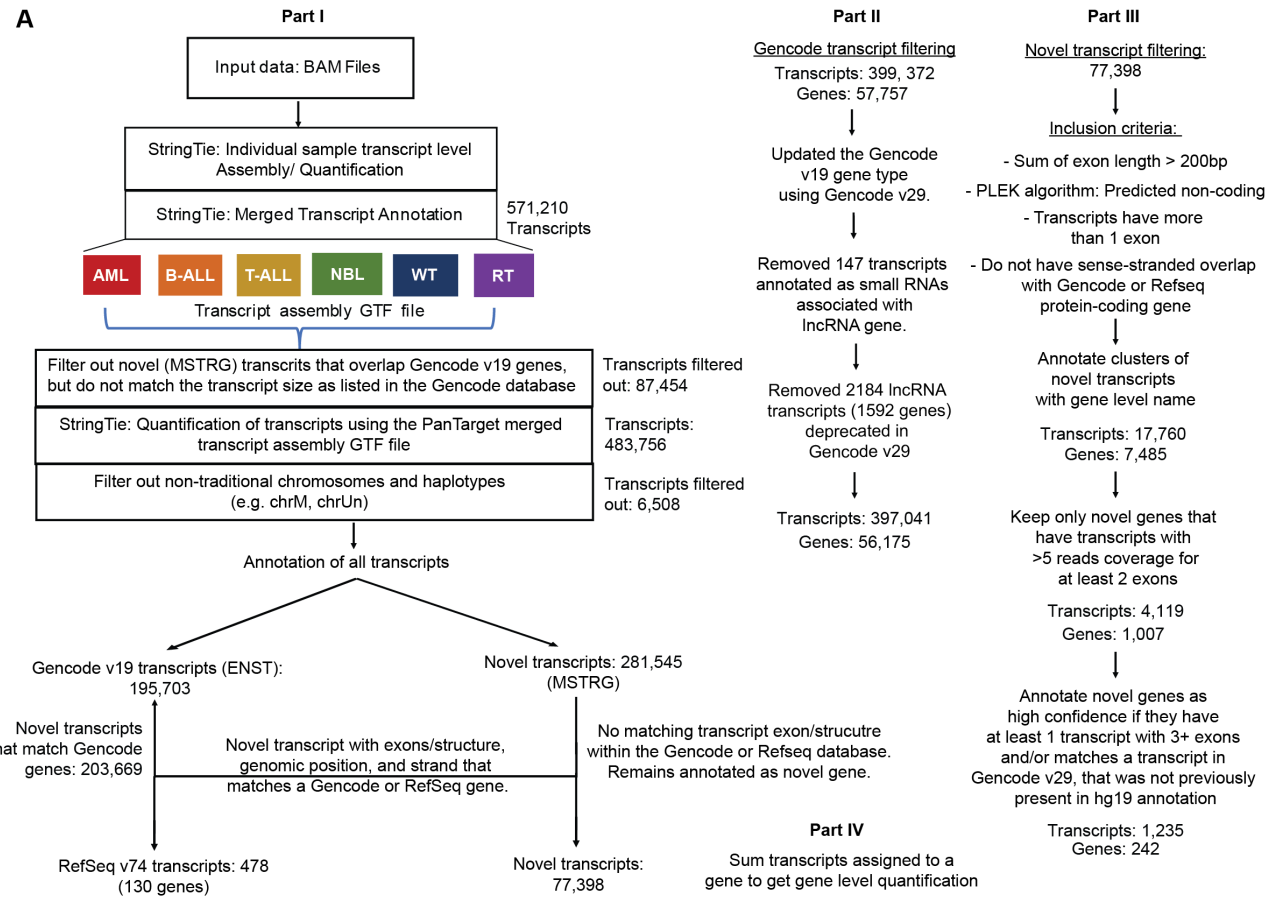
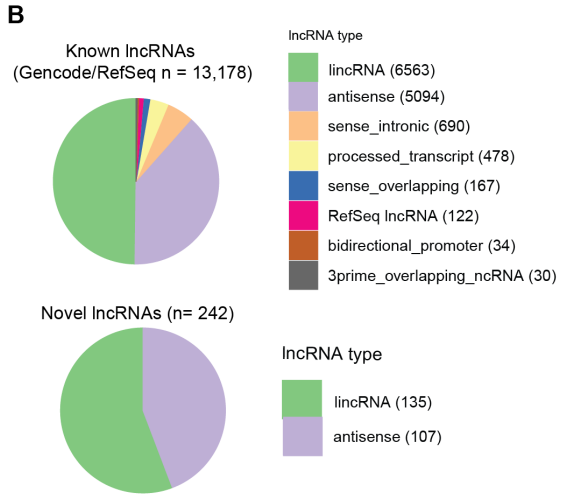


	Gencode	RefSeq	Novel (Non-coding)
# Transcripts	397,041	478	1,235
# Genes	56,175	130	242
# Protein Coding Gene	19,888	8	N/A
# lncRNAs	13,056	122	242



Supplementary Fig. S1: Workflow for RNA-seq gene mapping and quantification. (A) Workflow diagram showing how samples were processed using the StringTie program, which performs genes mapping and quantification. Custom scripts were then used for the following: Part I: Identified gene transcripts were assigned as a Gencode or Refseq gene, or as a novel gene. Part II: Gencode transcripts were further filtered including

based on gene type. Part III: Novel genes were further filtered based on non-coding potential, length, number of transcripts and exons, and read coverage per exon. Part IV: Gene-level expression was considered as the sum of associated transcript expression. Table summarizing the number of transcripts and genes per gene type post filtering used in this study. **(B)** High confidence selected novel lncRNA genes are primarily intergenic or antisense. Sense-overlapping novel lncRNAs were considered low confidence and not-considered in this analysis. The majority of known lncRNAs in the Gencode database are either intergenic or antisense.