

COMPUTER ANALYSIS OF GENOME CO-LOCALIZATION OF TRANSCRIPTION FACTOR BINDING SITES BASED ON CHIP-SEQ DATA

Dergilev A. I., Orlov Y. L.1

Novosibirsk State University, 630090, Novosibirsk, 2 Pirogova Str. arturd1993@yandex.ru
1Institute of Cytology and Genetics, 630090, Novosibirsk, Lavrentyeva 10,
orlov@bionet.nsc.ru

A scientific problem being solved is to study transcription factor binding sites (TFBS) co-localization in mammalian genomes using ChIP-seq data. Technology ChIP-seq, which combines chromatin immunoprecipitation (ChIP) and highly efficient DNA sequencing, allows to determine transcription factor binding sites in genome scale. The tasks of analyzing genome-wide ChIP-seq data rises are to identify the coordinates of TFBS and to compare their location with genomic annotation (relative location and distance to gene transcription start sites, promoter regions etc.). In addition to determining the location of binding sites for a transcription factor, there are problems of determining the cluster sites of different transcription factors, clusters together or located at a short (100-200 nt) distances on chromosomes assuming similar function and regulatory mechanisms. Programs processing huge amounts of text data (bed, wig files) identifying areas of intersection of genomic annotations (coordinates), adapted to the respective model genomes are technically necessary. High throughput sequencing technologies have enabled the identification of transcription factor (TF) binding sites in whole genomes. Important application is analysis of binding profiles in embryonic stem cells (ESCs). Somatic cells can be reprogrammed back to a pluripotent state by the combined introduction of transcription factors such as Oct4, Sox2, Klf4 and c-Myc. ChIP-seq data published recently and in the frames of ENCODE project allow construction of detailed genome binding maps. Such genome wide TF binding maps in mouse stem cells include Oct4, Sox2, Nanog, Tbx3, Smad2 as well as group of other factors. The signaling requirements for maintenance of human and murine embryonic stem cells (ESCs) differ considerably. Amongst the defined reprogramming factors, Oct4 is critical in inducing pluripotency.