Poster L-8

Detection of the candidates of Alternative promoters and TOP genes using Database of Transcription Start Sites (DBTSS)

Authors:
Riu Yamashita (*HGC, IMS,Univ. of Tokyo*)
Katsuki Tsuritani (*HGC, IMS, Univ. of Tokyo*)
Yutaka Suzuki (*Dep. of Med. Genome. Sci., Univ. of Tokyo*)
Sumio Sugano (*Dep. of Med. Genome. Sci., Univ. of Tokyo*)
Kenta Nakai (*HGC, IMS, Univ. of Tokyo*)

Short Abstract: We have constructed the Database of Transcription Start Sites to determine precise transcription start sites. Using this database, we tried to detect alternative promoters. We also predicted terminal oligo-pyrimidine (TOP) gene which are thought to be related with translation.

Long Abstract:
One of the major topics in the post-sequence era is the analysis of transcription regulation network. It is indispensable to define precise promoter regions and transcription start sites for this analysis. To support it, we have constructed the DataBase of Transcription Start Sites (DBTSS) (http://dbtss.hgc.jp), which contains accurate transcription start sites (TSSs) information based on experimentally determined 5'-end clones. It currently contains 1,359,000 clones corresponding to 15,262 human genes, as well as 364,487 clones corresponding to 14,162 mouse genes[1]. Here we report two elements from the analyses of these data: alternative promoters and Terminal Oligo-Pyrimidine rich genes (TOP genes).
It is getting clearer that alternative promoters (APs) play an important role in the spatio-temporal regulation of their downstream gene expression as well as in the isoformal production of their products. Recently, we have reported that genes having CpG islands are expressed in a housekeeping manner. In order to understand the relationship between the CpG islands and the APs, we searched APs for CpG islands. Candidate APs were obtained from the newly-constructed DBTSS ver.5: we could obtain 7,587 human and 3,816 mouse candidate genes. We also obtained a set of genes which are unlikely to have any APs (human: 7,041, mouse: 9,887). We predicted 8,195 CpG islands on the exact TSS regions in human genes and 7,706 in mouse genes using the 'newcpgreport' program in the EMBOSS package. In the negative group, more than half of the promoters had CpG islands (58 % in human, 55% in mouse). They occupy the majority of all identified CpG islands (66% in human, 70% in mouse). Interestingly, it is rare for all promoters of one gene to have CpG islands (4% in human, 5% in mouse), and most of the APs have no CpG islands around them. While we observed several genes containing no CpG islands in their promoter region at all (25% in human, 34% in mouse), no less than half of the genes had promoters with and without CpG islands. Altogether we concluded that the genes regulated by APs tend to have a combination of CpG+ promoters and CpG- promoters and that the latter promoters may be important for their tissue-specific expression[2].
It is known that the translation of most ribosomal proteins and translation factors is regulated by the same mechanism. These genes have terminal oligo-pyrimidine sequence at the 5'-end

of mRNA and are called TOP genes. However, how many TOP genes exist in the human and mouse genomes is still unknown. By using the accurate TSSs information provided by DBTSS, higher accuracy to detect them is expected. As a first step, we focused on genes that have fixed TSS positions, which will also reduce the number of potential false positives. By using a position specific weight matrix constructed from 48 known TOP genes and, we could screen 1,645 candidate TOP genes. Among them, the 8 previously characterized TOP genes were included. Moreover, 72 out of 75 ribosomal protein genes were also included. According to the number of 5'-end clones, the predicted TOP genes were expressed in housekeeping manner. 530 human TOP gene candidates also have mouse orthologs. Our result suggests that TOP genes are not restricted to translation-related genes but may include a wider variety of genes such as chaperones and transport proteins. We also experimentally validated some of the candidates, and these results were consistent with the characteristic of TOP genes.

[1]. Yamashita R. et al.(2006) DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. Nucleic Acids Res. Jan 1;34(Database issue):D86-9.
[2]. Kimura K. et al. (2006) Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. Jan;16(1):55-65. Epub 2005 Dec 12.