

## Extended transcriptome analysis reveals genome-wide lncRNA-mediated epigenetic dysregulation in colorectal cancer

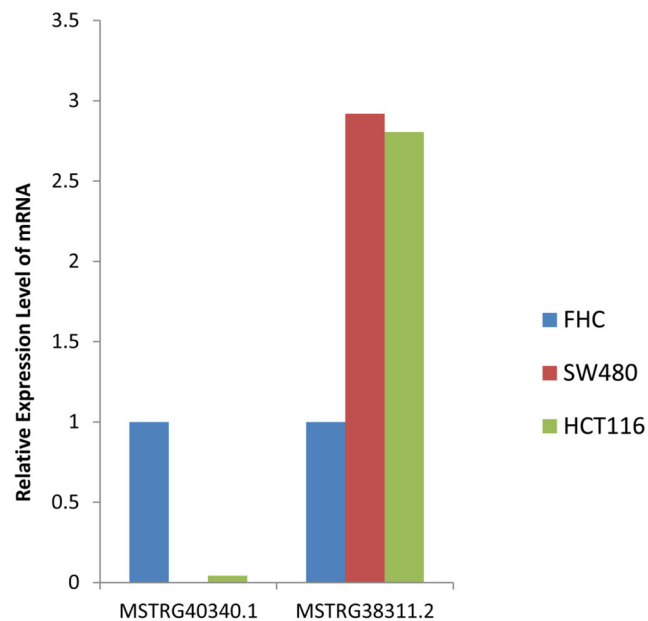
Sha He<sup>1#</sup>, Juanzhi Chen<sup>2,3#</sup>, Huan Gao<sup>1</sup>, Guixian Yang<sup>1</sup>, Feixiang Zhang<sup>2</sup>, Yanqing Ding<sup>2,3\*</sup>, Hao Zhu<sup>1\*</sup>

1. Bioinformatics Section, School of Basic Medical Sciences, Southern Medical University, Guangzhou, 510515, China

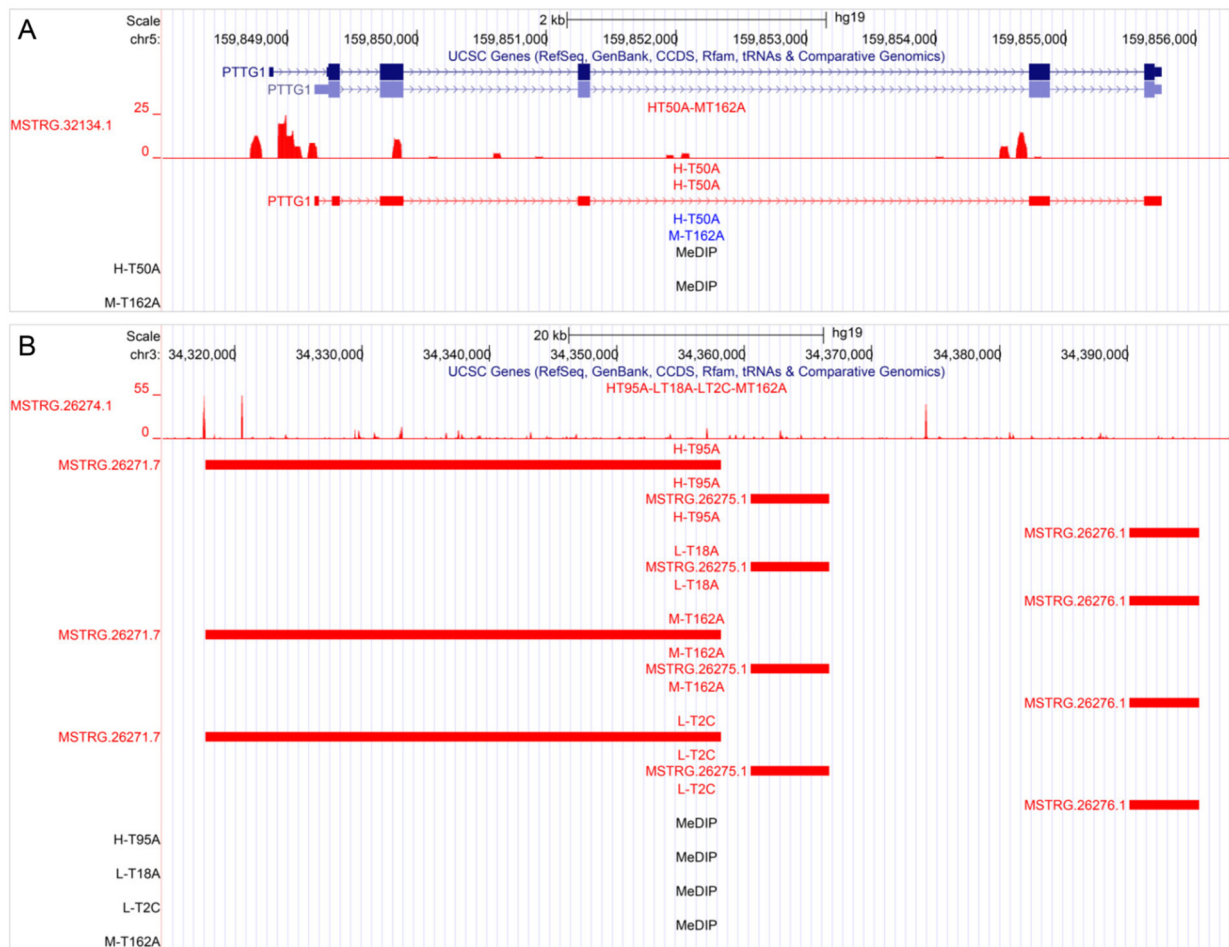
2. Department of Pathology, Southern Hospital, Southern Medical University, Guangzhou, 510515, China

3. Department of Pathology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, 510515, China

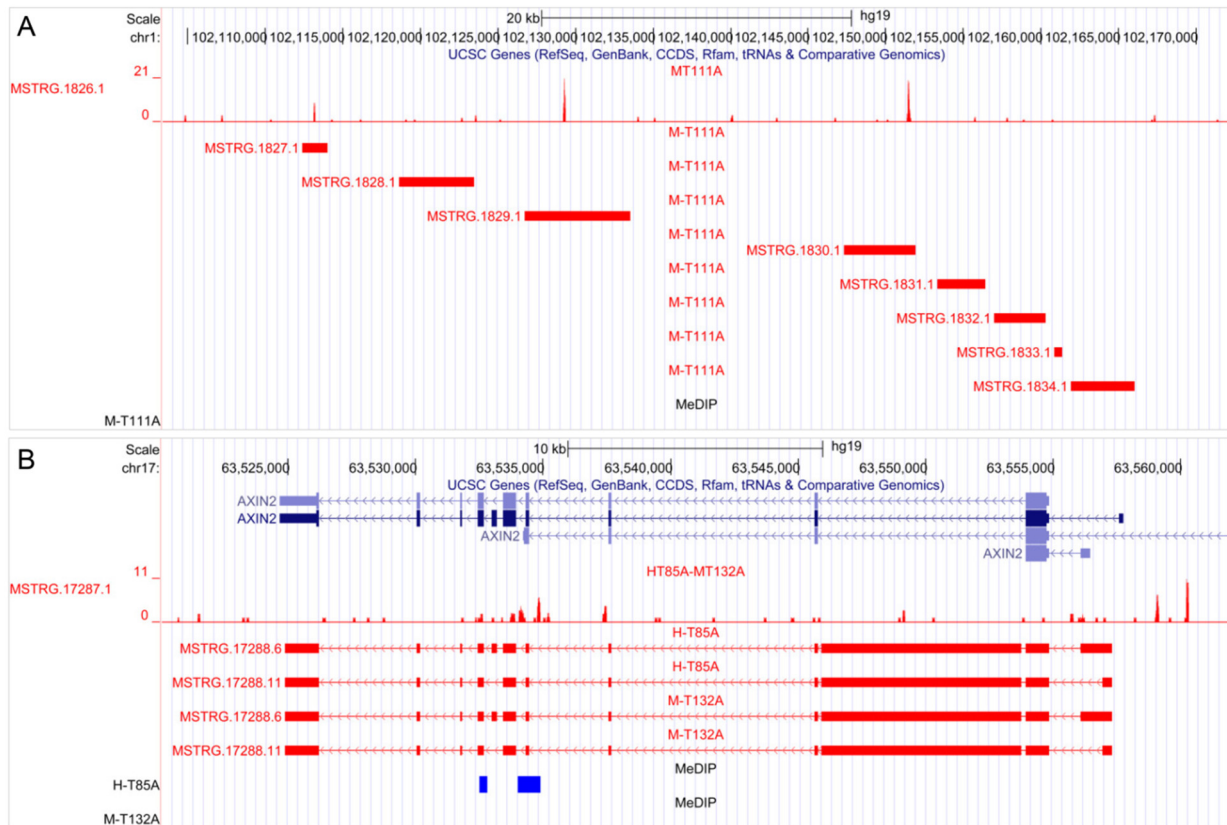
### Supplementary Figures



Supplementary Figure 1. MSTRG.40340.1 mRNA expression was high in the FHC cell line but very low in the SW480 and HCT116 cell lines ( $F=160.66$ ,  $P<0.001$  (F-test)). MSTRG.388311.2 mRNA expression was high in the SW480 and HCT116 cell lines ( $F=17.85$ ,  $P<0.005$  (F-test)).



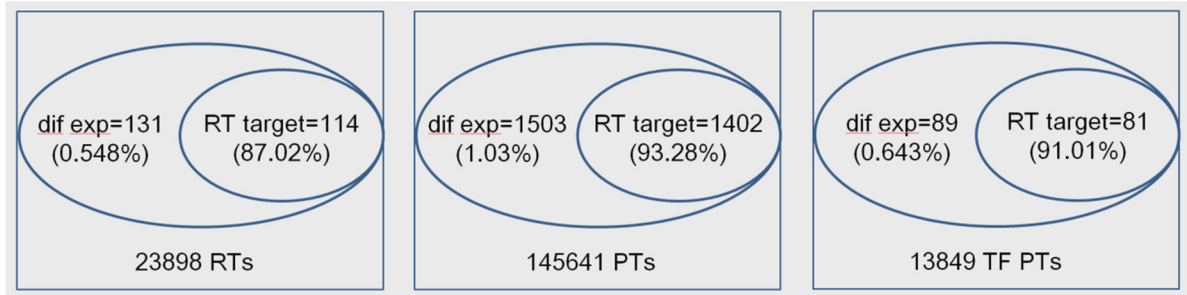
Supplementary Figure 2. Some MSTRG RTs have TTSs in their local genomic regions. (A) MSTRG.32134.1 (transcribed near *MIR3142HG*) has a TTS in the promoter region of *PTTG1* in T50A and T162A. *PTTG1* is up-regulated in T50A. (B) MSTRG.26274.1 (transcribed near *LINC01811*) has three TTSs in the local genomic region in T95A, T18A, T2C, and T162A. MSTRG.26271.7, MSTRG.26275.1, and MSTRG.26276.1 are generated in multiple samples.



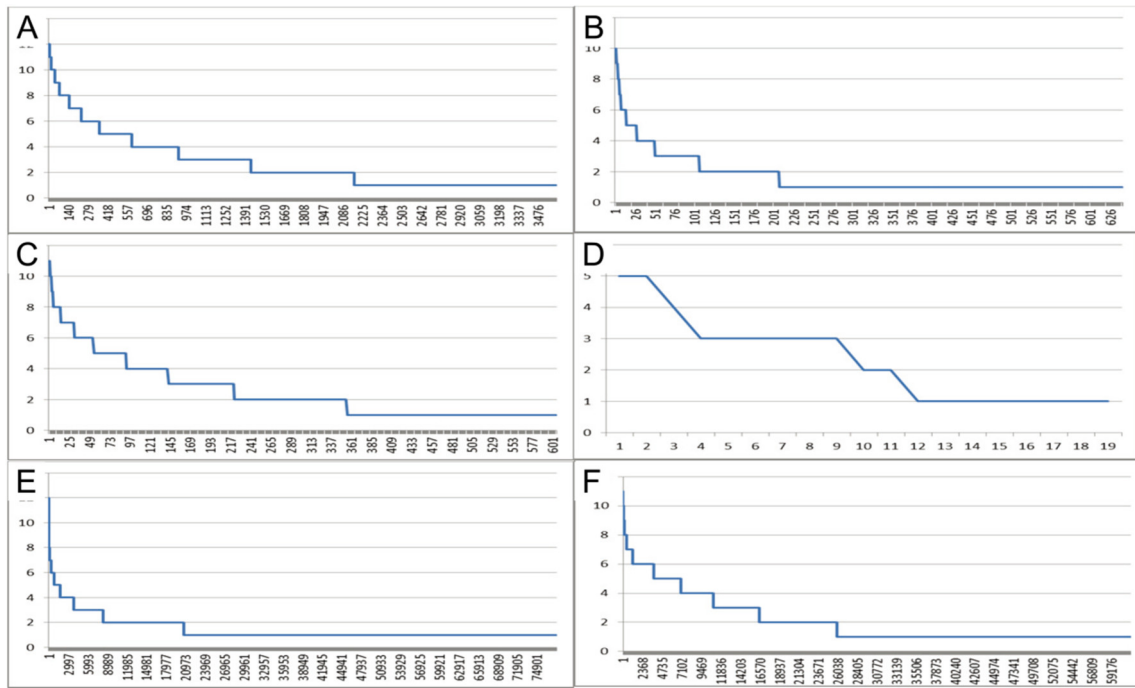
Supplementary Figure 3. (A) MSTRG.1826.1 has TTS in the regions of multiple MSTRG transcripts in T111A. (B) MSTRG.17287.1 has TTS in the region of AXIN2 and several MSTRG transcripts (which may be wrongly transcribed AXIN2).



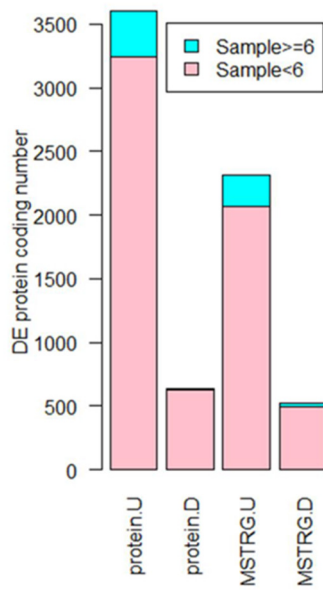
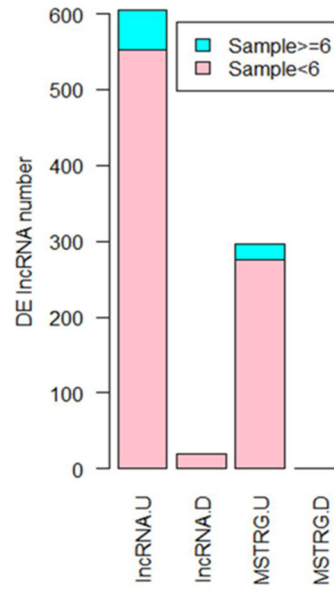
Supplementary Figure 4. (A) The lncRNA *LPP-AS2* has TTSS in the region of multiple *MSTRG* transcripts in multiple tumor samples, together with DMRs in the region. Some of these *MSTRG* transcripts may be wrongly transcribed *LPP* and *LPP-AS2*. (B) The lncRNA *RP11-254F7.2* has a TTS in the promoter region of *GRHL1*, which is up-regulated in the sample T-19A.



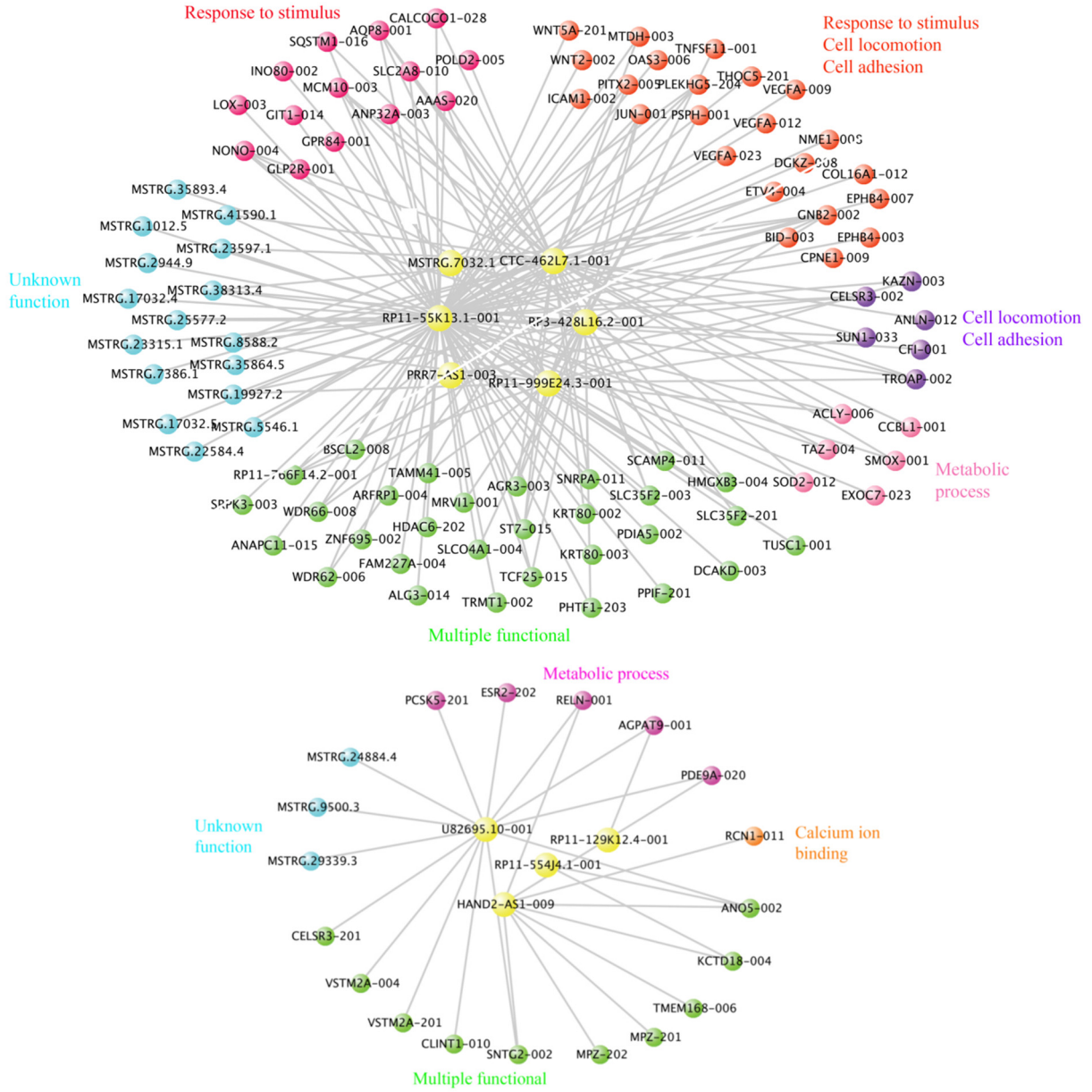
Supplementary Figure 5. Most differentially expressed RTs, PTs, and TF transcripts may undergo epigenetic regulation. In the 9 modules 23898 RTs, 145641 PTs, and 13849 TF transcripts are identified, and 0.548% RTs, 1.03% PTs, and 0.643% TF transcripts are differentially expressed. 87.02% differentially expressed RTs, 93.28% differentially expressed PTs, and 91.01% differentially expressed TF transcripts have TTS of the differentially expressed RTs in their genomic regions.



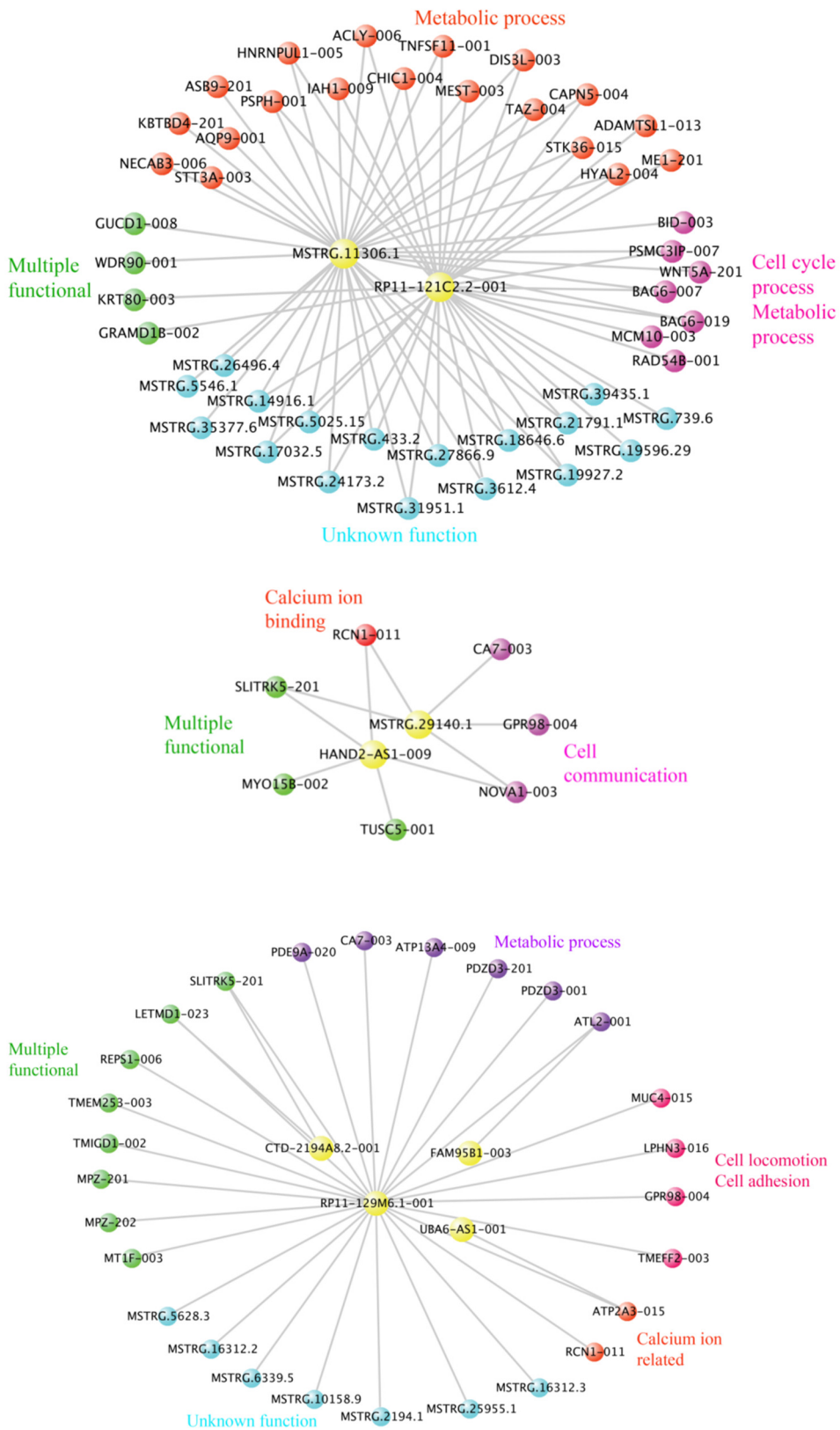
Supplementary Figure 6. Some differentially expressed transcripts occur in many samples but many differentially expressed transcripts occur in just few samples. Numbers in the Y axis indicate samples and in the X axis indicate differentially expressed genes or DMRs. (A) Up-regulated PTs in samples. (B) Down-regulated PTs in samples. (C) Up-regulated RTs in samples. (D) Down-regulated RTs in samples. (E) DMRs (increased and decreased DNA methylation) in samples.

**A****up/down protein coding transcripts****B****up/down lncRNA transcripts**

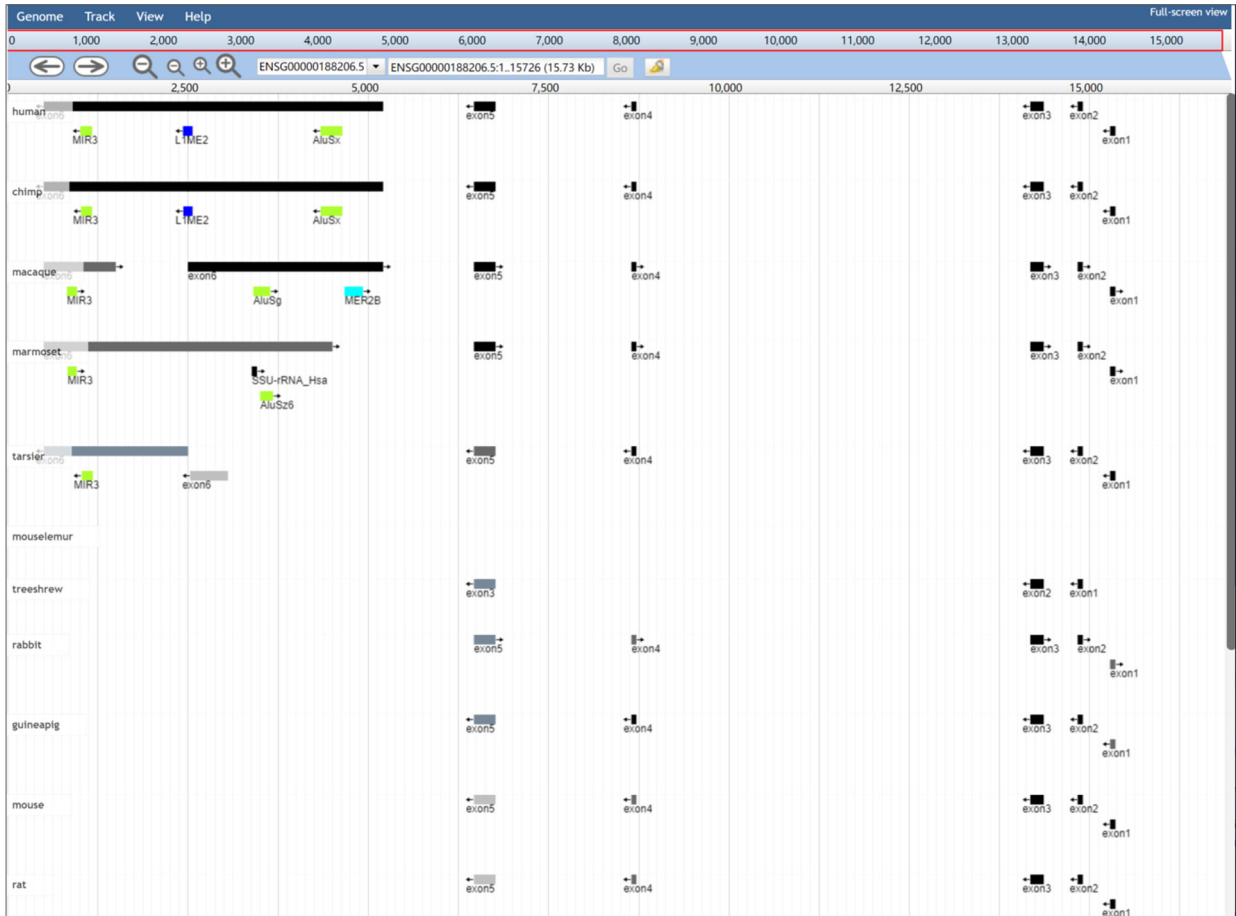
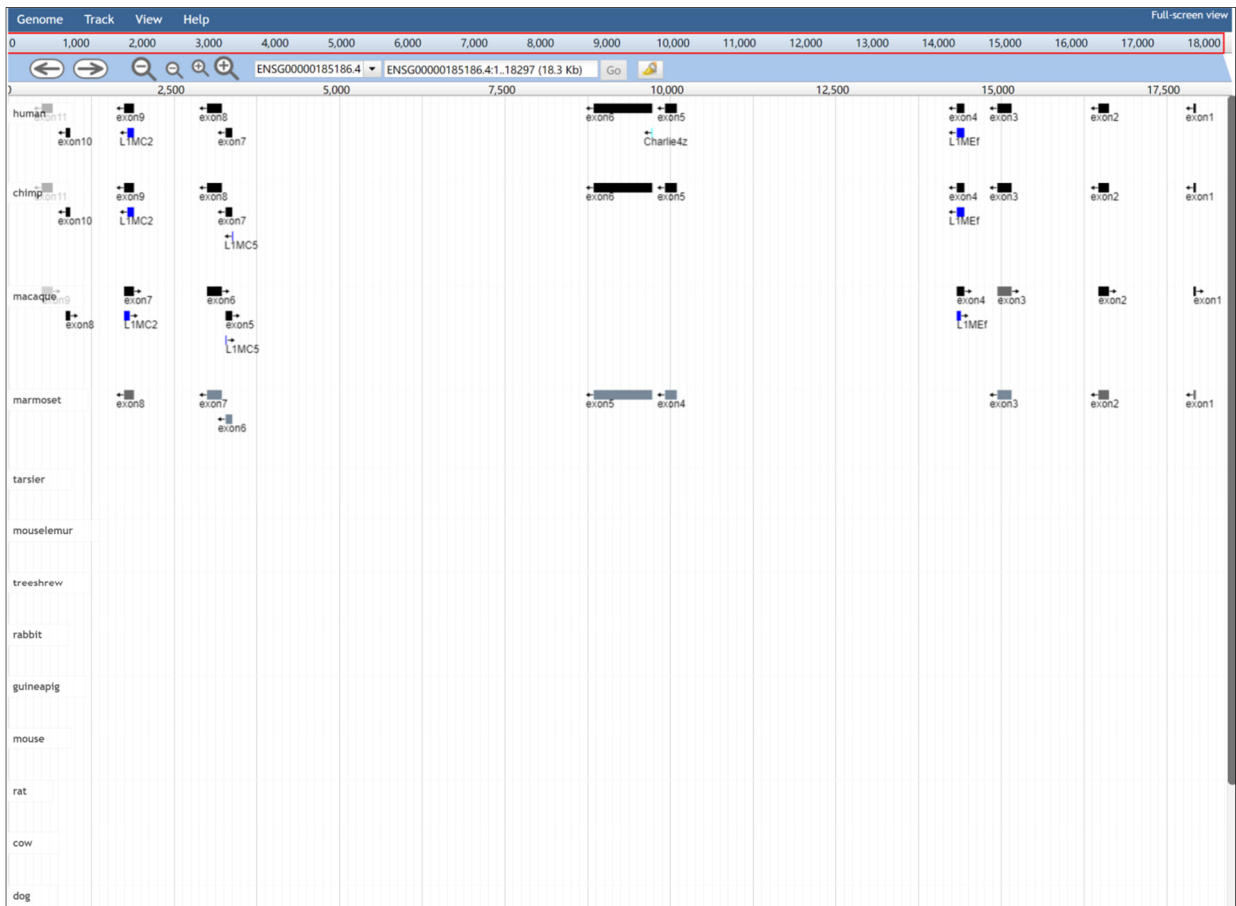
Supplementary Figure 7. (A) More PTs and MSTRG PTs are up-regulated than down-regulated. (B) More RTs and MSTRG RTs are up-regulated than down-regulated.

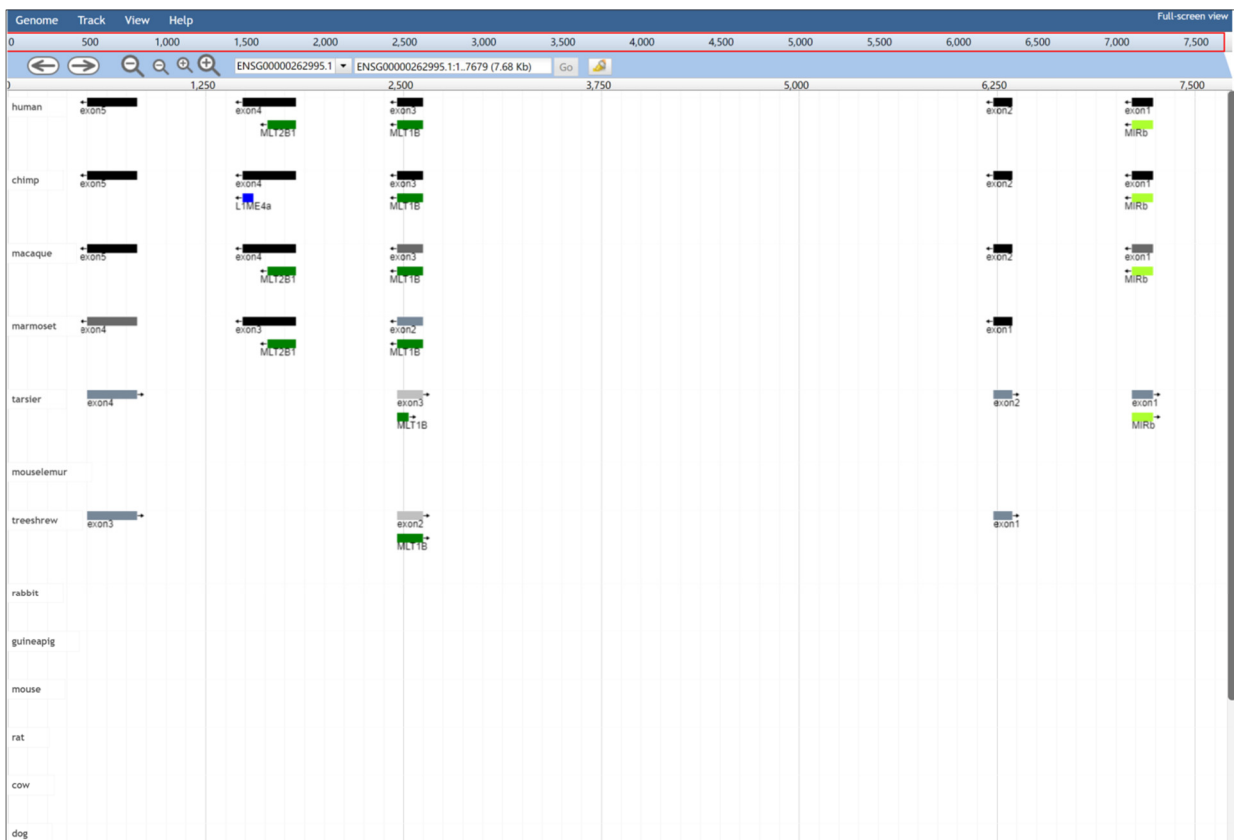
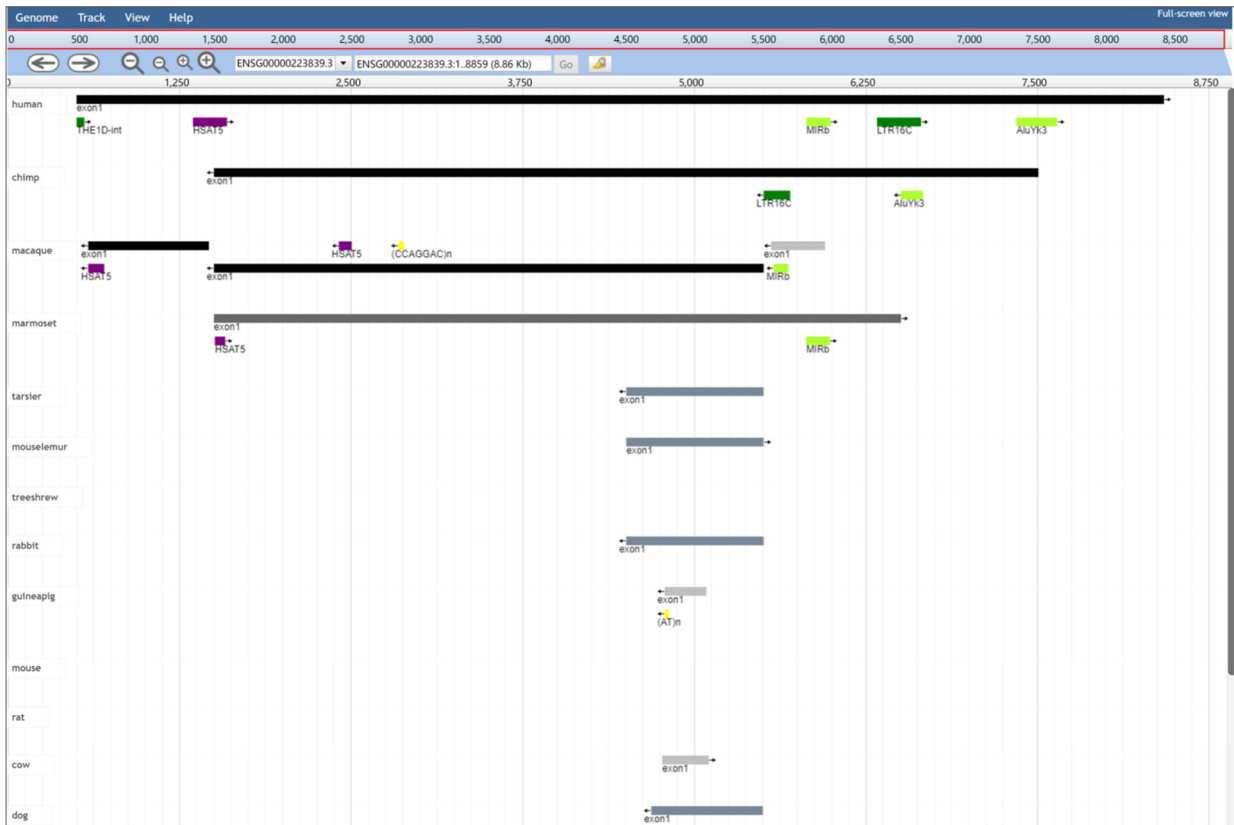


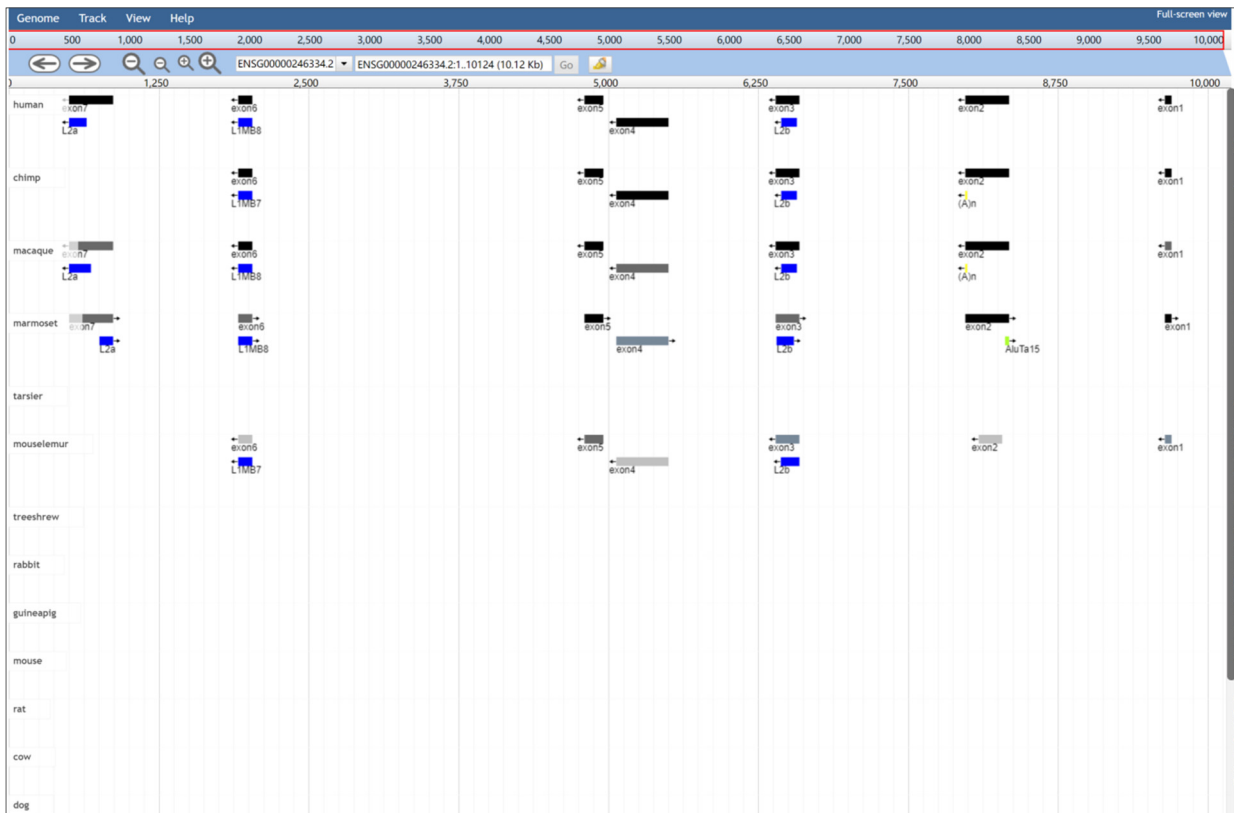


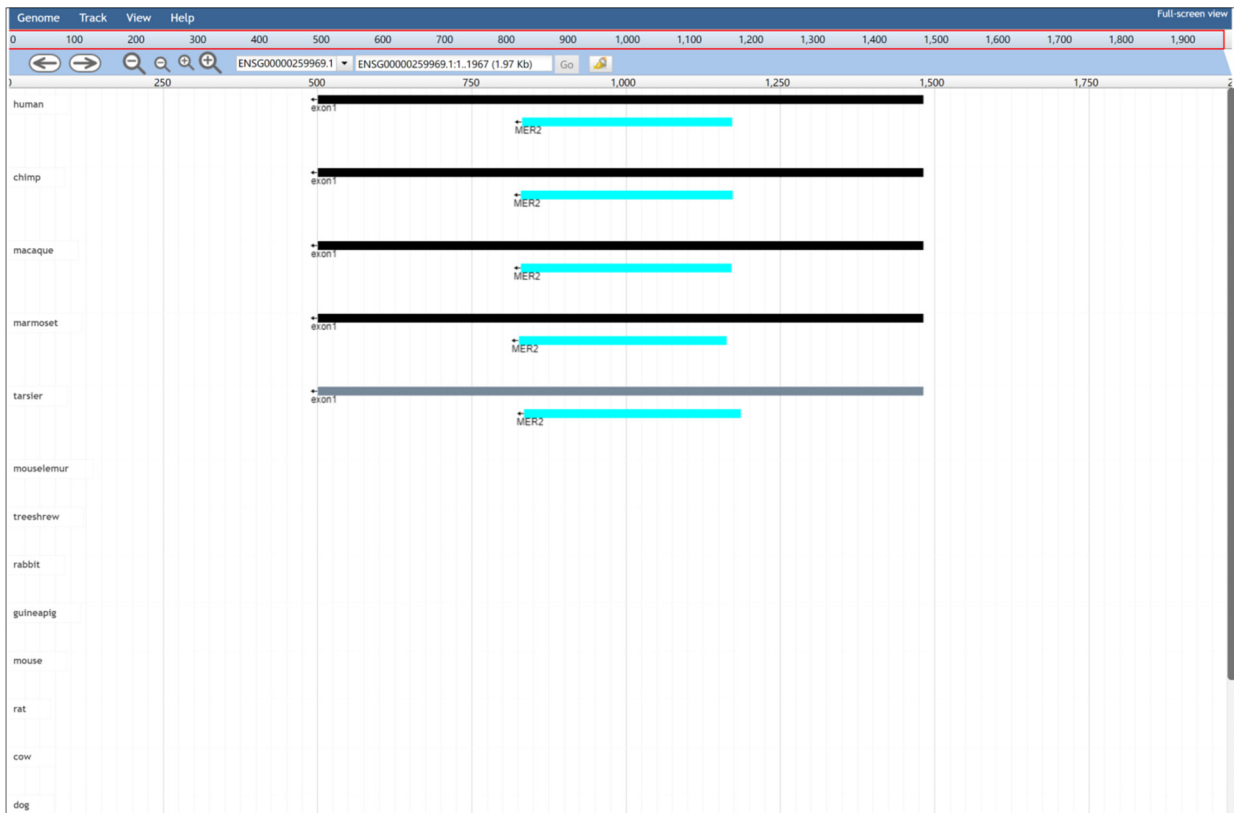
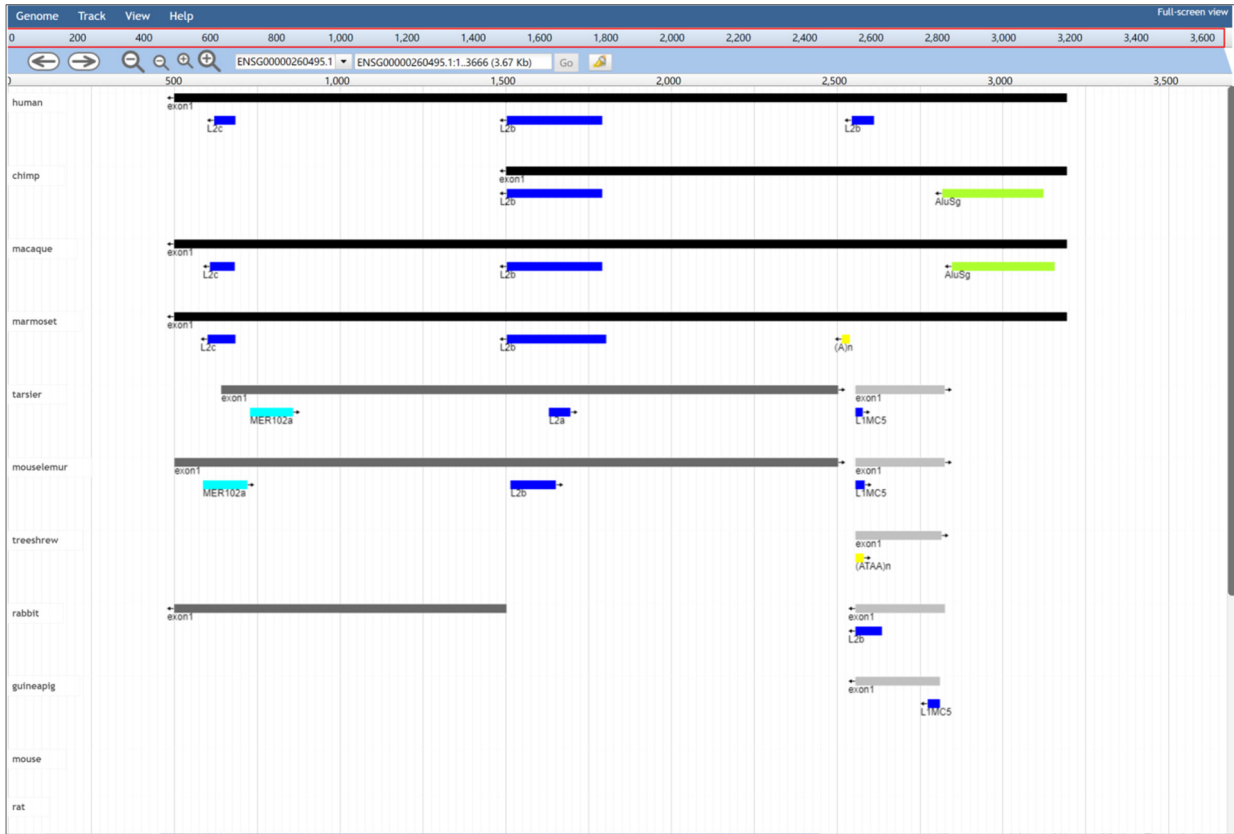


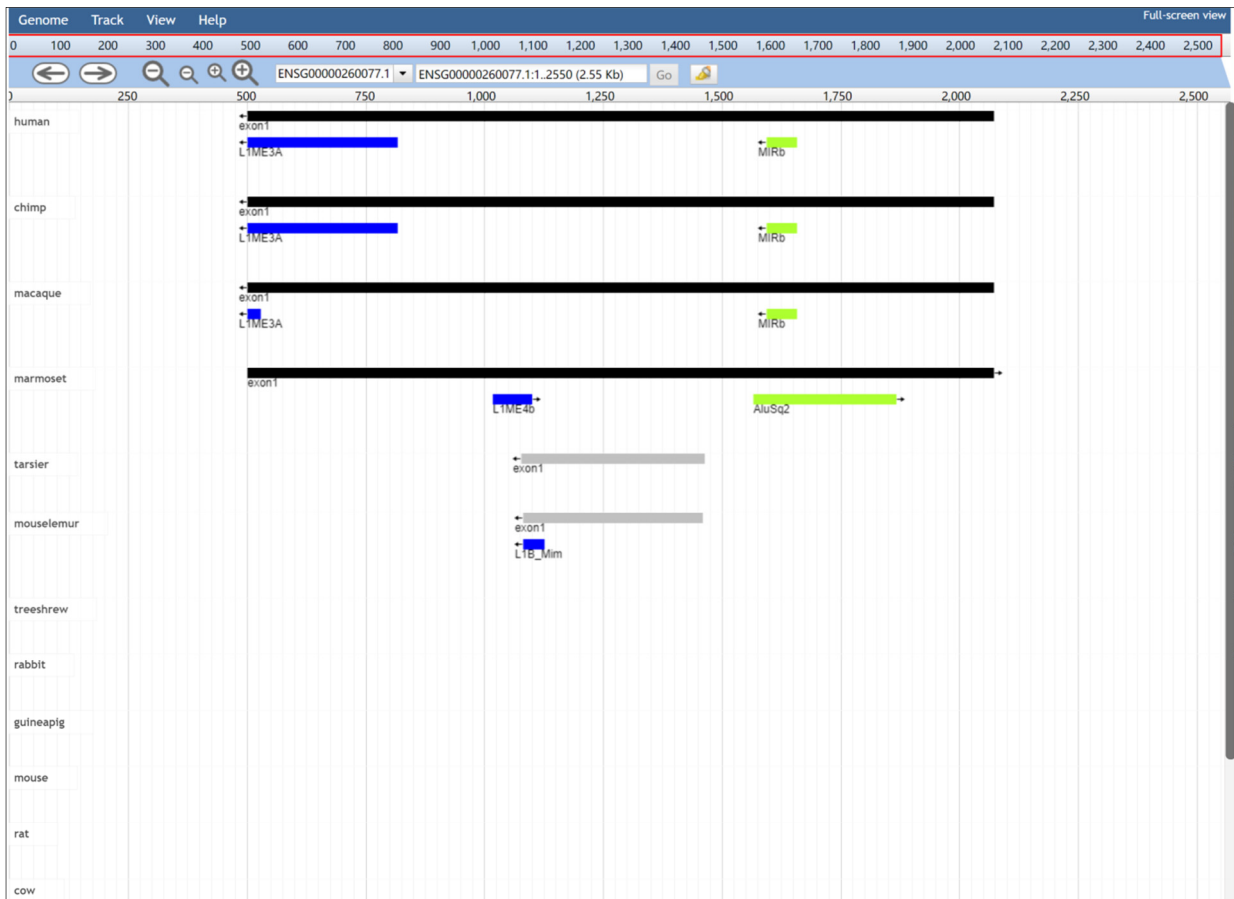
Supplementary Figure 8. From top to bottom are: The network that consists of the top 30% mostly connected RTs and PTs in the H1, H2, M2, M3, and L3 modules.











Supplementary Figure 9. Orthologous sequences of exons in most hub RTs (and the lncRNA RP11-254F7.2 in the scRNA-seq datasets) are only identified in primates. From top to bottom are LINC00313, HNRNPU-AS1, FAM95B1, CTD-2194A8.2, RP11-121C2.2, PRR7-AS1, RP11-55K13.1, RP11-999E24.3, and RP11-254F7.2 (from the lncRNA database *LongMan*) (<http://lncrna.smu.edu.cn/show/info>).