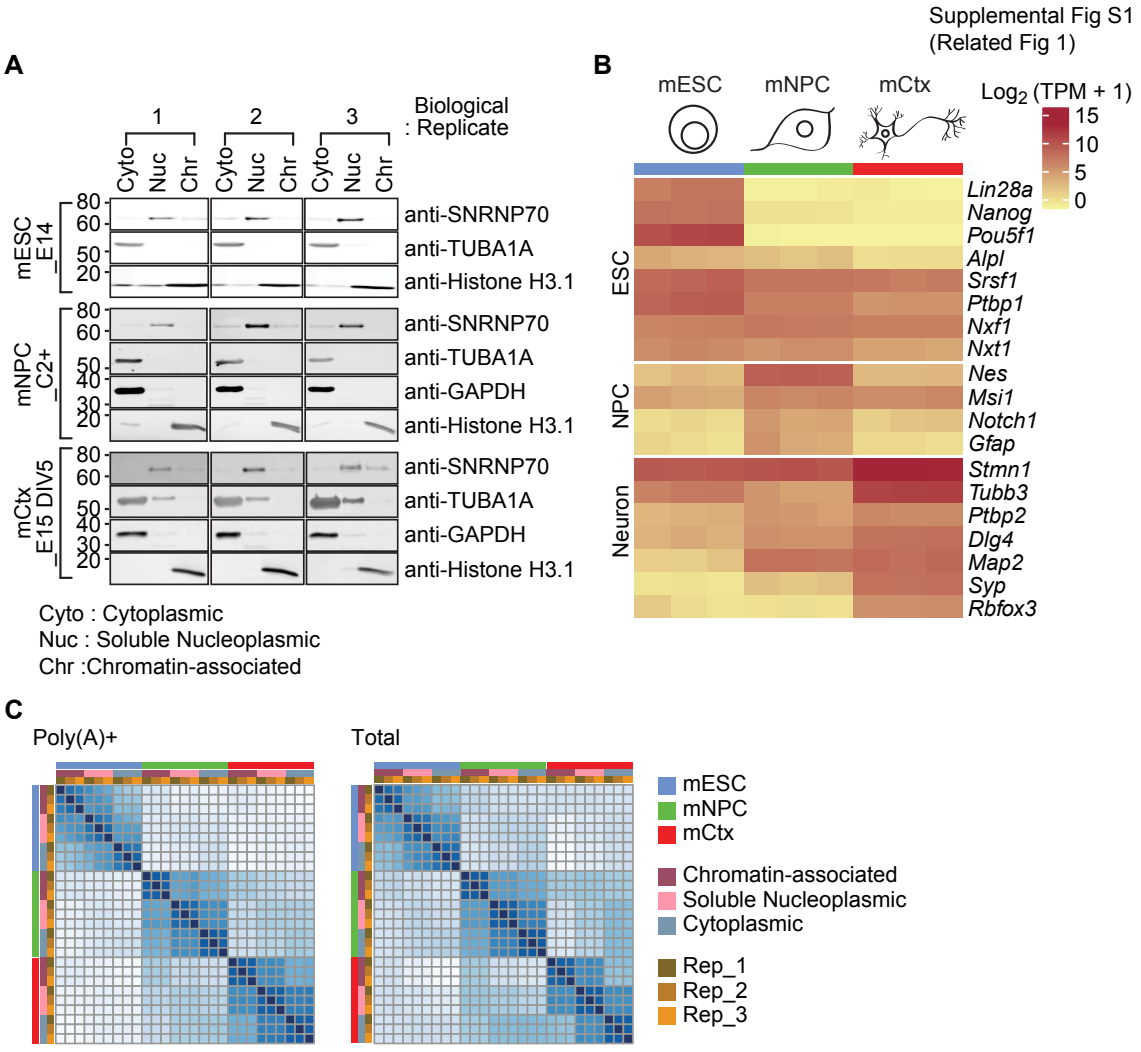
# Supplementary Materials for

## Tracking pre-mRNA maturation across subcellular compartments identifies developmental gene regulation through intron retention and nuclear anchoring.

Kyu-Hyeon Yeom[1,†], Zhicheng Pan[2,3,†], Chia-Ho Lin[1], Han Young Lim[1,4], Wen Xiao[1], Yi Xing[3,5,*], Douglas L. Black[1,*].

Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA 90095
Bioinformatics Interdepartmental Graduate Program, University of California, Los Angeles, CA 90095
Center for Computational and Genomic Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104
Molecular Biology Interdepartmental Doctoral Program, University of California, Los Angeles, Los Angeles, CA 90095
Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104

**This PDF file includes:**



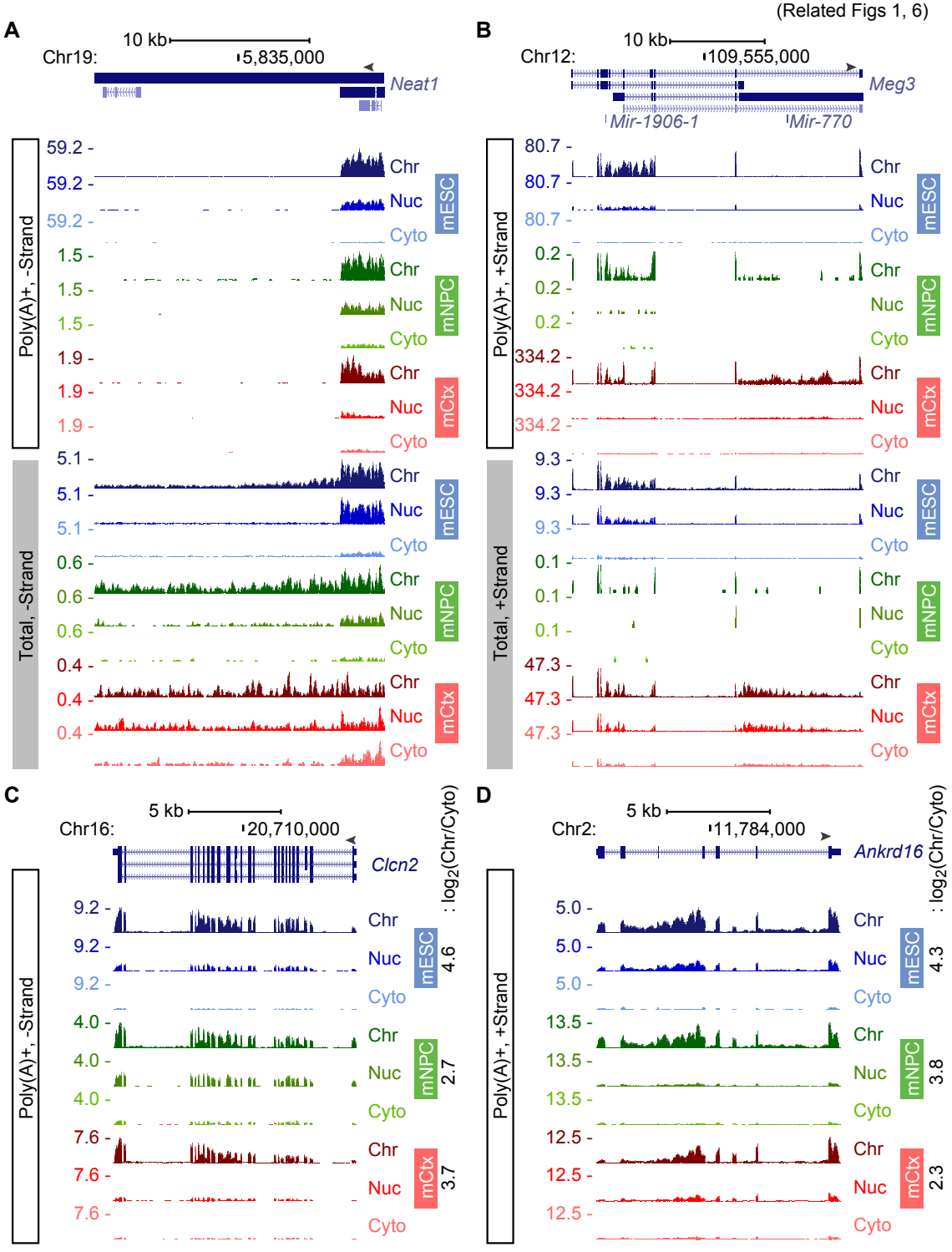

**Supplemental Figure S1. (Related to Figure 1)**

Validation of subcellular fractionation, cell type gene expression, and library consistency.

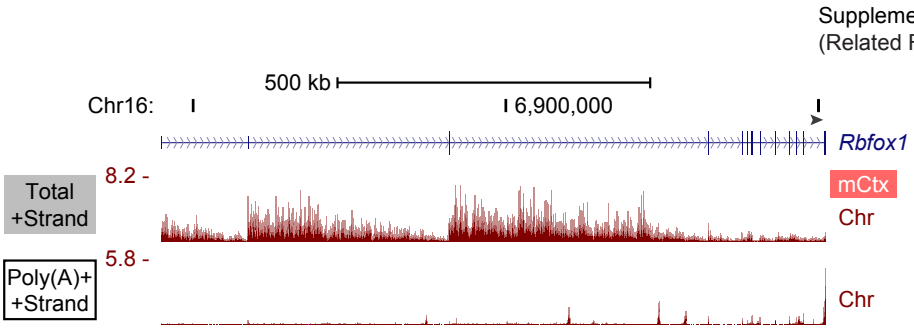
A. Confirmation of subcellular fractionation. Immunoblot analysis of diagnostic proteins in subcellular fractions. SNRNP70 for soluble nucleoplasm (Nuc), TUBA1A and GAPDH for cytoplasm (Cyto), and Histone H3.1 for chromatin pellet (Chr). Gel images include 3 biological replicates of mouse embryonic stem cells (line E14), mouse neuronal progenitor cell line C2+, and mouse cortical neurons after 5 days *in vitro* culture (E15DIV5; mCtx). Note that the immunoblot results of the third replicate of mESC_E14 are reprinted from Yeom et al. (Yeom et al. 2018).

B. Confirmation of cell type specific gene expression. Heatmap presents the cytoplasmic expression as measured by kallisto for the indicated mRNAs in each cell type and replicate.

C. Confirmation of library similarity. Heatmap displays similarity of gene expression between pairwise comparisons of all cell types, fractions, and replicates. Color codes are indicated on right.

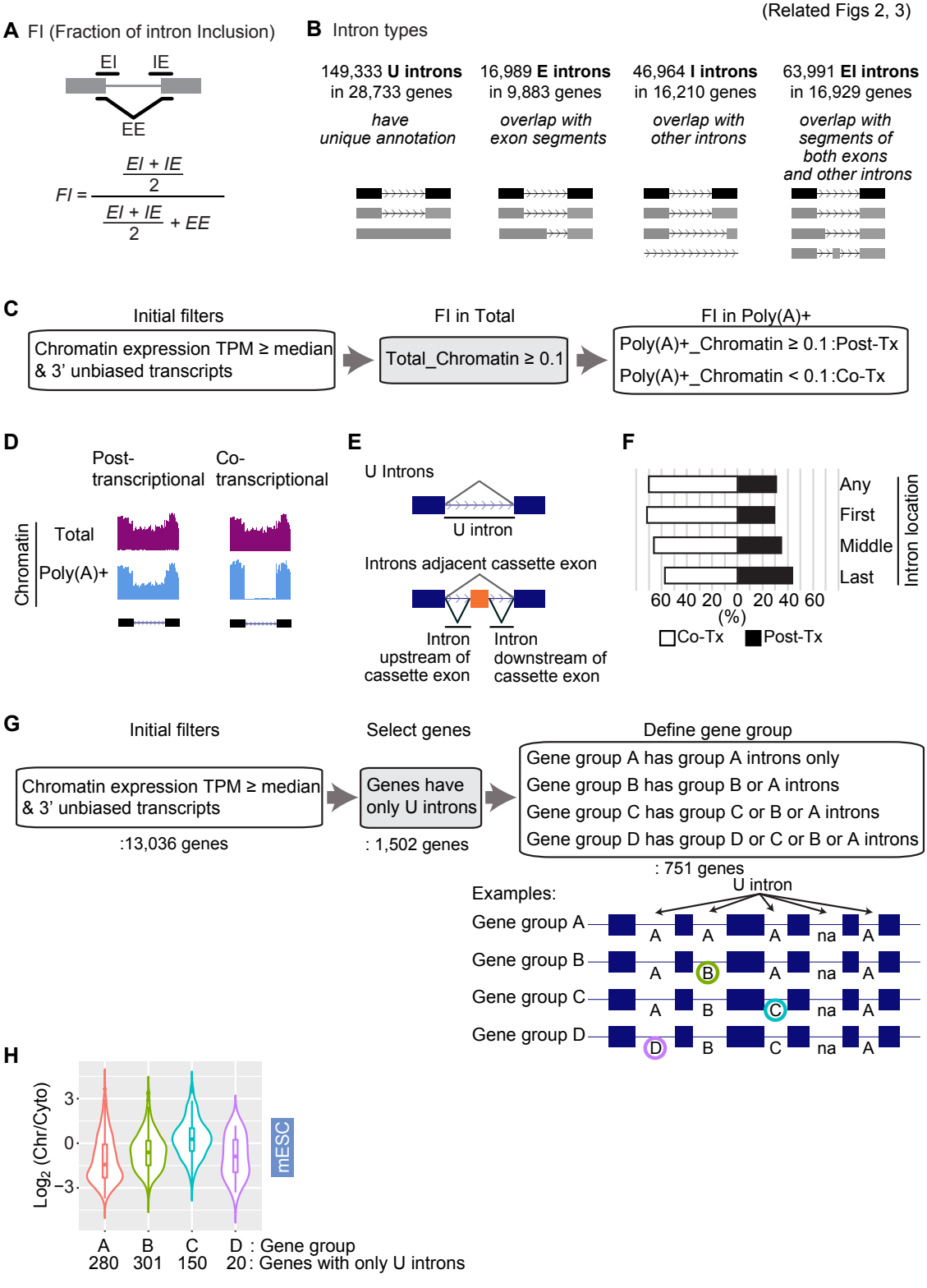

A
10 kb
Chr19:
5,835,000
Neat1
Poly(A)+, -Strand
Total, -Strand
Chr
Nuc
Cyto
mESC
mNPC
mCtx
B
10 kb
Chr12:
109,555,000
Meg3
Mir-1906-1
Mir-770
Poly(A)+, +Strand
Total, +Strand
C
5 kb
Chr16:
20,710,000
Clcn2
: log2(Chr/Cyto)
4.6
2.7
3.7
D
5 kb
Chr2:
11,784,000
Ankrd16
4.3
3.8
2.3

**Supplemental Figure S2. (Related to Figures 1 and 6)**
Example genome browser tracks of non-coding and coding RNAs.
A. *Neat1* expression in mESC, mNPC, and mCtx. Genome browser tracks of the *Neat1* locus for poly(A)+ and total libraries. Y axis shows RPM scaled to the highest value in the Chromatin-associated fraction.
B. *Meg3* expression in mESC, mNPC, and mCtx. Genome browser tracks of the *Meg3* locus in poly(A)+ and total libraries.
C. Genome browser tracks of the *Clcn2* locus in mESC, mNPC, and mCtx. Transcripts are enriched in the chromatin fraction and exhibit unspliced introns in poly(A)+ RNA. The partition index of *Clcn2* in each cell type is indicated on the right.
D. Genome browser tracks of the *Ankrd16* locus in mESC, mNPC, and mCtx. Transcripts are enriched in the chromatin fraction and exhibit unspliced introns in poly(A)+ RNA. Partition index of *Ankrd16* in each cell type is indicated on the right.




**Supplemental Figure S3. (Related to Figure 2)**
Very long introns exhibit declining reads 5' to 3' to create a sawtooth pattern. Genome browser tracks of the *Rbfox1* locus for poly(A)+ and total libraries. Y axis shows RPM in each library.

Supplemental Fig S4
(Related Figs 2, 3)



**Supplemental Figure S4. (Related to Figures 2 and 3)**

Computational definition of introns and splicing.

A. Determination of FI value using read numbers for the exon-intron junction (EI), intron-exon junction (IE), and exon-exon junction (EE).

B. Introns were categorized as one of four types based on their Ensembl v91 annotation. Introns that are not partly overlapped with either exons or other introns are classified as U type introns. Introns that partly overlap with exons but not with other introns are classified E type introns. Introns that overlap with other annotated introns but not exons are called I type introns. EI type introns overlap with both exons and introns of other annotated isoforms.

C. Determination of cotranscriptional and posttranscriptional splicing. FI values were determined for all U introns from total and poly(A)+ chromatin associated RNA. Genes with overall expression above the median (2.13 TPM) were analyzed. Genes showing a bias for reads in the 3' end in the poly(A)+ RNA, and introns exhibiting FI values in total RNA below 0.1 were removed. A posttranscriptional splicing event was then defined as an intron having an FI value in poly(A)+ RNA greater than or equal to 0.1 (Post-tx). Cotranscriptional splicing of an intron generates an FI of less than 0.1 in the poly(A)+ RNA (Co-tx).

D. Illustration of post and cotranscriptional splicing. Introns with high read numbers on chromatin in both the total and poly(A)+ libraries were defined as posttranscriptionally spliced. Cotranscriptional splicing events exhibited reads in the total but not the poly(A)+ RNA.

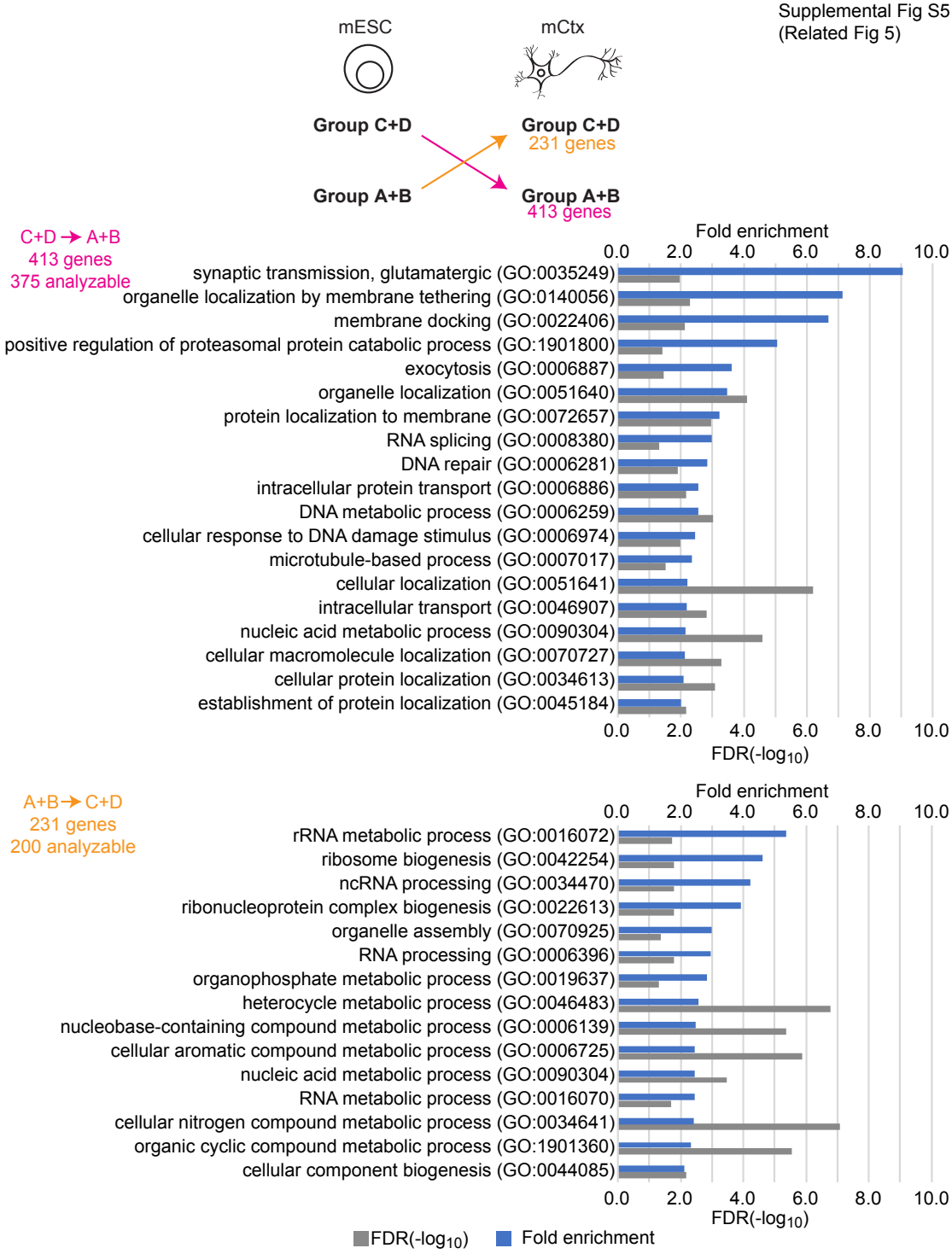
E. Diagrams of constitutive U introns and I introns adjacent to simple cassette exons that were assessed for co- and posttranscription splicing as presented in Figure 2C.

F. The proportions of co- and posttranscriptional splicing for all U introns and for first, middle and last introns in a transcript.

G. Transcripts with unspliced introns are enriched in the chromatin fraction.

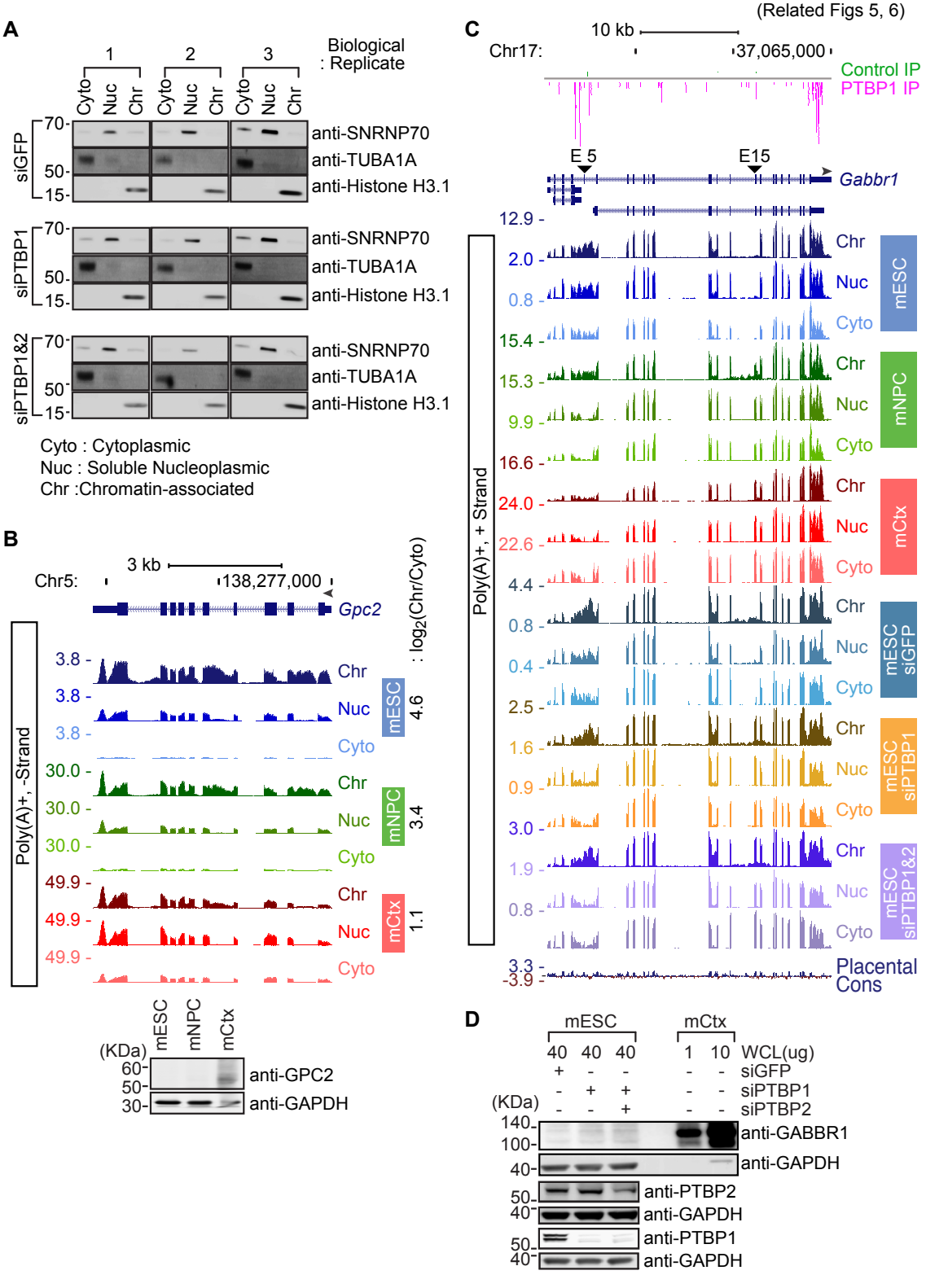
Genes having only U introns were selected from those whose overall expression was above the median (2.13 TPM). The gene group was then defined by the highest intron group within the gene (751 genes), where D > C > B > A. Introns marked 'na' indicate they were filtered by SIRI during X-means clustering.

H. Violin plots showing the distribution of chromatin partition indices ($Log_2$(Chr/Cyto)) of transcripts from different gene groups defined above. The number of genes in each gene group is indicated at the bottom.




**Supplemental Figure S5. (Related to Figure 5)**
GO analysis of genes containing introns that switch intron group during neuronal differentiation. Number of genes containing introns that changed group between mESC and mCtx is indicated at the top in yellow and pink. GO biological process enrichment these gene sets are listed at the bottom. Fold enrichment and FDR ($-\log_{10}$) shown in blue and grey bars, respectively.

Supplemental Fig S6
(Related Figs 5, 6)



**Supplemental Figure S6. (Related to Figures 5 and 6)**
Validation of subcellular fractionation after *Ptbp* knockdown in mESC and genome browser tracks of *Gabbr1*.
A. Confirmation of subcellular fractionation. Immunoblot analysis of diagnostic proteins in sub cellular fractions. SNRNP70 for soluble nucleoplasm (Nuc), TUBA1A and GAPDH for cytoplasm (Cyto), and Histone H3.1 for chromatin pellet (Chr). Gel images include 3 biological replicates of mouse embryonic stem cells (line E14).
B. (Upper Panel) Genome browser tracks of the *Gpc2* locus in mESC, mNPC, and mCtx. Transcripts are enriched in the chromatin fraction and exhibit unspliced introns in poly(A)+ RNA. The partition index of *Gpc2* in each cell type is indicated on the right. (Lower Panel) Immunoblot measuring expression of GPC2 protein relative to GAPDH control in mESC, mNPC and cortical neurons (mCtx). Gel image is one of 3 biological replicates.
C. Complete genome browser tracks of the *Gabbr1* locus in mESC, mNPC, and mCtx, and for *Ptbp1* knockdown and *Ptbp1/2* double knockdown in mESC. PTBP1 iCLIP tags in mESC are shown at the top (Linares et al. 2015). Intron 4-5 region is shown with a bracket, and exons 5 and 15 are marked with arrowheads.
D. Immunoblot measuring expression of GABBR1 protein relative to GAPDH in *Ptbp1 and Ptbp1/2* double knockdown samples in mESC and in mCtx as positive control. 40ug of whole cell lysate (WCL) were loaded on the gel for mESC, and 1 and 10 ug of WCL for mCtx. Gel image is one of 3 biological replicates.

## List of Supplemental Tables

Supplemental Table S1.
A. Primers used in this study, and PCR conditions.
B. Antibodies used in this study.

Supplemental Table S2.
A. Quantity and quality of RNA extracted from subcellular fractions.
B. Summary of RNA sequencing read alignment to genome_poly(A)+ library.
C. Summary of RNA sequencing read alignment to genome_total library.
D. Summary of RNA sequencing read alignment to genome_*Ptbp1* knockdown experiments_poly(A)+ library.
E. Summary of RNA sequencing read alignment to genome_Small RNA.

Supplemental Table S3.
Comparison of peak RPM and RNA quantity from RNA-Seq and qRT-PCR between subcellular fractions.

Supplemental Table S4.
Neuronal expression of MiR-770.

Supplemental Table S5.
Number of introns exhibiting Co- or Posttranscriptional splicing.

Supplemental Table S6.
A. Table of U introns with annotation classifications and FI values in poly(A)+ library.
B. Table of U introns with annotation classifications and FI values in total library.
C. Table of all introns with annotation classifications and FI values in poly(A)+ library.
D. Table of all introns with annotation classifications and FI values in total library.

Supplemental Table S7.
A. Introns containing premature termination codons (PTC).
B. Proportions of C and D introns located in UTRs.

Supplemental Table S8.

A. List of U introns reported as detained introns (DIs) in Boutz et al. (Boutz et al. 2015).

B. Proportion of introns assigned to groups.

C. List of genes containing PTBP1 dependent retained introns in chromatin fraction (871 genes). 87 genes out of 871 were reported upregulated by 10% after *Upf1* knockdown in data from Hurt et al. (Hurt et al. 2013).

D. Number of genes with different partition indices in mESC that were seen to be upregulated by 10% after *Upf1* knockdown. Data from Hurt et al. (Hurt et al. 2013).

E. Number of U introns and proportion of intron groups in mCtx that overlap with intron groups from Mauger et al. (Mauger et al. 2016).

Supplemental Table S9.

A. List of sequence features used for deep leaning model (1,387).

B. Predictive value of genomic features for distinguishing introns of different groups.

## Supplemental Methods

### Cell lines and tissue culture

The mouse embryonic stem cells (mESC) line, E14 (Hooper et al. 1987) was cultured on 0.1 % gelatin-coated dishes with mitotically inactivated mouse embryonic fibroblasts (MEF) (CF1, Applied StemCell, Inc.) in mESC media at least 2 passages from the initial thawing. Then, the mESCs were transferred onto 100 mm cell culture plates that contained only 0.1 % gelatin in preparation for RNAi and cell fractionation experiments. mESC media consisted of DMEM (Fisher Scientific) supplemented with 15 % ESC-qualified fetal bovine serum (Thermo Fisher Scientific), 1x non-essential amino acids (Thermo Fisher Scientific), 1x GlutaMAX (Thermo Fisher Scientific), 1x ESC-qualified nucleosides (EMD Millipore), 0.1 mM β-Mercaptoethanol (Sigma-Aldrich), and $10^3$ units/ml ESGRO leukemia inhibitor factor (LIF) (EMD Millipore). Mouse primary cortical neuron cultures were prepared from gestational day 15 C57BL/6 embryos (E15) (Charles River Laboratories), as described previously (Zheng et al. 2010). Briefly, cortices were dissected out into ice cold HBSS and dissociated after a 10 min digestion in Trypsin (Thermo Fisher Scientific), then plated with plating media consisted of 70 % Neurobasal (v/v), 20 % horse serum (v/v), 25 mM sucrose, and 0.25x GlutaMAX plated at 5 million cells per 78.5 $cm^2$ on 0.1 mg/ml poly-L-lysine coated dishes. The half media was replaced with feeding media consisted of 98 % Neurobasal (v/v), 1x B27 (with vitamin A, Thermo Fisher Scientific), and 0.25x GlutaMAX at day 1 and 3. The neuronal culture maintained for 5 days *in vitro* (DIV5). Neurons represented 70 – 80 % of the cells in the culture. A mouse neuronal progenitor cell line (mNPC) was established from cortical cells of gestational day 15 embryos generated by crossing homozygous Nestin-GFP transgenic mice (Mignone et al. 2004) to wild type C57BL/6. GFP positive cells were collected by FACS and plated on uncoated culture dishes. These NPCs were

grown in DMEM/F12 supplemented with B27 (without vitamin A, Thermo Fisher Scientific), 1x GlutaMAX and antibiotics. EGF and FGF (PeproTech) were added every day at 10 ng/ml concentration. All experiments were approved by the UCLA Institutional Animal Care and Use Committee (ARC# 1998-155-53).

**Knockdown of *Ptbp* in mESC.**

siRNAs that target EGFP (Silencer Select AM4626), PTBP1 (Silencer Select s72337), and PTBP2 (Silencer Select s80149) were transfected into mESCs twice using Lipofectamine RNAiMAX (Invitrogen). The first transfection was performed while the cells were in suspension, and the second transfection was performed 24 hours after when the cells had already attached to the surface of the cell culture dish. The cells were harvested and fractionated 24 hours after the second transfection.

**Subcellular fractionation, RNA isolation, and library construction in detail**

mESCs, mNPC, and cortical neurons (mCtx) were fractionated into cytoplasmic, soluble nuclear, and chromatin pellet as described previously (Wuarin and Schibler 1994; Pandya-Jones and Black 2009; Yeom and Damianov 2017; Yeom et al. 2018). Briefly, cells were washed twice with washing buffer (1X PBS/1 mM EDTA [pH 8.0]) and gently trypsinized to collect the cell pellet by centrifugation. 1 X $10^7$ cells from the pellet were retrieved in a 2.0 mL low adhesion microcentrifuge tube (USA Scientific) and washed twice with washing buffer to remove cell debris from prematurely-lysed cells. For neurons, the initial plating number on the day of dissection and culture was used for quantification, and cells were collected by scraping from the plate without trypsinization.

The cells were then incubated first in ice-cold hypotonic buffer (10 mM Tris–HCl [pH 7.5], 15 mM KCl, 1 mM EDTA [pH 8.0], 0.15 mM Spermine, 0.5 mM Spermidine, 0.5 mM DTT, and 1× Protease inhibitor, and 15 mM KCl (50 mM NaCl for mNPC and mCtx))

for 5 min and lysed in ice-cold lysis buffer (10 mM Tris–HCl [pH 7.5], 15 mM KCl, 1 mM EDTA [pH 8.0], 0.15 mM Spermine, 0.5 mM Spermidine, 0.5 mM DTT, and 1× Protease inhibitor, and 0.0375 %, 0.15 %, and 0.1 % of Igepal CA-630 for mESC, mNPC, and mCtx, respectively) for 1 min. The resulting lysate was immediately layered on top of a chilled 24 % sucrose solution (hypotonic buffer in 24 % (w/v) sucrose without detergent) and centrifuged for 10 min, 4 °C, 6000 x g. 10 % of the supernatant (cytoplasmic fraction) was used for immunoblot to check for contamination from nuclear materials, and the rest was mixed with TRIzol LS (Invitrogen) to extract cytoplasmic RNA. The nuclear pellet was washed once with washing buffer before being resuspended in 100 µl of chilled glycerol buffer (20mM Tris-HCl [pH 7.9], 75 mM NaCl, 0.5 mM EDTA [pH 8.0], 50 % glycerol (v/v), 0.85 mM DTT, and 0.125 mM PMSF). 100 µl of cold nuclear lysis buffer (20 mM HEPES [pH 7.6], 7.5 mM MgCl2, 0.2 mM EDTA [pH 8.0], 300 mM NaCl, 1M urea, 1 % Igepal CA-630, 0.3 mM Spermine, 1.0 mM Spermidine, 1× Protease inhibitor) was then added and the mixture incubated on ice for 2 min. After the 2 min incubation period, 1/8 of the nuclear lysate was transferred to a separate 1.5 mL microcentrifuge tube for each sample. After centrifugation for 2 min, 4 °C, 6000 x g, the supernatant (soluble nuclear fraction) from both tubes were pooled. 10 % of the pooled supernatant was used to check its purity by western blot and the rest was mixed with TRIzol LS to extract nucleoplasmic RNA. The resulting two insoluble nuclear pellets (one large, one small) were washed once with washing buffer before incubation of the large pellet in TRIzol at 50 °C until the pellet was completely solubilized to extract chromatin-associated RNA (from 7/8 nuclear lysate). After washing the small pellet (from 1/8 nuclear lysate) was resuspended in 5 % SDS sample buffer (5 % SDS, 30 % glycerol (v/v) and 150 mM Tris–HCl [pH 8.0]) at 55 °C for 15 min to extract proteins from the chromatin fraction. Each fractionation experiment was performed in triplicate.

RNA was treated DNase I (Takara) followed by phenol extraction. Before RNA quality check and the size selection steps, RNA was quantified on a NanoDrop spectrophotomer (Supplemental Table S2A), and RNA integrity was checked by either Bioanalyzer or RNA screen tape (Agilent) (Supplemental Table S2A). Note that the sum of the RNA from the chromatin, nucleus, and cytoplasmic fractions does not represent total cellular RNA due to extensive washing steps during fractionation.

5ug of RNA from each subcellular fraction was processed with the RNeasy MinElute Cleanup Kit (Qiagen). RNAs longer than 200 nt were eluted from the membrane to make long RNA libraries, while the flow through (< 200 nt) was also collected and ethanol precipitated to make short RNA libraries from each fraction. Half of the long RNA eluate was used to generate poly(A)+ libraries using the TruSeq Stranded mRNA Library Prep Kit (Illumina). The other half the long RNA eluate was used for total RNA library construction, using the ribosomal RNA removal kit (Ribo-Zero™ rRNA Removal Kits (Human/Mouse/Rat), MRZH116, Epicentre Biotechnologies) followed by library construction using TruSeq Stranded mRNA Library Prep Kit (Illumina) without the oligo-T bead purification step. Short RNA was used for small RNA library construction, using the ribosomal RNA removal kit (Ribo-Zero™ rRNA Removal Kits (Human/Mouse/Rat), MRZH116, Epicentre Biotechnologies) followed by library construction using the TruSeq Small RNA sample Prep Kit (Illumina). The small RNA library was further purified on a TBE gel by excising bands between 120 - 160 nt to generate small non-coding RNA reads that include mature-miRNAs (~ 22 nt) and piwi-interacting RNAs (piRNAs, ~ 30 nt).

For *Ptbp* knockdown experiments in mESC, poly(A)+ libraries were created using the TruSeq Stranded mRNA Library Prep Kit (Illumina).

**RNA sequencing and alignment**

The libraries of all fractions and cell types, were subjected to 100 nt paired-end sequencing at the UCLA Broad Stem Cell Center core facility on an Illumina HiSeq 4000 to generate about 21 million mapped reads per sample. The libraries of *Ptbp* KD samples in mESC were subjected to 75 nt paired-end sequencing at the UCLA Neuroscience Genomics Core on an Illumina HiSeq 4000 to generate about 25 million mapped reads per sample. Small RNA libraries were sequenced on an Illumina MiSeq machine using the MiSeq Reagent Kit v3 (Illumina). RNA-seq data were aligned using STAR (version 020210) (Dobin et al. 2013) on mm10.

Sequencing reads and mapping results are summarized in Supplemental Table S2B – S2D. Reads per million read values (RPM) were calculated and displayed in UCSC Genome Browser sessions including strand information. TPM values obtained from kallisto (version 0.43.0) were used in gene expression analyses (Bray et al. 2016). To examine the similarity of replicate samples, gene expression distances between samples were calculated using DESeq2 (Anders and Huber 2010), and plotted in heatmaps (Supplemental Fig S1) using ggplot2. For small RNA, sequencing reads and mapping results are summarized in Supplemental Table S2E. Small RNA expression was analyzed using the online tool SPAR (https://www.lisanwanglab.org/SPAR) (Kuksa et al. 2018).

**RT-PCR and RT-qPCR validation**

RNA was collected and extracted from subcellular fractions as described above. For the RT reaction, 0.8 - 1 µg of total RNA, 100 ng of random hexamers (or 250 ng of oligo(dT)), and 0.5 mM of dNTP mix (0.5 mM each) were incubated in a 7 µl reaction volume at 65 °C for 5 min. After 1min incubation on ice, 2 µl of 5x reaction buffer, 5 mM of DTT, and 100 units of SuperScript III RT (Thermo Fisher Scientific) were added. This

10 µl mixture was incubated at 25 °C for 10 min, 50 °C for 60 min, and 70 °C for 15 min. PCR was performed using Phusion DNA polymerase (Fisher Scientific). After the initial denaturing step at 98 °C for 1 min, amplification continued for 19 - 28 cycles of 98 °C for 1 min, 58 - 62 °C for 30 sec (annealing), and 72 °C for 20 - 25 sec (elongation). The reaction was completed with a final elongation at 72 °C for 10 min. Annealing temperature, elongation time, cycle numbers, and amplicon size are listed in Supplemental Table S1A. RT-PCR products were run on either agarose gels with EtBr staining or PAGE gels with SYBR Gold staining (Thermo Fisher Scientific), and visualized on a Typhoon imager (GE Healthcare) using the 492 nm excitation laser and 510 nm emission filters. For RT-qPCR the SensiFAST SYBR Lo-ROX Kit (Bioline) was used, and reactions contained 0.4ul of diluted cDNA (1:5 dilution in water) with 250nM each of forward and reverse primers in 6ul of RT-qPCR reactions. The mixtures were run on a QuantStudio 6 Real-Time PCR System (Thermo Fisher Scientific) with combined annealing and elongation cycles (Step1: 95 °C for 2 min, followed by 40 cycles of Step2: 95 °C for 5 sec and Step3: 55 or 60 °C for 30 sec. A list of primer sequences and PCR conditions is presented in Supplemental Table S1A.

**Co- and posttranscriptional splicing analysis**

For cotranscriptional splicing analysis, the U intron list was filtered for genes whose kallisto TPM in chromatin poly(A)+ RNA was over the median. Genes exhibiting 3' bias were identified by comparing the reads per nucleotide length of the 2$^{nd}$ exon to that of the 2$^{nd}$ to last exon for the longest transcript from the gene. If this ratio was less than 0.5 for all transcripts, the gene was excluded (Supplemental Fig. S4C). The same filters were applied to a list of simple cassette exons (SE) extracted from rMATS 4.0.1 (Shen et al. 2014). Cassette exons were selected that were located between I introns

when included, and become EI introns when skipped, as defined in Supplemental Fig. S4E.

To examine splicing at different intron locations, we compiled a list of 936 genes containing U introns as first, middle, and last introns. The middle intron was defined as intron number X. For genes with an odd number of total introns, X= (total intron number + 1) / 2. For genes with an even number of total introns, then X= total intron number / 2.

**Immunoblotting**

Protein samples for checking subcellular fractionation were prepared as described above. Whole-cell lysates (WCL) for checking GPC2, GABBR1, TUBB3, PTBP1 and PTBP2 were extracted in RIPA, quantified with the bicinchoninic acid (BCA) method (Smith et al. 1985), and loaded on 10 % SDS-PAGE gel. Images were taken on a Typhoon Imager (GE Healthcare) using fluorescent (Cy3 or Cy5) secondary antibodies. For GABBR1 detection HRP-linked secondary antibodies and a LAS-3000 Imaging System (Fuji) were used. All the experiments were repeated in three times. Antibodies used in this study were listed in Supplemental Table S1B.

**GO analysis**

Gene Ontology (GO) enrichments were determined using PANTHER (Protein ANalysis THrough Evolutionary Relationships) (Ashburner et al. 2000; Gene Ontology Consortium 2019). Molecular function terms were assessed for overrepresentation in genes containing introns that changed during neuronal differentiation from Groups C or D to Groups A or B, or from Groups A or B to Groups C or D compared to all mouse genes.

**Identification of NMD target genes**

A list NMD targets was extracted from a previous study in mESC (Hurt et al. 2013). We filtered for genes upregulated by greater than 10 % in two independent siRNA knockdowns of *Upf1* (siUpf1.1 and siUpf1.2). If a gene had multiple mRNA isoforms and one isoform was upregulated it was included as an NMD target on our list.

**Identifying Premature Termination Codons (PTC) in retained introns**

For each U intron, we retrieved all transcripts containing this intron from GTF. The transcripts without a clear open reading frame annotation or with termination codons after the intron were not considered. We also removed transcripts in which the intron was the last intron, because last intron retention should not cause Nonsense Mediated Decay (NMD). For each transcript that contained the intron, we inserted the intron sequence into the CDS of the transcript at junction of the intron. If the U intron was not the second last intron of the transcript, the intron was marked as PTC-containing if generated a stop codon inside the intron. If the intron was the second last intron of the transcript, the intron was defined as PTC-containing if an in frame stop codon within the intron was at least 50 nt upstream from the last exon-exon junction of the transcript. After checking all transcripts containing the intron, we defined the intron as a PTC-containing intron if it generated PTC in at least one transcript.

## References for Supplemental Materials


Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* **25**: 25–29.

Boutz PL, Bhutkar A, Sharp PA. 2015. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev* **29**: 63–80.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527.

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.

Gene Ontology Consortium T. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**: D330–D338.

Hooper M, Hardy K, Handyside A, Hunter S, Monk M. 1987. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**: 292–295.

Hurt JA, Robertson AD, Burge CB. 2013. Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Res* **23**: 1636–1650.

Kuksa PP, Amlie-Wolf A, Katanić Ž, Valladares O, Wang L-S, Leung YY. 2018. SPAR: small RNA-seq portal for analysis of sequencing experiments. *Nucleic Acids Res* **46**: W36–W42.

Linares AJ, Lin C-H, Damianov A, Adams KL, Novitch BG, Black DL. 2015. The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation ed. B.J. Blencowe. *eLife* **4**: e09268.

Mauger O, Lemoine F, Scheiffele P. 2016. Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity. *Neuron* **92**: 1266–1278.

Mignone JL, Kukekov V, Chiang A-S, Steindler D, Enikolopov G. 2004. Neural stem and progenitor cells in nestin-GFP transgenic mice. *J Comp Neurol* **469**: 311–324.

Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**: 1896–1908.

Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci* **111**: E5593–E5601.

Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC. 1985. Measurement of protein using bicinchoninic acid. *Anal Biochem* **150**: 76–85.

Wuarin J, Schibler U. 1994. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol Cell Biol* **14**: 7219–7225.

Yeom K-H, Damianov A. 2017. Methods for Extraction of RNA, Proteins, or Protein Complexes from Subcellular Compartments of Eukaryotic Cells. In *mRNA Processing: Methods and Protocols* (ed. Y. Shi), *Methods in Molecular Biology*, pp. 155–167, Springer, New York, NY https://doi.org/10.1007/978-1-4939-7204-3_12 (Accessed August 21, 2020).

Yeom K-H, Mitchell S, Linares AJ, Zheng S, Lin C-H, Wang X-J, Hoffmann A, Black DL. 2018. Polypyrimidine tract-binding protein blocks miRNA-124 biogenesis to enforce its neuronal-specific expression in the mouse. *Proc Natl Acad Sci* **115**: E11061–E11070.

Zheng S, Eacker SM, Hong SJ, Gronostajski RM, Dawson TM, Dawson VL. 2010. NMDA-induced neuronal survival is mediated through nuclear factor I-A in mice. *J Clin Invest* **120**: 2446–2456.