

Additional file 1: Supplementary figures

1

2

3

Supplementary figures

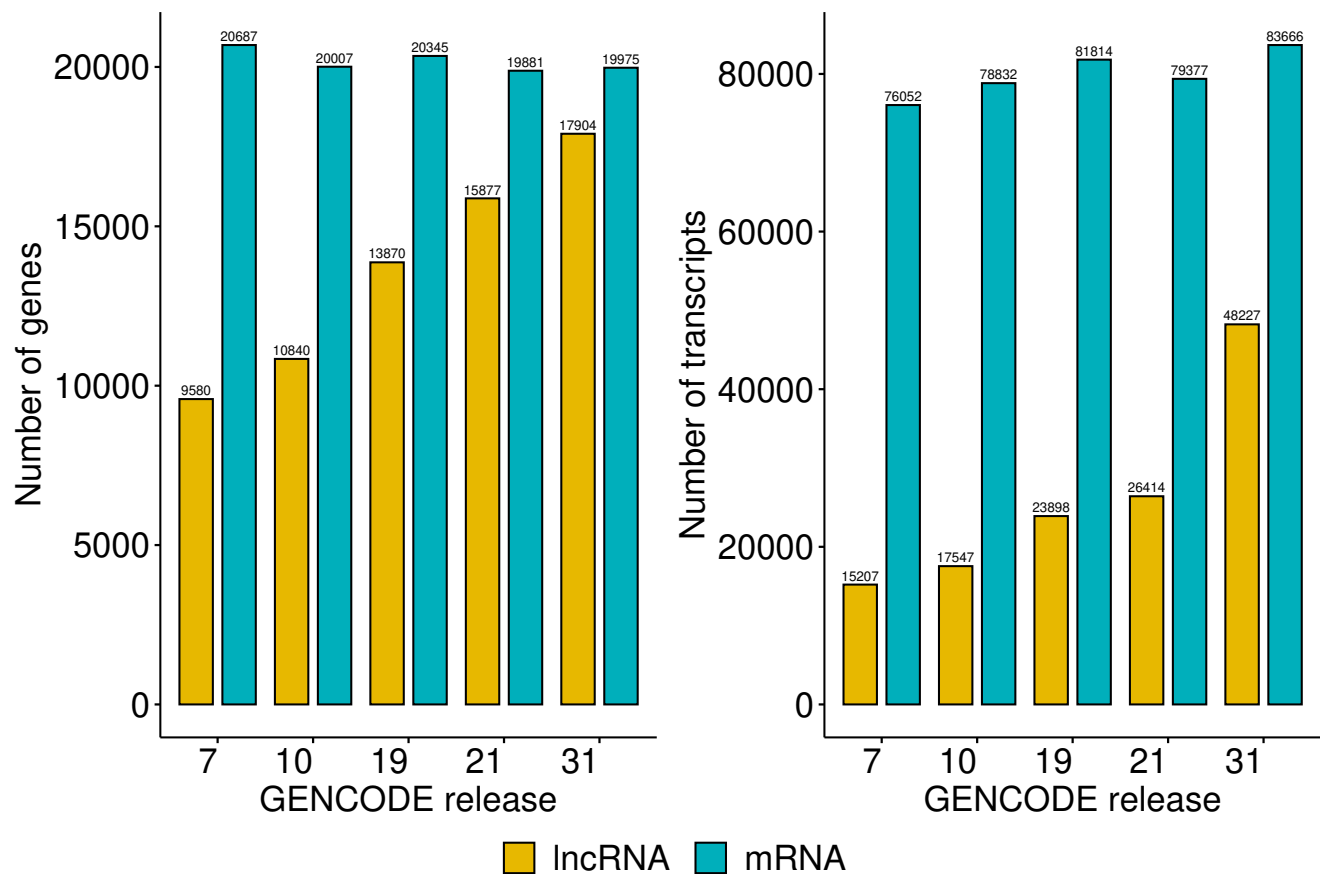


Figure S1. Growth of GENCODE gene annotation. The left and right panels show the number of gene loci and transcripts annotated in GENCODE gene annotation, respectively. Only the past major releases and the latest release 31 at the time of writing are included. Colors indicate RNA types.

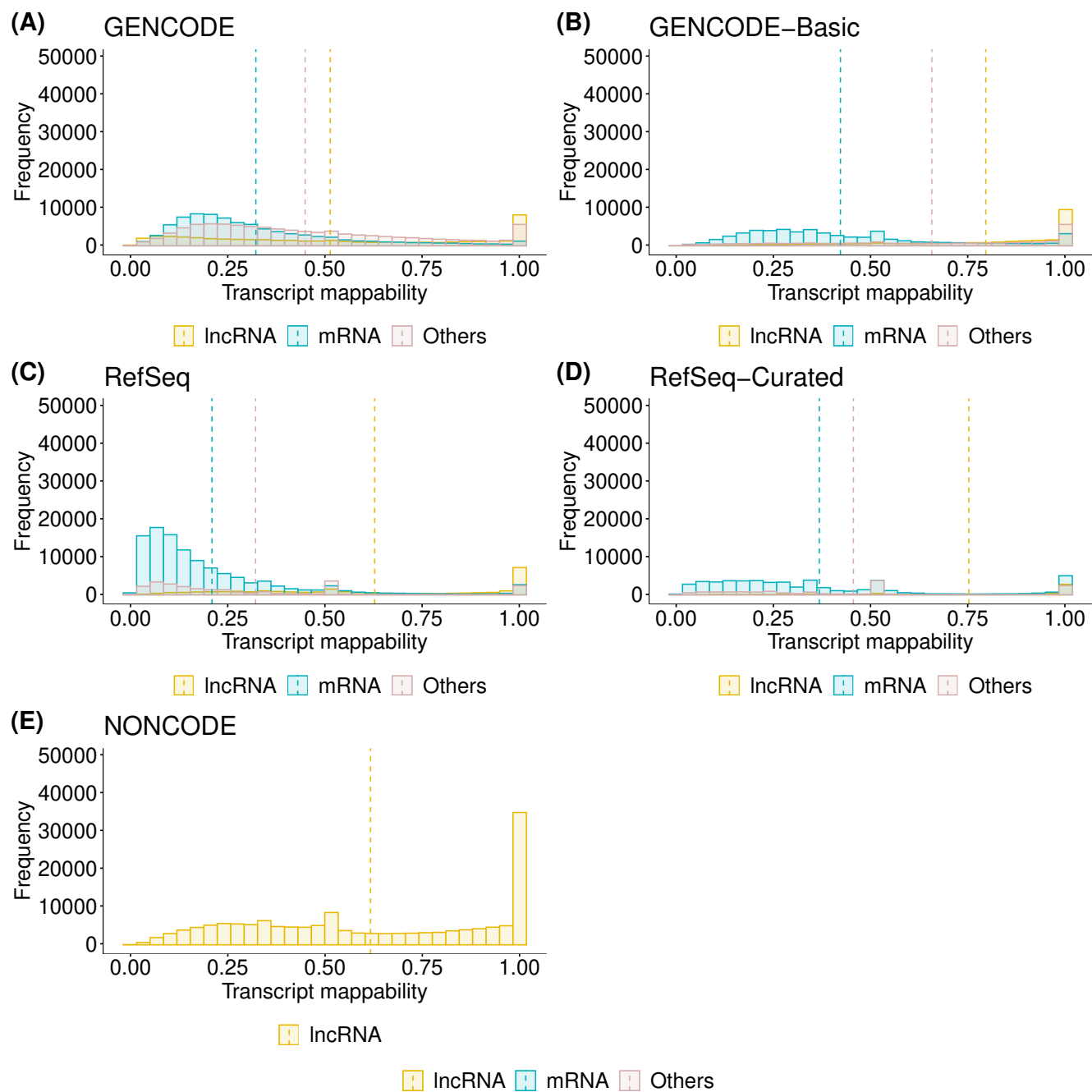


Figure S2. Complexity of gene models was significantly different among human gene annotations. We calculated mappability using 50 bases length reads. (A)–(E) show the distribution of transcript mappability for GENCODE, GENCODE-Basic (a subset of GENCODE), RefSeq, RefSeq-Curated (a subset of RefSeq), and NONCODE, respectively. Colored bars indicate the frequency of mRNAs (blue), lncRNAs (yellow), and other biotypes (red). Dotted vertical lines indicate the average mappability for each gene annotation.

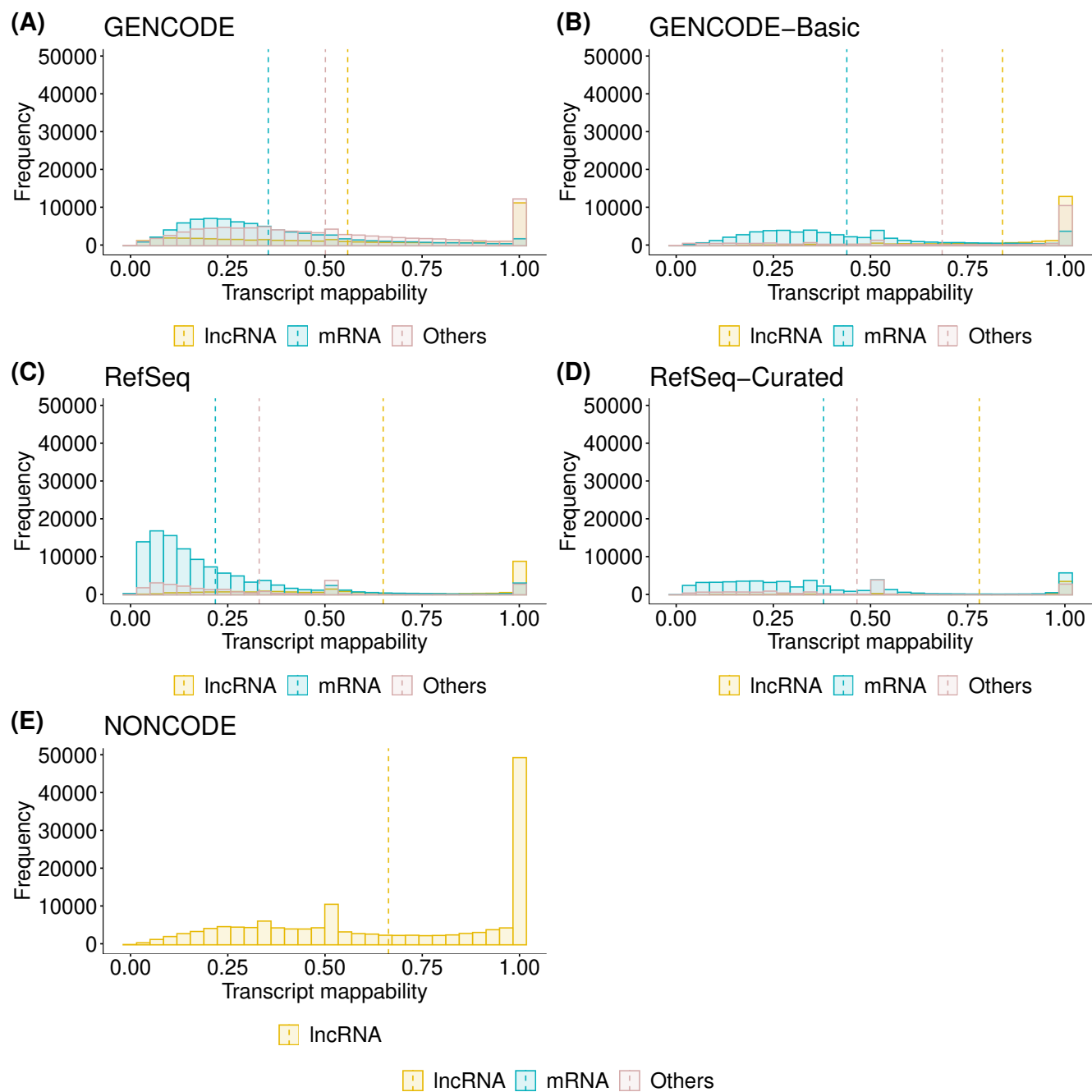


Figure S3. Complexity of gene models was significantly different among human gene annotations. We calculated mappability using 150 bases length reads. (A)–(E) show the distribution of transcript mappability for GENCODE, GENCODE-Basic (a subset of GENCODE), RefSeq, RefSeq-Curated (a subset of RefSeq), and NONCODE, respectively. Colored bars indicate the frequency of mRNAs (blue), lncRNAs (yellow), and other biotypes (red). Dotted vertical lines indicate the average mappability for each gene annotation.

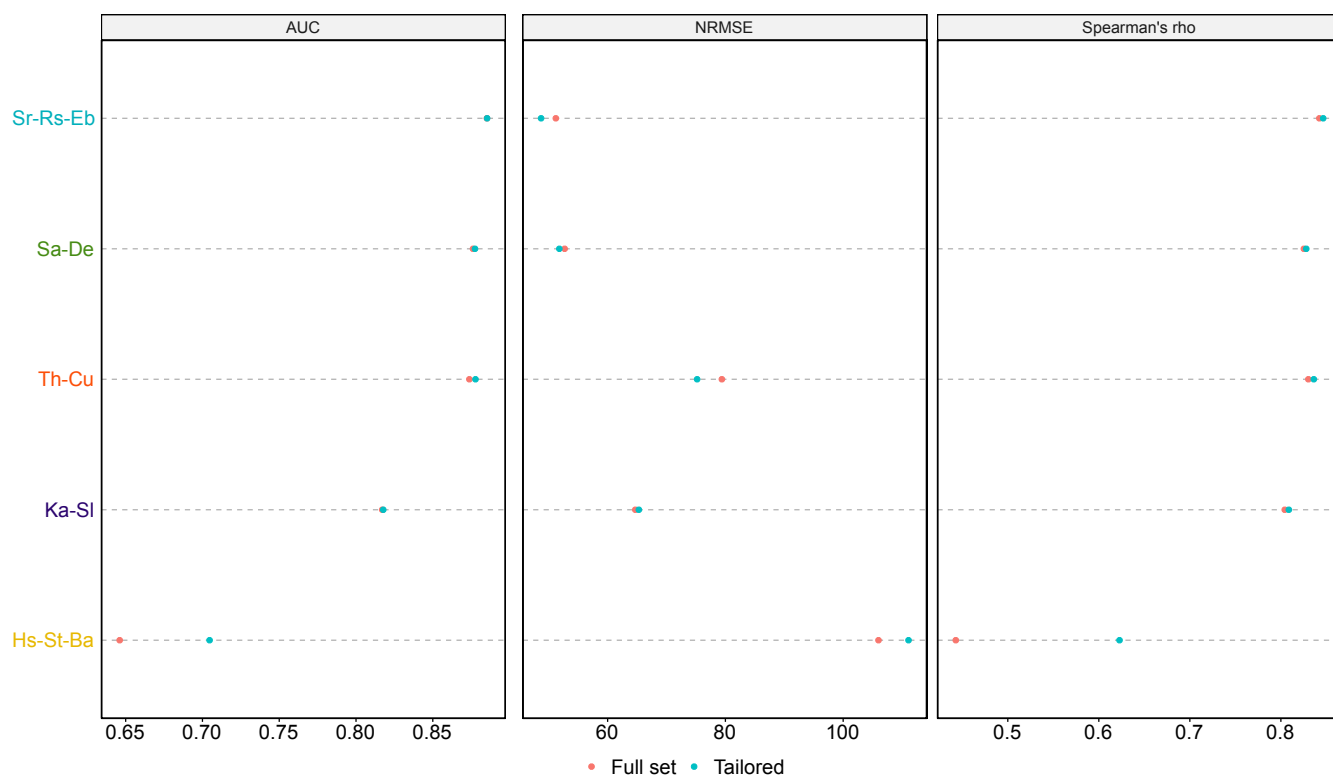


Figure S4. Comparison of metrics with full set and tailored annotation on DE step. Three metrics evaluated analysis results with full set gene annotations and tailored gene annotations: AUC, RMSE, and Spearman's rho. These metrics were calculated for all transcripts expressed. Colors indicate the annotation used. Abbreviations indicate combinations of the following tools: Hs-St-Ba; HISAT-StringTie-Ballgown, Sa-De; Salmon-DESeq2, Ka-Sl; Kallisto-Sleuth, Th-Cu; Tophat2-Cufflinks.