

**Genome Track Analyzer:  
New toolkit for genome-wide study of  
correlations between distributed genome features**

Galina I. Kravatskaya

*Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Vavilova str., 32, Moscow 119991,  
Russia, gk@eimb.ru*

Vladimir R. Chechetkin

*Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Vavilova str., 32, Moscow 119991,  
Russia, vladimir\_chechet@mail.ru*

Nikolai A. Tchurikov

*Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Vavilova str., 32, Moscow 119991,  
Russia, tchurikov@eimb.ru*

Yuri V. Kravatsky

*Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Vavilova str., 32, Moscow 119991,  
Russia, jiri@eimb.ru*

The broad class of tasks in genetics and epigenetics can be reduced to the study of various features that are distributed over the genome (genome tracks). The rapid and efficient processing of the huge amount of data stored in the genome-scale databases cannot be achieved without the advanced software based on analytical methods. However, strong inhomogeneity of genome tracks hampers the development of relevant statistics. We developed the analytical criteria [1] for the assessment of genome track inhomogeneity and correlations between two genome tracks. We also developed a software package, Genome Track Analyzer, based on this theory. It contains the following tools applicable to genomic track investigations:

- Correlations between point-wise and stretch-wise genomic tracks
- Correlations between profiles (including expression and DNA-protein binding profiles)
- Correlations between point-wise and stretch-wise genomic tracks and expression profiles
- Statistical Kolmogorov-Smirnov and entropy tests for assessment of distribution of genomic tracks over the chromosomes

The theory and software were tested on simulated data, and were applied to the study of correlations between CpG islands and transcription start sites in the *Homo sapiens* genome, between profiles of protein binding sites in chromosomes of *Drosophila melanogaster*, and between DNA double-strand breaks and histone marks in the *Homo sapiens* genome. Significant correlations between transcription start sites on the forward and the reverse strands were observed in genomes of *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus*, *Homo sapiens*, and *Danio rerio*. The observed correlations may be related to the regulation of gene expression in eukaryotes. Genome Track Analyzer is freely available at <http://ancorr.eimb.ru/>.

This work was supported by the Russian Fund of Basic Research (grant № 14-04-01638) and a grant from the Molecular and Cellular Biology Program RAS.

1. Y.V. Kravatsky *et al.* (2015) Genome-wide study of correlations between genomic features and their relationship with the regulation of gene expression, *DNA Research*, **22(1)**:109-119.