Integrative approaches for mode of action determination

Carolina Fierro¹, R. De Smet², Y. Wu¹, L. Cloots¹, D. De Maeyer¹, P. Zarrineh¹, Q. Fu¹, K. Engelen¹, K. Marchal¹,²

¹Department of Microbial and Molecular Systems, Katholieke Universiteit Leuven, Kasteelpark Arenberg 20, 3001 Leuven, Belgium

²Department of Plant Biotechnology and Bioinformatics, Ghent University, VIB, Technologiepark 927, 9052 Gent, Belgium

Problem statement:

In current wet-lab practice high-throughput experiments are becoming a standard to identify leads or genetic causes underlying complex phenotypes. High-throughput expression profiling techniques are used to find components of response pathways, e.g. mRNA profiling focuses on identifying genes that are differentially expressed following stimuli. Genetic screenings identify genetic hits of which manipulation alters the phenotype of the individual cell. With the availability of next-gen sequencing, complete genotyping in combination with expression analysis (eQTL) analysis allows finding genetic loci that underlie interesting phenotypic traits. Each of the experiments mentioned above eventually results in a list of candidate genes or loci that all relate to the same phenotype of interest.

Computational analysis of such unstructured lists will in most cases be restricted to a simple functional enrichment analysis to elucidate the potential pathways or functions the candidate genes belong to or to identify functionally related gene sets, but the interactions between the identified candidate genes or their role in the global molecular network of the cell remains unknown.

Insights in the relation between the different candidate genes and their role in the global regulatory network can only be obtained by exploiting the large body of publicly available “omics” datasets (expression datasets, DNA-protein interaction data, large scale protein-protein interaction studies and information on signaling networks) in addition to own experimental data.

Methods

In our work we developed an ensemble of tools that allow analyzing dedicated in-house experimental expression data and interrogating them with publicly available omics data and provided a proof of concept of our analysis flow on prokaryotic and eukaryotic model systems. To be able to maximally exploit the large body of publicly available expression data, we developed COLOMBOS a backend system to automatically generate cross-platform expression compendia for a particular model organism (Engelen et al., 2011). Such compendium consists of the expression value of all genes over thousands of conditions. To query such compendia with own gene lists, we developed an ensemble based query-based biclustering (De Smet et al., 2011). Identifying in a large compendium the genes coexpressed with the genes in the query list helps identifying the larger context of pathways and conditions under which
the genes of interest showed to be active (modules). Coclustering strategies (Van Deun et al., 2009) or explanatory models such as LeMoNe (DE Smet and Marchal, 2010; Michoel et al., 2009) help finding regulatory programs that can explain the expression profiles in these modules. Such programs typically consist of TFs or miRNAs of which the expression profile is related to that of the genes in the module. Integrative inference procedures such as DISTILLER (Lemmens et al., 2009) that combine information with expression data other sources of information such as regulatory motifs or ChIP-seq data complement previous approaches as they also allow identifying regulatory programs consisting of regulators of which the expression profile is not necessarily related to that of their targets. To detect to what extent identified modules are conserved across species we developed COMODO a coclustering strategy that combines homology relations and coexpression to identify conserved modules (Zarrineh et al., 2010). Lastly we developed a network based approach that allows overlaying coexpressed modules with a molecular interaction network of the organism of interest (Cloots et al., 2011).

References


