Personalized metabolic analysis of diseases

Ali Cakmak, Muhammed Hasan Celik
Department of Computer Science
Istanbul Sehir University

The phenotype of diseases have often reflections on the metabolism of patients. Certain pathways may be boosted, while some others may experience activity decrease. Collectively, such changes may explain the etiology of a disease. In this paper, we propose an algorithm, Metabolitics, which quantifies the changes in activity levels of pathways (and their reactions) given concentration fold changes for a set of metabolites. A number of past studies (e.g., Drier et al., 2013; Wang et al., 2013) focused on pathway/reaction level analysis of high throughput biological data. The initial set of studies mainly employed pathway enrichment analysis (Wang et al., 2013). The next set of methods (Drier et al., 2013) directly transfers the measured metabolite/gene changes to their corresponding pathways without any filtering. Then, statistical significance analysis is done on the pathways and their change levels to assign them a deregulation score. These methods consider each pathway as a collection of genes/metabolites, and ignore the interactions between these entities. An extension of the above set of methods also takes into account pathway topology (Vaske et al., 2010). More specifically, some measured genes/metabolites are given more weights in the statistical significance analysis based on the centrality of each gene/metabolite in the pathway topology.

None of the above summarized pathway analysis methods considers that pathways are part of a larger network, and they interact with each other. The main novelty of the proposed method is that we perform the analysis on the whole pathway network in a holistic manner, rather than considering each pathway in an isolated manner. The main advantage of such an approach is that, for a given disease, it allows to identify those key player pathways for which there are few or no associated gene/metabolite measurements in the analyzed omics data. Our method is not specific to the metabolomics domain. It may be easily extended to for the analysis of other types of omics data (e.g., mRNA) as well.

Metabolitics assigns a score for each pathway/reaction in each patient. This score represents how much the activity of the corresponding pathway/reaction differs from that of healthy individuals. In order to achieve this, Metabolitics works on the whole network of metabolic pathways. It turns the analysis task into an optimization problem (Orth et al., 2010) where the objective is dynamically set to maximize the flux for increasing metabolites' reactions, and minimize the flux for decreasing metabolites' reactions in proportion to their fold changes. Then, the optimization problem is solved with linear programming (Fernández-Castané et al., 2014). Since the optimization problem is under-determined (Orth et al., 2010), there are usually multiple solutions. In order to accommodate multiple solutions with a single score, we employ flux variability analysis (Labhsetwar et al., 2013) to identify the lower and upper bound flux values for each pathway. The average lower and upper flux values of healthy individuals are considered as reference values, and for each individual and pathway, Metabolitics compute how much the lower and upper flux values differ from the reference values in a given patient. The average of the changes in upper and lower flux bounds is assigned to each patient-pathway pair as the "diff" score.

In order to evaluate the Metabolitics algorithm, we apply it on a breast cancer metabolomics dataset. We demonstrate that Metabolitics (i) captures biologically relevant information in breast cancer, (ii)

accurately stratifies people with breast cancer, (iii) is robust to decrease in the amount of measurement data, and (iv) provides more coverage than the state of the art. Table 1 lists the top-10 significantly changing pathways with their diff scores in comparison to the state of the art.

Pathway	F-val	p-val	diff	Pathifier	Paradigm	Met. Cnt
Alanine and aspartate metabolism	200	1.70E-31	1300	0.16	0.0012	7
Arginine and Proline metabolism	160	1.90E-26	850	0.46	-0.019	8
Methionine and cysteine met.	130	2.50E-23	170	0.17	0.024	6
Taurine and hypotaurine met.	130	1.00E-22	970	-	0.18	1
CoA catabolism	120	2.50E-22	610	-	-	0
Fatty acid oxidation	120	2.40E-21	-1100	0.21	-	15
Nucleotide interconversion	110	2.70E-20	1300	-	-	1
Eicosanoid met.	80	6.30E-16	-730	0.5	-	4
Butanoate met.	69	3.40E-14	-670	-	-	0
Glycolysis	68	4.60E-14	770	0.11	-0.019	5
Std. Deviation			883.5	0.17	0.08	

Table 1. Significantly changing pathways in breast cancer with their diff scores computed by Metabolitics. The statistical significance values (i.e., F-val and p-val) are computed through ANOVA analysis based on computed diff values (corrected for multiple hypothesis testing using Benjamini/Hochberg). The last column reports the measured metabolite counts for the reported pathways. Columns 5-6 provides the scores computed by the state of art approaches Pathifier and Paradigm.

References

Drier, Y., Sheffer, M., & Domany, E. (2013). Pathway-based personalized analysis of cancer. Proceedings of the National Academy of Sciences, 110(16), 6388-6393.

Fernández-Castané, A., Fehér, T., Carbonell, P., Pauthenier, C., & Faulon, J. L. (2014). Computer-aided design for metabolic engineering. Journal of biotechnology, 192, 302-313.

Labhsetwar, P., Cole, J. A., Roberts, E., Price, N. D., & Luthey-Schulten, Z. A. (2013). Heterogeneity in protein ex-pression induces metabolic variability in a modeled Escherichia coli population. Proceedings of the National Academy of Sciences, 110(34), 14006-14011.

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis?. Nature biotechnology, 28(3), 245-248.

Vaske, C. J., Benz, S. C., Sanborn, J. Z. et al. (2010). Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics, 26(12), i237-i245.

Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). WEB-based gene set analysis toolkit (WebGestalt): update 2013. Nucleic acids research, 41(W1), W77-W83.