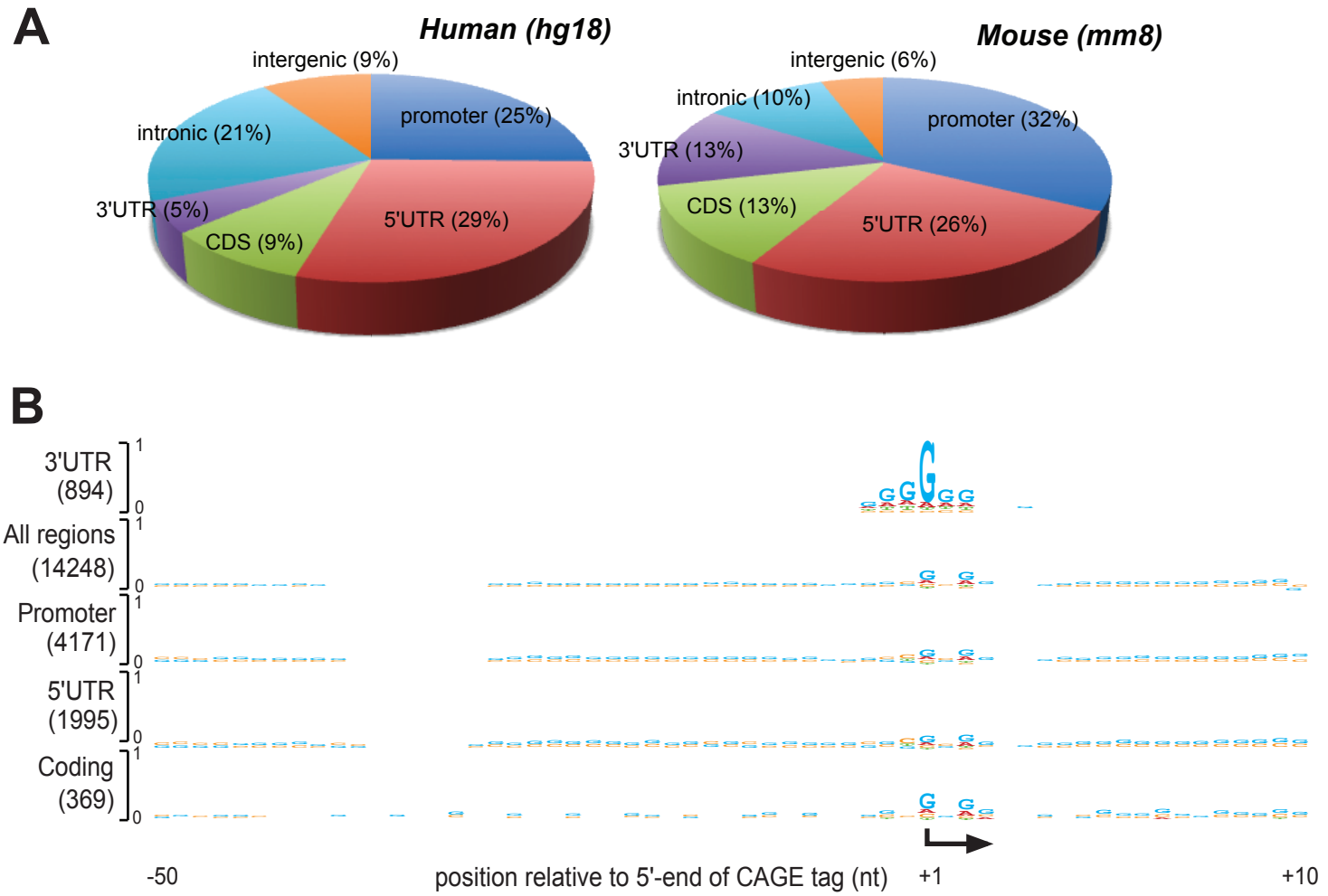

Fig. S1. Supporting bioinformatic analyses of CAGE and cDNA data. (A) Proportion of CAGE tags that map to different regions of the human and mouse genome. (B) Nucleotide composition associated with CAGE tag sites parsed according to genomic location. Genomic sequence 50 nt upstream and 10 nt downstream of the 5' end of CAGE tags were visualized using SeqLogo as described previously (18). CAGE tags were parsed into 3'UTRs, promoters, 5'UTRs and coding regions according to mouse RefSeq annotations. The number of CAGE tags associated with each region is shown in brackets.

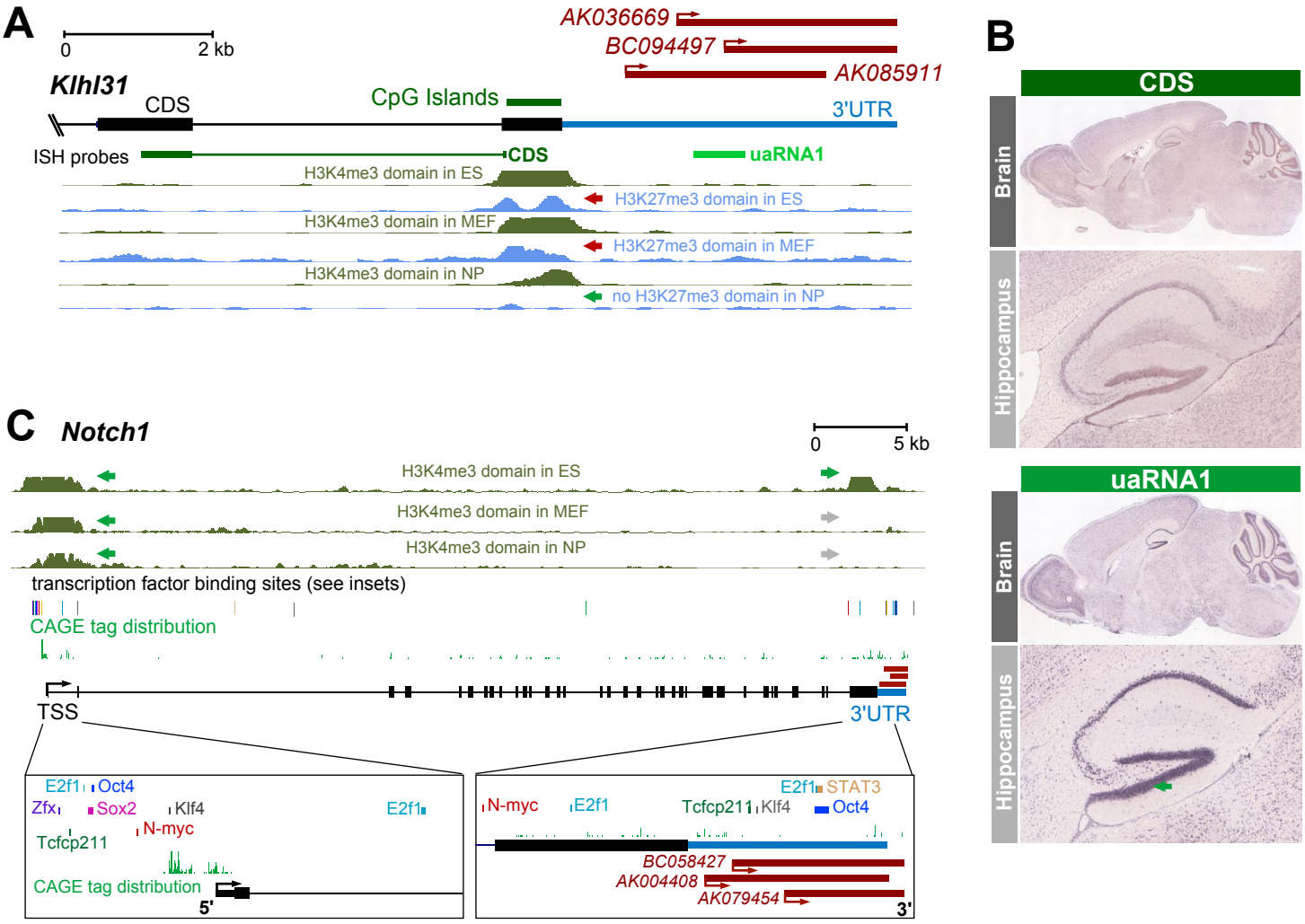

Fig. S2. Examples of active promoters within 3'UTRs. (A) Chromatin modifications at the 3' end of the *Klhl31* gene in embryonic stem (ES) cells, mouse embryonic fibroblasts (MEFs) and neuronal progenitors (NPs). Density of H3K4me3 (dark green) and H3K27me3 (light blue) chromatin marks are shown as histograms. The coding region (CDS) of *Klhl31* is shown as a thick black bar, the intron as a black line and the 3'UTR as a blue bar. Overlapping H3K4me3 and H3K27me3 marks (known as bivalent domains and are normally indicative of genes poised for transcription) are evident in ES cells and MEFs (red arrow), whereas in NPs these bivalent domains are resolved to permissive H3K4me3 marks (green arrow). *AK036669*, *BC094497* and *AK085911* are independent full-length cDNA sequences that map to the *Klhl31* 3'UTR. B. In situ hybridization (ISH) of probes targeting the 3'UTR (uaRNA1) and coding sequence (CDS) of *Klhl31* in the adult mouse brain (images courtesy of the Allen Brain Atlas). Genomic mappings of the ISH probes are shown in A. Consistent with the expression anticipated from the chromatin marks, the uaRNA1 probe shows strong and specific expression of the 3'UTR in the hippocampus (green arrow), whereas the CDS probe showed low expression. (C) Genomic context of *Notch1* showing H3K4me3 domains associated with the *Notch1* transcription start site (TSS) in ES cells, MEFs and NPs (green arrows). In contrast, H3K4me3 domains are associated with the 3'UTR in ES cells only (green arrow) and absent in MEFs and NPs (grey arrows) suggesting discordant regulation of the mRNA and 3'UTR at the chromatin level. The lower panel shows transcription factor binding sites associated with the 5' promoter region or the 3'UTR (see inset for higher resolution). Many similar transcription factors (including E2F1, OCT4, KLF4, N-MYC and TCFCP211) are associated with both the 5' promoter and 3'UTR. However, there are also differences (ZFX and SOX2 binding to the 5' promoter and STAT3 binding to the 3'UTR) suggesting discordant regulation by transcription factors between the 5' promoter and 3'UTR. *BC058427*, *AK004408* and *AK079454* are independent full-length cDNA sequences that map to the *Notch1* 3'UTR.

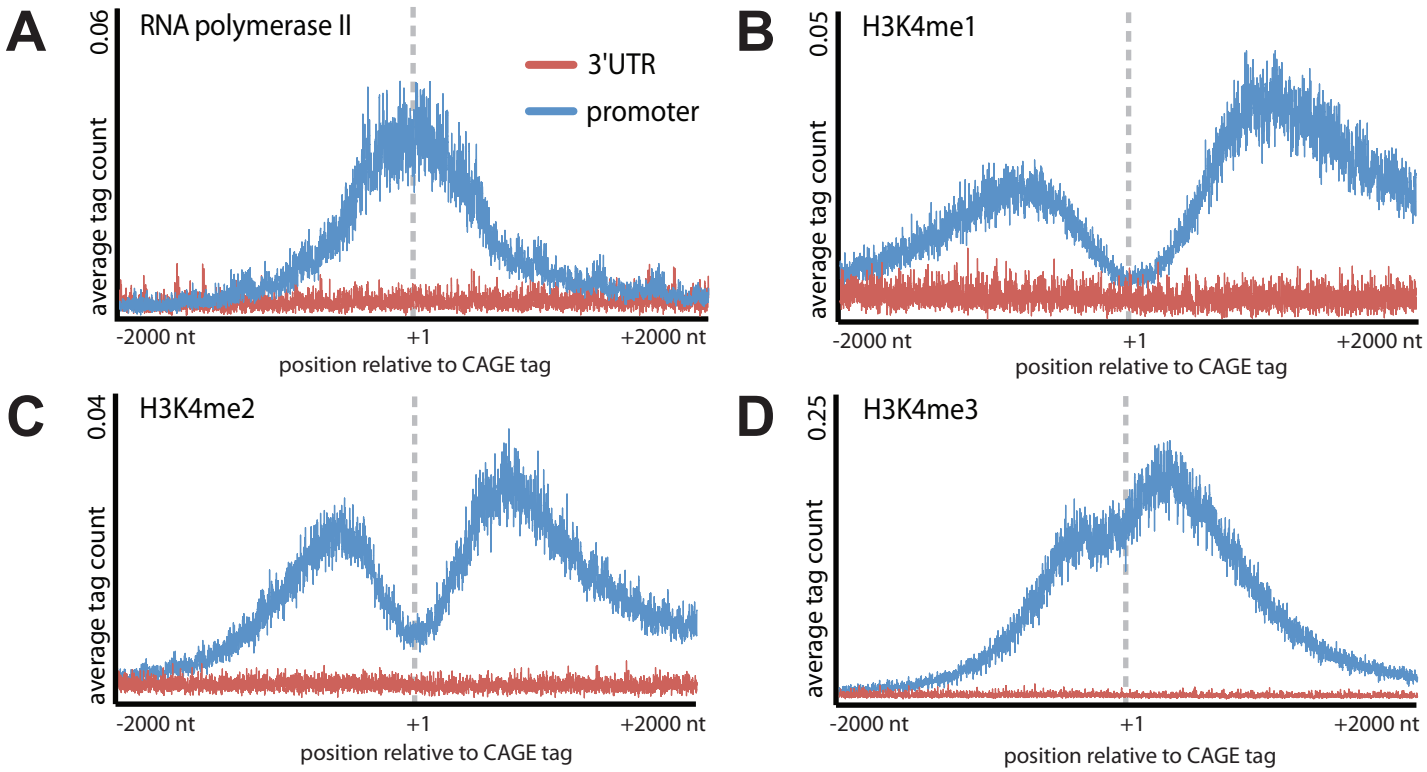

Fig. S3. Distribution of average tag frequency of RNA polymerase II occupancy and modified histones in human CD4+ T cells. (A-D) The average frequency of DNA tags immunoprecipitated with RNA polymerase II (A), H3K4me1 (B), H3K4me2 (C) and H3K4me3 (D) within a 4 kb window centered on CAGE tags at promoters (blue) and 3'UTRs (red).

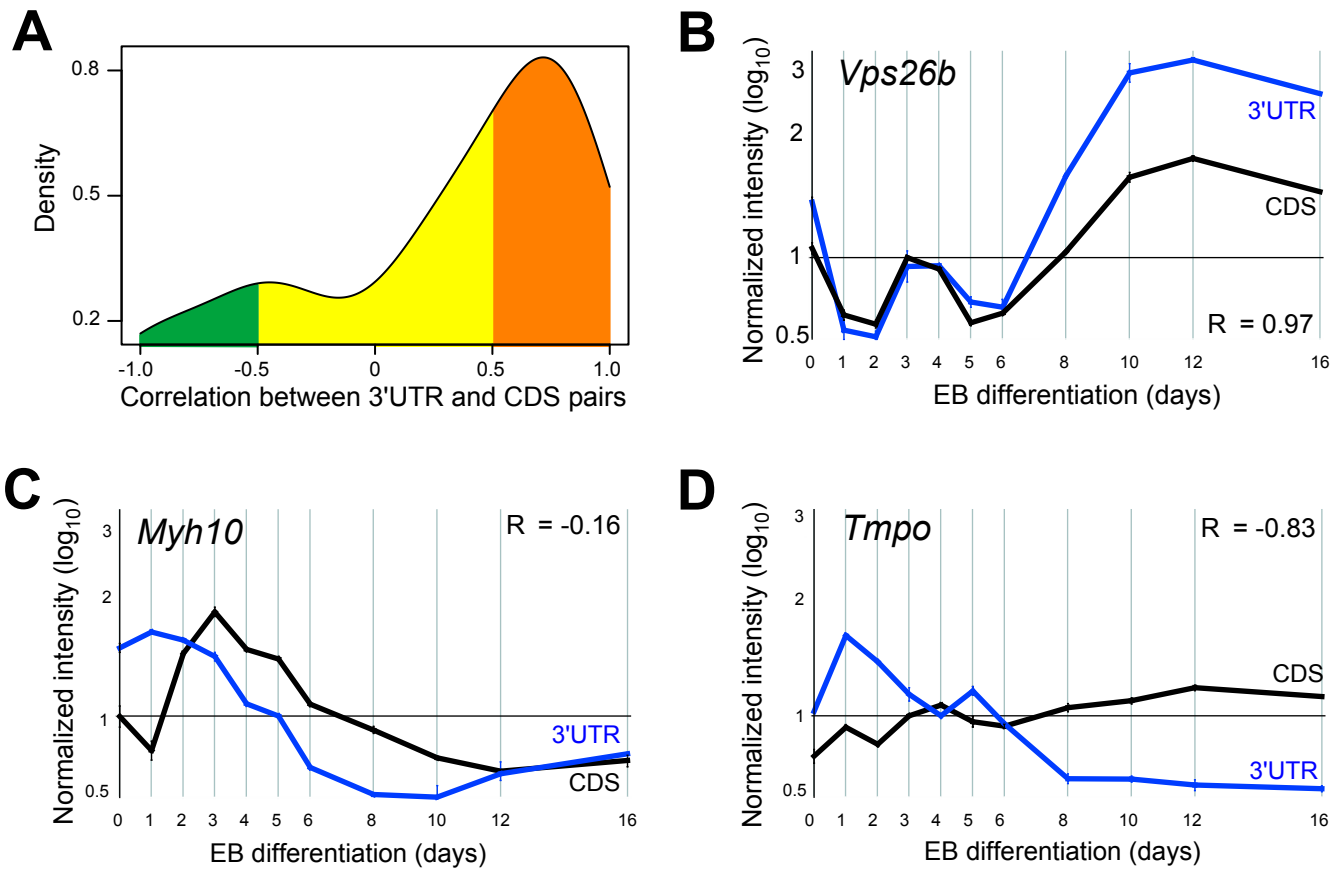

Fig. S4. Independent 3'UTR expression by microarray analysis of embryonic stem (ES) cell differentiation. (A) Kernel density plot of expression profile correlation (Pearson's correlation coefficient; *r*) between coding (CDS) and 3'UTR probe pairs. Probe pairs were defined as positively correlated ($r \geq 0.5$; orange), uncorrelated ($-0.5 < r < 0.5$; yellow) or negatively correlated ($r \leq -0.5$; green). (B-D) Examples comparing expression of coding and 3'UTR probe pairs. (B) Positive correlation between *Vps26b* coding and 3'UTR probe pairs. (C) No correlation of expression between *Myh10* coding and 3'UTR probe pairs. (D) Inverse correlation of expression between *Tmpo* coding and 3'UTR probe pairs.

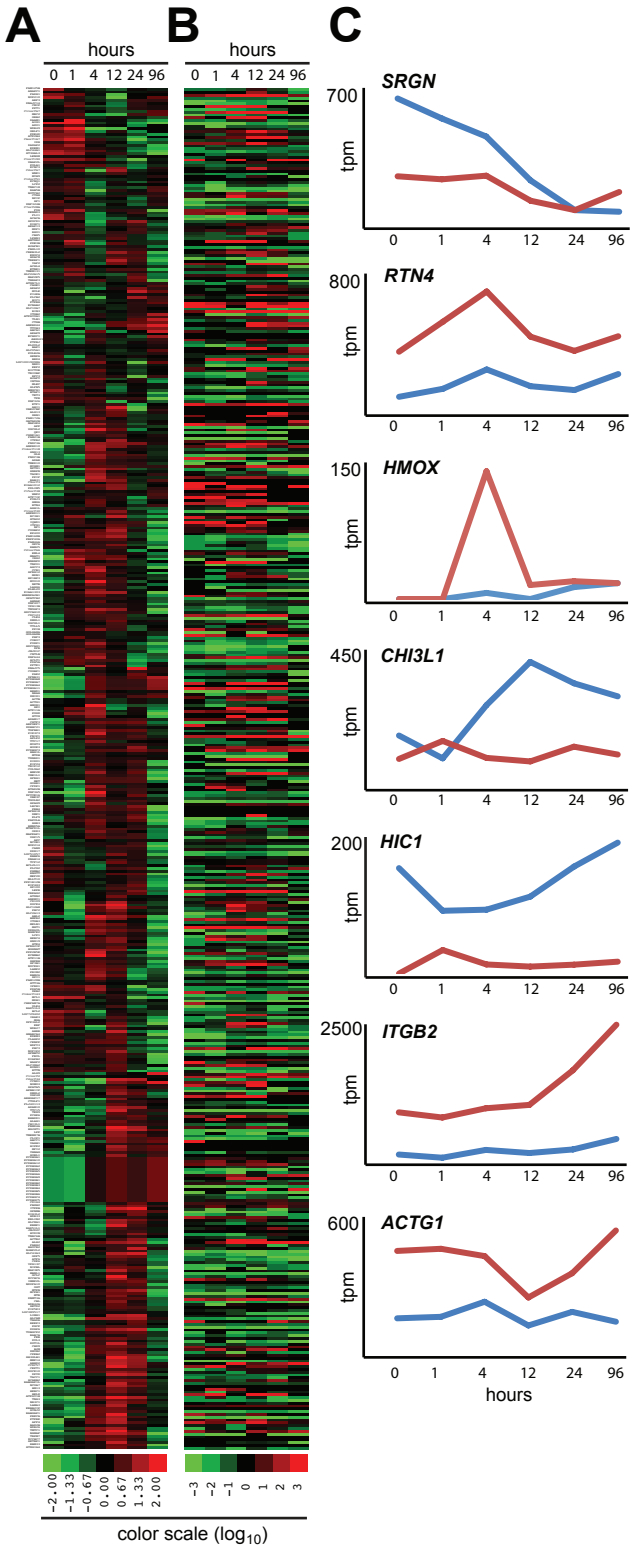

Fig. S5. Specific and independent expression profiles for 3'UTRs in human THP-1 cell differentiation. (A) Cluster analysis of the expression of the 500 genes containing the highest 3'UTR CAGE tag frequency shows the up (red) and down (green) regulation of 3'UTRs in human THP-1 cell differentiation. Expression level was determined as the normalized CAGE tag frequency mapping to 3'UTRs. (B) The diverse ratios of promoter to 3'UTR CAGE frequency for each time point indicate independent expression of mRNAs and 3'UTRs. (C) Illustrative examples indicating promoter (blue) and 3'UTR (red) CAGE frequency. tpm; tags per million.

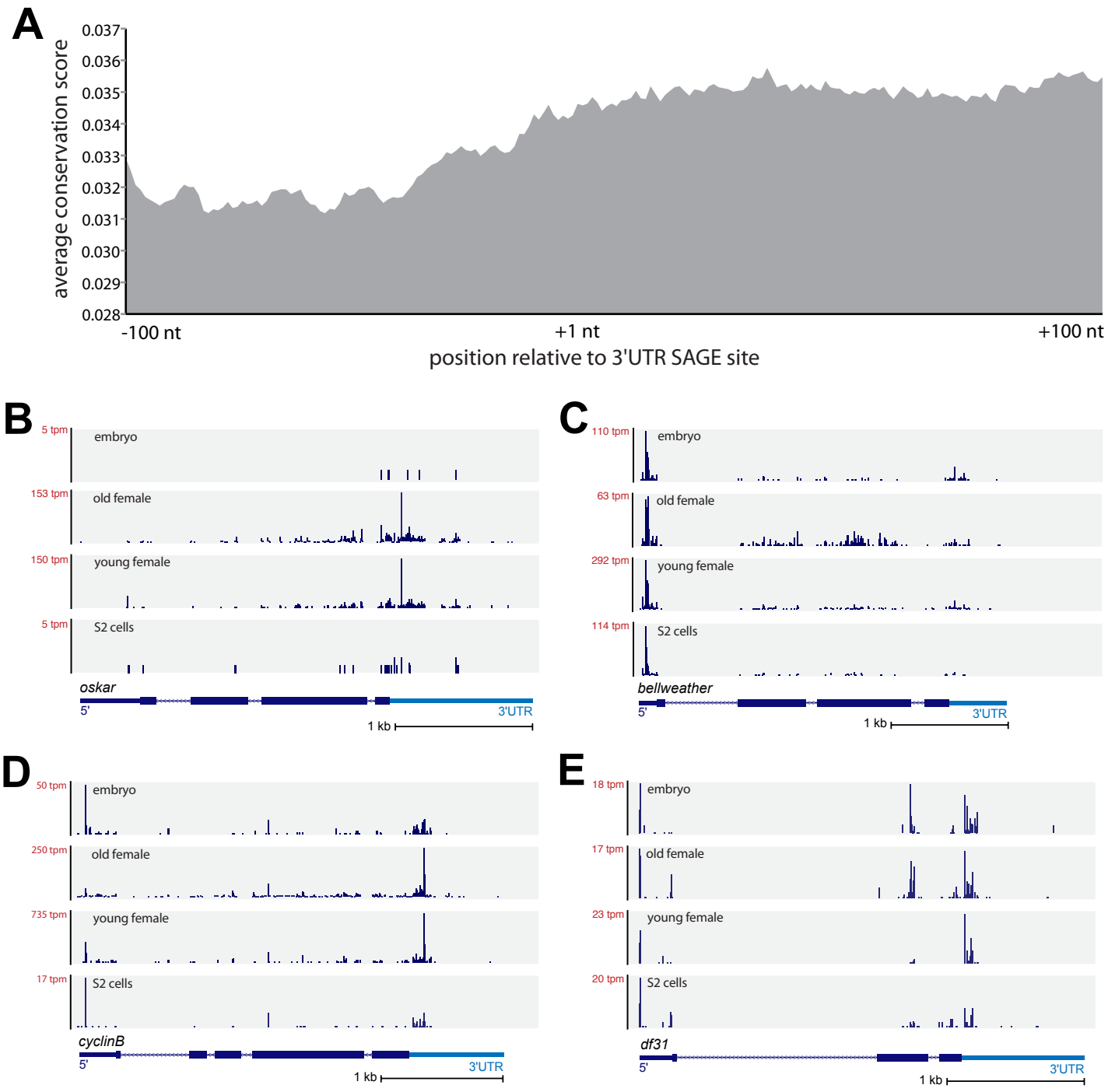

Fig. S6. Supplementary *Drosophila melanogaster* SAGE tag analysis. (A) Conservation profile surrounding 3'UTR SAGE tags in *Drosophila melanogaster*. Histogram showing that the conservation is higher downstream (+100 nt) of 3'UTR SAGE tags (+1 nt) relative to upstream (-100 nt). The conservation score was determined using the UCSC MultiZ alignment, which is based on 12 flies, mosquito, honeybee and beetle. (B-E) Examples of 3'UTR initiated transcripts in Drosophila melanogaster based on SAGE data. SAGE tag counts are represented as histograms for each developmental stage (old female, young female, S2 cells and embryo) across the length of four genes; *oskar* (B), *bellweather* (C), *cyclinB* (D) and *df31* (E). The 5'UTR is represented as a thin bar, the CDS as a thick bar, and the 3'UTR as a thin light blue bar. tpm; tags per million.

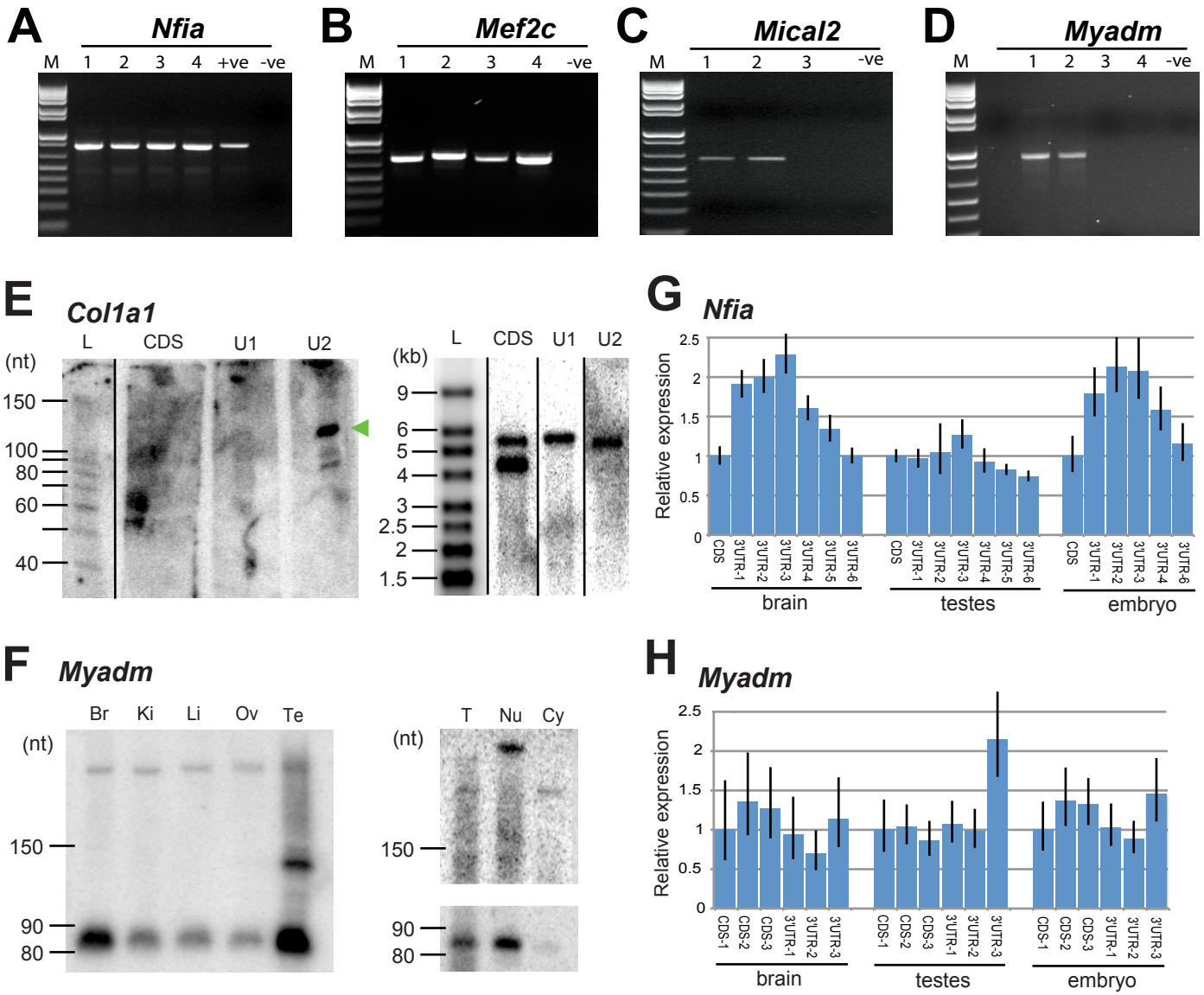

Fig. S7. Validation of 3'UTR connection to the associated mRNA and independent expression by Northern analysis. (A-D) Gene-specific cDNA synthesis with different primers (1 to 4) along the 3'UTR, followed by PCR for the terminal exon indicates that the 3'UTRs is contiguous with the terminal exon of the *Nfia* (A), *Mef2c* (B), *Mical2* (C) and *Myadm* (D) genes. M, 1 kb plus DNA ladder; +ve, random-primed positive control; -ve, no RT (reverse transcriptase) template negative control. (E) Northern blot analyses using the CDS, uaRNA1 and uaRNA2 Col1a1 ISH riboprobes on mouse embryo total RNA resolved on two gels to optimize separation of small RNAs (left panel) and large RNAs (right panel). The small RNA blot (left) show that the uaRNA2 probe detects a small RNA of ~120 nt (green arrowhead). The large RNA blot (right) shows that each of the riboprobes detect the *Col1a1* transcript. The CDS probe additionally detects an alternate smaller *Col1a1* isoform. (F) (left panel) Northern blot of RNAs from brain (Br), kidney (Ki), liver (Li), ovary (Ov), testis (Te) using the *Myadm* uaRNA1 ISH probe. The blot show the testis-specific expression of a ~140 nt transcript within the *Myadm* 3'UTR. (right panel) Northern blot of total, nuclear and cytoplasmic RNA fractions from C2C12 myoblast cell culture. The blot shows the nuclear-enriched expression of the ~85 nt transcript derived from the *Myadm* 3'UTR. (G,H) Quantitative RT-PCR comparing the expression of the terminal coding exon with various regions within the 3'UTR of *Nfia* (G) and *Myadm* (H) from total RNA extracted from mouse brain, testis and whole embryo. Reverse transcription was performed with random primers followed by amplification with sequence-specific primer pairs (SI Dataset I).

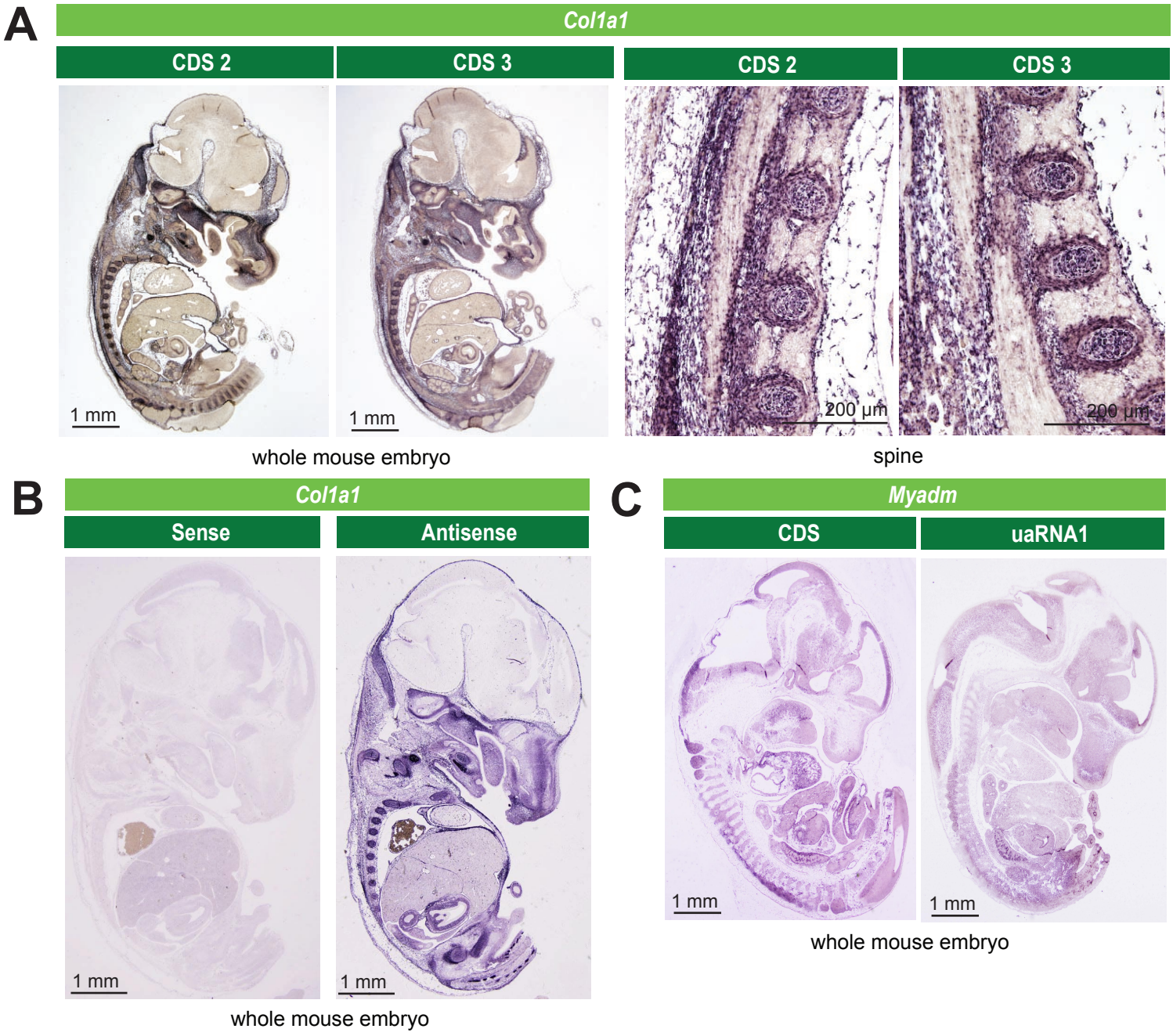

Fig. S8. Supplementary *in situ* hybridization (ISH) data for *Col1a1* and *Myadm*. (A) ISH of whole mouse embryos (left panels) and detail of the ribs (right panels) using two different riboprobes targeting constitutive coding exons (CDS2 and 3, dark green) of the *Col1a1* gene. The expression patterns observed with CDS2 and 3 probes are consistent with those derived from the CDS1 probe (Fig. 4). (B) ISH of whole mouse embryos using *Col1a1* CDS1 sense (left) and antisense (right) probes. The sense probe shows the level of background hybridization. Sense controls for all ISH probes used yielded similar results. (C) View of the whole mouse embryos for the *Myadm* CDS and uaRNA1 ISHs shown in Fig. 5.