Building directed networks from biological pathways to visualise and interpret metabolomics data

Denise Slenter[1], Martina Kutmon[1,2], Ryan Miller[1], Jonathan Melius[1], Chris T. Evelo[1,2] and Egon L. Willighagen[1]
[1]Dept. of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, The Netherlands
[2]Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, The Netherlands
E-mail: denise.slenter@maastrichtuniversity.nl

1. Introduction

Pathway diagrams are everywhere: in textbooks, articles, posters and whiteboards. Their utility to biologists as conceptual models is obvious, and can be useful for computational analysis and interpretation of large-scale experimental data, when properly modelled. Online pathway databases (e.g. WikiPathways [1], Reactome [2], KEGG [3]) provide rich, intuitive models of pathways.

Pathway analysis is widely adopted in the analysis of transcriptomics data. In experimental metabolomics data, however, many measured metabolites cannot be linked to the metabolites present in pathways databases. The resulting sparseness complicates the use of metabolomics data in pathway analysis.

2. Approach

Here, we present our approach (Figure 1) to calculate the shortest, directed paths between metabolites of interest. Using WikiPathways RDF [4], we created a directed network of all metabolic reactions from WikiPathways and Reactome. This network is stored in the graph database Neo4j (https://neo4j.com/) and enriched with knowledge from the ChEBI ontology [5] and Wikidata [6]. Using the cyNeo4j app [7] in Cytoscape [8], we are able to extract and visualise the smallest sub-network between the metabolites of interest and further study the processes involved.

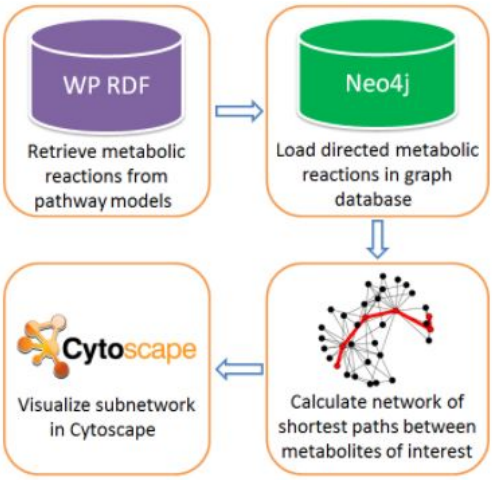

Figure 1. Workflow to visualise and interpret metabolic relevant sub-networks.

3. Discussion

We developed a solution to visualize the biological pathways involved in sparse metabolomics data, without losing their biological context. Using detailed models from online pathway databases and ontology-based approaches, we can extract the directed sub-networks between metabolites of interest. We tested our approach on several metabolomics datasets from the Metabolights repository [9], including measurements from mass-spectrometry and NMR and several biological fluids (e.g. blood, urine), related to metabolic changes due to aging. With the shortest path calculation and visualisation of the smallest sub-network, we aim to understand the biological mechanism underlying these metabolomics measurements separately, and try to find overlapping reactions and/or pathways for the combination of these datasets.

With the current approach however, the weight of the edges within the network is not taken into account, which could be calculated with protein kinetic data. This would account for the flux through the paths, and could give a biologically more plausible subnetwork [10]. We are exploring linking our metabolic pathways to flux and kinetics data [11], integrated with machine learning approaches to capture the dynamics of biological directed networks, and are interested in machine learning approaches to detect new pathways from experimental data.

References

1. Slenter *et al.* Nucleic Acids Res (2018) doi: 10.1093/nar/gkx1064
2. Fabregat *et al.* Nucleic Acids Res (2018) doi: 10.1093/nar/gkx1132
3. Kanehisa *et al.* Nucleic Acids Res (2000) PMID: 10592173
4. Waagmeester *et al.* PLoS Comput Biol (2016) doi: 10.1371/journal.pcbi.1004989
5. Degtyarenko *et al*. Nucleic Acids Res (2008) PMID: 17932057
6. Vrandečić *et al*. Communications of the ACM (2014) doi: 10.1145/2629489
7. Summer *et al.* Bioinformatics (2015) doi: 10.1093/bioinformatics/btv460
8. Shannon *et al*. Genome Research (2003) doi: 10.1101/gr.1239303
9. Haug *et al*. Nucleic Acids Res (2012) doi: 10.1093/nar/gks1004
10. Harush *et al*. Nat commun (2017) doi: 10.1038/s41467-017-01916-3
11. Benson *et al.* Brit J pharmacol. (2017) doi: 10.1111/bph.14037