Analysis of distal gene regulation based on ChIP-seq and ChIA-PET data and biophysical models of chromosome contacts

Arthur I. Dergilev, Ivan A. Sidorenko, Vladimir N. Babenko, Dmitri A. Afonnikov,

Institute of Cytology and Genetics SB RAS, Lavrentyeva 10, 630090, Novosibirsk, Russia,
arturd1993@yandex.ru{di,bob,ibragim,nat}@bionet.nsc.ru

Anton G. Bogomolov, Yuriy L. Orlov

Novosibirsk State University, Novosibirsk, Russia,
{mantis_anton,orlov}@bionet.nsc.ru

Computer studies of transcription regulation is important problem in molecular biology challenging growing volume of sequencing data. ChIP-Seq detects interactions between DNA and proteins; ChIA-PET (Chromatin Interaction Analysis with Paired-End-Tag) technology allows detect interactions between pairs of DNA sites affecting gene regulation. General scheme of chromosome contacts in interphase nuclei could be studied by 3C and Hi-C methods.In earlier work Fullwood et al. [1] used ChIA-PET technology to construct chromatin interaction network bound by estrogen receptor alpha from human breast cancer cell line and found long-range ER binding sites are mostly located at promoter regions. The problem of actual distance from a regulatory regionaffecting gene expression has been modeled using available chromosome contacts data. Li et al. [2] detected promoter-centered distant interactions bound by RNA Polymerase II in cancer cells using ChIA-PET technology. Important problem is analysis of methylation status of contacting chromosome regions.

We developed several computer programs for statistical data analysis and tested it on CTCF binding sites, genes and spatial topological domains. These data have been obtained experimentally by ChIP-seq, Hi-C, ChIA-PET methods.

We used data on the spatial domains in the genome of the mouse embryonic stem cells and in the human genome, data on the location of CTCF binding sites clusters obtained by ChIA-PET. Gene annotation was obtained from UCSC Genome Browser (http://genome.ucsc.edu). The result of the analysis is the distribution of CTCF transcription factor binding sites on domains of the human chromosomes and relative gene locations. The

distributions of human genes relative CTCF binding sites and a randomly generated list of such sites as the program output were used to estimate statistical significance of the associations found.

Chromatin interaction network is organized into "community", and genes within community perform related functions and respond to external stimuli in a coordinated manner. In all the promoter-nonpromoter interactions, more than 40% of the non-promoter regulatory elements didn't interact with their nearest promoters.

The work was supported by Russian Science Foundation (14-24-00123) (biophysical model of chromosome contacts) and ICG SB RAS budget project (0324-2016-0008) (computing support).

References

1. M.J. Fullwood et al. (2009) Anoestrogen-receptor-alpha-bound human chromatin interactome, *Nature*, 462(7269):58-64.

2. G. Li et al. (2014) Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application, *BMC Genomics*, 15(Suppl 12):S11.