Comprehensive Detection of Terminal Oligo-Pyrimidine (TOP) Genes in Human Genome

Riu Yamashita
ryamasi@hgc.jp

Yutaka Suzuki
ysuzuki@ims.u-tokyo.ac.jp

Hiroyuki Wakaguri
wakaguri@ims.u-tokyo.ac.jp

Sumio Sugano
ssugano@ims.u-tokyo.ac.jp

Kenta Nakai
knakai@ims.u-tokyo.ac.jp

1 Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
2 Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8562, Japan

Keywords: transcription, comparative genomics, full-length cDNA, promoter, terminal oligo-pyrimidine

1 Introduction
It is known that the translation of most ribosomal proteins and some of translation factors is regulated by the same mechanism. These genes have a terminal oligo-pyrimidine sequence at the 5'-end of mRNA, and are hence called TOP genes. Interestingly, these mRNAs are also known to encode translationally regulated genes. Under normal conditions, most of mRNAs have many ribosomes, and such status is called ‘polysome’. Once the cell is faced with starvation or appropriate chemical treatments such as 12-O-tetradecanoyl-1-phorbol-13-acetate (TPA), TOP genes release their ribosomes and changed their status into ‘subpolysome’, and are in a translationally inactive state. On the other hand, most of the other non-TOP mRNAs stays in polysome [1]. There have been some reports for these characteristics; however, how many TOP genes exist in the human genome is still unknown. We performed a detection of most of the TOP gene candidates using the accurate TSSs’ information provided by the Database of Transcription Start Sites (DBTSS). We also experimentally validated 83 of these candidates, and found 44 of them to be translationally regulated. Our results suggest that TOP genes are not restricted to translation-elongation factors genes but may include a wider variety of genes such as translation initiation factors and other proteins. We also found that this translational regulation is related with the length of mRNAs.

2 Methods
In DBTSS version 5.1, there are 425,117 human TSSs corresponding to 19,573 NCBI reference sequence (RefSeq) genes. Among them, we first used the set of 921 genes (921 TSSs) that have more than 10 clones, of which more than a half start from the same TSS. This dataset contained 48 known TOP genes and 873 presumed not-TOP genes. We constructed position specific weight matrix (PSWM) from these 48 genes. To increase the reliability of our data, we used only TSSs that were indicated by multiple clones, reducing the TSS number to 87,397 (13,717 RefSeq genes) for human.

We also used mouse TSS information from DBTSS. 149,876 of TSSs were registered corresponding to 14,745 mouse genes. Because the number of 5'-end clones of mouse is relatively small, we used all of them to detect TOP gene candidates.

3 Results and Discussion
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 the PSWM constructed from 48 known TOP genes, we could screen 1,645 candidate TOP genes. Among them, the eight previously characterized TOP genes [1] were included. Moreover, 72 out of 75 ribosomal protein genes were also included. We also obtained 239 genes which are defined as orthologous TOP genes both human and mouse.
For the experimental validation, we used the samples of mRNA from Human HL-60RG cells by Takeuchi and Ueda [3]. To avoid the influence of the amount of total RNA expression level, we therefore focused on only the genes which expression levels were not very different. We used 86 candidates including 34 known TOP (including ribosome proteins) genes, and 10 negative controls. We focused on this 96-gene group to validate ‘TOP-ness’, i.e., if these genes would shift their status before and after TPA treatment. As shown in Figure 1A, the expression level of most of the TOP gene candidates and all of the negative control genes were higher in the polysomal fraction. After the TPA treatment, most of the expression levels of TOP candidates changed to the subpolysomal fraction (Figure 1B). 44 genes showed higher expression in the subpolysomal fraction. On the other hand, all of the negative control genes still showed higher expression in polysomal fraction.

In Figure 1, we found that some of the detected genes did not behave in a TOP manner. Ledda et al. reported that the length of the 3'-UTR influences the translational regulation [4]. To test this idea, we checked the relationship between translational regulation and mRNA 3'-UTR-length, and a clear correlation was observed between them (Figure 2A). Moreover, we also observed correlations with translational regulation in total mRNA length, ORF length, and even 5'UTR length (Figure 2B-D). The RNA length showed the highest correlation coefficient ($r=-0.61$), and it is strongly related with other the three categories of length (3'-UTR: $r=-0.56$, 5'-UTR: $r=-0.042$, ORF: $r=-0.53$). These observations suggest that the total length of mRNA as well as the 5'-end pyrimidine tract is responsible for translational regulation. This means that TOP characteristics depend on the total length of mRNA.

**References**


[2] [http://dbtss.hgc.jp](http://dbtss.hgc.jp)
