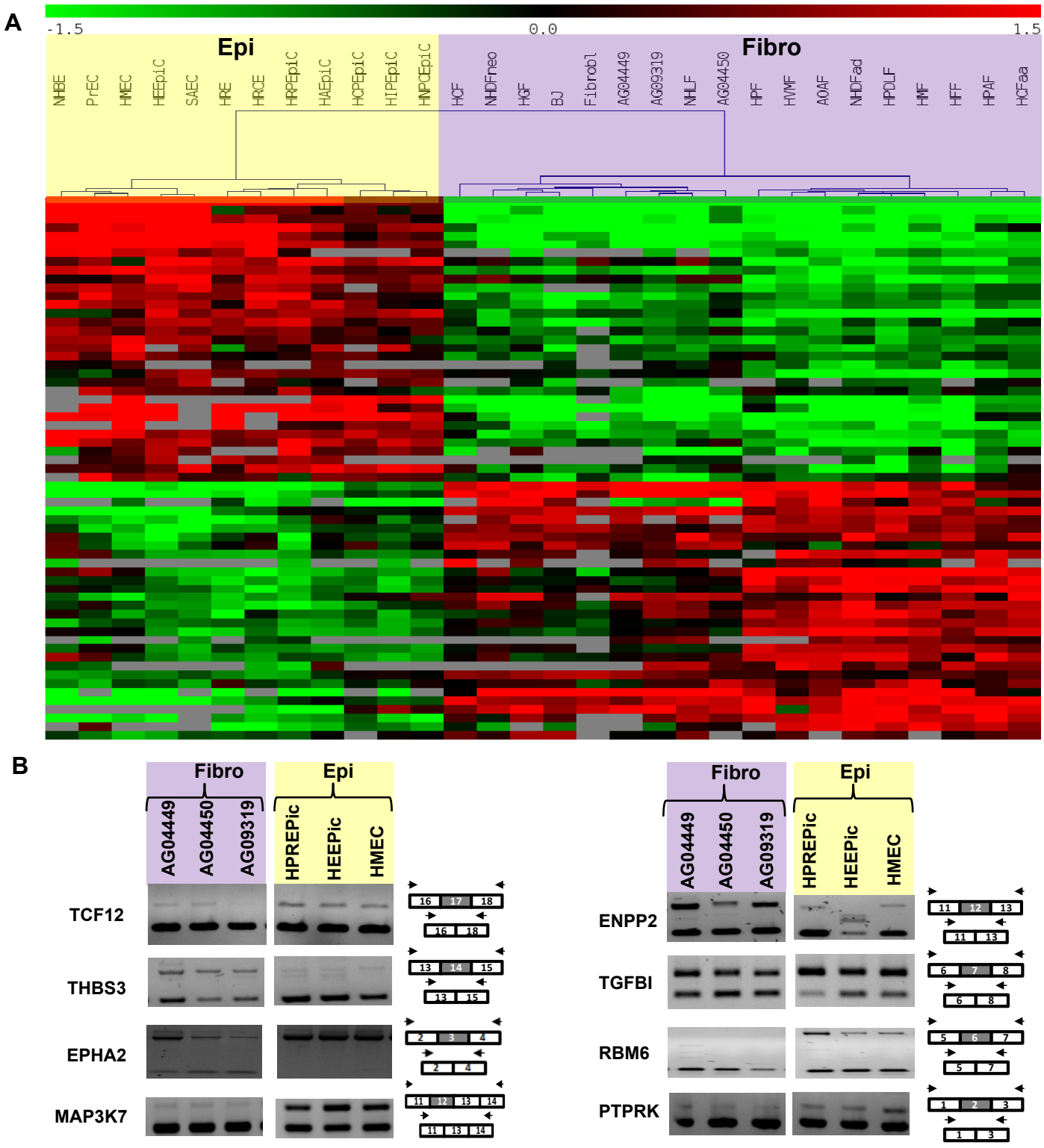
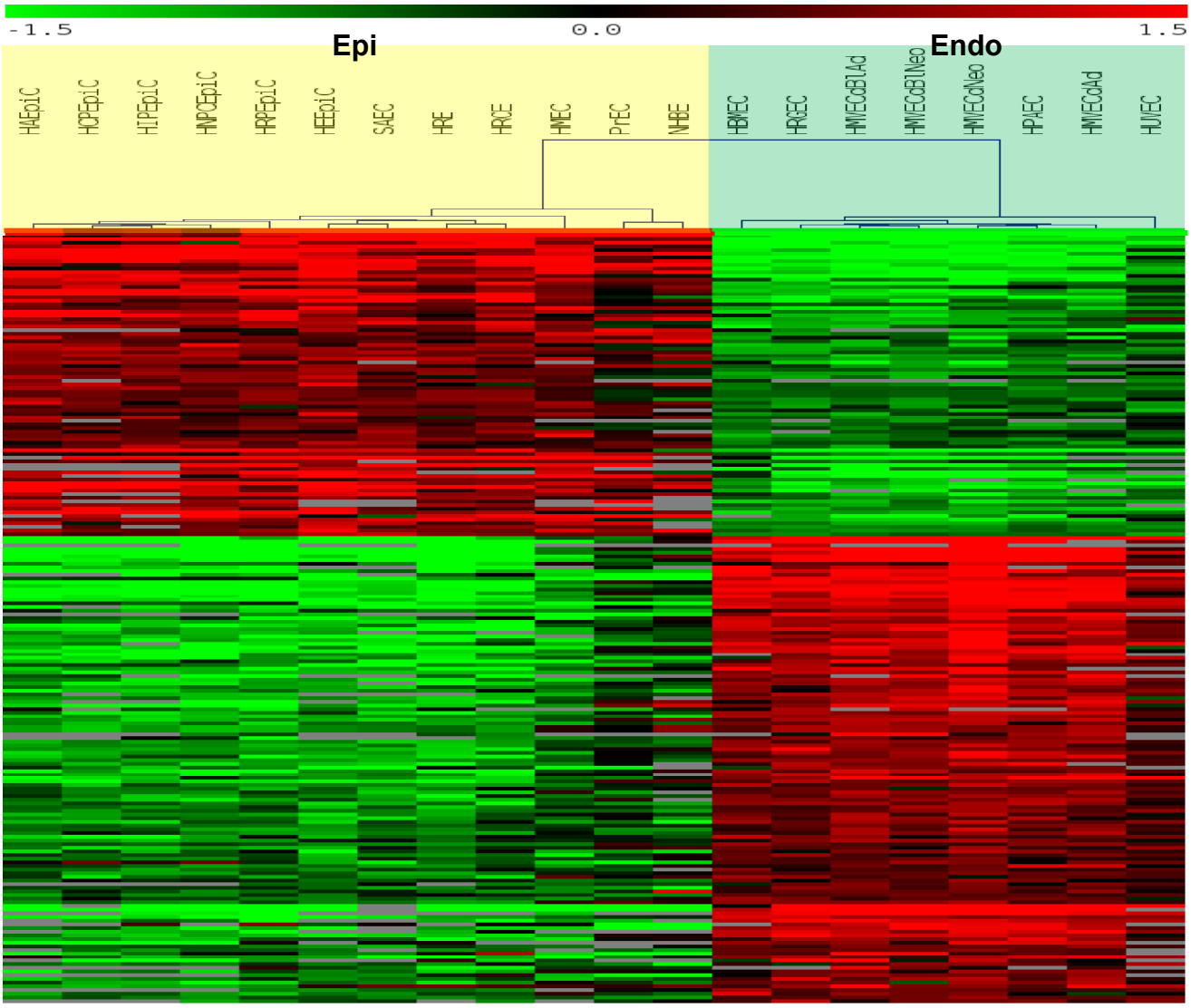


Supplementary Figure S1.

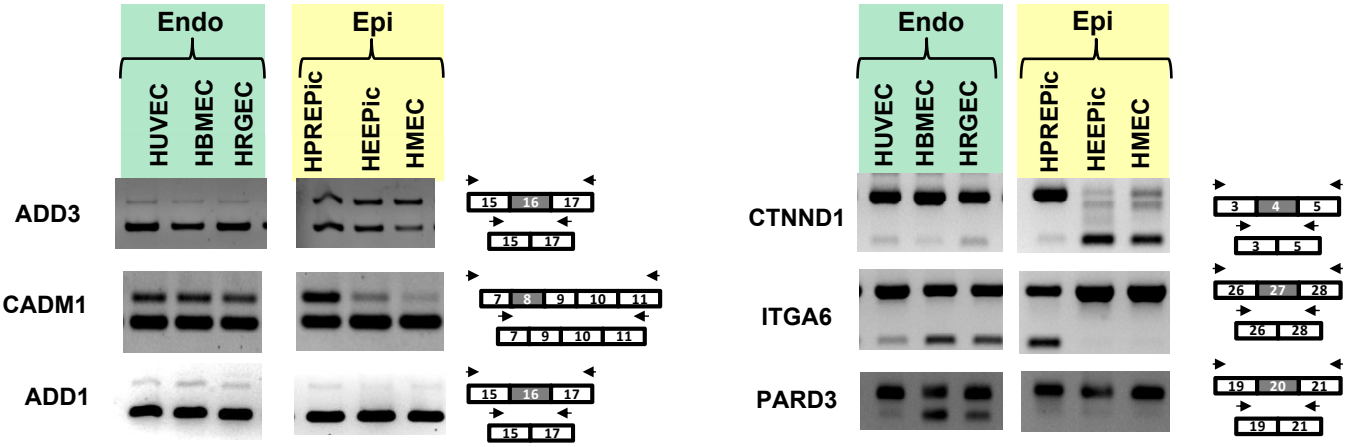
A. HeatMap presentation of the Splicing Index (SI) values for exons differentially spliced when comparing fibroblast to epithelial cells. Each line corresponds to a regulated exon while each column corresponds to a specific cell. Green boxes (-1.5 < SI < 0) correspond to low inclusion level in the cell compared to all the others; red boxes (0 < SI < 1.5) to high inclusion level; black boxes (SI=0) point no difference for exon inclusion between cells while grey boxes correspond to missing values. Exons were computationally split into several groups depending on their inclusion rate which correlate with the two major cell types.
(same as in Fig. 1B)

B. RT-PCR validations using RNAs from fibroblasts and epithelial cells, as indicated.

A



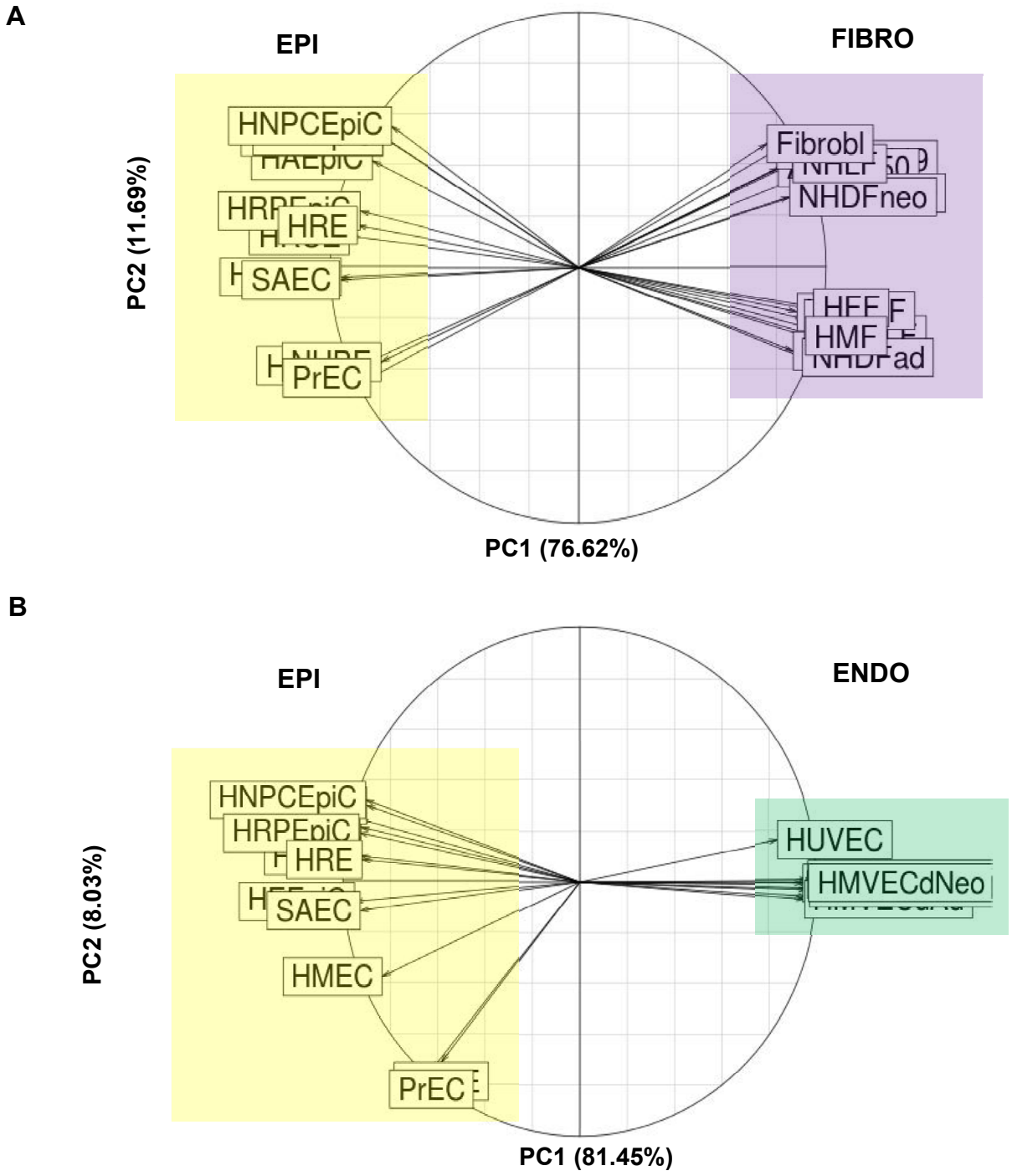
B



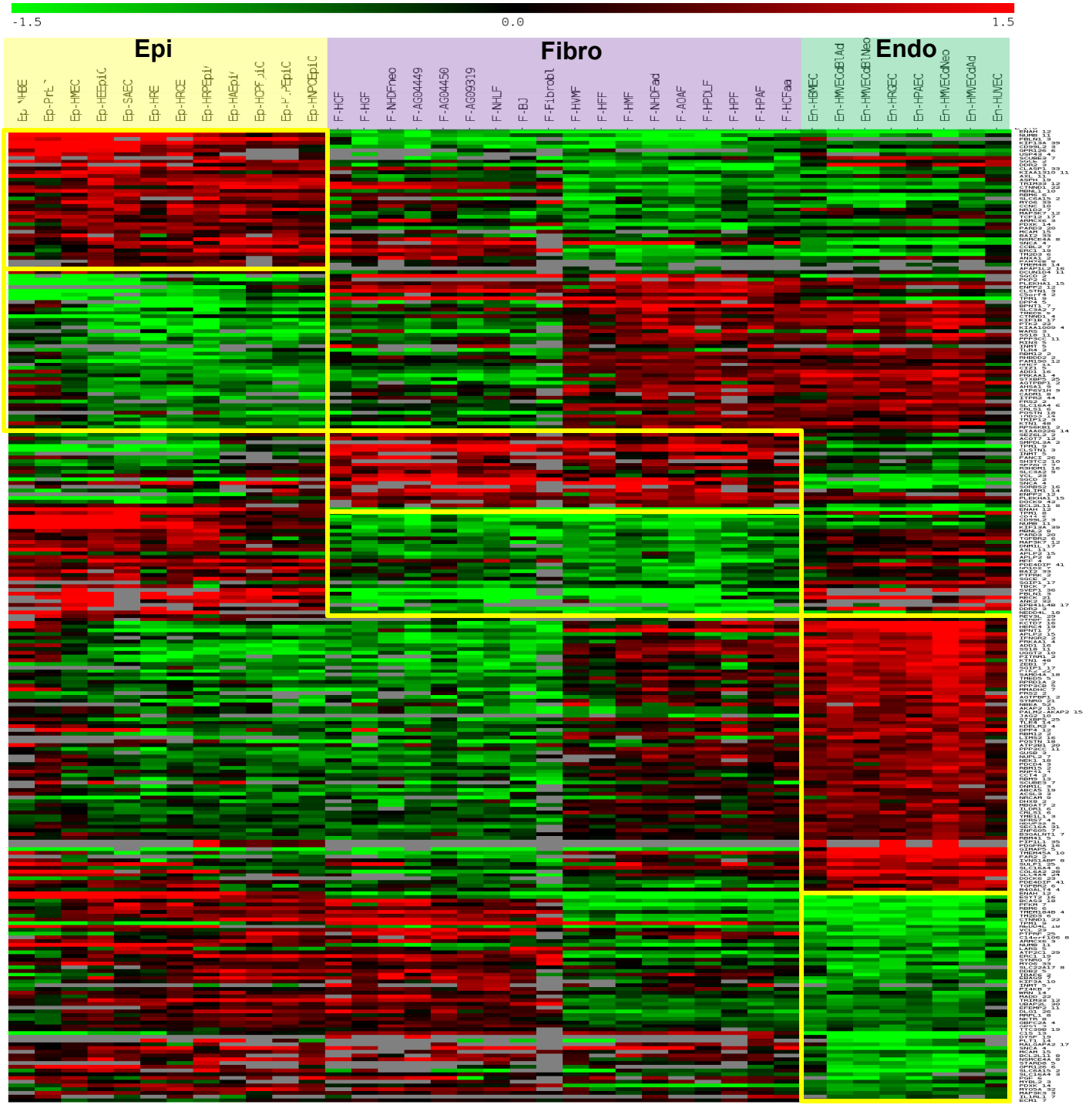
Supplementary Figure S2.

A. HeatMap presentation of the Splicing Index (SI) values for exons differentially spliced when comparing endothelial to epithelial cells. Each line corresponds to a regulated exon while each column corresponds to a specific cell. Green boxes (-1.5 < SI < 0) correspond to low inclusion level in the cell compared to all the others; red boxes (0 < SI < 1.5) to high inclusion level; black boxes (SI=0) point no difference for exon inclusion between cells while grey boxes correspond to missing values. Exons were computationally split into several groups depending on their inclusion rate which correlate with the two major cell types.
(same as in Fig. 2B)

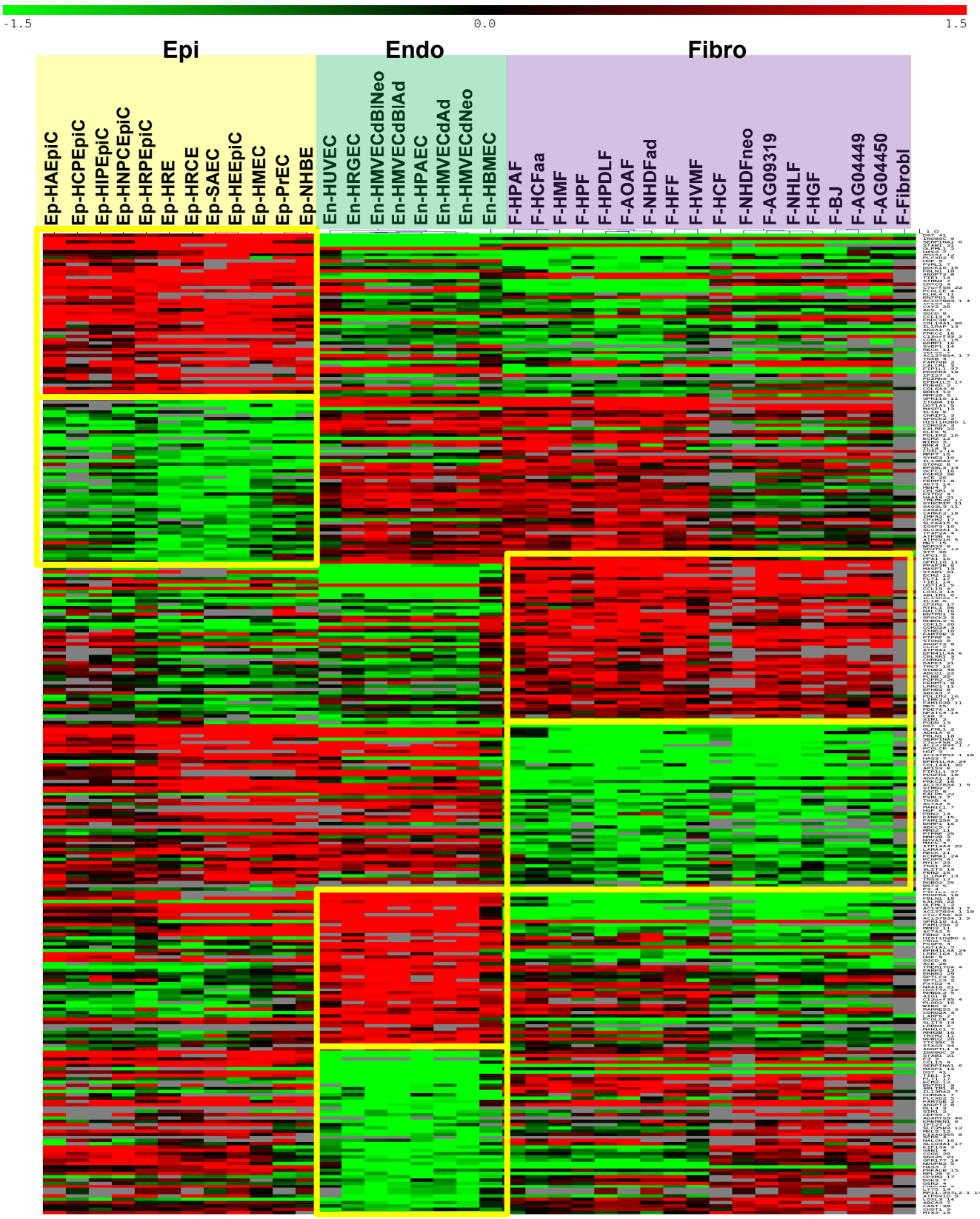
B. RT-PCR validations using RNAs from endothelial and epithelial cells, as indicated.



Supplementary Figure S3: Circles of correlations and arrows of the loadings of variables with principal components 1 and 2 for the fibroblastic vs. epithelial cell lines comparison (A) and the endothelial vs. epithelial comparison (B). The percentage of explained inertia by each component is also indicated for each of the three analyses. Epithelial cell lines are shown in yellow, endothelial in green and fibroblastic in purple. Principal component analysis was performed using the ade4 package of R (http://pbil.univ-lyon1.fr/ade4/).

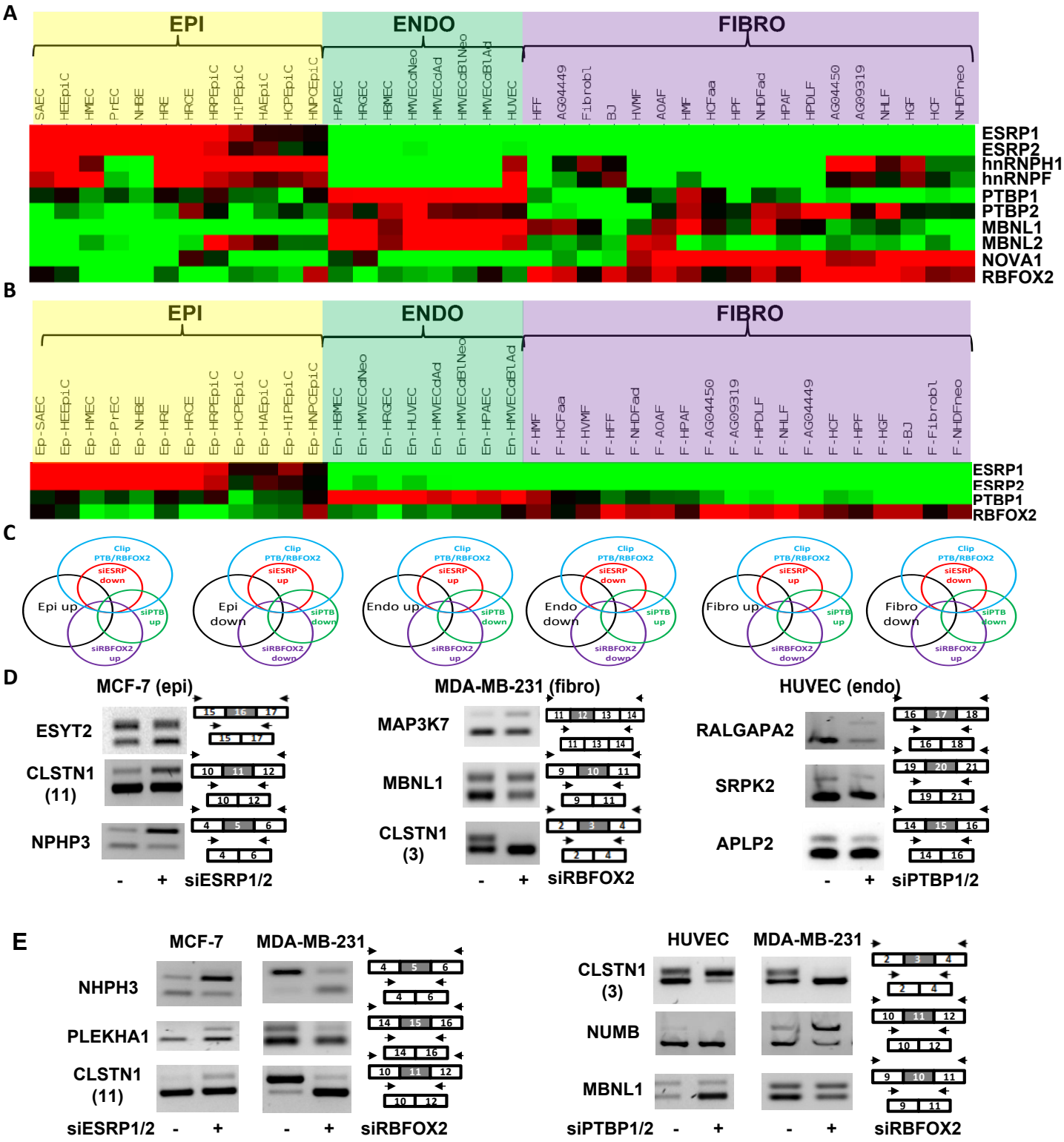


Supplementary Figure S4: HeatMap presentation of the Splicing Index (SI) values for exons differentially spliced across fibroblast, endothelial and epithelial cells. Exons were computationally split into several groups depending on their inclusion rate in the three major cell types (same as in Fig. 3A).



Supplementary Figure S5: HeatMap presentation of the Splicing Index (SI) values for Alternative Last Exons (ALE) differentially spliced across fibroblast, endothelial and epithelial cells. Exons were computationally split into several groups depending on their inclusion rate in the three major cell types (same as in Supplementary Fig. S4).

Supplementary Figure S6: HeatMap presentation of the Splicing Index (SI) values for Alternative First Exons (AFE) differentially spliced across fibroblast, endothelial and epithelial cells. Exons were computationally split into several groups depending on their inclusion rate in the three major cell types (same as in Supplementary Fig. S4).



Supplementary Figure S7.

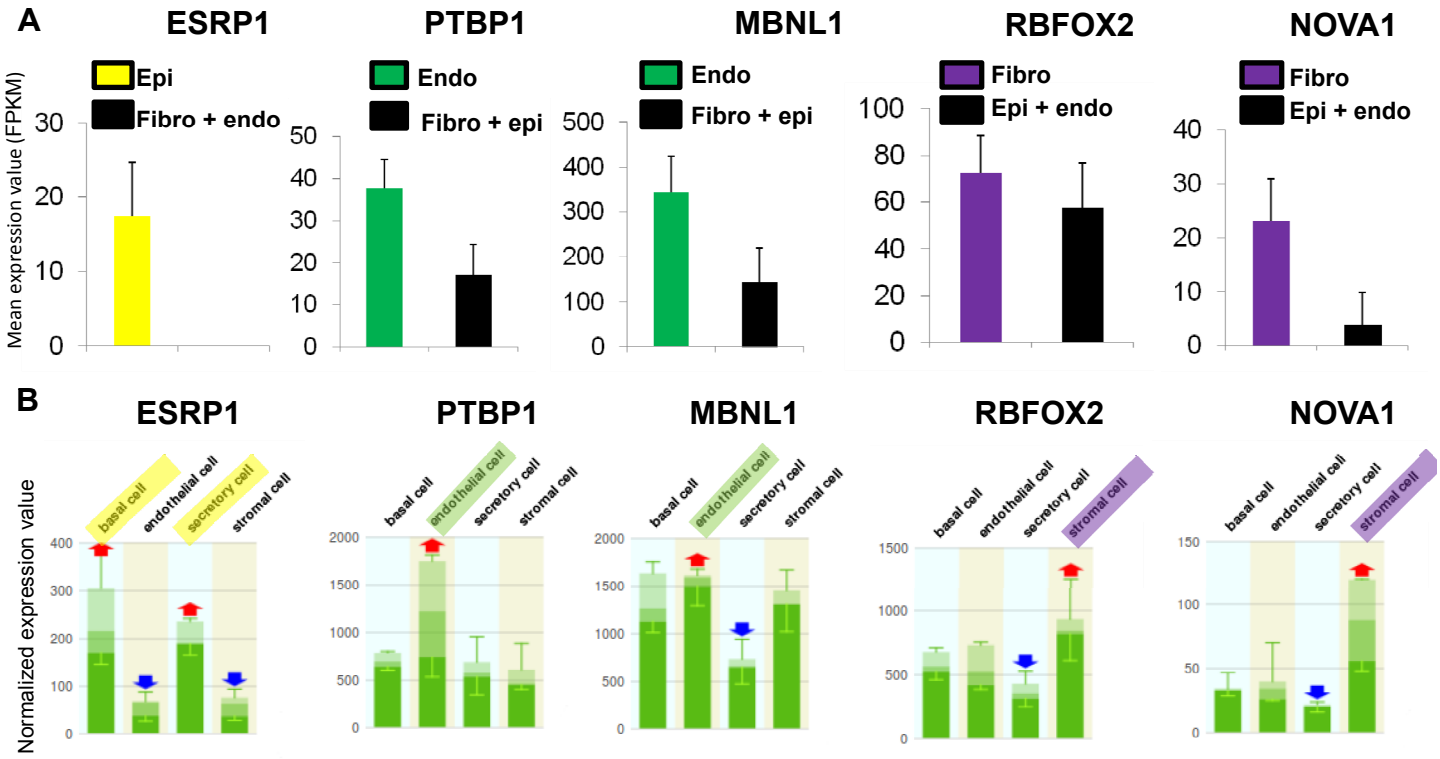
A. HeatMap of splicing factor expression level (same as Fig. 4A).

B. Same as above with a selection of splicing factors.

C. Strategy of splicing factor analyses. ASEs predicted to be regulated by a splicing factor using exon-array or RNA-seq datasets were compared to cell-type specific predicted ASEs. For each cell type, we paid attention to the ASE regulation sense. For example, in epithelial cells, as ESRP is up-regulated while PTB and RBFOX2 are down-regulated, an epithelial-included ASE is predicted to be regulated by one of these factors if it is skipped upon ESRP depletion or included by PTB or RBFOX2 depletion. To be considered as confident, CLIP-hits must be detected in the exon or within 100 nt upstream and downstream of the exon.

D. RT-PCR analysis of the effect of ESRP-, RBFOX2-, PTB-depletion on alternative splicing of selected genes in the MCF-7 epithelial cell line, MDA-MB-231 fibroblast-like cell line, and HUVEC endothelial cell, respectively.

E. Same as in D



Supplementary Figure S8.

A. Samples corresponding to the normal cell lines available for the epithelial, endothelial and fibroblast categories were selected from total long, whole cell RNA-Seq samples of the ENCODE project (ENCODE Project Consortium, 2007, 2011) Tier 3, sequenced by the Cold Spring Harbor Laboratory. Endothelial cells are Haoec, Human Aortic Endothelial Cells (2 replicates) and Hsavec, Human Saphenous Vein Endothelial Cells (2 replicates); Epithelial cells are Hmepc, Mammary Epithelial Cells and Hpiepc, Human Placental Epithelial Cells amniotic membrane (2 replicates); Fibroblasts are Nhdf, Normal Human Dermal Fibroblasts from temple / breast (2 replicates) and Haoaf, Human Aortic Adventitial Fibroblasts (1 replicate). Read alignments for the different cell lines in bam format have been downloaded using the ENCODE project Data Coordination Centre, from the ENCODE RNA Dashboard (http://genome.crg.es/encode_RNA_dashboard/). Cuffdiff, from the Cufflinks package (Trapnell et al. 2013) was used to analyse the gene expression, using the protocol described as "Differential analysis without gene and transcript discovery". We kept the default parameters, except that we specified the RNA-Seq library type (--library-type fr-firststrand), following the recommendations provided by the STAR program (Dobin et al. 2013), which was used for the alignment. The gene annotation file in gtf format was by Illlumina iGenomes, as suggested by the Cufflinks website. The mean value and standard deviation were computed for each gene of the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) expression values obtained by each replicates in a category. The figure shows the mean expression value in FPKM of selected genes in the different cell types. These results confirm the expression pattern across endothelial, epithelial and fibroblast cells of the selected splicing factors obtained using Exon Array datasets (main manuscript Fig. 4A) or by RT-qPCR (main manuscript Fig. 4B).

B. The experiment E-GEOD-3998, which corresponds to a transcription profiling of four human prostate cell types (Oudes et al. 2006) was selected from the Gene Expression Atlas website (http://www.ebi.ac.uk/gxa; Kapushesky et al. 2012). Prostate basal and secretory cells are two epithelial cell types, and stromal cells are composed of fibroblasts and smooth muscle cells. The figure shows the expression value of selected genes in the different cell types, and the up and down arrows show respectively the up- and down-regulated genes. These results confirm the expression pattern across endothelial, epithelial and fibroblast cells of the selected splicing factors obtained using Exon Array datasets (main manuscript Fig. 4A) or by RT-qPCR (main manuscript Fig. 4B).

References:

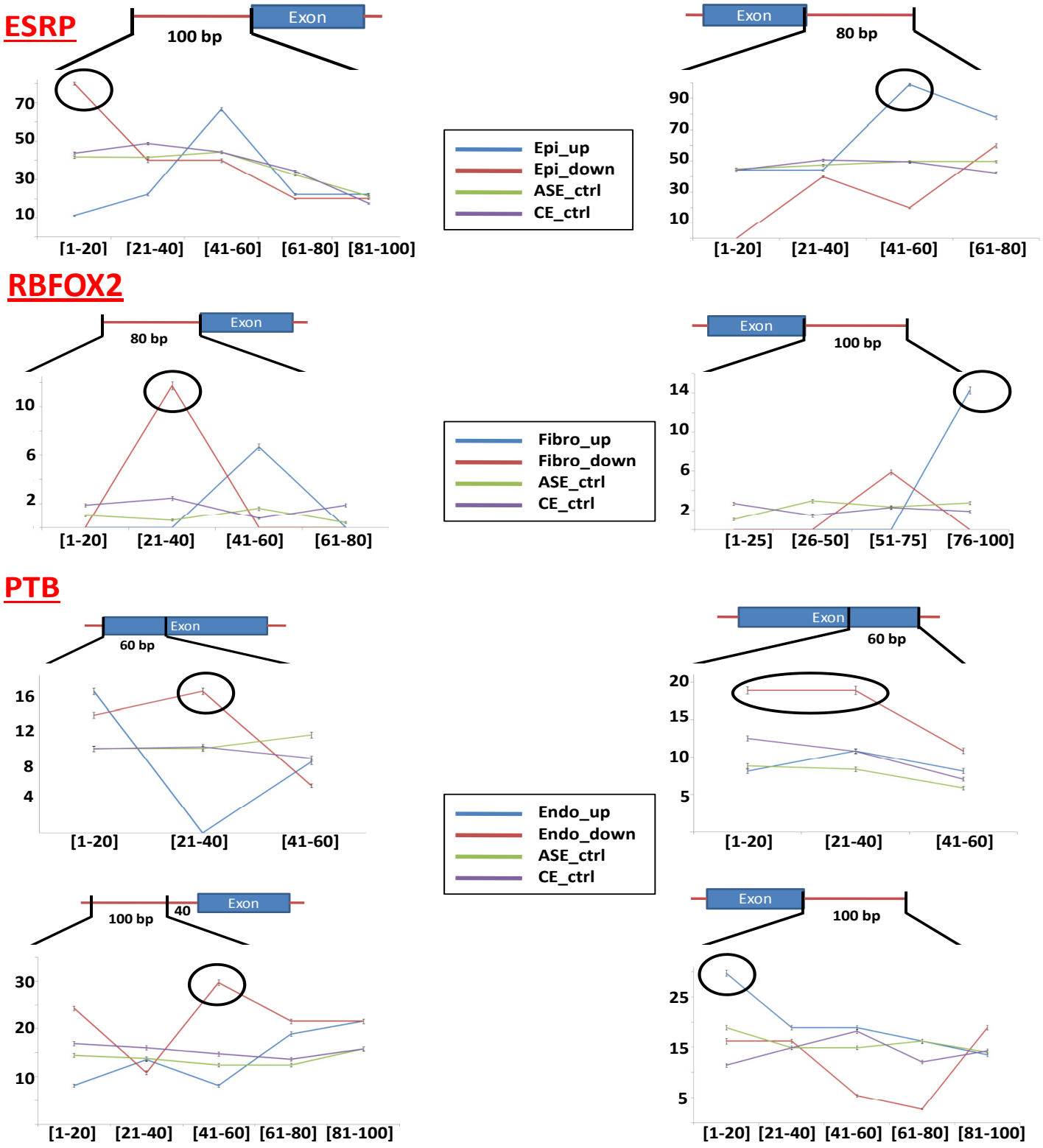
_ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799-816.

_ENCODE Project Consortium. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9**: e1001046. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**: 46-53.

_Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

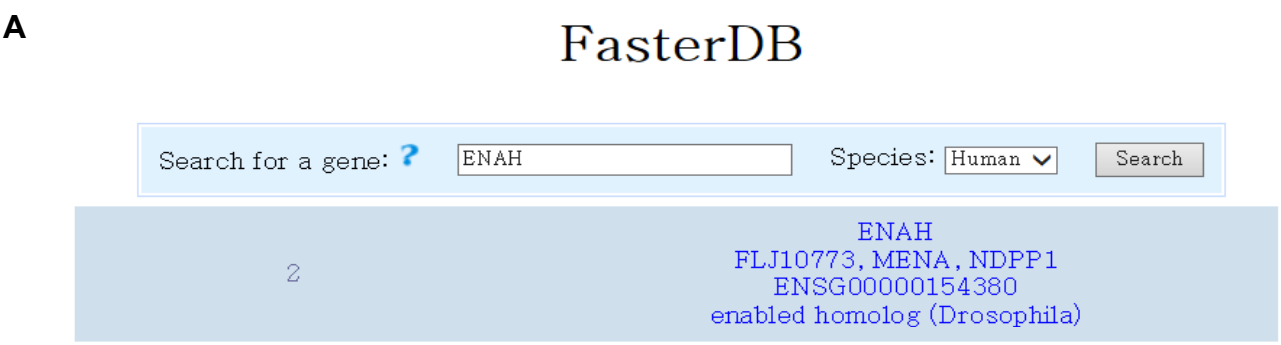
_Kapushesky M, Adamusiak T, Burdett T, Culhane A, Farne A, Filippov A, Holloway E, Klebanov A, Kryvych N, Kurbatova N, et al. 2012. Gene Expression Atlas update--a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic Acids Res* **40**:D1077-1081.

_Oudes AJ, Campbell DS, Sorensen CM, Walashek LS, True LD, Liu AY. 2006. Transcriptomes of human prostate cells. *BMC Genomics* **7**: 92.

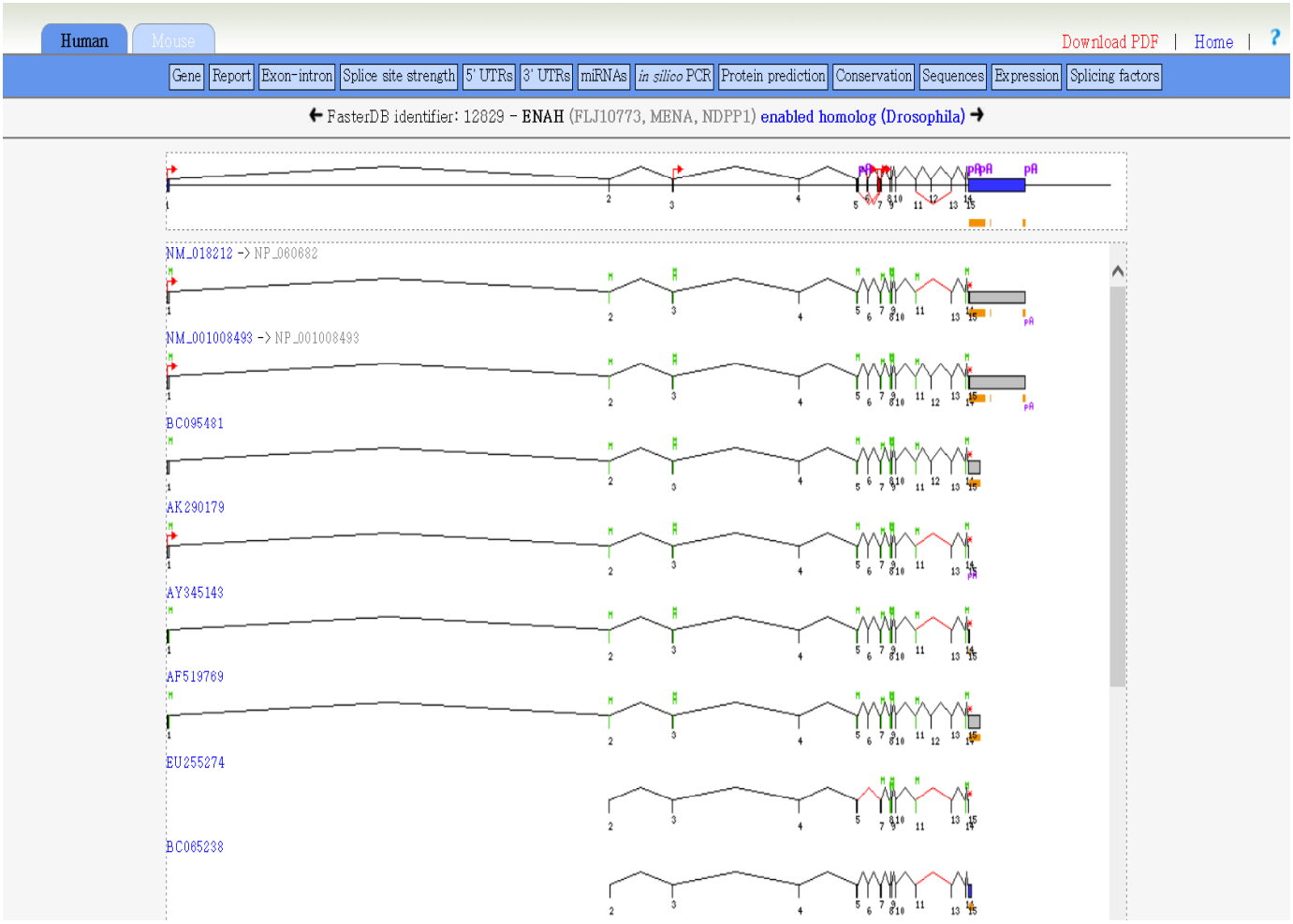


Supplementary Figure S9.
Splicing factor binding sites were searched using PatSearch Tool on the genomic sequences using previously defined splicing factor binding motifs (see Supplementary Table S8). A set of 1000 randomly selected alternative exons and a set of 1000 randomly selected constitutive exons were used as controls. Four regions were defined: 100 nt upstream and downstream of the exon, the first 60 nts of the exon and the last 60 nts of the exon. For each region, a sliding window of a specific length was considered and the enrichment score of the splicing factor was computed at each position as : (∑number of factor motifs at position *X*/Total number of analyzed sequences) * 100. The total number of binding sites found in each region of each exon group was used to compute the standard deviation value.

A



B



Supplementary Figure S10. FasterDB search and main pages corresponding to the human *ENAH* gene.
A. The search engine is based on keywords (e.g. Gene Symbol, ENSEMBL ID, HUGO gene name, etc.).
B. *ENAH* gene main page in FasterDB. Alternative splicing events are depicted by red broken lines, and microRNAs binding sites are depicted by orange under the gene structure. Known transcripts are schematically represented below the gene structure. Several buttons in the toolbar allow the user to get additional information (see below). The main page for the mouse orthologous gene can be displayed by clicking on the “Mouse” link.

A

Gene	
Symbol:	ENAH
Synonyms:	FLJ10773, MENA, NDPP1
Description:	enabled homolog (Drosophila)
EnsEMBL link:	ENSG00000154380
Chromosome:	1 - Reverse Strand
Chromosomal location: (UCSC link)	225674534 - 225840845 (166312bp)
Sequence:	sequence
Associated transcripts:	NM_018212 NM_001008493 BC095481 AK290179 AY345143 AF519769 EU255274 BC065238 BC010414 AK096246 AK126894 AK127561 AK001635 AL133059

B

Exon-Intron						
Position	**Type**	**Gene coordinates**	**Chromosomal coordinates**	**Length**	**Sequence**	**Features**
1	Exon	1 - 458	225840388-225840845	458	sequence	download
1	Intron	459 - 85729	225755117-225840387	85271	sequence	-
2	Exon	85730 - 85895	225754951-225755116	166	sequence	download
2	Intron	85896 - 98060	225742786-225754950	12165	sequence	-
3	Exon	98061 - 98238	225742608-225742785	178	sequence	download
3	Intron	98239 - 122505	225718341-225742607	24267	sequence	-
4	Exon	122506 - 122590	225718256-225718340	85	sequence	download
4	Intron	122591 - 133578	225707268-225718255	10988	sequence	-
5	Exon	133579 - 133946	225706900-225707267	368	sequence	download
5	Intron	133947 - 135837	225705009-225706899	1891	sequence	-
6	Exon	135838 - 135948	225704898-225705008	111	sequence	download
6	Intron	135949 - 138243	225702603-225704897	2295	sequence	-
7	Exon	138244 - 138548	225702298-225702602	305	sequence	download
7	Intron	138549 - 140127	225700719-225702297	1579	sequence	-
8	Exon	140128 - 140273	225700573-225700718	146	sequence	download
8	Intron	140274 - 140429	225700417-225700572	156	sequence	-
9	Exon	140430 - 140487	225700359-225700416	58	sequence	download
9	Intron	140488 - 141284	225699562-225700358	797	sequence	-
10	Exon	141285 - 141333	225699513-225699561	49	sequence	download
10	Intron	141334 - 145126	225695720-225699512	3793	sequence	-
11	Exon	145127 - 145193	225695653-225695719	67	sequence	download
11	Intron	145194 - 148090	225692756-225695652	2897	sequence	-
12	Exon	148091 - 148153	225692693-225692755	63	sequence	download
12	Intron	148154 - 152073	225688773-225692692	3920	sequence	-
13	Exon	152074 - 152152	225688694-225688772	79	sequence	download
13	Intron	152153 - 154739	225686107-225688693	2587	sequence	-
14	Exon	154740 - 154797	225686049-225686106	58	sequence	download
14	Intron	154798 - 155331	225685515-225686048	534	sequence	-
15	Exon	155332 - 166312	225674534-225685514	10981	sequence	download

Supplementary Figure S11. Gene (A) and Exon-Intron (B) tables. **A-** The gene table shows general information about the selected gene (official symbol, synonyms, description, chromosomal location and associated transcripts). Links to other resources are also available: clicking on the Ensembl ID of the gene leads to Ensembl while clicking on chromosomal coordinates directs to the UCSC page of the gene. Finally, clicking on sequence displays the sequences of each of the exons and introns of the gene. **B-** Summary table of the exon/intron structure of the gene. For each element, gene and chromosomal coordinates are displayed as well as its length. Clicking on sequence link displays the nucleotidic sequence of the element. For exons only, clicking on download link in the features column results in a file containing further information about the exon of interest.

A

Report	
Exon	**Number of supporting transcripts**
Transcription initiation & first exon(s)	
1	2
3	1
6	1
7	2
Transcription termination & last exon(s)	
5	1
15	6
Exon skipping	
6	1
12	9
Alternative 3' splice sites	
7	1
Alternative 5' splice sites	
6	1

B

Polyadenylation							
Identifier	**Position**	**Gene coordinates of cleavage site**	**Chromosomal coordinates of cleavage site**	**Pattern**	**PolyA tail**	**Accession number**	**Type**
PA1	5	134098	225706748	AATAAA	yes	BC010414	Cdna
PA2	15	155383	225685463		no	AK290179	Full_length
PA3	15	157666	225683180		no	AK096246	Full_length
PA4	15	157666	225683180		no	AK126894	Full_length
PA5	15	157666	225683180		no	AK001635	Full_length
PA6	15	166312	225674534	AATAAA	no	NM_018212	Cdna
PA7	15	166312	225674534	AATAAA	no	NM_001008493	Cdna

C

Exon skipping		
Identifier	**Position**	**Transcript**
ES1	6	EU255274
ES2	12	NM_018212 AK290179 AY345143 EU255274 BC065238 AK096246 AK126894 AK001635 AL133059

Supplementary Figure S12. Report (A), polyadenylation (B) and Exon skipping (C) tables. **A-** Features relative to alternative splicing events are summarized in the report table. For each type of event, the exon concerned and the number of transcripts having this event (number of supporting transcripts) are displayed. Clicking on a given event displays further information about it. Two examples of transcription end and exon skipping events tables are shown in **B** and **C** respectively. **B-** The "Transcription termination & last exon(s)" link shows polyadenylation sites. For each polyadenylation site, we display its exonic position, gene and chromosomal coordinates. Moreover information about the signal pattern, the accession number and the type of sequence (cDNa/Est) of the transcript used to define the polyadenylation site are displayed. Finally, the presence of a polyA tail is tested and notified when present in the sequence. **C-** The "Exon Skipping" table shows information about skipped exons displaying their position and the transcripts that have been used to make this annotation.

Splice site strength					
Identifier	**Site**	**Exon**	**Strength**	**Alternative**	**Sequence**
S1	5' (donor)	1	8.41	no	GAGgtgagg
S2	5' (donor)	2	11.08	no	CAGgtaagg
S3	3' (acceptor)	2	8.59	no	ttttattttattttttacagTGA
S4	5' (donor)	3	11.08	no	CAGgtaagg
S5	3' (acceptor)	3	6.9	no	aaagaattattatttttttagGTC
S6	5' (donor)	4	8.88	no	AAGgtaaat
S7	3' (acceptor)	4	11.04	no	aatgttttctccctctgtagGGC
S8	5' (donor)	5	9.45	no	CTGgtaaga
S9	3' (acceptor)	5	11.06	no	ttttttttttttttttccagACA
S10	5' (donor)	6	9.16	no	AGGgtaagg
S11	3' (acceptor)	6	12.94	no	tctcccttctccctctgcagCTG
S12	5' (donor)	6	-31.91	yes	CTCctctaa
S13	5' (donor)	7	6.15	no	CGGgtaaat
S14	3' (acceptor)	7	10.84	no	ttaacatttctctctctcagGCA
S15	3' (acceptor)	7	-5.83	yes	accgccccctcctcccccctcCTC
S16	5' (donor)	8	10.28	no	GAGgtaagg
S17	3' (acceptor)	8	4.63	no	gtctgtatattctcattaagATG
S18	5' (donor)	9	8.07	no	GGTgtaagt
S19	3' (acceptor)	9	7.13	no	cggatttgccttctttgtagGAG
S20	5' (donor)	10	10.65	no	CTGgtaagt
S21	3' (acceptor)	10	7.02	no	atatatatctgtatatttagGAA
S22	5' (donor)	11	7.33	no	CAGgtatca
S23	3' (acceptor)	11	6.54	no	tttgtttttttatttttaaagAAC
S24	5' (donor)	12	7.33	no	CAGgtatca
S25	3' (acceptor)	12	8.22	no	aatggttttccccttcatagACG
S26	5' (donor)	13	8.56	no	CAGgtgggt
S27	3' (acceptor)	13	4.74	no	tggaaatctcatttatacagACC
S28	5' (donor)	14	11.01	no	ATGgtaagt
S29	3' (acceptor)	14	10.55	no	atttttctctaatattttagGAC
S30	3' (acceptor)	15	8.11	no	acttgttcttttccatccagCAA

Supplementary Figure S13. Splice site strength table. All the splice sites of the *ENAH* gene are showed in this table. For each splice site, we display its type (donor or acceptor), its strength according to MaxEntScan computation method and if the site is alternative or not. Furthermore, the sequence (9 and 23 nucleotides for donor and acceptor sites respectively) is showed with capital letters corresponding to the exonic part of the site while intronic part is represented in lowercase letters.

A

	5' untranslated regions				
	UTR 1	**UTR 2**	**UTR 3**	**UTR 4**	**UTR 5**
Supporting GenBank accession identifers	NM_018212 NM_001008493 BC095481 AY345143 AF519769	AK290179	EU255274	AK127561	AK126894
Length	453	397	1029	727	353
Number of ATGs in frame	1	1	4	1	1
Number total of ATGs	1	1	14	10	2
Number of GTGs in frame	3	1	12	3	2
Number total of GTGs	10	8	24	13	13
Number of CTGs in frame	1	1	7	0	2
Number total of CTGs	5	5	14	2	5
Number of micro ORFs in frame	1	1	1	1	1
Number total of micro ORFs	1	1	1	7	1
GC percentage	76	76	54	42	58
Number of pyrimidine tracks	5	4	6	10	9

B

	3' untranslated regions
	UTR 1
Supporting GenBank accession identifers	NM_018212 NM_001008493 AL133059 BC095481 AK096246 AK126894 AK001635 AF519769 AK127561 BC065238 AY345143
Length	10943
Number of ATGs in frame	66
Number total of ATGs	197
Number of GTGs in frame	59
Number total of GTGs	207
Number of CTGs in frame	54
Number total of CTGs	154
GC percentage	38
Number of pyrimidine tracks	78

Supplementary Figure S14. 5'UTR (A) and 3'UTR (B) tables. **A-** Several informations about each of the 5' Untranslated regions (UTR) are given. For each UTR, the subset of transcripts having this UTR is displayed as well as the number of motifs ATG, GTG and CTG found either globally in the UTR sequence or in frame. The length and the GC content are also computed for each UTR. Finally, the number of pyrimidine tracks as well as the number of Micro ORF is displayed. **B-** Same information as in **A** for the 3'UTR region.

miRNAs				
Identifier	**Name**	**Gene coordinates**	**Chromosomal coordinates**	**Prediction algorithm**
Group 1				
M1	hsa-miR-600	155464 - 155470	225685376 - 225685382	pita
M2	hsa-miR-136	155475 - 155482	225685364 - 225685371	pita
M3	hsa-miR-320	155547 - 155554	225685292 - 225685299	pictar
M4	hsa-miR-545	155564 - 155571	225685275 - 225685282	pita
M5	hsa-miR-16	155566 - 155573	225685273 - 225685280	pictar
M6	hsa-miR-103	155566 - 155573	225685273 - 225685280	pictar
M7	hsa-miR-15b	155566 - 155573	225685273 - 225685280	pictar
M8	hsa-miR-195	155566 - 155573	225685273 - 225685280	pictar
M9	hsa-miR-424	155566 - 155573	225685273 - 225685280	targetscan
M10	hsa-miR-424	155566 - 155573	225685273 - 225685280	pita
M11	hsa-miR-107	155566 - 155573	225685273 - 225685280	pictar
M12	hsa-miR-195	155566 - 155573	225685273 - 225685280	targetscan
M13	hsa-miR-15b	155566 - 155573	225685273 - 225685280	targetscan
M14	hsa-miR-15a	155566 - 155573	225685273 - 225685280	pictar
M15	hsa-miR-497	155566 - 155573	225685273 - 225685280	targetscan
Group 2				
M1	hsa-miR-141	157688 - 157695	225683151 - 225683158	targetscan
M2	hsa-miR-200a	157688 - 157695	225683151 - 225683158	targetscan
M3	hsa-miR-200a	157689 - 157695	225683151 - 225683157	pita
M4	hsa-miR-141	157689 - 157695	225683151 - 225683157	pita
M5	hsa-miR-421	157691 - 157698	225683148 - 225683155	targetscan
M6	hsa-miR-374a	157761 - 157767	225683079 - 225683085	pita
M7	hsa-miR-374b	157761 - 157767	225683079 - 225683085	pita
M8	hsa-miR-216b	157804 - 157811	225683035 - 225683042	targetscan
M9	hsa-miR-216b	157805 - 157811	225683035 - 225683041	pita
M10	hsa-miR-548d-3p	157891 - 157898	225682948 - 225682955	pita
M11	hsa-miR-335	157900 - 157907	225682939 - 225682946	targetscan
M12	hsa-miR-335	157900 - 157907	225682939 - 225682946	pita
M13	hsa-miR-641	157916 - 157922	225682924 - 225682930	pita
M14	hsa-miR-208b	157917 - 157924	225682922 - 225682929	targetscan
M15	hsa-miR-208a	157917 - 157924	225682922 - 225682929	targetscan
M16	hsa-miR-499-5p	157917 - 157924	225682922 - 225682929	targetscan
M17	hsa-miR-512-5p	157925 - 157931	225682915 - 225682921	pita
M18	hsa-miR-548g	157931 - 157938	225682908 - 225682915	pita
M19	hsa-miR-128	157980 - 157988	225682858 - 225682866	targetscan
M20	hsa-miR-27a	157981 - 157988	225682858 - 225682865	targetscan

Supplementary Figure S15. Micro RNA table. This table summarizes information about micro RNAs binding sites found among the *ENAH* gene sequence. For each miRNA, we display its ID, gene and chromosomal coordinates, and the algorithm which allowed predicting this miRNA (Miranda, pita, targetscan). When a polyadenylation site is present between miRNA binding sites, a blue bar is displayed between miRNA binding sites because the binding of the different groups of miRNAs could be affected by alternative polyadenylation usage. (NB: Group1 and Group2 are not complete but were cropped in order to get both on the same page).

Conserved exons		
Gene identifier mouse	**Position Human**	**Position mouse**
5106	2	2
5106	3	3
5106	4	5
5106	6	7
5106	8	9
5106	9	10
5106	10	11
5106	11	12
5106	13	13
5106	14	14
5106	15	15
5106	15	16

Supplementary Figure S16. Conservation table. For each human gene, we display, in the first column, the FasterDB id of its orthologous in the mouse genome (as provided by Ensembl). Furthermore, we display for each human exon its orthologous, if exists, in the mouse genome using the FasterDB numbering.

A

Expression			
Exon position	**Tissues**	**Cell lines**	**Cancer cell lines**
1	Expression	Expression	Expression
2	Expression	Expression	Expression
3	Expression	Expression	Expression
4	Expression	Expression	Expression
5	Expression	Expression	Expression
6	Expression	Expression	Expression
7	Expression	Expression	Expression
8	Expression	Expression	Expression
9	Expression	Expression	Expression
10	Expression	Expression	Expression
11	Expression	Expression	Expression
12	Expression	Expression	Expression
13	Expression	Expression	Expression
14	Expression	Expression	Expression
15	Expression	Expression	Expression

B

Experiments	Expressed	Gene expression level	NI	Global SI	Local SI	p-value
HEK293	Yes	1637	1.40	10.28	3.89	0.0001
hESCT0	Yes	1637	1.49	10.92	4.93	0.0000
Medullo	Yes	1365	-7.14	1.03	-2.86	0.0006
NHA	Yes	1357	-10.00	-1.35	-4.35	0.0003
HConF	Yes	934	-10.00	-1.43	-4.17	0.0002
HPAF	Yes	911	-14.29	-1.96	-6.25	0.0006
Fibrobl	No	893	-25.00	-3.45	-7.14	0.0003
LNCAP	Yes	889	1.23	9.04	2.88	0.0001
HVMF	Yes	841	-14.29	-2.04	-6.67	0.0003
AoAF	Yes	836	-7.14	1.02	-3.70	0.0003
HMEC	Yes	487	-1.52	4.82	2.31	0.0009
BJ	No	476	-16.67	-2.22	-5.00	0.0006
HPAEC	Yes	470	-10.00	-1.32	-3.85	0.0155
HIPEpiC	Yes	466	-9.09	-1.19	-2.13	0.0096
MCF7	Yes	463	-5.00	1.45	-1.12	0.2966
SKNS	Yes	462	-7.14	1.06	-3.33	0.0010
HRCE	Yes	436	-1.32	5.61	1.88	0.0064
HAEpiC	Yes	425	-7.69	-1.05	-1.35	0.0254
HCF	No	413	-16.67	-2.22	-2.08	0.0041
HMVEC_dAd	Yes	409	-12.50	-1.67	-3.23	0.0008
HCPEpiC	Yes	403	-10.00	-1.43	-2.33	0.0019
AG09309	No	403	-9.09	-1.20	-1.92	0.0040
NHBE	No	384	-1.28	5.75	2.63	0.0001
NHDFneo	No	337	-6.67	1.09	-1.85	0.0228
JURKAT	No	210	-5.88	1.26	1.05	0.4361
NB4	No	178	-7.69	-1.06	-1.02	0.4624
TH2	No	158	-7.69	-1.06	0.00	0.0000
K562	No	156	-7.69	-1.04	-1.09	0.1798
GM12878	No	130	-6.25	1.19	0.00	0.0000
TH1	No	130	-6.67	1.13	0.00	0.0000
HLAF	No	128	-6.25	1.18	1.50	0.0419
GM69	No	107	-5.26	1.41	0.00	0.0000

Supplementary Figure S17.

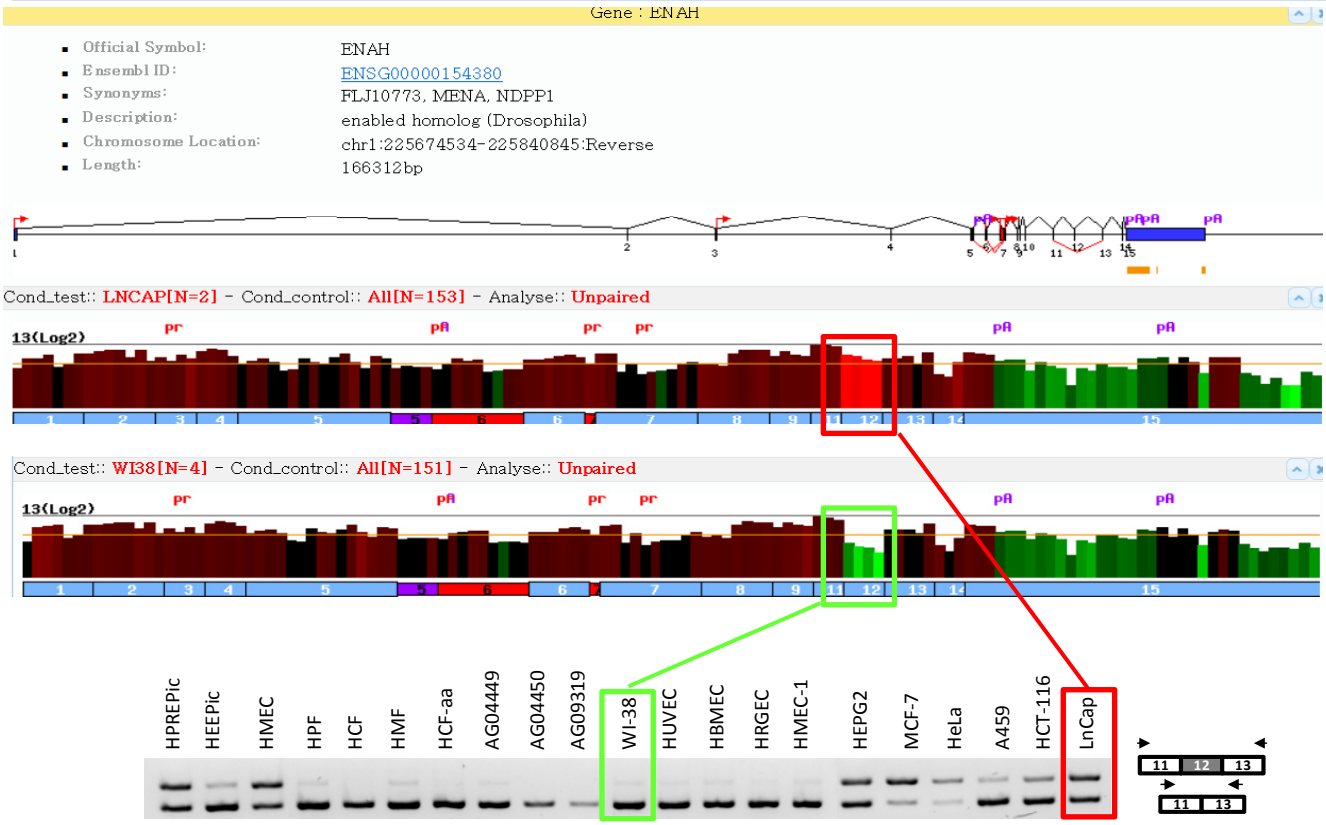
A. Expression table. In this table, we display three links: the first one leads to a table summarizing the expression of the exon in the tissues (column “Tissue”), the second one displays a table summarizing exon expression in the cell lines (column “Cell lines”) while the third link displays exon expression in cancer cell lines only (column “Cancer cell lines”). See Material and Methods for more information about tissues and cell lines included in these analyses.

B. Clicking on any “Expression” link in the “Cell lines” column for example, allows users to know the relative expression level of the gene across the cell line collection based on Encode Exon Array datasets. For each cell line, we computed gene expression level which represents the median of the intensity of all the probes associated to the exons of the gene and determine if the gene is expressed or not. This also allow to know which cell expresses a high or low level of the selected gene. For example, the *ENAH* gene expression level is higher in LnCAP than in GM69 cells.

A

Experiments	Expressed	Gene expression level	NI	Global SI	Local SI	p-value
hESCT0	Yes	1637	1.49	10.92	4.93	0.0000
HEK293	Yes	1637	1.40	10.28	3.89	0.0001
CACO2	Yes	558	1.17	8.64	3.47	0.0001
PrEC	Yes	795	1.21	8.93	3.34	0.0000
LNCAP	Yes	889	1.23	9.04	2.88	0.0001
NHBE	No	384	-1.28	5.75	2.63	0.0001
SAEC	Yes	624	-1.15	6.40	2.37	0.0004
HMEC	Yes	487	-1.52	4.82	2.31	0.0009
HUVEC	Yes	758	-10.00	-1.37	-3.57	0.0014
AoAF	Yes	836	-7.14	1.02	-3.70	0.0003
HPAEC	Yes	470	-10.00	-1.32	-3.85	0.0155
AG04449	Yes	1017	-16.67	-2.44	-4.00	0.0020
WI38	Yes	803	-14.29	-1.89	-4.17	0.0022
HConF	Yes	934	-10.00	-1.43	-4.17	0.0002
NHA	Yes	1357	-10.00	-1.35	-4.35	0.0003

B



Supplementary Figure S18. Exon array visualization web interface.

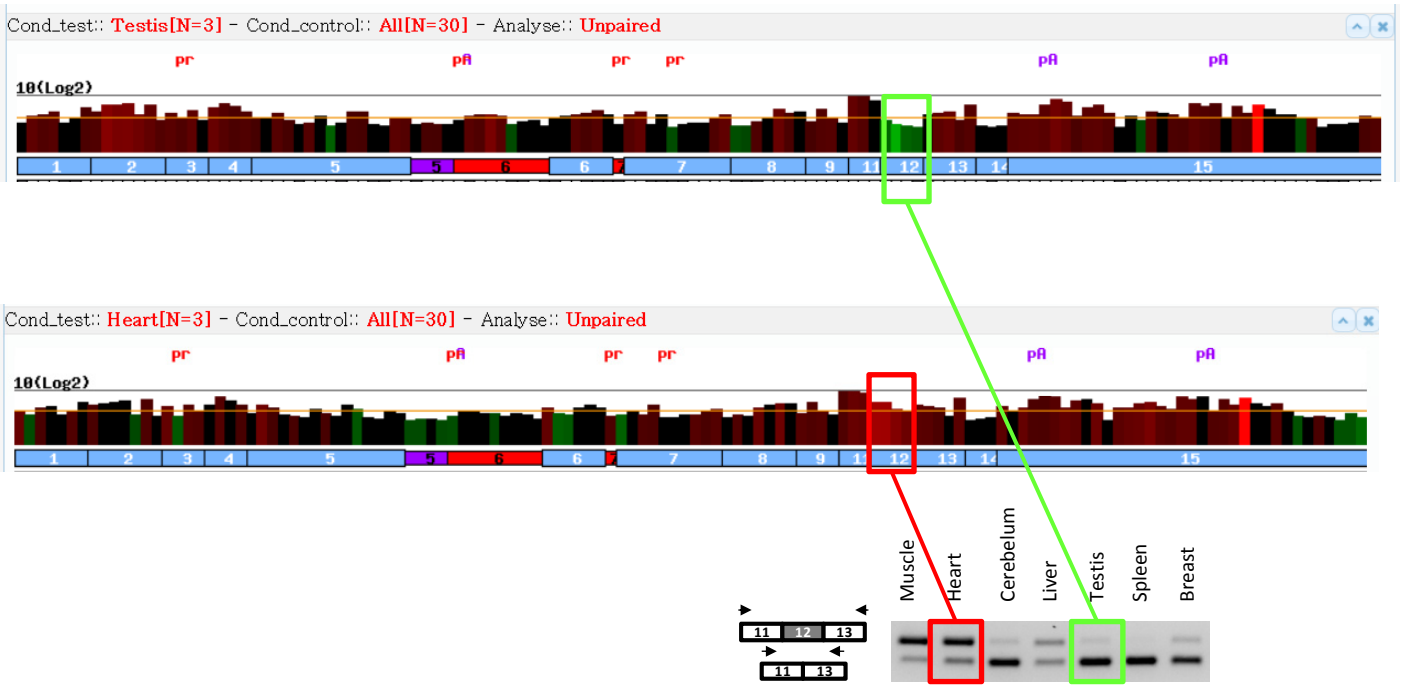
A. Inclusion rate of *ENAH* exon 12 across human cell lines. Clicking on "Expression" button corresponding to one exon (e.g. ENAH exon 12, **Supplementary Figure S17A**) generate a table containing the "Ni" (exon normalized intensity), "Global SI", and "Local SI" that are three different ways to calculate the inclusion rate of the selected exon in each sample. Higher "Global SI" or "Local SI" positive values indicate increased inclusion rates of the selected exon, while negative values indicate high exclusion rates. For instance, the ENAH exon 12 had high Global SI and Local SI values in the LnCAP cell line but negative ones in WI38 cells, suggesting that it has a high inclusion rate in LnCAP cells as compared to WI38 cells .

B. Clicking on any cell line open the ELEXIR interface allowing to Visualize exon array probe intensities in the selected cell line compared to all other cells (as shown here, LnCAP and WI38 cell lines as compared to other cells). Each rectangle corresponds to one probe above the corresponding exon. The height of each probe corresponds to its intensity level. Red probes indicate that the test condition had a higher intensity than the control condition, while green probes indicate a lower intensity than the control condition. Black probes indicate no difference between the conditions. As shown in Supplementary Figure S18B, most of the probes corresponding to the *ENAH* gene are in red, indicating that ENAH is more expressed in LnCAP cells as compared to most other cell lines (Supplementary Fig. S17). In addition, in agreement with the high SI values computed in LnCAP cells (Supplementary Fig. S18A), *ENAH* exon 12 probes are more intensively red than probes targeting other gene regions, suggesting that the ENAH exon 12 is more included in LnCAP cells as compared to other cells. Meanwhile, in agreement with low SI values calculated in WI38 cells (Supplementary Fig. S18A), exon 12 probes were green when comparing WI38 to other cells, indicating that exon 12 was skipped in WI38 cells (Supplementary Fig. S18B). These results were validated by RT-PCR (Supplementary Fig. S18B, lower panel).

A

Experiments	Expressed	Gene expression level	NI	Global SI	Local SI	p-value
Testis	Yes	90	-3.57	-1.54	-2.44	0.0031
Spleen	Yes	86	-4.55	-1.92	-2.00	0.0253
Breast	No	81	-3.33	-1.45	-1.61	0.0038
Prostate	No	90	-2.86	-1.22	-1.32	0.0266
Cerebellum	Yes	52	-2.63	-1.09	-1.14	0.2873
Liver	No	53	-2.13	1.26	1.19	0.0326
Muscle	No	103	-1.75	1.52	1.23	0.0785
Pancreas	No	80	-1.75	1.51	1.23	0.1171
Kidney	No	59	-2.78	-1.18	1.24	0.2472
Thyroid	No	95	-1.19	2.23	1.58	0.0017
Heart	No	108	-1.14	2.36	1.62	0.0109

B

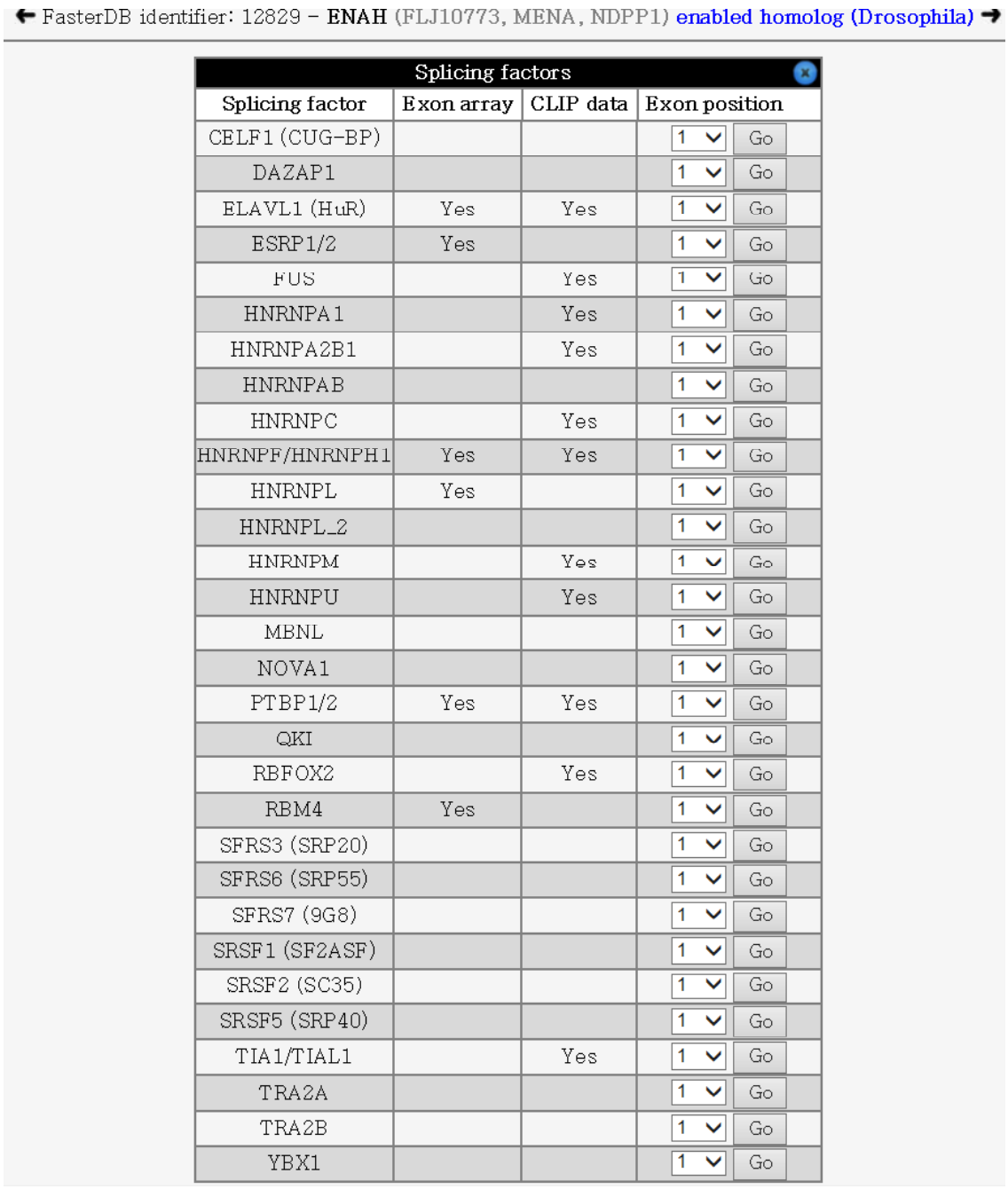


Supplementary Figure S19 *ENAH* exon 12 was also predicted and validated to be more included in human testis as compared to heart.

A. Inclusion rate of *ENAH* exon 12 across human tissues. For each tissue, we compute Normalized Intensity (NI), Global Splicing Index (SI) and Local Splicing index (Si) values which represent different methods to calculate the inclusion level of an exon. Finally, we show the p-value of the student test based on the local Si values. The table can be sorted according to any of the columns. This table indicate that *ENAH* exon 12 gives rise to higher global or local SI values in the heart than in the testis, suggesting that *ENAH* exon 12 is included in heart but not in testis.

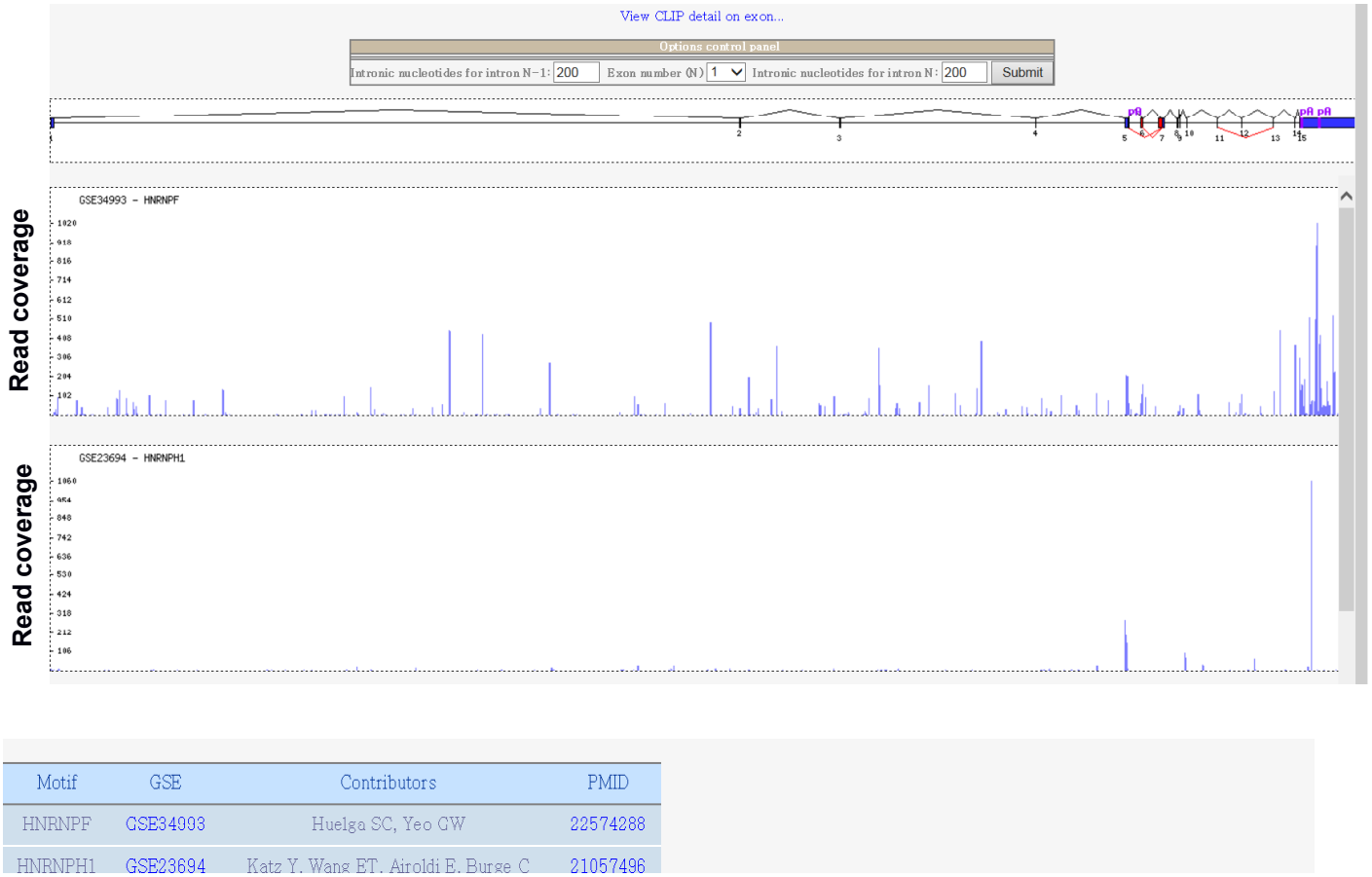
B. Visualization of exon array probe intensities in Testis and Heart compared to other tissues. The height of each probe corresponds to its intensity level. Red probes indicate inclusion in test condition while green probes indicate its exclusion when compared to controls. Black probes point no difference between both conditions. *ENAH* exon 12 probes are green in testis and red in heart suggesting that *ENAH* exon 12 is included in heart but not in testis as validated by RT-PCR.

.



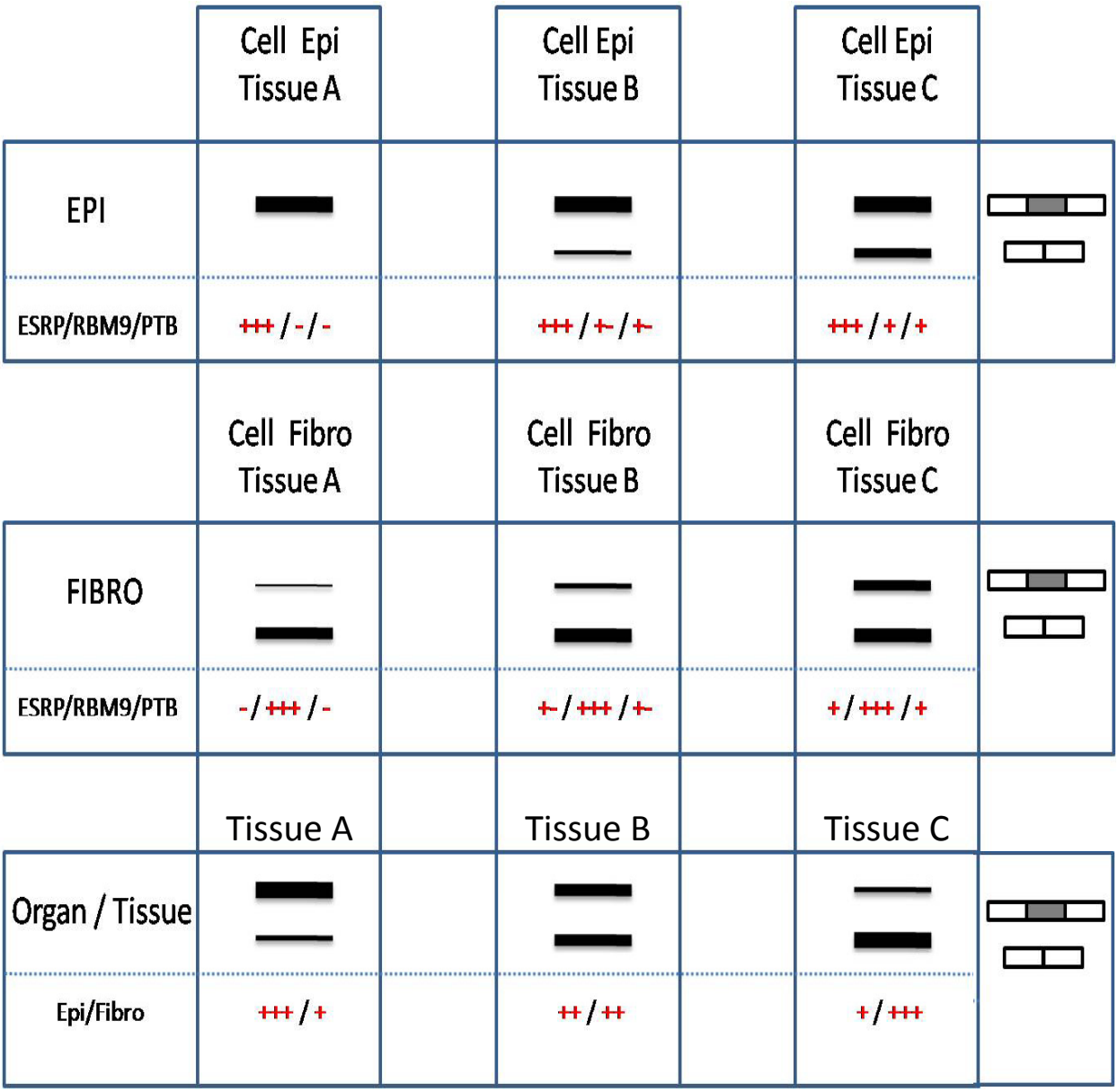
Splicing factor	Exon array	CLIP data	Exon position
CELF1 (CUG-BP)			1 Go
DAZAP1			1 Go
ELAVL1 (HuR)	Yes	Yes	1 Go
ESRP1/2	Yes		1 Go
FUS		Yes	1 Go
HNRNPA1		Yes	1 Go
HNRNPA2B1		Yes	1 Go
HNRNPAB			1 Go
HNRNPC		Yes	1 Go
HNRNPF/HNRNPH1	Yes	Yes	1 Go
HNRNPL	Yes		1 Go
HNRNPL_2			1 Go
HNRNPM		Yes	1 Go
HNRNPU		Yes	1 Go
MBNL			1 Go
NOVA1			1 Go
PTBP1/2	Yes	Yes	1 Go
QKI			1 Go
RBFOX2		Yes	1 Go
RBM4	Yes		1 Go
SFRS3 (SRP20)			1 Go
SFRS6 (SRP55)			1 Go
SFRS7 (9G8)			1 Go
SRSF1 (SF2ASF)			1 Go
SRSF2 (SC35)			1 Go
SRSF5 (SRP40)			1 Go
TIA1/TIAL1		Yes	1 Go
TRA2A			1 Go
TRA2B			1 Go
YBX1			1 Go

Supplementary Figure S20. Splicing factor table allows to get for each selected exon, the motifs found for a list of splicing factors in the exon and 200 nucleotides in the flanking introns. This table indicates whether 'Exon array'' and ''CLIP data'' datasets are available for each splicing factor. Clicking on "Go" allows then to get further information as detailed d in the main manuscript (see Supplementary Fig. S21).



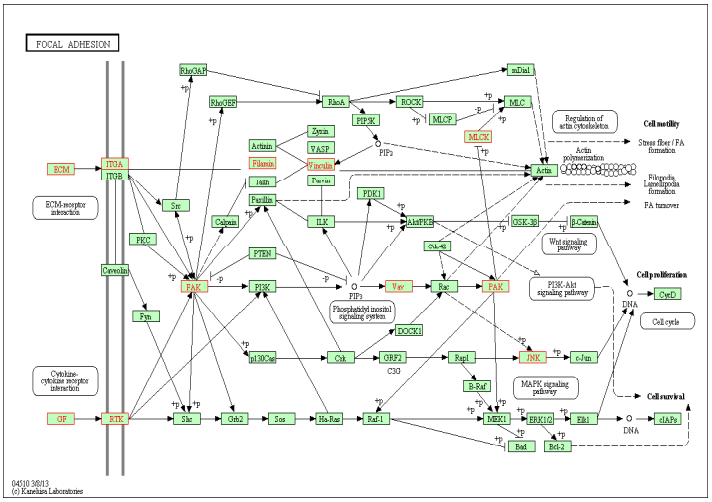
Motif	GSE	Contributors	PMID
HNRNPF	GSE34993	Huelga SC, Yeo GW	22574288
HNRNPH1	GSE23694	Katz Y, Wang ET, Airoldi E, Burge C	21057496

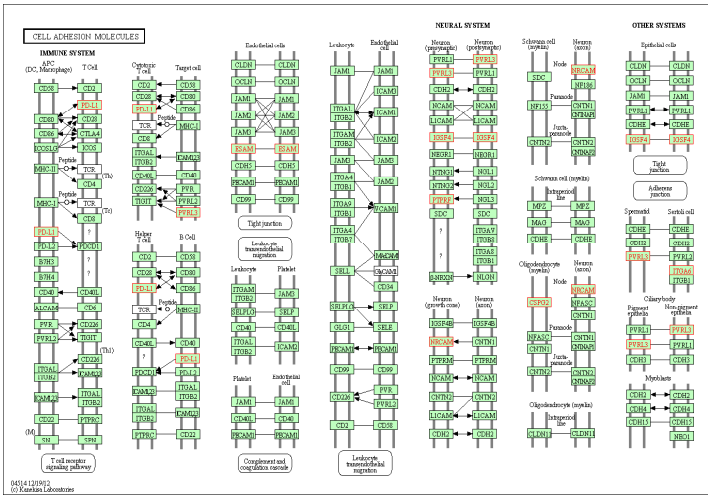
Supplementary Figure S21: CLIP dataset visualization web interface.
After selecting a splicing factor in the "Splicing Factor" table (see Supplementary Figure S20), users can visualize the read coverage of the selected splicing factor in CLIP-seq experiments referenced in the bottom screen. This allows to determine in which regions of the selected gene are enriched for the binding sites of the selected splicing factor. Selecting Exon and indicating the window in the flanking introns in the "Option control panel" allows to zoom in on the selected exon as described in the main manuscript (Fig. 5B).

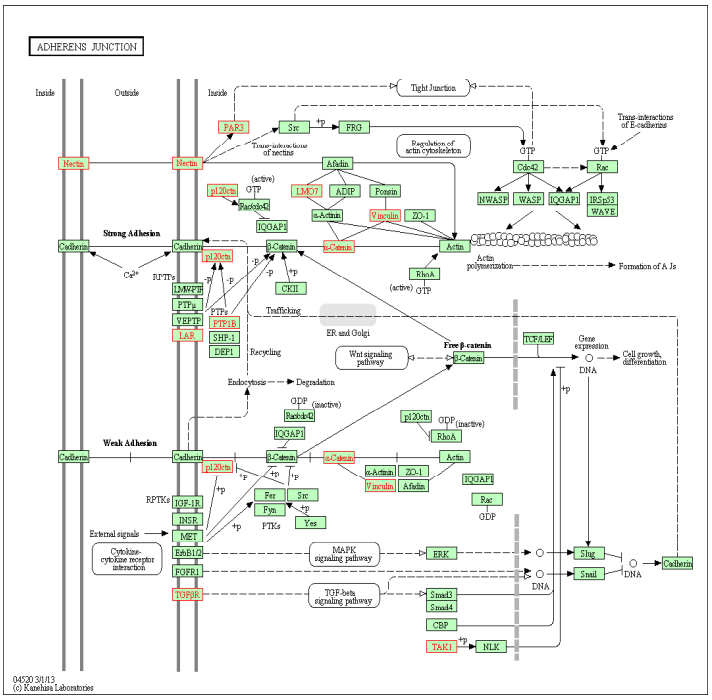


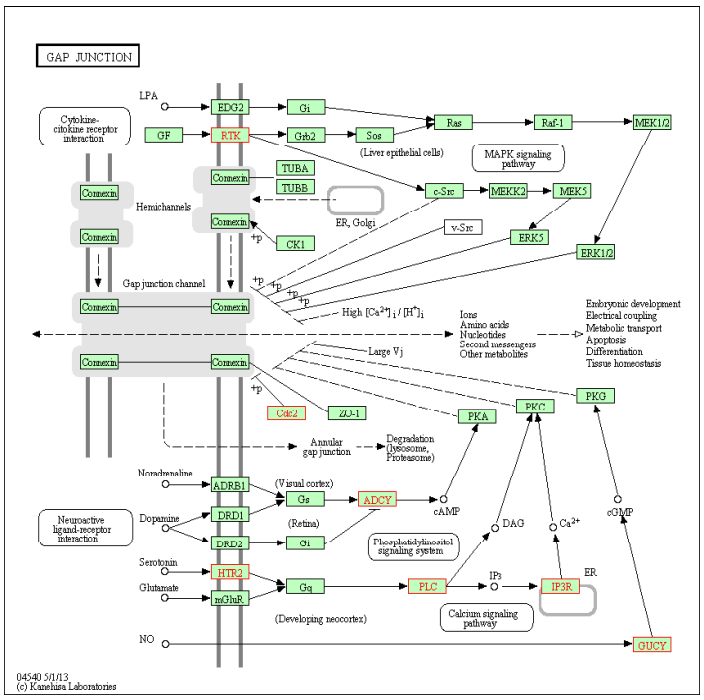
Supplementary Figure S22.
We demonstrate that each major cell type (e.g. epithelial cells or fibroblasts) expresses a specific splicing pattern owing to the differential expression of splicing factors (comparing "EPI" to "FIBRO"). However, depending on the tissue origin, the ratio between the expression level of the different splicing factors may slightly changes leading to cell type- and tissue-specific splicing pattern (e.g. comparing Tissue A, B and C in "EPI" and Tissue A, B and C in "FIBRO"). In addition, the splicing pattern observed in an organ or a tissue that contain different cell populations could reflect a different composition of different cell types as schematized in the "Organ/Tissue" panel.

Sublist	Category	Term	RT	Genes	Count	%	P-Value
☐	KEGG_PATHWAY	Calcium signaling pathway	RT		17	2,2	3,8E-3
☐	KEGG_PATHWAY	Adherens junction	RT		10	1,3	5,8E-3
☐	KEGG_PATHWAY	Endocytosis	RT		17	2,2	5,9E-3
☐	KEGG_PATHWAY	Phosphatidylinositol signaling system	RT		8	1,0	4,1E-2
☐	KEGG_PATHWAY	MAPK signaling pathway	RT		19	2,4	4,1E-2
☐	KEGG_PATHWAY	Focal adhesion	RT		15	1,9	5,3E-2
☐	KEGG_PATHWAY	mTOR signaling pathway	RT		6	0,8	7,4E-2
☐	KEGG_PATHWAY	Complement and coagulation cascades	RT		7	0,9	7,8E-2
☐	KEGG_PATHWAY	ErbB signaling pathway	RT		8	1,0	8,3E-2
☐	KEGG_PATHWAY	Gap junction	RT		8	1,0	9,1E-2









Supplementary Figure S23.
Several genes that are differentially spliced across fibroblasts, epithelial and endothelial cells are involved in adherens or gap junctions and in focal and cell adhesion as predicted by KEGG_PATWAY enrichment analysis performed on the "DAVID ontology" web site (http://david.abcc.ncifcrf.gov/). Genes that are differentially spliced across fibroblasts, epithelial and endothelial cells are indicated in red in the KEGG Pathways.