SUPPLEMENTAL INFORMATION

Supplemental Figures

Supplemental Tables

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure S1. Detection and prevalence of intron retention in mouse data

A, Percentages of total introns retained at PIR ≥ 10 in 52 human samples from different cell and tissue types. Adult brain, ESC, and muscle samples are highlighted in blue, green and red, respectively.

B, As in A, for 65 mouse samples.

C, Percentage of total mouse introns detected as retained at different PIR thresholds as increasing numbers of cell and tissue samples are randomly sampled. Numbers in parentheses are estimates for percentages of total introns retained at different PIR thresholds, as derived from extrapolation (see Methods). Circles, means from 1,000 iterations; lines, fitted function used for extrapolation.

D, Percentage of total mouse genes with retained introns at different PIR thresholds as increasing numbers of cell and tissue samples are randomly sampled. Numbers in parentheses are estimates for percentages of total genes with retained introns at different PIR thresholds, as derived from extrapolation. Circles and lines as in C.

Supplemental Figure S2. RT-PCR validation of IR events

RT-PCR validation assays for 25 IR events detected by RNA-Seq data analysis. Validation assays were performed for the same tissue and cell types as those profiled by RNA-Seq, but using samples from independent sources. Icons (left of the gel images) indicate intron-retained and spliced isoforms. Colored squares next to the gene name indicates whether an event was predicted to be differentially retained in the following cell/tissue types: red, muscle; green, embryonic stem cells (ESCs); blue, neural; black, general. PIR values estimated from RNA-Seq data and gel image quantification are indicated below each panel. *Gapdh* RT-PCR assays employed primers spanning three constitutively spliced introns and were used as a loading/recovery control. These are shown below each set of IR events that were assayed in parallel. Event coordinates are given above each gel in the format: chromosome, strand, upstream exon, downstream exon (1-based). Differential PIR between all sample pairs for all events as measured by RNA-Seq and RT-PCR and Pearson correlation coefficient are plotted on the right of the page.

Supplemental Figure S3. Properties associated with retained introns

A, Distributions of intron length and C/G content in human and mouse retained introns and, for comparison, constitutively spliced introns. Retained introns compared in the analysis have a PIR ≥ 10 in ≥ 10% of the profiled samples for each species; constitutive introns have PIR < 2 in all samples where PIR could be determined. P-values are from one-sided Mann-Whitney *U* tests.

B, Comparison of donor splice site strength (measured using maximum entropy) distributions of mouse retained introns and of constitutively spliced introns as in A. P-values are from one-sided Mann-Whitney *U* tests.

Supplemental Figure S4. Distinct types of retained introns and associated properties

A, Fractions of total mouse retained introns belonging to each evolutionary type at different PIR thresholds, of all introns that were assigned to a class and that were retained at the indicated PIR thresholds in ≥ 10% of the samples.

B, Cumulative distribution of median PIR levels for each retained intron type in mouse cells and tissues. Only introns where PIR could be determined in ≥ 10% of the samples are shown.

C, Intron length and C/G content for each retained intron type in human and mouse genomes. Retained introns compared have a PIR ≥ 10 in ≥ 10% of the samples where PIR could be determined; 'constitutive' introns are introns with PIR < 2 in all samples where PIR could be determined. Dashed lines indicated group medians.

D, Comparison of donor and acceptor splice site strength (measured using maximum entropy) distributions of human and mouse retained introns and of constitutively spliced introns. Black lines indicate group medians. Intron groups as in C.

E, Fraction of all mouse introns retained in each genic region that is retained with PIR ≥ 10 in ≥ 10% of the samples where PIR could be determined. UTR, untranslated region; CDS, coding region of gene; PTC, premature termination codon that can elicit nonsense mediated mRNA decay. The total number of retained introns in each region is indicated.

Supplemental Figure S5. Conservation of intron retention in eleven vertebrate species

A, As for Figure 3A but with all data points shown. Proportion of total orthologous introns retained (PIR ≥ 10) in an organ of one species that are also retained in the same organ of one of eleven vertebrate species being compared (refer to main text and Panel B). Lines connect mean values for each evolutionary distance. Asterisks indicate significance of differences between each organ and the average of all other organs (*, $p < 0.05$; **, $p < 0.001$, see Methods for details.)

B, Principal component analysis of introns with a PIR range of ≥ 10 in at least three species (n = 4,853). Top panel, % of variance captured by principal components; Bottom, projections of samples onto the first three principal components.

Supplemental Figure S6. Positional bias of IR detection in polyA+ RNA from total cell and subcellular fractionations

A, Fraction of introns that are retained (PIR ≥ 10), in ≥ 10% of the samples where PIR could be determined, in each of ten equal-sized positional bins from the 5' to the 3'-end of the respective transcripts.

B, Mean PIR in each of ten positional bins as in A, in polyA+ RNA from subcellular fractions of three human cell lines .

C, Cumulative distribution of PIR levels in the same samples as in B. +PTC, introns in the coding sequence and that introduce a PTC predicted to elicit NMD; -PTC, introns in UTRs, in CDS and not predicted to introduce a PTC predicted to elicit NMD, or in non-coding transcripts.

Supplemental Figure S7. Role of IR in gene regulation

A, Distributions of correlation between intron PIR and gene expression across cell and tissue types for introns of Types A, B and C. Only introns with a PIR value of 0 in at most one sample are included.

B, Expression difference between cytosolic and nuclear fractions of three human cell lines for transcripts that do or do not contain PTC-introducing introns, as measured at different PIR thresholds in the nuclear fraction. Shaded boxes indicate upper and lower quartiles of distributions of expression level differences, and colored lines indicate median values. Expression level differences for transcripts harboring retained introns that do or do not introduce PTCs are indicated by red and green, respectively. Asterisks indicate a significance difference between median values for expression level differences for transcripts with and without PTC-introducing retained introns (*, $p < 0.05$; **, $p < 0.001$; one-sided Mann-Whitney *U* tests after Bonferroni correction).

Supplemental Figure S8. RT-PCR validation of regulated IR events during neuronal differentiation

RT-PCR validation of RNA-seq detected events with increasing PIR and decreasing spliced mRNA expression during differentiation of glutamatergic neurons from mouse ES cells. Quantification of PIR and mRNA expression are shown beneath each panel. For relative expression, the spliced band was quantified and normalized to *Gapdh*, and Day -8 was set to 1.

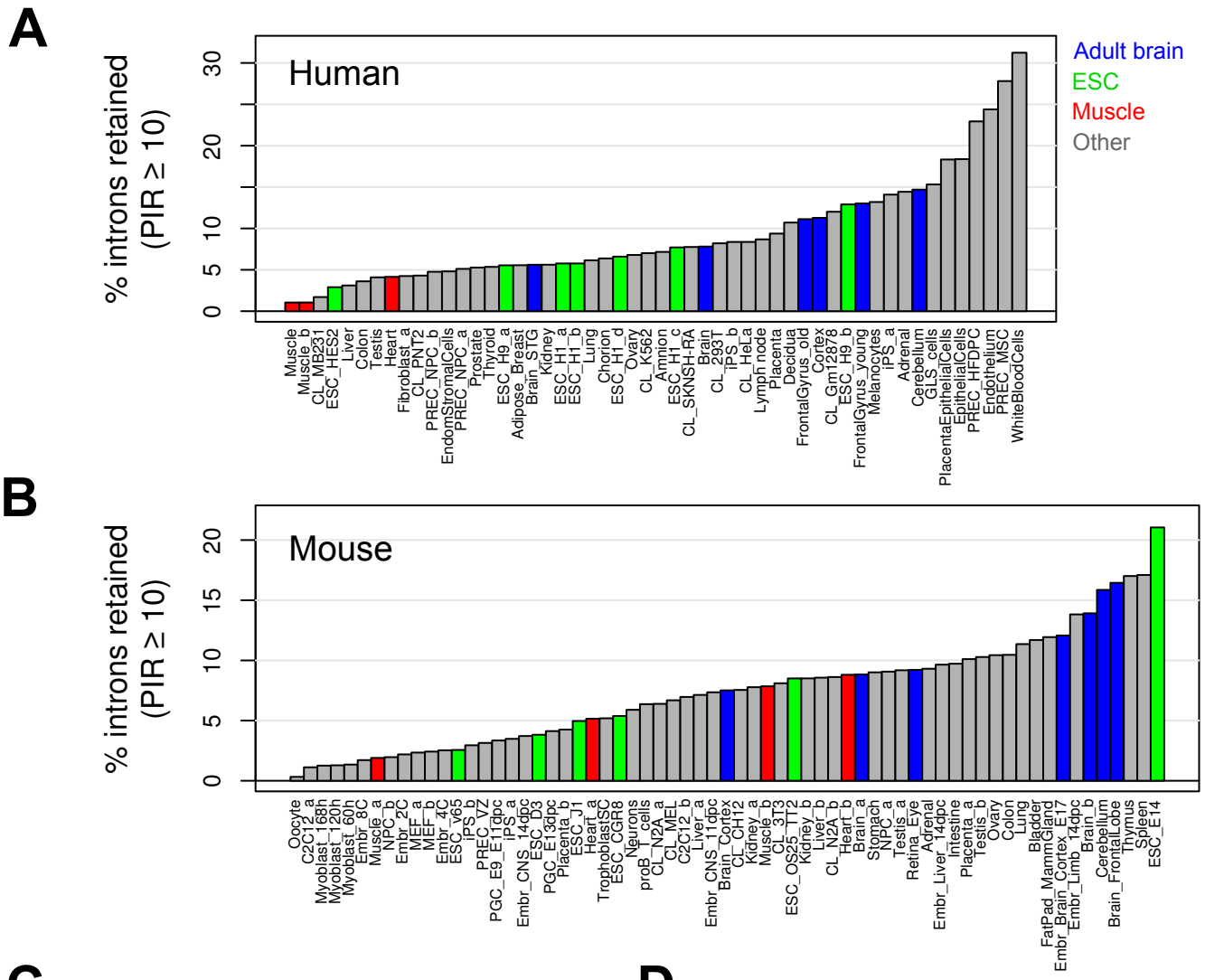
Supplemental Figure S9. Chromatin features across introns in K562 and CH12 cells

A, Average ChIP-Seq read counts across intron groups in human K562 and mouse CH12 cells, for ten chromatin features. Retained (PIR ≥ 10) and constitutive (PIR < 2) introns analyzed are in genes with matched expression levels (cRPKM, see Methods) as indicated by brackets in the legend. Gray bars, flanking exons; y-axes represent input-subtracted average ChIP fragment density per million reads (FPM-input) at each aligned bp. Grey dashed lines mark the 2- and 98-percentile of all values in the plot region before averaging to give an indication of the dynamic range. Left column for each cell line shows retained (orange) and constitutive (purple) introns. Right column shows retained (bright colors) and non-retained (dark colors) introns that are within 2 kb of the transcription start site (TSS, red); in the interior of the transcript (green); or within 2 kb of the transcription termination site (TTS, blue).

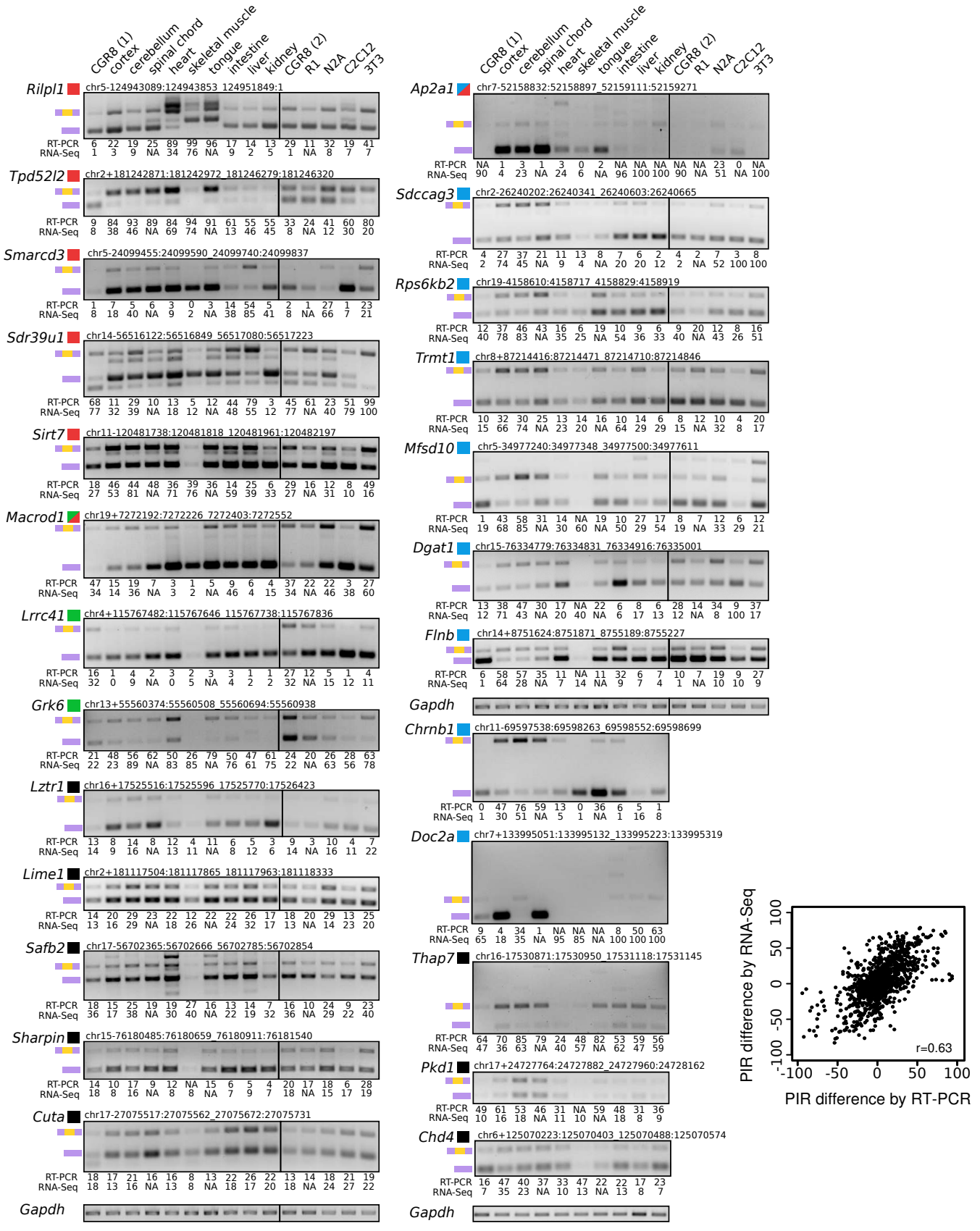
B, Average ChIP-Seq signal of Ser2-phosphorylated RNA Pol II over introns with different PIR thresholds in CH12 cells. Labels and plots as in (A).

Supplemental Figure S10. DRB-induced IR changes in CGR8 mouse ES cells

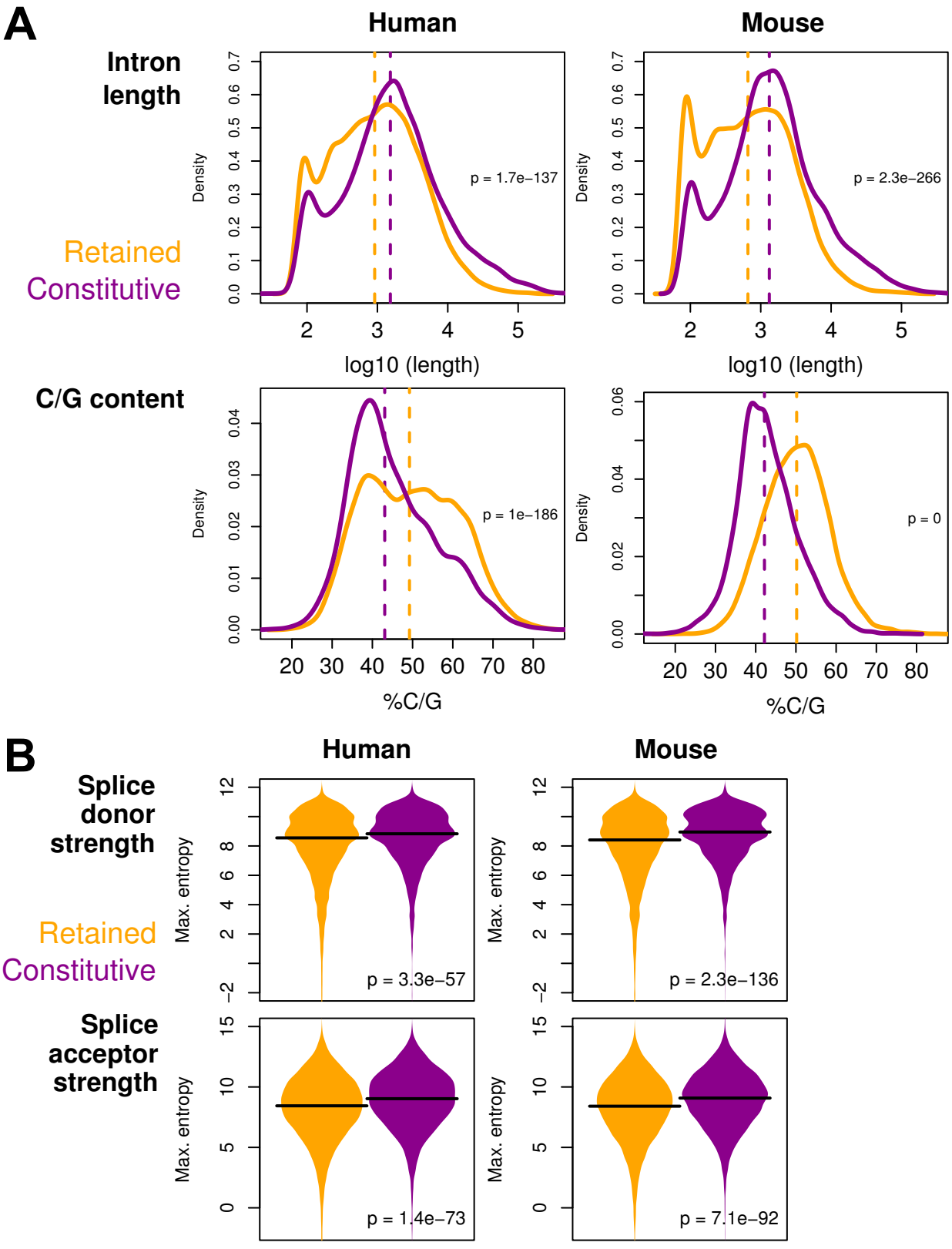
RT-PCR assays of IR events in CGR8 mouse ES cells with and without treatment with the RNA Pol II elongation inhibitor DRB. Retained introns with low PIR in ES cells relative to other cell types were analyzed. Lanes are as indicated for *Gapdh*. Quantitations are shown beneath each panel. The second row shows those events with a difference between control 25 μg/ml of ≥ 10 PIR. Note that loading was higher for treatment with 25 μg/ml of DRB to compensate for reduced transcript levels.



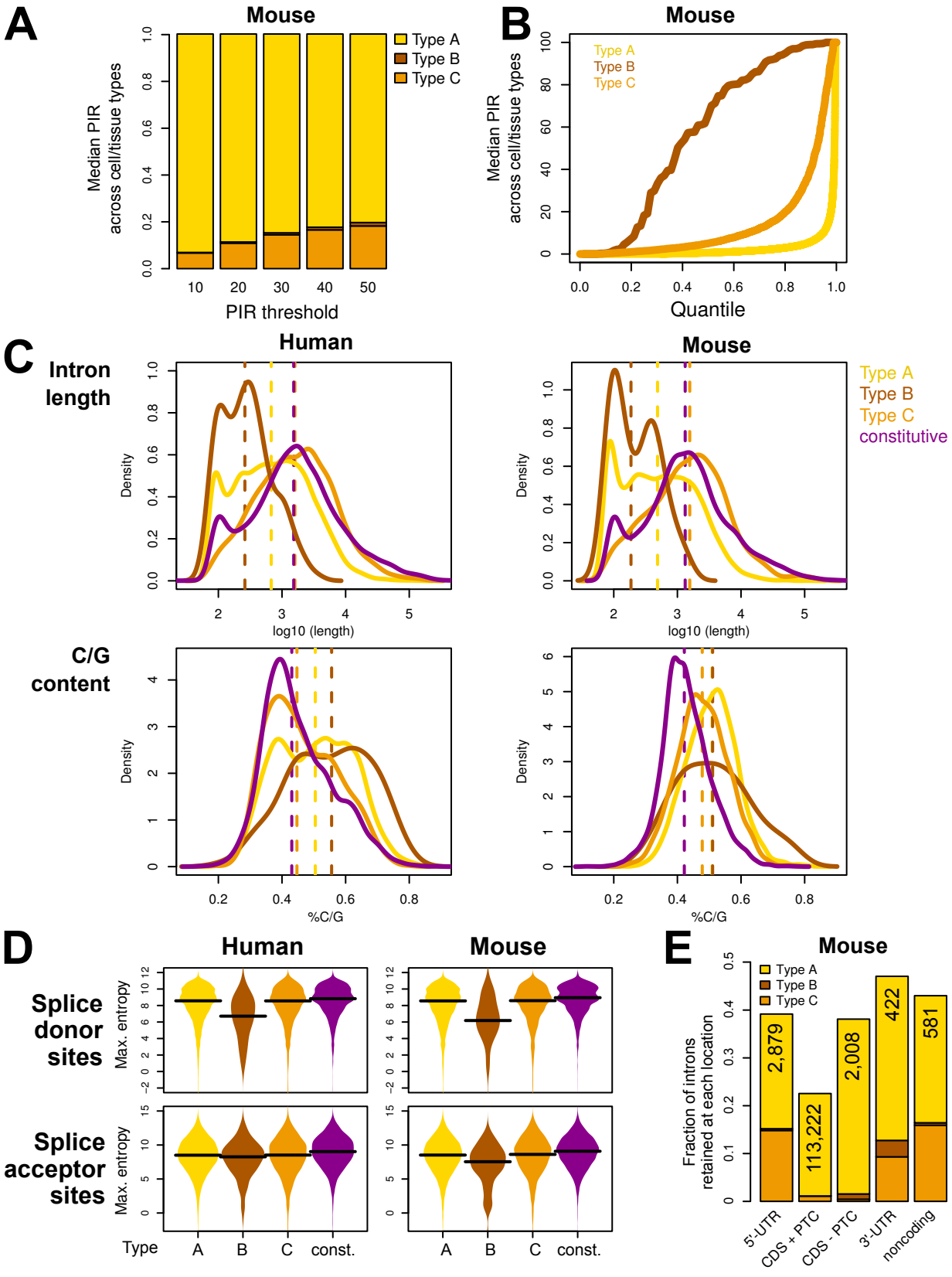
Braunschweig et al., Supplemental Figure S1



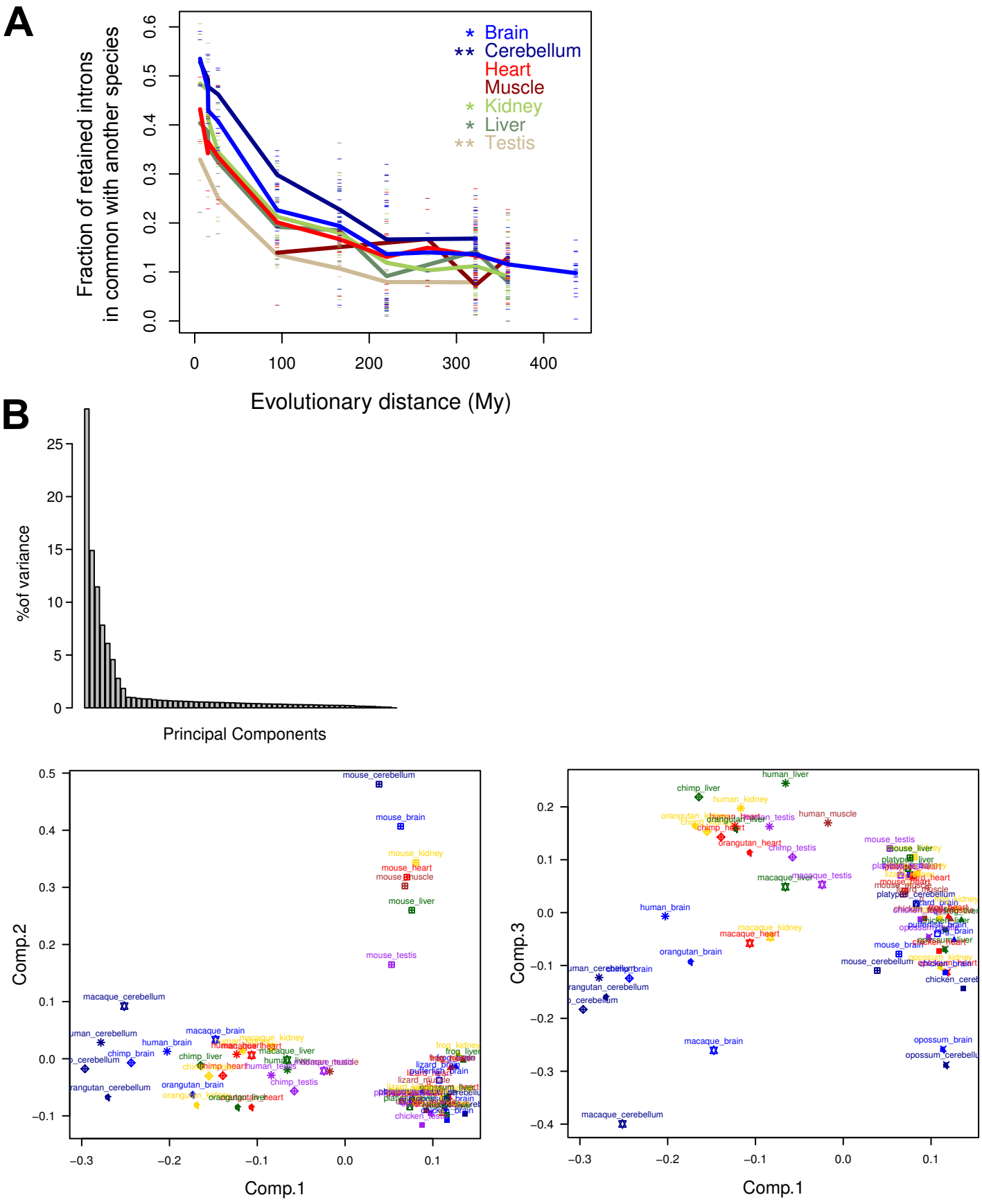
Braunschweig et al., Supplemental Figure S2



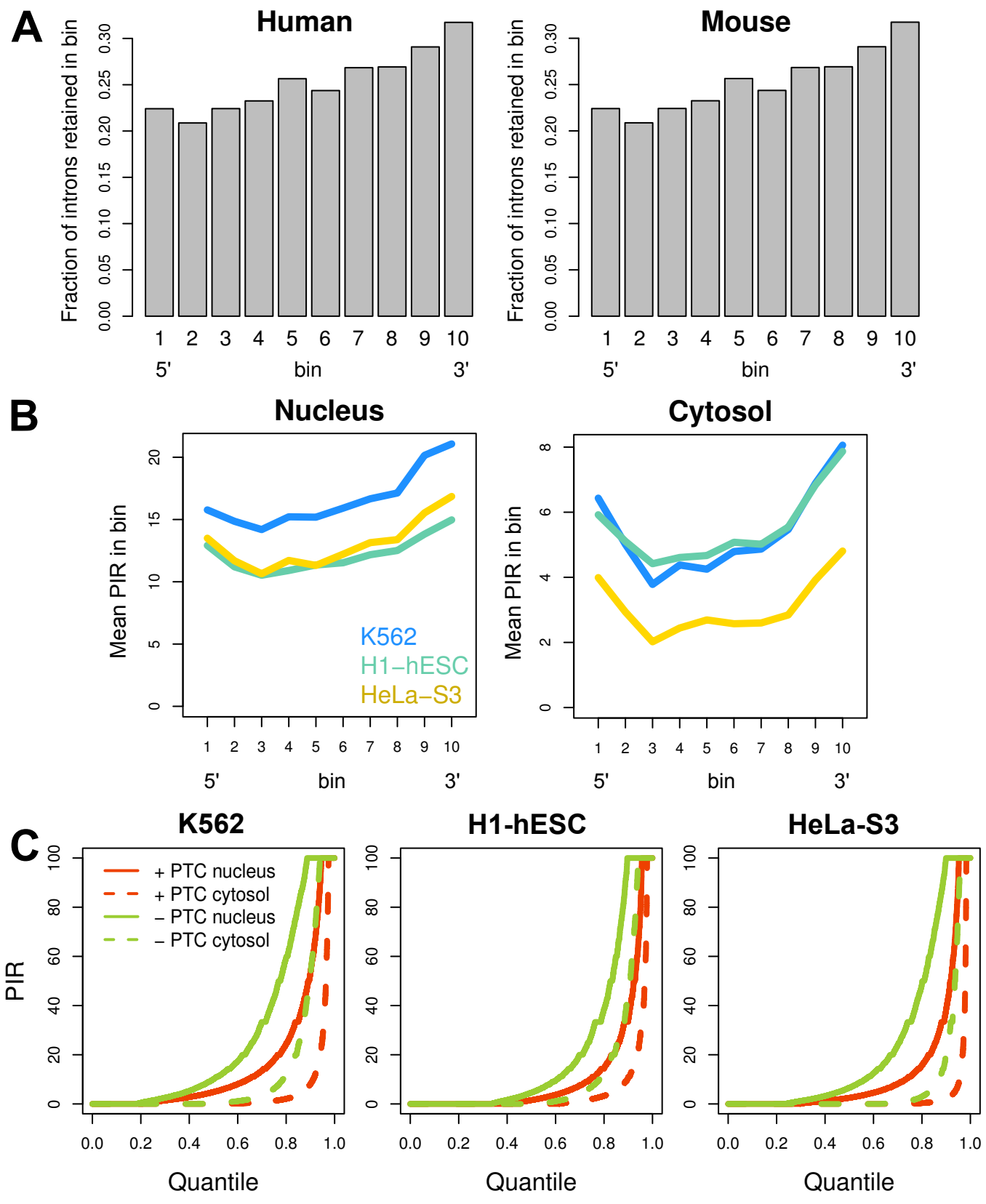
Braunschweig et al., Supplemental Figure S3



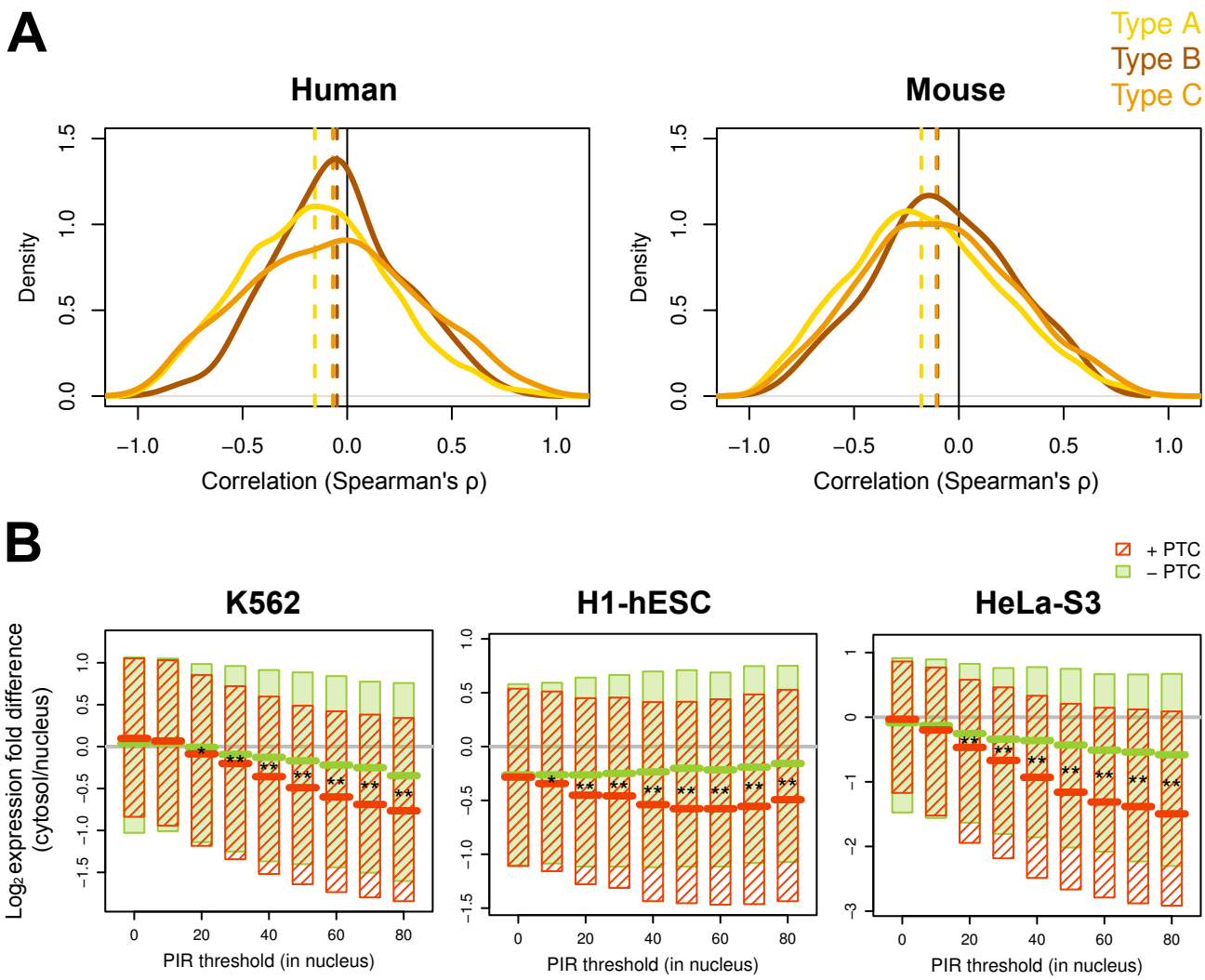
Braunschweig et al., Supplemental Figure S4



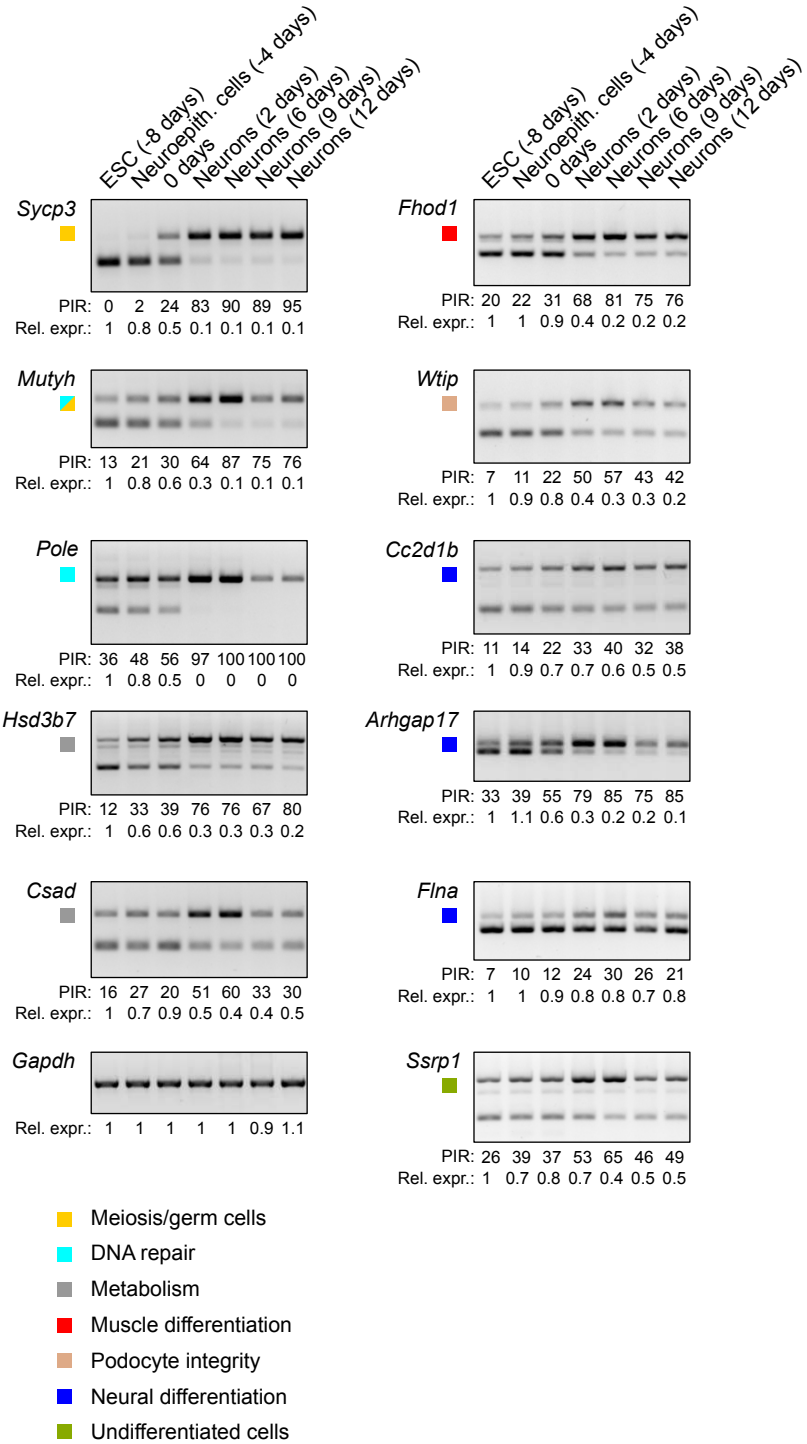
Braunschweig et al., Supplemental Figure S5



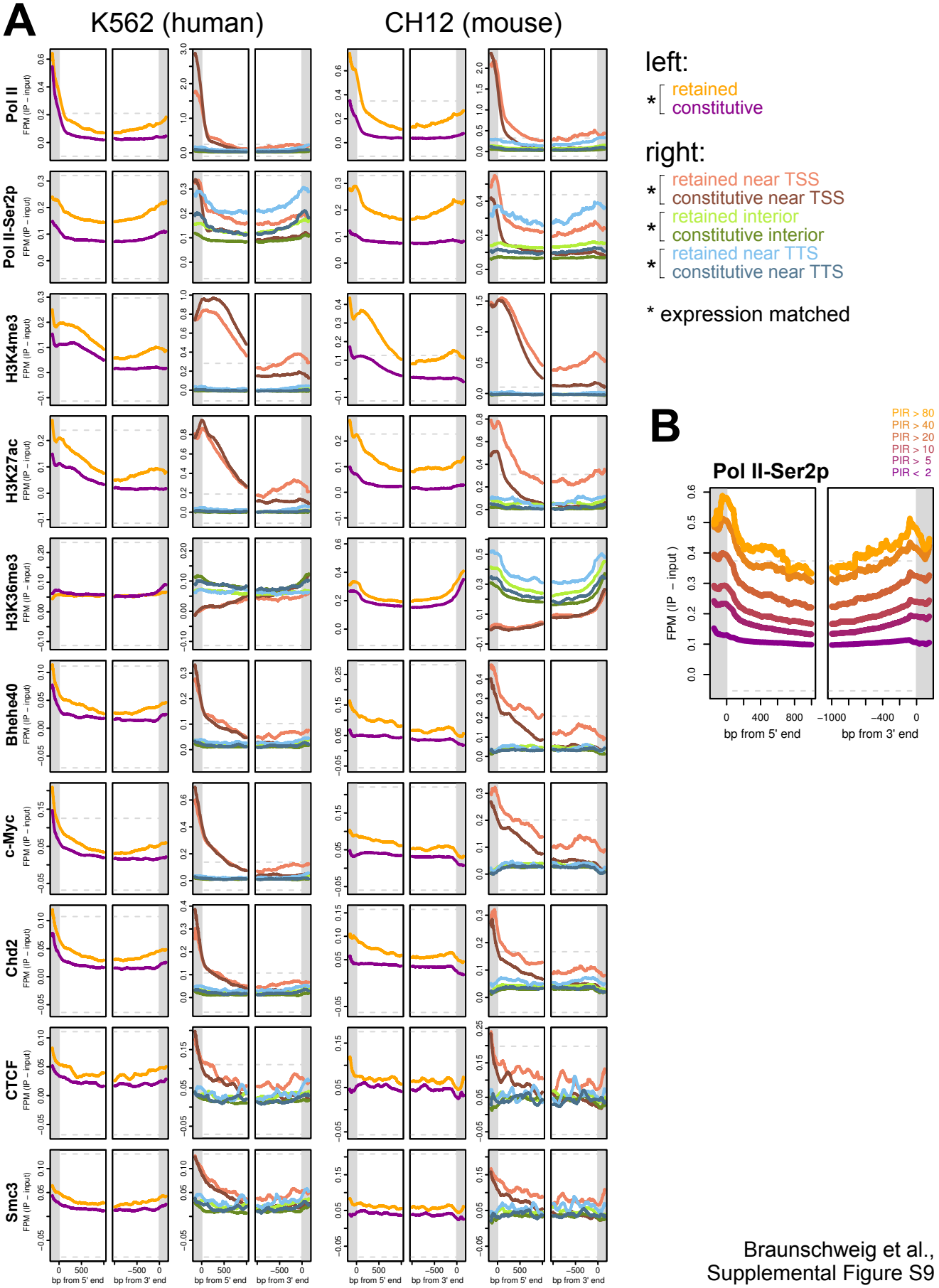
Braunschweig et al., Supplemental Figure S6



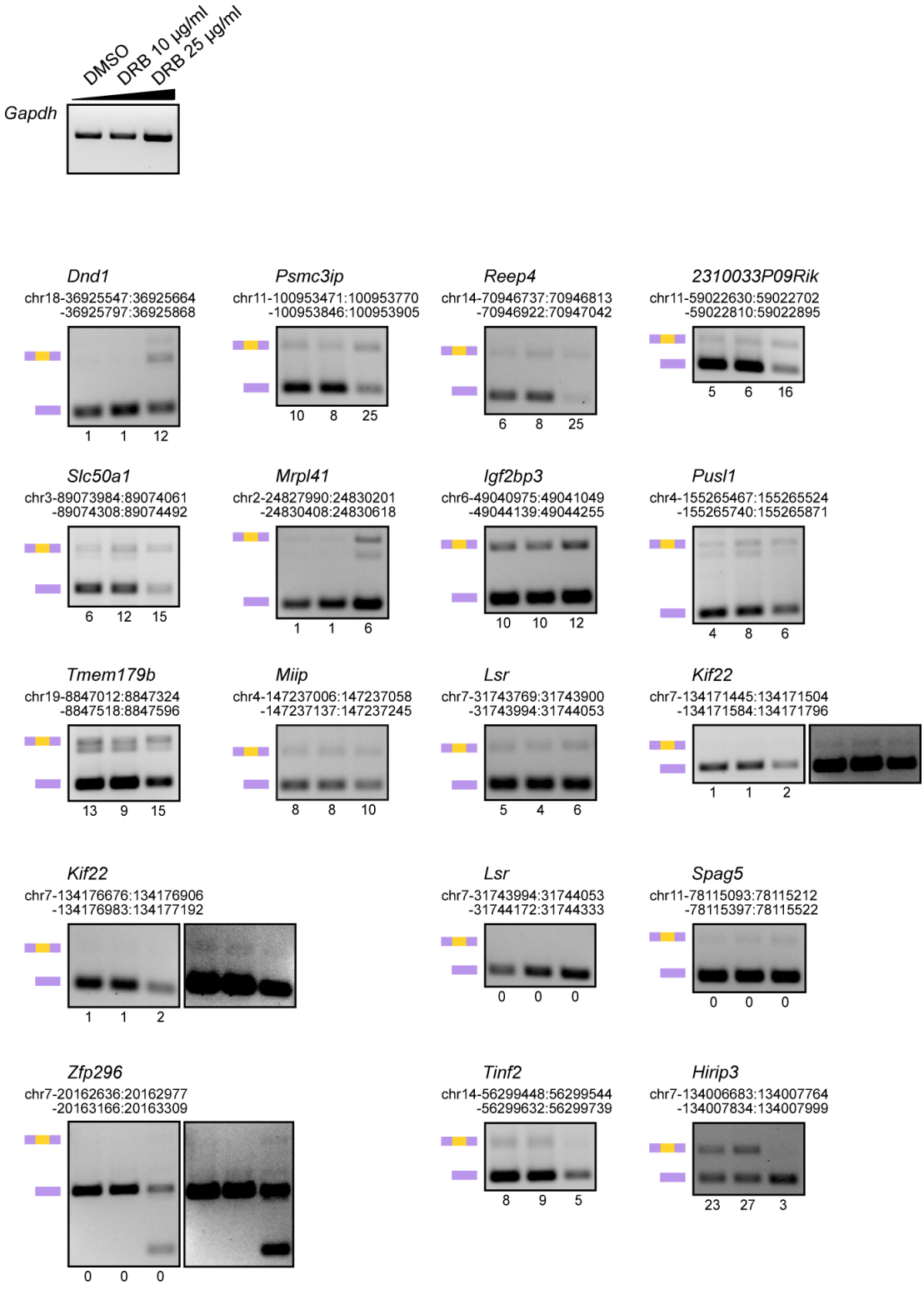
Braunschweig et al., Supplemental Figure S7



Braunschweig et al., Supplemental Figure S8



Braunschweig et al.,
Supplemental Figure S9



Braunschweig et al., Supplemental Figure S10