Evolution of core promoters in hominids towards the increased "norm of reaction" of gene regulation in the lineages of *Homo sapiens*.

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

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, genkvg@gmail.com

Novosibirsk State University, Novosibirsk, Russia

F. Gusev, E.I. Rogaev

University of Massachusetts Medical School, Worcester, Massachusetts, USA

Vavilov Institute of General Genetics RAS, Moscow, Russia

The adaptability to variety of environmental conditions is a prominent feature of *Homo sapiens* whereas other primates are adjusted to specific environments and restricted in their adaptation capacity. We hypothesize that the broader response of humans to environmental stresses can be explained by evolutionary changes leading to a more flexible gene regulation network especially underlying CNS functioning and behavior. The genotype-dependent range of expression of genes ("norm-of-reaction") essential for rapid adaptation to diverse conditions can be broader in humans than in other higher primates. Thus, we sought for specific signatures of evolutionary changes in promoter architectures of hominid genes that may indicate a widening the "norm of reaction" of gene expression rather than just changes in level of expression, such as down- or up-regulation of the gene.

In this study, we have focused on the structural and functional evolution of the core promoters of human genes, especially of those expressed in the brain prefrontal cortex. To identify functionally active transcription start sites (TSS), potential TSS locating near annotated 5' termini of all human gene transcripts (GENCODE v. 19) were searched on the basis of CAGE data from FANTOM5 database. After that, we were searched for H3K4me3 Chip-seq peaks located near these TSS, which were obtained on three separate tissues: lymphoid tissue [1, 2], neuronal and non-neuronal tissues [3-5]. Using this approach we revealed the transcriptionally active promoters in a certain genomic regions and subdivided promoters by its tissue specificity. Additionally, to define the evolutionary stable promoters through the hominid evolution we aligned experimentally detected regions (using H3K4me3 RNA-Seq peaks [5]) corresponded to transcriptionally active neuronal core promoters between human, chimpanzee

and macaque genomes. The multiple alignments of gene promoter regions and the chromosome positions of aligned regions were taken from Ensembl rel. 75. From these alignments we excluded alignments containing many N's or gaps in region corresponding roughly to wide TBP binding site ([-10; -100 bp]). In each internal node of the Hominidae tree, the ancestral nucleotide sequences were reconstructed using baseml program from the PAML 4.7 package and the ancestral pattern of indels, using PRANK v.121218 program. Additionally, all reconstructed ancestral sequences were corrected for any parsimony inappropriate states.

In this work we made a comparative analysis of the rates of molecular evolution of the three features of the broad [-600; -1] and/or narrow [-200; -1] upstream region of core promoter (appearance or disappearance of GC dinucleotides, nucleosomal packing level of core promoter, and predicted promoter/TBP affinity) on a different branches of the hominid tree. It is of importance that this analysis does not require the absolute length of the branches of the phylogenetic tree, just the relative length. Recent literature regarding the matter suggests that the relative lengths of the branches of the hominid tree are subject to very little variation.

At the first step, to analyze nucleotide variation in promoters during the evolution, we have conducted both a general study of promoter variation, which takes into account all relevant single-nucleotide substitutions and indels, and a study of evolutionary variation in the number of GC dinucleotides. The general analysis of promoter variation was shown a slight increase in the evolutionary rates specifically for broad regions of the promoters [-600; -1] on the human and chimpanzee branches and on the branch leading to human-chimpanzee ancestor. The analysis of variation in the number of GC dinucleotides revealed clear differentiation in the manner of the substitutions in Hominidae tree branches: a highly statistically significant (p<0.0001) prevalence of GC dinucleotides lost observed on all Hominidae tree branches except for branch leading to human-chimpanzee ancestor and for human branch. Because GC-dinucleotide rich promoters are known to adopt a transcriptionally variable state within which initiation can occur at a number of locations in contrast with TATA-box containing promoters, it is possible that there were a decrease in the norm-of-reaction spectrum for gene regulation on the majority of Hominidae tree branches. On the contrary, the branch leading to human-chimpanzee ancestor and the human branch can be characterized by maintaining or even

widening the norm-of-reaction spectrum for gene expression regulation.

According to modern views the magnitude of transcription "is a promoter-specific property that is relatively robust to sequence mutations but is strongly dependent on the interaction between the TATA box and promoter nucleosomes" [6]. Considering this circumstance, we thoroughly analyzed the evolution of the TBP affinity [7] and nucleosome potential [8] in the core-promoter region of human and higher hominid genes. The comparison of the "ancestral" (branches leading to human-chimpanzee and human-chimpanzee-gorilla ancestors, human branch) and "terminal" (orangutan, gorilla, and chimpanzee branches) groups of the hominid tree branches clearly demonstrated that negative changes in the promoter/nucleosome affinity were much less frequent in the "terminal" than in "ancestral" group and vice versa positive changes in the promoter/nucleosome affinity were much more frequent in the "terminal" than in "ancestral" group. This fact convinces us that widening of the norm-of reaction for gene expression regulation in the human lineage is a major evolutionary tendency.

At the last step of our investigation we conducted functional enrichment analysis separately on each hominids tree branch using two contrast promoter samples (the promoters functioning in neurons and non-neuronal tissue of the prefrontal cortex and lymphoid tissue; the promoters functioning only in neurons of the prefrontal cortex) and two well-known informational resources (DAVID 6.7 and REVIGO). Only tiny statistically significant results available for "ancestral" (see above) branches and for promoter sets with general functions. This fact again convinces us that the ancestral promoter state preserved in the human lineage. On the contrary, if neuronal-specific promoter set took into account the highly statistically significant (p<0.00001) and nearly identical results for "ancestral" branches appeared immediately despite the fact that the selected threshold do not allow to select large number (>2000) of genes in these branches for such analysis. This fact allowed us to conclude that the main target for natural selection on the "ancestral" branches is the same; this target is the set of various neuronal tissues. Unexcitingly, functional categories of genes with evolved promoters acting only in neurons tightly related with skeletal system development and locomotion. This fact can be interpreted by the following hypotheses: the selective pressure on locomotion during hominid evolution can promote changes in hominid brains or vice versa, the selective pressure

on behavior can lead to changes in locomotion.

Thus, our data imply that the origin of modern man has been associated with widening the "norm of reaction" of gene expression regulation. In contrast, after splitting with ancestral lineages of *Homo sapiens*, the evolution of ape species is characterized by reduced flexibility of promoters and narrowing down the "norm of reaction" of promoters.

The work were partly supported by the grant 14.B25.31.0033 from Russian Federation Government (Resolution No.220), by the grant 14-04-00485 from RFBR and by Integration Project of the Presidium RAS No.28.

References:

1. A. Barski, et al. (2007) High-resolution profiling of histone methylations in the human genome, *Cell*, **129**:823-837.
2. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome, *Nature*, **489**:57-74.
3. I. Cheung, et al. (2014) Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex, *Proc Natl Acad Sci U S A,* **107**:8824-8829.
4. H.P. Shulha, et al. (2012) Epigenetic signatures of autism: Trimethylated h3k4 landscapes in prefrontal neurons, *Arch Gen Psychiatry*, **69**:314-324.
5. H.P. Shulha, et al. (2013) Coordinated Cell Type–Specific Epigenetic Remodeling in Prefrontal Cortex Begins before Birth and Continues into Early Adulthood, *PLoS Genet* **9**:e1003433.
6. P. Shi, et al. (2006) Did brain-specific genes evolve faster in humans than in chimpanzees? *Trends Genet*, **22**:608-613.
7. L.K. Savinkova, et al. (2013) An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein, *PLoS ONE*, **8**:e54626.
8. V.G. Levitsky, et al. (1999) Nucleosomal DNA property database, *Bioinformatics*, **15**:582-592.