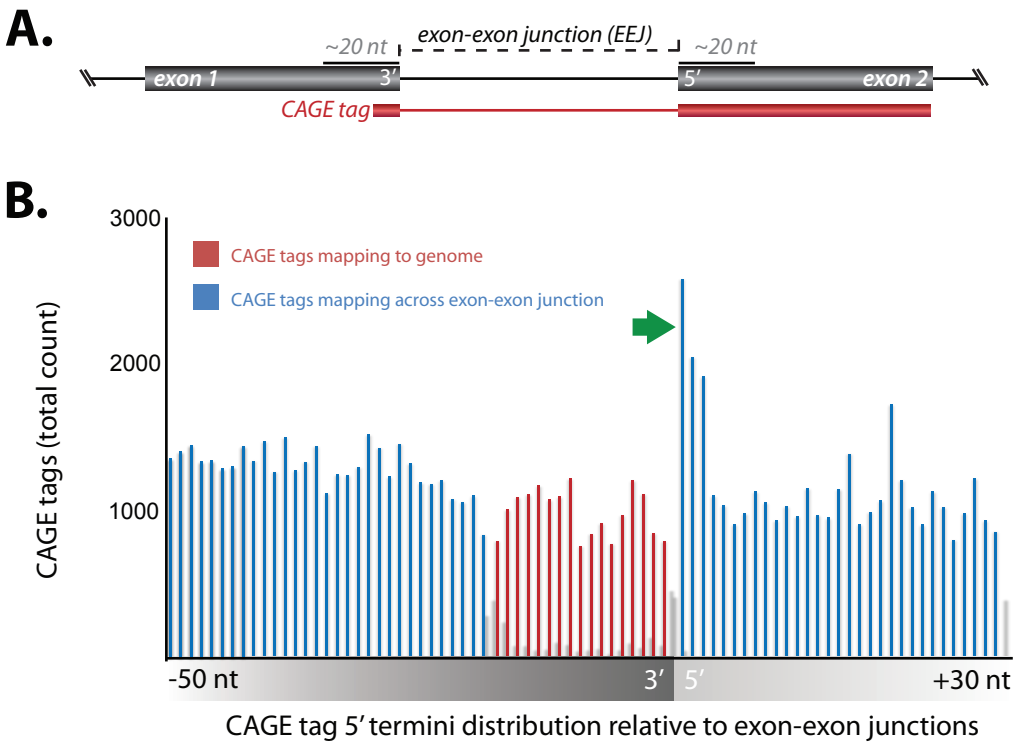
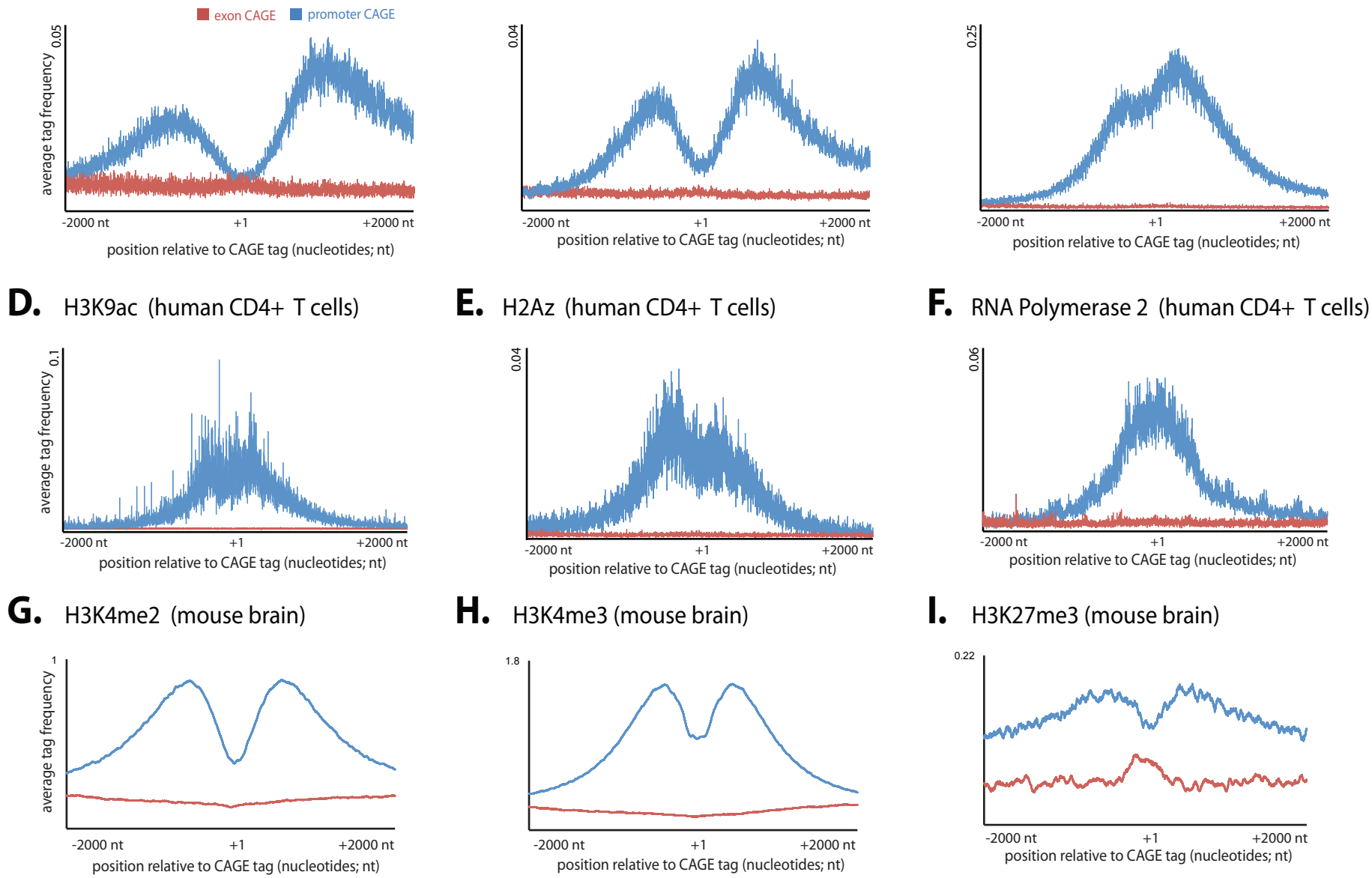
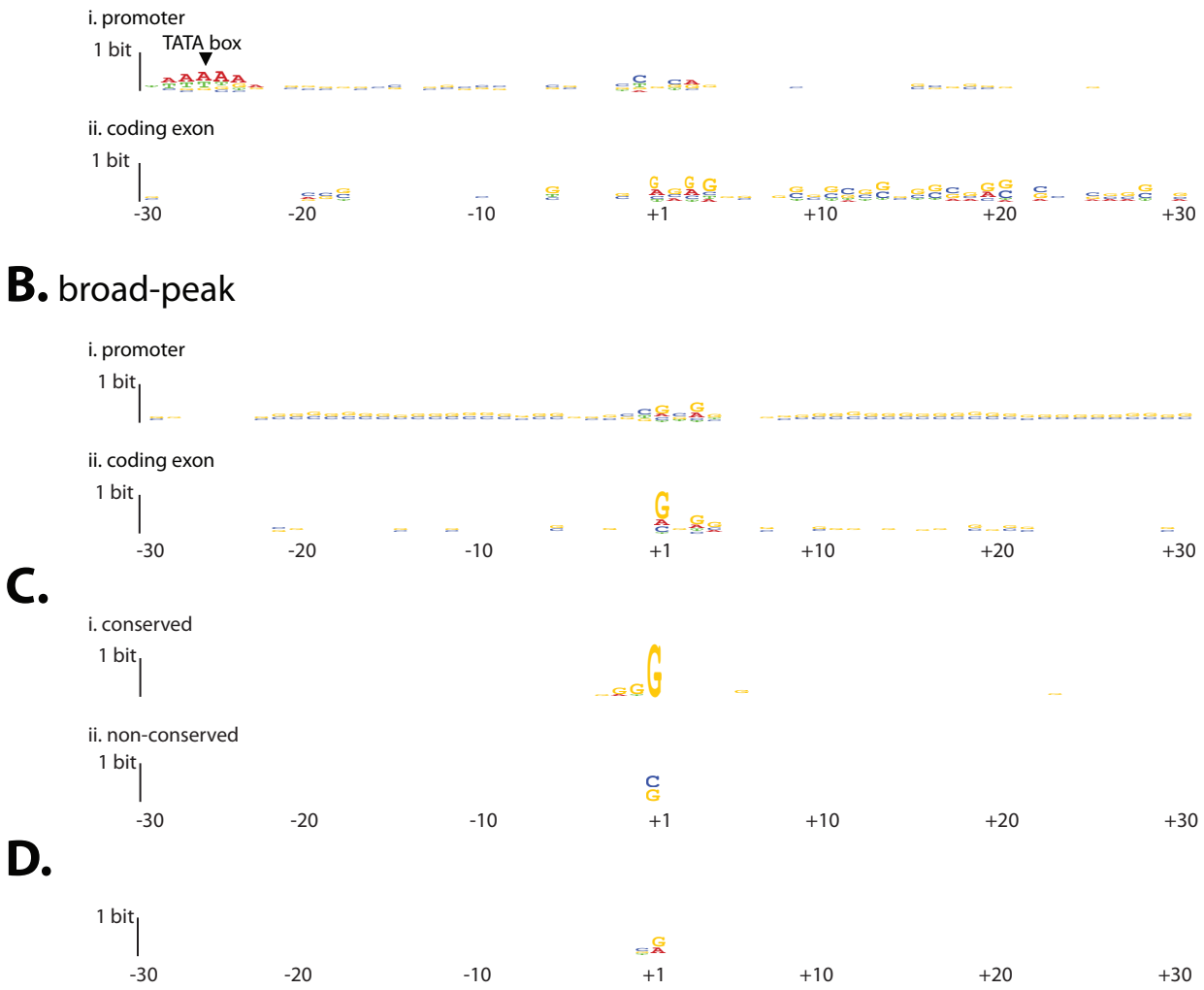


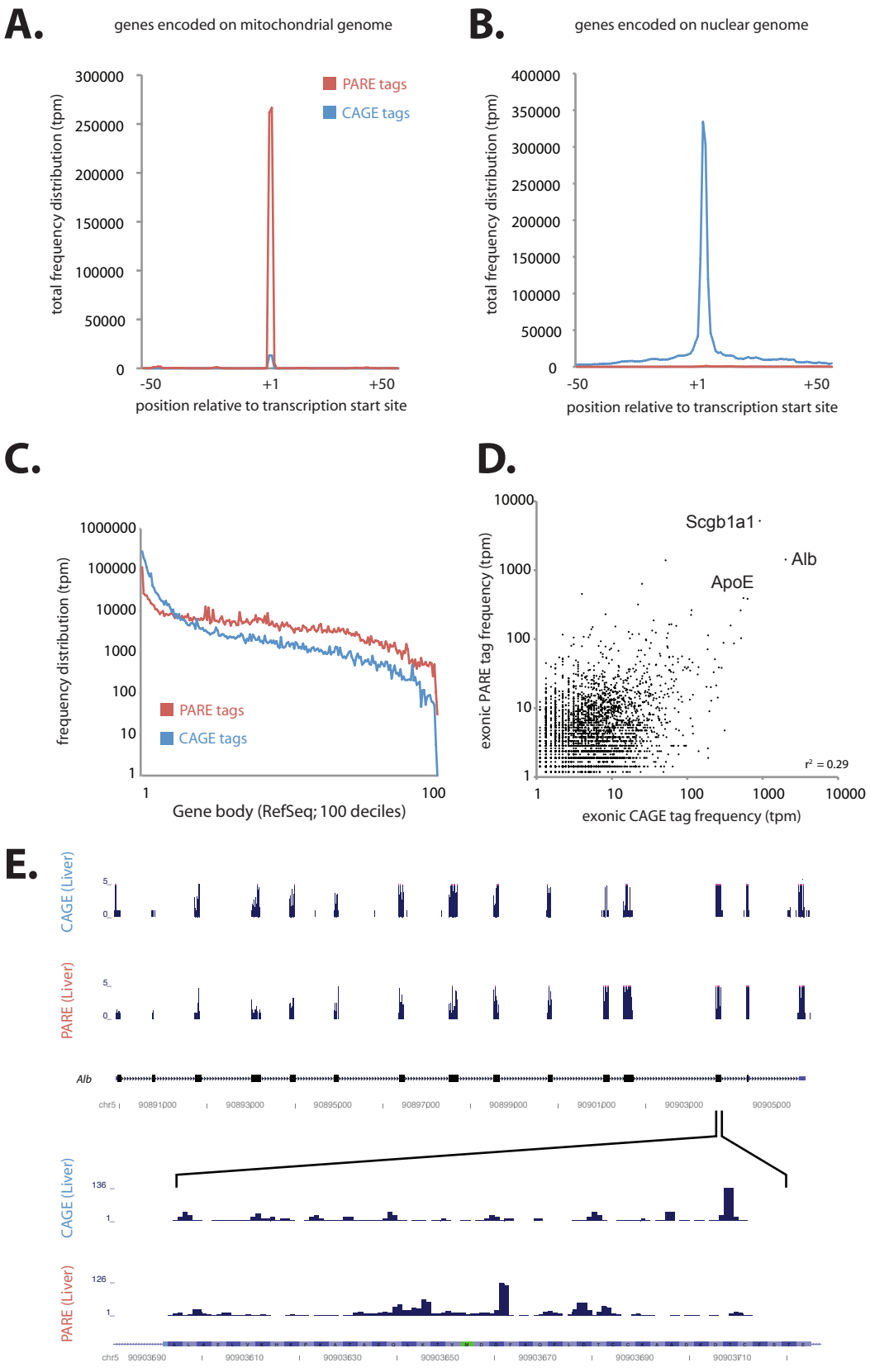
Supplemental Fig. S1. Evidence of post-transcriptional cleavage from CAGE tag analysis. (A) Schematic diagram illustrating the mapping strategy employed to discern CAGE tags that span exon-exon junctions (EEJs). **(B)** Distribution of mouse CAGE tags mapping proximal to RefSeq EEJs. The frequency of CAGE tags that map uniquely and exactly to the genome but do not span EEJs (blue) are under-represented adjacent to exon boundaries. This under-representation can be reconciled with CAGE tags that map uniquely and exactly across EEJs (red). The distribution also shows a distinct peak of enrichment of CAGE tags immediately downstream of EEJs (green arrow).



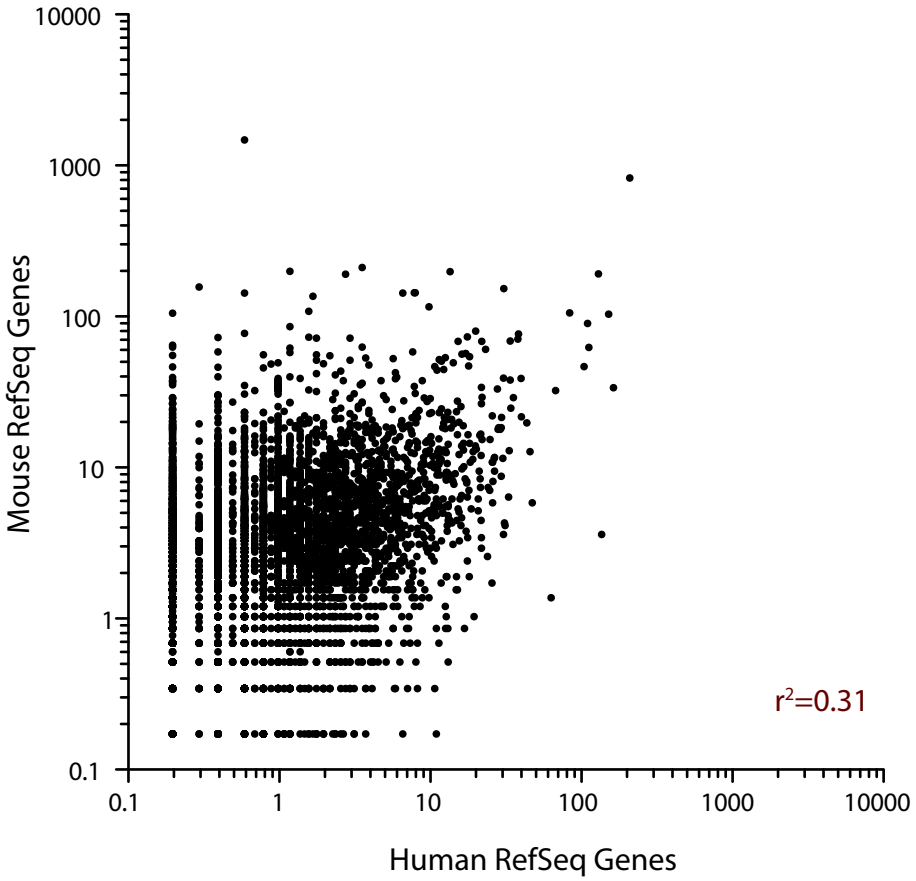
Supplemental Fig. S2. Frequency distribution for modified chromatin immunoprecipitated tags relative to CAGE sites associated with gene promoters and exons. (A-I) The average frequency of DNA tags immunoprecipitated within a 4 kb window centered on CAGE tags mapping to promoters (blue) and coding exons (red) with antibodies to H3K4me3 **(A)**, H3K2me2 **(B)**, H3K4me3 **(C)**, H3K9ac **(D)**, H2Az **(E)** and RNA polymerase II **(F)** in human CD4+ T cells (Wang et al., 2008) and H3K4me2 **(G)**, H3K4me3 **(H)** and H4K27me3 **(I)** in adult mouse brain (Mikkelsen et al., 2007).



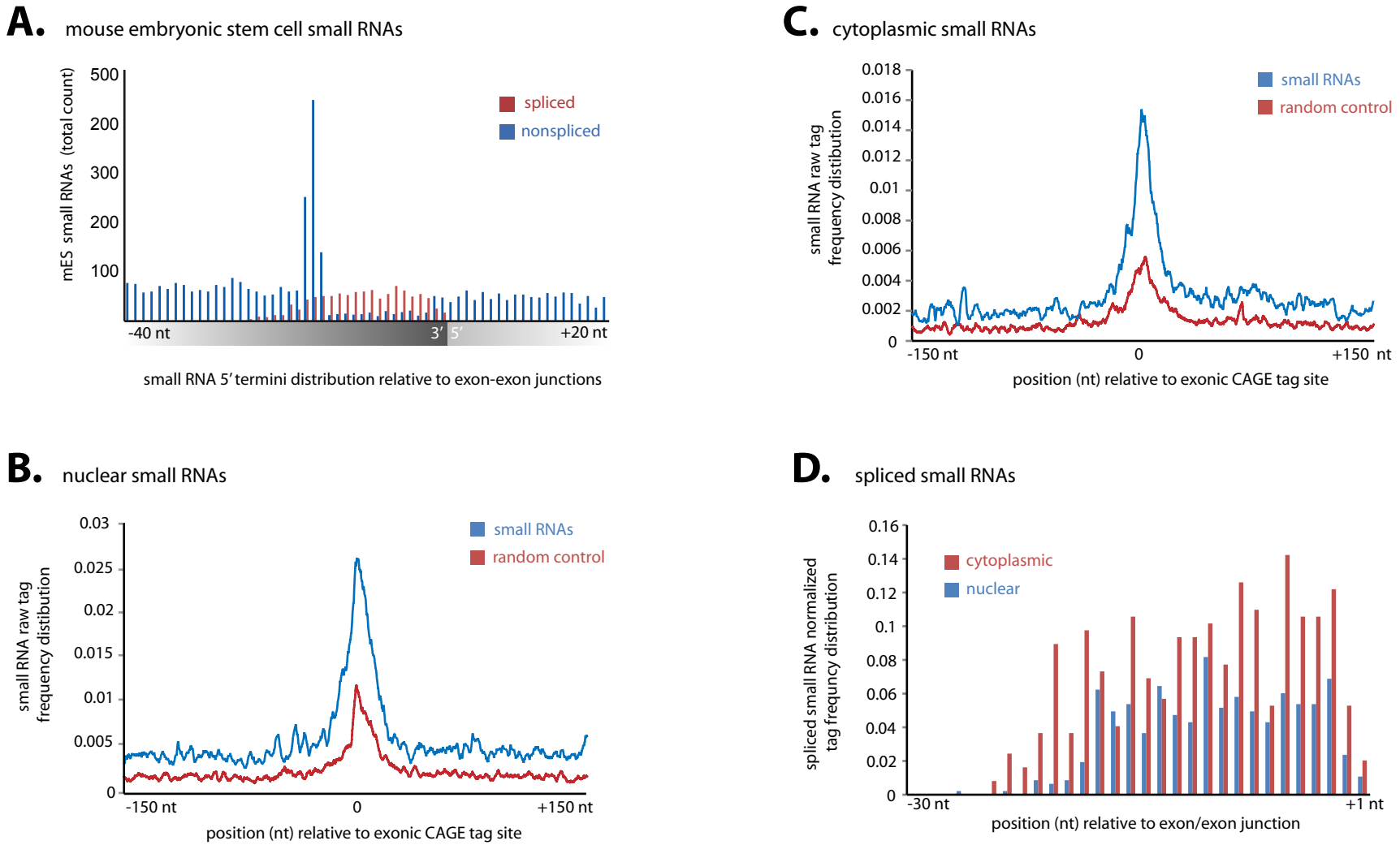
Supplemental Fig. S3. Nucleotide motifs associated with exonic CAGE tags. (A) Sequence composition for a 60 nt window centered on single-peak CAGE tag clusters mapping to promoter (i) or coding exons (ii). **(B)** Sequence composition for a 60 nt window centered on broad-peak CAGE tag clusters mapping to promoter (i) or coding exons (ii). **(C)** Sequence composition for a 60 nt window centred on exonic CAGE tags that are conserved (i) or nonconserved (ii) between human and mouse. **(D)** Sequence composition for a 60 nt window centred on CAGE tag cluster after removal of terminal 5′ nucleotide from CAGE tags to control for potential 5′ guanine non-template nucleotide addition.



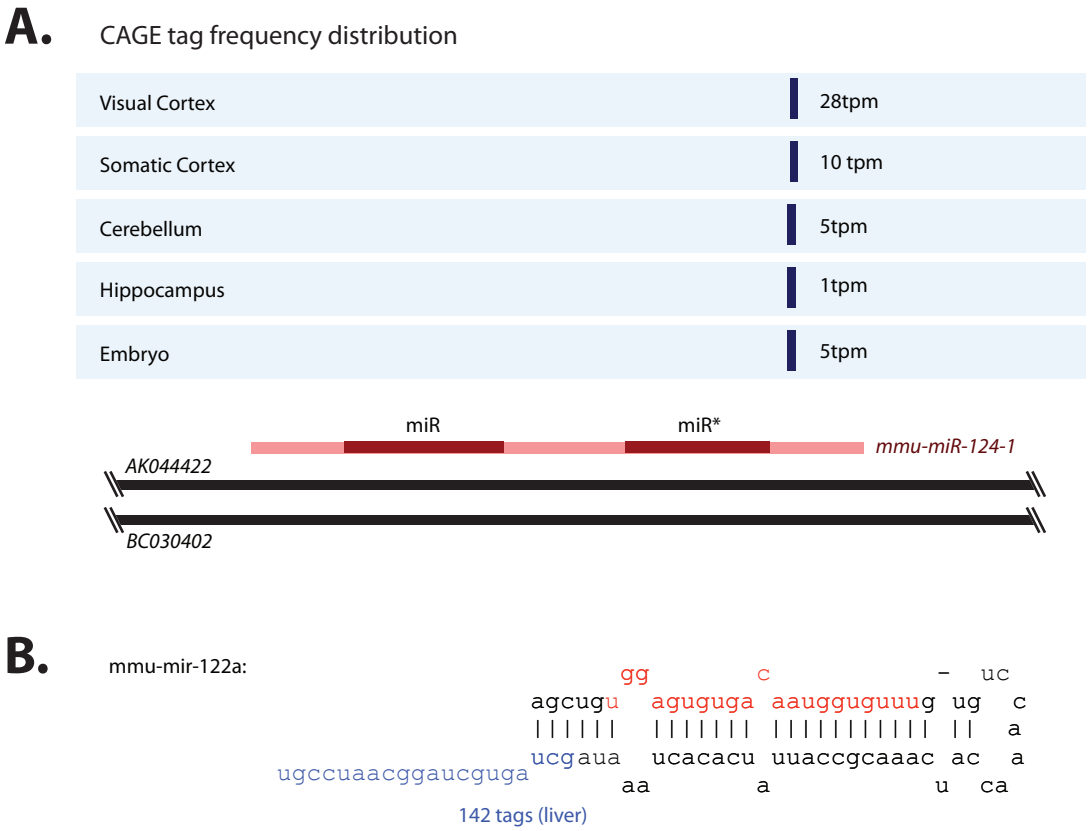
Supplemental Fig S4. Comparative distribution of PARE and CAGE tags. (A,B). Frequency distribution of PARE (red) and CAGE (blue) tags across a 100 nt window centred on transcription start sites in the mitochondrial **(A)** and nuclear **(B)** genome. **(C)** Frequency distribution of PARE (red) and CAGE (blue) tags across gene body. **(D)** Scatter plot comparing CAGE and PARE tag frequencies mapping to RefSeq gene exons. **(E)** Genome browser view of the Alb gene (top panel) with detail shown (bottom panel). The Alb gene is prevalently tiled with both CAG and PARE tags.



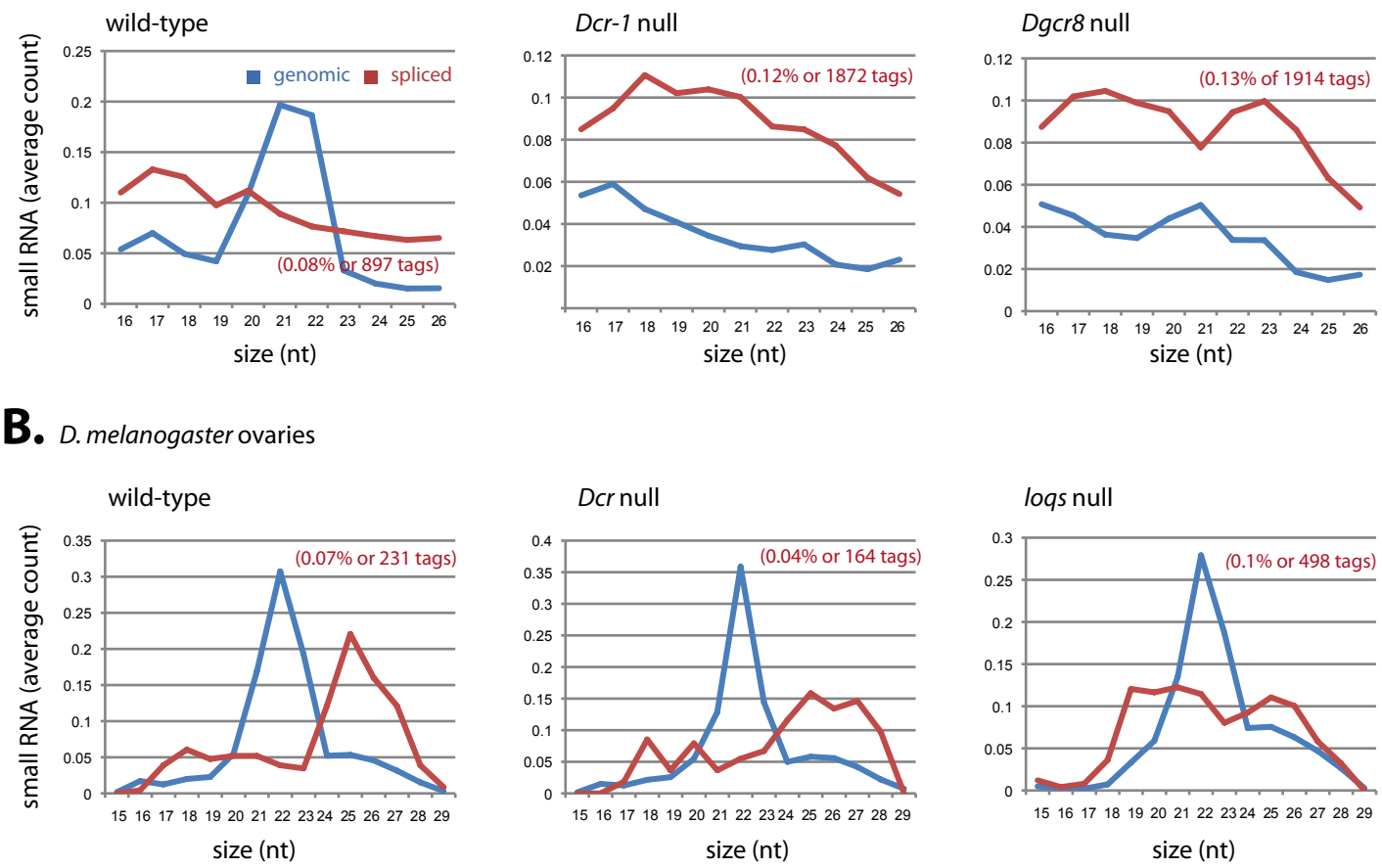
Supplemental Fig. S5. Conservation of cleavage frequency between human and mouse. Scatter-plot graph for cleavage frequency of homologous genes between human and mouse (Pearson's correlation $r^2 = 0.31$).



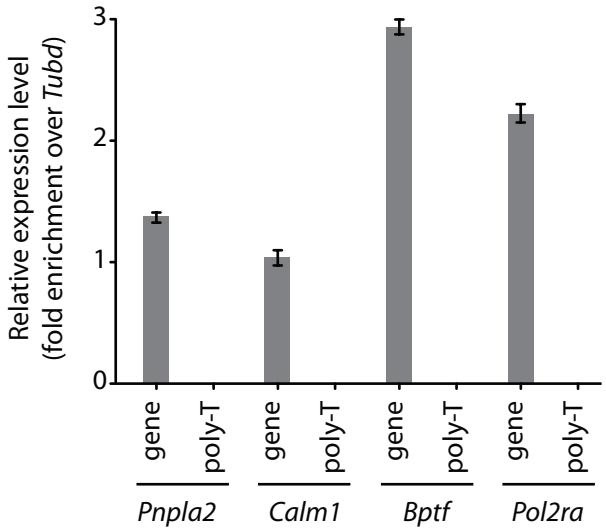
Supplemental Fig. S6. Evidence that post-transcriptional cleavage generates small RNAs. (A) Schematic diagram illustrating the mapping strategy employed to discern small RNAs that span EEJs. Distribution of small RNA tags derived from mouse embryonic stem cells (Babiarz et al. 2008) mapping proximal to RefSeq EEJs. The frequency of small RNAs that map uniquely to the genome (blue) are under-represented adjacent to EEJs. This under-representation may be reconciled with small RNA tags that map across EEJs (red). Average frequency distribution for small RNA tags derived from human THP1 cells nuclear **(B)** and cytoplasmic **(C)** preparations (Taft et al., 2009) within a 200 nt window relative to exonic CAGE tags (blue) and random nucleotides within matched exons (red). **(D)** Frequency distribution of small RNAs derived from THP1 cytoplasmic (red) and nuclear (blue) across EEJs.



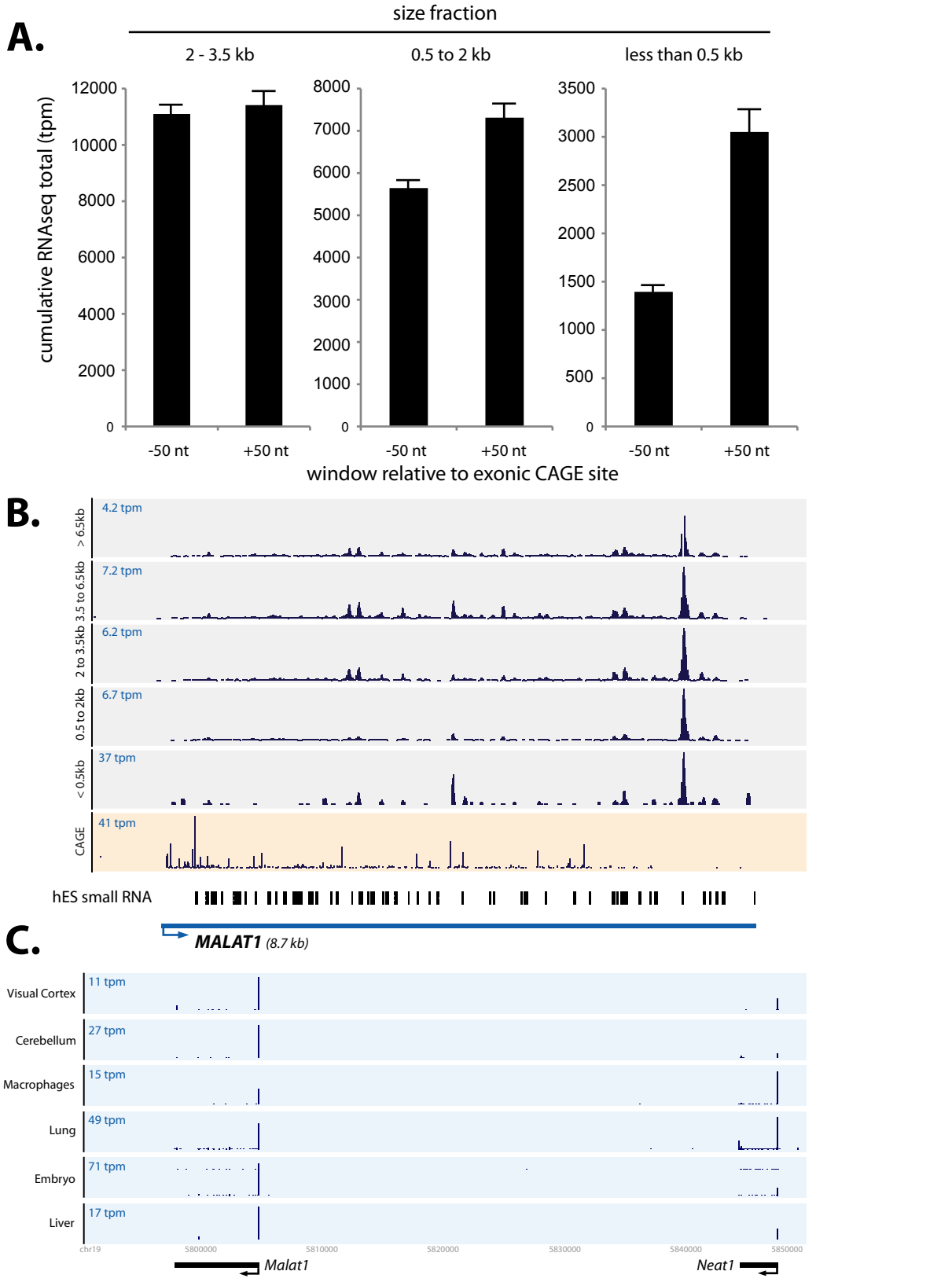
Supplemental Fig. S7. Secondary capping of microRNA precursors. (A) Genome browser view of *mmu-miR-124-1* hairpin (light red bar) showing CAGE tags immediately downstream of star sequence (dark red bar) in mouse tissues (blue panels). The *mmu-miR-124-1* hairpin is hosted within full-length RNA transcripts (black bars). **(B)** Schematic diagram of the *mmu-miR-122a* hairpin showing mature microRNA sequence (red) and CAGE tag locations (blue) in mouse liver.



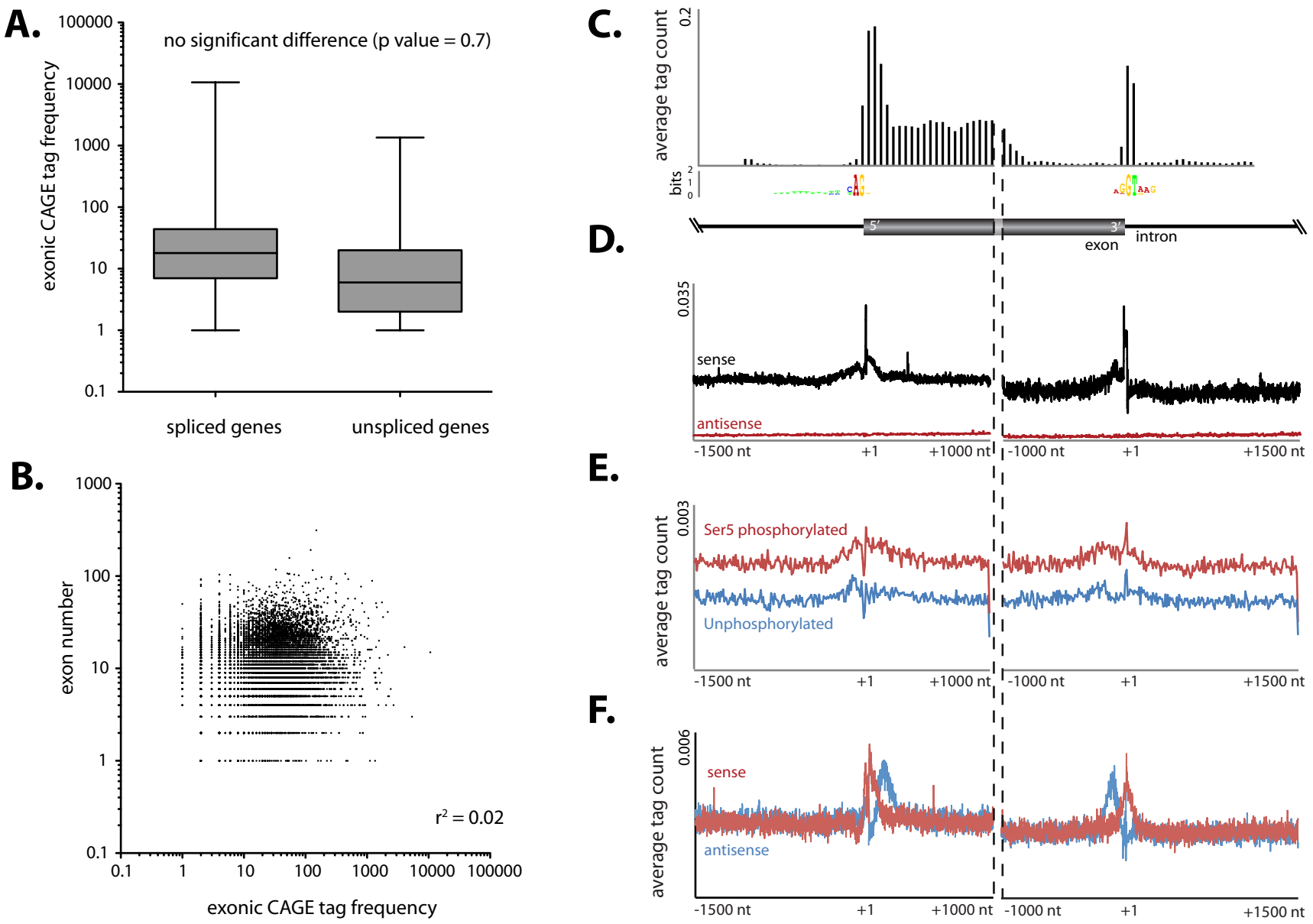
Supplemental Fig. S8. Characterization of exon-derived small RNAs. (A) Average small RNA length frequency from wild-type, *Dcr-1* null and *Dgcr8* null mutant mouse embryonic stem cells (Babiarz et al. 2008) that map to the genome (blue) or across EEJs (red; total count indicated). **(B)** Average small RNA length frequency from wild-type, *Dcr* null or *loqs* null mutant flies that map to the genome (blue) or across EEJs (red; total count indicated).



Supplemental Fig. S9. Detection of polyadenylated transcripts upstream of exonic cleavage sites. Quantitative PCR using gene-specific reverse primers (gene) and poly-T primers (poly-T) indicates the absence of an upstream polyadenylated transcript for exonic cleavage sites within the *Pnpla2*, *Calm1*, *Bptf* and *Pol2ra* genes. Fold enrichment of gene relative to *Tubd* is indicated with gene specific primers (error bars indicate standard deviation).

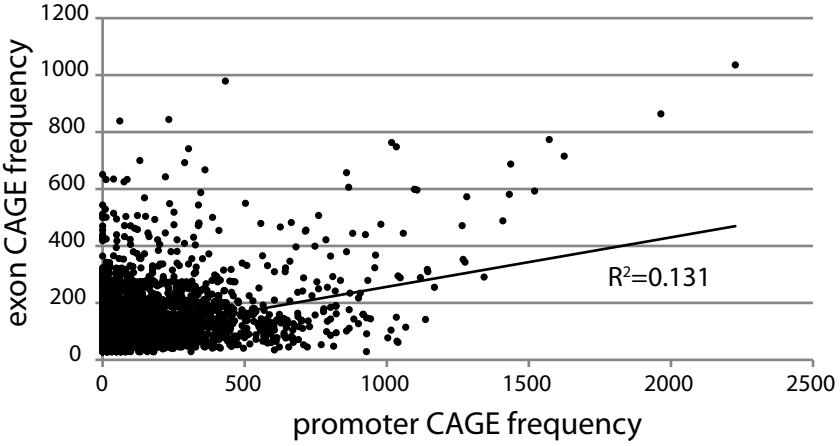


Supplemental Fig. S10. Size-fractionated RNA deep sequencing analysis. (A) The sum total RNAseq tags for 50 nt windows centered around exonic CAGE tags that occur within genes larger than 2 kb shows progressive enrichment for downstream windows in smaller size fractions. **(B)** The *MALAT1* ncRNA encompass prevalent exonic CAGE tags (orange panel) and are expressed in all size fractionated RNAseq preparations (gray panels). Furthermore, recent small RNA libraries derived from human embryonic stem cells indicate numerous sense small RNAs associated with the *MALAT1* transcript (Lister et al. 2009). **(C)** Both the *Malat1* ncRNA and the associated *Neat1* ncRNAs (black bars) are subject to prevalent cleavage as indicated by exonic CAGE tags in mouse tissues (blue panel).

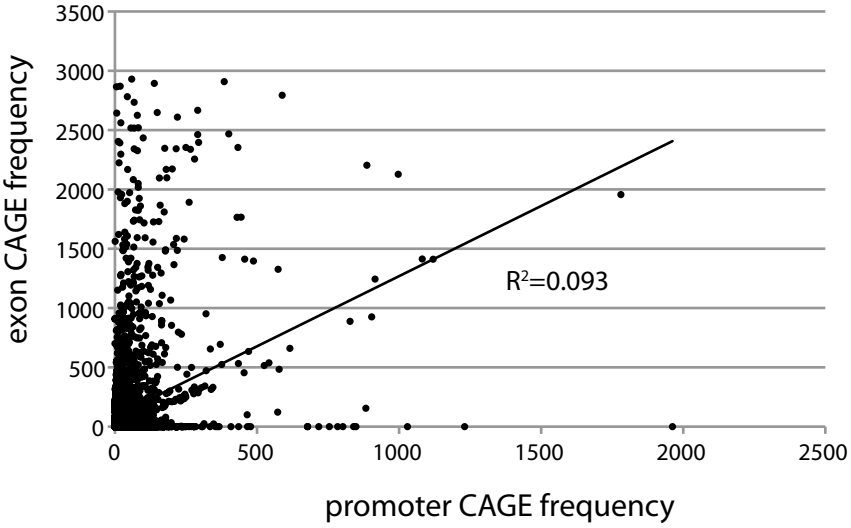


Supplemental Fig. S11. Enrichment of CAGE tags at 5′ end of exons. (A) Comparison of exonic CAGE tag frequency for spliced versus unspliced genes indicates no significant difference (p-value = 0.7, unpaired t-test) **(B)** Scatter-plot indicates no correlation ($r^2 = 0.02$) between number of splicing events and exonic CAGE tag frequency. **(C)** Average CAGE tag frequency distribution within 20 nt window centered on splice junctions with accompanying nucleotide composition. **(D)** Average GRO-seq sense (black) and antisense (red) frequency distribution within a 2.5 kb window centred on splice junction (Core et al. 2008). **(E)** The average frequency of DNA tags immunoprecipitated with antibodies to phosphorylated (red) and unphosphorylated (blue) RNA polymerase II across splice junctions. **(F)** The average frequency of MNase digested sense (red) and antisense (blue) DNA tags show nucleosome positioning across splice junction (Wang et al. 2008).

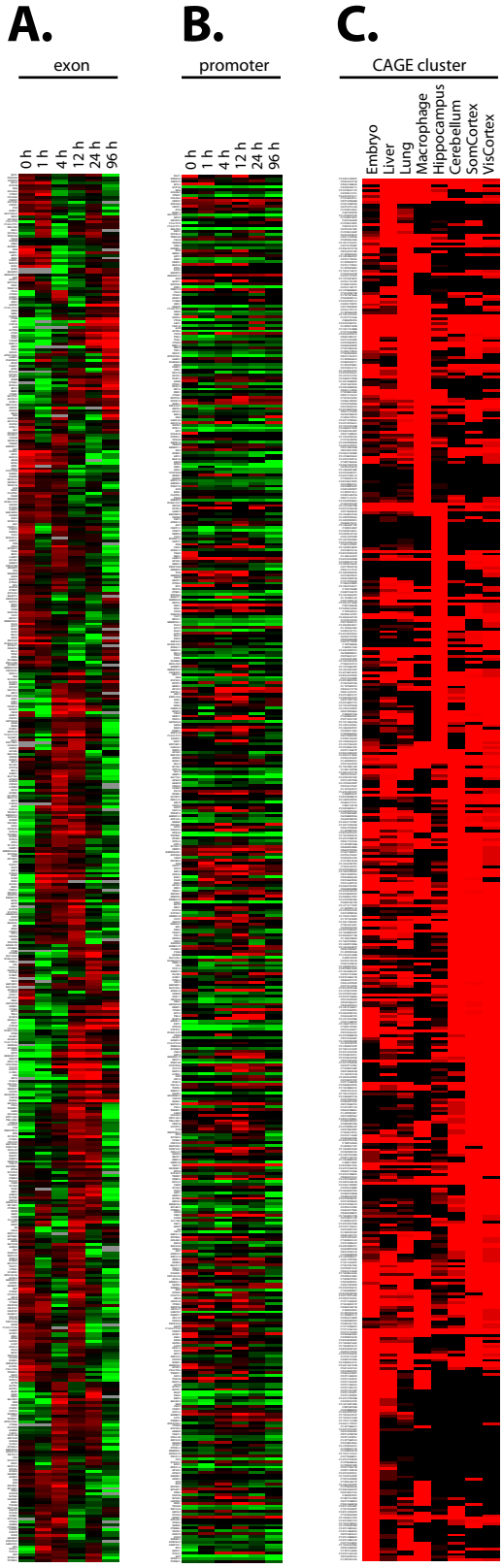
A. human (THP1 differentiation)



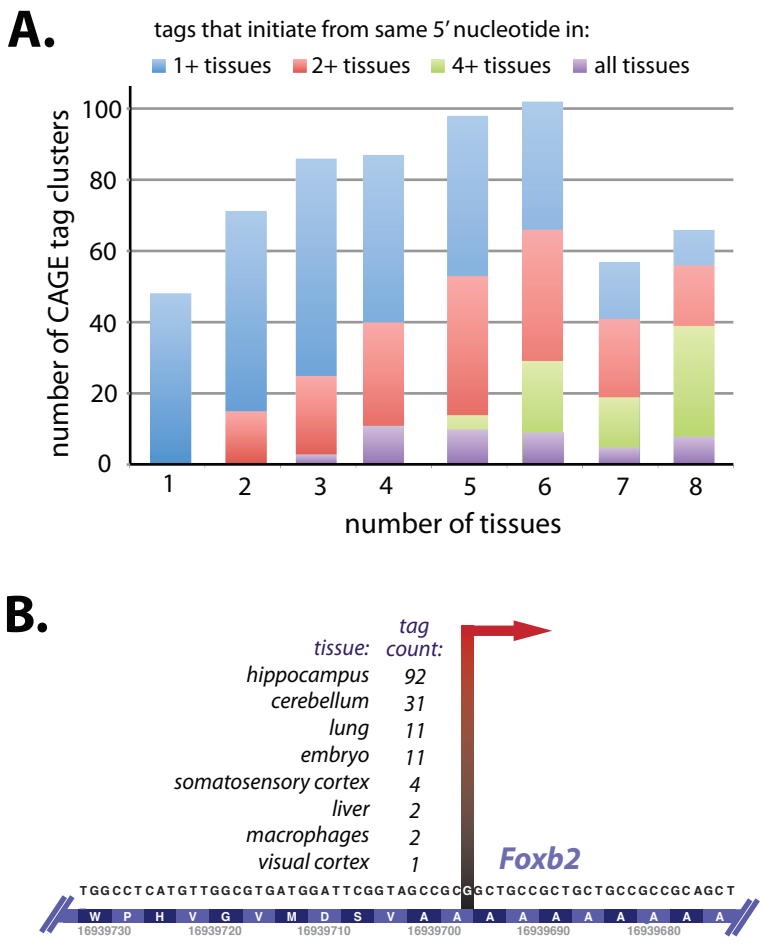
B. mouse (tissue)



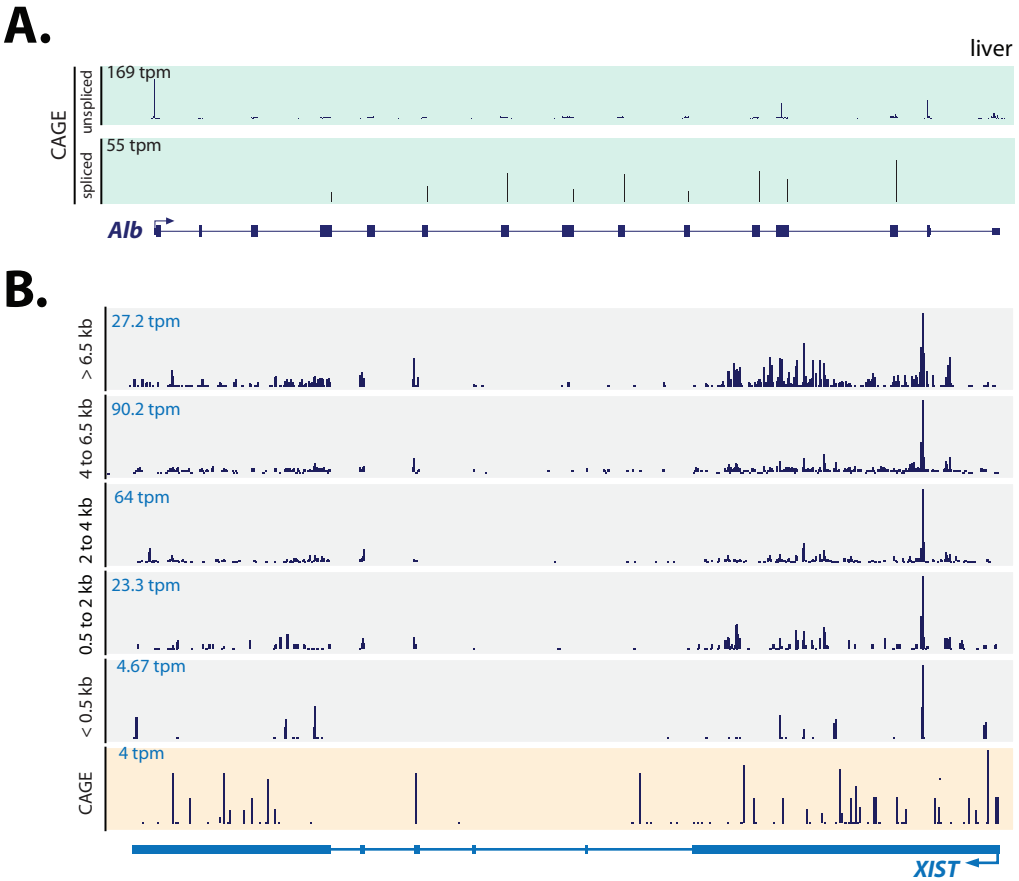
Supplemental Fig. S12. Correlation between promoter and exonic CAGE tags. (A,B) Comparison of CAGE tags mapping to promoter relative to coding exons for the top 500 selected genes in human THP1 cell differentiation **(A)** or mouse tissue **(B)**. Low Pearson's correlation (r^2) values suggest little correlation between promoter expression and cleavage levels.



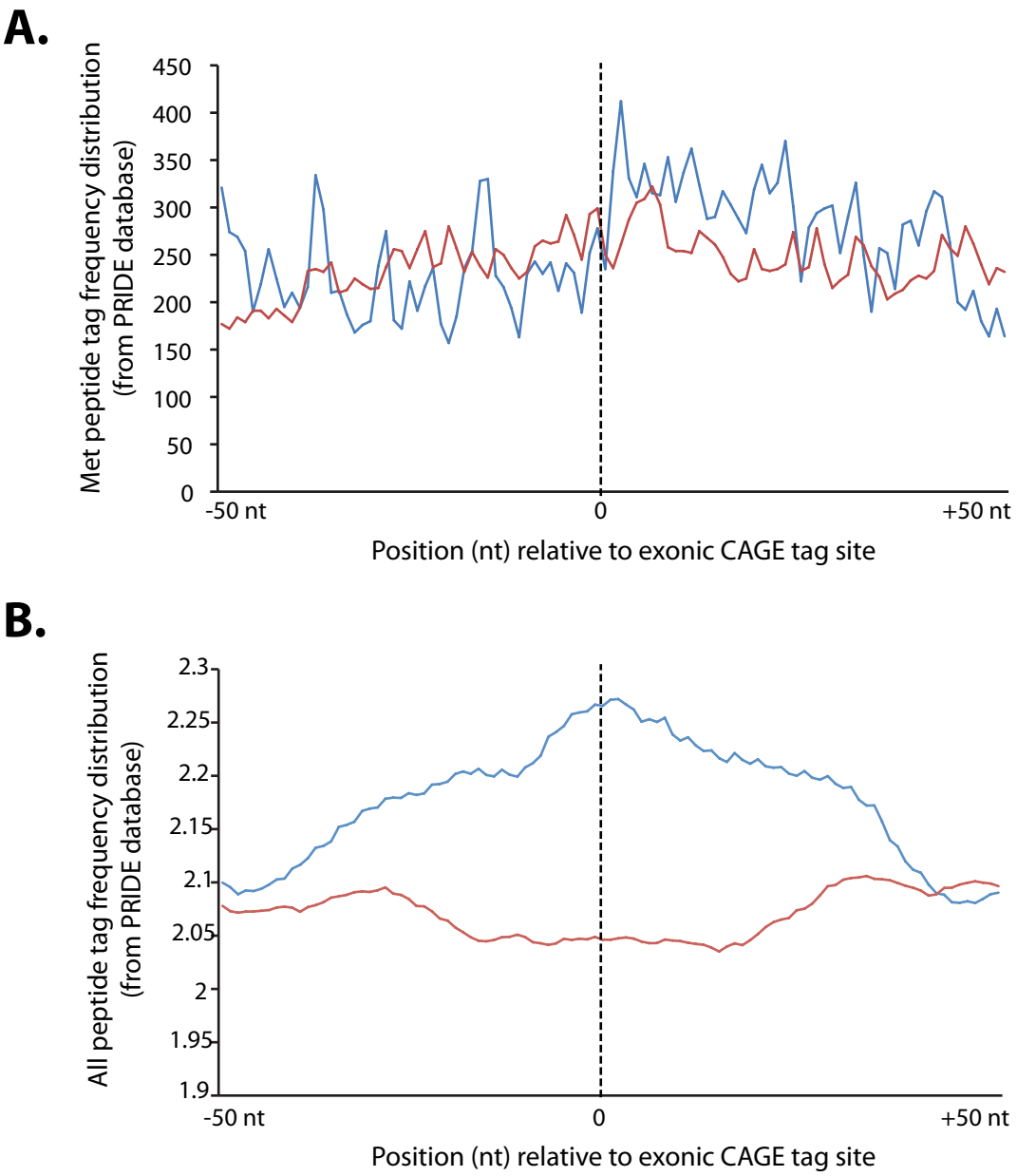
Supplemental Fig. S13. Differential post-transcriptional cleavage during THP-1 cell differentiation. **(A)** Cluster analysis of the expression of the 500 genes containing the highest exonic CAGE tag frequency shows the up (red) and down (green) regulation of mRNA cleavage in human THP-1 cell differentiation. Expression level was determined as the normalized CAGE tag frequency mapping to coding exons. **(B)** The diverse ratios of CAGE tags mapping to promoter relative to exons for each time point indicates relative independence of phenomenon. **(C)** Cluster analysis of the expression of the 500 CAGE tag clusters (Valen et al. 2009) shows differential regulation of mRNA cleavage across mouse tissues.



Supplemental Fig. S14. Tissue specificity of post-transcriptional cleavage. (A) Proportion of CAGE tag clusters (Valen et al. 2009) where the preferred nucleotide to which the highest frequency of CAGE tags map is preferred across more than one tissue. **(B)** CAGE tags from eight different tissues map to a single nucleotide (red arrow) within the mouse *Foxb2* gene.



Supplemental Fig. S15. Examples of small RNA genesis by post-transcriptional cleavage. **(A)** Liver CAGE tags (green panel) map prevalently to the mouse *Alb* gene coding exons. **(B)** The *XIST* ncRNA encompass prevalent exonic CAGE tags (orange panel) and are expressed in all size fractionated RNAseq preparations (gray panels).



Supplemental Fig. S16. Enrichment of start codons (Met) downstream of post-transcriptional cleavage sites. (A) Frequency distribution of peptide tags initiating with methionine (Met) residue across a 100 nt window centered on exonic CAGE tags (blue) and random exonic nucleotides (red, control). **(B)** Frequency distribution of peptide tags across a 100 nt window centered on exonic CAGE tags (blue) and random exonic nucleotides (red, control). Peptide tags downloaded from PRIDE database (www.ebi.ac.uk/pride/).