Detecting regulatory protein complexes that define pluripotency

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 by molecular logic circuits that integrate regulatory signals of many different factors. In particular, complexation of transcription factors (TFs) and other regulatory proteins is a prevailing and evolutionary conserved mechanism of signal integration within critical regulatory pathways. Knowledge on the assembly of such complexes can enable us to infer the target genes that are cooperatively controlled as well as the exerted regulatory mechanisms of all proteins involved, including potentially recruited coregulators.

We demonstrated for TF complexes in yeast that combining protein interaction data with domain–domain interaction data by our algorithm DACO yields superior predictions of the combinatorial manifold of TF complexes compared to existing methods that are designed to detect self-contained functional modules. 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. Currently we are upscaling and expanding the capabilities of our software tools. To generate sample-specific interactome data as the input for DACO, for example, we subsequently developed the tool PPIXpress that exploits expression data at the transcript-level and is able to construct contextualized protein and domain interaction networks with isoform-resolution that even account for the effects of alternative splicing.

By inferring such specific interactomes for public data on human embryonic stem and iPS cells as well as other samples of the ENCODE and ROADMAP projects, we could predict the TF complexomes found in those cell states. Our most recent developments finally allow quantifying the abundance of the complexes per sample and enabled us to pin down a set of differential TF complexes of significantly higher abundance in pluripotent cells. Those particular complexes contain many known drivers of pluripotency and allowed us to construct a gene regulatory network of pluripotency that even considers cooperativity between proteins.