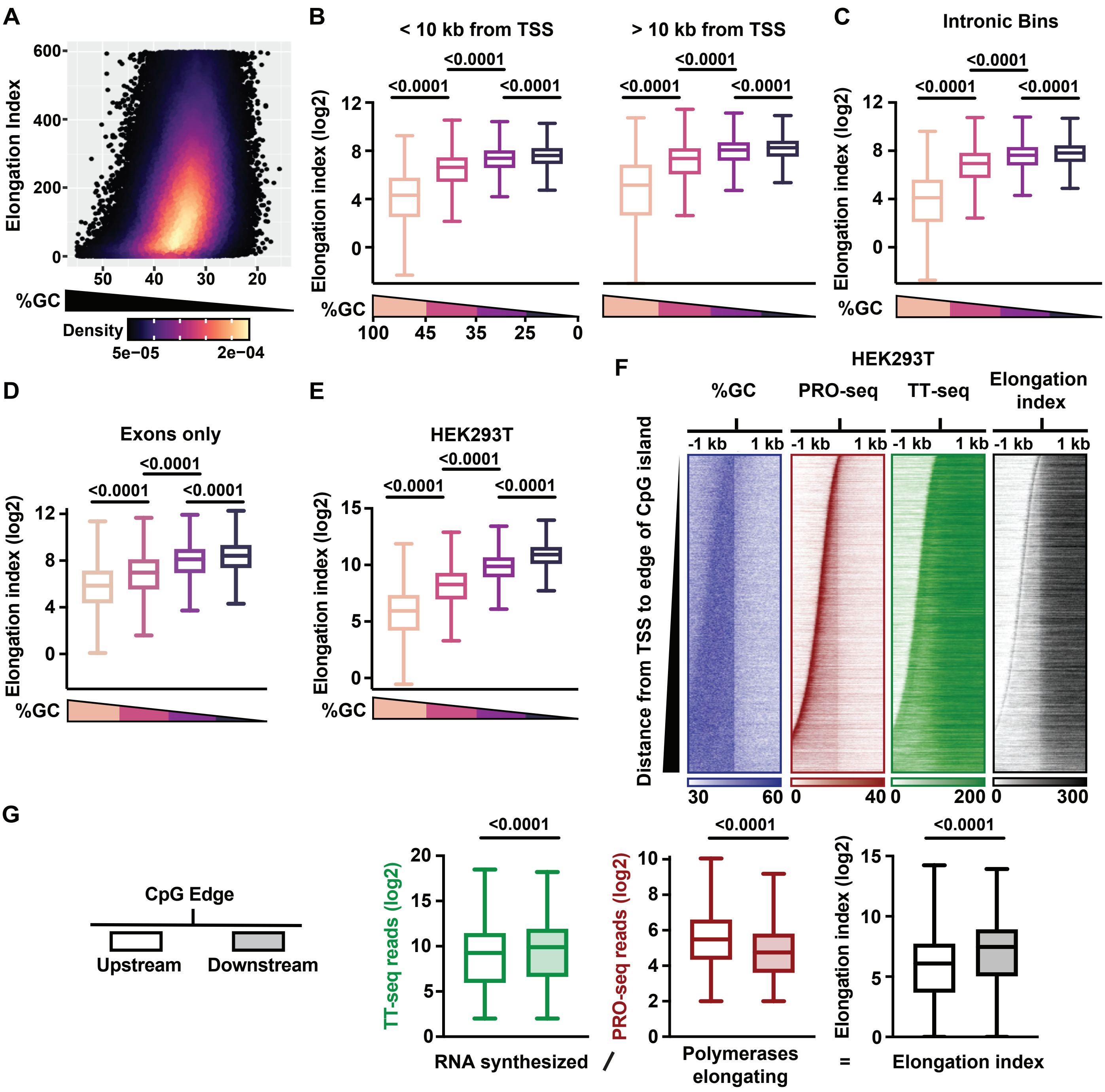

Figure S1. RNAPII elongation rate is influenced by GC content in both mouse and human cells. Related to Figure 1

(A) The relationship between elongation index and %GC across active protein coding genes in mESCs (Intron-containing genes > 1kb; N = 12,327) is shown as a density scatter plot. Genes were divided into 500 nt bins as described in Figure 1B (N = 109,621).

(B) 500 nt bins, as in A, were divided into two groups based on the location of the bin within the gene body (Left, within 10 kb of the TSS, N = 53,335; Right, > 10 kb from TSS, N = 56,296). Bins were then separated into four groups based on %GC (Highest to lowest %GC: Left N = 1619, 19821, 30272, 1623. Right N = 833, 21161, 32468, 1834). Box plots have a line at the median, and whiskers depict 1.5 times the interquartile range. P-values from the Mann-Whitney test.

(C) 500 nt bins, as in A, that overlapped an exon were removed, and the remaining bins were separated into four groups based on %GC (Highest to lowest %GC: N = 893, 23267, 47549, 2541). Box plots and p-values as in B.

(D) Same as (C), but for internal exons in WT mESCs (Highest to lowest %GC: N = 4380, 13478, 8628, 1138)

(E) Active protein coding genes in HEK293T cells were divided into 500 nt bins as described for mESCs. Bins were separated into four groups based on %GC (Highest to lowest %GC: N = 6036, 29679, 30006, 2878). Box plots and p-values as in B.

(F) Heatmap of indicated data aligned around the edges of CpG islands. Shown are active intron-containing genes in HEK293Ts whose promoter overlaps a CpG island (N = 10,079). Heatmaps are ranked by increasing distance from the TSS to the CpG edge. Read counts were summed in 25 nt bins.

(G) Box plots depict the distribution of TT-seq and PRO-seq read densities and elongation index in mESCs in windows located upstream (-125 nt to -25 nt from CpG edge) and downstream (CpG edge +25 nt to +125 nt) of the CpG edge. To avoid biases from promoter proximal RNAPII signal, only genes where the CpG Edge ≥ 400 nt from the TSS are shown (N = 4,906). P-values from the Wilcoxon test.

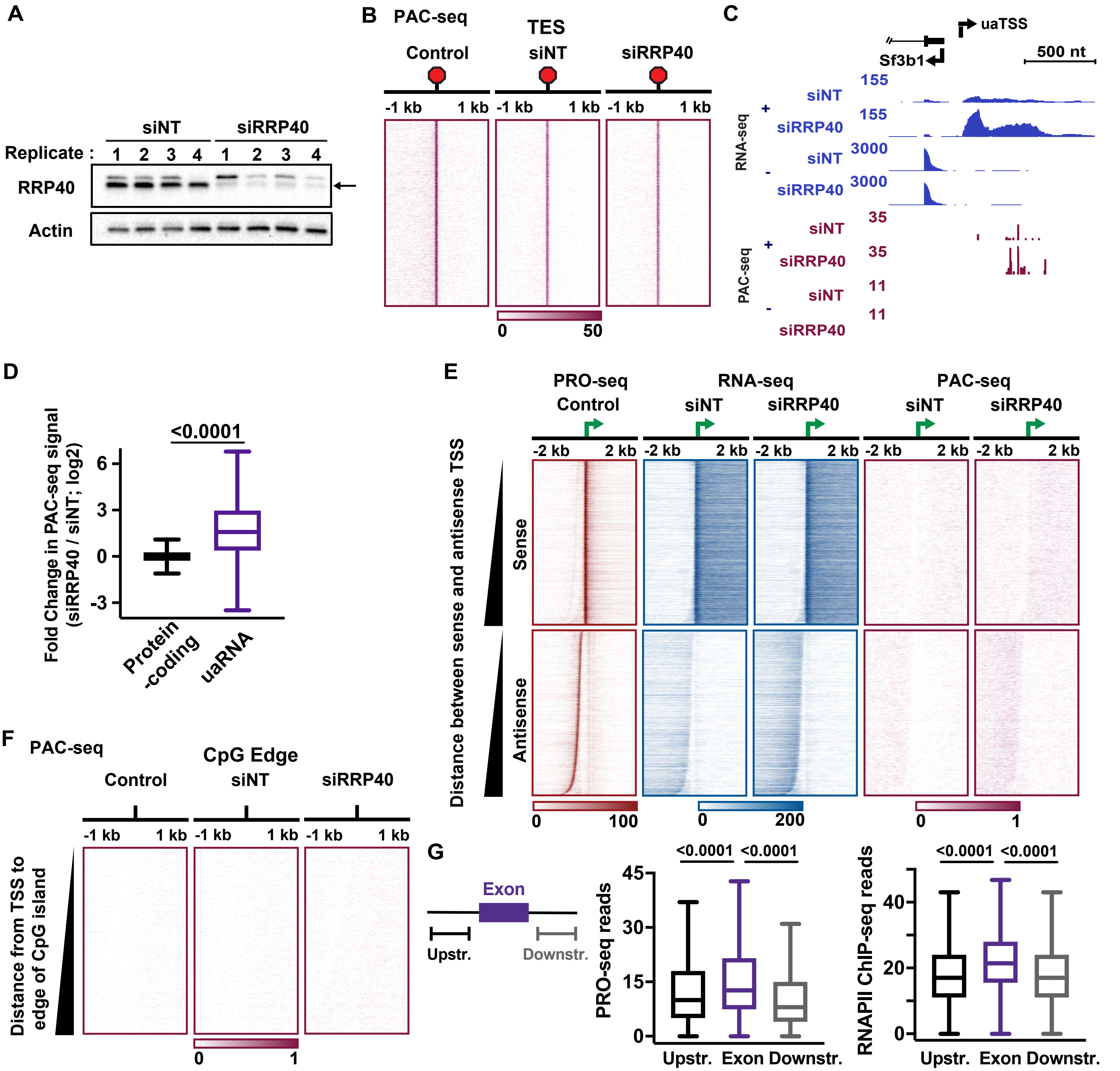

Figure S2. PAC-seq provides evidence for termination within uaRNAs, but not at the CpG edge. Related to Figure 1

(A) mESCs were transfected with siRNAs targeting exosome subunit RRP40 or a non-targeting control (NT) for 48 hrs. RRP40 (lower band, indicated with an arrow) and Actin protein levels were visualized via Western blot. Actin is shown as a loading control.

(B) Heatmap of PAC-seq signal from indicated conditions, aligned around TESs. Sense strand reads shown in 25 nt bins.

(C) Sense (+) and antisense (-) strand RNA-seq and PAC-seq signal is shown for the uaRNA upstream of Sf3b1, a defined target of exosome mediated degradation in mESCs.[S1]

(D) Sense strand PAC-seq reads were summed between the TSS and TES of active protein coding genes, and between the TSS to +2 kb at uaRNA TSSs. The fold changes between siRRP40 and siNT PAC-seq counts are shown per biotype. Box plots have a line at the median, and whiskers depict 1.5 times the interquartile range. P-values were calculated using the Mann-Whitney test.

(E) Heatmap representation of PRO-seq, RNA-seq and PAC-seq signal is shown for active protein coding genes with defined upstream antisense TSSs (N = 9,191). Heatmaps are centered on the sense TSS and ranked by increasing distance between the sense and antisense TSSs. PRO-seq and RNA-seq signal shown in 25 nt bins, and PAC-seq signal was summed in 100 nt bins.

(F) Heatmap representation of PAC-seq signal from indicated cells, aligned around the edge of the CpG island. Genes (N = 9,768, as in Figure 1C) are ranked by increasing distance from the TSS to the CpG island edge. Read counts were summed in 25 nt bins.

(G) The distribution of PRO-seq and RNAPII ChIP-seq read densities within internal exons is shown, as compared to density upstream (-150 to -50 nt) and downstream (+50 to +150 nt) of the exon. Box plots were generated as in D. P-values were calculated using the Wilcoxon test.

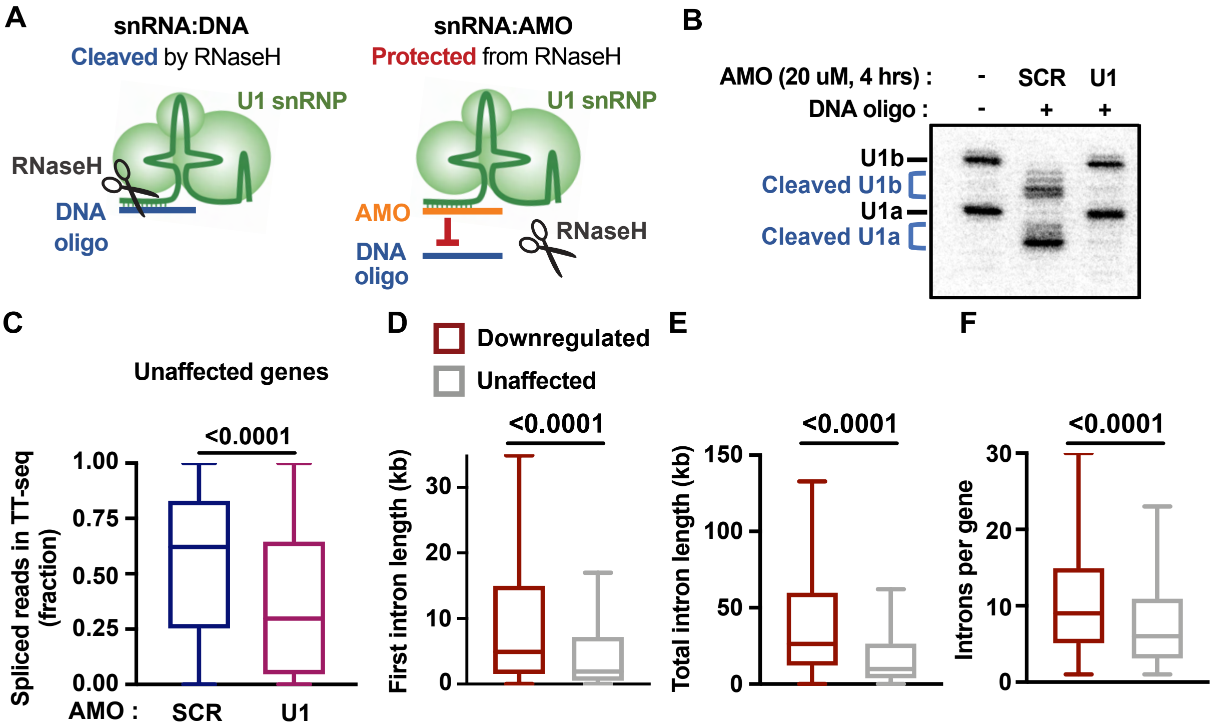

Figure S3. Antisense morpholino (AMO) targeting U1 preferentially disrupts expression of long genes with numerous, long introns. Related to Figure 2.

(A-B) An RNaseH protection assay (as described in[S2]) was performed after electroporating mESCs with scrambled (SCR) or U1 AMO to determine the concentration of U1 AMO needed to fully outcompete 5'SS binding. (A) Free U1 in solution will bind to a DNA oligo, rendering the snRNA:DNA duplex susceptible to cleavage by RNaseH. In contrast, snRNA:AMO duplexes are protected from RNaseH dependent cleavage. (B) Cells were electroporated with 20 uM of AMO. Four hours after electroporation, cells lysates were subjected to the RNaseH protection assay. U1 snRNA products were visualized by Northern blot. mESCs contain two dominant isoforms of the U1 snRNA (U1a and U1b).[S3] Full length and cleaved U1 snRNA products are labeled. Optimized conditions are shown to confirm full U1 protection.

(C) Splicing efficiencies in SCR and U1 AMO cells are shown for first introns at unaffected genes, as described in Figure 2D. Box plots have a line at the median, and whiskers depict 1.5 times the interquartile range. P-values were calculated using the Wilcoxon matched-pairs signed rank test.

(D-F) The distribution of (D) first intron and (E) total intron lengths and (F) the number of introns per gene is reported for downregulated and unaffected genes. Box plots are shown as in C. P-values were calculated using the Mann-Whitney test.

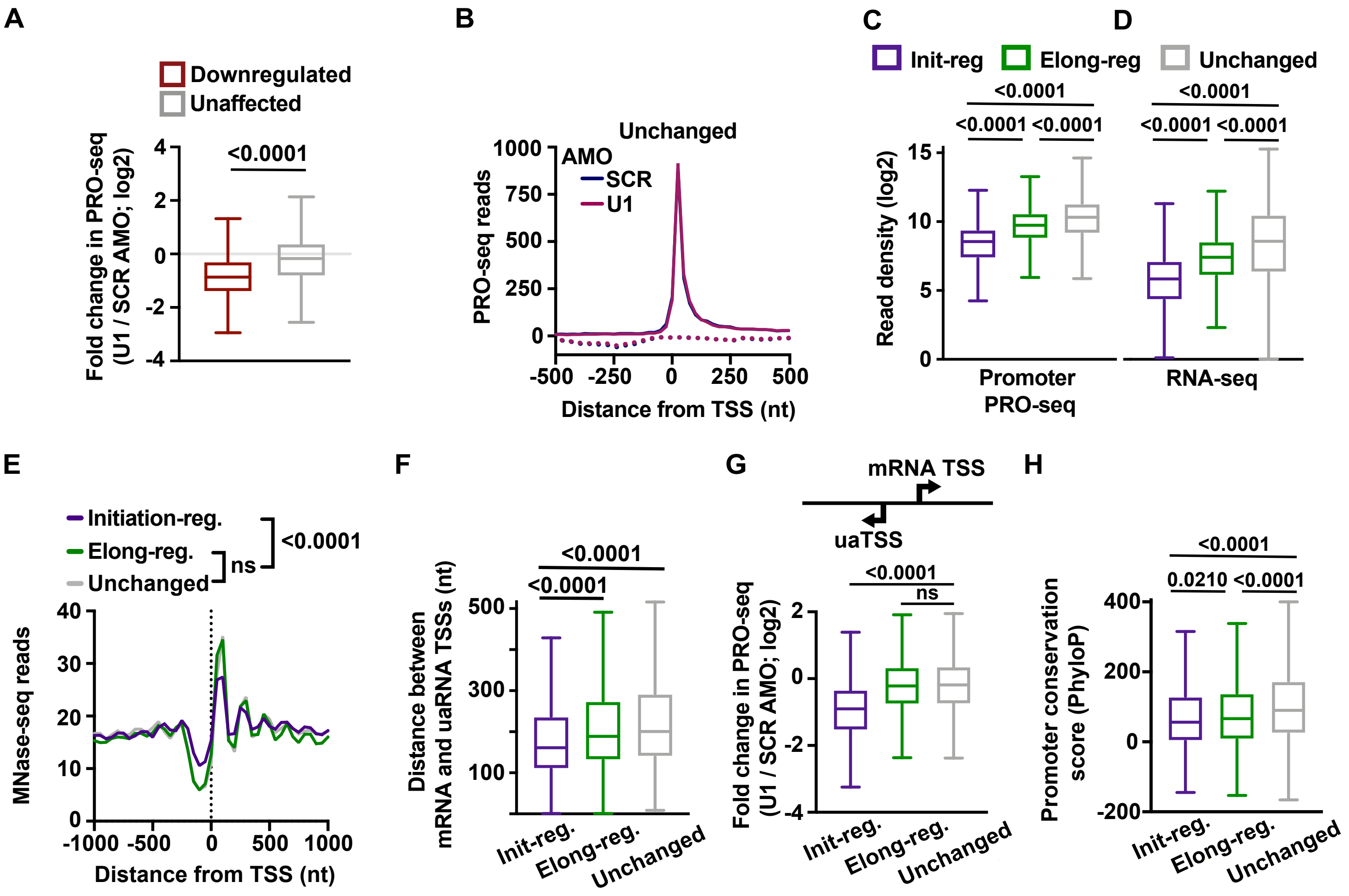

Figure S4: Genes at which U1 AMO affects initiation show weaker, smaller nucleosome depleted promoter regions and lower expression levels. Related to Figure 4

(A) PRO-seq signal was summed between the TSS and TES in SCR and U1 AMO conditions at downregulated and unaffected genes, as defined in Figure 2D. The distribution of fold changes in PRO-seq signal is shown as a box plot with a line at the median, and whiskers depict 1.5 times the interquartile range. P-values were calculated using the Mann-Whitney test. 81% of genes downregulated in TT-seq were also downregulated in PRO-seq, see STAR methods.

(B) Average PRO-seq signal in SCR and U1 AMO conditions is shown at genes classified as unchanged (N = 1,004). Read counts for sense (solid lines) and antisense (dotted lines) strands are shown in 25 nt bins, centered on the sense TSS. There is no significant difference in the PRO-seq reads in the promoter (TSS to +100 nt) of these genes upon U1 AMO treatment. P-values were calculated between SCR and U1 AMO conditions using the Wilcoxon test.

(C-D) The distribution of (C) PRO-seq promoter reads (TSS to +100 nt) and (D) RNA-seq read densities over full transcript models is shown at initiation-regulated (N = 1,398), elongation-regulated (N = 2,696) and unchanged (N = 1,004) genes under control conditions as a box plot, as in A. P-values were calculated using the Mann-Whitney test.

(E) Metagene plots of average MNase-seq signal in control mESCs is shown at each gene group. Read counts were summed in 50 nt bins, centered on the TSS. For statistical testing, MNase-seq signal was summed (from 500 nt upstream of TSS to TSS) and p-values were calculated using the Mann-Whitney test.

(F) The distribution of distances between sense and upstream antisense TSSs is shown for initiation-regulated (N = 902), elongation-regulated (N = 2,215) and unchanged (N = 790) genes at which an upstream antisense TSS could be identified. Box plots are shown as in A. P-values were calculated using the Mann-Whitney test.

(G) PRO-seq reads were summed on the antisense strand from -500 nt to the sense mRNA TSS. The fold change in antisense PRO-seq counts is shown per gene list as a box plot, as in A. P-values were calculated using the Mann-Whitney test.

(H) Promoter conservation was calculated as the total PhyloP conservation score in the region TSS +/- 100 nt. The distribution of promoter conservation scores is shown per gene list as a box plot, as in A. Mann-Whitney test used to calculate p-values.

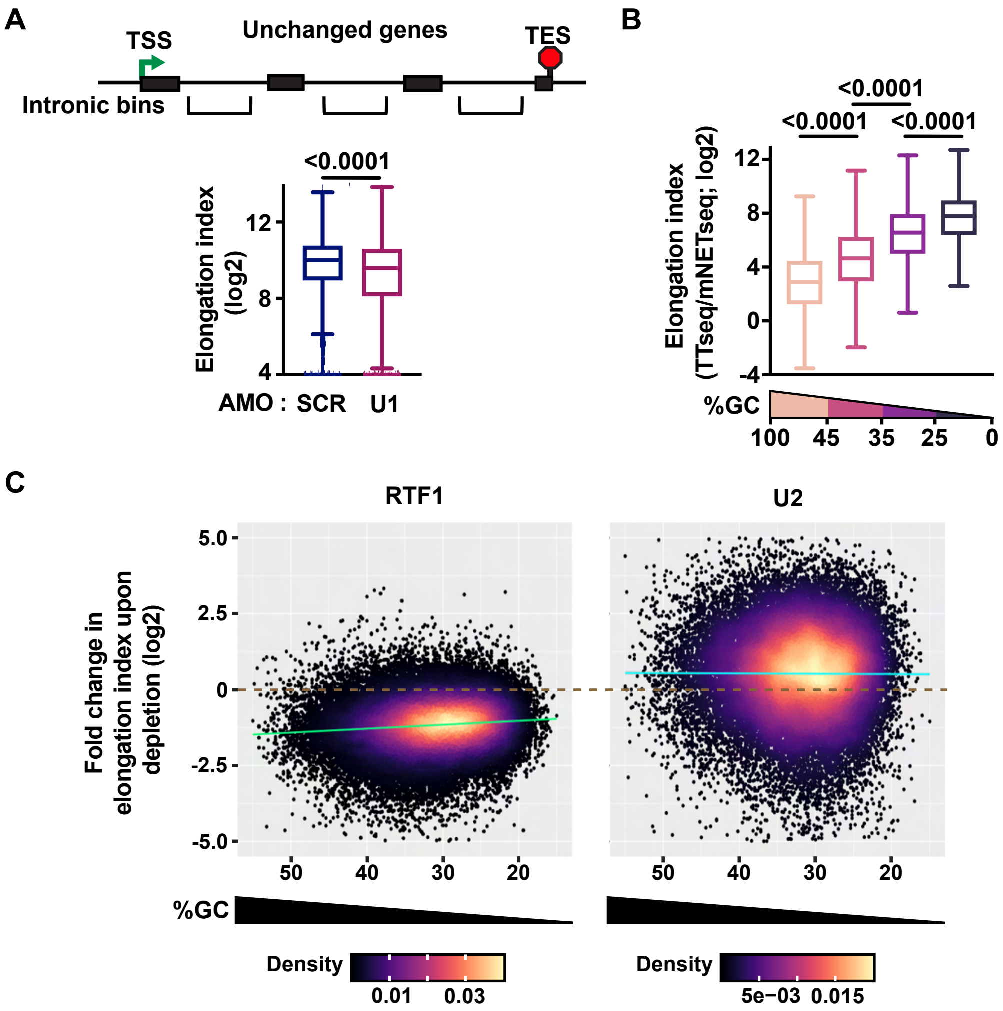

Figure S5: U1 stimulates elongation rate, particularly in AT-rich sequences. Related to Figure 5

(A) Elongation rate at unchanged genes in SCR and U1 AMO conditions is reported for 500 nt bins that contain only introns (N = 14,796). Box plots have a line at the median and whiskers depicting 1.5 times the interquartile range. P-values were calculated using the Wilcoxon test. We note that intron retention, which would elevate the TT-seq signal within introns, would be anticipated to increase, rather than decrease the Elongation index calculated using this method.

(B-C) Active protein coding genes in K562 cells were divided into 500 nt bins. **(B)** Elongation index in control K562 cells was calculated for each %GC group (highest to lowest %GC: N = 10360, 29586, 26717, 5750). Box plots were generated as in A and p-values were calculated using the Mann-Whitney test. **(C)** Fold change in elongation index was calculated per bin between control conditions and cells wherein RTF1 (left) or U2 (right) was inhibited. Bins within unchanged genes (defined for each dataset) were retained (RTF1, N = 65,953 ; U2, N = 32,078). The relationship between the fold change in elongation index upon factor depletion and %GC is shown as a density scatter plot. Scatter plots include a solid line representing the linear fit trend line and a dotted line indicating no change (i.e, a log2 fold change of 0).

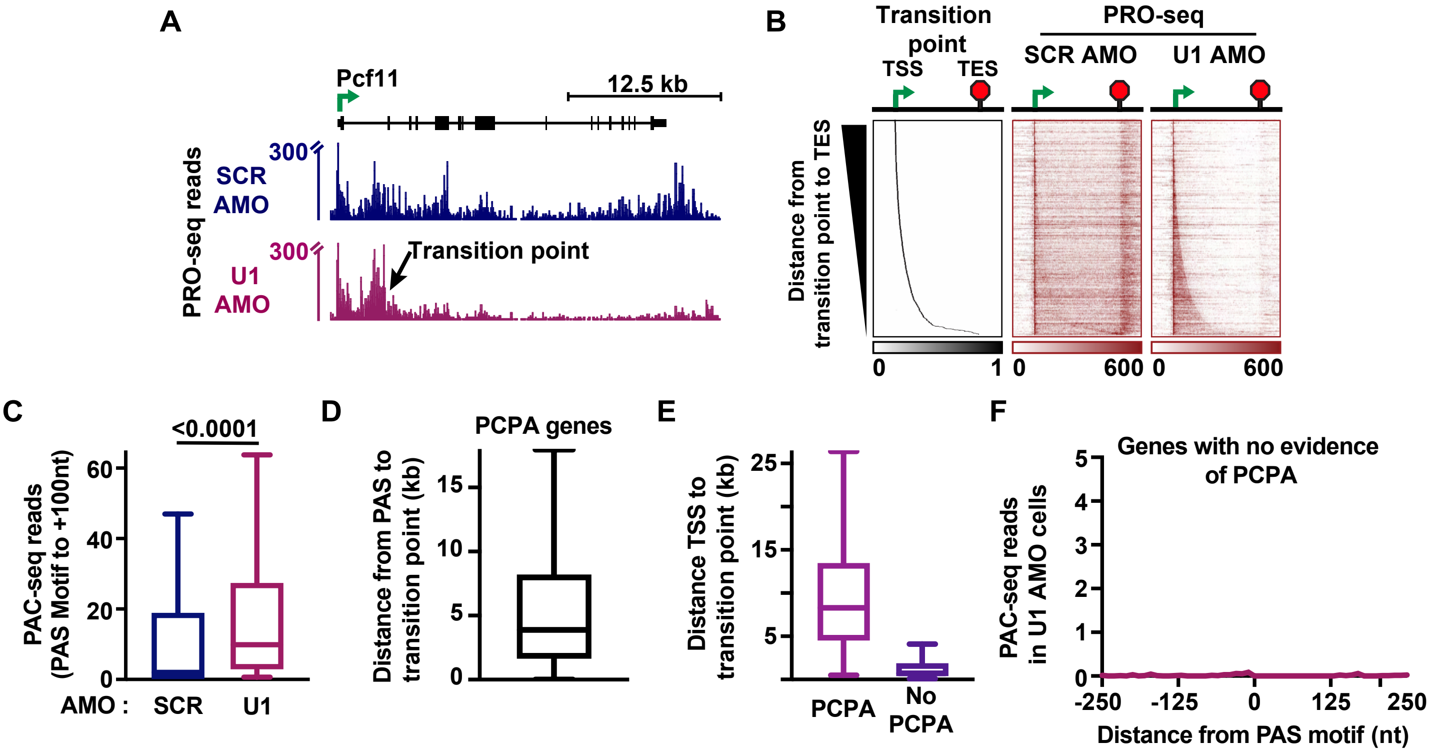

Figure S6: Premature cleavage and polyadenylation (PCPA) is more common in the absence of U1. Related to Figure 6

(A) Sense strand PRO-seq signal is shown at an example elongation-regulated gene with a defined transition point. PRO-seq signal is shown in 25 nt bins and the y-axis is truncated to highlight gene body signal.

(B) Heatmaps of the indicated data are shown for genes with a TP (N = 1,162). The region shown extends from 2 kb upstream of the TSS (arrow) to 2 kb downstream of the TES (stop sign). The region between the TSS and TES was scaled by length into 400 bins. Genes are ranked by decreasing distance between the TP and TES.

(C) PAC-seq reads were summed from the actionable PAS motif to 100 nt downstream under SCR and U1 AMO conditions, for genes with evidence of PCPA (N = 541). Box plots have a line at the median, and whiskers depict 1.5 times the interquartile range. P-values were calculated using the Wilcoxon test.

(D) The distribution of distances between the actionable PAS motif and the TP is shown at genes with evidence of PCPA (N = 541) as a box plot (as in C).

(E) The distribution of distances between the TSS and the TP is shown for genes that undergo PCPA (N = 541) or do not show evidence of PCPA (N = 621; right). Box plots were generated as in C

(F) Metagene plot of average PAC-seq signal in U1 AMO conditions at PAS motifs found between the TSS and the TP at genes lacking PCPA (N = 621). Read counts were summed in 10 nt bins.

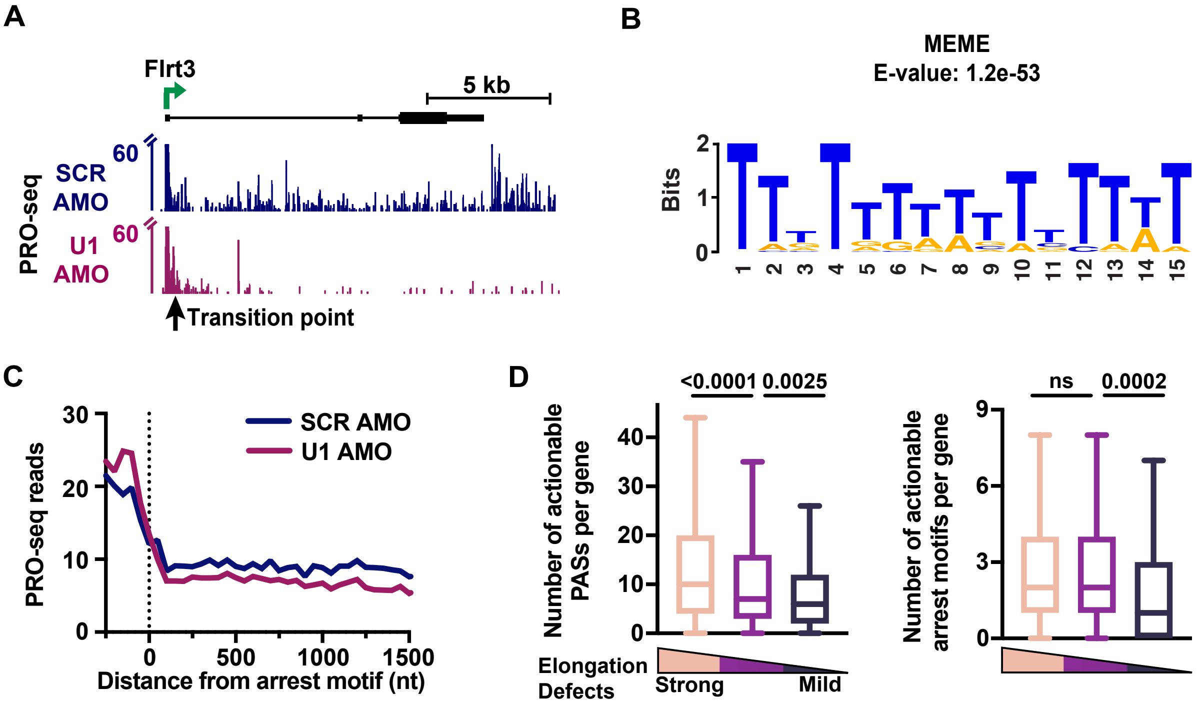

Figure S7: RNAPII is more prone to transcriptional arrest in the absence of U1. Related to Figure 7

(A) PRO-seq signal at a gene with a TP that lacks evidence of PCPA. Reads are shown in 25 nt bins and the y-axis is truncated to highlight gene body signal.

(B) Top enriched motif within a 100 nt window of the TP at genes without PCPA (N = 621), identified by MEME.

(C) Metagene plot of PRO-seq signal in SCR and U1 AMO conditions at arrest motifs near TPs of genes without PCPA. To avoid biases from promoter proximal RNAPII signal, only genes where the arrest motif is $\geq$ 400 nt from the TSS are shown (N= 395). Reads were aligned to the final nucleotide of the arrest motif and summed in 50 nt bins.

(D) Genes were separated into three groups based on the fold change in elongation index after U1 inhibition using the late gene body window (from TSS +2250 to TES). Groupings were: Strong log2 Fold Change (FC) < -1, N = 2264; Medium -1 $\leq$ log2 FC < -0.5, N = 568; Mild -0.5 $\leq$ log2 FC $\leq$ 0, N = 367). Box plots depict the number of actionable PAS (left) and arrest (right) motifs per gene in U1 AMO treated cells. P-values from the Mann-Whitney test.

Supplementary Tables

Table S1. Oligonucleotide sequences used in this study. Related to STAR Methods.

Table S1. Oligonucleotide sequences used in this study. Related to STAR Methods.

Antisense Morpholino (AMO)	Sequence
SCR AMO	CCTCTTACCTCAGTTACAATTTATA
U1 AMO	GGTATCTCCCCTGCCAGGTAAGTAT

Northern Blot Probe	Sequence
U1 snRNA	CAAATTATGCAGTCGAGTTTCCCACATTTG

Supplemental References

[S1] Chiu, A.C., Suzuki, H.I., Wu, X., Mahat, D.B., Kriz, A.J., and Sharp, P.A. (2018). Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. Mol. Cell *69*. 10.1016/j.molcel.2018.01.006.

[S2] Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664–668. 10.1038/nature09479.

[S3] Lund, E., Kahan, B., and Dahlberg, J.E. (1985). Differential Control of U1 Small Nuclear RNA Expression During Mouse Development. Science (80-.). *229*, 1271–1274. 10.1126/SCIENCE.2412294.