

Phylostratigraphic analysis of human cancers transcriptomic data

Ivanov R.^{1*}, Mustafin Z.^{1,2,3}, Lashin S.^{1,2,3}

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

³ Novosibirsk State University, Novosibirsk, Russia

* ivanovromanart@bionet.nsc.ru

Key words: cancer, transcriptomics, phylostratigraphy

Motivation and Aim: The emergence and evolution of human cancer is a topical issue of modern biology. One of the promising modern approaches to the evolutionary analysis of genes is phylostratigraphic analysis. It allows to determine the important stages of genome evolution, at which there was a rapid increase in the number of new genes, to identify genes specific to certain organisms [1]. In particular, with the help of phylotranscriptome analysis, which correlates phylostratigraphic indices with the level of gene expression, a pattern was revealed between the level of gene expression and their age in the process of embryogenesis, both in plants and animals [2, 3]. In the course of our work, we set ourselves the task to analyze the available transcriptomic data of various types of cancer using the phylotranscriptomic method.

Methods and Algorithms: Raw expression data from tumor RNA-seq were processed using STAR and SAMtools. Phylostratigraphic and phylotranscriptomic analysis of PAI (Phylostratigraphic Age Index), DI (Divergence Index), TAI (Transcriptome Age Index) and TDI (Transcriptome Divergence Index) indices were carried out with Orthoweb web service.

Results: In our study we decided to analyze data of lung, uterus, prostate, thyroid, kidney and colon adenocarcinomas, breast ductal cancer and glioma. We obtained several gene expression datasets of these tumor tissues from open-sources, mainly TCGA and GEO NCBI. Raw counts were normalized by calculating TPM and FPKM. Phylostratigraphic indices shows the enrichment peak of genes formed early in the evolution of Eukaryote and the peak at the age of Vertebrata occurrence. The evolutionary indices data were also overlaid on pathway networks obtained from KEGG databases.

Acknowledgements: The study was supported by the budgetary projects of the Institute Cytology and Genetics of SB RAS (numbers of State registration: AAAA-A19-119101090031-8, 121031300209-8).

References

1. Domazet-Lošo T., Brajković J., Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* 2007;23:533-539. doi: 10.1016/j.tig.2007.08.014.
2. Quint M., Drost H.-G., Gabel A., Ullrich K.K., Bönn M., Grosse I. A transcriptomic hourglass in plant embryogenesis. *Nature.* 2012;490:98-101. doi: 10.1038/nature11394.
3. Domazet-Lošo T., Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature.* 2010;468:815-819. doi: 10.1038/nature09632.