Protein-protein interactions improve multiple transcription factor binding site prediction

Kirsti Laurila[1], Heidi Pukkila[1], Olli Yli-Harja[1], Harri Lähdesmäki[1,2]

[1]Department of Signal Processing, Tampere University of Technology, Finland;
[2]Department of Information and Computer Science, Helsinki University of Technology, Finland.

An important step in gene expression processes is transcriptional regulation. This regulation is largely controlled by transcription factors binding to DNA sites, chromatin structure and other epigenetic factors. As the identification of transcription factor binding sites with experimental methods can be very laborious, prediction methods are needed. Several prediction algorithms and programs have been developed but most of them either predict binding sites for a single factor at a time or they rely on prior knowledge of co-operating transcription factors [1]. Here, we present a probabilistic model that predicts binding of several transcription factors simultaneously with the help of transcription factor binding specificity models and protein-protein interactions [2].

Our method tries to mimic the situation in the nucleus by including in prediction several transcription factors and thus explicitly modeling the competition between the factors. As factors are known to interact with each other, we also add the prior knowledge of existing interactions to the model.

Our results show that our protein-protein interactions guided method performs better than the method without interactions or predictions where individual binding prediction results of separate TFs have been combined. The number of false positives is reduced remarkably compared with the individual predictions. Moreover, binding sites that were unpredictable with other methods could be identified with our protein-protein interaction guided method.

We have also studied how additional data sources, such as evolutionary conservation and nucleosome locations, affect the predictions.

[1] Hannenhalli . (2008), Eukaryotic transcription factor binding sites – modeling and integrative search methods. Bioinformatics, 24:1325-31.

[2] Laurila K, Yli-Harja O, Lähdesmäki H. (2009) A protein-protein interaction guided method for competitive transcription factor binding improves target predictions, *Nucleic Acids Res.*