Poster L-24

Fuzzy puzzle of cis-elements in co-regulated promoters – the blueprint of key nodes in signal transduction networks.

Authors:
Nico Voss (*BIOBASE GmbH Halchtersche Str. 33, D-38304 Wolfenbüttel, Germany*)
Tatyana Konovalova (*Institute of Cytology and Genetics, Novosibirsk, Russia*)
Philip Stegmaier (*BIOBASE GmbH Halchtersche Str. 33, D-38304 Wolfenbüttel, Germany*)
Olga Kel-Margoulis (*r. 33, D-38304 Wolfenbüttel, Germany*)
Evelin Katona (*University Hospital Regensburg Franz-Josef-Strauss-Allee 11, D-93053 Regensburg, Germany*)
Thomas Langmann (*University Hospital Regensburg Franz-Josef-Strauss-Allee 11, D-93053 Regensburg, Germany*)
Gerd Schmitz (*University Hospital Regensburg Franz-Josef-Strauss-Allee 11, D-93053 Regensburg, Germany*)
Edgar Wingender (*Dept. Bioinformatics, UKG/Univ. Göttingen, Goldschmidtstr. 1, 37077 Göttingen, Germany*)
Alexander Kel (*BIOBASE GmbH Halchtersche Str. 33, D-38304 Wolfenbüttel, Germany*)

Short Abstract: Novel method for causal interpretation of gene expression data identifies composite cis-regulatory elements in promoters of co-regulated genes and reveals key transcription factors and “upstream” signaling molecules – perspective drug targets. Regulatory regions are modeled as fuzzy puzzle of TF binding sites providing blueprints to explain signal transduction networks.

Long Abstract:
Motivation: Cellular signal transduction networks of multicellular organisms are enormously complex though very robust in providing fast and appropriate response to any extracellular signal. This is achieved through combinatorial usage of a rather limited set of signaling molecules and pathways. These combinatorics must be mirrored by the structure of gene promoters as combinations of transcription factor binding sites (composite modules). Different signal transduction pathways leading to the activation of transcription factors converge at key molecules that master the regulation of certain cellular processes. Such crossroads of signaling networks often appear as “Achilles Heels” causing a disease when not functioning properly.

Results: We developed an integrated computational method for causal interpretation of gene expression data. It analyzes microarray data and proposes complexes of transcription factors as well as “upstream” key signaling molecules that master the observed gene expression profile. The method utilizes data from two databases (TRANSFAC® and TRANSPATH®) and integrates two programs:
1) Composite Module Analyst (CMA) analyzes 5’-upstream regions of co-expressed genes and applies a genetic algorithm to reveal composite modules (CMs) consisting of co-occurring single TF binding sites and composite elements. CMA models the structure of the long gene regulatory regions by a Boolean function that joins several local CMs. Having

as an input a set of co-regulated genes, CMA builds the promoter model and optimizes the parameters of the model automatically by applying a genetic-regression algorithm.
2) ArrayAnalyzer™ is a fast network analysis tool, to identify “upstream” key signaling molecules that are potentially involved in activation/inhibition of the “target set” of transcription factors. The program searches for so called “common nodes” located upstream in the signal flow in such a way that a signal from the common nodes can reach the maximal number of target molecules through a minimal number of reaction steps. We refer to these common nodes as “key signaling molecules”. The program proposes molecules that may provide coordinated signal transduction to the target transcription factors based on biological knowledge and can thereby identify “master regulators” of differentially expressed genes.

The method was applied to micrarray data on Pseudoxanthoma elasticum (PXE), a genetic disorder of the elastic fiber system which results in a skin disorder. Gene expression levels were measured in fibroblasts obtained from punch biopsies of non-affected areas from three PXE patients and compared to expression levels in normal fibroblasts (the comparison was conducted with Affymetrix software). The results obtained suggest a number of highly interesting biological hypotheses about molecular mechanisms of pathological genetic deregulation. First, with the help of CMA we constructed a promoter model for the PXE differentially expressed genes. The promoter model includes motifs of transcription factors: SOX-9, MAZ, P53, IRF, NF-kB, whose expression changed in PXE as well as of factors for which no significant expression change was observed, such as CREB, AP-2, DEC1, and SRF. We believe that differential activation of such factors is accomplished through changes in the expression of upstream components of signal transduction pathways leading to these factors.

Next, with the help of ArrayAnalyzer™ we explored the signal transduction pathways upstream of the TFs composing the promoter model. This analysis aimed at revealing key nodes of the network that can facilitate coordinate regulation of the target TFs and, hence, may resemble protopathogenic “master regulators”. A dense “network cluster” of transcription factors whose binding sites were overrepresented in promoters of PXE differentially expressed genes are reachable in the PKAc kinase pathway. PKAc expression differed 1.7-fold (down) in PXE compared to normal fibroblasts of our dataset. We speculate that these changes in PKAc expression, though rather small, can nevertheless alter the activity of downstream transcription factors SOX-9, MAZ, AP-1, NF-kB, CREB, HNF-4, which in turn may cause genetic disregulation leading to pathogenesis.

Conclusion: Here we describe a new method for analysis and interpretation of gene expression data that integrates classification of promoters of differentially expressed genes with analysis of signal transduction pathways. The logic of our strategy originates from the consideration that cellular sys-tems organize adequate responses to a large (possibly infinite) number of external signals on the basis of a rather limited set of genes, so that combinatorial actions of signal transducers have to be reflected in composite, “fuzzy” promoter structures. Combinations of transcription factors and composite regulatory elements/modules in their target promoters are the keys to this “fuzzy puzzle”. Hence, by analyzing promoters of genes disregulated in a certain disease we can attempt to “reverse engineer” the signal transduction network involved in regulating these genes and propose a cause of the disease at the molecular level.

We think that “fuzzy puzzle” principle of regulation is the result of genome evolution of multicellular organisms that shall overcome evolutionary bottlenecks caused by the requirement of multiple ontogenetic programs to be encoded in a single genome.

Availability: The CMA program is freely available for non-commercial users. URL: http://www.gene-regulation.com/pub/programs.html#CMAnalyst. It is also a part of the computer system ExPlain™ (www.biobase.de) designed for causal analysis of gene expression data.