

### **Description of additional supplementary file**

File name: Supplementary Data 1

Description: RNA embedding similarities of human–mouse homologous genes and their CDS DNA sequence similarities.

File name: Supplementary Data 2

Description: Human–mouse homolog and random gene pairs with RNA embedding similarity, DNA sequence similarity, GC content, transcript metrics, and phenotype/disease associations.

File name: Supplementary Data 3

Description: DNA characteristics of 1 kb regions upstream of transcription start sites (TSS) for homolog and random gene pairs.

File name: Supplementary Data 4

Description: DNA characteristics of 1 kb regions downstream of transcription termination sites (TTS) for homolog and random gene pairs.

File name: Supplementary Data 5

Description: Gene set enrichment analysis results for four groups of homologous genes classified by RNA embedding and DNA sequence similarity.