

Mapping of Chemical and Biochemical Relationships of Mass Spectrometry-based Metabolomics Data

Dinesh Kumar; Tobias Kind; OLIVER FIEHN

UC Davis, Davis, CA

Keywords: Bioinformatics; Computational Methods; Database; Internet-Based; Network;

Novel Aspect: Use of open access biochemical databases and Cytoscape for mapping and visualization of MS-based metabolomics data.

Introduction

Mass spectrometry-based metabolomic studies generate quantitative data of the differential regulation of metabolites in response to genetic, environmental or physiological perturbations. Improved algorithms, larger mass spectral libraries and better instrumentations enable the identification of a larger number of metabolites in unbiased GC-MS or LC-MS runs. However, the evaluation of such data sets must not stop at the level of applying statistical tests to the data sets and delivering tables of regulated metabolites. Instead, metabolomics results must additionally convey a biological view of the data sets, enabling biochemical interpretations and structure the metabolites with respect to their chemical, biochemical and physiological roles in an analogous manner to gene ontologies in microarray studies.

Methods

Publicly available data sets of metabolite profiles were used that were downloaded from the SetupX/BinBase database (<http://fiehnlab.ucdavis.edu/>). This database comprises data matrices of approximately 300 metabolites and up to 1,400 chromatograms per study using a Leco PegasusIV GC-TOF mass spectrometer. A variety of open access tools from KEGG, Reactome, MetaCyc and ARM were used to map differential regulation of metabolites to biochemical network graphs. Pathway networks were downloaded as SBML and Biopax formats or created in Cytoscape SIF network format. Additionally, chemical structures were employed to calculate chemical similarity indices which were visualized by their resulting distances.

Preliminary results

Depending on the samples, 250-450 metabolic signals can be reliably and consistently detected in GC-TOF chromatograms. Of these, up to 150 metabolites were identified using retention index based mass spectral libraries and were referenced by KEGG and PubChem database identifiers. Three different approaches were compared for visualizing results from mass spectrometry-based metabolomic data sets: mapping to standard biochemical databases, visualization by chemical structure similarities and a hybrid approach employing atomic reconstruction of metabolism (ARM). The visualization results were then evaluated by the number of metabolic nodes that were represented in network graphs and the clarity of network clusters that should aid the interpretation of differential expression of metabolites. Common for mapping to biochemical databases was that not all identified metabolites were actually comprised by these databases. Approximately only 50% of the identified metabolites were represented by Reactome DB queries, 60% by MetaCyc queries and over 75% by KEGG and ARM queries. Specifically, lipids were often not correctly represented by these databases. Conversely, network graphs that were based on chemical distances can map 100% of the identified compounds, because each PubChem DB entry is associated with encoded chemical structures. Next, the cluster density and network graph layout was investigated. It was found that for any biochemical mapping approach, highly connected metabolites (e.g. ATP, CoA) have first to be removed to yield meaningful clusters. We can further conclude that graphs based on shortest-path mapping between two metabolite nodes do not yield biologically relevant clusters. One of the best and biologically most relevant graph representations was found when using the ARM distance maps, resulting in metabolite clusters according to known biochemical modules (lipids, carbohydrates, TCA cycle intermediates, aromatics, amino acids and urea cycle compounds). Interestingly, similar clusters were obtained when networks based on chemical structure distances were constrained by similarity thresholds.