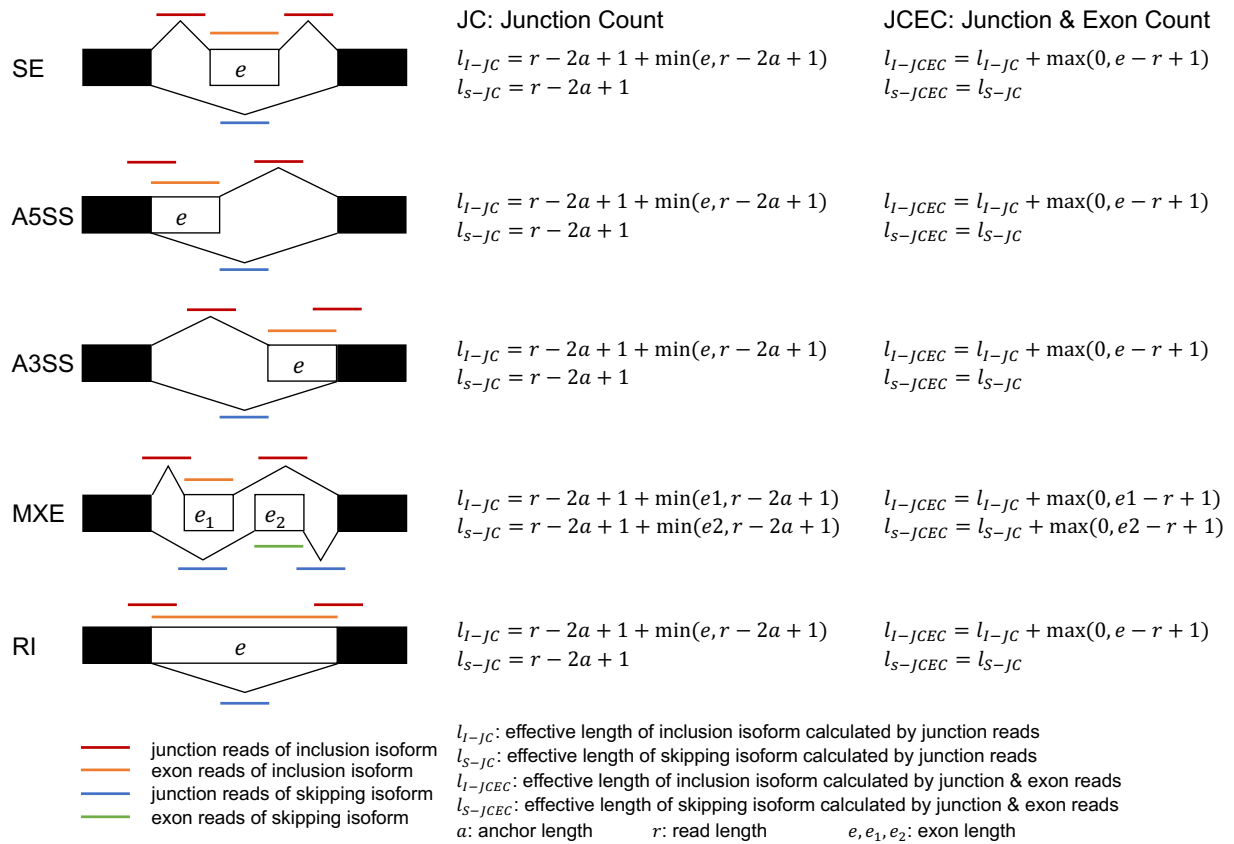
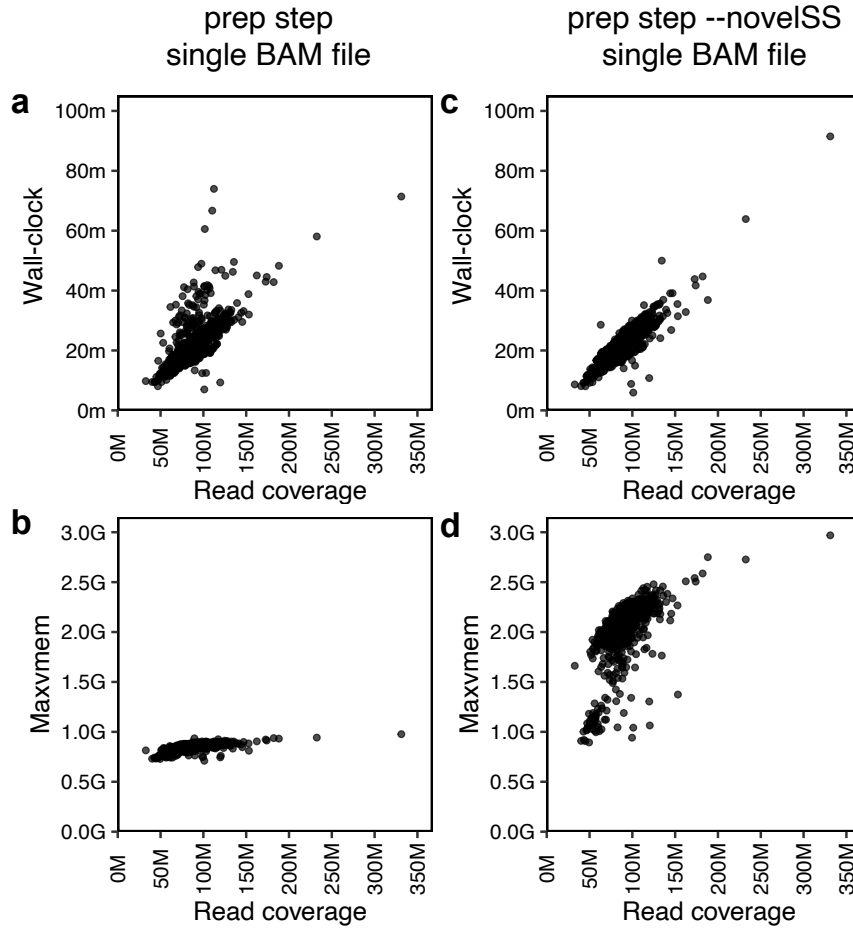


rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data

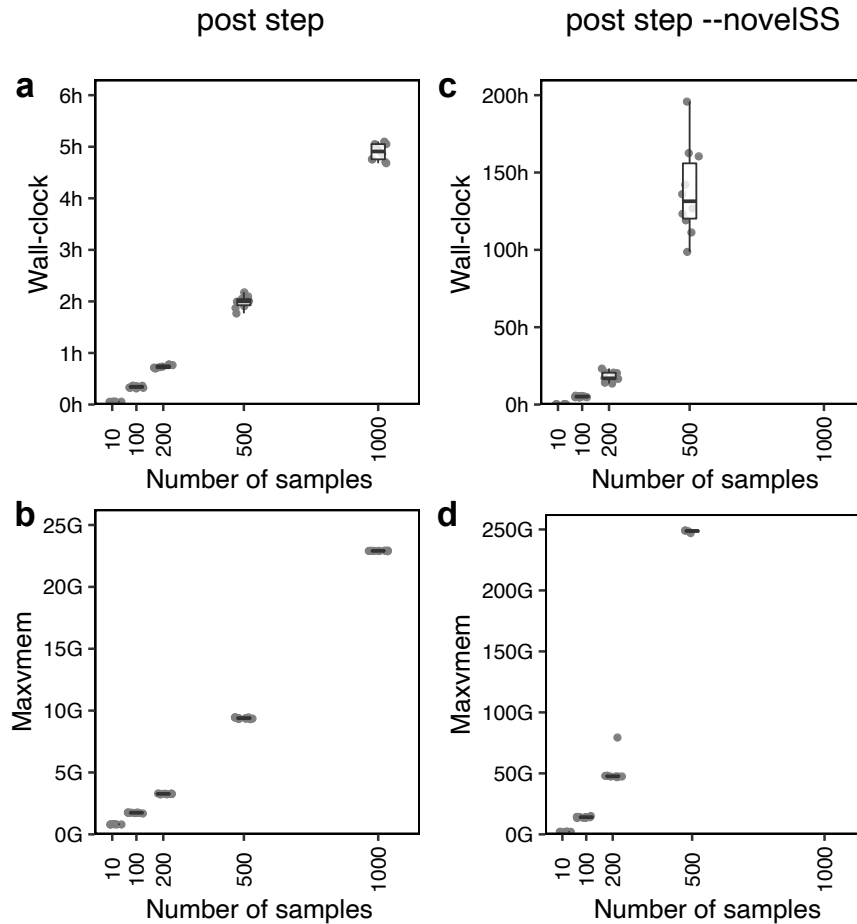
In the format provided by the authors and unedited



Supplementary Figure 1. Schematic illustration of supporting RNA-seq reads and effective lengths of different alternative splicing patterns. The alternative splicing events of skipped exon (SE), alternative 5' splice sites (A5SS), alternative 3' splice sites (A3SS), and retained intron (RI) have two junctions for the inclusion isoform and one junction for the skipping isoform. The alternative splicing events of mutually exclusive exons (MXE) have two junctions for the inclusion isoform of the first exon and two junctions for the skipping isoform of the first exon (i.e., the inclusion isoform of the second exon). The exon reads are RNA-seq reads mapped to the genomic regions of the target exons. rMATS-turbo uses either the junction counts only (JC) or the junction counts plus the exon counts (JCEC) to calculate PSI values, normalized by the effective lengths of the inclusion and skipping isoforms. Modified from Figure S1 of Shen et al., 2014¹⁸.



Supplementary Figure 2. Runtime and memory usage for the prep step of rMATS-turbo running on the CCLE dataset, with one input file and different running options. (a, b) Runtime (a) and memory usage (b) for the prep step of rMATS-turbo with the '--novelSS' option turned off. **(c, d)** Runtime (c) and memory usage (d) for the prep step of rMATS-turbo with the '--novelSS' option turned on. Each dot represents one input file, with varying read coverage. Maxvmem is the maximum amount of RAM used by a job when running.



Supplementary Figure 3. Runtime and memory usage for the post step of rMATS-turbo running on the CCLE dataset, with different numbers of input files and different running options. (a, b) Runtime (a) and memory usage (b) for the post step of rMATS-turbo with the ‘--novelSS’ option turned off, based on the number of selected input files summarized in the post step, repeated 10 times. **(c, d)** Runtime (c) and memory usage (d) for the post step of rMATS-turbo with the ‘--novelSS’ option turned on, based on the number of input files summarized in the post step. We did not test running the post step on 1,000 samples with the ‘--novelSS’ option turned on, due to the expected large runtime and memory usage. Maxvmem is the maximum amount of RAM used by a job when running. For each boxplot, the top and bottom of the box represent the third and first quartiles, respectively. The band in the middle of the box represents the median. The whiskers of each boxplot extend to the most extreme data points within 1.5 times the interquartile range from each box.