

## Protein-protein and protein-DNA interaction are not obviously positively correlated

Sasha Belostotsky

*Institute for Information Transmission Problems (Kharkevich Institute) of Russian Academy of Sciences  
(Mathematic methods and models in bioinformatics lab), Bolshoy Karetny per. 19, Moscow, 127994, Russia,  
[alexbel.system@gmail.com](mailto:alexbel.system@gmail.com), [belostotsky@iitp.ru](mailto:belostotsky@iitp.ru)*

General point of view is that protein-protein interaction and protein–DNA interactions are correlated. It means that if protein-protein interaction data is known than it is possible to reconstruct binding sites on DNA. Opposite is supposed to be true as well. So, protein complexes and binding sites compositions must be positively correlated.

In order to check this many approaches using binding sites predictions for different transcription factors were applied. Protein-protein interaction data was taken from protein-protein interaction databases: IntAct, String, MINT, BIND. Genome regions were scanned using different scanning tools. One of most applied was Jaspar. All data was integrated in local MySQL database and all operation and queries were done by using Perl DBI.

Also ChIP-seq data was used from several sources. One of the most intensively used sources was UCSC Browser. It integrates all types of genomic data. In case of this study UCSC Genes (Knowngene) table was used, ChIP-seq regions (annotated as peaks) were exported from Yale directory for different cell lines (mostly for GM12878, K562, HeLa), DNase1 and FAIRE data was also used. All data was pre-mapped on hg18 assembly.

Neighboring predicted binding sites and number of Intersections between ChIP-seq regions were searched and protein-protein interaction data was taken into account. It occurred that interacting proteins bind neighboring binding sites and intersected ChIP-seq regions as much as not-interacting proteins bind it. Several possible explanations are discussed. The most possible two explanations are that protein-protein interaction data estimated by existing experimental methods is not valid or that post-translation modifications, underestimated in protein-protein interaction databases, lead to specificity of interactions of proteins with DNA (or both).

Acknowledgements: Author is grateful to Drs. Mironov AA and Gelfand MS for good recommendations.