

Tissue Specificity Dependent Sequence Conservation

Cornero A, Squillario M, Ikeo K, Gojobori T

Center for Information Biology and DNA Data Bank of Japan,
National Institute of Genetics, Mishima, Japan

Motivation

To understand how the regulation of gene expression works, functional identification of the non-coding regions is of essential importance, because a large part of the regulatory elements that control gene expression are located in these sequences. In the last few years, various research groups have identified relationships between functions of the gene and the degree of sequence conservation of its 5' flanking region. For example, Iwama and Gojobori (2004) discovered that the degree of sequence conservation in the upstream region between human and the mouse is significantly correlated with gene function: The genes coding for transcription factors are more conserved than other genes. Similarly, Lee et al. (2005) found that there was a significantly higher conservation of the 5' flanking regions of genes involved in development compared to those of genes involved in metabolism. However, any reasonable explanation for those observed relationships has not been given yet.

Results

To better understand how the degree of sequence conservation in non-coding regions is related with gene function, we took into account the gene expression patterns and conducted an analysis of the sequence conservation of the non-coding regions, particularly in two classes of genes; housekeeping genes (HK) and genes coding for transcription factor (TF). In our analysis, we selected specifically these two classes of genes because of the obvious difference in gene expression patterns. First, we found that there was a statistically significant difference in the degree of sequence conservation of the upstream non-coding regions between the two gene classes. Next, since a variability of tissue specificities exist among TF genes, we performed a correlation analysis between the degree of sequence conservation of the 5' flanking regions and the tissue broadness of expression of TF coding genes. As a result, taking advantage of using qRT-PCR data, we successfully identified a significant negative correlation between the degree of sequence conservation of the upstream regions and the broadness of the gene expression. In particular, we found that this negative correlation was amplified when we adopted both the 4 Kbp non-coding sequence upstream to the transcription starting site and the 5' untranslated region as the non-coding regions. The results obtained from our analysis suggest that the transcription control for the genes expressed in specific tissues may have undergone more dynamic changes, through sequence changes of the 5'-flanking regions, than that for the genes expressed in a wide range of different tissues. References: - Iwama,H. and Gojobori,T. (2004) Highly conserved upstream sequences for transcription factor genes and implications for the regulatory network. *Proc Natl Acad Sci U S A*, 101:17156-61. - Lee,S., Kohane,I. and Kasif,S. (2005) Genes involved in complex adaptive processes tend to have highly conserved upstream regions in mammalian genomes. *BMC Genomics*, 6(168).

Contact : tgojobor@genes.nig.ac.jp