

Figure S1: Distribution of the pair numbers in 100 groups of 5,000 random sequences (bar). The pair number of 5,000 randomly selected human transcripts (*).

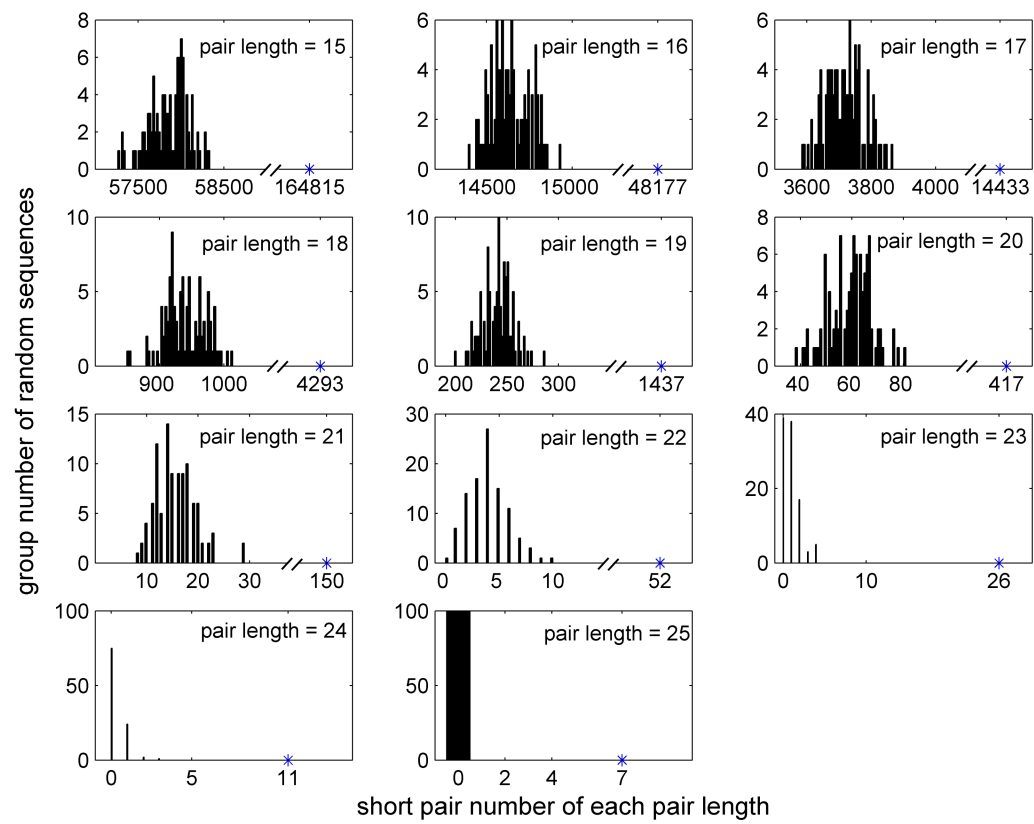


Figure S2: The kernel density distribution of natural AT and CG pairs (blue) and artificial AG and CT pairs (red) of mRNA.

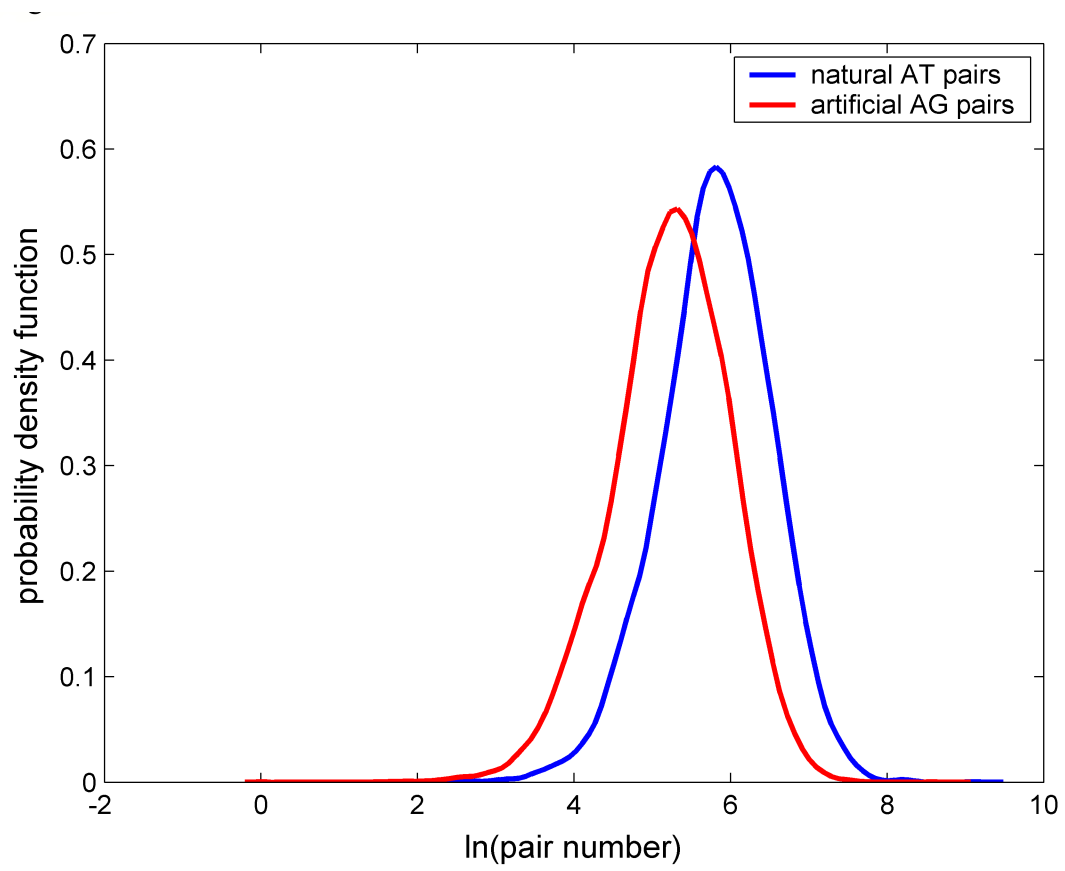


Figure S3: The non-gap pair percentage of flanking sequences of short 22bp pairs. (A) non-Alu pairs; and (B) Alu pairs.

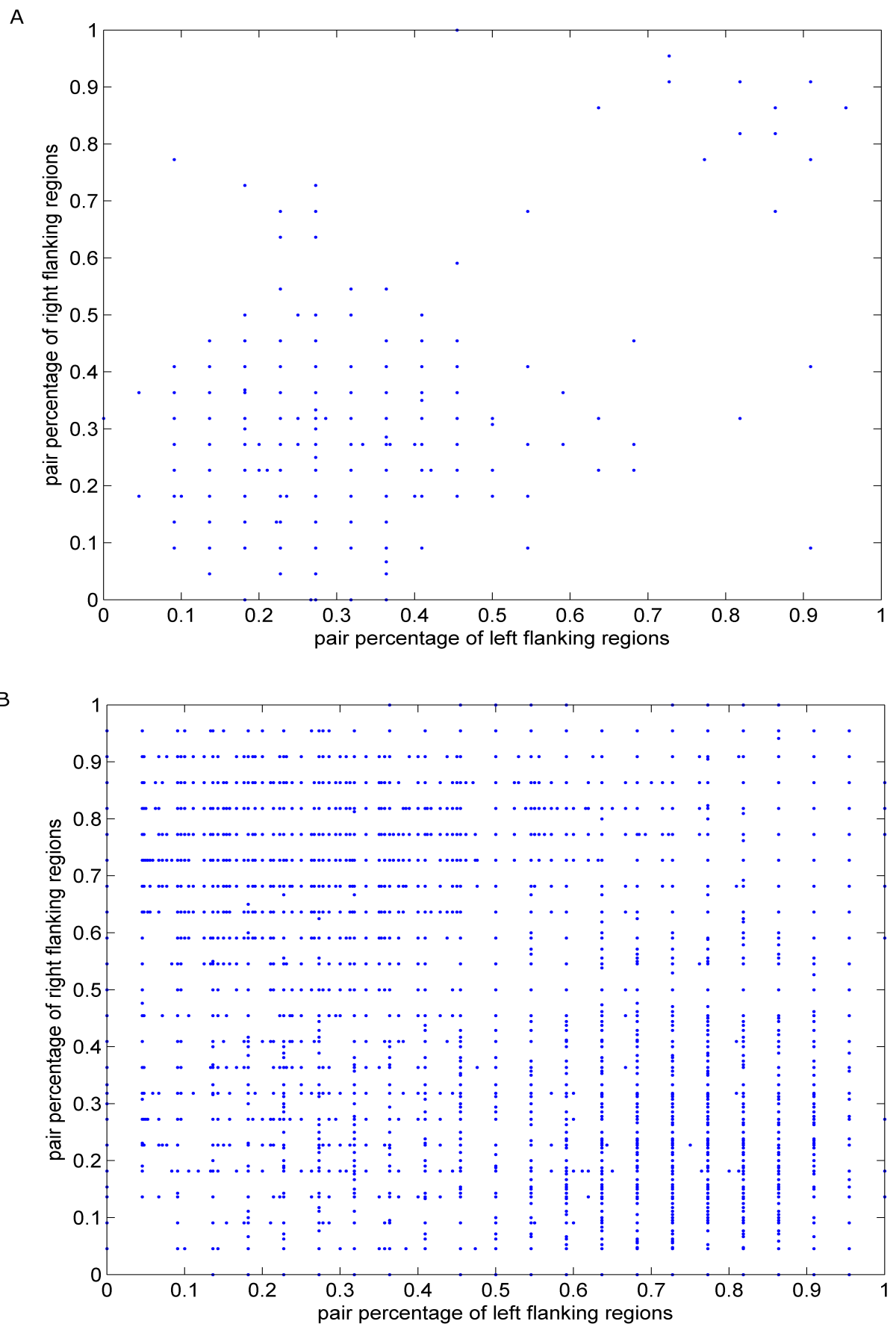


Figure S4: The mean SNP density difference between the pair region and the flanking region. (A) Short pairs without repetitive element pairs and both expressed in at least one tissue. (B) Short pairs with repetitive element pairs and both expressed in at least one tissue. *, Significant difference between pair region and flanking region, $p < 0.01$ (wilcoxon sign-rank test).

