Visualizing metabolomics data in directed biological networks

Denise Slenter[1], Martina Kutmon[1,2], **Ryan Miller**[1], Jonathan Melius[1], Georg Summer[3], Chris T. Evelo[1,2] and Egon L. Willighagen[1]

[1]Department of Bioinformatics – BiGCaT, NUTRIM Research School, Maastricht University, The Netherlands
[2]Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, The Netherlands
[3]Department of Cardiology, Maastricht University, The Netherlands
E-mail: ryan.miller@maastrichtuniversity.nl

Background

Metabolomics data describes the state of a biological system at a phenotypic level. Unfortunately, not all measured metabolites can be linked to metabolite identities present in biological pathway models. The resulting sparseness makes it more complicated to use metabolomics data in pathway and network analysis.

In 2014, Posma *et al.* introduced MetaboNetworks [1], a Matlab toolbox to create sub-networks between sets of metabolites using the reactions from the KEGG database [2] by calculating the shortest paths between them. Such networks overcome the metabolic data sparseness by focussing on the paths between the metabolites of interest. To upscale this approach, we need to be able to combine different pathway knowledge bases and introduce detailed directionality information, ensuring the shortest paths follow one-directional biological cause-and-effect paths.

Materials & Methods

The presented work creates subnetworks of shortest, directed pathways between active metabolites. First, with the WikiPathways RDF [3], we created a directed network of all metabolic reactions from the WikiPathways [4] and Reactome [5] pathway knowledgebase, see Figure 1. In the next step, we identified the location(s) of the active metabolites in the network, in which we match data with nodes in the network using knowledge from the ChEBI ontology [6] and Wikidata [7]. This ontological linking generalizes the more limited exact matching based on metabolite identifiers. Finally, using the cyNeo4j app for Cytoscape [8] we extracted the smallest connected subnetwork between the changed metabolites using the functionality of the graph database Neo4j.

We will apply the described approach to study the metabolic changes in diabetes patients reported in a publicly available dataset in the MetaboLights repository [9].

Conclusion

We developed a new solution to visualize the biological pathways involved in sparse metabolomics data. Using knowledge from two pathway resources and ontology-based approaches, we can show the directed networks between active metabolites from metabolomics data. The data from both resources is made interoperable by collapsing metabolites in the pathways into single nodes in the biological networks using ontological approaches. This explicit ontological linking allows for precise biological interpretation of the paths. By using Neo4j [10] and Cytoscape [11], we ensure the computational calculation environment for larger networks as well as advanced visualization functionality to investigate the identified subnetworks. The generic nature of this approach opens up the option to combine with other omics data sources, such as proteomics and transcriptomics.

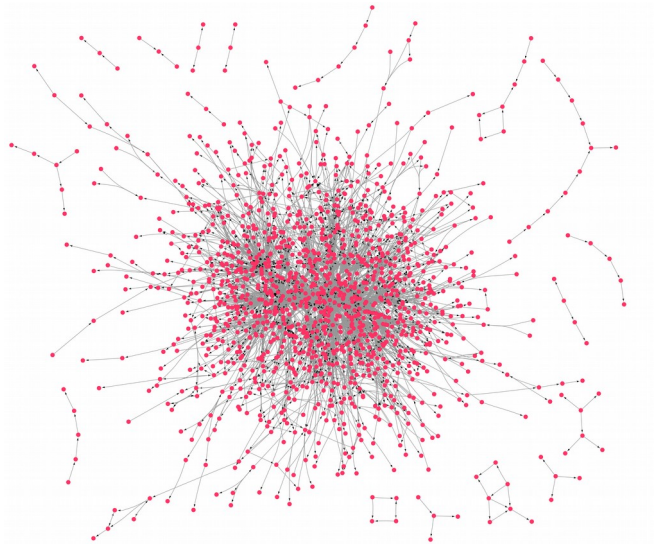

Figure 1: Directed network of metabolic reactions.

References

1. J.M. Posma *et al.* "MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG." Bioinformatics (2013): btt612.
2. M. Kanehisa *et al.* "KEGG: Kyoto Encyclopedia of Genes and Genomes." NAR 28.1 (2000): 27-30.
3. A. Waagmeester *et al.* "Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources." PLoS Comput Biol 12.6 (2016): e1004989.
4. M. Kutmon *et al.* "WikiPathways: capturing the full diversity of pathway knowledge." Nucleic acids research (2015): gkv1024.
5. G. Joshi-Tope *et al.* "Reactome: a knowledgebase of biological pathways." NAR 33.suppl 1 (2005): D428-D432.
6. K. Degtyarenko *et al.* "ChEBI: a database and ontology for chemical entities of biological interest." Nucleic acids research 36.suppl 1 (2008): D344-D350.
7. D. Vrandečić *et al.* "Wikidata: a free collaborative knowledgebase." Communications of the ACM 57.10 (2014): 78-85.
8. G. Summer *et al.* "cyNeo4j: connecting Neo4j and Cytoscape." Bioinformatics 31.23 (2015): 3868-3869.
9. K. Haug *et al.* "MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data." NAR (2012): gks1004.
10. J. Partner *et al.* Neo4j in action. Manning,, 2015.
11. P. Shannon *et al.* "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Research 13.11 (2003): 2498-2504.