Supporting Online Material for

Regulation of Alternative Splicing by Histone Modifications

Reini F. Luco, Qun Pan, Kaoru Tominaga, Benjamin J. Blencowe, Olivia M. Pereira-Smith, Tom Misteli*

*To whom correspondence should be addressed. E-mail: mistelit@mail.nih.gov

Published 4 February 2010 on *Science* Express
DOI: 10.1126/science.1184208

This PDF file includes:

MATERIALS AND METHODS

Cell lines and plasmids

Human normal prostate epithelial PNT2 cells (Sigma), immortalized human mesenchymal hMSC stem cells[S1] and human embryonic kidney HEK 293 cells were grown at 37^{o}C, 5% CO_2 in RPMI (Invitrogen) or DMEM (Invitrogen), respectively, both supplemented with 10% fetal bovine serum. Plasmids were nucleoporated in PNT2 and hMSC cells following the manufacturer's recommendations. Briefly, 5 µg of endo-free plasmid Maxi prep (Qiagen) were electroporated using an Amaxa nucleoporator into 10^6 cells resuspended in 100 µl of either solution V (T30 program) for PNT2 cells or solution hMSC (C17 program) for hMSC cells. HEK 293 were transfected with FuGENE HD (Roche) following the manufacturer's instructions. 40-60% of cells survived typically achieving ~45-60% transfection efficiency. After 48h in culture, transfected cells were used for experimentation. pCDNA3 Flag-hASH2-HA was provided by Dr. D. Skalnik, University School of Medicine, Indianapolis USA, pPTB-HA was provided by Dr. J. Patton, Vanderbilt University, Nashville, Tennessee USA, pEGFPN1-MRG15 and pMRG15-V5/His were generated as described[S2;S3]. Yeast pN823-Flag-Set2 (provided by Dr. B. Strahl, University of North Carolina-Chapel Hill, USA) was subcloned in the mammalian expression plasmid pCDNA3.1 (+). pEGFP-C1 (Clontech) was used as a control.

Quantitative RT-PCR

Total RNA was extracted with the RNeasy mini kit (Qiagen) and 500 ng of total RNA was retro-transcribed (RT) by random priming into cDNA using the High Capacity cDNA RT kit (Applied Biosystems) following the manufacturers' instructions. Quantification of the amount of cDNA was done with IQ-custom SYBR-Green supermix for real-time qPCR (MyIQ, BioRad) with oligonucleotide sequences that specifically recognize FGFR2-IIIb, FGFR2-IIIc, TPM2-e6, TPM2-e7, TPM1-e2, TPM2-e3, PKM2-e9, PKM2-e10 splicing variants. CD44 exon variant v6, HMGA1 exon e2 and VDR exon e2 were used as controls for PTB-independent alternative splicing events. GAPDH exons e6 and e7 were used as controls for constitutive splicing. For validation of the PTB, MRG15 and SETD2 co-dependent AS events predicted by high-throughput sequencing, specific primers against the alternatively spliced exon and a constant region of the candidate gene were designed for assessment of the percentage of exon inclusion. The oligonucleotide sequences used are listed in Supplementary Table S2.

RNAi knockdown

Downregulation of PTB and nPTB was performed using ON-TARGET plus SMARTpool siRNA oligonucleotides and downregulation of MRG15, MRGX and SETD2 was performed using Accell siRNA oligonucleotides (Dharmacon). siRNA oligos against human PTB (L-3528), nPTB (L-21323), SETD2 (A-12448), MRG15 (MORF4L1, A-6379) and MRGX (MORF4L2, A-6380) were delivered to cells following the manufacturer's in-

structions for 72h. Accell siRNA anti-human Cyclophilin B (D-1920) or Accell non-targeting siRNA (D-001910) were used as controls.

Chromatin immunoprecipitation (ChIP)

Approximately 2.10^6 cells per sample were crosslinked for 10 min in 1% formaldehyde at room temperature. Cells were washed twice with cold PBS and swelled on ice for 10 min in 25 mM HEPES pH 8, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% NP-40, 1 mM DTT and 1× protease inhibitor cocktail set III (Calbiochem). Following dounce homogenization, the nuclei were collected and resuspended in 1 ml sonication buffer (50 mM HEPES pH 8, 140 mM NaCl, 1 mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS and 1× protease inhibitor cocktail). DNA was sonicated in an ultrasonic bath (Bioruptor Diagenode) to an average length of 200–500 bp. After addition of 1% Triton X-100, samples were centrifuged at 15,000 xg. Supernatants were immunoprecipitated o/n with 40 µl of pre-coated anti-IgG magnetic beads (Dynabeads M-280 Invitrogen) previously incubated with the antibody of interest for 6h at 4°C. The antibodies used were: rabbit anti-MRG15[S2] (10 µL), rabbit anti-H3-K4me1 (4 µg, Abcam ab8895), rabbit anti-H3-K4me2 (2 µg , Upstate 07-030), rabbit anti-H3-K4me3 (5 µg, Upstate 05-745), rabbit anti-H3-K9me1 (5 µg, Abcam ab9045), rabbit anti-H3-K36me3 (3 µg, Abcam ab9050), rabbit anti-H3-K9acetyl (7,5 µg, Upstate 06-942), rabbit anti-H4acetyl (5 µg, Upstate 06-598), mouse anti-H3-K27acetil[S4] (10 µL), mouse anti-H3-K4me3[S4] (10 µL), mouse anti-H3-K9me1[S4] (50 µL), mouse anti-H3-K36me3 (50 µl, provided by Dr. H. Kimura, University of Osaka), mouse anti-H3-K27me3 (10 µl, provided by Dr. H. Kimura, University of Osaka). Control im-

munoprecipitations were performed with no antibody and input between samples was normalized with rabbit anti-histone H3 ChIP (1 µg, Abcam ab1791). Beads were washed sequentially for 5 min each in Low-salt (20 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), High-Salt (20 mM Tris-HCl pH 8, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS) and LiCl buffer (10mM Tris pH 8.0, 1mM EDTA, 250mM LiCl, 1% NP-40, 1% Na-deoxycholate) for 5 min at 4°C and then twice in TE 1x for 2 min at room temperature. Beads were eluted in 1% SDS and 100mM $NaHCO_3$ buffer for 15 min at 65°C and crosslinking was reversed for 7h after addition of NaCl to a final concentration of 200 mM. Chromatin was precipitated with ethanol o/n, treated with 20 µg proteinase K and purified with Qiaquick PCR purification columns (Qiagen) as indicated by the manufacturer. Immunoprecipitated DNA (1.5 µl) and serial dilutions of the 10% input DNA (1:4, 1:20, 1:100 and 1:500) were analyzed by SYBR-Green real-time qPCR. The oligonucleotide sequences used are listed in Supplementary Table 2.

Co-immunoprecipitation and Western-blotting

Crosslinked chromatin was isolated and sonicated as described in the ChIP protocol. Protein concentration of clarified samples was performed by Bio-Rad Protein Assay using BSA as a standard. Three mg of pre-cleared protein were immunoprecipitated with either 4 µg of mouse anti-PTB antibody (Invitrogen, 32-4800) or 1 µg of rabbit anti-His (H-15) (Santa-Cruz, sc-803) in 500 µl sonication buffer for 3 h at 4°C. Precipitation with no antibody was performed in parallel as a control and 30 µg of protein was kept

as 1% input. Immune complexes were captured by incubation with 30 µl of Dynabeads-protein G o/n at 4°C. After washing 4x in 500 µl sonication buffer, immune complexes were eluted in 20 µl SDS sample buffer and boiled for 5 min. Proteins were separated on 10% SDS-PAGE and transferred onto nitrocellulose membrane. Successful transfer of proteins was confirmed by Ponceau S staining. The membranes were blocked in PBS with 0.05% Tween-20, 2% goat serum, and 0.5% skimmed milk, incubated with the appropriate primary (anti-PTB, anti-MRG15, anti-HA or anti-V5) and secondary antibodies, and washed with PBS-Tween 20. Horseradish peroxidase conjugated secondary antibodies were detected by SuperSignal Chemiluminescent Substrate system (Thermo Scientific).

RNA immunoprecipitation (RNA-IP)

Approximately 5.10^6 cells per IP were crosslinked with 1% formaldehyde for 15 min at room temperature. After quick PBS-washing, cells were harvested in 3 ml of ice-cold Lysis buffer (50 mM Tris-HCl pH 8, 100 mM NaCl, 5 mM $MgCl_2$, 0.5% NP-40). Nuclei were pelleted at 2,000xg for 4 min at 4°C. Pellets were resuspended in 500 µl of IP buffer (50mM Tris-HCl pH 8, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) for sonication in an ultrasonic bath (Bioruptor Diagenode). After spinning at maximum speed for 25 minutes to remove cell debris, the supernatants were incubated o/n with 40 µl of pre-coated anti-IgG magnetic beads (Dynabeads M-280 Invitrogen) previously incubated with the antibody of interest for 6h at 4°C. We used 10 µl of two different kinds of rabbit anti-PTB for confirmation of the results (provided by Dr.

D. Black, UCLA and Dr. M. Garcia-Blanco, Duke University), 10 µL rabbit anti-MRG15[S2] or 25 µl mouse anti-SM Y12 (provided by Dr. J. Steitz, Yale). Beads were washed sequentially in RIPA buffer (50mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS and 0.1% Na-deoxycholate) and high-salt RIPA (RIPA at 500 mM NaCl) for 5 min in a rotation wheel at 4°C and then twice in TE 1x at room temperature for 2 min. Beads were eluted for 15 min at 65°C in 200 µl of elution buffer (1% SDS and 100mM $NaHCO_3$) and after addition of NaCl to a final concentration of 200 mM, cross-linking was reversed for 4.5 h at 65°C. 1 ml of Trizol (Invitrogen) and 200 µl of chloroform were added and spun for 15 min at 12,000 xg for separation of the phases. 700 µl of RNAse-free ethanol 70% was added to the supernatant and loaded onto RNAeasy mini columns (Qiagen) for RNA extraction following manufacturer's instructions. DNAse I digestion was performed to avoid DNA contamination. 30 µl of eluted RNA was concentrated in a SpeedVac to 10 µl final volume and the entire sample was used for RT-PCR in a final volume of 20 µl. Immunoprecipitated RNA (3 µl) and serial dilutions of the 10% input RNA (1:2, 1:10 and 1:50) were analyzed by SYBR-Green real-time qPCR. We used the same oligonucleotide sequences as in the ChIP assay. All buffers were sterile, RNAse-free and supplemented with 100 U/ml of RNAse inhibitor (RNAsine, Promega) and 1x protease inhibitor cocktail set III (Calbiochem).

RNA deep-sequencing analysis

50 mer reads (>2 GB data per sample) were generated using an Illumina GA starting from polyA+ RNA isolated from hMSC cells transfected with control and targeting

siRNAs. The sequence reads were first aligned to human genomic sequence (NCBI Genome Assembly Build #36) using Bowtie[S5] and reads mapping to multiple locations were removed. The remaining reads were mapped to a database of human exons and splicing junctions, as previously described[S6], allowing up to two mismatches/indels. The database comprised 27,088 human alternative splicing events identified from alignments of human mRNA/EST sequences to genomic sequence (NCBI Genome Assembly Build #36), essentially as previously described[S7;S8]. Alternative exon inclusion levels (%in) were calculated using the number of reads mapping to exons and splice junctions by the following formula:

%in = Normalized C1:A-A-A:C2 / (Normalized C1:A-A-A:C2 + Normalized C1:C2)

C1:A-A-A:C2 represents the exon-included isoform and refers to the sequence of the last 50 nucleotides of the upstream constitutive exon (C1) followed by the alternative exon sequence (A) followed by the first 50 nucleotides of the downstream constitutive exon (C2). C1:C2 represents the exon-skipped isoform and refers to the 50 nucleotides of the upstream constitutive exon joined to the 50 nucleotides of the downstream constitutive exon. The number of reads that mapped to the two isoform sequences were normalized by sequence length.

12,235 alternative splicing events with ten or more reads that map to one or more of the three junction sequences (C1-A, A-C2, C1-C2), in addition to exon mapping reads, were

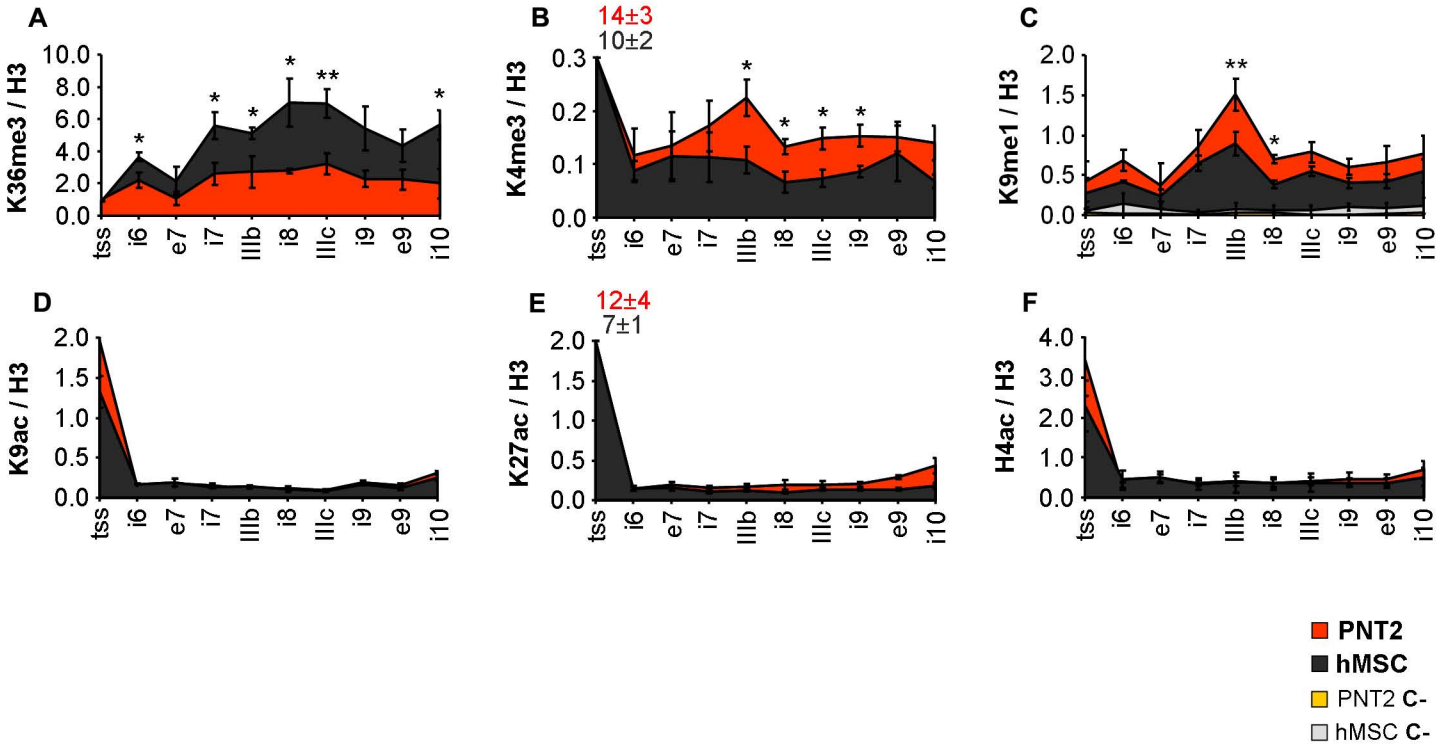
analyzed further. Only events with a % in change between the factor knockdown and control knockdown of at least 15% were analyzed. Significance of the overlaps between factor knockdowns was calculated using the hypergeometric test.

Statistical Analysis

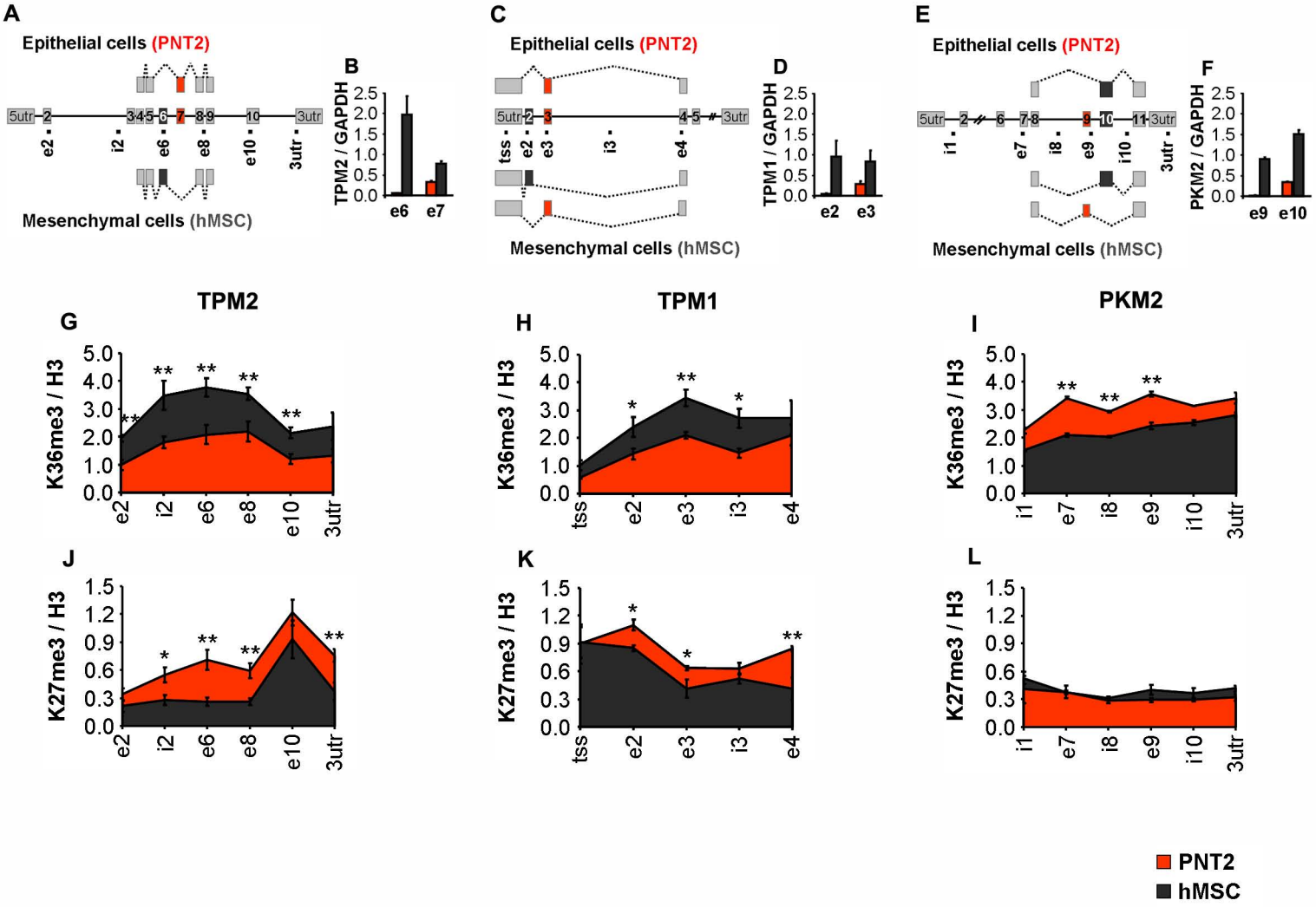
All values represent means ± S.E.M of three to six biological replicas. Statistical significance of the differences between tissues was estimated using the two-tailed paired t-student test comparing either the two cell types or relative to control.

SUPPLEMENTARY FIGURES

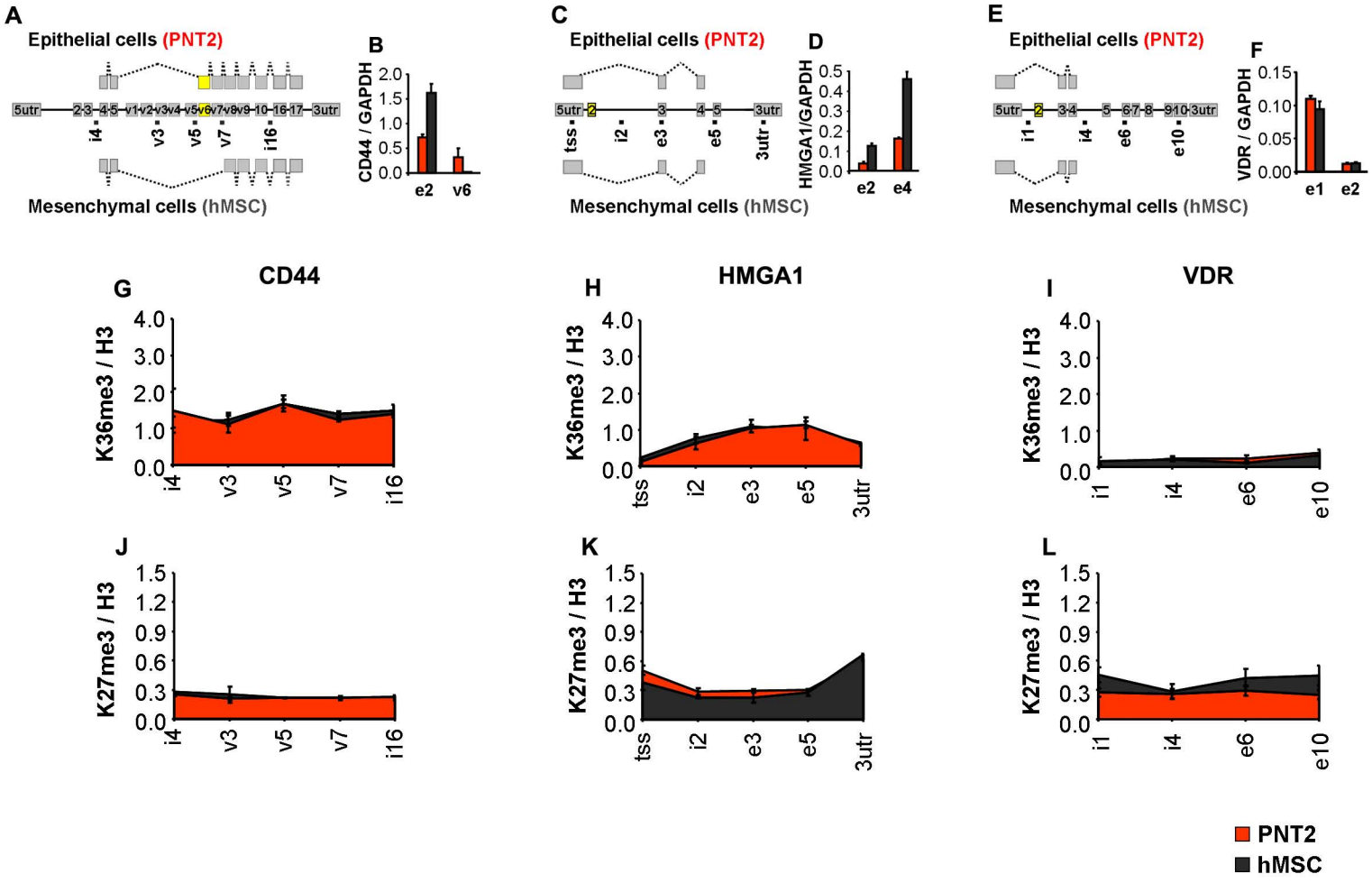
Figs. S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13



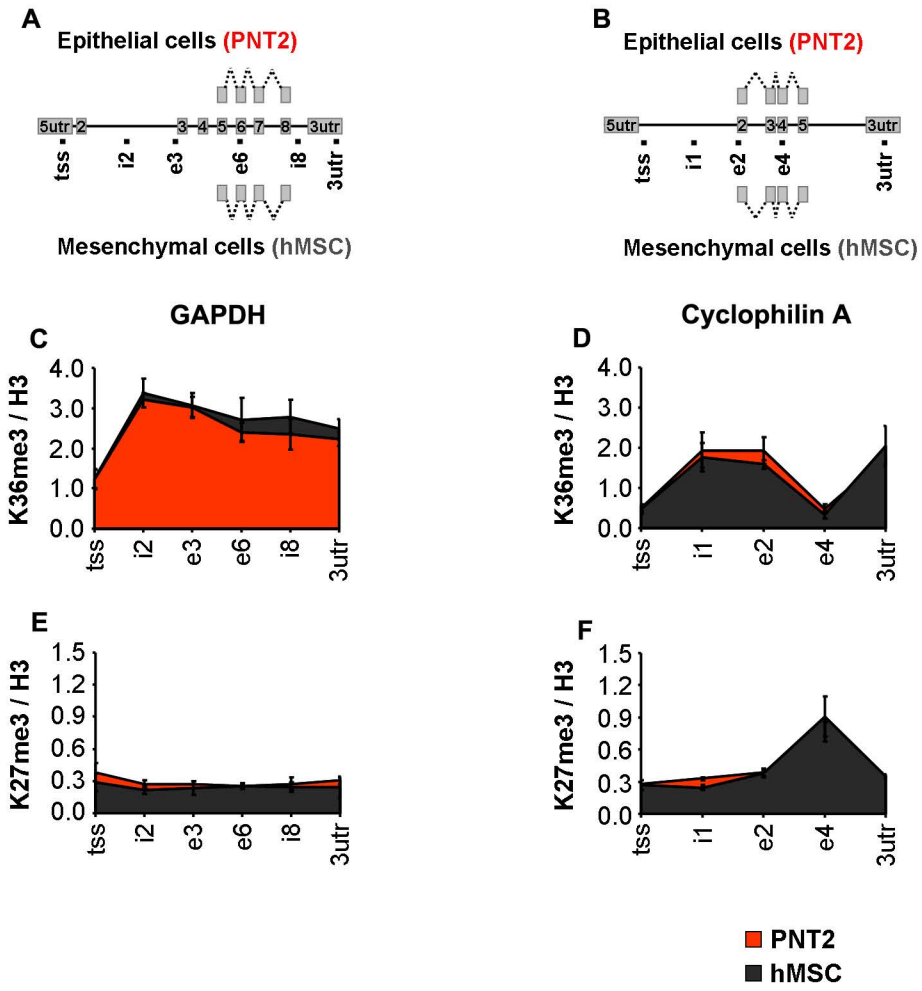
Supplementary Fig. 1. Histone and chromatin properties along FGFR2. (**A-F**) The indicated histone modification was mapped along the FGFR2 gene in PNT2 (red) or hMSC (black) cells using monoclonal (**A-C**) and polyclonal (**D-F**) antibodies (see Methods). A no-antibody control is shown in (**C**) and was typically ~ 1/10 of the signal for a specific antibody. The % input was normalized to unmodified-H3 and values represent means ± S.E.M. from 3-5 independent experiments. * indicates $p<0.05$ and ** indicates $p<0.01$ in t-student test comparing the two cell types.



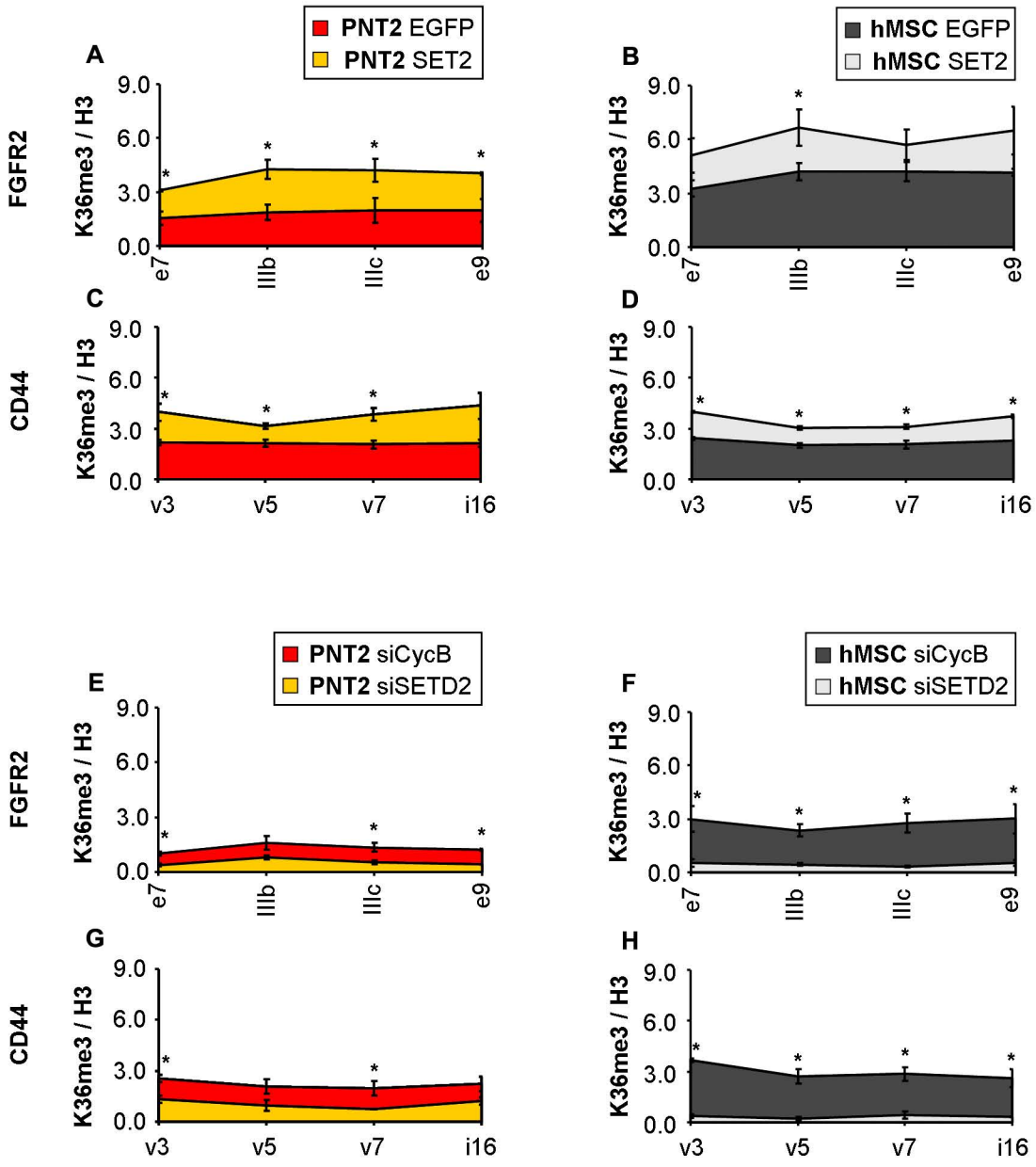
Supplementary Fig. 2. Histone modifications in the PTB-dependent alternatively spliced TPM2, TPM1 and PKM2 genes. (**A,C,E**) Schematic representation of human TPM2, TPM1 and PKM2 genes. Symbols as in Fig. 1. PTB-dependent exon (in red) e7 in TPM2 and e3 in TPM1 are preferentially included in PNT2 cells, whereas e9 in PKM2 is preferentially included in hMSCs. (**B, D, F**) Expression levels of TPM2 exons e6 and e7, TPM1 e2 and e3 and PKM2 e9 and e10 relative to GAPDH in PNT2 (red) and hMSC (black) cells determined by qPCR. (**G-L**) Mapping of H3-K36me3 (**G-I**) and H3-K27me3 (**J-L**) along TPM2, TPM1 and PKM2 in PNT2 (red) and hMSC (black) cells, using quantitative chromatin immunoprecipitation. The % input was normalized to unmodified-H3 and values represent means ± S.E.M. from 5 independent experiments. * indicates $p<0.05$ and ** indicates $p<0.01$ in t-student test comparing the two cell types.



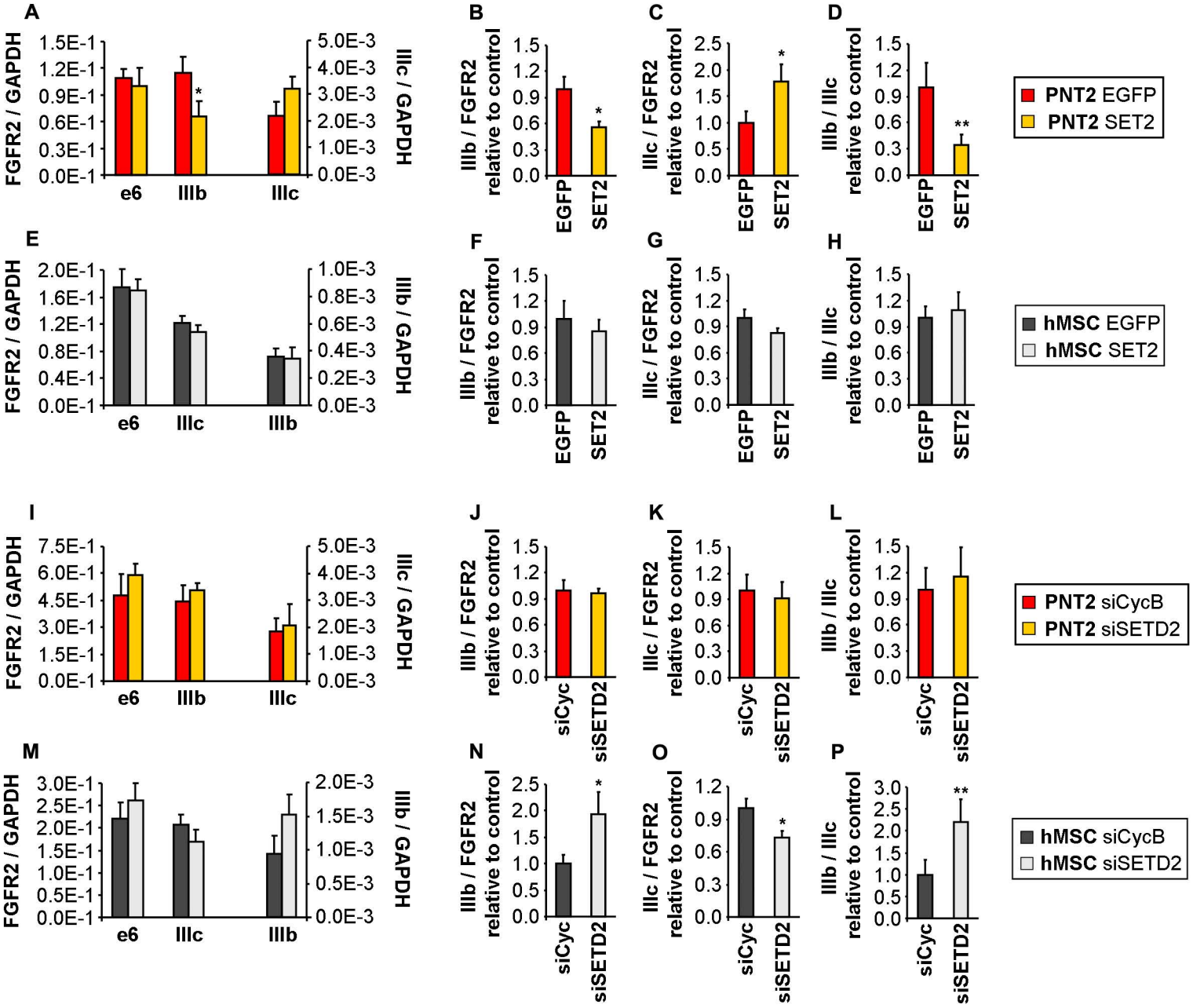
Supplementary Fig. 3. Histone modifications in PTB-independent alternatively spliced genes. (**A,C,E**) Schematic representation of PTB-independent alternatively spliced human CD44, HMGA1 and VDR genes. Symbols as in Fig. 1 except alternatively spliced exons are shown in yellow. (**B,D,F**) Levels of CD44 constitutive exon e2 and variant 6 (v6), HMGA1 alternative e2 and constitutive e4 and VDR constitutive e1 and alternative e2 in PNT2 (red) and hMSC (black) cells determined by qPCR. CD44 exon v6 is only included in PNT2, whereas HMGA1 e2 and VDR e2 are not included in either cell types. (**G-L**) Mapping of H3-K36me3 (**G-I**) and H3-K27me3 (**J-L**) along CD44, HMGA1 and VDR in PNT2 (red) and hMSC (black) cells, using quantitative chromatin immunoprecipitation. The % input was normalized to unmodified-H3 and values represent means ± S.E.M. from 3-4 independent experiments. No significant differences were observed in a t-student test comparing the two cell types.



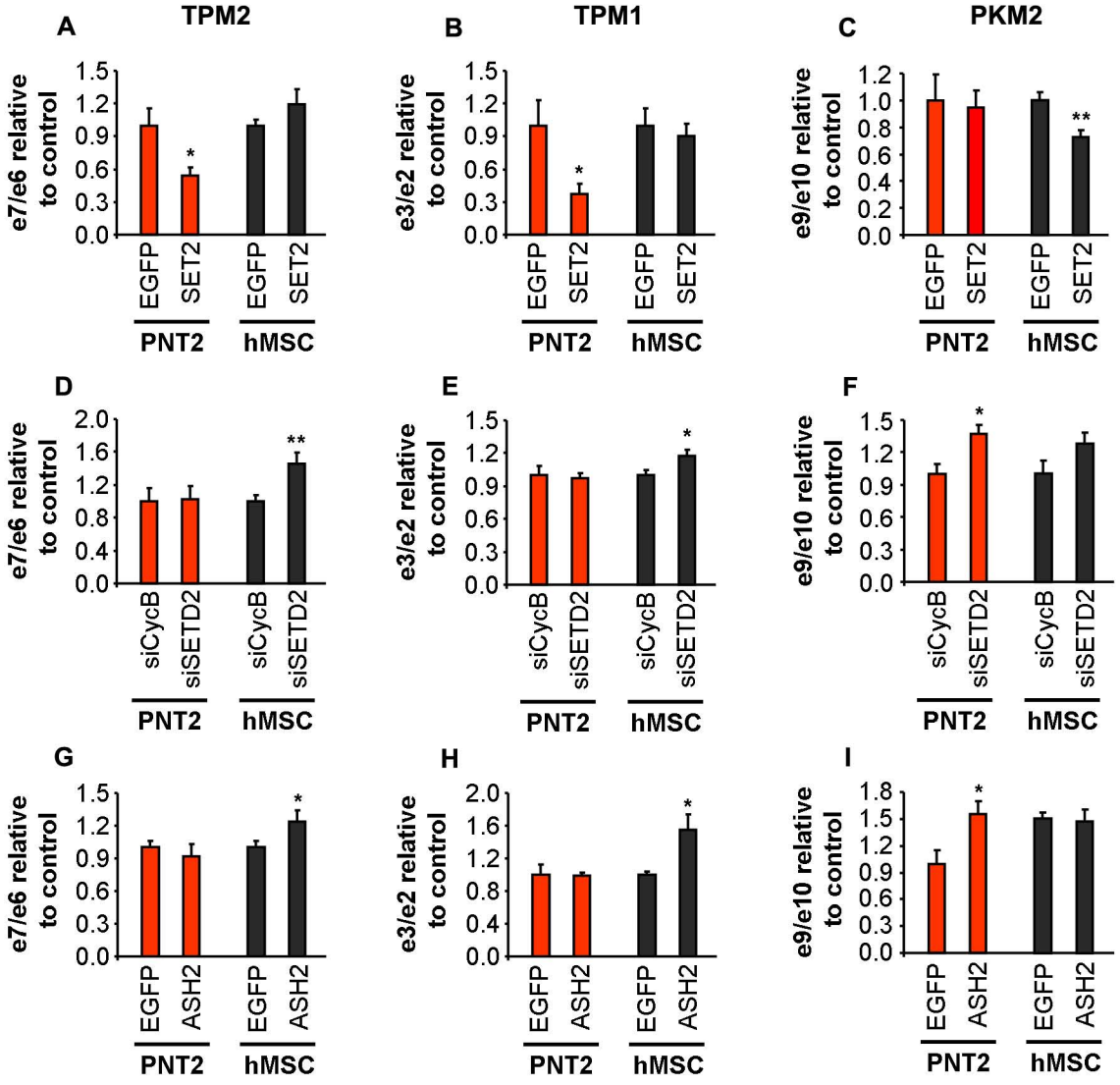
Supplementary Fig. 4. Histone modifications in the constitutively spliced GAPDH and Cyclophilin A genes. (**A-B**) Schematic representation of human GAPDH and Cyclophilin A genes. Symbols as in Fig. 1. (**C-F**) Mapping of H3-K36me3 (**C-D**) and H3-K27me3 (**E-F**) along GAPDH and Cyclophilin A in PNT2 (red) and hMSC (black) cells, using quantitative chromatin immunoprecipitation. The % input was normalized to unmodified-H3 and values represent means ± S.E.M. from 3-4 independent experiments. No significant differences were observed in a t-student test comparing the two cell types.



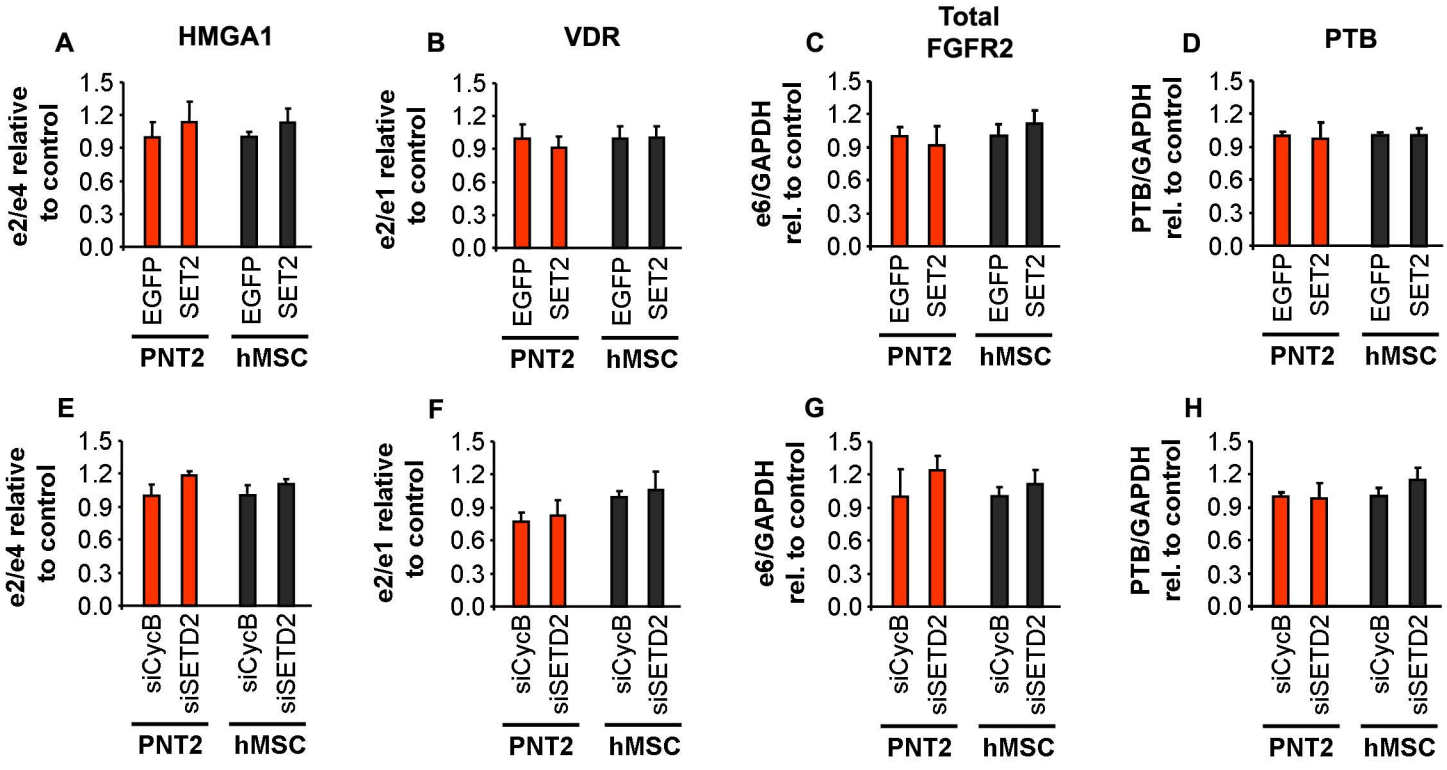
Supplementary Fig. 5. Confirmation of changes in H3-K36me3 enrichment at FGFR2 and CD44 loci after SET2 modulation. Quantitative chromatin immunoprecipitation of H3-K36me3 FGFR2 and CD44 loci after (**A-D**) overexpression of Flag-Set2 (SET2) or (**E-H**) downregulation of SETD2 (siSETD2), respectively, in PNT2 (red and yellow) or hMSC (black and grey) cells. Overexpression of EGFP and downregulation of Cyclophilin B (siCycB) were used as controls. The % input was normalized to unmodified-H3 and values represent means ± S.E.M. from 3 independent experiments. * indicates $p<0.05$ in t-student test compared to control.



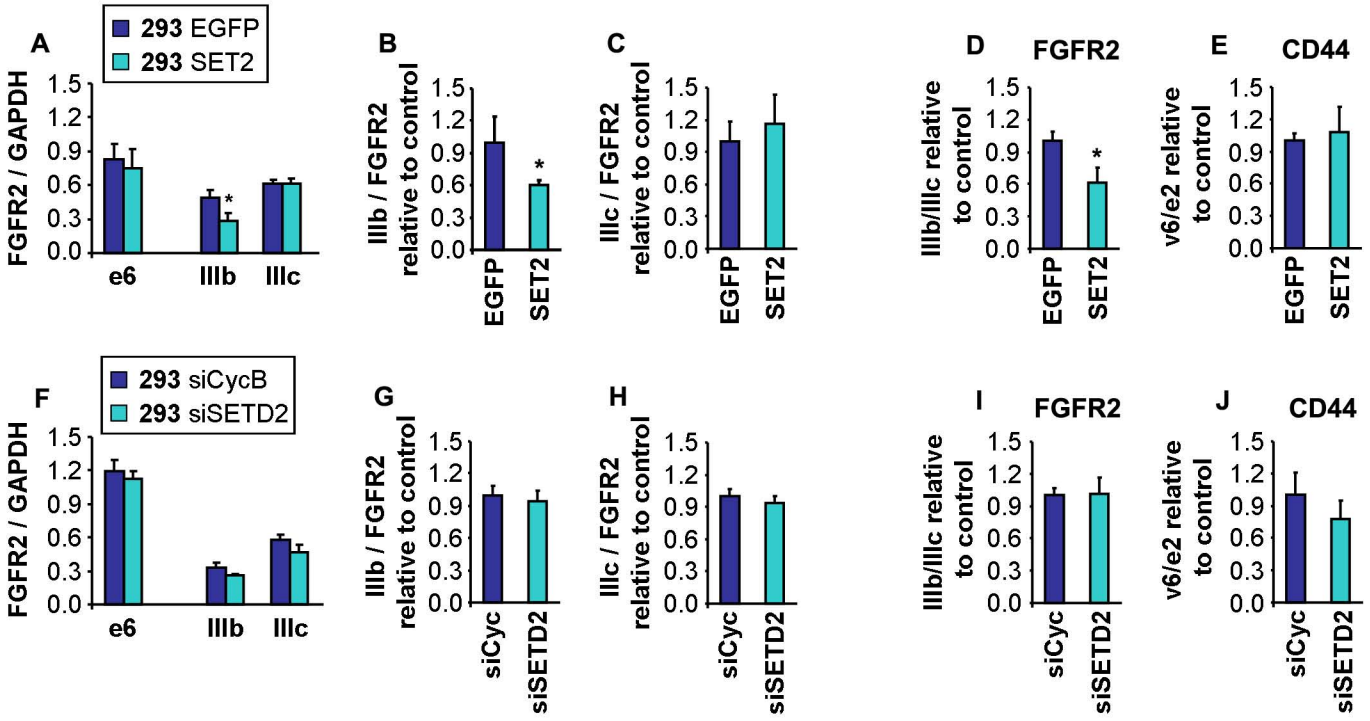
Supplementary Fig. 6. Quantitation of FGFR2 exon usage upon SET2 expression. Detection by qPCR of FGFR2 e6, IIIb and IIIc after (**A-H**) overexpression of Flag-Set2 (SET2) or (**I-P**) downregulation of SETD2 (siSETD2) in PNT2 (red and yellow) or hMSC (black and grey) cells. (**A,E,I,M**) Levels of FGFR2 exons e6, IIIb and IIIc normalized to GAPDH. (**B,C,F,G,J,K,N,O**) Levels of exons IIIb and IIIc normalized to FGFR2 e6 and shown relative to EGFP or Cyclophilin B (siCycB) controls. (**D,H,L,P**) Exon ratio IIIb/IIIc relative to control. Values represent means ± S.E.M from 4-6 independent experiments. * indicates $p<0.05$ and ** indicates $p<0.01$ in a t-student test compared to control.



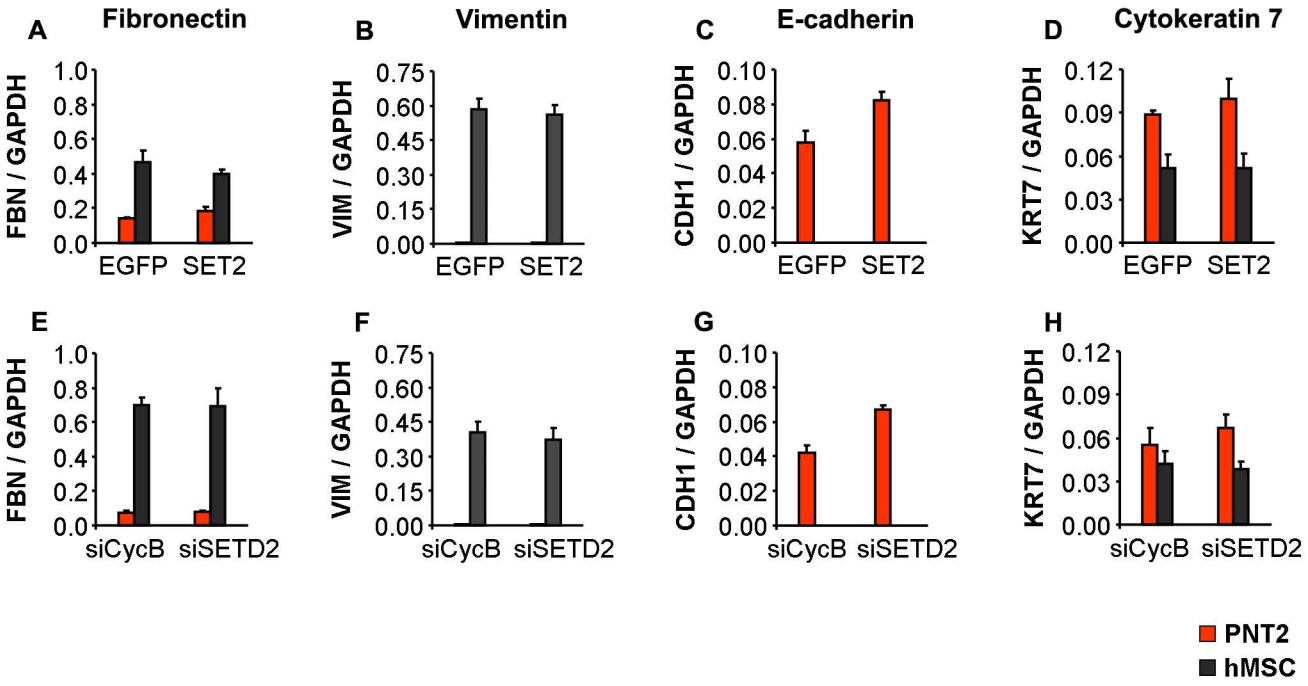
Supplementary Fig. 7. Quantitative RT-PCR analysis of exon ratios in PTB-dependent alternatively spliced genes. (**A-C**) Overexpression of Flag-Set2 (SET2), (**D-F**) downregulation of SETD2 (siSETD2) or (**G-I**) overexpression of Flag-ASH2 (ASH2) in PNT2 (red) or hMSC (black) cells. Ratios are: exon e7/e6 for TPM2, exon e3/e2 for TPM1 and exon e9/e10 for PKM2. Values represent means ± S.E.M of percentages relative to overexpression of EGFP or downregulation of Cyclophilin B (siCycB) used as controls from 4-6 independent experiments. * indicates $p<0.05$ and ** indicates $p<0.01$ in a t-student test compared to control.



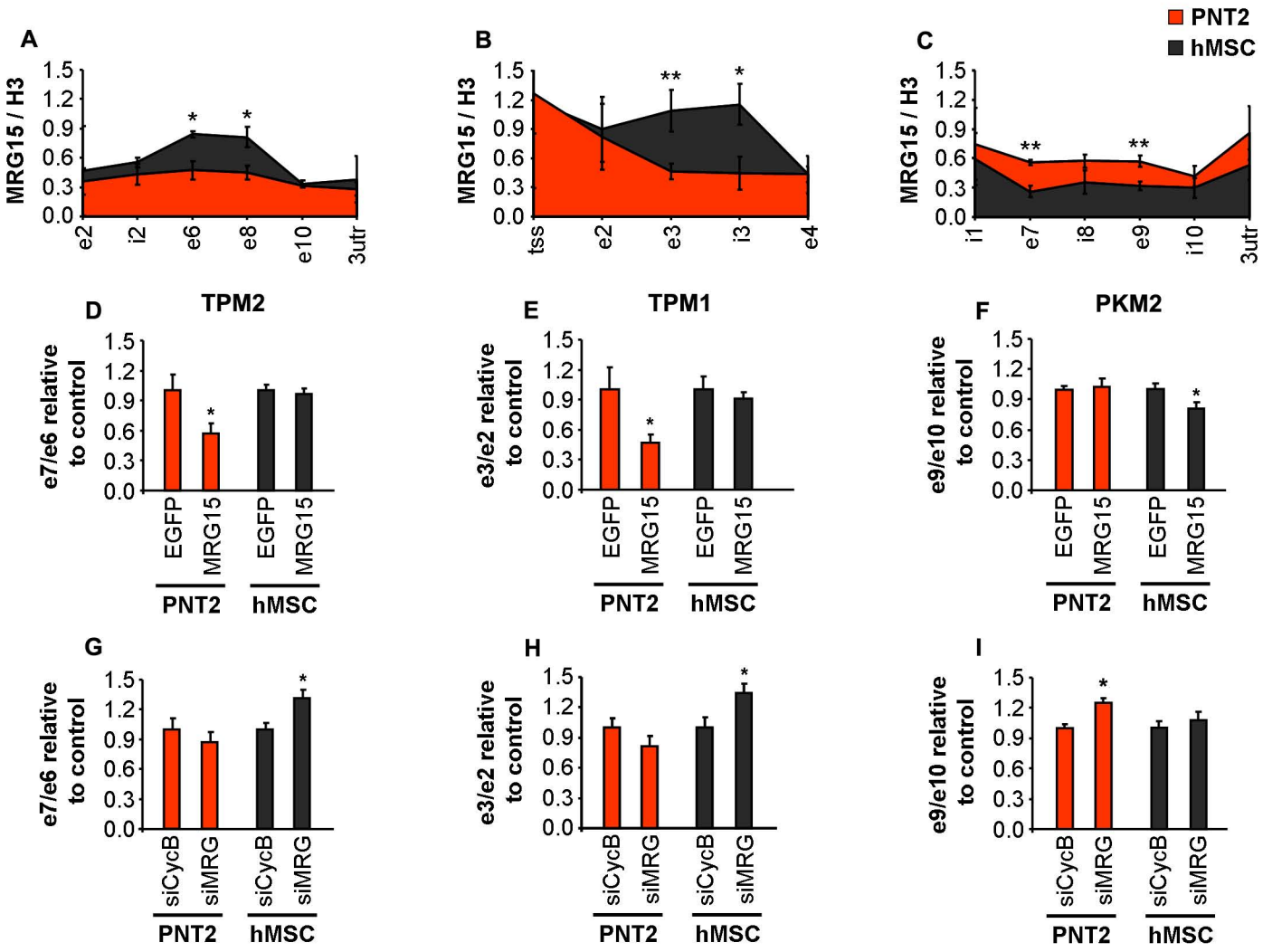
Supplementary Fig. 8. Quantitative RT-PCR analysis of exon ratios in non PTB-dependent alternatively spliced genes and total levels of FGFR2 and PTB. (**A-D**) Overexpression of Flag-Set2 (SET2) or (**E-H**) downregulation of SETD2 (siSETD2) in PNT2 (red) or hMSC (black) cells. Ratios are: exon e2/e4 for HMGA1, exon e2/e1 for VDR and the constitutively included FGFR2 exon e6 or PTB relative to GAPDH. Values are percentages relative to overexpression of EGFP or downregulation of Cyclophilin B (siCycB) used as controls. Values represent means ± S.E.M from 3-4 independent experiments. No significant differences were observed in a t-student test compared to control.



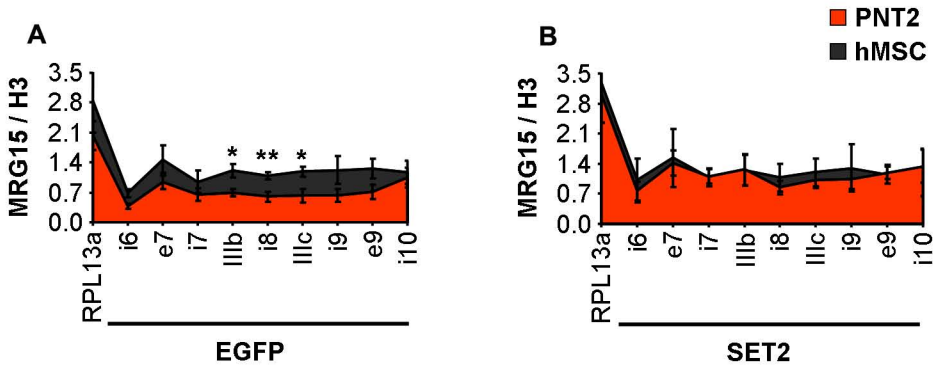
Supplementary Fig. 9. Quantitation of FGFR2 exon usage in HEK 293. Detection by qPCR of FGFR2 e6, IIIb, IIIc, and CD44 exon v6, e2 after (**A-E**) overexpression of Flag-Set2 (SET2) or (**F-J**) downregulation of SETD2 (siSETD2) in HEK293 cells (dark and light blue). (**A,F**) Expression levels of FGFR2 exons e6, IIIb and IIIc normalized to GAPDH. (**B,C,G,H**) Expression levels of exons IIIb and IIIc normalized to FGFR2 e6 and shown relative to EGFP or Cyclophilin B (siCycB) controls. (**D,E,I,J**) Exon ratio IIIb/ IIIc for FGFR2 and v6/e2 for CD44 relative to control. Values represent means ± S.E.M from 3-6 independent experiments. * indicates $p<0.05$ in a t-student test compared to control.



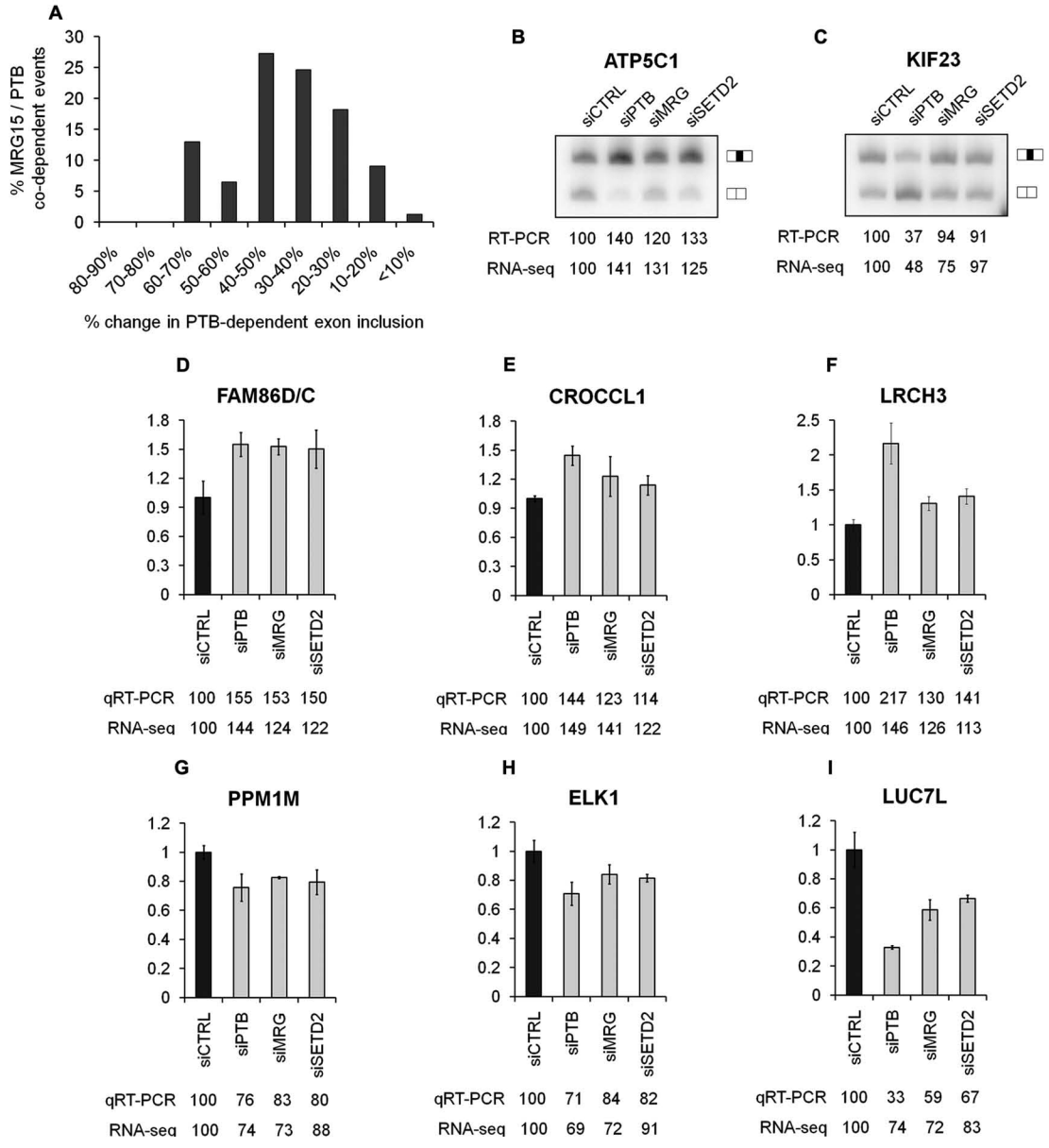
Supplementary Fig. 10. Absence of epithelial-to-mesenchymal transition upon SET2 modulation. Quantitative RT-PCR analysis of mesenchymal (Fibronectin and Vimentin) and epithelial (E-caherin and Cytokeratin 7) cell markers after (**A-D**) overexpression of Flag-Set2 (SET2) or (**E-H**) downregulation of SETD2 (siSETD2) in PNT2 (red) or hMSC (black) cells. Overexpression of EGFP and downregulation of Cyclophilin B (siCycB) were used as controls. Values were normalized to GAPDH and represent means ± S.E.M from 3 independent experiments. No significant differences were observed in a t-student test.



Supplementary Fig. 11. Modulation of TPM2, TPM1 and PKM2 splicing upon MRG15 modulation. (**A-C**) Chromatin immunoprecipitation of MRG15 along TPM2, TPM1 and PKM2 genes in PNT2 (red) and hMSC (black) cells. The % input was normalized to unmodified-H3 and values represent means ± S.E.M from 4 independent experiments. * indicates $p<0.05$ and ** indicates $p<0.01$ in t-student test comparing the two cell types. (**D-I**) Quantitative RT-PCR analysis of exon ratios after (**D-F**) overexpression of EGFP-MRG15 or (**G-I**) downregulation of both MRG15 and MRGX (siMRG) in PNT2 (red) or hMSC (black) cells. Ratios are: exon e7/e6 for TPM2, exon e3/e2 for TPM1 and exon e9/e10 for PKM2. Values represent means ± S.E.M of percentages relative to overexpression of EGFP or downregulation of Cyclophilin B (siCycB) used as controls from 4-6 independent experiments. * indicates $p<0.05$ in t-student test compared to control.



Supplementary Fig. 12. H3-K36me3 recruits MRG15 to FGFR2. Chromatin immunoprecipitation of MRG15 along FGFR2 after overexpression of EGFP (**A**) or SET2 (**B**) for 48h in PNT2 (red) or hMSC (black) cells. % input was normalized to unmodified-H3. Values represent means ± S.E.M from 4 independent experiments.* indicates $p<0.05$ and ** indicates $p<0.01$ in t-student test comparing the two cell types.



Supplementary Fig. 13. (**A**) Percentage of MRG15/PTB co-dependent alternatively spliced events versus the degree of PTB-dependence (percent of change in exon inclusion upon downregulation of PTB). Events that are moderately PTB-dependent are the most sensitive to MRG15 loss. RNAcompete analysis[59] of RNA binding specificities to determine the relative preferences of RNA binding proteins for short RNAs (k-mers) confirmed that PTB binding sites in alternative exons and their flanking intron sequences within 500 nts were on average stronger (7-mer score -0.02) in PTB-dependent events that undergo >50% change in splicing than in PTB-dependent splicing events of 15-50% change (7-mer score -0.04; p=6.4e-05, Wilcoxon Rank Sum Test). Furthermore, PTB/MRG15 co-dependent events were in general associated with weaker PTB binding sites (7-mer score -0.048) than PTB-dependent events that are MRG15 independent (7-mer score -0.036; p=9.3e-03, Wilcoxon Rank Sum Test). (**B-I**) Validation of predicted PTB, MRG15 and SETD2-dependent alternative splicing events by radioactive (**B-C**) and quantitative real-time PCR (**D-I**). 8 out of 11 candidate genes that were predicted by high-throughput RNA sequencing to be PTB, MRG15 and SETD2-dependent in hMSC downregulated cells were successfully validated. The percentage of change in exon inclusion between non-targeting siRNA control (siCTRL) and targeting siRNA (siPTB, siMRG or siSETD2) in both quantitative RT-PCR (qRT-PCR) and RNA-deep sequencing (RNA-seq) are given below each graph for comparison.

SUPPLEMENTARY TABLES

Supplementary Table 1. PTB, MRG15 and SETD2 co-dependent events. List of predicted PTB and MRG15-dependent alternative splicing events that change the percentage of exon inclusion ≥15% upon specific downregulation in hMSC cells. The changes in alternative splicing in SETD2 downregulated cells are also shown. C1 and C2: flanking constitutive exon sequence, A: alternative exon sequence.

Assay	Name	FOR	REV	Prod. length (bp)	Ta (°C)
RT-PCR	hGADPH	TTCGACAGTCAGCCGCATCTTCTT	CAGGCGCCCAATACGACCAAATC	110	60
	hMRG15	AGGAGGAGGCGGCGAATCACTT	AGAGGCCCATGAAAGCACAGCAC	97	60
	hMRGX	TTGGAAGGGAAAGAAGGAAGATTG	GATTGCTGTCCACGAGGTTGAG	93	60
	hPTB (PTBP1)	ACGTCACGGAGGGGGAAGTCATCT	GTTGGCAGCCTCCTCCGTGTTCAT	119	60
	hPTBP2	TCCTTCTGCCACCCTTCACCTATC	CTTTCACAGTGCCCCCAGTGTTAG	95	60
	hSETD2	AGCTCCCTCTCACCACCCTCTT	CAACTTCCGGCGTTCCTCTGT	138	60
	hGAPDH (e6)	GTGGAGTCCACTGGCGTCTTC	CTGATGATCTTGAGGCTGTTGTCA	158	60
	hGAPDH (e7)	CAATGCCTCCTGCACCACCAACT	AGCGCCAGTAGAGGCAGGGATGAT	196	60
	hCD44	GACAAGTTTTGGTGGCACG	CACGTGGAATACACCTGCAA	105	60
	hCD44 (v6)	ATCAAGCAGGAAGAAGGATGGATA	TGAGAATTACTCTGCTGCGTTGTC	139	60
	hVDR	GAACAGCTTGTCCACCCG	GCAGGTAAGTGGAGCCCAG	97	60
	hVDR (e2)	TAGCCCAGCTGGACGGAGAAATG	CTGAAGGAGCAGGGGGCAGGTAA	191	60
	hHMGA1	CCCAGCGAAGTGCCAACACCTAAG	GCCCTCCTCTTCCTCCTTCTCCAGT	150	60
	hHMGA1 (e2)	GACTCCGAGCCGGGGCTATTTCTG	CGGTGCTGGGCGCTGAGGAC	111	60
	hFGFR2(e6)	TCCATCAATCACACGTACCACCTG	ACTCTACGTCTCCTCCGACCACTG	112	60
	hFGFR2 (IIIb)	AAGTGCTGGCTCTGTTCAATGT	GTAGTCTGGGGAAGCTGTAATCTC	161	60
	hFGFR2(IIIc)	TGAGGACGCTGGGGAATATACG	TAGTCTGGGGAAGCTGTAATCTCCT	123	60
	hTPM2 (e6)	CTGGAGCGCTCGGAGGAGAG	TGAGGCCATCAGGGACTTGAGG	105	60
	hTPM2 (e7)	GGAGGCCCAGGCGGACAA	GCCACAGACCTCTCGGCAAACTC	120	60
	hTPM1 (e2)	AGCAGGCGGAGGCCGACAAGA	TCCTCCGACACCCGCAGCAACT	97	60
	hTPM1 (e3)	CTCAAGGGCACCGAAGATGAACTG	CTCAACCAGCTGGATGCGTCTGTT	141	60
	hPKM2 (e9)	GGAGAAACAGCCAAAGGGGACTAT	GAGGCTCGCACAAGTTCTTCAAAC	119	60
	hPKM2 (e10)	TGCCGTGGAGGCCTCCTTCAAGT	GGGGCACGTGGGCGGTATCTG	102	60
	hCROCCL1	CGGCGGAAGGCTGAGAAGGAA	AGGCTGTGCCGTGTACCCATCAAC	167	**65**
	hCROCCL1 (AS)	CTGCAGGCCGAGAAGACCGAGG	GAGGGCGCTCAGCTTGGACAGG	138	**65**
	hLRCH3	CAAAAAGCATCACAAAGTCCACAA	CATGCACAAGGCACTATCATCAGT	114	**60**
	hLRCH3 (AS)	GCTGCTACCCTGCCTCATTCTTC	CGCTGAGGCTGATTTCTCTGGATA	149	**60**
	hFAM86D/C	AGAACGCGGGGAGCGAACTCT	AAATATCCCGCAGCAGCTCAGAATC	138	**60**
	hFAM86 (AS)	GGTGCTGGTGGAGACCCTGATG	ACCCCGACCAGCGACACGA	104	**60**
	hPPM1M	GGCCATCTTGGTGCGGAGAG	CACCAGCCAGAAGCTCAGGATAGA	116	**65**
	hPPM1M (AS)	TCTGCCCCAGTGACCCTCAGTTTG	GCTGTGCAGCCGCCCATCTG	148	65
	hLUC7L	TTCCATGCCTGCATCCAGTTTTCA	GCCAGGCGACGGTCATTGTCA	96	**60**
	hLUC7L (AS)	GGGGGACCCTACCCTCAGAA	TCTCCGATCACATTCAGCAATAAA	102	**60**
	hELK1	CTCGGGGAGCAGGAGCACCAGT	GCCCAAACCCGGCTCCACATTA	145	60
	hELK1 (AS)	AACGCCACCGCCACGCTACA	CGCCACGGATTGATTCGCTACG	103	**60**
	hE-Cadherin	TTGACGCCGAGAGCTACAC	GACCGGTGCAATCTTCAAA	93	60
	hKRT7	CTGCCTACATGAGCAAGGTG	GGGACTGCAGCTCTGTCAAC	108	60
	hVimentin	CTTCAGAGAGAGGAAGCCGA	ATTCCACTTTGCGTTCAAGG	97	60
	hFibronectin	CCATAAAGGGCAACCAAGAG	ACCTCGGTGTTGTAAGGTGG	91	60

Assay	Name	FOR	REV	Prod. length (bp)	Ta (°C)
ChIP	SAT	CATCGAATGGAAATGAAAGGAGTC	ACCATTGGATGATTGCAGTCAA	160	60
	hEVX1.tss	TTCGCTGTGGCAGACGTTTCTATT	AAGCCCCATTGCCCTCTTCTTT	121	60
	GAPDH.tss	CCCAACTTTCCCGCCTCTC	CAGCCGCCTGGTTCAACTG	293	65
	GAPDH.i2	CCAAGCCGGGAGAAGCTGAGTCAT	AAGGGCGGGGCCGTGAGGT	116	60
	GAPDH.e3	AGGTGGCCTAGGGCTGCTCACATA	GGCGCCCAATACGACCAAATCTAA	93	60
	GAPDH.e6	ATCTCTGCCCCCTCTGCTGATGC	GGCCCTGCCTTCCTCACCTGAT	100	60
	GAPDH.e9	GAATCTCCCCTCCTCACAGTTG	GGTTGAGCACAGGGTACTTTATTG	105	60
	GAPDH.3UTR	CTGGGGAGGGACCTGGTATGTTC	TGCCAGCTTCCTGTAGCACTCAAG	119	60
	CyclinA.e1	GCGGCGCCATTTCCTGACGA	CGCCCCGCCTGCCTTTGTC	92	60
	CyclinA.i1	CCCCACCCCACCTATGAGTGTAGT	ACCCCTCCATTCTCATCAAGACCT	149	60
	CyclinA.e2	TAAGGGGCCTGGATACCAAGAAGT	ACTGCTGCACGATCAGGGGTAAA	129	60
	CyclinA.e4	TATGTTGACAGGGTGGTGACTTCA	CAGGACCCGTATGCTTTAGGATG	108	60
	CyclinA.3UTR	ATTCCCTGGGTGATACCATTCAAT	ATGACAACGTGGTGAGGCTATTCT	145	60
	VDR.i1	TCTGTGCAGACTCGGCAGTTGGTT	CCTCGGGCCTCATGGGTATTCC	133	60
	VDR.i4	CCCCTCAGTGACCAGAGCGAAATG	CTCAGGCCTCAGAATGCCCGTAAT	135	60
	VDR.e6	CTGAGGAGCAGCAGCGCATCATTG	TCCTCCCAGCAGGCAGACATACCC	111	60
	VDR.e9	ACCTGGGGAGCGGGGAGTATG	GGGCTCTGCAAACCAGCAAAGTAG	123	60
	CD44.i4	TTCCCTGCTGAGGCTATTAGTCTG	CCCACTTGTTTCCCTACTCTCC	115	60
	CD44.v3	AGTCACCCTCCGCTTTCCCTCCA	CTTCGCCACTTGTCCCGCACAT	111	60
	CD44.v5	TCACCTGAACAGGAATGGATACAA	AGGGCCTCTGCTACAATCTCACTA	128	60
	CD44.v7	TGGAATGGTGCTATGTGGCTTAC	AGAAATGAAGGGAAAATCAGGTTG	102	60
	CD44.i16	TCCTTTGCCCCAGCTTGCCTAA	TGGGAAGAATCTTGCTGCCTGATG	121	60
	HMGA1.tss	CCCCGCCCCCTGAGTGACAC	CCGCGCCAGCGCCAGAAATA	141	55
	HMGA1.i2	GGAGCAGGATAAATCACCAACCA	GCATCAAGAGACCCTCCAGACAG	108	60
	HMGA1.e3	GGGCAGACCCCTCCATCC	CATCTCGCTTCAGCCACTTTCTAC	120	60
	HMGA1.e5	AACCCTCTAGAAAACCACCACAAC	ACTCAACACCCTCAACTCAAAAGA	122	60
	HMGA1.3UTR	CCAGGATTCCCCCAGCCAAACT	CACCCCTCCTGCCTTCCTGTAACC	126	60
	FGFR2.tss	GGCAACCCTCCCCGCAGTATCAAG	GCTGGCCAACGGCTCGCTGAG	139	**60**
	FGFR2.i6	TATCTTAAAACGCACCCACACTAC	CAAGGTCACAAACTATGCTCCTAT	162	60
	FGFR2.e7	TGTTATTTCAAAGGTGTCAGCCA	GAGAGGGAAGAAAGGAGGAGTG	131	60
	FGFR2.i7	GTGTTGATTGTTACTCTGATGTTGTT	TATTGTCTAAATCACCTCTGAATGG	159	60
	FGFR2.eIIIb	AGCCCTTTAATGCCGCTGTTTAGA	TAACGGCCAACCAGGAAGGTCTTAG	102	65
	FGFR2.i8	TGATCCTTATGAGTTGCTGTTCTTG	TTGCCTTTAGTAGCGTCCAGTAGTA	93	60
	FGFR2.eIIIc	GAATATACGTGCTTGGCGGGTAAT	TAAAAGGGGCCATTTCTGATAACA	166	60
	FGFR2.i9	GGAGCGGCACCTCTGAATGTCA	CAAGCCCAAGATGGCAGGAGCAT	139	65
	FGFR2.e9	CCTGGGTCCTGGTGGTCAAAT	TAAGCAGGCCATAGAGTTAGCACAC	158	65
	FGFR2.i10	GGAATGAAAAGCCCCAGAAA	GCCTCATAAATGCCTAGAAACAAC	134	60
	FGFR2.i13	TGTGGCTGGACAACTGGAGGTA	AAAGAAAACGCAGTGTGGCTCTAA	175	60
	FGFR2.e14	TTCCTTTTTGTTCTGGCGGTGTT	CTCTGGCGAGTCCAAAGTCTGCTA	148	60
	FGFR2.e15	AGGGGCGGCTTCCAGTCAAGT	CAAGCCCAGGAAAAAGCCAGAGAA	112	60

FGFR2.i15	TCCCAGAGATTGAGCCTCCTGAAC	AATCACATGCCCCAGATGAGTTGC	128	60
FGFR2.3UTR	CCGTGCGTACTGGCTGTG	CTGCATTTGTGCTCTGTAAGTGTG	121	60
TPM1.e1	CGGCTCCGCGCTCGCACTCC	TTCTTGTCGGCCTCCGCCTGCTCAG	181	65
TPM1.e2	CCGAGGGGCCCCAGCCAACC	GGTCCCGCTCGTCCTCCGACACC	155	65
TPM1.e3	GGATGCCGCCTCTGACTGCTA	GGCATTCTCTACAAACCTTGGTG	213	60
TPM1.i3	TTCTGGGGTGGTGGCTTTCAGT	GTGCCTCTCCCACTTGACCTGTC	158	60
TPM1.e4	TACATGGCTGAGCCTGGGTAAGTC	TTGCCTGTTAAACAAGAGCGTCAT	94	60
TPM2.e2	CCGTGAAGGAGGCCCAGGAGAAACT	GCGCTGGGGACATAAGGGGATAAGGT	162	60
TPM2.i2	GCTTGATTCTGTGACGGGCTATTCTT	GGTCTGATGCAGCTCTGATTCCTACA	117	60
TPM2.e6	GGGCTGAGGTGGCCGAGAGGTAAAGA	GCAGCGGGCAGGGGTCAGAGAAC	142	60
TPM2.e8-9	AGTCTCCCCTCAAGCCACCCTAATC	GCCCTCCCACAGACTACTGCCTAAG	108	60
TPM2.e10	AGGGGAGCAGGCAGGAGGAGCAG	AGCAAAGGAGGGTGGAAGGGGATAGG	141	60
TPM2.3UTR	CAGGCTCTGTGAAGCAGGGAGTCTG	TCACCCTGCTCAAGGAAAATGTGTCA	175	60
PKM2.i1	GCCTTTGGTGTTCCGCTGTA	GACCCTGGCTCCAATCCTAATC	102	60
PKM2.e7	CCCCCAACTTTGTCCATCAGG	GCACCGTCCAATCATCATCTTC	135	60
PKM2.i8	GCCTGTCAGTGGGGTCAGAGACTA	TCCAAACCTTGCAGCCCTAAAGTA	118	60
PKM2.e9	GCAGCAGCTTTGATAGTTCTGACGG	ACCTCCCTGGCGGTGTTCCTAC	172	60
PKM2.i10	GTCACTGTCATTGGCAGCGAGAG	CAGAACCCCTCCTACACCCTGAAC	94	60
PKM2.3UTR	AAGATGGGAAAGAGCAGGGACAAG	GGGATACGCTAATGCAGGAAGACC	115	60

Supplementary Table 2. List of oligonucleotide pair sequences used in real-time PCR. Ta: recommended annealing temperature.

SUPPLEMENTARY REFERENCES

S1. T. Okamoto et al., *Bioch. Biophys. Res. Commun* **295** (2002)

S2. P.S. Pardo et al., *J. Biol. Chem.* **277** (2002)

S3. K. Tominaga et al., *Gene* **294** (2002)

S4. H. Kimura et al., *Cell Struct.Funct.* **33** (2008)

S5. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, *Genome Biol.* **10**, R25 (2009).

S6. Q. Pan, O. Shai, L. J. Lee, B. J. Frey, B. J. Blencowe, *Nat. Genet.* **40**, 1413 (2008).

S7. Q. Pan et al., *Mol. Cell* **16**, 929 (2004).

S8. Q. Pan et al., *Trends in Genet.* **21**, 73 (2005).

S9. D. Ray et al., *Nature Biotech.* **27**, 667 (2009).