



Figure R1. Characteristics of simulated RNA-seq data. (A) Precision of APA site prediction for 1000 genes with two isoforms. The read coverage per gene of the simulated data (total 1,113,000 reads) is the same as that of the human MAQC brain data (1113 reads per gene). (B) Prediction precision of differential APA site usage for 1000 genes in (A). (C) The ratio of read coverage between downstream 100 bp and upstream 100 bp around predicted APA sites identified from human MAQC brain data. A smaller ratio means a higher difference; this is because shorter isoform is embedded in the longer one, so that the upstream part is of higher expression level than the downstream part. Here we used the predicted poly(A) sites rather than the reference ones is because that the true poly(A) sites in real data are unknown while the whole annotated sites are far more than the sites in the investigated real data. We used two tools, APAtrop and DaPars, to examine the difference of read coverage change between simulated data and real data, as both tools predict APA sites based on the read coverage change. (D) As in (C) but the simulated dataset was used.