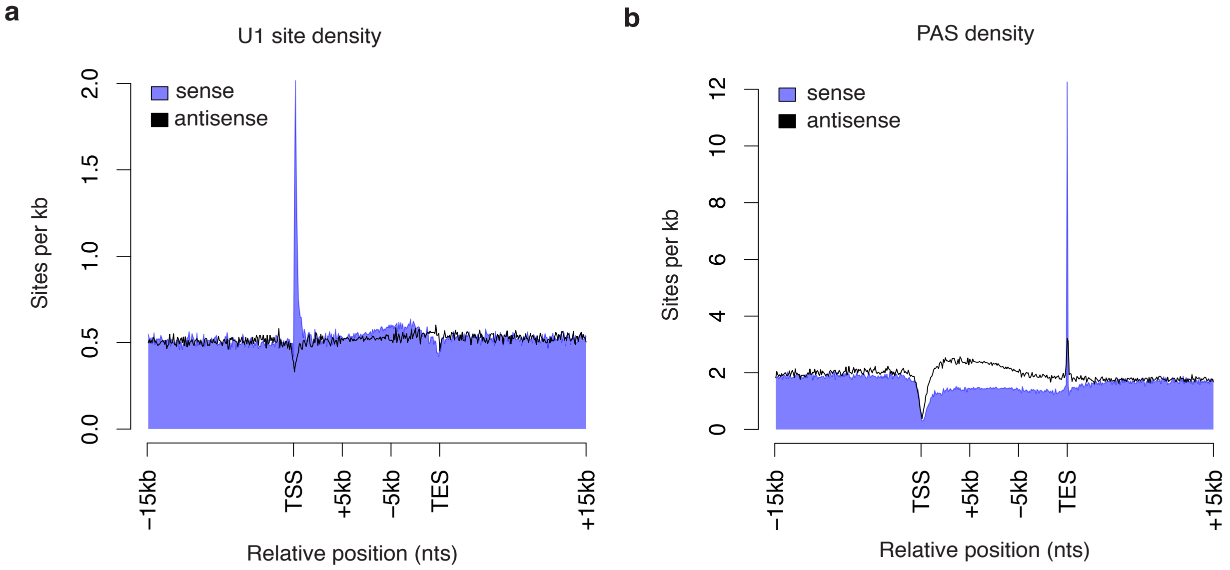

Supplementary Figure 9. **Density of U1 and PAS signals at coding genes and intergenic regions.** The density of strong U1 sites (a) and AAUAAA polyadenylation signals (b) in sites per kb for protein-coding genes longer than 15 kb and flanking 15 kb of intergenic sequences. U1 or PAS signals located on sense or antisense regions are depicted in purple and black, respectively. In addition to the strong U1 enrichment in the proximal sense direction of the gene, we observe a modest increase in the frequency of strong U1 signals internal to genes. We also observed a strong strand bias of PAS in coding transcription units, both exon and intron sequences, as compared to intergenic regions. Specifically, PAS are depleted on the sense strand when compared to the antisense strand throughout coding genes prior to the TES. In absolute terms, the genome background has a relatively high density of PAS (~ 2 sites per kb on average) but lower density of strong U1 sites (~0.5 sites per kb on average). Together, the observed distributional patterns support a general model of a U1-PAS axis favoring elongation to produce long transcripts such as precursors to mRNA but limiting transcription from antisense and intergenic regions.