Supplementary Materials

Supplementary Figure 1. a. Isolation of mesenchymal and epithelial tissues from fetal human colon. Representative phase-contrast pictures of the luminal intestine after 10 h incubation in Matrisperse. (i) Before shaking, the intestinal epithelium (arrowheads) remains in place. The tip of villus mesenchyme is visible by transparency (arrows). After gentle shaking, the complete dissociation of the entire epithelial lining (ii) from the remaining underlying mesenchyme (iii) is obtained. Bar = 100 μm. Total RNA was extracted and used for assessing the alternative splicing profiles of 283 alternative exons by endpoint RT-PCR. **b.** Quantitative RT-PCR analysis of *E-cadherin* and *vimentin* from the mesenchymal (Mes) and epithelial (Epi) colon samples. Fold change in expression levels of each gene in each tissue type are shown relative to the HCT-116 cell line, arbitrarily set to 1.0. Errors reported for technical triplicate experiments derived using the qBase relative expression method and 3 reference genes (GAPDH, MRLP19 and YWHAZ) **c**. CDH1 and VIM relative mRNA expression in Fallopian tube epithelium and ovarian stroma. The relative expression level of the epithelial markers CDH1 and the stromal mesenchymal markers VIM was monitored in four dissected Fallopian tube epithelium and four dissected ovarian stroma samples by quantitative RT-PCR (Brosseau et al., manuscript submitted). The geometrical mean of three housekeeping (GAPDH, PSMC4 and YWHAZ) was used to calculate the normalization factor (53). Bar graphs represent the mean of four samples and error bars represent standard deviation.

Supplementary Figure 2. Comparing epithelial/mesenchymal-specific alternative exons common to the colon/ovary with breast cell line EMT splicing markers. The Venn diagrams illustrate the distribution of the epithelial/mesenchymal alternative exons that are common to ovarian and colon tissues and that shift in the same direction at various cut-off values. Only simple cassette exons from the Shapiro et al. study (31) were considered as EMT markers.

Supplementary Figure 3. Splicing profiles in a spectrum of cell lines with varying epithelial to mesenchymal characteristics. a. Unsupervised hierarchical clustering of 9 cancer cell lines relative to epithelial and mesenchymal tissues using 199 alternative exons in diverse genes. Percent spliced in values derived from endpoint RT-PCR are in bins of 5 percentage points indicated in shades of yellow (exon skipping) to blue (exon inclusion), as indicated in the color key histogram. **b.** Scatter plot comparing the percent-spliced-in (Ψ) values of 47 alternative splicing events (ASEs) in epithelial and mesenchymal fetal colon tissues. Four switch-like splicing differences are labeled. This set of 47 ASEs is made up of 36 cassette exons, 3 multiple exon splice switches, 4 alternative 3' splice sites, 1 complex alternative splicing event and 3 intron retention events (for primer design see http://palace.lgfus.ca/data/related/1452). These ASEs had previously been associated with cancer and/or apoptosis (33-36) and partially overlap with the 283 alternative exons used in Figure 1b. **c.** Unsupervised hierarchical clustering of cancer cell lines using the 4 alternative exons that are common to the ovarian stroma, the colon mesenchyme and the Twist-induced breast cell line. Color codes are as in panel a.

Supplementary Figure 4. Screen of the effect on alternative splicing of RNA-

binding proteins. The shifts in splicing values for 47 alternative splicing events (ASEs) ($\Delta\Psi = \Psi_{knockdown} - \Psi_{control}$) shown are averages of two knockdowns with different siRNAs (both achieving knockdown over 50%) in up to 5 cell lines for 78 splicing factor candidates. Splicing shifts are in bins of 10% from <-30 percentage points (brightest yellow) to >+30 percentage points (darkest blue). The targeted splicing factors and the corresponding cell lines are indicated on the right in alphabetical order followed by the lowest knockdown efficiency of the two siRNAs. The ASEs are automatically clustered (Ward clustering) on the *x* axis. The large cluster on the left (box *a)* are exons that are downregulated (yellow) when members of the SF3 family of U2-associated core spliceosomal components are knocked down. Box *b* contains introns and 3' splice sites that move in the opposite direction. A subset of the SF3-dependent exons are also dependent on RBFOX2 as seen in box *c.* Splicing events are named on the *x* axis with just gene name for simple cassette exons, except for the two fibronectin exons which are named EDA or EDB. Complex events that include alternative 3' splice sites, facultative introns and multiple exon splicing events are indicated after the gene name.

Supplementary Figure 5. Identification of RBFOX2 as a regulator of mesenchymal-specific splicing. Splicing shifts ($\Delta\Psi$) for 14 of the 47 alternative splicing events (ASEs) that change between epithelium and mesenchyme are shown. Line 1 depicts the differences for epithelium – mesenchyme, and the subsequent lines show the impact of knockdowns in MCF-7, OVCAR3, MDA-MB-231, SKOV3ip1 and PC-3 cell lines affecting at least one ASE by more than 20 percentage points. Shifts ($\Delta\Psi = \Psi_{knockdown} - \Psi_{control}$) of greater than 20% are shown in yellow and blue according to the legend. The labels on the *y* axis indicate the protein knocked down followed by the cell line and a Pearson correlation value (R) of the splicing profile relative to the profile on line 1.

Supplementary Figure 6. Additional regulated ASEs. Median splicing values for the 27 RBPs which shifted splicing across two or more cell lines by more than 20 percentage points (median value) for at least one ASE. We previously showed that hnRNP K affected many exons using a set of splicing events in apoptotic genes (36). In the current set, hnRNP K also acts as an enhancer or a repressor for 14 of the 47 ASEs tested. MBNL1 inhibits three exons (the FN1-EDA exon and exons in *LGALS9* and *FGFR1). FN1-EDA* is also inhibited by RBM10. RBM39 (RNPC2/CAPER) had a strong splicing signature, shifting 10 ASEs. It enhanced some exons, most notably *KITLG,* and inhibited *AKAP13* and *IGSF4.* RBM39 also favored the downstream *BCL2L1* 3' splice site. SRSF3 (SRp20) had the strongest splicing signature of the SR proteins in our cell lines; promoting the use of all four upstream 3' splice sites in our set of 47 alternative splice events. SRSF3 also inhibited *AKAP13, RUNX2* and *SMG7* exons and enhanced *PTK2B* and *BCL2L12.* TIAL1 strongly inhibited an exon in *OPA1.* Finally of note, knockdown of U2AF1 compromised the inclusion of exons in *CHEK2* and *PPP3CB.*

Supplementary Table 1. A total of 43 hits (24 mesenchyme-specific exons in blue and 19 epithelium-specific exons in green) that shifted by more than 50 percentage points between fetal colon epithelium and mesenchyme are shown with an abbreviated description of their function extracted from the NCBI database (http://www.ncbi.nlm.nih.gov/gene). The first 29 hits were uncovered in the initial screen of 283 exons, the next three from the cancer/apoptosis set and the last 11 from our

previously identified RBFOX2 targets (33).

Supplementary Table 2.

Tab a. Splicing data for the 283 alternative exons. PCRs for these exons passed the quality control test (>75% of PCR products found at their expected mobilities, >20 nM total in expected peaks) in fetal colon samples; the following data are listed in the following columns: **A** Gene symbol, **B** Alternative splice event unique symbol (gene name followed by exon size for genes with more than one alternative exon in this study). The 12 mesenchyme-specific and 14 epithelial-specific exons are colored yellow and blue, respectively, **C-E** Sequences of the alternative spliced exon (D) and 150 nucleotides on either side in the introns (C,E), **F,G** Primer sequences used for PCR, **H,I** PCR product sizes for each isoform **J, K,** Ψ values for epithelial and mesenchymal fetal colon, **L** Splicing shift between epithelium and mesenchyme, **M** Difference in exon inclusion for the 178 exons tested in laser-captured and microdissected ovarian stromal tissues (OVN) and fallopian tubes epithelium (FTE), **N** Average splicing shift between 25 cancer and 21 normal samples (33), **O-W** Ψ values for 9 cell lines (3 epithelial (orange) and 6 mesenchymal (yellow)), **X-AA** Splicing shifts for the 26 switch-like exons in additional knockdowns for candidate epithelial and mesenchymal splicing factors, **AB** 'In-house' specific primer names. NA = data not available, where shaded because PCR was not performed.

Tab b Splicing data for 47 ASEs. A Gene symbol, **B** Alternative splice event unique symbol (gene name followed by type of splicing event when not a simple cassette exon). The 2 mesenchyme-specific and 2 epithelial-specific exons are colored yellow and blue, respectively, **C** Type of splicing event, **D-F** The sequences of the alternative spliced exon and 150 nucleotides either side in the introns or the extra sequence and flanking regions for non cassette exons, **G,H** Primer sequences used, **L,J** PCR product sizes for each isoform, **K,L** Percent spliced in (Ψ) values for epithelial and mesenchymal fetal colon, **M** Splicing shift between epithelium and mesenchyme, **N-V** Ψ values for 9 cell lines (3 epithelial and 6 mesenchymal), **W** Splicing shifts between median epithelial and mesenchymal cell line values, **X** Splicing shift in RBFOX2 knockdown (shifts less than 10% shown as zero), **Y** 'In-house' specific primer names. NA = data not available.

Tab c. Full data for potential splicing factor knockdown screen. A Protein knocked-down, **B** Cell line treated, **C** siRNA sequence used, **D** Percentage knockdown achieved as assessed by qPCR, **E-AY** Shift in splicing (knockdown-control) for 47 alternative splicing events detailed in tab **b**.

Tab d Average shifts in percent spliced in values (2 siRNAs-control) for 227 combinations of RBP and cell lines shown in columns **A and B** are given for the 47 ASEs (from tab **b**) shown in columns **C-AW.** Shifts in Ψ values for fetal epithelium-mesenchyme are shown in row 2. NA = data not available.

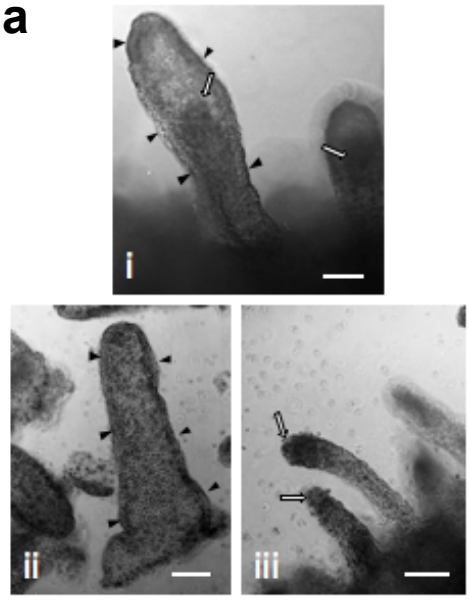
Tab e Median splicing shifts for 78 RBPs across 1-5 cell lines. The protein knocked down is shown in column **A** along with a brief description of its function and known presence or absence of RNA-binding doemain in columns B and C. The number of cell

lines used to take the median value is indicated in column **D.** The splicing shifts (knockdown - control) for the 47 ASEs (from tab b) are shown in columns **E-AY.** Shifts in Ψ values for fetal epithelium-mesenchyme are shown in row 2. NA = data not available.

Tab f Splicing results for 48 RBFOX2 targets (from reference 33). Column **A** gene name. Mesenchyme-specific and epithelial-specific exons are colored yellow and blue, respectively, **B-D** the sequences of the alternative spliced exon and 150 nucleotides either side in the introns, **E-F** primer sequences used, **G-H** PCR product sizes for each isoform, **J** average RBFOX2 knockdown splicing shift values in two cell lines (knockdown - control) (33), **J,K** Ψ values for epithelial and mesenchymal fetal colon. **L** Splicing shift between epithelium and mesenchyme. **M '**In-house' specific primer names.

Tab g qPCR primers for detection of E-cadherin, vimentin and *Rbfox2.*

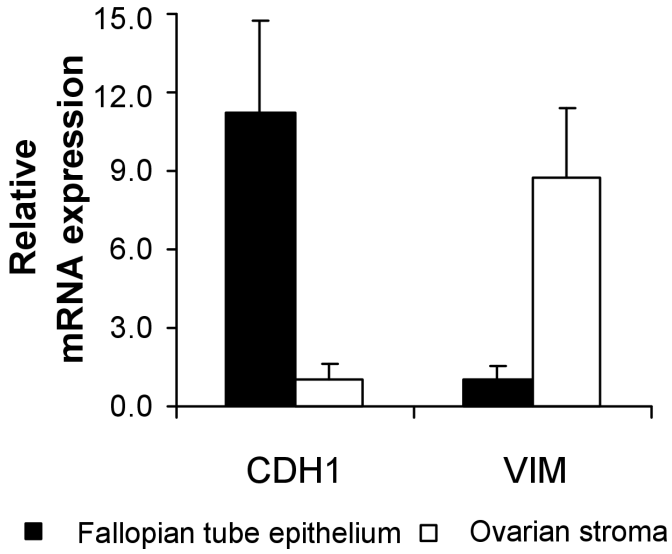
a



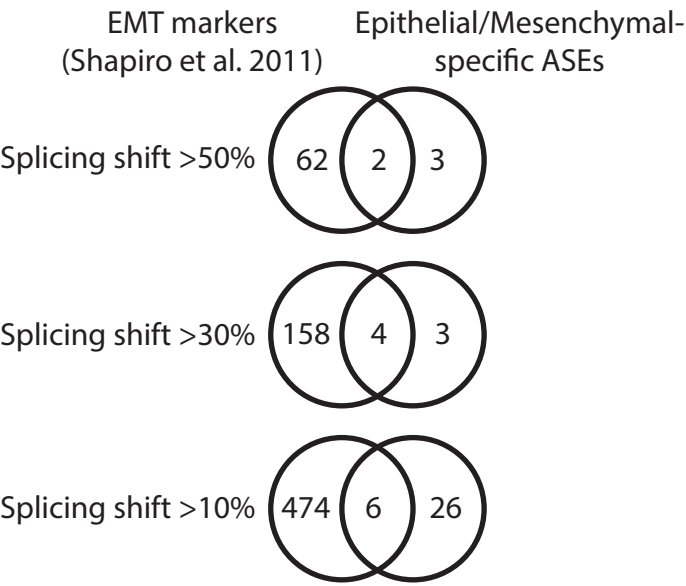
b

	Mes	Epi
E-cadherin	1.0 ± 0.1	145.9 ± 5.1
vimentin	540.1 ± 33.3	3.8 ± 0.2

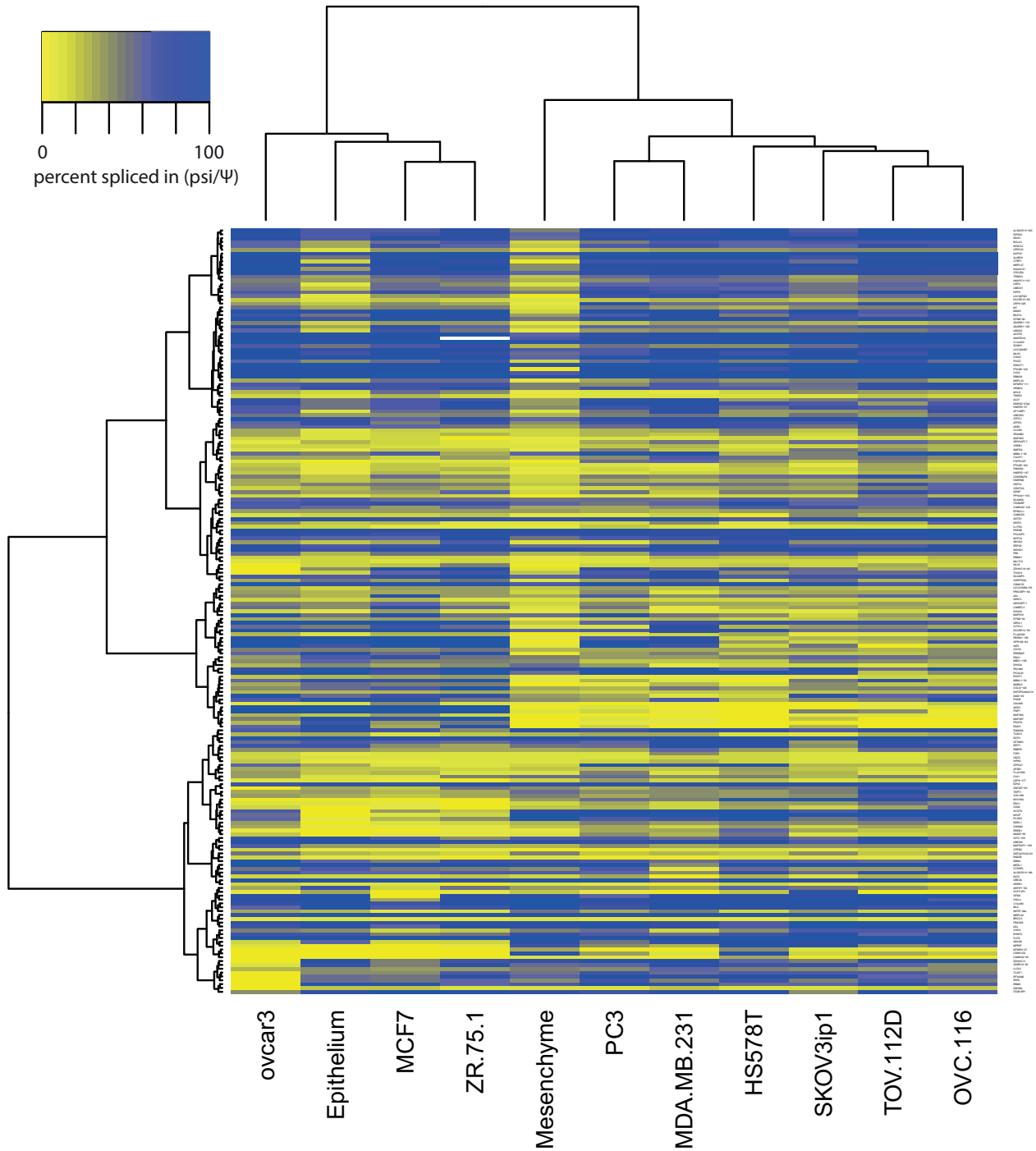
c



Supplementary Figure 1

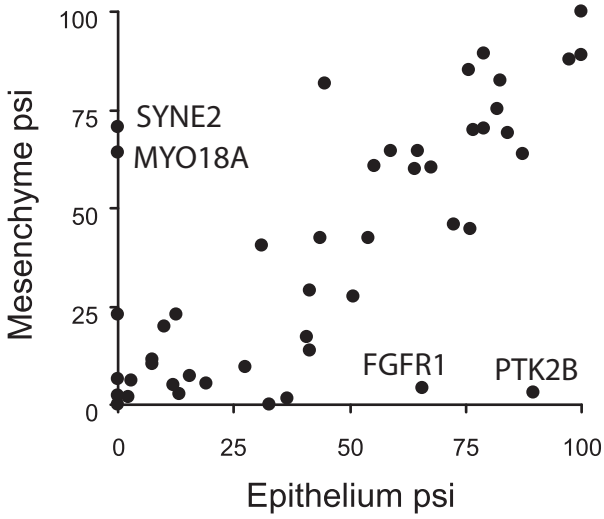


Supplementary Figure 2

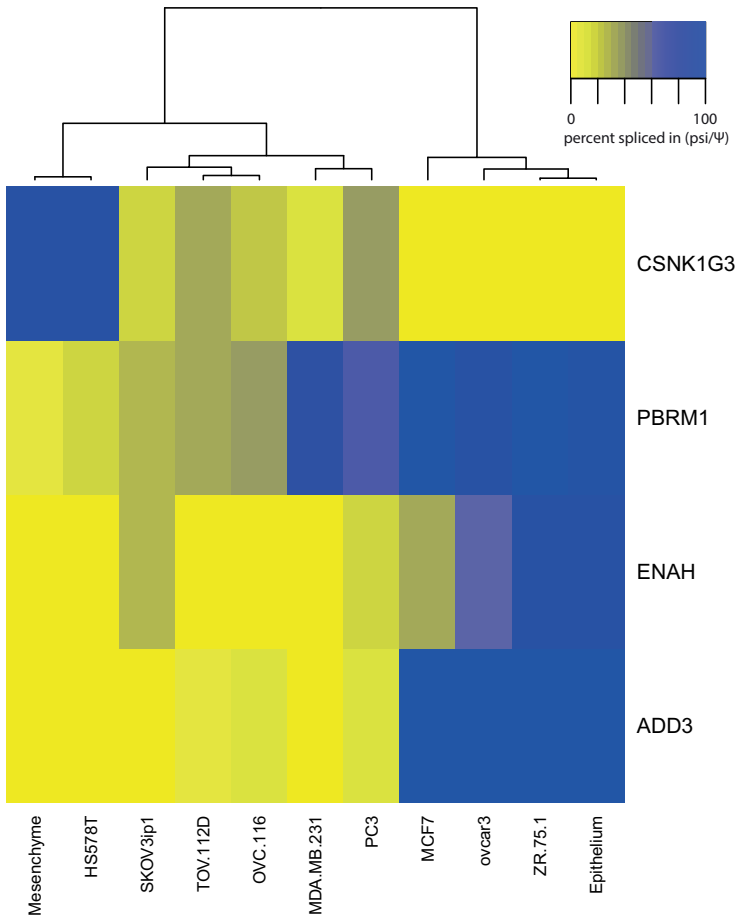


Supplementary Figure 3a

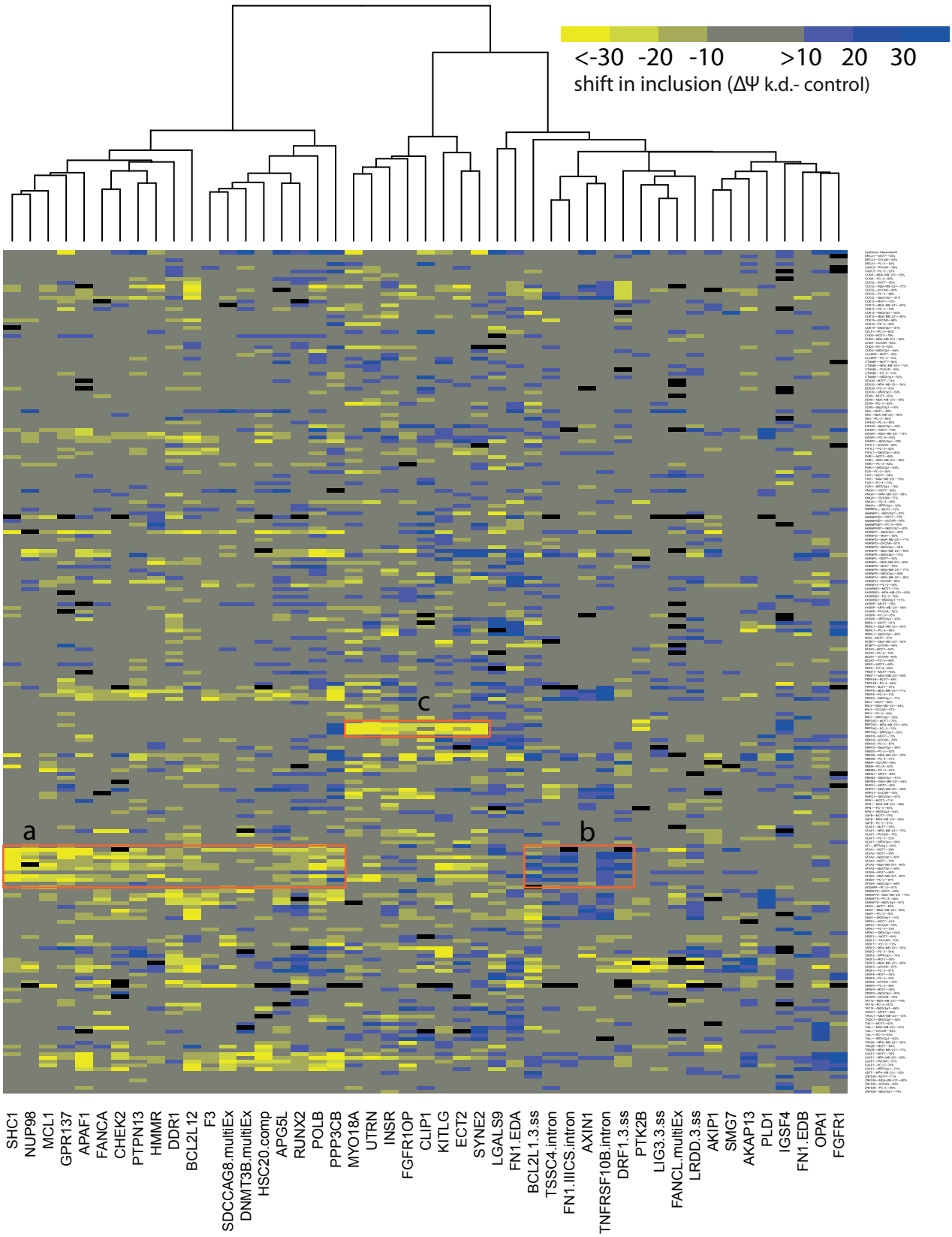
b



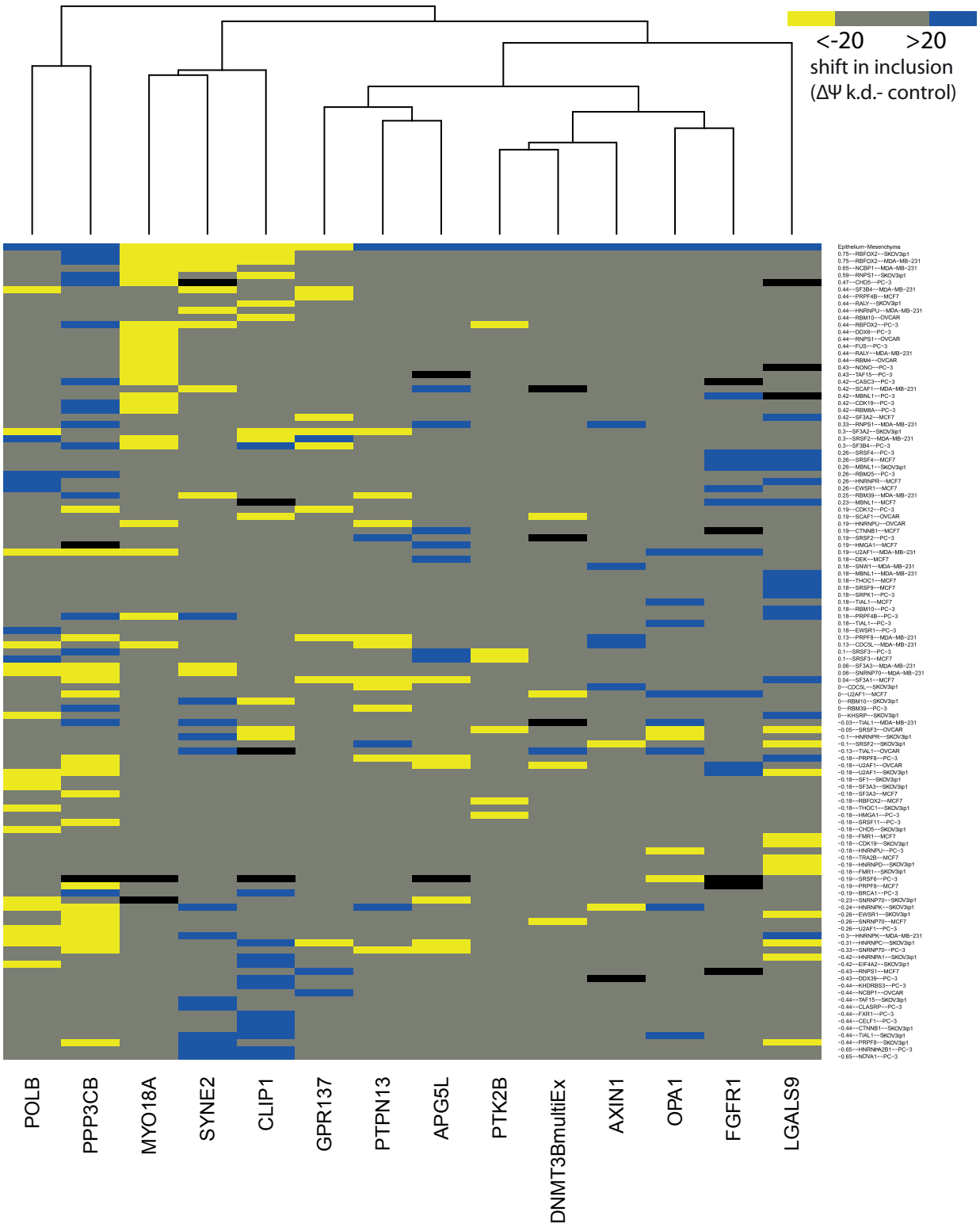
c



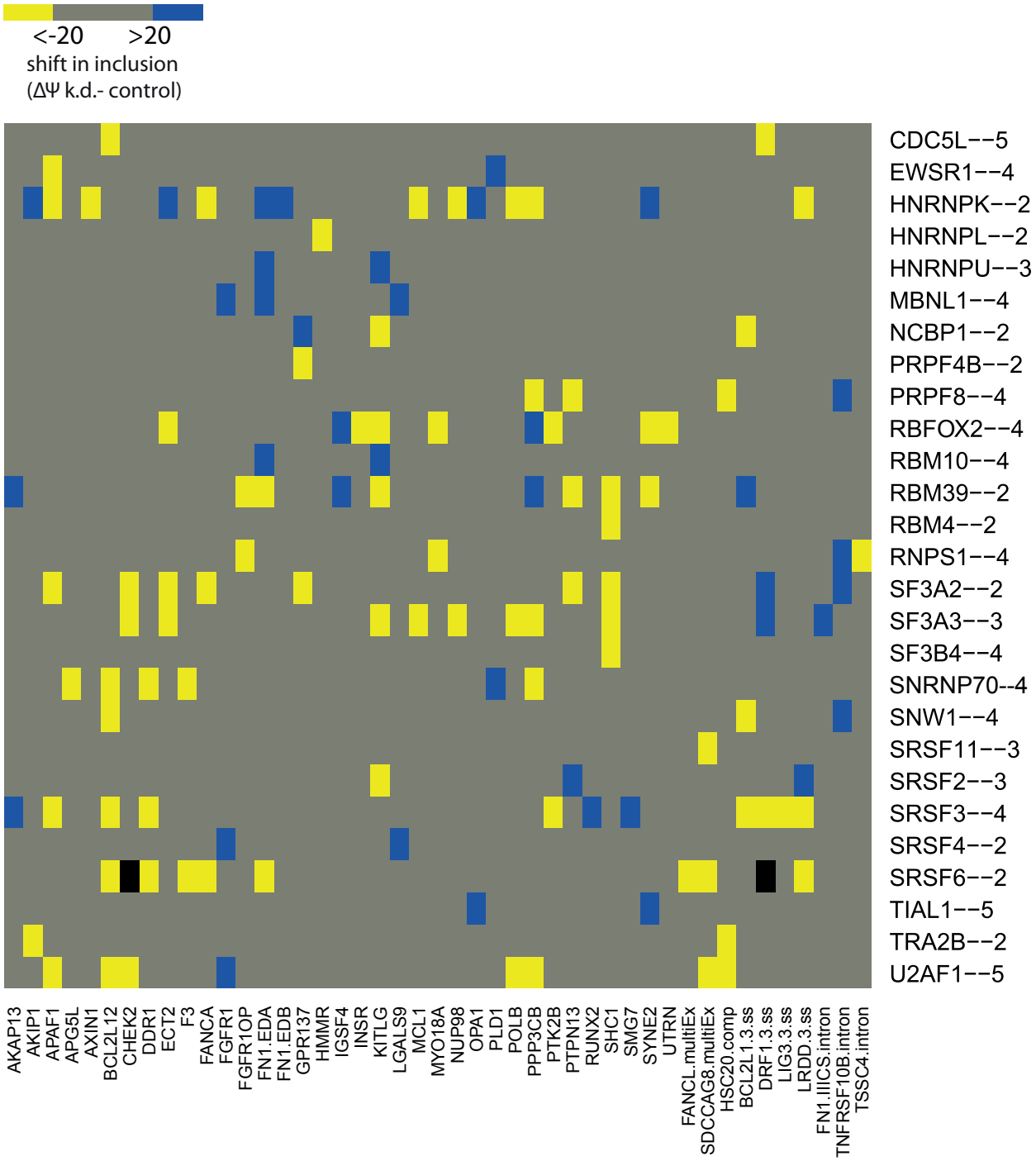
Supplementary Figure 3 (cont)



Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 6

Gene name	Summary
ADD3	Adducins are heteromeric proteins composed of different subunits referred to as adducin alpha, beta and gamma. The three subunits are encoded by distinct genes and belong to a family of membrane skeletal proteins involved in the assembly of spectrin-actin network in erythrocytes and at sites of cell-cell contact in epithelial tissues.
GPR126	This gene, which is upregulated in human umbilical vein endothelial cells, encodes a G protein-coupled receptor.
MAP3K7	The protein encoded by this gene is a member of the serine/threonine protein kinase family. This kinase mediates the signaling transduction induced by TGF beta and morphogenetic protein (BMP), and controls a variety of cell functions including transcription regulation and apoptosis.
ENAH	This gene encoded actin-associated protein involved in a range of processes dependent on cytoskeleton remodeling and cell polarity such as axon guidance and lamellipodial and filopodial dynamics in migrating cells.
FNIP1	May be involved in energy and/or nutrient sensing through the AMPK and mTOR signaling pathways.
PBRM1	This locus encodes a subunit of ATP-dependent chromatin-remodeling complexes. The encoded protein has been identified as in integral component of complexes necessary for ligand-dependent transcriptional activation by nuclear hormone receptors. Mutations at this locus have been associated with primary clear cell renal cell carcinoma.
PTK2B	This gene encodes a cytoplasmic protein tyrosine kinase which is involved in calcium-induced regulation of ion channels and activation of the map kinase signaling pathway.
ERBB2IP	The encoded protein binds to the unphosphorylated form of the ERBB2 protein and regulates ERBB2 function and localization. It has also been shown to affect the Ras signaling pathway by disrupting Ras-Raf interaction.
MBNL1	This gene encodes a splicing factor of the muscleblind-like family. Acts either as activator or repressor of splicing on specific pre-mRNA targets. Antagonizes the alternative splicing activity pattern of CELF proteins.
GRHL1	This gene encodes a member of the grainyhead family of transcription factors. The encoded protein can exist as a homodimer or can form heterodimers with sister-of-mammalian grainyhead or brother-of-mammalian grainyhead. This protein functions as a transcription factor during development.
PDGFA	The protein encoded by this gene is a member of the platelet-derived growth factor family. The four members of this family are mitogenic factors for cells of mesenchymal origin.
INF2	This gene represents a member of the formin family of proteins. It is considered a diaphanous formin. Involved in polymerization and depolymerization of actin filaments.
MINK1	This gene encodes a serine/threonine kinase that activates the cJun N-terminal kinase (JNK) and the p38 signalling pathways. Involved in the regulation of actin cytoskeleton reorganization, cell-matrix adhesion, cell-cell adhesion and cell migration
MARK2	This gene encodes a member of the Par-1 family of serine/threonine protein kinases. The protein is an important regulator of cell polarity in epithelial and neuronal cells, and also controls the stability of microtubules through phosphorylation and inactivation of several microtubule-associating proteins. The protein localizes to cell membranes.
SNX14	Members of this family contain a phox (PX) domain, which is a phosphoinositide binding domain, and are involved in intracellular trafficking. The encoded protein also contains a regulator of G protein signaling (RGS) domain. Regulator of G protein signaling family members are regulatory molecules that act as GTPase activating proteins for G alpha subunits of heterotrimeric G proteins.
SULF2	Heparan sulfate proteoglycans (HSPGs) act as coreceptors for numerous heparin-

	binding growth factors and cytokines and are involved in cell signaling.
ARNT	The aryl hydrocarbon (AH) receptor has also been identified as the beta subunit of a heterodimeric transcription factor, hypoxia-inducible factor 1. A t(1;12)(q21;p13) translocation, which results in a TEL-ARNT fusion protein, is associated with acute myeloblastic leukemia.
ACOT9	The protein encoded by this gene is a mitochondrial acyl-CoA thioesterase of unknown function.
MPRIP	Targets myosin phosphatase to the actin cytoskeleton. Required for the regulation of the actin cytoskeleton by RhoA and ROCK1. Depletion leads to an increased number of stress fibers in smooth muscle cells through stabilization of actin fibers by phosphorylated myosin.
ST7	The function of this gene product has not been determined.
RALGAPA1	Catalytic subunit of the heterodimeric RalGAP1 complex which acts as a GTPase activator for the Ras-like small GTPases RALA and RALB (by similarity).
CSNK1G3	CK1 is involved in a number of cellular processes including DNA repair, cell division, nuclear localization and membrane transport. CK1 isoforms also have key roles in the developmentally important Wnt and Hedgehog (Hh) signaling pathways.
LRRFIP2	May function as activator of the canonical Wnt signaling pathway, in association with DVL3, upstream of CTNNB1/beta-catenin.
PTPRD	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation.
PLOD2	The protein encoded by this gene is a membrane-bound homodimeric enzyme that is localized to the cisternae of the rough endoplasmic reticulum. The enzyme (cofactors iron and ascorbate) catalyzes the hydroxylation of lysyl residues in collagen-like peptides. The resultant hydroxylysyl groups are attachment sites for carbohydrates in collagen and thus are critical for the stability of intermolecular crosslinks.
MYOF	It is a member of the ferlin family and associates with both plasma and nuclear membranes. The protein contains C2 domains that play a role in calcium-mediated membrane fusion events, suggesting that it may be involved in membrane regeneration and repair.
FGFR1	FGFR family members differ from one another in their ligand affinities and tissue distribution. A full-length representative protein consists of an extracellular region, composed of three immunoglobulin-like domains, a single hydrophobic membrane-spanning segment and a cytoplasmic tyrosine kinase domain. signaling, ultimately influencing mitogenesis and differentiation.
SYNE2	Nuclear outer membrane protein that binds cytoplasmic F-actin. This binding tethers the nucleus to the cytoskeleton and aids in the maintenance of the structural integrity of the nucleus.
MYO18A	May be involved in the maintenance of the stromal cell architectures required for cell to cell contact (by similarity). In concert with LRP35A and CDC42BPA/CDC42BPB, is involved in modulating lamellar actomyosin retrograde flow that is crucial to cell protrusion and migration.
DIAPH2	The product of this gene belongs to the diaphanous subfamily of the formin homology family of proteins. This gene may play a role in the development and normal function of the ovaries. Defects in this gene have been linked to premature ovarian failure 2.
GIT2	GTPase-activating protein. This gene has been shown to repress lamellipodial extension and focal adhesion turnover, and is thought to regulate cell motility.
NPHP3	Required for normal ciliary development, and it functions in renal tubular development.

NFYA	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters, for example in type 1 collagen, albumin and beta-actin genes.
FMNL3	Plays a role in the regulation of cell morphology and cytoskeletal organization. Required in the control of cell shape and migration.
PLEKHM2	May play a role in the regulation of conventional kinesin activity.
APBB2	The protein encoded by this gene interacts with the cytoplasmic domains of amyloid beta (A4) precursor protein and amyloid beta (A4) precursor-like protein 2. This protein contains two phosphotyrosine binding (PTB) domains, which are thought to function in signal transduction.
EXOC1	The protein encoded by this gene is a component of the exocyst complex, a multiple protein complex essential for targeting exocytic vesicles to specific docking sites on the plasma membrane. At least eight components of the exocyst complex, including this protein, are found to interact with the actin cytoskeletal remodeling and vesicle transport machinery.
CA12	carbonic anhydrase XII is overexpressed in 10% of clear cell renal carcinomas
ATP11C	Encodes a member of the P4 ATPase family thought to serve as 'flippases' that concentrate aminophospholipids in the cytoplasmic leaflet of cell membranes.
MACF1	Possibly coupling the microtubule network to cellular junctions.

SUPPLEMENTARY TABLE 1