Published in final edited form as:

Curr Metabolomics. 2013 January 1; 1(1): 83–91. doi:10.2174/2213235X11301010084.

Analyzing LC/MS metabolic profiling data in the context of existing metabolic networks

Tianwei Yu and

Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA

Yun Bai

Department of Pharmaceutical Sciences, School of Pharmacy, Philadelphia College of Osteopathic Medicine, Suwanee, GA

Tianwei Yu: tianwei.yu@emory.edu; Yun Bai: yunba@pcom.edu

Abstract

Metabolic profiling is the unbiased detection and quantification of low molecular-weight metabolites in a living system. It is rapidly developing in biological and translational research, contributing to disease mechanism elucidation, environmental chemical surveillance, biomarker detection, and health outcome prediction. Recent developments in experimental and computational technology allow more and more known metabolites to be detected and quantified from complex samples. As the coverage of the metabolic network improves, it has become feasible to examine metabolic profiling data from a systems perspective, i.e. interpreting the data and performing statistical inference in the context of pathways and genome-scale metabolic networks. Recently a number of methods have been developed in this area, and much improvement in algorithms and databases are still needed. In this review, we survey some methods for the analysis of metabolic profiling data based on metabolic networks.

Background

The complex biological system consists of multiple functional levels – genome, transcriptome, proteome, and metabolome. Probing the metabolome offers a unique perspective in understanding the linkage between genotype and phenotype¹. It is the most informative level in the study of environmental effects on the organism²⁻⁴. In clinical settings, metabolic profiling offers critical information about a person's metabolic status, which compensates genomic, proteomic and other information to paint the overall picture of a person's health status and disease risk⁵. When a pharmacological compound is involved, its interaction with the human system can be revealed by metabolomics studies⁶⁻⁷. The field of metabolomics has been rapidly developing in the post-genome era⁸. In addition to utilizing metabolomic technology alone to study mechanisms and find biomarkers, more and more studies integrate metabolomics with other omics technologies to bring insights into the complex regulatory networks of the multi-level system⁹⁻¹².

There are two major platforms in metabolomic studies – nuclear magnetic resonance (NMR) and mass spectrometry (MS)¹³. In NMR data, many of the signals overlap, making it difficult to identify and quantify individual metabolites¹⁴. The signal overlap issue can be reduced by multidimensional techniques. Still the method is very useful in measuring abundant metabolites and following their dynamics. On the other hand, mass spectrometry, when coupled with separation techniques such as liquid chromatography, can provide quantification to thousands of metabolites in a single sample¹³. Due to its feasibility for

complex biological samples, e.g. plasma and urine, liquid chromatography – mass spectrometry (LC/MS) has become a favored technique in metabolic profiling^{8, 13, 15-16}.

Bottom-up and top-down metabolomics

Bottom-up metabolomics start with known metabolites and pathways. By quantifying the metabolites longitudinally or measuring their fluxes in a targeted manner, the bottom-up approaches attempt to build dynamic models that can simulate and predict complex behaviors of key pathways¹⁷. Complementing the bottom-up approaches, the top-down approaches start with untargeted profiling of low molecular weight components in a system. The identified features may correspond to various ionic forms of known metabolites, previously uncharacterized metabolites, food components, and environmental chemicals. After detection and quantification, they can be matched to databases to find their identities, or further characterized using techniques such as LC/MS/MS. The top-down approach helps discover previously unknown metabolites and their relations with components of the known metabolic network, as well as previously undiscovered patterns between metabolites¹⁸⁻¹⁹.

Recently, top-down profiling of human metabolome is gaining wider and wider use in biomedical research. In a typical liquid chromatograph-Fourier Transform mass spectrometry (LC/FTMS) analysis, roughly 4,000 features can be identified from a single human plasma sample¹⁷. While a large portion of the features are not matched to the KEGG database, those that match to KEGG cover about 25% of known human metabolites in the KEGG database²⁰. When a dual chromatography LC/FTMS procedure is used, a greater number (~7,000) features can be detected¹⁷. Major efforts have been made to improve the feature extraction, quantification, and identification. A number of algorithms and software have been developed that greatly facilitate the data processing of LC/MS data²¹⁻³⁷. As this is not the focus of the review, we will skip the details. In the following discussion, we assume the data has been properly preprocessed, and high confidence matches between detected features and known metabolites are available.

Although the capability of metabolic profiling of complex samples has been greatly improved, there are still major difficulties in the study of human metabolome. Firstly, the plasma/urine metabolome is heavily impacted by food intake, the presence of medicine and environmental chemicals. Secondly, when studying metabolomics in a cohort of subjects, variations of the metabolome along the time of day, such as diurnal variation, may cause an increase in measurement noise. Thirdly, the plasma/urine metabolome represent an average effect of multiple tissues, rather than being the product of a homogeneous group of cells. These and other factors cause large variations between subjects, as well as difficulty in mechanistic interpretations of results. With these problems, the analysis of metabolic profiling data could greatly benefit from the utilization of existing knowledge on genome-scale metabolic network and pre-defined metabolic pathways. By utilizing such information, information is pooled from metabolites that are functionally related, and the impact of measurement noise can be reduced.

Pathway/metabolite set testing

In the related field of gene expression analysis, a number of methods were designed for the so-called “gene set analysis”. Groups of functionally related genes are first determined based on functional annotations. Then at the gene set level, statistical testing is conducted in order to determine if a gene set is associated with the outcome variable³⁸⁻⁴⁴. In the past few years, similar strategies have been applied to metabolomics data (Fig. 1). The methods are largely borrowed from gene set analysis. Such methods use functional grouping of metabolites, but ignore the detailed topological relations within a pathway.

In the field of gene (or feature) set analysis, there are generally two types of methods (Fig. 2). A number of methods were designed to compare the distribution of certain test statistics that reflect the gene set-wise behavior in response to a certain outcome variable⁴¹⁻⁴⁴. Gene sets are not homogeneous, i.e. not all genes in a gene set are regulated at the expression level in a given physiological condition. Some methods were developed to select subsets of genes simultaneously with the generation of gene set statistics⁴⁵⁻⁴⁶, utilize covariance structure in testing⁴⁷, or incorporate regulatory network connectivity⁴⁸. Some of the gene set analysis methods have been previously reviewed and compared^{38-40, 49-52}. Most of the gene set differential expression analysis methods use permutation tests to assess the significance of gene sets. They can be further classified into two sub-classes – those based on gene label permutation (competitive hypotheses), and those based on sample permutation (self-contained hypotheses)³⁸⁻⁴⁰. The methods using competitive hypotheses have been criticized for their properties of ignoring within-gene set correlation structure, tendency to yield false-positives on gene sets with high internal correlation, and difficulty in statistical interpretation³⁸⁻⁴⁰. On the other hand, methods using self-contained hypotheses require certain sample size to reach enough significance level, and could identify too many gene sets as significant when the number of differentially expressed genes is large⁴⁰. A second class of gene set analyses, which we call gene set differential coordination analysis, tries to capture relationships that are not revealed by the analysis of differential expression, by testing within-gene set co-expression changes between treatment groups⁵³, differential coexpression between pairs of gene sets under different conditions⁵⁴, or changes in coordination between a gene set with all other genes⁵⁵.

All the aforementioned gene set methods can potentially be applied to the analysis of metabolite sets. Some software packages have been published specifically for metabolomics data, most of which involve functionality of matching features to known metabolites by database search. MBRole applies the enrichment analysis to metabolite sets⁵⁶. MPEA considers the many-to-many relations when matching features to known metabolites⁵⁷. MSEA and MetaboAnalyst use multiple gene set analysis methods⁵⁸⁻⁵⁹. MetPA applies gene set analysis methods, and considers network topology by incorporating centrality measures⁶⁰. PAPI is a new algorithm specifically designed for metabolic pathways, which defines a pathway activity score for each treatment condition⁶¹.

Testing on the network

The methods mentioned in the previous section consider each pathway or metabolite set independently from other parts of the genome-scale metabolic network. However, the metabolic network is highly inter-connected, and some pathways are artificially carved out from the entire network. Thus there is a strong interest in analyzing high-throughput data in the context of the entire network, without pre-determining sub-networks to study (Fig. 1). To achieve this goal, computational techniques have been developed in both transcriptomics and metabolomics. Again, the two disciplines can borrow models and insights from each other.

In order to incorporate network topology into the analysis of gene expression data, a number of methods were proposed. A graph-structured two sample test finds if a subgraph is differentially expressed under different treatment conditions⁶². A Markov Random Fields model was proposed to use the pathway structures in identifying differentially expressed genes and important sub-networks⁶³. This work was later extended to consider longitudinal behavior of genes to analyze time course gene expression data⁶⁴. Bayesian linear regression models were developed to use the dependence structure of transcription factors and/or genes in the network to aid gene selection⁶⁵⁻⁶⁶. Network-based penalized regression model was developed for subnetwork selection⁶⁷. The Omis Response Group (ORG) method starts

with a list of selected genes, potentially generated from between-treatment group testing, and seeks pathways (response groups) that are highly connected in a certain direction to the list of query genes. Using metabolomics data, Cakir *et al* proposed a method that finds “reporter reactions”, which are reactions whose neighboring genes on average show high levels of change between treatment groups ⁶⁸.

With current technology, there are still major drawbacks in testing metabolomics data on the genome-scale metabolic network. One major issue is the limited coverage. Cakir *et al* attempted to address this issue by producing a reduced genome-scale model, both eliminating unmeasured metabolites and reducing enzyme sets into virtual single reactions ⁶⁸. Another issue is the uncertainty in the mapping of detected features to known metabolites. Due to the sharing of m/z values, there are often instances of one-to-multi or multi-to-multi matching. This issue is yet to be addressed in network testing. A third issue is sample origin. Metabolomics data in translational research is most often generated from blood plasma. While each tissue type activates different portions of the overall metabolic network ⁶⁹, the plasma metabolome is influenced by multiple tissues. Thus it becomes questionable whether the behaviors of two metabolites on the genome-scale network are indeed mechanistically linked.

Joint analysis of metabolomics and transcriptomics data

Metabolite concentrations are regulated by the activities of enzymes that catalyze corresponding reactions. Both metabolites and enzymes are intrinsic parts of the metabolic network. Gene expression data is also linked to other networks, such as the transcriptional regulatory network ⁷⁰⁻⁷¹. With the development of both metabolic and gene expression profiling techniques, there is a strong interest in analyzing metabolomics data jointly with gene expression data, in order to boost the statistical power to detect perturbed pathways, and shed light on the regulation resulting in the changes of metabolite levels (Fig. 3).

A simple approach is analyzing the two types of data separately, match each to pathways, and then merge the results at the pathway level ⁷²⁻⁷³. Joint testing of pathways using the hypergeometric distribution and independence assumption between different data types was also proposed ⁷⁴. In the analysis of time course data, the Granger causality was used to find time lag effects between gene expression and metabolite levels ⁷⁵. A method using the orthogonal partial least square discriminant analysis (OPLS-DA) to combine multiple coresponse statistics between transcripts and metabolites was proposed to detect relations between genes and metabolic pathways ⁷⁶. The integrated metabolome and interactome mapping (iMIM) method incorporates metabolic network with protein interaction network and protein-RNA interactions, to reach a model linking a set of differential metabolites and a target gene ⁷⁷. A web service, IMPaLA, is available for the joint enrichment analysis of pathways using user-specified metabolite and gene lists ⁷⁸.

Besides the methods mentioned above, a number of publications deal with interpreting transcriptomics data in conjunction with the genome-scale metabolic network. In term of their goals, these methods partially overlap with those mentioned in the previous two sections. However these methods are focused on properties of the metabolic network, and are more adapted to its characteristics. The mixture model on graphs (MMG) is a Bayesian method for sub-network selection that can incorporate weights and directionality of the network ⁷⁹. The KEGG spider maps differentially expressed gene list onto the KEGG network, and finds sub-graphs by connecting the genes through the shortest path between them on the graph ⁸⁰. Tissue-specific metabolic behavior was predicted by maximizing the agreement between the expression levels of the enzyme genes and the predicted flux activity

in the tissue ⁸¹. Patil and Nielsen defined “reporter metabolites” as those surrounded by differentially expressed enzymes ⁸².

Metabolic network databases and network reconstruction efforts

A number of databases are available summarizing current knowledge of the genome-scale metabolic network of multiple organisms. The Kyoto Encyclopedia of Genes and Genomes (KEGG) ²⁰ may be the most widely used in metabolomics data analysis. Its pathway database contains small molecule reactions, as well as organism-specific enzymes that are known to catalyze the reactions. The ConsensusPathDB is an integration of different types of interactions, built from various sources ⁸³. It contains not only metabolic reactions, but also protein interactions, signal transduction pathways etc. The small molecule pathway database (SMPDB) is focused on human pathways ⁸⁴. Detailed information, such as tissue specificity, subcellular locations, and cofactors *etc* are available in the database. Pathway Commons is a comprehensive pathway database incorporating multiple types of biological networks ⁸⁵. The UniPathway ⁸⁶ links metabolic reactions to Uniprot ⁸⁷ enzyme database. Other relevant databases that can help annotate metabolites include the human metabolome database (HMDB) ⁸⁸, Madison Metabolomics Consortium Database (MMCD) ⁸⁹, and the METLIN Metabolite Database ⁹⁰.

The existing network databases represent the current state of the knowledge on the metabolic networks. There are on-going efforts to refine and improve the knowledge based on new experimental and computational developments. An important theme is to utilize genomic sequence information to refine the knowledge on metabolic networks. To name a few examples, Duarte et al reported a human metabolic network reconstruction based on both the literature and genomic information, in which compartmentalization was predicted by targeting signals in gene sequences and protein localization data ⁹¹. Hao *et al* reconstructed the Edinburgh Human Metabolic Network (EHMN) by including transport reactions and compartmentalization predicted by subcellular location of enzymes ⁹². For the human metabolic network, efforts were made in tissue-specific network reconstruction ⁹³⁻⁹⁴. Herrgard et al built a consensus yeast metabolic network by merging information about yeast metabolism from several sources ⁹⁵.

Visualization of metabolomics data using the genome-scale metabolic network

Visual examination is an intuitive approach for the analysis of high-throughput data. Several software are available for the visualization of metabolomics data, often in conjunction with relatively simple testing capabilities. Metscape ⁹⁶⁻⁹⁷ is a visualization tool associated with the popular Cytoscape software ⁹⁸. It allows users to upload gene and metabolite lists with measurement values, and builds network based on the relations between the entities stored in the database. It also conducts enrichment analysis using pathway information. Another software, VANTED, conducts data analysis such as Pearson's correlation, self-organizing map (SOM), and two-sample testing ⁹⁹. Then the results are visualized in the network context. The Omics Viewer is a component of the Pathway Tools visualization software ¹⁰⁰. The software includes metabolic, signal transduction and transport pathways. The Omics Viewer paints metabolomics measurements and/or gene expression onto the network. Paintomics is a web service that visualizes gene expression and metabolomics data jointly onto the KEGG pathway maps ¹⁰¹.

Conclusion

The booming field of metabolomics represents tremendous opportunities of new knowledge discovery, including potential disease biomarkers, drug targets, and environmental impacts on human health. Improvements in metabolic profiling technologies yield more and more comprehensive and accurate coverage of the metabolome. Being able to analyze the high-throughput data from the systems biology perspective is critical to maximize information retrieval from such data. The development and application of new computational techniques generate new insights into the regulation of the metabolic system and its association with diseases. Advances in other omics areas are opening up new opportunities for the development of joint modeling methods.

Acknowledgments

This work was partially supported by NIH grants P20 HL113451 and P01 ES016731.

References

1. Fiehn O. Metabolomics--the link between genotypes and phenotypes. *Plant Mol Biol.* 2002; 48(1-2): 155–71. [PubMed: 11860207]
2. Viant MR, Bearden DW, Bundy JG, Burton IW, Collette TW, Ekman DR, Ezernieks V, Karakach TK, Lin CY, Rochfort S, de Ropp JS, Teng Q, Tjeerdema RS, Walter JA, Wu H. International nmr-based environmental metabolomics intercomparison exercise. *Environ Sci Technol.* 2009; 43(1): 219–25. [PubMed: 19209610]
3. Pluskal T, Nakamura T, Villar-Briones A, Yanagida M. Metabolic profiling of the fission yeast *S. Pombe*: Quantification of compounds under different temperatures and genetic perturbation. *Mol Biosyst.* 6(1):172–88.
4. Park YH, Lee K, Soltow QA, Strobel FH, Brigham KL, Parker RE, Wilson ME, Sutliff RL, Mansfield KG, Wachtman LM, Ziegler TR, Jones DP. High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. *Toxicology.* 2012; 295(1-3):47–55. [PubMed: 22387982]
5. Yang SY, Adelstein J, Kassisi AI. Putative molecular signatures for the imaging of prostate cancer. *Expert Rev Mol Diagn.* 10(1):65–74.
6. Boudonck KJ, Mitchell MW, Nemet L, Keresztes L, Nyska A, Shinar D, Rosenstock M. Discovery of metabolomics biomarkers for early detection of nephrotoxicity. *Toxicol Pathol.* 2009; 37(3):280–92. [PubMed: 19380839]
7. Sun J, Schnackenberg LK, Beger RD. Studies of acetaminophen and metabolites in urine and their correlations with toxicity using metabolomics. *Drug Metab Lett.* 2009; 3(3):130–6. [PubMed: 19702550]
8. Dunn WB. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys Biol.* 2008; 5(1):11001.
9. Fan TW, Higashi RM, Lane AN. Integrating metabolomics and transcriptomics for probing anticancer mechanisms. *Drug Metab Rev.* 2006; 38(4):707–32. [PubMed: 17145697]
10. Moreno-Risueno MA, Busch W, Benfey PN. Omics meet networks--using systems approaches to infer regulatory networks in plants. *Curr Opin Plant Biol.* 2009
11. Waterman CL, Currie RA, Cottrell LA, Dow J, Wright J, Waterfield CJ, Griffin JL. An integrated functional genomic study of acute phenobarbital exposure in the rat. *BMC Genomics.* 11(1):9. [PubMed: 20053287]
12. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, Hamalainen E, Jousilahti P, Kangas AJ, Mannisto S, Savolainen MJ, Jula A, Leiviska J, Palotie A, Salomaa V, Perola M, Ala-Korpela M, Peltonen L. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol.* 2010; 6:441. [PubMed: 21179014]
13. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev.* 2007; 26(1):51–78. [PubMed: 16921475]

14. Watkins SM, German JB. Toward the implementation of metabolomic assessments of human health and nutrition. *Curr Opin Biotechnol.* 2002; 13(5):512–6. [PubMed: 12459346]
15. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. Analytical and statistical approaches to metabolomics research. *J Sep Sci.* 2009; 32(13):2183–99. [PubMed: 19569098]
16. Griffin JL, Kauppinen RA. A metabolomics perspective of human brain tumours. *Febs J.* 2007; 274(5):1132–9. [PubMed: 17298437]
17. Jones DP, Park Y, Ziegler TR. Nutritional metabolomics: Progress in addressing complexity in diet and health. *Annu Rev Nutr.* 2012; 32:183–202. [PubMed: 22540256]
18. Bruggeman FJ, Westerhoff HV. The nature of systems biology. *Trends Microbiol.* 2007; 15(1):45–50. [PubMed: 17113776]
19. Mardinoglu A, Nielsen J. Systems medicine and metabolic modelling. *J Intern Med.* 2012; 271(2):142–54. [PubMed: 22142312]
20. Kanehisa M, Goto S. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28(1):27–30. [PubMed: 10592173]
21. Sturm M, Bertsch A, Gropl C, Hildebrandt A, Hussong R, Lange E, Pfeifer N, Schulz-Trieglaff O, Zerck A, Reinert K, Kohlbacher O. Openms - an open-source software framework for mass spectrometry. *BMC bioinformatics.* 2008; 9:163. [PubMed: 18366760]
22. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem.* 2003; 75(18):4818–26. [PubMed: 14674459]
23. Windig W, Smith WF. Chemometric analysis of complex hyphenated data. Improvements of the component detection algorithm. *J Chromatogr A.* 2007; 1158(1-2):251–7. [PubMed: 17418223]
24. Idborg-Bjorkman H, Edlund PO, Kvalheim OM, Schuppe-Koistinen I, Jacobsson SP. Screening of biomarkers in rat urine using lc/electrospray ionization-ms and two-way data analysis. *Anal Chem.* 2003; 75(18):4784–92. [PubMed: 14674455]
25. Katajamaa M, Miettinen J, Oresic M. Mzmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics.* 2006; 22(5):634–6. [PubMed: 16403790]
26. Tolstikov VV, Lommen A, Nakanishi K, Tanaka N, Fiehn O. Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Anal Chem.* 2003; 75(23):6737–40. [PubMed: 14640754]
27. Bellew M, Coram M, Fitzgibbon M, Igra M, Randolph T, Wang P, May D, Eng J, Fang R, Lin C, Chen J, Goodlett D, Whiteaker J, Paulovich A, McIntosh M. A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution lc-ms. *Bioinformatics.* 2006; 22(15):1902–9. [PubMed: 16766559]
28. Hastings CA, Norton SM, Roy S. New algorithms for processing and peak detection in liquid chromatography/mass spectrometry data. *Rapid Commun Mass Spectrom.* 2002; 16(5):462–7. [PubMed: 11857732]
29. Aberg KM, Torgrip RJ, Kolmert J, Schuppe-Koistinen I, Lindberg J. Feature detection and alignment of hyphenated chromatographic-mass spectrometric data. Extraction of pure ion chromatograms using kalman tracking. *J Chromatogr A.* 2008; 1192(1):139–46. [PubMed: 18378252]
30. Stolt R, Torgrip RJ, Lindberg J, Csenki L, Kolmert J, Schuppe-Koistinen I, Jacobsson SP. Second-order peak detection for multicomponent high-resolution lc/ms data. *Anal Chem.* 2006; 78(4):975–83. [PubMed: 16478086]
31. Yu T, Park Y, Johnson JM, Jones DP. Aplcms--adaptive processing of high-resolution lc/ms data. *Bioinformatics.* 2009; 25(15):1930–6. [PubMed: 19414529]
32. Yu T, Peng H. Quantification and deconvolution of asymmetric lc-ms peaks using the bi-gaussian mixture model and statistical model selection. *BMC bioinformatics.* 2010; 11:559. [PubMed: 21073736]
33. Takahashi H, Morimoto T, Ogasawara N, Kanaya S. Amdorap: Non-targeted metabolic profiling based on high-resolution lc-ms. *BMC bioinformatics.* 2011; 12:259. [PubMed: 21702951]
34. Yang ZR, Grant M. An ultra-fast metabolite prediction algorithm. *PLoS One.* 2012; 7(6):e39158. [PubMed: 22745711]

35. Brown M, Wedge DC, Goodacre R, Kell DB, Baker PN, Kenny LC, Mamas MA, Neyses L, Dunn WB. Automated workflows for accurate mass-based putative metabolite identification in lc/ms-derived metabolomic datasets. *Bioinformatics*. 2011; 27(8):1108–12. [PubMed: 21325300]
36. Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S. Camera: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical chemistry*. 2012; 84(1):283–9. [PubMed: 22111785]
37. Varghese RS, Zhou B, Nezami Ranjbar MR, Zhao Y, Ressom HW. Ion annotation-assisted analysis of lc-ms based metabolomic experiment. *Proteome Sci*. 2012; 10(Suppl 1):S8. [PubMed: 22759585]
38. Goeman JJ, Buhlmann P. Analyzing gene expression data in terms of gene sets: Methodological issues. *Bioinformatics*. 2007; 23(8):980–7. [PubMed: 17303618]
39. Liu Q, Dinu I, Adewale AJ, Potter JD, Yasui Y. Comparative evaluation of gene-set analysis methods. *BMC Bioinformatics*. 2007; 8:431. [PubMed: 17988400]
40. Nam D, Kim SY. Gene-set approach for expression pattern analysis. *Brief Bioinform*. 2008; 9(3): 189–97. [PubMed: 18202032]
41. Efron B, Tibshirani R. On testing the significance of sets of genes. *Ann Appl Stat*. 2007; 1:107–29.
42. Keller A, Backes C, Gerasch A, Kaufmann M, Kohlbacher O, Meese E, Lenhof HP. A novel algorithm for detecting differentially regulated paths based on gene set enrichment analysis. *Bioinformatics*. 2009; 25(21):2787–94. [PubMed: 19713416]
43. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. Gage: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*. 2009; 10:161. [PubMed: 19473525]
44. Oron AP, Jiang Z, Gentleman R. Gene set enrichment analysis using linear models and diagnostics. *Bioinformatics*. 2008; 24(22):2586–91. [PubMed: 18790795]
45. Chen X, Wang L, Smith JD, Zhang B. Supervised principal component analysis for gene set enrichment of microarray data with continuous or survival outcomes. *Bioinformatics*. 2008; 24(21):2474–81. [PubMed: 18753155]
46. Wu MC, Zhang L, Wang Z, Christiani DC, Lin X. Sparse linear discriminant analysis for simultaneous testing for the significance of a gene set/pathway and gene selection. *Bioinformatics*. 2009; 25(9):1145–51. [PubMed: 19168911]
47. Tsai CA, Chen JJ. Multivariate analysis of variance test for gene set analysis. *Bioinformatics*. 2009; 25(7):897–903. [PubMed: 19254923]
48. Bussemaker HJ, Ward LD, Boorsma A. Dissecting complex transcriptional responses using pathway-level scores based on prior information. *BMC Bioinformatics*. 2007; 8(Suppl 6):S6. [PubMed: 17903287]
49. Abatangelo L, Maglietta R, Distaso A, D'Addabbo A, Creanza TM, Mukherjee S, Ancona N. Comparative study of gene set enrichment methods. *Bmc Bioinformatics*. 2009; 10:275. [PubMed: 19725948]
50. Dinu I, Liu Q, Potter JD, Adewale AJ, Jhangri GS, Mueller T, Einecke G, Famulsky K, Halloran P, Yasui Y. A biological evaluation of six gene set analysis methods for identification of differentially expressed pathways in microarray data. *Cancer Inform*. 2008; 6:357–68. [PubMed: 19259416]
51. Song S, Black MA. Microarray-based gene set analysis: A comparison of current methods. *Bmc Bioinformatics*. 2008; 9:502. [PubMed: 19038052]
52. Dinu I, Potter JD, Mueller T, Liu Q, Adewale AJ, Jhangri GS, Einecke G, Famulski KS, Halloran P, Yasui Y. Gene-set analysis and reduction. *Brief Bioinform*. 2009; 10(1):24–34. [PubMed: 18836208]
53. Choi Y, Kendzierski C. Statistical methods for gene set co-expression analysis. *Bioinformatics*. 2009; 25(21):2780–6. [PubMed: 19689953]
54. Cho SB, Kim J, Kim JH. Identifying set-wise differential co-expression in gene expression microarray data. *BMC Bioinformatics*. 2009; 10:109. [PubMed: 19371436]
55. Yu T, Bai Y. Capturing changes in gene expression dynamics by gene set differential coordination analysis. *Genomics*. 2011
56. Chagoyen M, Pazos F, Mbrole. Enrichment analysis of metabolomic data. *Bioinformatics*. 2011; 27(5):730–1. [PubMed: 21208985]

57. Kankainen M, Gopalacharyulu P, Holm L, Oresic M. Mpea--metabolite pathway enrichment analysis. *Bioinformatics*. 2011; 27(13):1878–9. [PubMed: 21551139]
58. Xia J, Wishart DS. Msea: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res*. 2010; 38(Web Server issue):W71–7. [PubMed: 20457745]
59. Xia J, Wishart DS. Web-based inference of biological patterns, functions and pathways from metabolomic data using metaboanalyst. *Nat Protoc*. 2011; 6(6):743–60. [PubMed: 21637195]
60. Xia J, Wishart DS. Metpa: A web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics*. 2010; 26(18):2342–4. [PubMed: 20628077]
61. Aggio RB, Ruggiero K, Villas-Boas SG. Pathway activity profiling (papi): From the metabolite profile to the metabolic pathway activity. *Bioinformatics*. 2010; 26(23):2969–76. [PubMed: 20929912]
62. Jacob L, Neuvial P, Dudoit S. More power via graph-structured tests for differential expression of gene networks. *Ann Appl Stat*. 2012; 6:561–600.
63. Wei Z, Li H. A markov random field model for network-based analysis of genomic data. *Bioinformatics*. 2007; 23(12):1537–44. [PubMed: 17483504]
64. Wei Z, Li H. A hidden spatial-temporal markov random field model for network-based analysis of time course gene expression data. *Ann Appl Stat*. 2008; 2(1):408–429.
65. Stingo F, Chen Y, Tadesse M, Vannucci M. Incorporating biological information into linear models: A bayesian approach to the selection of pathways and genes. *Ann Appl Stat*. 2011; 5:1978–2002. [PubMed: 23667412]
66. Li F, Zhang N. Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *Journal of the American Statistical Association*. 2010; 105:1202–1214.
67. Pan W, Xie B, Shen X. Incorporating predictor network in penalized regression with application to microarray data. *Biometrics*. 2010; 66(2):474–84. [PubMed: 19645699]
68. Kaker T, Patil KR, Onsan Z, Ulgen KO, Kirdar B, Nielsen J. Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol Syst Biol*. 2006; 2:50. [PubMed: 17016516]
69. Sugimoto M, Ikeda S, Niigata K, Tomita M, Sato H, Soga T. Mmmdb: Mouse multiple tissue metabolome database. *Nucleic Acids Res*. 2012; 40(Database issue):D809–14. [PubMed: 22139941]
70. Kim TM, Park PJ. Advances in analysis of transcriptional regulatory networks. *Wiley Interdiscip Rev Syst Biol Med*. 2011; 3(1):21–35. [PubMed: 21069662]
71. Yu T, Li KC. Inference of transcriptional regulatory network by two-stage constrained space factor analysis. *Bioinformatics*. 2005; 21(21):4033–8. [PubMed: 16144806]
72. Kaplan F, Kopka J, Sung DY, Zhao W, Popp M, Porat R, Guy CL. Transcript and metabolite profiling during cold acclimation of arabidopsis reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content. *Plant J*. 2007; 50(6):967–81. [PubMed: 17461790]
73. Kresnowati MT, van Winden WA, Almering MJ, ten Pierick A, Ras C, Knijnenburg TA, Daran-Lapujade P, Pronk JT, Heijnen JJ, Daran JM. When transcriptome meets metabolome: Fast cellular responses of yeast to sudden relief of glucose limitation. *Mol Syst Biol*. 2006; 2:49. [PubMed: 16969341]
74. Cavill R, Kamburov A, Ellis JK, Athersuch TJ, Blagrove MS, Herwig R, Ebbels TM, Keun HC. Consensus-phenotype integration of transcriptomic and metabolomic data implies a role for metabolism in the chemosensitivity of tumour cells. *PLoS Comput Biol*. 2011; 7(3):e1001113. [PubMed: 21483477]
75. Walther D, Strassburg K, Durek P, Kopka J. Metabolic pathway relationships revealed by an integrative analysis of the transcriptional and metabolic temperature stress-response dynamics in yeast. *OMICS*. 2010; 14(3):261–74. [PubMed: 20455750]
76. Redestig H, Costa IG. Detection and interpretation of metabolite-transcript coresponses using combined profiling data. *Bioinformatics*. 2011; 27(13):i357–65. [PubMed: 21685093]
77. Davidovic L, Navratil V, Bonaccorso CM, Catania MV, Bardoni B, Dumas ME. A metabolomic and systems biology perspective on the brain of the fragile × syndrome mouse model. *Genome Res*. 2011; 21(12):2190–202. [PubMed: 21900387]

78. Kamburov A, Cavill R, Ebbels TM, Herwig R, Keun HC. Integrated pathway-level analysis of transcriptomics and metabolomics data with impala. *Bioinformatics*. 2011; 27(20):2917–8. [PubMed: 21893519]
79. Sanguinetti G, Noirel J, Wright PC. Mmg: A probabilistic tool to identify submodules of metabolic pathways. *Bioinformatics*. 2008; 24(8):1078–84. [PubMed: 18292114]
80. Antonov AV, Dietmann S, Mewes HW. Kegg spider: Interpretation of genomics data in the context of the global gene metabolic network. *Genome Biol*. 2008; 9(12):R179. [PubMed: 19094223]
81. Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol*. 2008; 26(9):1003–10. [PubMed: 18711341]
82. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A*. 2005; 102(8):2685–9. [PubMed: 15710883]
83. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. Consensuspathdb: Toward a more complete picture of cell biology. *Nucleic Acids Res*. 2011; 39(Database issue):D712–7. [PubMed: 21071422]
84. Frolkis A, Knox C, Lim E, Jewison T, Law V, Hau DD, Liu P, Gautam B, Ly S, Guo AC, Xia J, Liang Y, Shrivastava S, Wishart DS. Smpdb: The small molecule pathway database. *Nucleic Acids Res*. 2010; 38(Database issue):D480–7. [PubMed: 19948758]
85. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res*. 2011; 39(Database issue):D685–90. [PubMed: 21071392]
86. Morgat A, Coissac E, Coudert E, Axelsen KB, Keller G, Bairoch A, Bridge A, Bougueleret L, Xenarios I, Viari A. Unipathway: A resource for the exploration and annotation of metabolic pathways. *Nucleic acids research*. 2012; 40(Database issue):D761–9. [PubMed: 22102589]
87. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. The universal protein resource (uniprot). *Nucleic Acids Res*. 2005; 33(Database issue):D154–9. [PubMed: 15608167]
88. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong Y, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythe I. Hmdb: A knowledgebase for the human metabolome. *Nucleic acids research*. 2009; 37(Database issue):D603–10. [PubMed: 18953024]
89. Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, Westler WM, Eghbalian HR, Sussman MR, Markley JL. Metabolite identification via the madison metabolomics consortium database. *Nat Biotechnol*. 2008; 26(2):162–4. [PubMed: 18259166]
90. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. Metlin: A metabolite mass spectral database. *Ther Drug Monit*. 2005; 27(6):747–51. [PubMed: 16404815]
91. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*. 2007; 104(6):1777–82. [PubMed: 17267599]
92. Hao T, Ma HW, Zhao XM, Goryanin I. Compartmentalization of the edinburgh human metabolic network. *BMC bioinformatics*. 2010; 11:393. [PubMed: 20649990]
93. Gille C, Bolling C, Hoppe A, Bulik S, Hoffmann S, Hubner K, Karlstadt A, Ganeshan R, Konig M, Rother K, Weidlich M, Behre J, Holzthutter HG. Hepatonet1: A comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol Syst Biol*. 2010; 6(411)
94. Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using init. *PLoS Comput Biol*. 2012; 8(5):e1002518. [PubMed: 22615553]
95. Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo ML, Oliveira

- AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasic I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttila M, Klipp E, Palsson BO, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol.* 2008; 26(10):1155–60. [PubMed: 18846089]
96. Karnovsky A, Weymouth T, Hull T, Tarcea VG, Scardoni G, Laudanna C, Sartor MA, Stringer KA, Jagadish HV, Burant C, Athey B, Omenn GS. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics.* 2012; 28(3):373–80. [PubMed: 22135418]
97. Gao J, Tarcea VG, Karnovsky A, Mirel BR, Weymouth TE, Beecher CW, Cavalcoli JD, Athey BD, Omenn GS, Burant CF, Jagadish HV. Metscape: A cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. *Bioinformatics.* 2010; 26(7):971–3. [PubMed: 20139469]
98. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics.* 2011; 27(3):431–2. [PubMed: 21149340]
99. Junker BH, Klukas C, Schreiber F. Vanted: A system for advanced data analysis and visualization in the context of biological networks. *BMC bioinformatics.* 2006; 7:109. [PubMed: 16519817]
100. Paley SM, Karp PD. The pathway tools cellular overview diagram and omics viewer. *Nucleic Acids Res.* 2006; 34(13):3771–8. [PubMed: 16893960]
101. Garcia-Alcalde F, Garcia-Lopez F, Dopazo J, Conesa A. Paintomics: A web based tool for the joint visualization of transcriptomics and metabolomics data. *Bioinformatics.* 2011; 27(1):137–9. [PubMed: 21098431]

Figure 1. An illustration of the typical workflow of metabolite set analysis and network testing using metabolic profiling data.

Figure 2.
Types of feature set level analyses.

Figure 3.

An illustration of the multi-layer network subject to joint-analysis. For simplicity, this picture is an over-simplified illustration. The gene/protein level can be further separated into several layers, e.g. genetic variations, gene expression, and protein level.