Molecular Re-Classification of renal disease through approximate graph matching, clustering and pattern mining.

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

Classification of patients with a chronic disease course, such as kidney diseases, uses mainly descriptive disease definitions. To develop molecular based disease stratification, we aimed to define patient subgroups by conserved transcriptional networks. Defining similarity of patients on a regulatory network level, rather than on an individual gene level, might yield more robust indicators of function. Network nodes for each patient were derived from Affymetrix microarrays of kidney biopsies compared to healthy controls. Subsequently, relations between the nodes were established by natural language processing of PubMed abstracts and automated promoter analysis for transcription factor binding sites. The resulting networks are typically noisy or incomplete in nature; therefore network similarities are determined through an approximate graph-matching tool, allowing a degree of mismatching (within a preset threshold) in the displayed transcriptional networks. Based on a similarity score the patient networks are clustered - with the goal of attaining high intra-cluster similarity (networks within a cluster are highly similar) and low inter-cluster similarity (networks from different clusters are dissimilar). To extract underlying biological mechanism inside each cluster, we employ graph mining techniques and search for frequently occurring motifs (recurring subnetworks) within each cluster, indicative of characteristic disease processes (commonly occurring phenomenon within each cluster). Motifs across each cluster are compared to define mechanistic similarities and differences between network clusters. Finally, both clusters and motifs are matched back to the established descriptive clinical classifications to compare molecular and clinical classification.