Towards a Molecular Classification of Kidney Diseases Based on Network

Felix Eichinger[1], Ramakrishna Varadarajan[2], Jignesh Patel[2] and Matthias Kretzler[1].
University of Michigan, Ann Arbor, MI[1] and University of Wisconsin-Madison, Madison, WI [2].

Classification of patients based on conventional criteria such as histology and laboratory values in comprehensive datasets often show a significant discrepancy with patterns in gene expression. While this disconnect can easily be explained by the several layers of cellular machinery between gene expression and phenotype, even a description of the discrepancies is missing. To address this problem, we perform a molecular classification of a comprehensive gene expression dataset of 226 patients with 11 kidney related diseases states and link the results back to clinical and histological data. In detail, we select the regulated genes from each patient by comparison to a pool of healthy controls. To control for noise in the data and redundancy in gene function, we extend the gene lists to networks by adding edges representing cocitations of genes in PubMed abstracts and comparison of the resulting networks with an approximate graph matching algorithm (TALE). Subsequently we cluster the networks by similarity. Each cluster is analyzed for patterns either specific to the cluster or shared across clusters, and tested for homogeneity by appearance of patterns in the patients. Since patterns are hypothesized to indicate biological processes active in a subset of patients, we investigate the genes for interactions and strive to assign functional annotation to the patterns. Based on function assignment and knowledge of connections between biological processes and phenotype, we hypothesize about phenotypic effects and test on the clinical data. Embedding the cluster specific patterns in cross-cluster patterns enables integration into the common biological context.