclusion from each reporter determined by RT-PCR using primers sets specific for each minigene. This analysis revealed that only seven of the 18 confirmed hits required FGFR2 intron fragment 3 (IF3) to significantly enhance exon 40B inclusion (right panel, arrows). Three of these seven genes, CUGBP2, Brunol4, and BRUNOL5, are members of the CELF/Brunol family of splicing regulators (Ladd et al., 2001). Despite their function in the heterologous reporter, transient co-transfections of Brunol4 and BRUNOL5 cDNAs with FGFR2 minigenes did not show significant changes in exon IIIb or exon IIIc splicing (data not shown). CUGBP2 displayed some silencing of exon IIIc, but this effect was less robust than other factors shown to silence exon IIIc, such as Fox-2, hnRNP M, and hnRNP H1 (data not shown). BOLL expression is largely restricted to testes and was thus felt unlikely to be generally required for FGFR2 spicing regulation.

Figure S2. Amino acid sequence of mouse Esrp1 (Rbm35a).
Amino acid sequence of the longest mouse Refseq Esrp1 protein (NP_918944). The three RRM domains are underlined. The dotted box indicates N-terminal amino acids missing in the MGC clone used in the screen (BC031468.1), but which are present in the cDNAs used in most of the subsequent figures. The solid boxes indicate amino acid sequences whose presence or absence is determined by alternative splicing. Inclusion of amino acids CKLP is determined by alternative 5' splice site use at the end of exon 12. The boxed sequences at the C-terminus of the protein are encoded by alternative exons 14 and 15. A full length mouse Rbm35a cDNA (Esrp1-2A) was assembled from existing cDNA clones and corresponds to the single mouse NCBI refseq transcript (NM_194055.2) except that it results from use of the proximal 5' splice site in exon and thus does not encode a protein containing CKLP. A second Rbm35a cDNA was constructed in which both alternative exons 14 and 15 are not represented (Esrp1-NA). The asterisk denotes a Valine encoded after S607 and immediately prior to the stop codon in the shorter protein encoded by splice variants that skip exons 14 and 15.

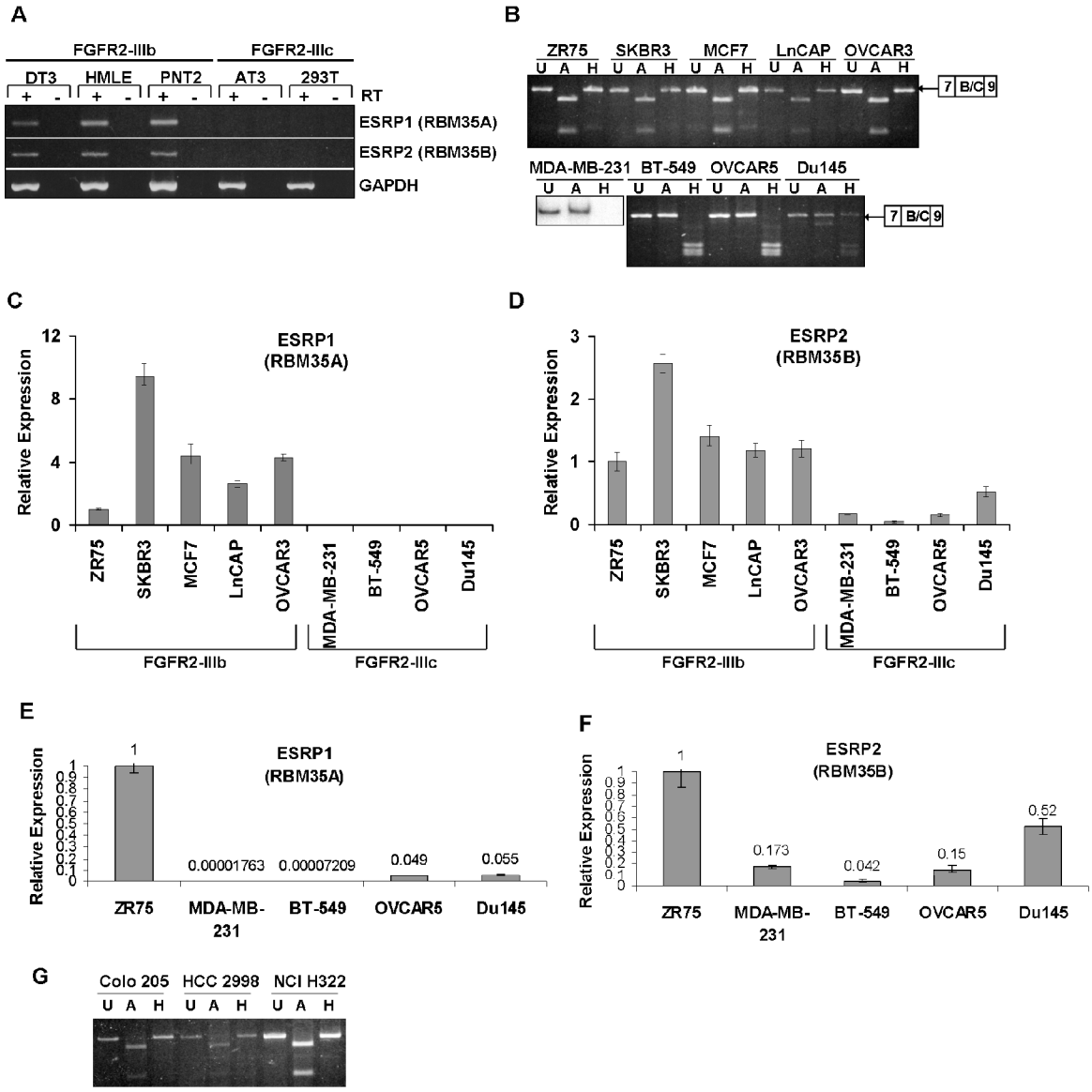


Figure S3. RBM35A and RBM35B are cell type-specific splicing regulators with mRNA expression limited to cell that express the epithelial FGFR2-IIIb isoform.

(A) Semiquantitative RT-PCR analysis of RBM35A (ESRP1) and RBM35B (ESRP2) expression in DT3, HMLE, PNT2, AT3, and 293T cell lines. Expression of the ESRP mRNAs is limited to cells expressing FGFR2-IIIb. Expression of GAPDH was used as a loading control.

(B) Splicing analysis of endogenous FGFR2 transcripts in a panel of human cancer cell lines. MDA-MB 231 cells expressed minute quantities of FGFR2-IIIc transcripts that required radiolabelling during PCR for visualization. FGFR2 splice variant analysis is as described in the legend to Figure 1E.

(C) The expression values of ESRP1 relative to ZR75 cells in the panel of cells shown in (C) by qRT-PCR.

(D) Relative ESRP2 expression as described for (D).

(E) Due to the lower expression of ZR75 ESRP1 mRNA relative to the high levels in some of the other FGFR2-IIIb expressing cell line, a subset of the same data in Figure S2C is shown to indicate that ZR75 mRNA levels of RBM35A are still significantly higher than those in the FGFR2-IIIc expressing cell lines.

(F) Relative mRNA levels of ESRP2 in ZR75 cells relative to the FGFR2-IIIc expressing cell lines using a subset of the same data shown in Figure S2D. Note that expression of mRNAs for either RBM35A or RBM35B was very low to undetectable in all FGFR2-IIIc expressing cell lines tested except for in DU-145, which had modest RBM35B expression that corresponded with a small fraction of exon IIIb containing transcripts.

(G) Demonstration that three additional NCI60 cell lines with high mRNA levels of ESRP1 and/or ESRP2 based on microarray analysis (Figure S4) express FGFR2-IIIb.

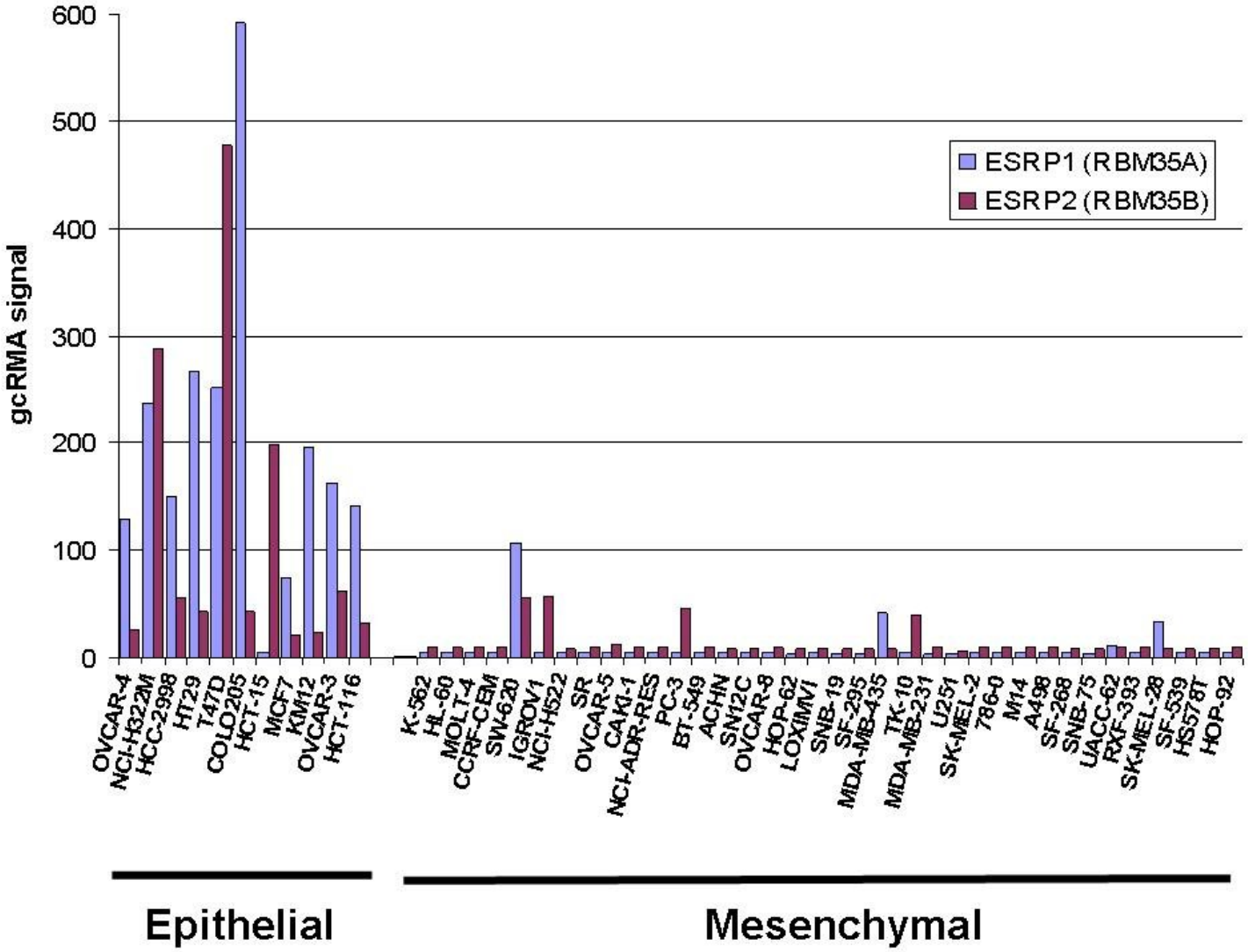


Figure S4. ESRP1 and ESRP2 expression is specifically observed in NCI60 cell lines classified as "epithelial".

Microarray data for ESRP1 and ESRP2 mRNA expression in cell lines that were classified as epithelial or mesenchymal based on the E-Cadherin/Vimentin ratio as described (Park et al., 2008). The gene profiling dataset of the NCI60 cell using the Affymetrix U133A and U133B was performed as described and was downloaded from the Cellminer website http://discover.nci.nih.gov/cellminer (Shankavaram et al., 2007). Shown are gcRMA normalized signal intensity in each cell line using probesets 219121_s_at (ESRP1/RBM35A) and 219395_at (ESRP2/RBM35B). The colon carcinoma cell line SW620 was the only "mesenchymal" cell line with significant expression of both ESRP1 and ESRP2. However, this classification might be called into question as it displayed very low protein levels for both E-Cadherin and Vimentin and furthermore showed significant expression of epithelial-specific miR-200 family members (Park et al., 2008).

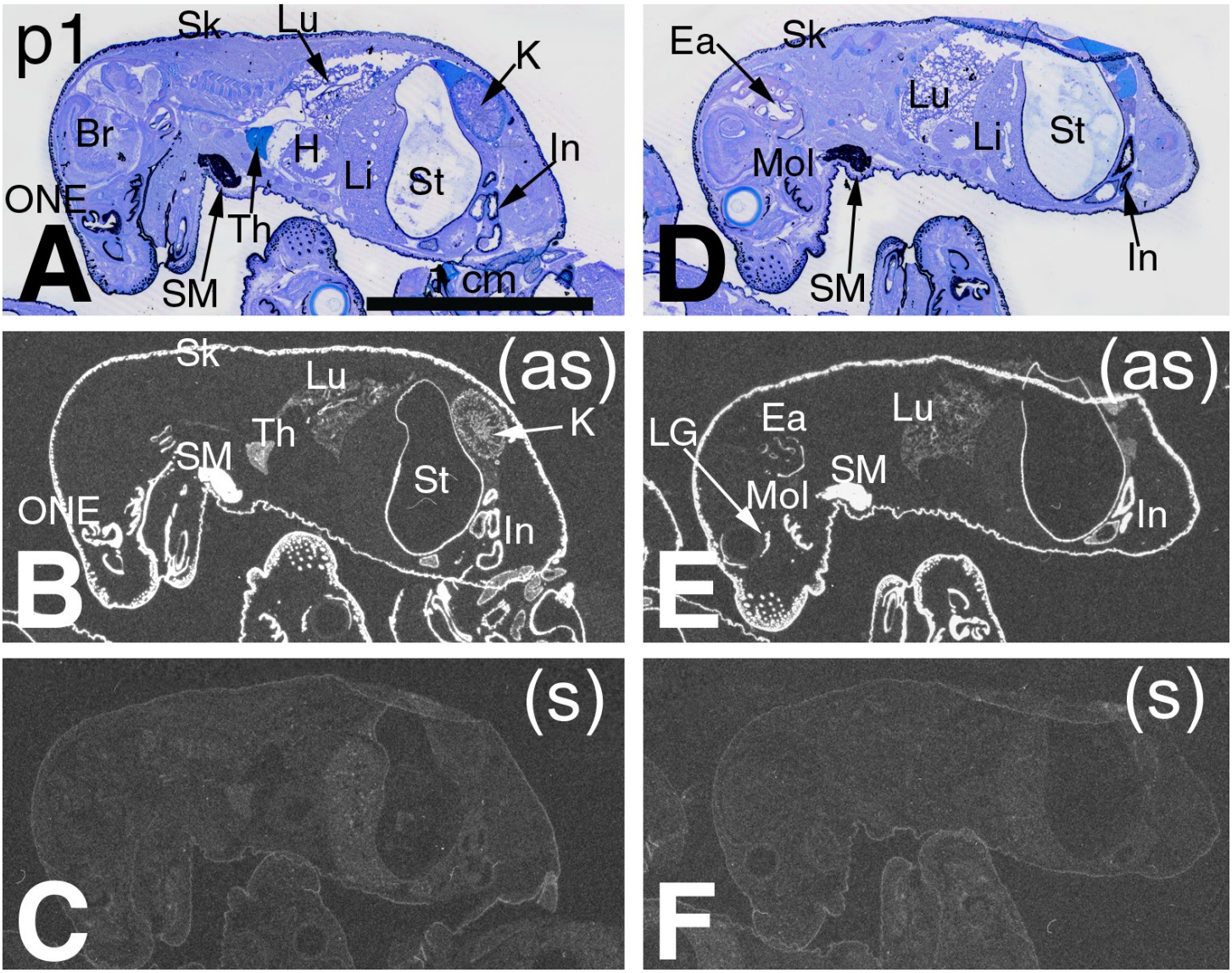


Figure S5. Rmb35a expression in postnatal mouse (day 1)
(A, D) Whole-body sections of a postnatal mouse following staining with cresyl violet. A medial section (A) and a more lateral section in the plane of the eye (D) are shown.
(B, E) X-ray film autoradiography detection of Rmb35 mRNA, seen as bright labeling. High-level labeling is observed in the skin, salivary glands, molars, stomach and intestine. Moderate-level labeling is seen in the kidney, lung and thymus.
(C, F) Control (sense) hybridization in an adjacent section comparable to (B and E).
Abbreviations: Br – brain; Ea – ear; H – heart; K – kidney; In – intestine; Li – liver; LG – lacrimal gland; Lu – lung; Mol – molar; ONE – olfactory neuroepithelium; Sk – skin; SM – submaxillary gland; St – stomach; Th – thymus; (as) – antisense; (s) – sense

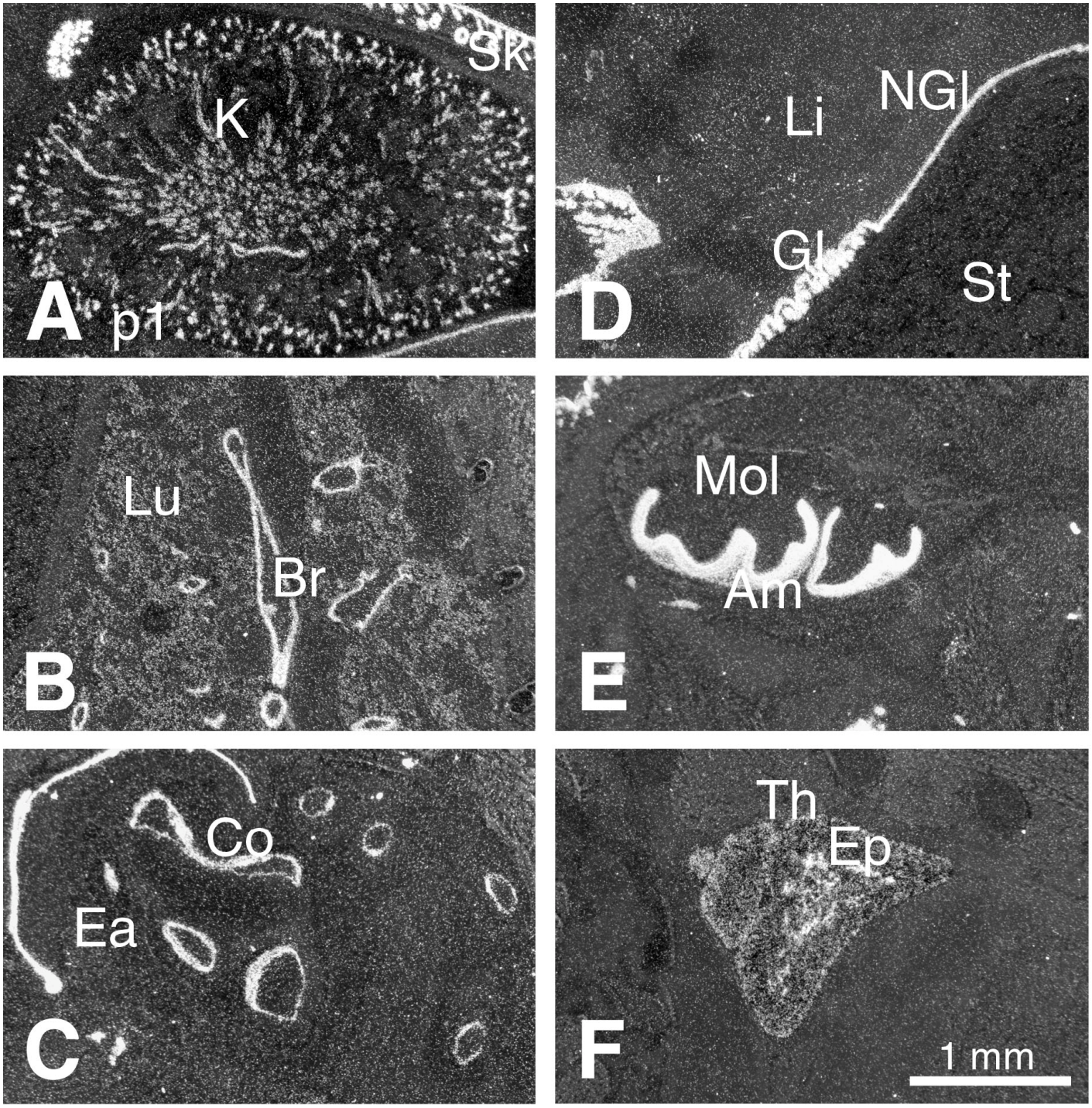


Figure S6. Rmb35a expression in postnatal mouse (p1); partial survey. Emulsion autoradiography detection of Rbm35a mRNA in postnatal mouse tissues, seen as bright labeling under darkfield illumination.

(A) Kidney tubules
(B) Lung bronchi
(C) Ear organ of Corti
(D) Stomach wall; glandular and non-glandular regions
(E) Molar ameloblasts
(F) Thymus epithelial cells

Abbreviations: Am – ameloblasts; Br – bronchus; Co – organ of Corti; Ea – ear; Ep – epithelium; K – kidney; Gl – glandular region of the stomach wall; Li – liver; Lu – lung; Mol – molar; NGl – non-glandular region of the stomach wall; Sk – skin; St – stomach; Th – thymus

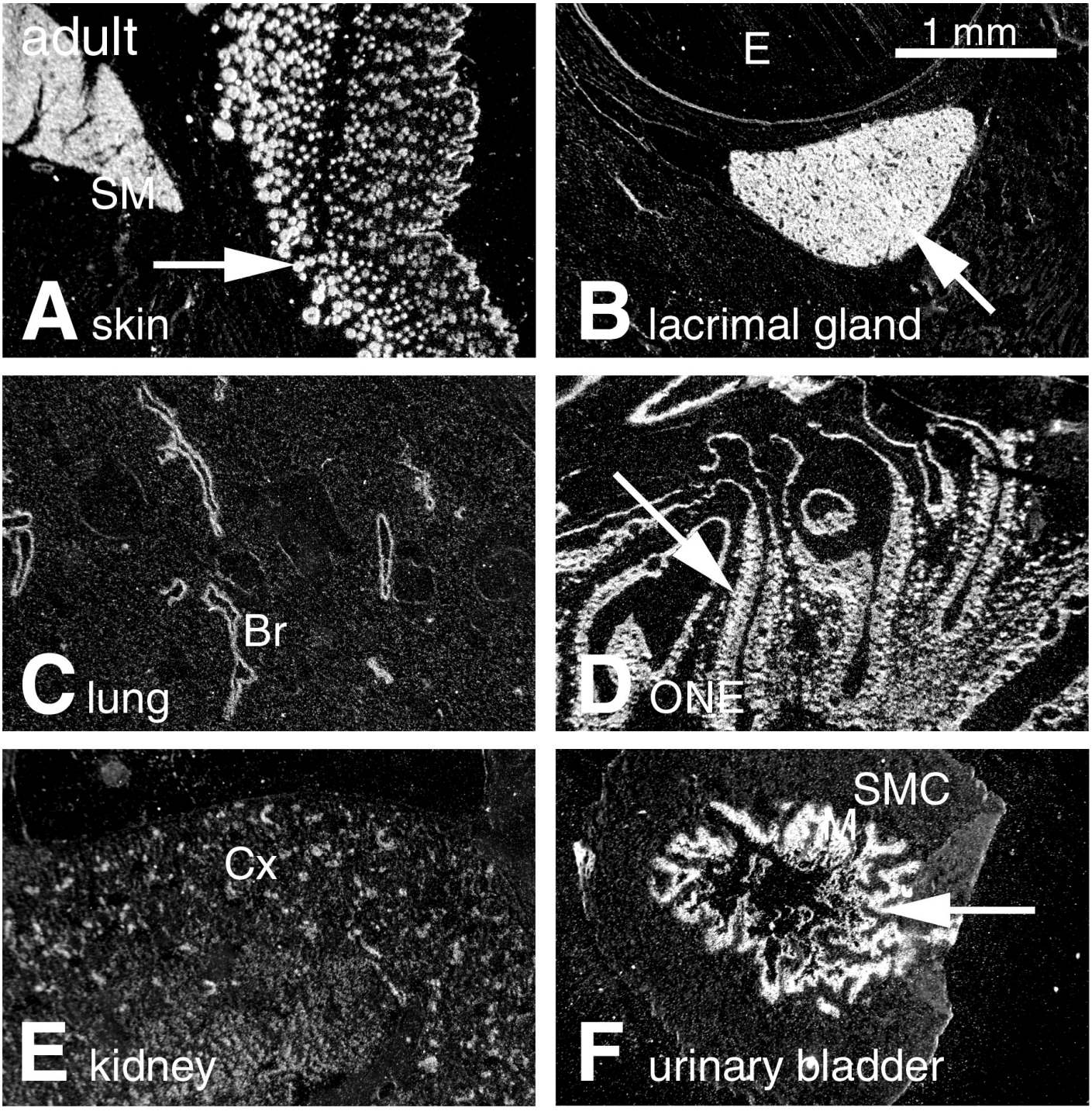


Figure S7. Rmb35a expression in adult mouse tissues; partial survey
(A) Skin with hair bulb labeled (arrow)
(B) Lacrimal gland with epithelial cells labeled (arrow)
(C) Lung with bronchi labeled
(D) Olfactory turbinates with olfactory neuroepithelium labeled (arrow)
(E) Kidney with labeled tubules in the kidney cortex (arrow)
(F) Urinary bladder with labeled epithelial cells (arrow) and unlabeled smooth muscle cells
Abbreviations: Br – bronchi; Cx – cortex, kidney; E – eye; ONE – olfactory neuroepithelium; SM – submaxillary gland; SMC – smooth muscle cells; (as) – antisense; (s) – sense

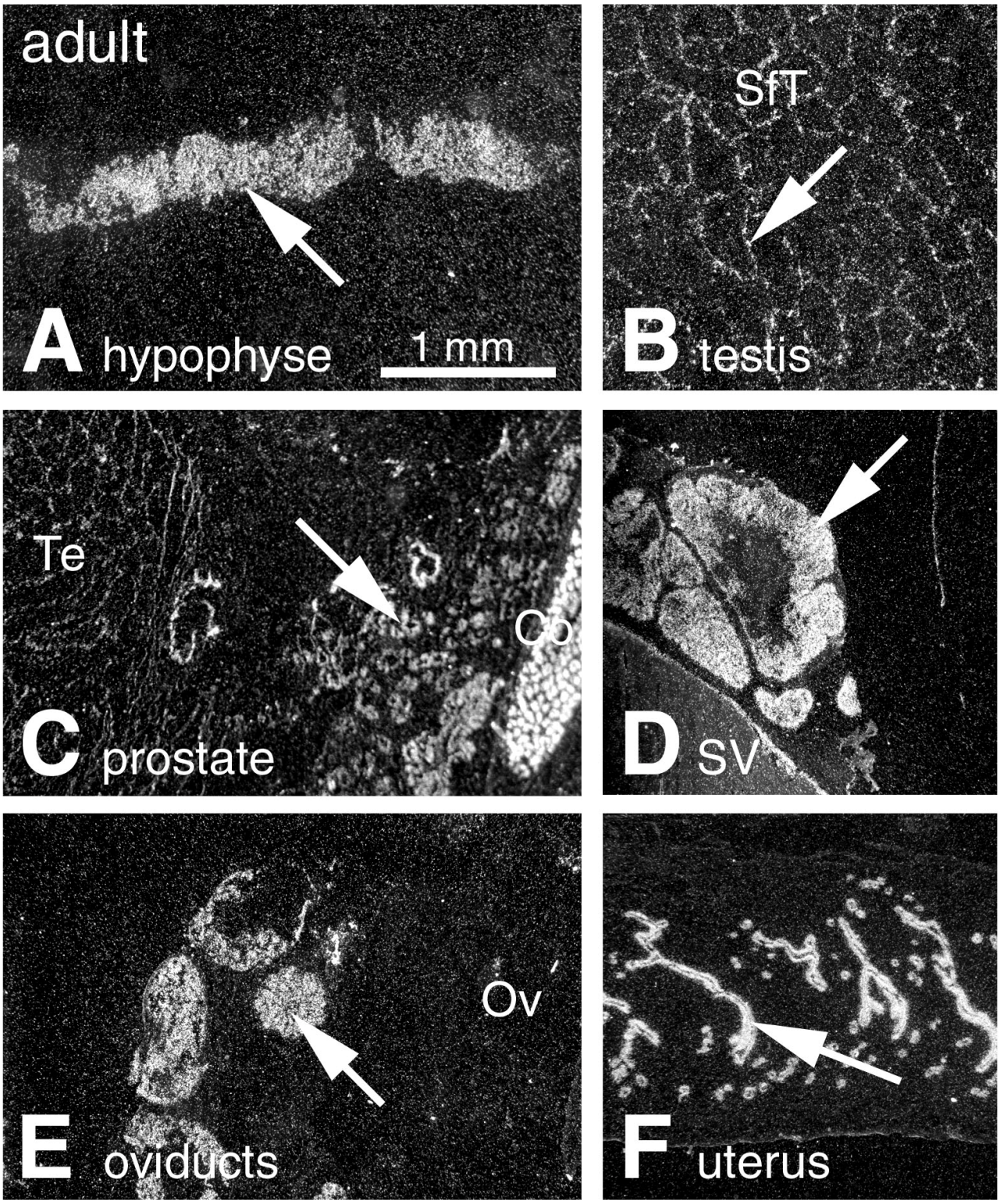


Figure S8. Rmb35a expression in adult mouse tissues; partial survey
(A) Pituitary gland with labeled adenohypophyseal cells (arrow)
(B) Testis with labeling observed around the seminiferous tubules (arrow)
(C) Prostate with labeled epithelial cells (arrow)
(D) Seminal vesicles with labeled epithelial cells (arrow)
(E) Oviducts with labeled epithelial cells (arrow); note the unlabeled ovary
(F) Uterus with labeled epithelial cells (arrow)
Abbreviations: Co – colon; Ov - ovary; SfT – seminiferous tubule; SV – seminal vesicle; Te – testis; (as) – antisense; (s) – sense

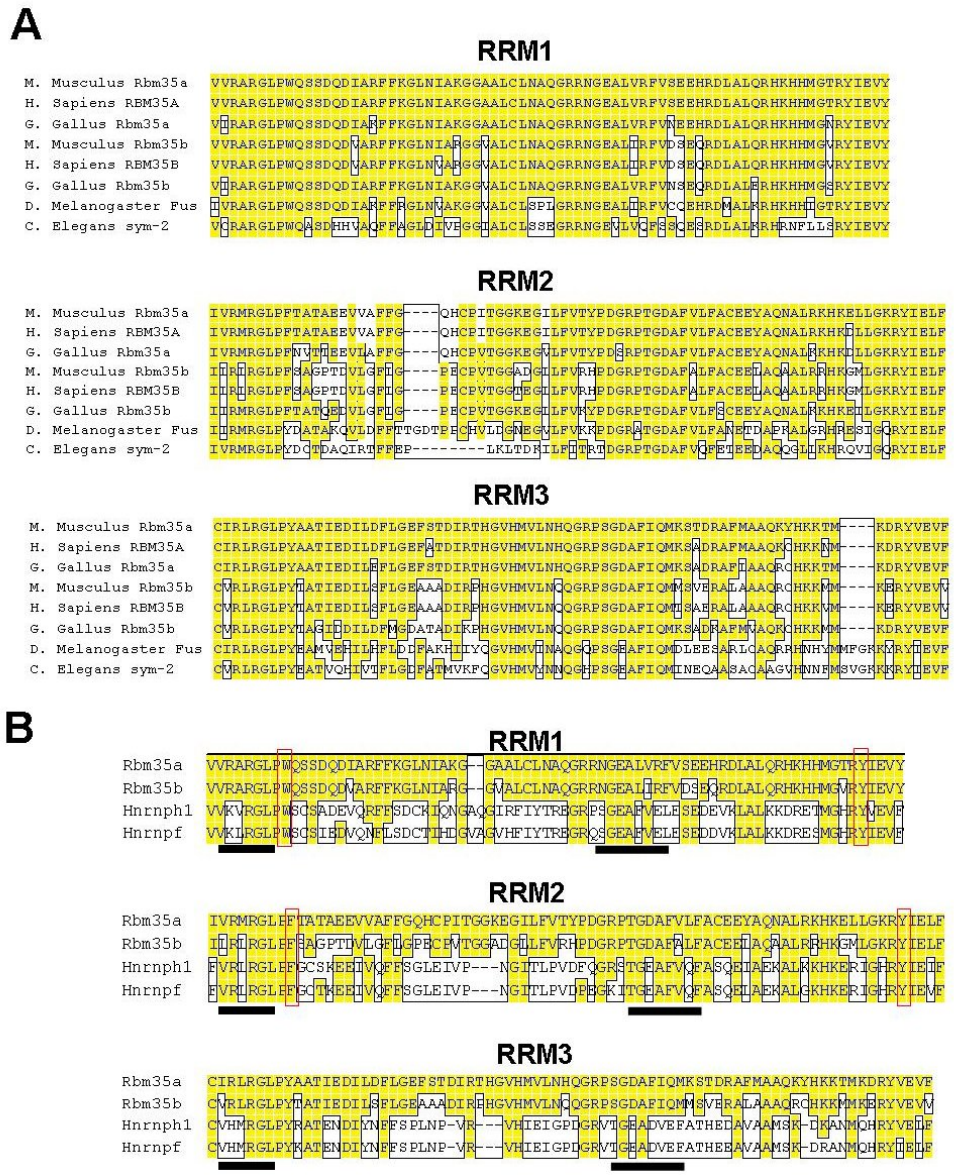


Figure S9. ESRP1 and ESRP2 are evolutionarily conserved RNA binding proteins with homology to the hnRNP F/H family.

(A) Clustal W (DNASTAR) alignment of the RRM domains of the vertebrate ESRPs and the *D. melanogaster* and *C. elegans* orthologs. Conserved amino acids are shaded in yellow. Black boxes indicate amino acid residues that are not identical to mouse Esrp1. The accession numbers of the sequences represented are NP_918944 (Mm Esrp1), NP_789808 (Mm Esrp2), NP_060167 (Hs ESRP1), NP_524691 (Dm Fusilli), NP_495960 (Ce Sym-2) XP_418338 (Gg Esrp1), NP_001025737 (Gg Esrp2). The boundaries of the RRM domains shown are as defined by SMART (http://smart.embl-heidelberg.de/).

(B) Alignment of the RRM domains of mouse Rbm35a (Esrp1) and Rbm35b (Esrp2) with those of mouse hnrnph1 (hnRNP H; NP_067485) and hnrnpf (RNP F; NP_598595). The short and long underlined sequences indicate RNP-2 and RNP-1 domains, respectively. The conserved amino acids in the red boxes indicate residues in hnRNP H and hnRNP F that have been shown to be required for RNA binding and/or splicing activity (Dominguez and Allain, 2005; Mauger et al., 2008)

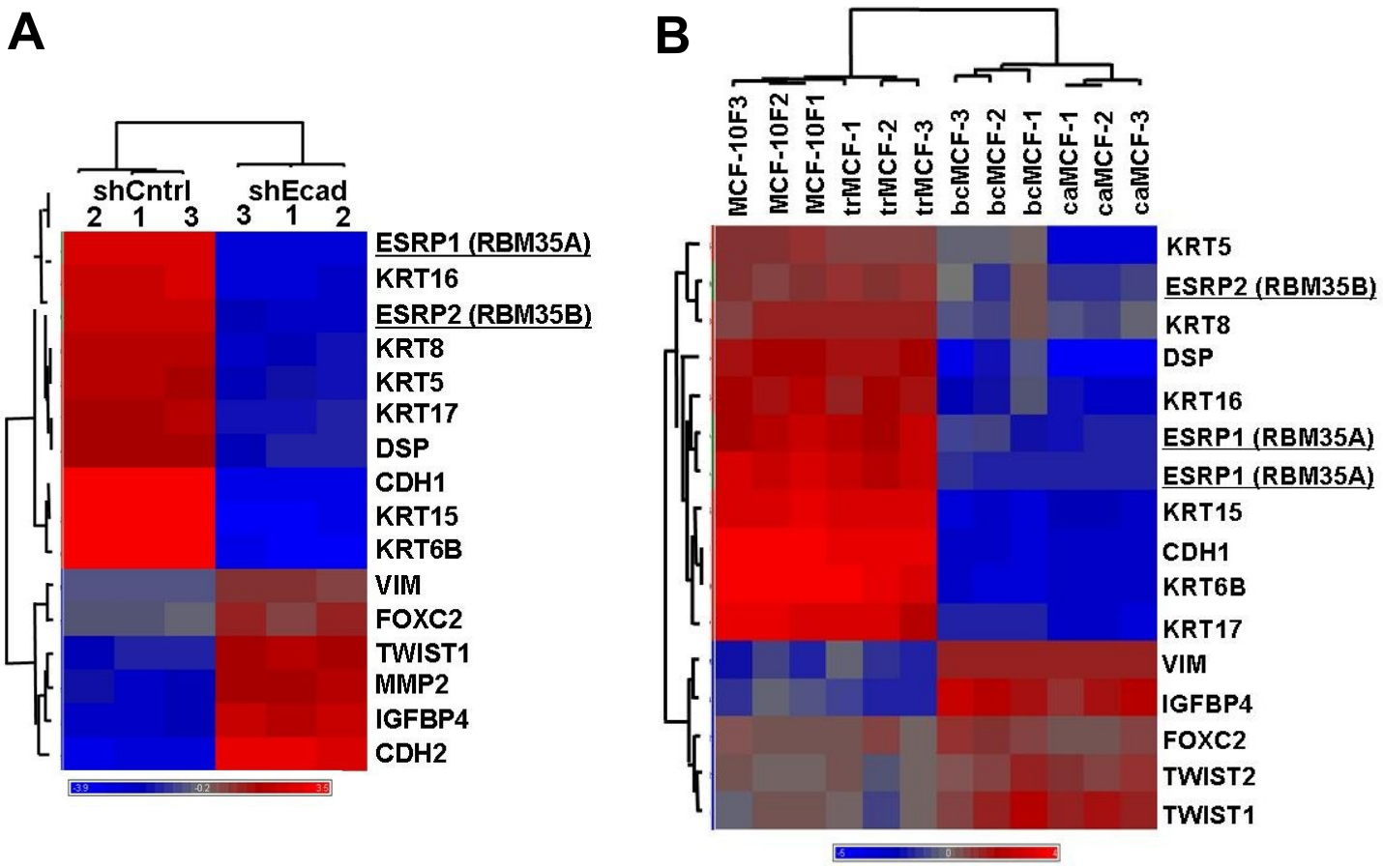


Figure S10. ESRP1 and ESRP2 are components of an epithelial-specific gene expression signature that is lost during the EMT.

(A) Hierarchical gene cluster analysis of selected epithelial and mesenchymal markers in HMLE cells undergoing an EMT following treatment with shRNAs leading to loss of E-Cadherin (CDH1) (shEcad) compared to control shRNAs (shCntrl) (Onder et al., 2008). Triplicate data are shown and gene symbols for downregulated genes (blue) and upregulated genes (red) are indicated.

(B) Gene cluster analysis of epithelial and mesenchymal markers in an EMT observed in 17β-estradiol (E_2) transformed human MCF10F breast cancer cells (Huang et al., 2007). Microarray analysis of the respective cells was performed using Affymetrix U133 Plus 2.0 Arrays. MCF10F and E_2 transformed cells (trMCF) represent three separate passages of each cell type. bcMCF data represent three different clones of transformed cells that were selected for invasiveness in matrigel. caMCF cells represent invasive cells that were furthermore capable of tumor formation in SCID mice. Gene symbols for downregulated genes (blue) and upregulated genes (red) are shown at right.

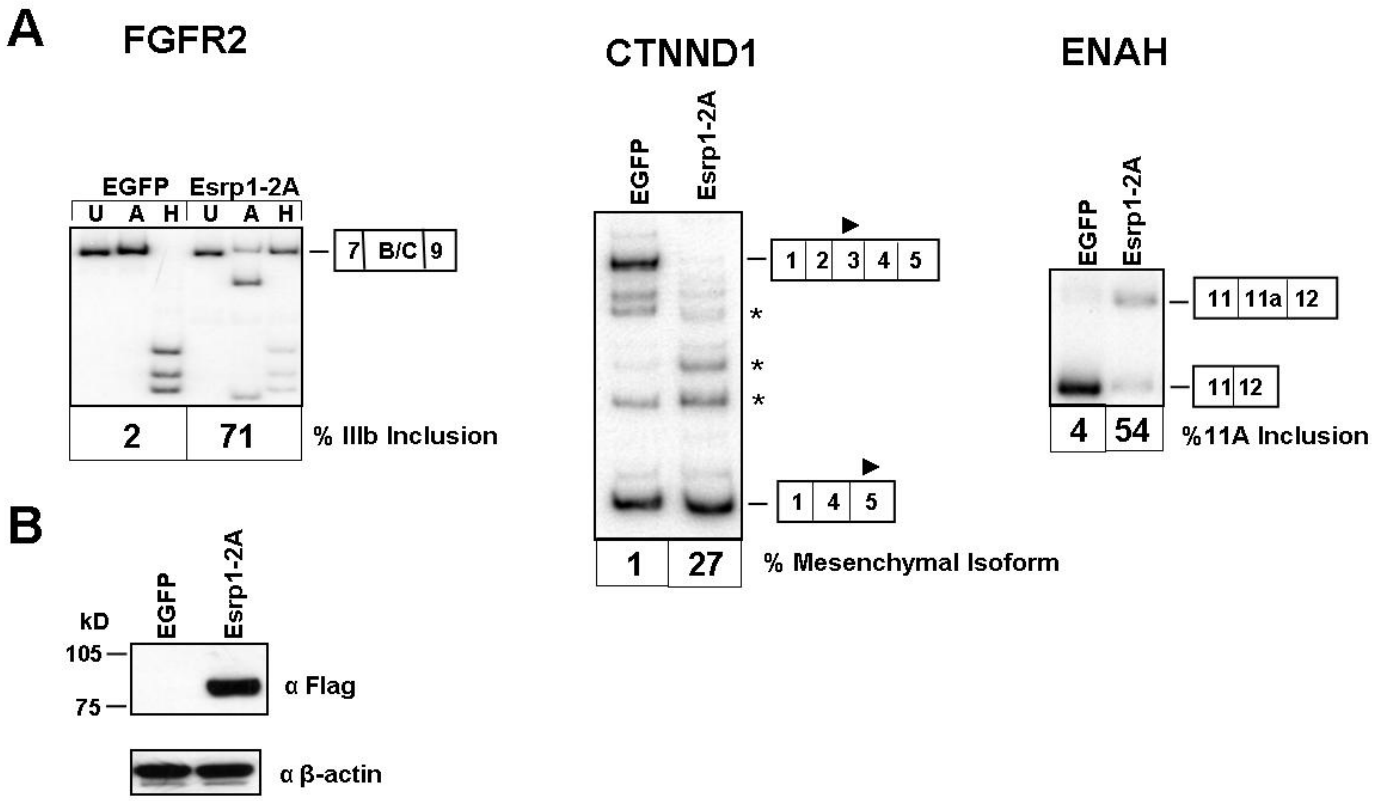


Figure S11. Ectopic expression of Esrp1 in mesenchymal MDA-MB-435 cells causes a switch towards an epithelial splicing pathway.
(A) Splicing analysis of FGFR2, CTNND1, and ENAH in MDA-MB-435 cells transduced with control cDNAs for EGFP and Esrp1. Analysis of CD44 in control cells showed predominant expression of the standard (CD44s) isoform, but in this case a significant downregulation of CD44 expression precluded simple analysis of changes in splicing (data not shown).
(B) Expression of FLAG-tagged Esrp1 in as determined by immunoblotting with anti-FLAG antibodies. An immunoblot for β-actin is used as a loading control.

	Gene Symbol	Gene Title	r	p-value (correlation)	Probeset ID
1	SPINT1	serine peptidase inhibitor, Kunitz type 1	0.828404	5.66E-16	202826_at
2	STARD10	StAR-related lipid transfer (START) domain containing 10	0.794743	5.72E-14	223103_at
3	**RBM35A**	RNA binding motif protein 35A	0.788257	1.26E-13	225846_at
4	CLDN7	claudin 7	0.788238	1.26E-13	202790_at
5	**RBM35A**	RNA binding motif protein 35A	0.78733	1.41E-13	219121_s_at
6	RAB25	RAB25, member RAS oncogene family	0.78085	3.01E-13	218186_at
7	STARD10	StAR-related lipid transfer (START) domain containing 10	0.779361	3.57E-13	232322_x_at
8	MARVELD3	MARVEL domain containing 3	0.778058	4.15E-13	239148_at
9	QKI	quaking homolog, KH domain RNA binding (mouse)	-0.76907	1.12E-12	212263_at
10	TJP3	tight junction protein 3 (zona occludens 3)	0.767045	1.40E-12	35148_at
11	FLJ23044	CDNA: FLJ23044 fis, clone LNG02454	0.764283	1.88E-12	226185_at
12	VIM	vimentin	-0.76286	2.18E-12	201426_s_at
13	C1orf172	chromosome 1 open reading frame 172	0.761817	2.44E-12	236058_at
14	MSN	moesin	-0.75324	5.88E-12	200600_at
15	CXCL16	chemokine (C-X-C motif) ligand 16	0.739818	2.17E-11	223454_at
16	CGN	cingulin	0.739654	2.21E-11	223233_s_at
17	TMEM125	transmembrane protein 125	0.73872	2.41E-11	225822_at
18	ST14	suppression of tumorigenicity 14 (colon carcinoma)	0.733554	3.89E-11	216905_s_at
19	QKI	quaking homolog, KH domain RNA binding (mouse)	-0.72845	6.19E-11	212262_at
20	GRTP1	growth hormone regulated TBC protein 1	0.727606	6.67E-11	229377_at
21	TJP3	tight junction protein 3 (zona occludens 3)	0.72689	7.11E-11	213412_at
22	ST14	suppression of tumorigenicity 14 (colon carcinoma)	0.724814	8.55E-11	202005_at
23	LLGL2	lethal giant larvae homolog 2 (Drosophila)	0.723714	9.42E-11	203713_s_at
24	MANSC1	MANSC domain containing 1	0.723113	9.93E-11	220945_x_at
25	CDS1	CDP-diacylglycerol synthase 1	0.723059	9.98E-11	205709_s_at
26	STAP2	signal transducing adaptor family member 2	0.721444	1.15E-10	221610_s_at
27	HOOK2	hook homolog 2 (Drosophila)	0.719269	1.39E-10	218780_at
28	RBM47	RNA binding motif protein 47	0.714585	2.07E-10	222496_s_at
29	F11R	F11 receptor	0.713529	2.26E-10	224097_s_at
30	SACS	spastic ataxia of Charlevoix-Saguenay (sacsin)	-0.71053	2.91E-10	213262_at
31	DDR1	discoidin domain receptor tyrosine kinase 1	0.708789	3.36E-10	210749_x_at
32	LIX1L	Lix1 homolog (mouse)-like	-0.70868	3.39E-10	225793_at
33	SH3YL1	SH3 domain containing, Ysc84-like 1 (S. cerevisiae)	0.707711	3.67E-10	204019_s_at
34	SLC37A1	solute carrier family 37 , member 1	0.707268	3.81E-10	218928_s_at
35	ARRDC1	arrestin domain containing 1	0.705195	4.51E-10	226405_s_at
36	QKI	quaking homolog, KH domain RNA binding (mouse)	-0.70346	5.19E-10	212265_at
37	S100A14	S100 calcium binding protein A14	0.703324	5.24E-10	218677_at
38	GOLT1A	golgi transport 1 homolog A (S. cerevisiae)	0.697095	8.61E-10	229372_at
39	GRHL2	grainyhead-like 2 (Drosophila)	0.689799	1.52E-09	219388_at
40	DDR1	discoidin domain receptor tyrosine kinase 1	0.6883	1.70E-09	1007_s_at
41	BSPRY	B-box and SPRY domain containing	0.687102	1.86E-09	218792_s_at
42	MPZL2	myelin protein zero-like 2	0.684946	2.19E-09	203780_at
43	CLDN3	claudin 3	0.684686	2.23E-09	203953_s_at
44	**RBM35B**	RNA binding motif protein 35B	0.68414	2.32E-09	219395_at
45	BSPRY	B-box and SPRY domain containing	0.683773	2.39E-09	222746_s_at
46	DDR1	discoidin domain receptor tyrosine kinase 1	0.683324	2.47E-09	207169_x_at
47	MARVELD2	MARVEL domain containing 2	0.678766	3.46E-09	235141_at
48		Transcribed locus	0.677361	3.83E-09	229223_at
49	EHF	Ets homologous factor	0.67702	3.93E-09	225645_at
50	MYH14	myosin, heavy chain 14	0.676855	3.97E-09	226988_s_at

Table S1. Correlation of mRNA expression with E-cadherin across the NCI60 panel of tumor cell lines ranked by p-value. ESRP1 and ESRP2 expression across NCI60 cell lines is highly correlated with the epithelial cell marker E-Cadherin. NCI60 GCRMA expression values were $\log_2$ transformed and a Pearson Correlation was performed between E-Cadherin (CDH1; 201131_s_at) and all other probesets from U133A and U133B $\log_2$ transformed microarray data (www.Cellminer.com). Genes with both positive and negative correlations with E-cadherin expression are shown and were ranked together by correlation p-value. The two probesets for RBM35A (ESRP1) and single probeset for RBM35B (ESRP2) are indicated in bold red type.

Supplementary Experimental Procedures

High throughput cDNA retrotransfection screening

High throughput screening using the latest collection of the MGC clone collection of approximately 15,000 full length human and mouse cDNAs in pCMV-Sport6 was performed as described with several modifications (Chanda et al., 2003). A mix of 0.12 µl of Transit 293 transfection reagent (Mirus) and 20 µl of serum free media was applied using a Wellmate handler (Matrix) to each well of a 384 well plate containing 40 ng of individual cDNAs. After 30-60 minutes, 1000 cells of a 293T cell clone stably expressing the luciferase splicing reporter were mixed with DMEM containing 20% FBS and applied to each well. Each plate also has wells containing DNA of the empty pCMV-Sport6, as well as 4 wells containing cDNAs encoding mFox-1 and Tia-1 proteins as positive controls. 48-60 hours after retrotransfection, 40 µl of BriteLite luciferin reagent (Perkin Elmer) was added to each well and luminescence determined using an Analyst HT multimode plate reader (Molecular Devices).

Plasmid construction

Bicistronic expression plasmids pIPX and pINX were derived from the previously described pIRESpuro3ΔInt and pIRESneo3 ΔInt , respectively, by removing a 5' portion of the MCS with Eco RV and Age I and insertion of annealed oligos ESXA-F and ESXA-R (Newman et al., 2006). A bicistronic plasmid containing the blasticidin resistance gene (BSD), pIRESblast ΔInt, was constructed by amplification of the BSD coding sequence with primers Blasticidin-F and Blasticidin R and insertion of the resulting insert in place of the neomycin resistance cassette in Xma-I and Xba I digested pIRESneo3 ΔInt. The related pIRESblast2 ΔInt was similarly constructed except that the BSD coding sequence was amplified with Blasticidin-F2 as the forward primer. The pIBX plasmid was derived as described for pIPX and pINX. The C-terminal FLAG tag vectors pIPX-C-FF-C, pINX-C-FF-B, and pIBX-C-FF-B that contain two tandem copies of the FLAG tag were generated by insertion of annealed oligos C-FF-B-anneal-F and C-FF-B-anneal-R downstream of the MCS in the vectors described above. The pINX-C-FF-ZZ-B was generated by amplification of pIP-WT1-FF-ZZ with primers FF-C-B-F and FF-ZZ-Nsi-R. The resulting PCR product carrying tandem C-terminal FLAG and ZZ tags separated by a TEV cleavage site was inserted into the Not I and Nsi I sites in pINX (Geng and Carstens, 2006). C-terminal FLAG tagged vectors containing cDNAs for Esrp1, Esrp2, and Fusilli (Isoform D) were constructed by amplification of the respective full length cDNAs with primers ESRP-FL-Stu-F and 35a-clone-R, 35b-clone-F and 35b-clone-R, and Fus 5' Koz EcoR1-F and Fus 3' NotI-R, respectively. The cDNAs for Esrp1 and Esrp2 were inserted in Stu I and Not I or EcoRV and Not I digested pIBX-C-FF-B, respectively. The resulting cDNA construct for Esrp1 (pIBX-Esrp1-FF-NA) represented the splice variant that did not contain either of the alternative exons 14 and 15 and was amplified from cDNA clone BC031468. An Esrp1 splice variant that contained both exons 14 and 15 (pIBX-Esrp1-FF-2A) was constructed by PCR amplification from Riken clone AF018848 with primers Rbm35a alt exons-F and Esrp-FL-Not-R. The PCR product containing the 3' end of the coding sequence was digested with PshA1 and Not I and used to replace the same region in the Esrp1-NA plasmid. The cDNA for fusilli was inserted into EcoR1 and Not I digested pIBX-C-FF-B to create pIBX-Fusilli-FF. The cDNA for Fox-1 was previously described (Jin et al., 2003). The pIRESpuro-D-GP plasmid was created by insertion of an amplified GAG-POL insert that contained a Kozak consensus site at the 5' end from pMD-gag-pol (kindly provided by S. Ananth Karumanchi) into the Nhe I and EcoR1 sites in pIRESpuro3 ΔInt.

To construct the retroviral pMXs-IRES-Puro vector, pMXs (kindly provided by Toshio Kitamura, Kyoto University) was digested with Sal I, blunted with Pfu polymerase, and then digested with Not I (Kitamura et al., 2003). An insert containing the IRES and puro resistance coding sequence was digested with Xba I and blunted, followed by Not I digestion and ligation into pMXs. To generate pMXsIRESblast2 a cassette containing the IRES and BSD CDS was amplified with primers IRES-Resis-Not-F and IRES-Resis-Sal-R and inserted into the same sites in pMXs. Retroviral vectors for expression of FLAG tagged Esrp1 and Esrp2 were generated by digestion of pMXs-IRES-puro and pMXs-IRESblast2 with EcoR1, blunting with Pfu and Nsi I digestion. A Stu I-Nsi insert from pIBX-Esrp1-NA-FF or pIBX-Esrp1-2A-FF was inserted into the respective

vector to create the respective constructs for Esrp1. An Eco RV-Nsi insert from pIBX-Esrp2-FF was inserted for the Esrp2 retroviral constructs. Similarly both retroviral vectors were digested with EcoR1 (unblunted) and Nsi I for the insertion of the FLAG-tagged Fusilli cDNA described above.

Minigene PKC-neg-40B-IF3-Luc-Puro was derived from PKC-neg-40B-IF3-EGFP-Puro (Newman et al., 2006) by removal of the EGFP CDS with Pac I and Not I and insertion of a firefly luciferase CDS amplified from pGL4 with primers Luc-F and Luc-R. PKC-neg-40B-EGFP-Hygro was also previously described and PKC-neg-40B-IF5-mRFP-puro was derived through insertion of IF5 downstream of the test exon as described (Hovhannisyan and Carstens, 2005).

Plasmids used for in situ hybridization (ISH) were constructed by the insertion of Hind III and Eco R1 digested PCR products amplified from a cDNA for Rbm35a and inserted into the respective sites in the pDP19 vector (Ambion). Antisense strand RNAs were generated as runoff T7 transcripts from Hind III digested plasmids. Sense strand RNA controls were generated as T3 runoff transcripts from EcoR1 digested plasmids. The plasmid used for ^{S}35-labelled probes for Rbm35a, pDP19-probe-35a-M, was generated using PCR products using primers Rbm35a-Phyl-in situ H3-F and Rbm35a-Phyl-in situ R1-R.

Additional cell culture and transfection

DU-145 and KATO III cells were maintained in DMEM with 5-10% FBS. ZR75, and LNCaP cells were maintained in RPMI1640 with 10% serum. SKBR3 were maintained in DMEM with 10% FBS and 10 µg/ml insulin. MDA-MB-231, MCF7, BT549, Colo205, MDA-MB-435, NCC2998, and NCI H322 were maintained in RPMI1640 with 10% FBS and 10 µg/ml insulin. All cells other than 293T cells were transfected with Lipofectamine 2000 (Invitrogen) according the manufacturers' protocols. The 293T Clone 2 and Clone 3 cells stably expressing two different minigenes were established by sequential transfection and selection for stable clones in 1.0 µg/ml Puromycin (Sigma) or 400 µg/ml Hygromycin B (Roche). Single cell clones of each were obtained by serial dilution. The 293T-N16 packaging cell line expressing gag-pol from the pIRES-puro-D-GP plasmid was generated by transfection, stable selection in 1.0 µg/ml Puromycin and selection of a stable clone.

Antibodies

Antibodies used were anti-β-actin clone AC-15 (Sigma; 1:1000 dilution), anti-Vimentin Ab-2 Clone V9 (Thermo Scientific; 1:150 dilution), anti-E-Cadherin (BD Transduction Laboratories; 1:500 dilution), anti-RBM35A (anti-ESRP1 #1) MaxPab polyclonal antibody (Abnova;1:1,000), anti-RBM35A (anti-ESRP1 #2; Aviva Systems Biology; 1:1,000), and anti-RBM35B (Aviva Systems Biology; 1:1,000). Except where specified otherwise, anti-ESRP1 antibodies refer to the Abnova anti-RBM35A (anti-ESRP1 #1) antibodies.

.**UV-crosslinking, immunoprecipitation, nuclear extract preparation, and RNA affinity chromatography**
The constructs and protocol for UV crosslinking and immunoprecipitation were done largely as described (Hovhannisyan and Carstens, 2007). [^{32}P]UTP-radiolabelled RNAs were used at a specific activity of 2.7 X 10^{13} CPM/µmol and 500,000 CPM were incubated with a 20 µl binding reaction with 10-20 µg of nuclear extract and the addition of 4 ul of buffer D, 4 ul of 5X Binding Buffer (50 mM HEPES (pH 7.9), 10 mM MgCL2, 5 mM dithiothreitol, 325 mM KCl), and heparin (Sigma #H-3393) to a final 0.5 ug/ul. The RNA pull down assays were also done as described (Hovhannisyan and Carstens, 2007). KATO III cell nuclear extracts from large scale suspension cultures were prepared using a standard protocol (Dignam et al., 1983). Smaller scale nuclear extracts from monolayer cells were performed as described, with the following exceptions (Tsai and Carstens, 2006). Cell lysis was performed by passing cells 5-10 times through a $25^{3/8}$ gauge needle in the absence of detergents. Nuclear extracts were then dialyzed against buffer D (20 mM HEPES, pH 7.9, 20% (v/v) glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF) for 2-4 hours.

Microarray data cluster analysis
Gene cluster microarray data was downloaded from NCBI Geo Acc. # GSE9691 (Onder et al., 2008) and #GSE5116 (Huang et al., 2007) and processed using Partek Genomics Suite v. 6.3 using the RMA algorithm. The data was median shifted on a per gene basis such that the median signal across all samples was zero. For the selected gene markers we clustered the gene samples using Euclidean distance as the similarity

measure. The results of clustering are displayed with the color representing the median shifted signal of each sample.

Validation of cDNAs identified in the primary screen in independent transfections
Selected cDNAs identified in the high throughput screen and controls were transfected in triplicate in 293T luciferase reporter cells and luciferase activity was assayed as described for the primary screen except the setup was done manually and luciferase activity was determined using a DTX 880 Multimode Detector (Beckman Coulter).

In situ hybridization

In situ hybridization for Rbm35a was performed as a custom service by Phylogeny Inc. (Columbus, Ohio). C57BL/6 mice were sacrificed in a CO2 chamber according to GLP. Whole-bodies and individual tissues were flash frozen in isopentane and stored at -80°C. Frozen sections, 10 µm thick were mounted on gelatin-coated slides and stored at -80°C.Before ISH, sections were fixed in 4% formaldehyde (freshly made from paraformaldehyde; Sigma Aldrich P6148) in phosphate buffered saline (PBS), treated with triethanolamine/acetic anhydride, then washed and dehydrated in a series of ethanols. Prior to Rmb35a ISH, the tissues were validated with riboprobes to LDL receptor mRNA (data not shown). The cRNA transcripts were synthesized *in vitro* according to the manufacturer's recommended conditions (Ambion) and labeled with ^{35}S-UTP (>1000 Ci/mmol; Perkin-Elmer LAS Canada, Inc.). Rbm35a riboprobes were synthesized from the pDP-19-probe 35a-M DNA templates. Sections were hybridized overnight at 55 ˚C in 50% deionized formamide, 0.3 M NaCl, 20mM Tris-HCl pH 7.4, 5 mMEDTA, 10 mM NaH2PO4, 10% dextran sulphate, 1X Denhardt's, 50 µg/ml total yeast RNA, and 50-80,000 cpm/µl ^{35}SlabeledcRNA probe. The tissue was subjected to stringent washing at 65 ˚C in 50% formamide, 2X SSC, 10 mM DTT and washed in PBS before treatment with 20 µg/ml RNAse A at 37 ˚C for 30 minutes. Following washes in 2X SSC and 0.1XSSC for 10 minutes at 37 ˚C, the slides were dehydrated, exposed to Kodak BioMax MR x-ray film (3 days), then dipped in Kodak NTB nuclear track emulsion and exposed in light-tight boxes with desiccant at 4 ˚C (14 days). Photographic development was carried out in Kodak D-19. Slides were counterstained lightly with cresyl violet and analyzed using brightfield and darkfield microscopy. Sense (control) cRNA probes indicate the background levels of hybridization signal.

Oligonucleotide sequences used in this study

Luc-F	5'-TTAATTAATATCGAAGATGCCAAAAACATTAA-3'
Luc-R	5'-GCGGCCGCTTACACGGCGATCTTGCCGCCCT-3'
ESXA-F	5'-ATCAGGCCTCCTCGAGGGA-3'
ESXA-R	5'-CCGGTCCCTCGAGGAGGCCTGAT-3'
Blasticidin-F	5'-CCCCGGGATAATTCCTGCAGCCAATATGGCCAAGCCTTTG TCTCA-3'
Blasticidin-F2	5'-CCCCGGATCCATGGCCAAGCCTTTGTCTCA-3'
Blasticidin-R	5'-CTCTAGATTAGCCCTCCCACACATAAC-3'
C-FF-B-anneal-F	5'-GGCCGCCGATTACAAGGATGACGACGATAAGGCAGATT ACAAGGATGACGACGATAAGTAAATGCA-3'
C-FF-B-anneal-R	5'-TTTACTTATCGTCGTCATCCTTGTAATCTGCCTTAT CGTCGTCATCCTTGTAATCGGC-3'
ESRP-FL-Stu-F	5'-AGGCCTGCCACCATGACGGCGTCTCCGGATTA-3'
ESRP-FL-Not-R	5'- GGCGGCCGCAATACAAACCCATTCTTTGG-3'
35a-clone-R	5'-GCGGCCGCAACACTGGGATAGTAGGCTG-3'
35b-clone-F	5'-GATATCGCCACCATGACTCCGCCGCCGCCGCC-3'
35b-clone-R	5'-GCGGCCGCCAAACACACCCACTCCTTAG-3'
Fus 5' Koz EcoR1-F	5'-GGAATTCGCCACCATGCAGGTTCCCGAGCACGT-3'
Fus 3' NotI-R	5'-GGCGGCCGCGATTCCCGAATTATAGTACGA-3'
Rbm35a 3' alt exons-F	5'-TGGAGTGCACATGGTATTGA-3'
EGFP-R2	5'-GGTCAGGGTGGTCACGAGGG-3'

mRFP-R	5’-TTGGAGCCGTACTGGAACTGAGG-3’
hFGFR2-F	5’-TGGATCAAGCACGTGGAAAAGA-3’
hFGFR2-R	5’-GGCGATTAAGAAGACCCCTATGC-3’
IRES-Resis-Not-F	5’-GGCGGCCGCAGCATGCATCTAGGGCGGCC-3’
IRES-Resis-Sal-R	5’-GGTCGACGATTATGATCAGTTATCTAGA-3’
FF-C-B-F	5’-GCGGCCGCcGATTACAAGGATGACGACGATAAGGCA-3’
FF-ZZ-Nsi-R	5’-ATGCATTTAATTCGCGTCTACTTTCGGC-3’
Esrp1-F	5’-CAGAGGCACAAACATCACAT-3’
Esrp1-R	5’-AGAAACTGGGCTACCTCATTGG-3’
Esrp2-F	5’-TGGTGTGGCCCTCTGTCTCAAC-3’
Esrp2-R	5’-GCCCCCTGCAATCTTTACGAA-3’
ENAH-F	5’-TGCTGGCCAGGAGGAGAAGAAT-3’
ENAH-R	5’-ACTGGGCTGTGATAAGGGTGTGG-3’
CD44 5-15-F	5’-AGGAGCAGCACTTCAGGAGGTTAC-3’
CD44 5-15-R	5’-ACTGGGGTGGAATGTGTCTTGGTC-3’
hFOXC2-F	5’-ATCCGCCACAACCTCTCGCTCAAC-3’
hFOXC2-R	5’-GTGGGCCCGCTCCTCCTTCTCCTT-3’
hKRT15-F	5’-GGGGCGTTGGTGGGGGTTTTG-3’
hKRT15-R	5’-GCGCAGGGCCAGCTCATTCTCA-3’
hECAD-F	5’-CCGCCGGCGTCTGTAGGAAG-3’
hECAD-R	5’-GATTGGCAGGGCGGGGAAGATA-3’
hmNCAD-F	5’-GGATGAAACGCCGGGATAAAGA-3’
hmNCAD-R	5’-TGCTGCAGCTGGCTCAAGTCATAG-3’
GAPDH-F	5’-GGYCGTATTGGSCGCCTGGTCACC-3’
GAPDH-R	5’-GAGGGGCCATCCACAGTCTTCT-3’
Rbm35a-Phyl-in situ H3-F	5’- GAAGCTTAAGTGAGGAGCACAGAGATCTA-3’
Rbm35a-Phyl-in situ R1-R	5’- GGAATTCTAGGGAAGACCTCGCAGACGTA-3’

siRNA sequences

Sense strand:

ESRP1/RBM35A #3	5’-CACAAUGACAGAGUAUUUAAA-3’	(Qiagen)
ESRP1/RBM35A #4	5’- CCGGTATATTGAGGTTTACAA-3’	(Qiagen)
ESRP2/RBM35B #1	5’- AGCCCGAGGUGAUAAAGCAuu-3’	(Ambion)
ESRP2/RBM35B #2	5’- GACUUAAUCCUCCUAGUUUuu-3’	(Ambion)

Supplemental References

Chanda, S. K., White, S., Orth, A. P., Reisdorph, R., Miraglia, L., Thomas, R. S., DeJesus, P., Mason, D. E., Huang, Q., Vega, R.*, et al.* (2003). Genome-scale functional profiling of the mammalian AP-1 signaling pathway. Proc Natl Acad Sci U S A *100*, 12153-12158.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Dominguez, C., and Allain, F. H. (2005). Resonance assignments of the two N-terminal RNA recognition motifs (RRM) of the human heterogeneous nuclear ribonucleoprotein F (HnRNP F). J Biomol NMR *33*, 282.

Geng, J., and Carstens, R. P. (2006). Two methods for improved purification of full-length mammalian proteins that have poor expression and/or solubility using standard Escherichia coli procedures. Protein Expr Purif *48*, 142-150.

Hovhannisyan, R. H., and Carstens, R. P. (2005). A novel intronic cis element, ISE/ISS-3, regulates rat fibroblast growth factor receptor 2 splicing through activation of an upstream exon and repression of a downstream exon containing a noncanonical branch point sequence. Mol Cell Biol *25*, 250-263.

Hovhannisyan, R. H., and Carstens, R. P. (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.
Huang, Y., Fernandez, S. V., Goodwin, S., Russo, P. A., Russo, I. H., Sutter, T. R., and Russo, J. (2007). Epithelial to mesenchymal transition in human breast epithelial cells transformed by 17beta-estradiol. Cancer Res *67*, 11147-11157.
Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K., and Inoue, K. (2003). A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. Embo J *22*, 905-912.
Kitamura, T., Koshino, Y., Shibata, F., Oki, T., Nakajima, H., Nosaka, T., and Kumagai, H. (2003). Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. Exp Hematol *31*, 1007-1014.
Ladd, A. N., Charlet, N., and Cooper, T. A. (2001). The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol Cell Biol *21*, 1285-1296.
Mauger, D. M., Lin, C., and Garcia-Blanco, M. A. (2008). hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. Mol Cell Biol *28*, 5403-5419.
Newman, E. A., Muh, S. J., Hovhannisyan, R. H., Warzecha, C. C., Jones, R. B., McKeehan, W. L., and Carstens, R. P. (2006). Identification of RNA-binding proteins that regulate FGFR2 splicing through the use of sensitive and specific dual color fluorescence minigene assays. Rna *12*, 1129-1141.
Onder, T. T., Gupta, P. B., Mani, S. A., Yang, J., Lander, E. S., and Weinberg, R. A. (2008). Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. Cancer Res *68*, 3645-3654.
Park, S. M., Gaur, A. B., Lengyel, E., and Peter, M. E. (2008). The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. Genes Dev *22*, 894-907.
Shankavaram, U. T., Reinhold, W. C., Nishizuka, S., Major, S., Morita, D., Chary, K. K., Reimers, M. A., Scherf, U., Kahn, A., Dolginow, D.*, et al.* (2007). Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. Mol Cancer Ther *6*, 820-832.
Tsai, A., and Carstens, R. P. (2006). An optimized protocol for protein purification in cultured mammalian cells using a tandem affinity purification approach. Nat Protoc *1*, 2820-2827.