Human brain origin and the evolution of TATA-boxes of protein-coding genes expressed in brain

K.V. Gunbin, M.P. Ponomarenko,

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, genkvg@bionet.nsc.ru, pon@bionet.nsc.ru

D.A. Afonnikov

Institute of Cytology and Genetics SB RAS, Novosibirsk State University, Novosibirsk, Russia, ada@bionet.nsc.ru

The great enigma of contemporary evolutionary biology is how tiny changes of hominid genes could result in the unique human brain. In order to answer this question we performed a complex evolutionary analysis of hominid (*H. sapiens, P. troglodytes, G. gorilla, P. pygmaeus*) genes whose human orthologs are expressed in the brain and other organs.

We extracted multiple alignments of upstream regions of protein coding genes from the EnsEmbl Rel. 70. We took human transcription start sites (TSS) from GENCODE rel. 15 (only validated and manually annotated mRNA starts were used). For each alignment, ancestral nucleotide sequences were reconstructed in each internal node of the Hominidae tree using the Maximum likelihood approach implemented in the PAML 4.7 program package. That was done in order to focus on only those positions, whose evolution could be reconstructed at a probability higher than 99%. To predict TATA-box activity we implement our algorithm described in [1], which used an empirical equilibrium equation for three-step for TBP/TATA binding: TBP slides along DNA, TBP stops in the TATA box, the TBP/TATA-complex is stabilized due to deformations in DNA. Then, mRNAs with biochemically significant changes in TATA-box activity (negative or positive) were chosen for each Hominidae tree branch.

Annotation of mRNAs with biochemically significant changes in the activity of their TATA boxes was performed based on the information contained in the 3 sources of gene expression data: the Allen Human Brain Atlas, Human Brain Transcriptome from Sestan laboratory and the Gene Expression Barcode 2.0. Only statistically significant gene

expression values compared with the background were included in the analysis. For the statistical evaluation of the contribution of mRNAs with evolving TATA-boxes to tissue functioning we used the following randomization test. On the basis of the full set of mRNAs for each tissue/structure, 10^5 random samples of mRNAs were generated, the size of each of these samples was equal to the size of the "evolving" set of mRNAs; than for each of the random samples, the sum of mRNA expression values was calculated; the amount (Q) of such sums equal to or greater than the sum of the expression values for mRNA targets belonging to the "evolving" set (A) was calculated. Thus, the probability of observing A by random chance is $p=Q/10^5$.

Two global evolutionary trends for hominid genes whose human orthologs are expressed in brain were found: (1) changes in TATA-box activity most characteristic for hominid genes whose human orthologs are highly expressed in brain cortex regions (especially in prefrontal cortex); (2) the relative conservation of the ancestral TATA-box organization (20-90 b.p. upstream TSS) and activity in human genes compared with intensive negative changes in TATA-box activity in primate orthologs. It is of interest that the first trend elucidates the evolving of prefrontal cortex in hominids, while the second describes the potential mechanism of brains specializations.

The work supported by RFBR 13-04-00359 and 13-04-01008, SciSh-5278.2012.4, RAS Programs "Molecular and cellular biology", RAS program A.II.6.8, SB RAS Integration projects 39 and 130.

1. Ponomarenko PM, Savinkova LK, Drachkova IA, Lysova MV, Arshinova TV, Ponomarenko MP, Kolchanov NA. (2008) A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism, *Dokl Biochem Biophys.* **419**:88-92.