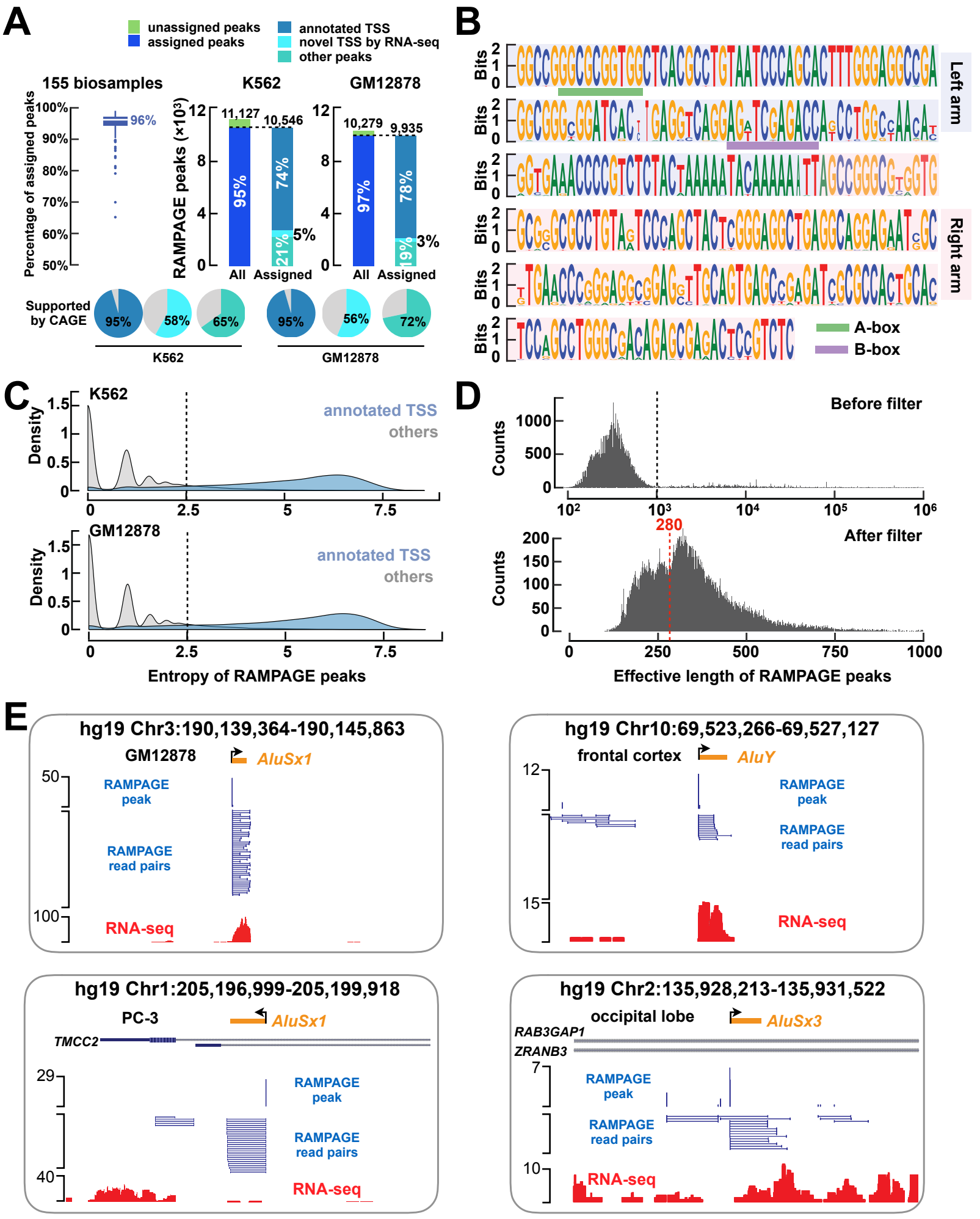
Zhang et al., Supplemental Figure S2



Supplemental Figure S2. Identification of expressed *Alu* elements using RAMPAGE data.

A. RAMPAGE data enable the assignment of TSSs to genes. Aided by the paired-end RAMPAGE reads, roughly 96% of RAMPAGE peaks could be assigned to GENCODE-annotated genes or transcripts that we *de novo* assembled using RNA-seq data in the same cell line (left boxplot), and 79–81% of the assigned peaks were located within ± 50 bp of GENCODE-annotated or RNA-seq assembled TSSs (right bar in right panels). More than 50% of assigned peaks could also be annotated by CAGE, with CAGE peaks (Abugessaisa et al. 2016) within their ± 50 bp regions (pie charts).

B. The sequence logo of *Alu* elements based on the multiple-sequence alignment of 37 main subfamilies of *Alu* elements (Price et al. 2004). A typical *Alu* element consists of a left arm (shaded blue) and a right arm (shaded red), with A-box (green bar) and B-box (purple bar) elements located in the left arm functioning as Pol III promoter recognition elements.

C. Entropy distribution of the reads that constitute RAMPAGE peaks overlapping annotated TSSs and the other RAMPAGE peaks. Note that an entropy cutoff of 2.5 effectively distinguishes TSS peaks from non-TSS peaks.

D. Histogram of the effective lengths of RAMPAGE peaks for *Alu* elements before and after filtering out peaks with effective lengths longer than 1 k nucleotides.

E. Example *Alu* elements expressed in GM12878, PC-3, occipital lobe, and frontal cortex. RAMPAGE peaks mark the TSSs of expressed *Alu* elements, with RAMPAGE read pairs linking the TSS to downstream positions. RNA-seq data further confirm the expression of *Alu* elements in the corresponding cells or tissues.