

Figure S1

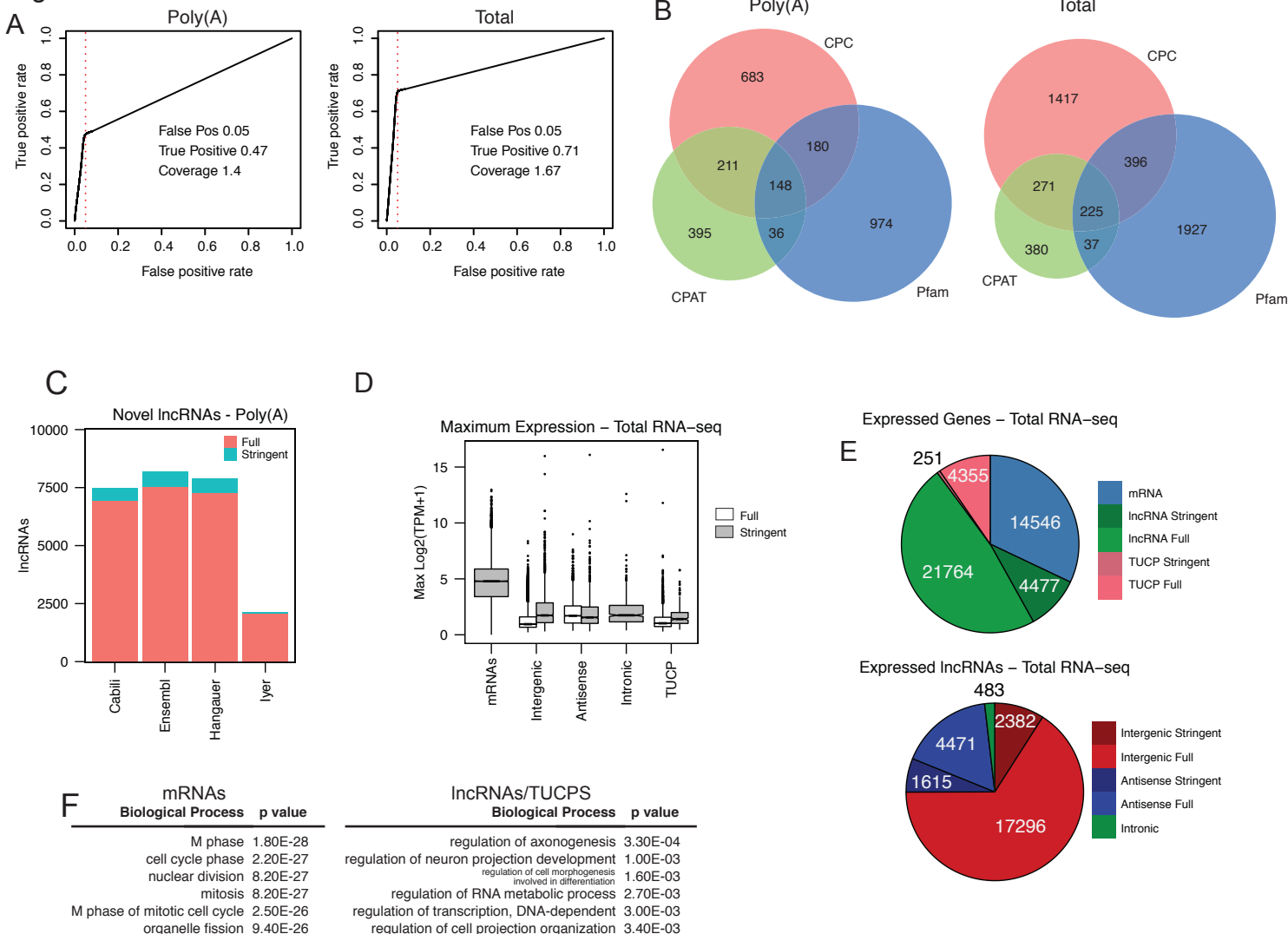


Figure S1. Computational pipeline of lncRNA identification

A) ROC analysis of coverage thresholds required to fully annotate known transcripts in poly(A) and total RNA-seq data. Coverage thresholds were selected for which false positive rates were 0.05, corresponding to the optimal balance of false positive and true positive rates. B) Venn diagrams of novel transcripts classified as protein coding by CPC, CPAT, and Pfam search. The union of the transcripts were annotated TUCPs. C) Number of novel poly(A) lncRNAs expressed in the Full (red) and Stringent (blue) references compared to Cabili et. al. 2013, Ensembl 75/Gencode 19, Hangauer et. al. 2013, and Iyer et. al. 2015 (MiTranscriptome). D) Maximum expression levels of transcripts described in the Full (white) and Stringent (gray) references derived total RNA-seq across all samples. TPM, Transcripts per Million E) Numbers of expressed mRNAs, lncRNAs, and TUCPs during neocortex development in bulk tissues, according to total RNA-seq libraries. F) Top gene ontology terms for differentially expressed mRNAs (left) and mRNA neighbors of lncRNAs/TUCPs (right).