

Developing of web-service WebMCOT for finding cooperative site-binding TF DNA-motifs

Mukhin A.M.^{1,2*}, Levitsky V.G.^{1,2}, Lashin S.A.^{1,2}

¹Novosibirsk State University, Novosibirsk

²Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* e-mail: mukhin@bionet.nsc.ru

Key words: transcription factors, transcription factor binding sites, composite element

Motivation and Aim: There are transcription and regulatory regions in eukaryotes genes. Transcription factors (TF) are proteins which regulate DNA transcription, and transcription factor binding sites (TFBS) are DNA sites for TF. TF action is cooperative. DNA sites are modelled by DNA motifs – position weight matrices where columns designate nucleotides, rows designate positions in a motif. Combinatorial binding of multiple TFs to a regulatory region provides fine-tuning of the gene expression; two TFBS in a composite element (CE) [1] are needed for starting transcription. Motifs can be overlapped or be separated with a spacer. The aim of this work is to develop a web-service WebMCOT (Web Motifs Co-Occurrence Tool [2]) for prediction of CEs in ChIP-seq data.

Materials and Methods: We used the following programming languages for each module: JavaScript and HTML for making web site, Python for web-service building, C++ for computational core. Also we used additional libraries: Vue.JS for making single page application, Flask for building REST API service, Celery for distribute task to another servers and cache-database Redis for optimal working with Celery. For checking C++ program and optimization memory, we used Jemalloc allocator with Jemalloc heap profiling.

Results: We have developed a web-service WebMCot for prediction of CEs in ChIP-seq data. There are 3 modules of this web-service: SPA, REST API web-service and statistical module. The SPA and the REST API web-service sending messages via HTTP-packets. The statistical module has been installed into the Institute of Cytology and Genetics HPC cluster. The web-service and the statistical module are connected via Redis database by the Celery library.

Conclusions: The GUI-based web-service has been implemented that can predict novel CEs.

Acknowledgements: The work was supported by Russian Foundation for Basic Research (project No. 18-29-13040).

References

1. Kel O.V., Romaschenko A.G., Kel A.E., Wingender E., Kolchanov N.A. A compilation of composite regulatory elements affecting gene transcription in vertebrates. *Nucleic Acids Res.* 1995;23:4097-4103.
2. Levitsky V.G., Oshchepkov D.Y., Zemlyanskaya E.V., Mironova V.V., Ignatieva E.V., Podkolodnaya O.A., Merkulova T.I. The overlapped motifs co-occurrence in ChIP-seq data. *Proceedings of the 11th International Conference on Bioinformatics of Genome Regulation and Structure\Systems Biology (BGRS\SB-2018)*. 2018;58.