Identifying transcription factor complexes and their roles

T. Will[1,2] and V. Helms[1]

[1]Center for Bioinformatics, and [2]Graduate School of Computer Science, Saarland University, Saarbrücken, Germany

Eukaryotic gene expression is controlled through molecular logic circuits that combine regulatory signals of many different factors. In particular, complexation of transcription factors and other regulatory proteins is a prevailing and highly conserved mechanism of signal integration within critical regulatory pathways and enables us to infer controlled genes as well as the exerted regulatory mechanism. Common approaches for protein complex prediction that only use protein interaction networks, however, are designed to detect self-contained functional complexes and have difficulties to reveal dynamic combinatorial assemblies of physically interacting proteins. We developed the novel algorithm DACO that combines protein-protein interaction networks and domain-domain interaction networks with the cluster-quality metric cohesiveness. The metric is locally maximized on the holistic level of protein interactions and connectivity constraints on the domain level are used to account for the exclusive and thus inherently combinatorial nature of the interactions within such assemblies. When applied to predicting transcription factor complexes in yeast, the proposed approach outperformed popular complex prediction methods by far. Furthermore, we were able to assign many of the predictions to target genes, as well as to a potential regulatory effect in agreement with literature evidence.

[1] Will, T. and Helms, V., Bioinformatics, Vol. 30 ECCB 2014, i415-i421 (2014).