Supplementary Information

PTBP1 and PTBP2 repress nonconserved cryptic exons

Jonathan P. Ling, Resham Chhabra, Jonathan D. Merran, Paul M. Schaughency,

Sarah J. Wheelan, Jeffry L. Corden, Philip C. Wong

Correspondence to P.C.W. (wong@jhmi.edu)

Supplementary Data

- Supplemental Experimental Procedures
- Supplemental Figures 1 to 8
- Supplemental Tables 1 and 2

Supplemental Experimental Procedures

RT-PCR primers used for Fig. 2 are listed below:

FERMT2 (108/141 bp)

Forward: AGGGGCCTCTTATCACTCCT

Reverse: GTGACAAGGGGCTTCCATCA

ANKS6 (168/289 bp)

Forward: CTTCGAGGTTGCACTGGACT

Reverse: CTGGGGTCTTCATCGGCAAT

PHLDB2 (150/171 bp)

Forward: CATCACCCCAAAGGCCCATC

Reverse: TGCCTGTGTCCTTCACTCAT

PBRM1 (175/295 bp)

Forward: ACTCTCCATTCCAGCTCTGTG

Reverse: CCATCATCAAAGTCCCCGCT

qRT-PCR primers used for Supplemental Fig. 1 are listed below:

ANKS6 (166 bp)

Forward: TCTGAACTGAACGCAGGCAA

Reverse: GGATTCAGAGAGCTCACGCT

IQGAP1 (178 bp)

Forward: TGCTGAAGGACTCGTTGCAT

Reverse: AGATTTCGGCGTTGGTCTGT

LEPRE1 (71 bp)

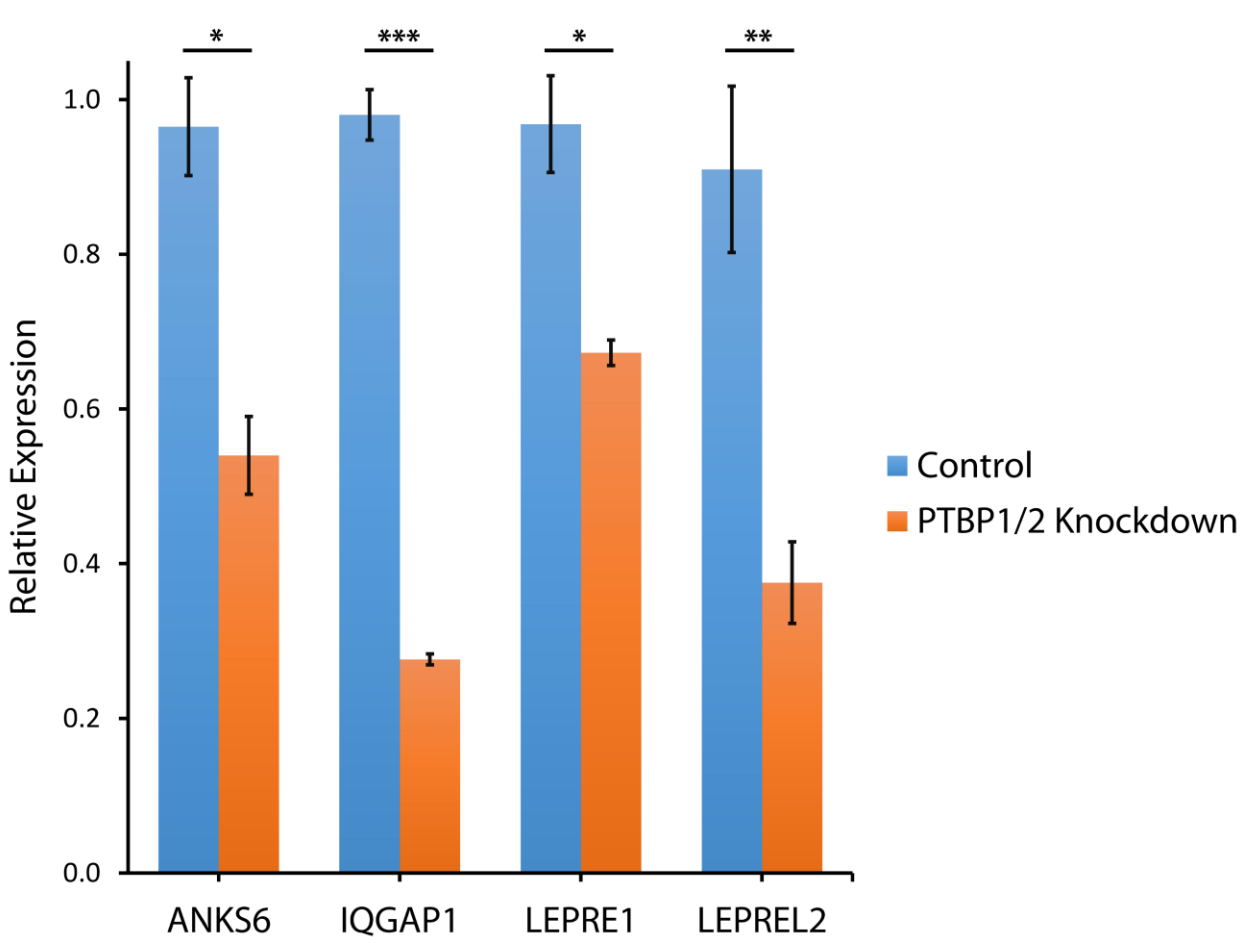
Forward: CACAGCGAGCGGGACAG

Reverse: AGGTCCATCTCTTCTGGGCT

LEPREL2 (156 bp)

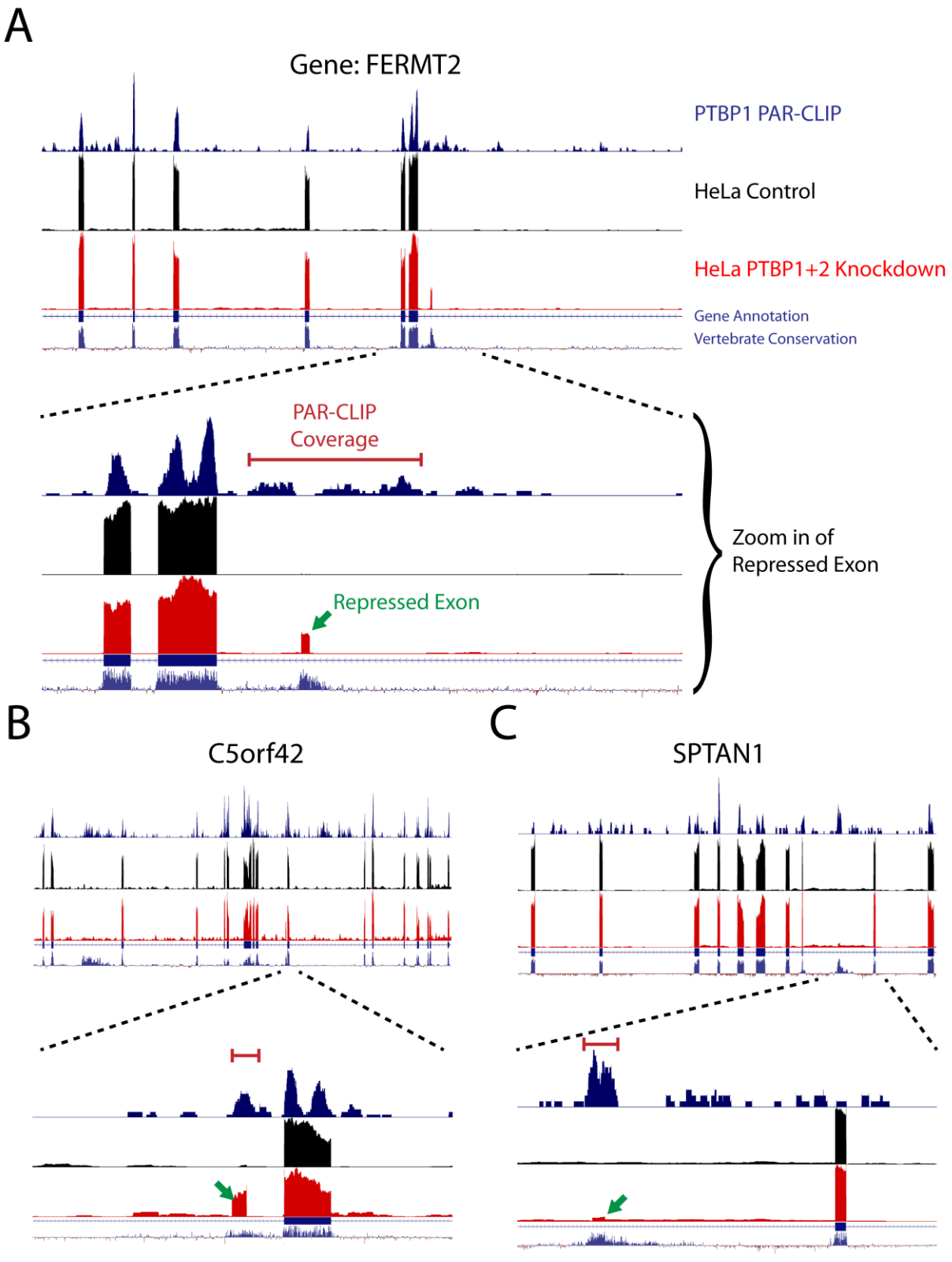
Forward: TATCGGGACTACAGCGGACT

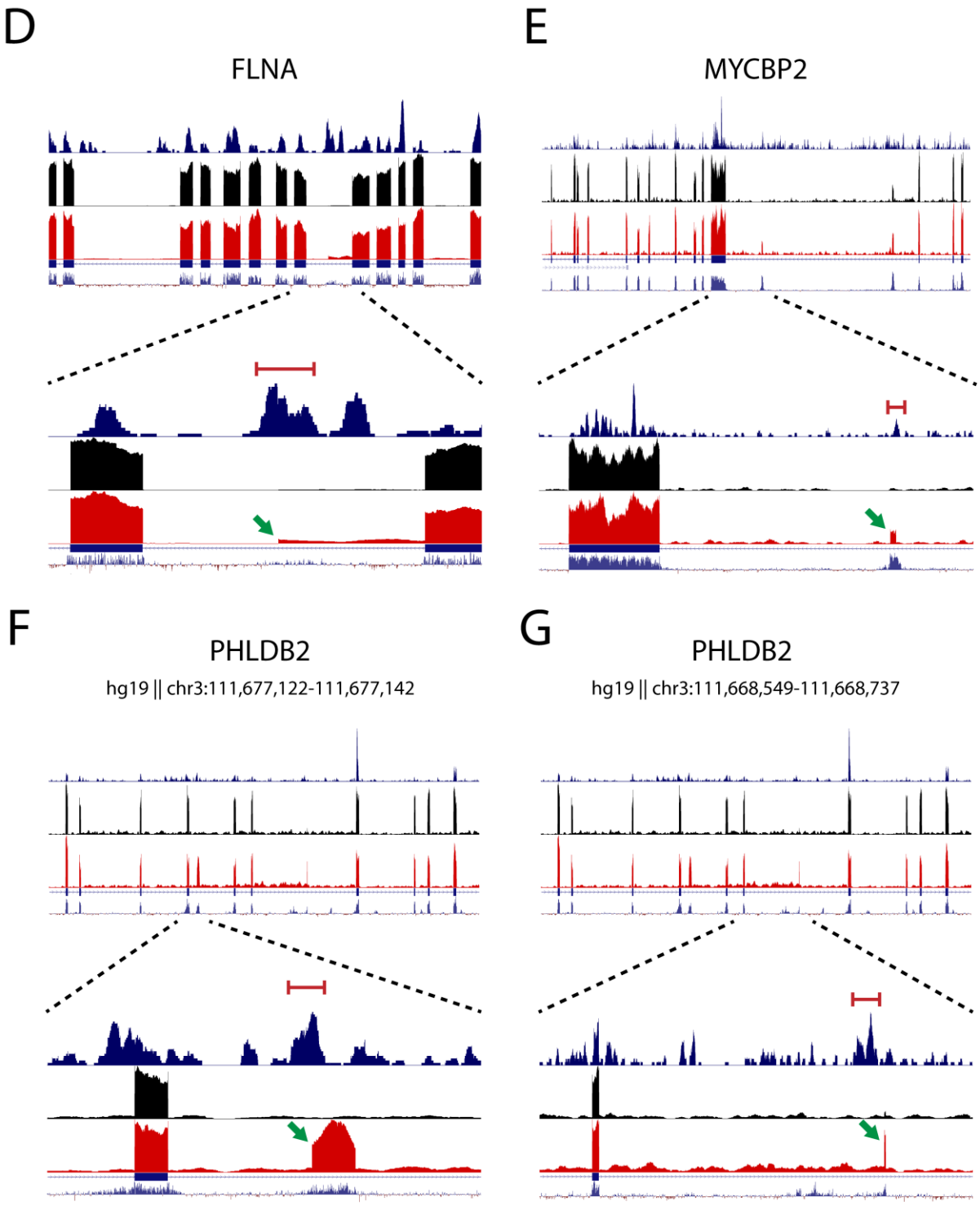
Reverse: CCCATGGGGATTCTCGACAC

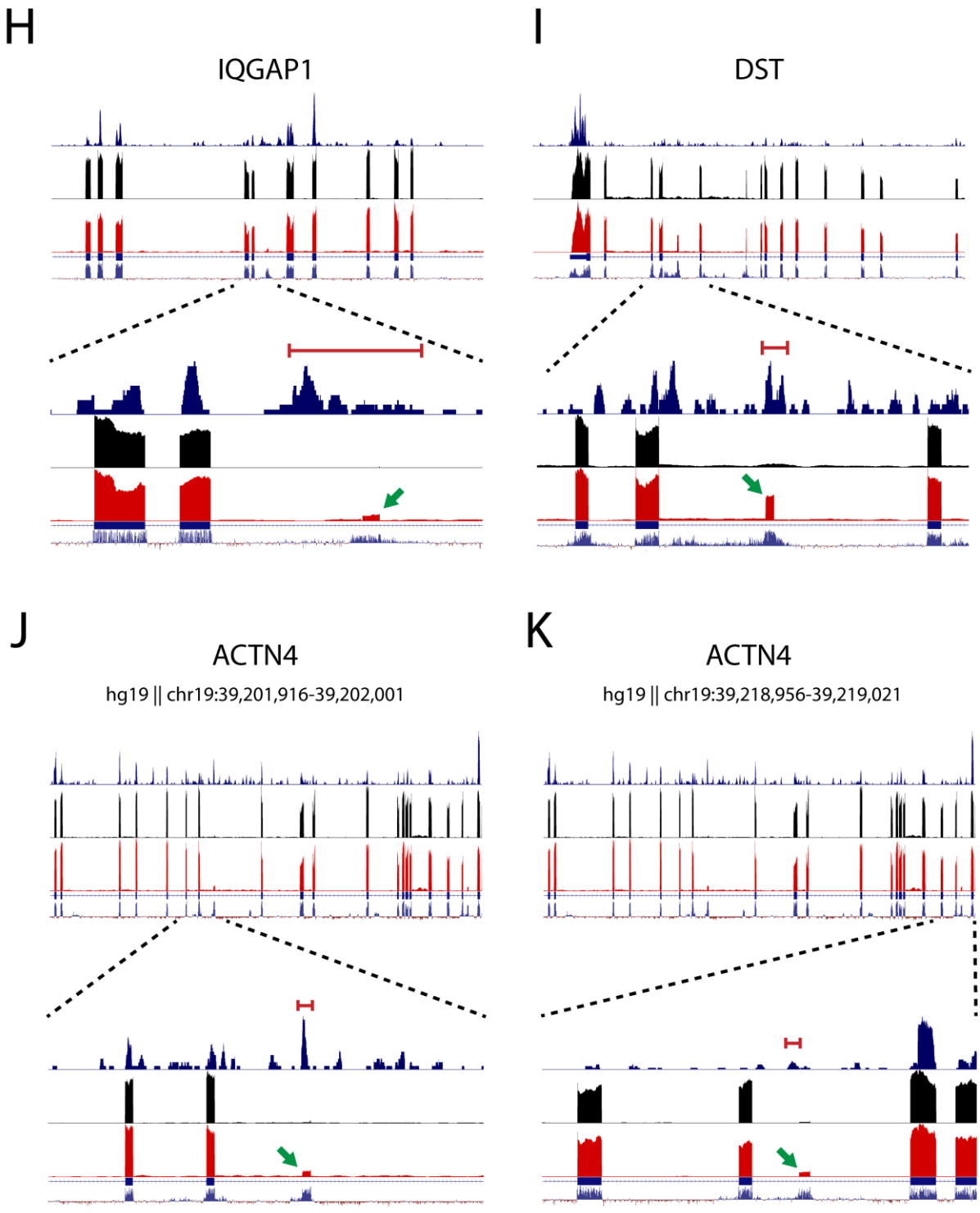


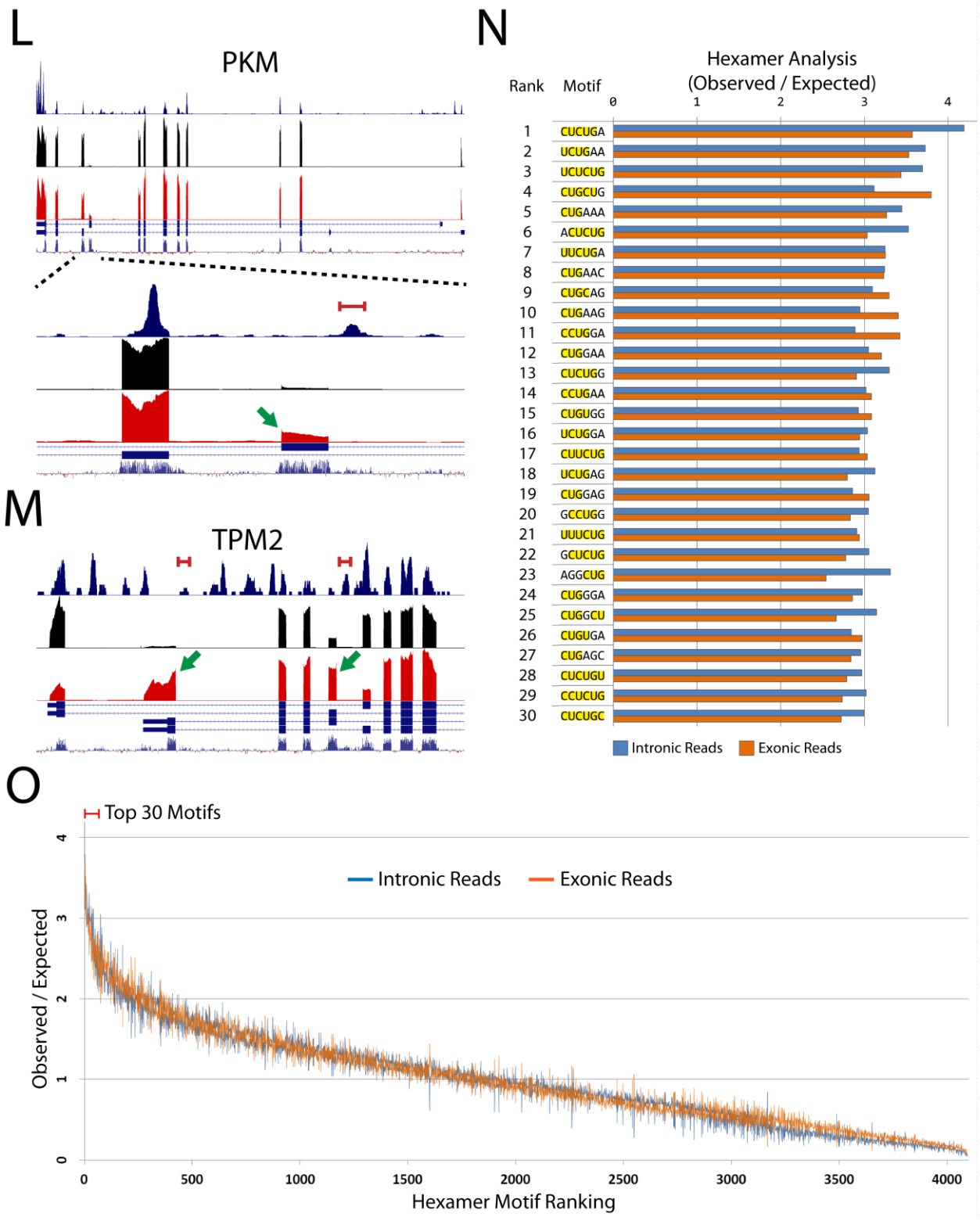
Supplemental Figure 1, related to Figure 2 | PTBP1/2 conserved and nonconserved repressed exons that induce nonsense-mediated decay lead to the downregulation of associated transcripts

qRT-PCR confirmation that PTBP1/2 repressed exons predicted to induce NMD result in downregulation of mRNA transcripts. Data are represented as mean ±SEM, n=3. *p<0.05; **p<0.005; ***p<0.001.


A
Gene: FERMT2
PTBP1 PAR-CLIP
HeLa Control
HeLa PTBP1+2 Knockdown
Gene Annotation
Vertebrate Conservation
PAR-CLIP
Coverage
Repressed Exon
Zoom in of
Repressed Exon
B
C5orf42
C
SPTAN1

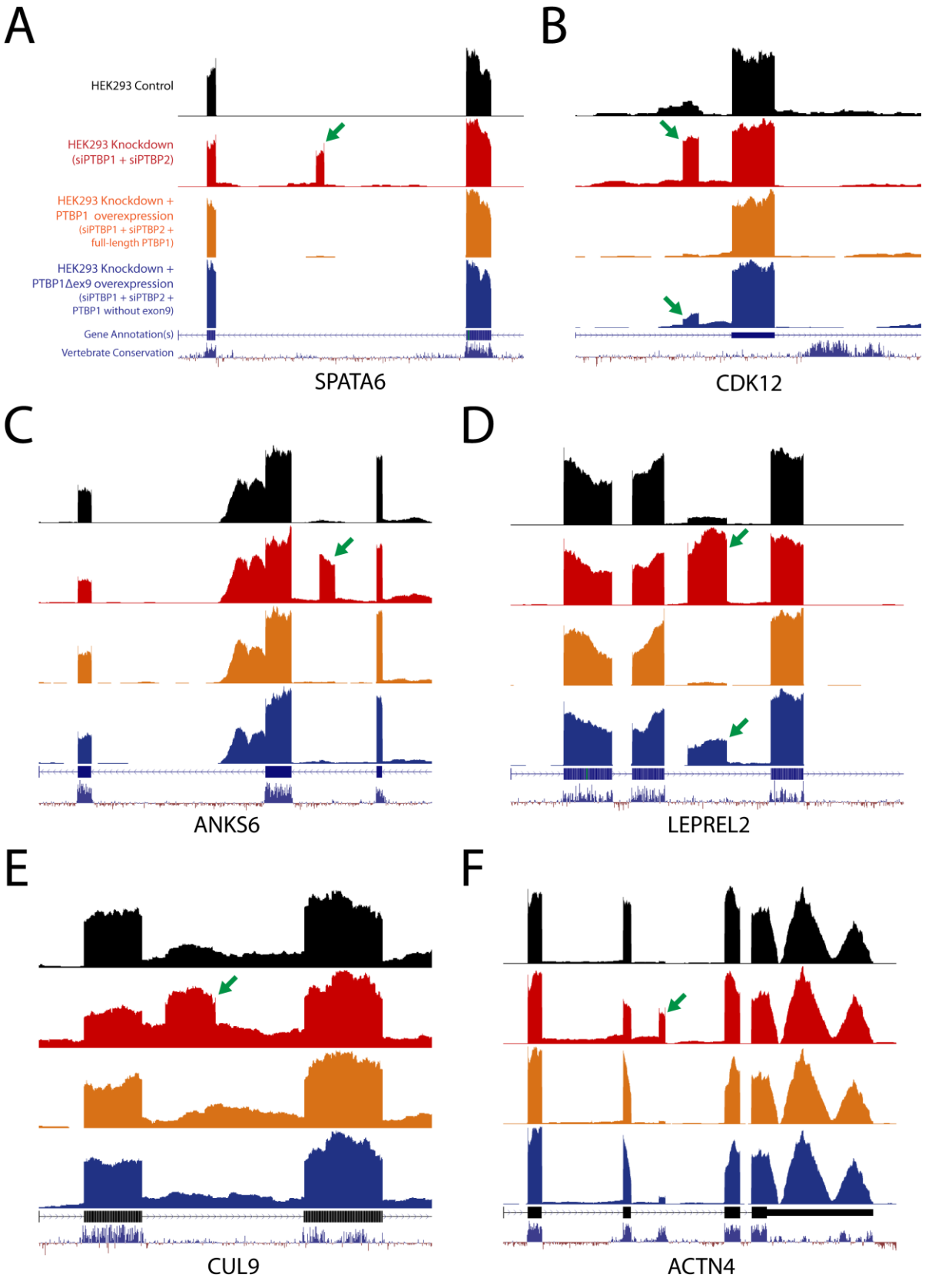

D
FLNA
E
MYCBP2
F
PHLDB2
hg19 || chr3:111,677,122-111,677,142
G
PHLDB2
hg19 || chr3:111,668,549-111,668,737

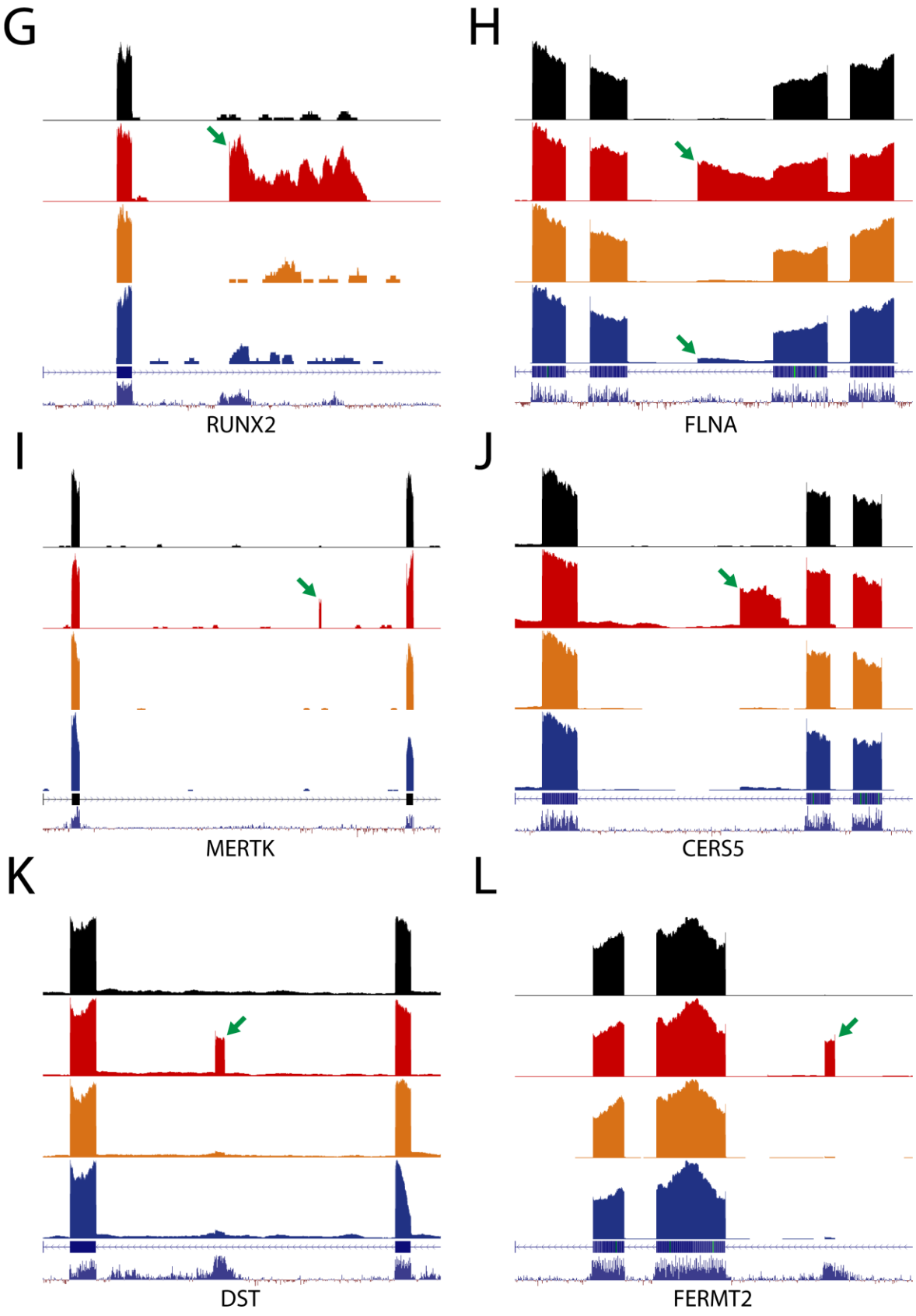

H
IQGAP1
I
DST
J
ACTN4
hg19 || chr19:39,201,916-39,202,001
K
ACTN4
hg19 || chr19:39,218,956-39,219,021


L
PKM
M
TPM2
N
Hexamer Analysis
(Observed / Expected)
Rank
Motif
0
1
2
3
4
1 CUCUGA
2 UCUGAA
3 UCUCUG
4 CUGCUG
5 CUGAAA
6 ACUCUG
7 UUCUGA
8 CUGAAC
9 CUGCAG
10 CUGAAG
11 CCUGGA
12 CUGGAA
13 CUCUGG
14 CCUGAA
15 CUGUGG
16 UCUGGA
17 CUUCUG
18 UCUGAG
19 CUGGAG
20 GCCUGG
21 UUUCUG
22 GCUCUG
23 AGGCUG
24 CUGGGA
25 CUGGCU
26 CUGUGA
27 CUGAGC
28 CUCUGU
29 CCUCUG
30 CUCUGC
Intronic Reads
Exonic Reads
O
Top 30 Motifs
Intronic Reads
Exonic Reads
Observed / Expected
0
1
2
3
4
0
500
1000
1500
2000
2500
3000
3500
4000
Hexamer Motif Ranking

Supplemental Figure 2, related to Figure 2 | PTBP1 PAR-CLIP analysis.

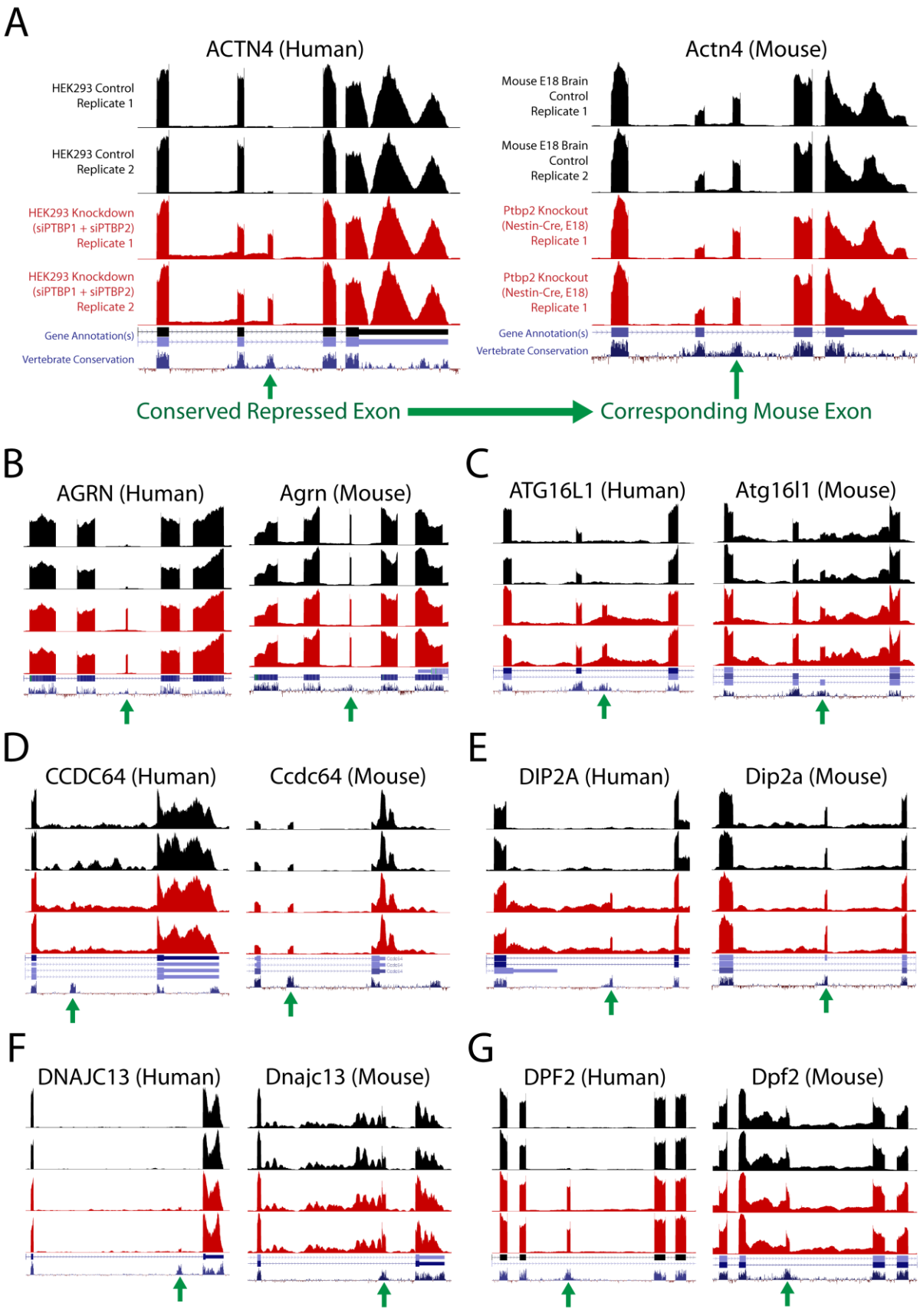
(**A**) Reads from PAR-CLIP were mapped to the human hg19 reference genome to visualize the direct RNA targets PTBP1 (top row, blue). Cross referencing with RNA-seq data from HeLa control (black) and HeLa double knockdown of PTBP1 and PTBP2 (red) revealed PAR-CLIP peaks (red bar) associated with the repressed exon of FERMT2 (green arrow). (**B**–**K**) Similar PAR-CLIP peaks can be found across other repressed exons, although many reads are also associated with exonic sequences. PTBP1 has been documented to have many other functions in mRNA processing[10-13] in addition to the role of splicing repression. Given the low complexity consensus sequence of PTBP1, it is hypothesized that PTBP1 binds ubiquitously to CU-rich elements and that its function is dictated more by the local genomic sequence and splicing factor environment than by PTBP1's RNA-binding domains. (**L** and **M**) Peaks can be observed at CU microsatellites directly upstream PKM exon 9a (**L**) and TPM2 exon 6a (**M**), both found to interact with PTBP1 from previous CLIP experiments performed in HeLa cells (Xue et al., 2009). (**N**) The top 30 hexamers present in intronic and exonic reads as compared to shuffled control. As expected pyrimidine rich motifs are over represented. Interestingly, many motifs also contain a trinucleotide CUG sequence. (**O**) All hexamer motifs ranked by the observed/expected ratio.

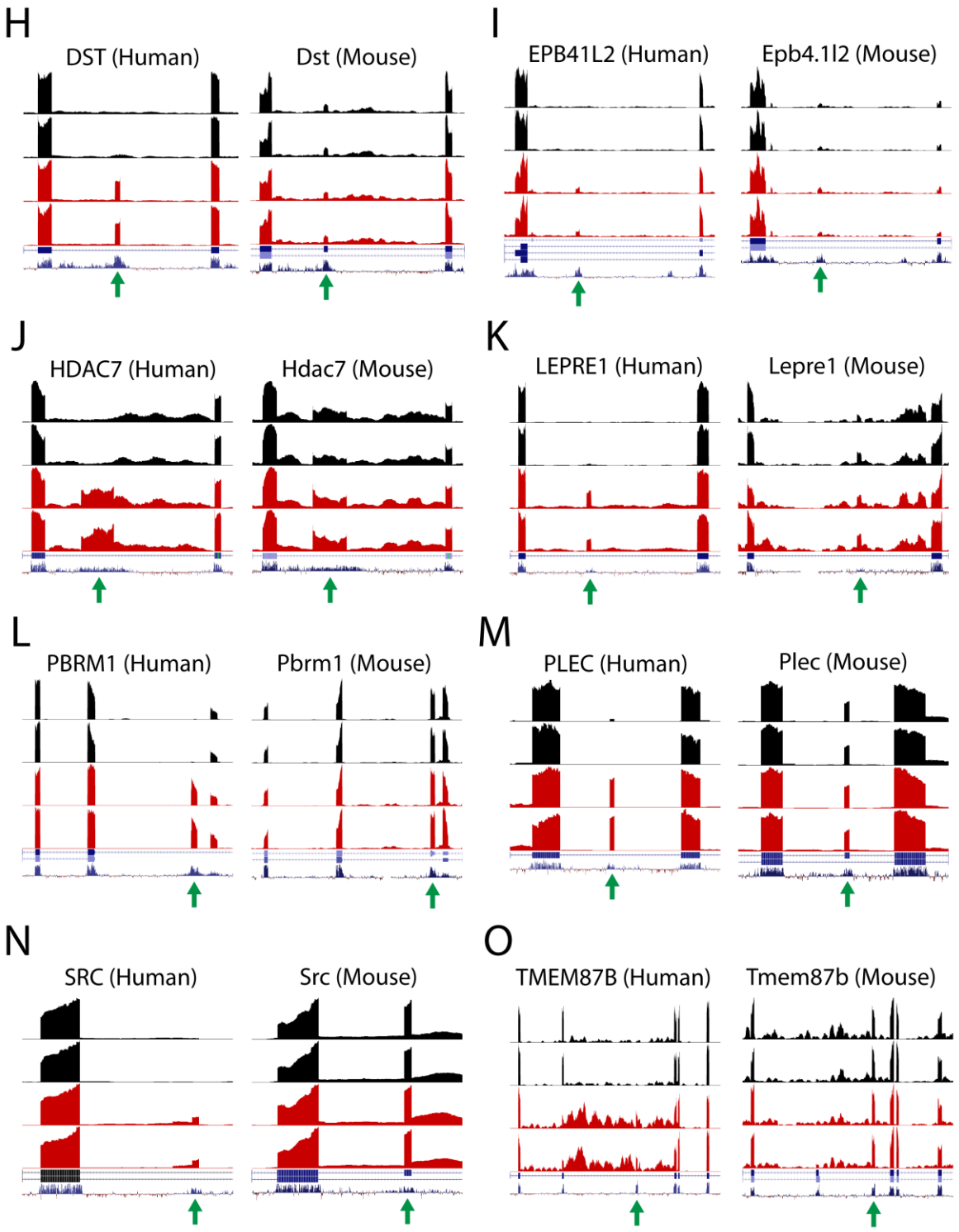

A
HEK293 Control
HEK293 Knockdown
(siPTBP1 + siPTBP2)
HEK293 Knockdown +
PTBP1 overexpression
(siPTBP1 + siPTBP2 +
full-length PTBP1)
HEK293 Knockdown +
PTBP1Δex9 overexpression
(siPTBP1 + siPTBP2 +
PTBP1 without exon9)
Gene Annotation(s)
Vertebrate Conservation
SPATA6
B
CDK12
C
ANKS6
D
LEPREL2
E
CUL9
F
ACTN4


G
RUNX2
H
FLNA
I
MERTK
J
CERS5
K
DST
L
FERMT2

Supplemental Figure 3, related to Figure 2 | Full-length PTBP1 and PTBP1Δex9 are capable of restoring exon repression.

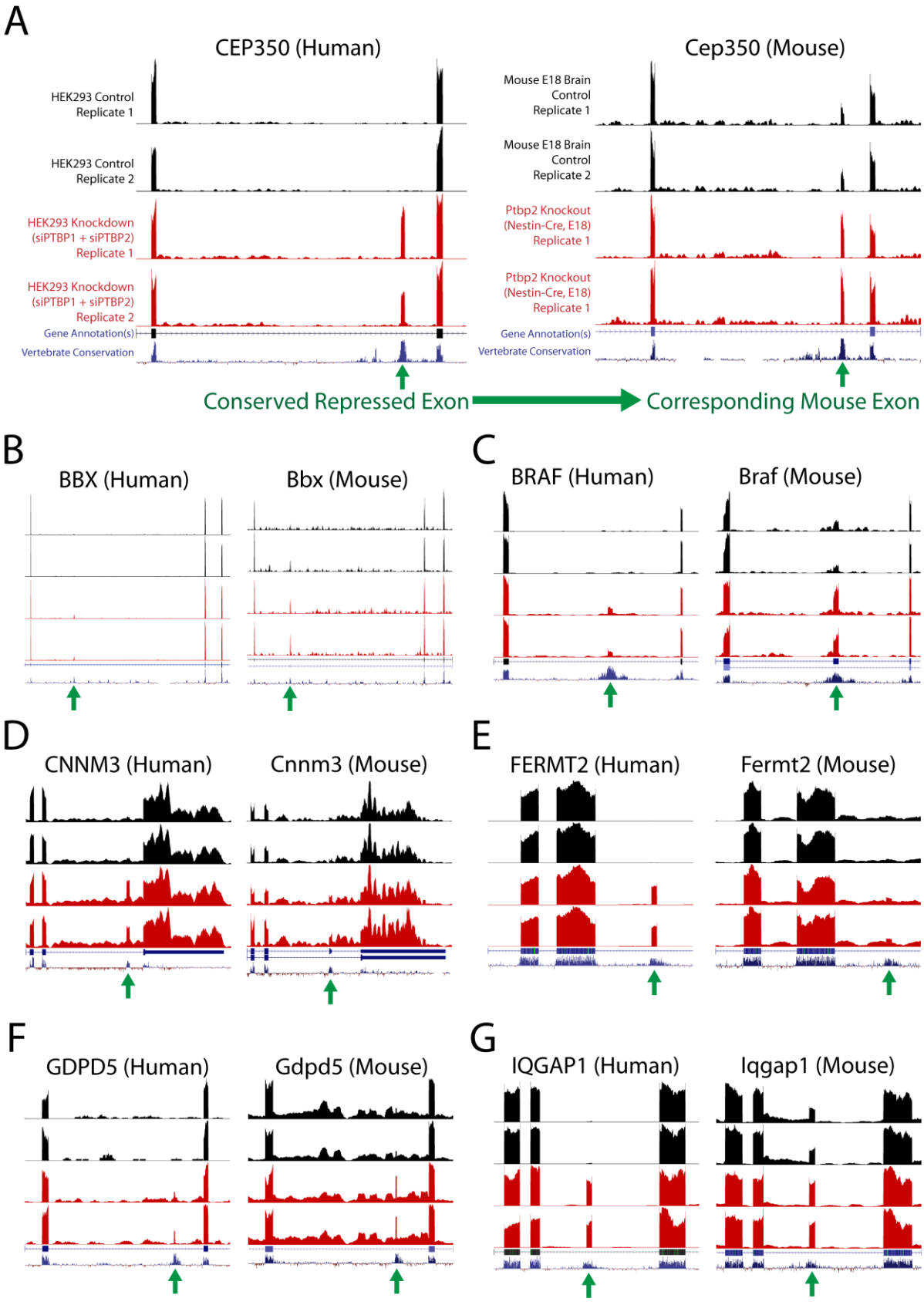
(**A**) RNA-seq coverage from HEK293 control (black), HEK293 double knockdown of PTBP1 and PTBP2 (red), HEK293 double knockdown with concurrent transfection of full-length PTBP1 (orange), and HEK293 double knockdown with concurrent transfection of PTBP1 without exon 9 (PTBP1Δex9, blue). Exon 9 of PTBP1 has been previously shown to alter the alternative splicing activity of PTBP1 (Gueroussov et al., 2015). Reintroducing full-length PTBP1 or PTBP1Δex9 in the double knockdown setting restores the repression of PTBP1/2 exons (green arrows). (**B**–**L**) Similar repression can be seen for many other cryptic exons. Interestingly, while the majority of exons are repressed by PTBP1Δex9, certain cryptic exons appear to be more resistant to PTBP1Δex9 overexpression (e.g. **B**, **D**, **F**, **H**).

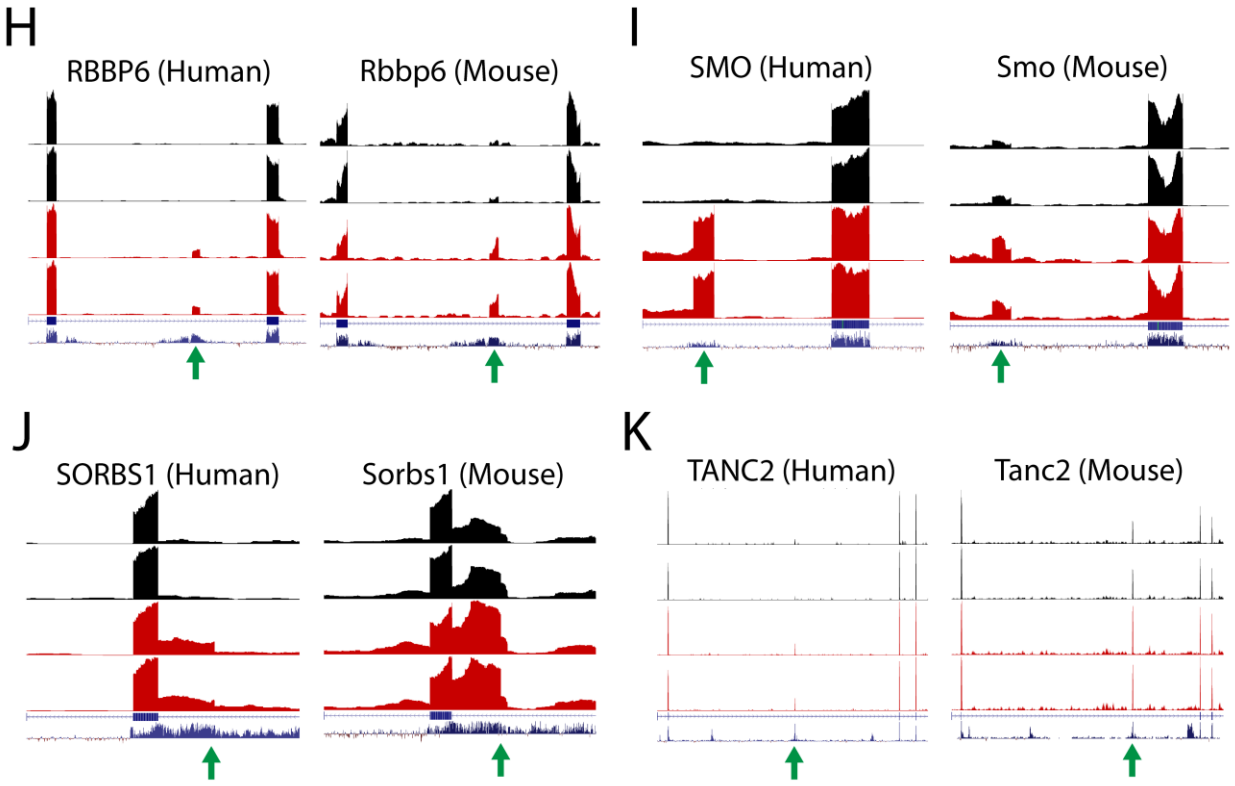

A
ACTN4 (Human)
HEK293 Control Replicate 1
HEK293 Control Replicate 2
HEK293 Knockdown (siPTBP1 + siPTBP2) Replicate 1
HEK293 Knockdown (siPTBP1 + siPTBP2) Replicate 2
Gene Annotation(s)
Vertebrate Conservation
Actn4 (Mouse)
Mouse E18 Brain Control Replicate 1
Mouse E18 Brain Control Replicate 2
Ptbp2 Knockout (Nestin-Cre, E18) Replicate 1
Ptbp2 Knockout (Nestin-Cre, E18) Replicate 1
Gene Annotation(s)
Vertebrate Conservation
Conserved Repressed Exon
Corresponding Mouse Exon
B
AGRN (Human)
Agrn (Mouse)
C
ATG16L1 (Human)
Atg16l1 (Mouse)
D
CCDC64 (Human)
Ccdc64 (Mouse)
E
DIP2A (Human)
Dip2a (Mouse)
F
DNAJC13 (Human)
Dnajc13 (Mouse)
G
DPF2 (Human)
Dpf2 (Mouse)



Supplemental Figure 4, related to Figure 4 | Ptbp2 knockout does not affect some repressed exons that are activated in neurons.

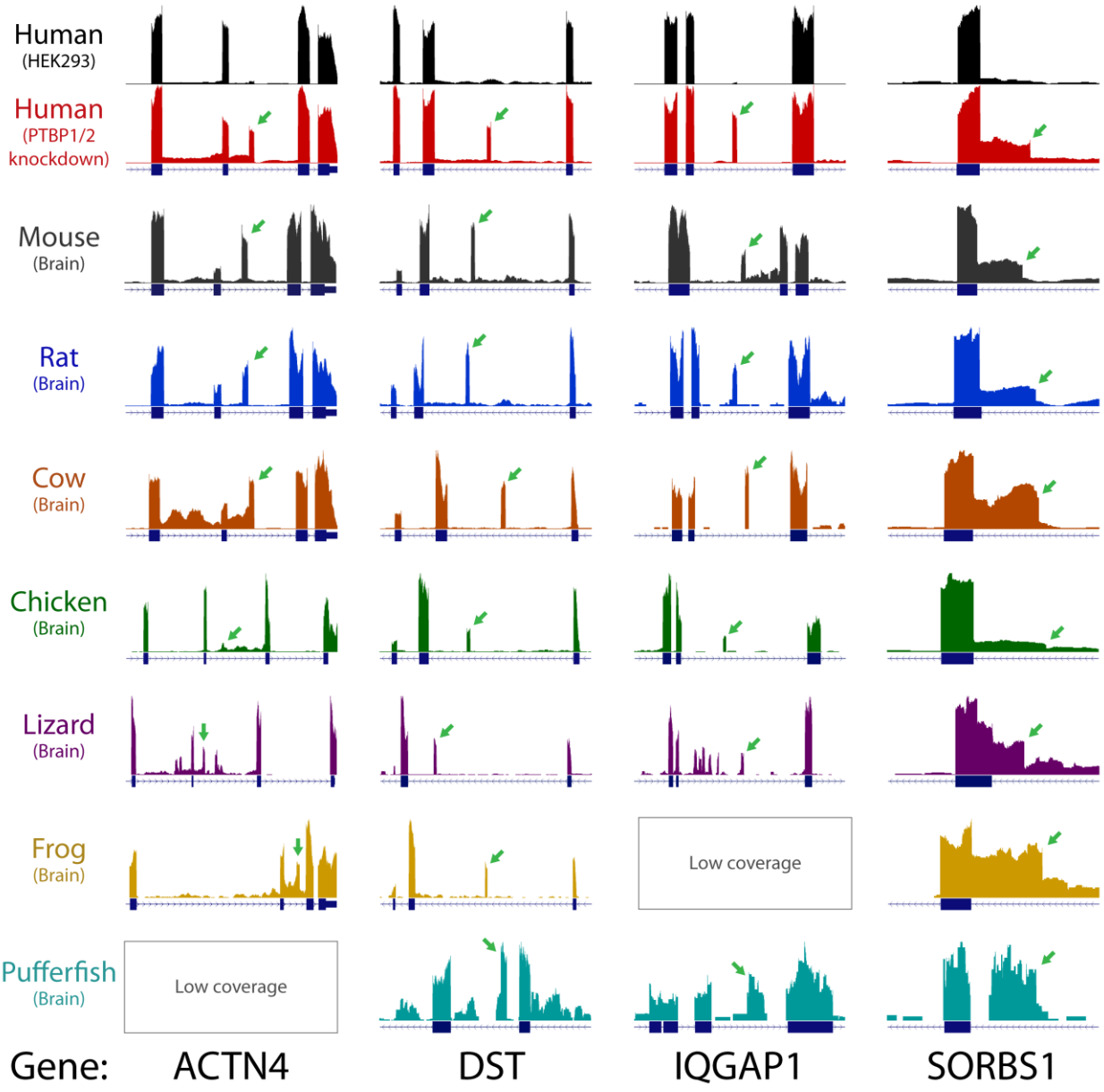
(**A**–**O**) Examples of mouse repressed exons that are activated in neurons where Ptbp2 knockout has no impact on exon splicing.

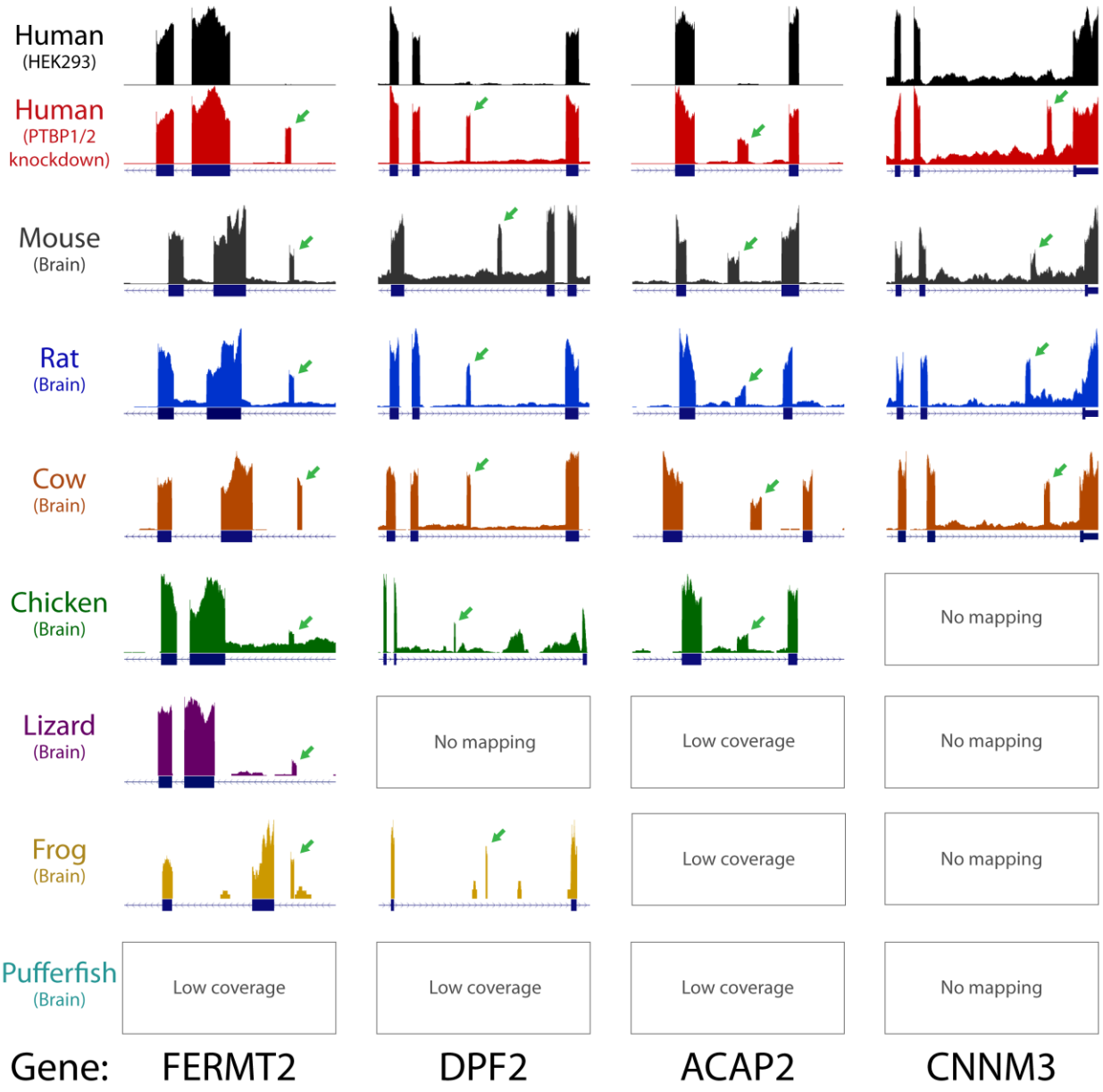

A
CEP350 (Human)
HEK293 Control Replicate 1
HEK293 Control Replicate 2
HEK293 Knockdown (siPTBP1 + siPTBP2) Replicate 1
HEK293 Knockdown (siPTBP1 + siPTBP2) Replicate 2
Gene Annotation(s)
Vertebrate Conservation
Cep350 (Mouse)
Mouse E18 Brain Control Replicate 1
Mouse E18 Brain Control Replicate 2
Ptbp2 Knockout (Nestin-Cre, E18) Replicate 1
Ptbp2 Knockout (Nestin-Cre, E18) Replicate 1
Gene Annotation(s)
Vertebrate Conservation
Conserved Repressed Exon
Corresponding Mouse Exon
B
BBX (Human)
Bbx (Mouse)
C
BRAF (Human)
Braf (Mouse)
D
CNNM3 (Human)
Cnnm3 (Mouse)
E
FERMT2 (Human)
Fermt2 (Mouse)
F
GDPD5 (Human)
Gdpd5 (Mouse)
G
IQGAP1 (Human)
Iqgap1 (Mouse)



Supplemental Figure 5, related to Figure 4 | Ptbp2 knockout modestly increases splicing of some repressed exons that are activated in neurons.

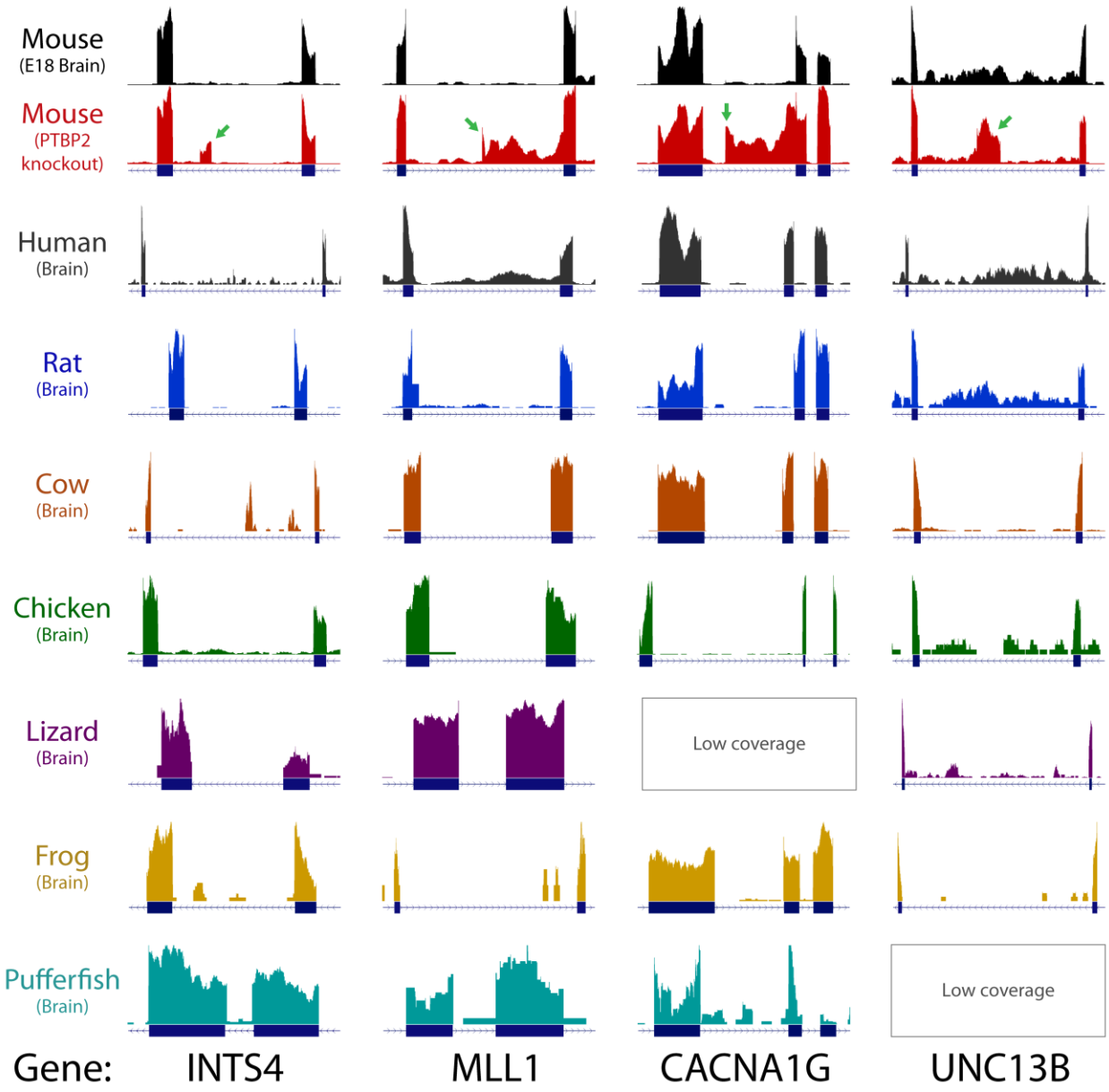
(**A**–**K**) Example of mouse repressed exons that are activated in neurons where Ptbp2 knockout results in a modest increase in exon splicing.

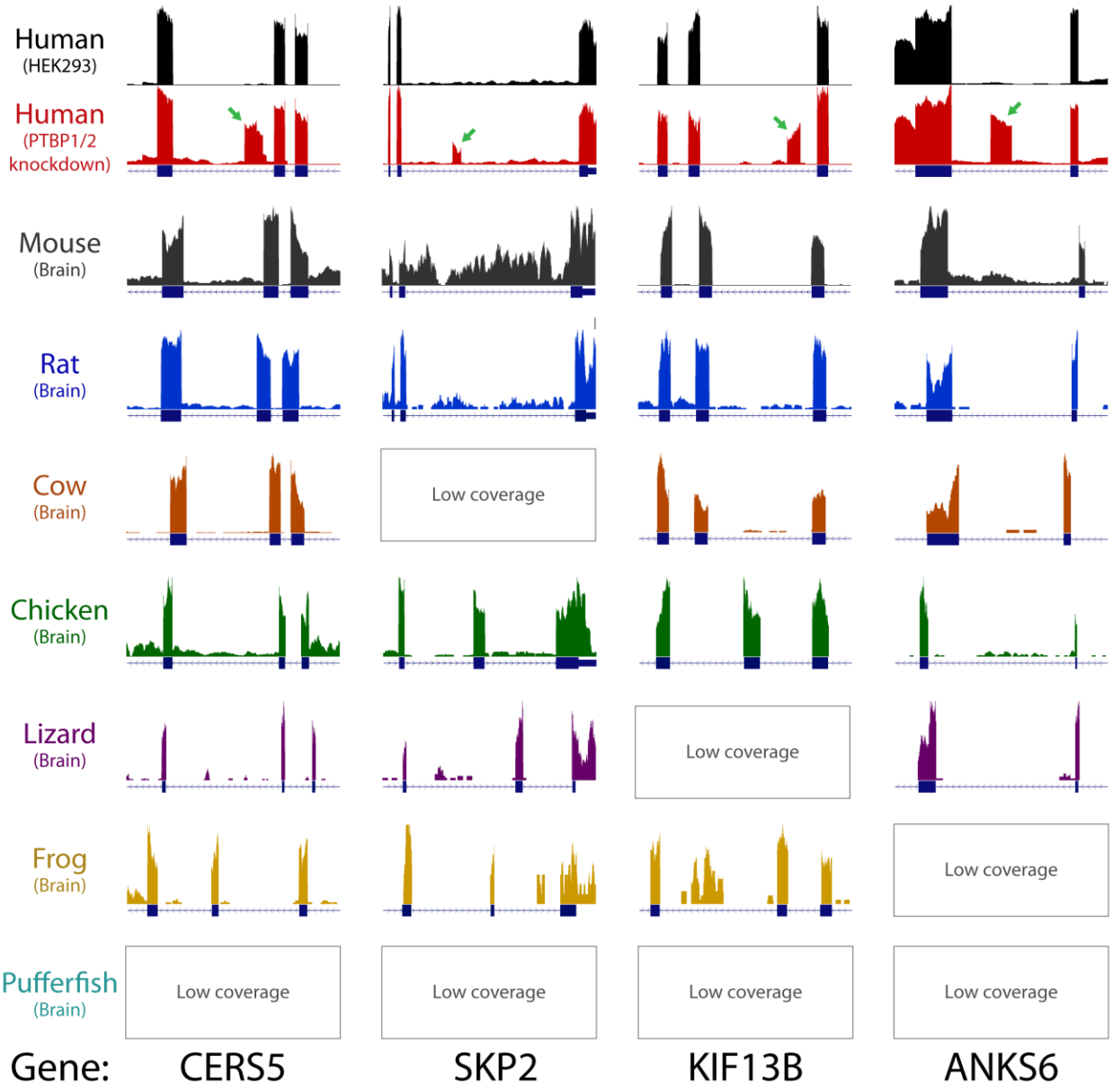

Human
(HEK293)
Human
(PTBP1/2
knockdown)
Mouse
(Brain)
Rat
(Brain)
Cow
(Brain)
Chicken
(Brain)
Lizard
(Brain)
Frog
(Brain)
Pufferfish
(Brain)
Low coverage
Low coverage
Gene:
ACTN4
DST
IQGAP1
SORBS1



Supplemental Figure 6, related to Figure 4 | Conserved PTBP1/2 repressed exons are found across vertebrate brains

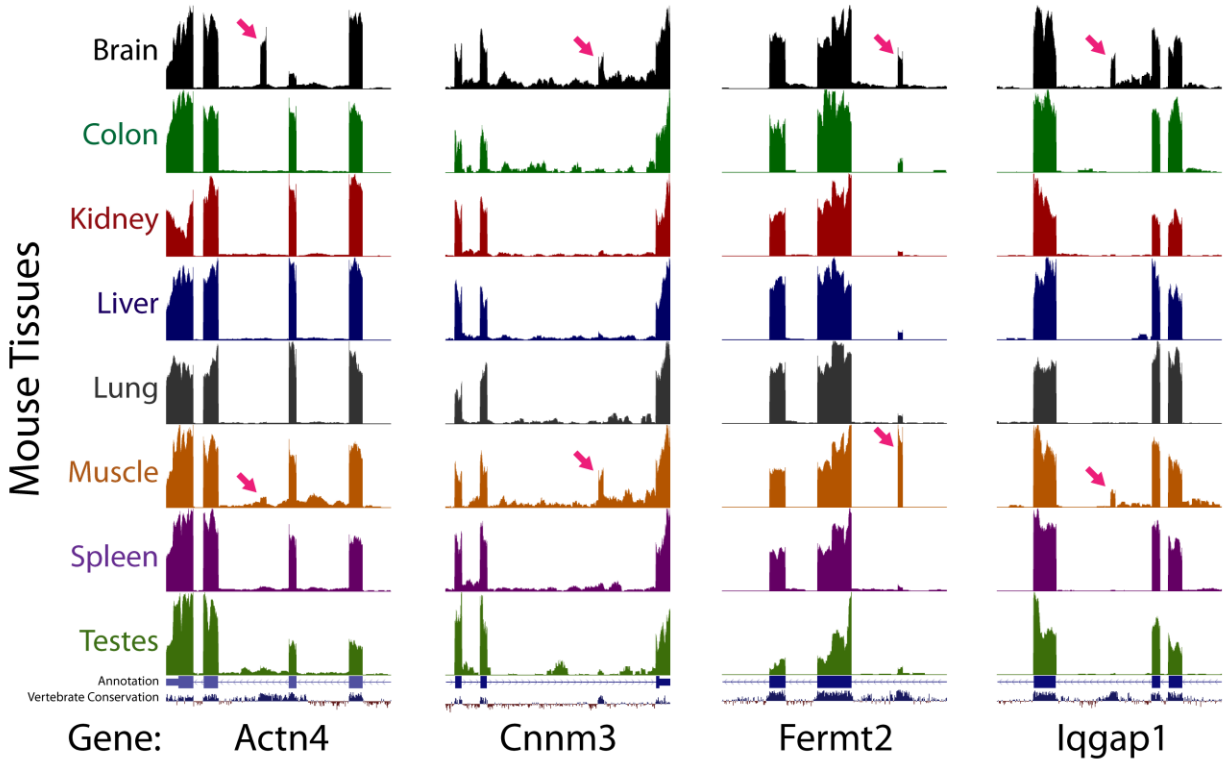
Analysis of conserved PTBP1/2 repressed exons and the equivalent genomic regions of other species confirms that these tissue-specific exons (green arrows) are present across many vertebrate brains. 'Low coverage' refers to insufficient read alignments in the genomic locus. 'No mapping' indicates that no homologous transcript can be identified in the species. The specific genome coordinates used for each species and gene are listed in the Supplemental Excel File.

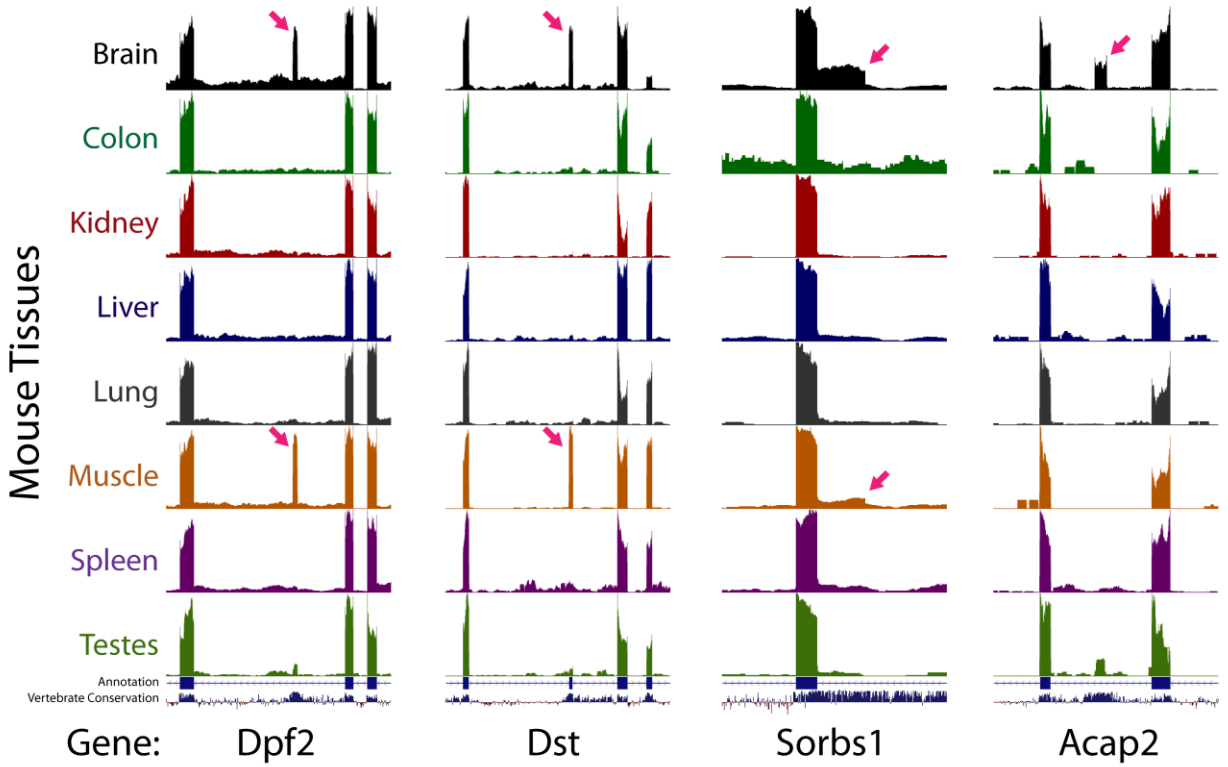

Mouse
(E18 Brain)
Mouse
(PTBP2
knockout)
Human
(Brain)
Rat
(Brain)
Cow
(Brain)
Chicken
(Brain)
Lizard
(Brain)
Frog
(Brain)
Pufferfish
(Brain)
Low coverage
Low coverage
Gene:
INTS4
MLL1
CACNA1G
UNC13B



Supplemental Figure 7, related to Figure 4 | Nonconserved PTBP1/2 cryptic exons are not present in the brains of other vertebrate species

Analysis of nonconserved PTBP1/2 cryptic exons and equivalent genomic regions of other species confirms that nonconserved cryptic exons (green arrows) are not present in other vertebrate brains. 'Low coverage' refers to insufficient read alignments in the genomic locus. 'No mapping' indicates that no homologous transcript can be identified in the species. The specific genome coordinates used for each species and gene are listed in the Supplemental Excel File.


Mouse Tissues
Brain
Colon
Kidney
Liver
Lung
Muscle
Spleen
Testes
Annotation
Vertebrate Conservation
Gene:
Actn4
Cnnm3
Fermt2
Iqgap1


Mouse Tissues
Brain
Colon
Kidney
Liver
Lung
Muscle
Spleen
Testes
Annotation
Vertebrate Conservation
Gene:
Dpf2
Dst
Sorbs1
Acap2

Supplemental Figure 8, related to Figure 4 | Conserved tissue-specific exons are present in brain and muscle but not in other tissues

Analysis of RNA-seq data from various mouse tissues reveals that conserved tissue-specific exons that are normally repressed by PTBP1/2 in undifferentiated cells, can instead be found in brain and muscle—tissues that express low levels of PTBP1 (Lilleväli et al., 2001)—but not in other tissues such as colon, kidney, liver, lung, spleen, or testes.

Gene Symbol	Full Name	Human Cryptic Exon Location (hg19)	Inf. Ins.	5' UTR or ncRNA	Exon Ext.	NMD	UCSC ID (HEK293)	HEK293 CTRL (FPKM)	HEK293 KD (FPKM)	HEK293 CTRL vs KD (FC)
ACAP2	Centaurin, Beta 2	chr3:195,015,925-195,016,029	Y	--	--	--	uc003fun.4	5.6	6.1	1.1
ACTN4	Actinin, Alpha 4	chr19:39,218,956-39,219,021	Y	--	--	--	uc002oja.2	132.3	68.2	-1.9
ACTN4	Actinin, Alpha 4	chr19:39,204,349-39,204,426	Y/N	--	--	--	uc002oja.2	132.3	68.2	-1.9
ADAM22	ADAM metallopeptidase domain 22	chr7:87,808,358-87,808,444	Y	--	Y	--	uc003ujo.3	2.3	1.2	-1.9
AGRN	Agrin	chr1:986,412-986,423	Y	--	--	--	uc001ack.2	26.8	21.2	-1.3
ARHGEF11	Rho guanine nucleotide exchange factor (GEF) 11	chr1:156,957,502-156,957,534	Y	--	--	--	uc001fqo.3	5.1	2.5	-2.1
ATG16L1	Autophagy Related 16-Like 1 (S. Cerevisiae)	chr2:234,182,637-234,182,687	Y	--	--	--	uc002vua.3	4.6	3.2	-1.5
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase	chr7:140,485,489-140,485,608	Y	--	--	--	uc003vwc.4	5.2	6.4	1.2
CEP350	Centrosomal Protein 350kDa	chr1:180,042,921-180,043,043	Y	--	--	--	uc001gnt.3	3.6	3.1	-1.2
DIP2A	Disco-Interacting Protein 2 Homolog A	chr21:47,979,747-47,979,776	Y	--	--	--	uc002zjo.2	7.2	5.6	-1.3
DPF2	D4, Zinc And Double PHD Fingers Family 2	chr11:65,112,051-65,112,092	Y	--	--	--	uc001odm.3	25.5	24.6	-1.0
DST	Dystonin	chr6:56,329,483-56,329,554	Y	--	--	--	uc003pcy.4	11.1	12.1	1.1
FERMT2	Fermitin Family Member 2	chr14:53,348,514-53,348,546	Y	--	--	--	uc001xac.3	3.1	20.0	6.4
GDPD5	Glycerophosphodiester Phosphodiesterase 2	chr11:75,150,385-75,150,399	Y	--	--	--	uc001owo.4	0.5	0.5	1.1
PLEC	Plectin	chr8:145,012,569-145,012,583	Y	--	--		uc003zaj.2	2.2	1.8	-1.2
PUM2	Pumilio-2	chr2:20,524,811-20,524,882	Y/N	--	--	--	uc002rdu.1	12.2	7.4	-1.7
RBBP6	Retinoblastoma Binding Protein 6	chr16:24,569,742-24,569,855	Y	--	--	--	uc002dmh.3	14.6	13.0	-1.1
SORBS1	Sorbin And SH3 Domain Containing 1	chr10:97,081,779-97,081,910	Y	--	Y	--	uc010qoe.2	1.5	2.1	1.4
SRC	SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase	chr20:36,014,863-36,014,880	Y	--	--	--	uc002xgy.3	13.0	9.9	-1.3
VTI1A	Vesicle Transport V-SNARE Protein Vti1-Like 2	chr10:114,293,289-114,293,309	Y	--	--	--	uc001kzz.3	5.2	4.2	-1.2
PBRM1	Polybromo 1	chr3:52,715,711-52,715,836	--	5'UTR	--	--	uc003dew.2	6.3	4.7	-1.3
PBRM1	Polybromo 1	chr3:52,715,720-52,715,836	--	5'UTR	--	--	uc003dew.2	6.3	4.7	-1.3
ACTN4	Actinin, Alpha 4	chr19:39,201,916-39,202,001	--	--	--	N	uc002oja.2	132.3	68.2	-1.9
CCDC64	Coiled-Coil Domain Containing 64	chr12:120,528,715-120,528,823	--	--	--	N	uc001txl.1	1.7	2.2	1.3
CNNM3	Cyclin-M3	chr2:97,497,706-97,497,805	--	--	--	N	uc002swy.3	11.3	7.5	-1.5
DNAJC13	RME-8	chr3:132,256,002-132,256,110	--	--	--	N	uc003eor.3	5.9	4.2	-1.4
LIMS1	LIM And Senescent Cell Antigen-Like Domains 1	chr2:109,297,425-109,297,616	--	--	--	N	uc002tel.3	12.2	7.8	-1.6
PICALM	Phosphatidylinositol Binding Clathrin Assembly Protein	chr11:85,671,718-85,671,820	--	--	--	N	uc001pbm.3	8.8	13.7	1.6

Supplemental Table 1, related to Figure 4 | Neuron-included exons regulated by PTBP1 and PTBP2 are often in-frame insertions.

Many conserved cryptic exons identified in the human PTBP1 + PTBP2 double knockdown RNA-seq datasets are actively spliced in mouse brain. The examples listed here illustrate the high percentage of exons that are in-frame insertions. Refer to the Supplemental Excel File for a complete list.

Gene Symbol	Full Name	Human Cryptic Exon Location (hg19)	Inf. Ins.	5' UTR or ncRNA	Exon Ext.	NMD	UCSC ID (HEK293)	HEK293 CTRL (FPKM)	HEK293 KD (FPKM)	HEK293 CTRL vs KD (FC)
IQGAP1	IQ Motif Containing GTPase Activating Protein 1	chr15:91,025,934-91,025,981	--	--	--	Y	uc002bpl.1	29.4	3.6	-8.2
SNAP23	Synaptosomal-Associated Protein, 23kDa	chr15:42,813,812-42,813,906	--	--	--	Y	uc001zpz.2	23.0	4.9	-4.7
LEPRE1	Prolyl 3-Hydroxylase 1	chr1:43,222,072-43,222,126	--	--	--	Y	uc001chv.2	17.1	6.3	-2.7
SMO	Smoothened, Frizzled Class Receptor	chr7:128,842,464-128,842,577	--	--	--	Y	uc003vor.3	19.0	8.2	-2.3
FLNB	Filamin B, Beta	chr3:58,085,365-58,085,462	--	--	--	Y	uc010hnf.2	11.3	5.1	-2.2
HDAC7	Histone Deacetylase 7	chr12:48,183,989-48,184,246	--	--	--	Y	uc001rqj.4	5.0	3.2	-1.6
FMNL3	Formin-Like 3	chr12:50,052,034-50,052,118	--	--	--	Y	uc001ruu.1	1.4	1.2	-1.2

Supplemental Table 2, related to Figure 4 | Certain neuron-included exons regulated by PTBP1 and PTBP2 appear to undergo NMD.

Many conserved cryptic exons identified in the human PTBP1 + PTBP2 double knockdown RNA-seq datasets are actively spliced in mouse brain. The exons identified here are predicted to undergo NMD. Refer to the Supplemental Excel File for a complete list.

Supplemental Excel File

Sheet 1. Human PTBP1/2 repressed exons (conserved and nonconserved)

Sheet 2. Mouse Ptbp1/2 repressed exons (conserved and nonconserved)

Sheet 3. List of human PTBP1/2 repressed exons that are actively spliced in mouse neurons

Sheet 4. Metadata for Supplemental Figures 6-8