S59-1 FLJ human full-length cDNAs and the mRNA diversity in human genes

◯Takao ISOGAI[1], Ai WAKAMATSU[1]

[1]Univ. Tokyo, Grad. Sch. Pharm. Sci.

Human gene number was predicted to be about 20 thousand. But the number of the mRNA was predicted to be several times of gene number. Those varieties of mRNAs were thought to be caused by splicing and transcription start site (TSS). In our FLJ human cDNA project, we sequenced 55,402 of human full-length cDNAs (FLJ cDNAs) and also obtained about 1.5 million of 5'-EST of full-length cDNAs from about 100 kinds of cDNA libraries consist of human tissues and cells constructed by oligo-capping method. We predicted the number of human genes transcribed into protein-coding mRNAs by using the sequence information of full-length cDNAs and 5'-ESTs of ours and public database, and obtained 23,241 of such human genes. Our FLJ cDNA sequences were covered about 76% of those. Then we constructed FLJ Human cDNA Database ver. 3.0, http://flj.lifesciencedb.jp, focusing on variations of mRNAs by splicing and TSS. Using these genes, we analyzed the mRNA diversity and consequently sequenced and identified 11,769 human full-length cDNAs whose predicted ORFs were different from other known full-length cDNAs. Especially, 30% of the cDNAs we identified contained variation in the TSS. Our analysis, which particularly focused on multiple variable first exons (FEVs) formed due to the alternative utilization of TSSs, led to the identification of 261 FEVs expressed in tissue-specific manner.