frontiers in
MOLECULAR BIOSCIENCES

REVIEW ARTICLE
published: 06 November 2014
doi: 10.3389/fmolb.2014.00022

Linking metabolomics data to underlying metabolic regulation

Thomas Nägele*

Department of Ecogenomics and Systems Biology, University of Vienna, Vienna, Austria

Edited by:
Guowang Xu, Chinese Academy of Sciences, China

Reviewed by:
Adam James Carroll, The Australian National University, Australia
Michal Jan Markuszewski, Medical University of Gdansk, Poland

****Correspondence:***
Thomas Nägele, Department of Ecogenomics and Systems Biology, University of Vienna, Althanstraße 14, 1090 Vienna, Austria
e-mail: thomas.naegele@univie.ac.at

The comprehensive experimental analysis of a metabolic constitution plays a central role in approaches of organismal systems biology. Quantifying the impact of a changing environment on the homeostasis of cellular metabolism has been the focus of numerous studies applying various metabolomics techniques. It has been proven that approaches which integrate different analytical techniques, e.g., LC-MS, GC-MS, CE-MS and H-NMR, can provide a comprehensive picture of a certain metabolic homeostasis. Identification of metabolic compounds and quantification of metabolite levels represent the groundwork for the analysis of regulatory strategies in cellular metabolism. This significantly promotes our current understanding of the molecular organization and regulation of cells, tissues and whole organisms. Nevertheless, it is demanding to elicit the pertinent information which is contained in metabolomics data sets. Based on the central dogma of molecular biology, metabolite levels and their fluctuations are the result of a directed flux of information from gene activation over transcription to translation and posttranslational modification. Hence, metabolomics data represent the summed output of a metabolic system comprising various levels of molecular organization. As a consequence, the inverse assignment of metabolomics data to underlying regulatory processes should yield information which—if deciphered correctly—provides comprehensive insight into a metabolic system. Yet, the deduction of regulatory principles is complex not only due to the high number of metabolic compounds, but also because of a high level of cellular compartmentalization and differentiation. Motivated by the question how metabolomics approaches can provide a representative view on regulatory biochemical processes, this article intends to present and discuss current metabolomics applications, strategies of data analysis and their limitations with respect to the interpretability in context of biological processes.

Keywords: metabolomics, GC-MS, LC-MS, systems biology, metabolic regulation, cellular compartmentalization, multivariate statistics, mathematical modeling

INTRODUCTION

Systems biology has become a rapidly growing research field aiming at a comprehensive representation of complex biological systems. Metabolomics plays a central role in systems biology as it provides essential information about the metabolome, i.e., the metabolic constitution and the dynamic behavior of metabolite levels. The combination of chromatographic techniques and mass spectrometry detection has enabled the rapid and precise high-throughput analysis of up to hundreds, or even thousands, of metabolic compounds from the same sample (Hall et al., 2002). Yet, the full scientific potential of metabolomics techniques is still limited due to a considerable variation in annotation confidence (Creek et al., 2014). As discussed by Creek and coworkers, one possibility to gain relatively high confidence is the comparison of multiple physicochemical properties of an authentic pure chemical standard to those of the metabolite of interest. Techniques like the comprehensive GCxGC—*time of flight mass spectrometry* (GCxGC-TOFMS), where two columns with orthogonal separation characteristics are combined, yield a much higher peak capacity (Almstetter et al., 2012) and may help increase the identification confidence by resolving co-eluting compounds. But also on the level of mass spectrometry such co-elutions might be resolved, for instance by applying techniques of tandem-MS or MS^n (Mei-Ling et al., 2006). The resulting data matrix, which typically comprises compounds of the central carbon/nitrogen metabolism, i.e., sugars, amino acids, organic acids and lipids, characterizes a metabolic homeostasis or its perturbation-induced dynamics (Kaplan et al., 2004; Leon et al., 2013; Aldridge and Rhee, 2014). Chemical derivatization broadens the spectrum of metabolites which can be assessed by GC techniques making them become volatile and thermally stable (Roessner et al., 2000). A commonly used method is a two-step derivatization comprising oximation and silylation. While the in oximation step sugars are stabilized in an open ring conformation, the silylation step stabilizes molecules by replacing hydrogen in functional polar groups, e.g., the hydroxyl group, by a trimethylsilyl group [$-Si(CH_3)_3$] (Hill and Roessner, 2013). While GC-MS particularly enables the quantification of volatile and uncharged compounds, Liquid chromatography coupled to mass spectrometry (LC-MS) is the method of choice to analyse

semi- or non-volatile and thermally unstable compounds (Hopfgartner and Varesio, 2013). Both GC-MS and LC-MS might be applied to analyse isomeric compounds, but the high chromatographic resolution of GC, and particularly GCxGC, make it a suitable analytical technique to resolve structurally closely related compounds (Meinert and Meierhenrich, 2012). In LC-MS, structural information about molecules can be obtained by collision-induced dissociation (Jennings, 2000). Different techniques in mass spectrometry as well as their characteristic features are summarized elsewhere (e.g., see Weckwerth, 2011), but it can be generalized that the combination of those techniques is expected to increase the coverage of a metabolome significantly.

Despite identification confidence clearly being the limiting factor, the output of a GC- and LC-MS platform yields a comprehensive view of a metabolic homeostasis and its response to genomic or environmental perturbation. Beyond, there are further analytical techniques, such as UV, IR, FT-IR, and FT-ICR spectroscopy (Van Agthoven et al., 2013), nuclear magnetic resonance (NMR) (Simmler et al., 2014) or capillary electrophoresis (Kuehnbaum and Britz-Mckibbin, 2013), which can even enlarge this metabolic information space and increase its confidence. The metabolic coverage yielded by metabolomics approaches depends on the organism and sample type being analyzed. While the prokaryotic metabolome of *E. coli* comprises about 750 metabolites (Nobeli et al., 2003), eukaryotic metabolomes have been described to range from more than 1000 (e.g., yeast Herrgard et al., 2008) over several thousands (e.g., humans Duarte et al., 2007) up to tens or even hundreds of thousands in plants (Hall et al., 2002). Particularly in eukaryotic organisms, the interpretation of metabolomics data is complex not only due to the high number of metabolic compounds but also because of a high level of cellular compartmentalization, different cell types, tissues and organs. In a metabolomics study on the alga *Chara australis* the subcellular localization and dynamics of 125 metabolites was analyzed revealing a stress-induced asynchronous fluctuation of metabolite levels in the vacuole and cytosol (Oikawa et al., 2011). Based on their findings the authors suggested that metabolite levels are regulated separately in intracellular compartments. Also in higher plants, several studies have focused the subcellular analysis of metabolite dynamics and underlying fluxes (Masakapalli et al., 2010, 2013; Klie et al., 2011; Krueger et al., 2011; Nägele and Heyer, 2013; Szecowka et al., 2013; Arrivault et al., 2014). As it is a characteristic feature of eukaryotic cells, the experimental analysis of subcellular organization of metabolism cannot be over assessed. Due to the high information content of studies resolving the subcellular level, it is possible to unravel unexpected features of metabolic regulation. A concrete example for the need of subcellular resolution is the existence of plastidial and cytosolic pathways for carbohydrate oxidation, i.e., glycolysis or the oxidative pentose phosphate pathway (PPP). In a study of subcellular flux analysis in a heterotrophic Arabidopsis cell suspension using steady-state stable isotope labeling, it has been shown that multiple data sets can be fitted successfully to models with an altered subcellular compartmentation of the PPP (Masakapalli et al., 2010). With their approach, the authors provide evidence for the importance of experimental data on the subcellular level in order to reduce the uncertainty about the interpretation of biochemical regulatory processes. Another comprehensive example for subcellular analysis of leaf metabolism in Arabidopsis was provided recently by using a strategy of $^{13}CO_2$-labeling to resolve time-dependent patterns and kinetics in the metabolome (Szecowka et al., 2013). Changes in isotope patterns of 40 primary metabolites were analyzed using LC- or GC-MS comprising central carbohydrates, organic, and amino acids as well as phosphorylated intermediates, e.g., from the Calvin-Benson cycle. While many of the experimental findings were according to previous expectations, several unexpected features of labeling kinetics could be unraveled due to the subcellular resolution (Szecowka et al., 2013). Finally, the authors conclude that information about metabolite compartmentation is a prerequisite for modeling photosynthetic and/or other metabolic processes in multicellular eukaryotic tissues.

The above mentioned studies present reasonable and comprehensive approaches to assess complex systems in a proportionally detailed manner. However, there exists a clear discrepancy between the number of metabolites which are absolutely quantified (∼10–100) or identified and (relatively) quantified (∼100–1000), and which are expected to be found in a metabolome based on genome-wide predictions (∼10^3–10^5). Beyond the objective of increasing the coverage and confidence of analytical metabolomics platforms it is a particular challenge to interpret the resulting experimental information in a biochemical meaningful way which is the prerequisite for the successful generation of a testable hypothesis. As outlined above, this is mainly due to the high information content of underlying molecular and structural organization being compiled in a metabolomics data set. In this context, the following chapter strives to outline open questions and different data evaluation strategies applied and developed in the metabolomics research field.

DERIVING REGULATORY STRATEGIES FROM METABOLIC SNAPSHOTS

The difficulty of interpreting comprehensive metabolomics data sets can easily be retraced by a simple example: an experiment comprising samples of two groups, e.g., a wild type (control group) and a *knock out* mutant (treatment group), typically results in m independent biological replicates each with n technical replicates for each group. Let p denote the number of metabolites being quantified in the metabolomics experiment. Then the summary of all experimental data within this experimental design ends up in two data arrays each with $p \times n \times m$ dimensions. Assuming that technical variance of the method or platform has been shown to be much lower than biological variance ($var_{tech} << var_{biol}$), a technical mean value can be built for each biological replicate and the data arrays reduce to data matrices with $p \times m$ dimensions. Although "basic" univariate statistical methods and tests immediately allow for a direct comparison of different data sets and provide a first idea of underlying biological mechanisms, generated hypotheses frequently suffer from ambiguity due to various biochemical explanations for one metabolomic feature. The covariance matrix of a metabolomics data set helps to quantify such an ambiguity. It is a statistical measure for how components, i.e., metabolites, are related to each other (Equation 1).

$$Cov(x, y) = \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y}) \quad (1)$$

x and y denote variables (*here:* metabolite levels) with a sample size m and a mean value $\bar{x}$ and $\bar{y}$. Accordingly, a $p \times p$ covariance matrix is derived from a $p \times m$ data matrix. This quadratic relationship between the number of metabolites and possible (metabolic) interactions exemplarily demonstrates why even small metabolomics data sets, i.e., $p < 10$, are difficult to interpret without appropriate mathematical methods: increasing the metabolomic coverage of experimental approaches is automatically linked to an exponential increase in possible explanations which have to be considered. Thus, to find the most reasonable and comprehensive biochemical explanation for all resolved experimental data, the simultaneous application of various statistical methods which confirm or even complement each other has become a common approach. Basic statistical methods are well-established and numerous platforms and graphical user interfaces (GUIs) have been developed to enable and facilitate the application of such methods [for an overview of available tools and software packages see e.g., (Sugimoto et al., 2012) or (Sheth and Thaker, 2014)]. Most of these statistical platforms even go beyond and are capable of analysing, resolving and visualizing multivariate problems, applying methods like the principal component analysis (PCA), the independent component analysis (ICA) (Steuer et al., 2007), or even the independent PCA (Yao et al., 2012). Based on the (co-)variance information, these multivariate techniques reduce the dimensionality of the dataset while retaining as much as information, i.e., variance, as possible. This yields abstract variables, i.e., components, which imply the most pronounced and characteristic features of multivariate data sets, thus enabling and facilitating the generation of biological hypotheses. Complemented with other unsupervised and supervised statistical methods, e.g., hierarchical clustering, *k*-means clustering and partial least squares discriminant analysis (PLS-DA), molecular compounds contributing to the separation of the samples by a varying degree, can be identified (Okada et al., 2010; Westerhuis et al., 2010; Le Cao et al., 2011; Korman et al., 2012; Sun and Weckwerth, 2012; Bellaire et al., 2014; Madala et al., 2014; Uarrota et al., 2014).

The abovementioned strategies of (multivariate) data analysis are an integral part of most systems biology approaches comprising not only metabolomics but, mostly, also other data sets derived from various "omics" techniques, for example transcriptomics and proteomics (Weckwerth, 2008). Special software packages have been developed for the integrative analysis of various experimental high-throughput data sets aiming at a comprehensive regression analysis, correlation and visualization. An excellent overview of software tools, platforms and workflows of computational tasks in systems biology was provided previously (Ghosh et al., 2011). Although these platforms provide a wide variety of statistical and mathematical methods, the final outcome which a system biologist is interested in may always be similar: the identification of specific metabolic clusters and patterns, i.e., samples with similar statistical characteristics, separating the control from a treatment group. This finally provides us with an idea about which metabolic steps or pathways are most likely affected by the introduced perturbation and conclusions about biochemical or molecular biological mechanisms can be drawn resulting in a testable hypothesis (**Figure 1**). Numerous studies from various fields of biological research have proven the applicability and usefulness of such computational workflows (Altaf-Ul-Amin et al., 2014). But contemperaneously, deriving information on the regulatory interaction between levels of molecular organization has been recognized to be accessible only to a very limited degree. This is mainly due to the complex and non-linear relationship which exists between levels of transcripts, proteins, and metabolites. Hence, if the level of a metabolite changes due to a genomic or environmental perturbation this may have various reasons: (i) feed-back/forward regulation of enzymes (with non-linear kinetics), (ii) posttranslational modification of enzymes/proteins involved in the metabolite synthesis, interconversion or transport, (iii) changes in the rate of protein biosynthesis and/or degradation, and so on. As a consequence, drawing an unambiguous conclusion directly from a set of metabolite levels on regulatory strategies in a metabolic network is hardly possible—except for the particular case that (A) all components of the metabolic network are known and biochemically characterized, or (B) all relationships between levels of molecular organization can—due to simplification—be considered as linear. Although during recent years, a lot of information about genome-wide metabolic network structures and regulatory principles has been unraveled, comprising, to name only a few, molecular systems of prokaryotes (Carrera et al., 2014), yeast (Sanchez et al., 2014), algae (Chang et al., 2011), higher plants (Mintz-Oron et al., 2012; Hill et al., 2013), human metabolism (Mardinoglu et al., 2013), and disease and medicine (Vandamme et al., 2013), we are still far away from being able to fulfill all—or at least most of the—requirements being necessary for approach (A). In contrast, strategies of mathematical modeling enable the procedure of linearization which is explained and discussed in the following paragraph.

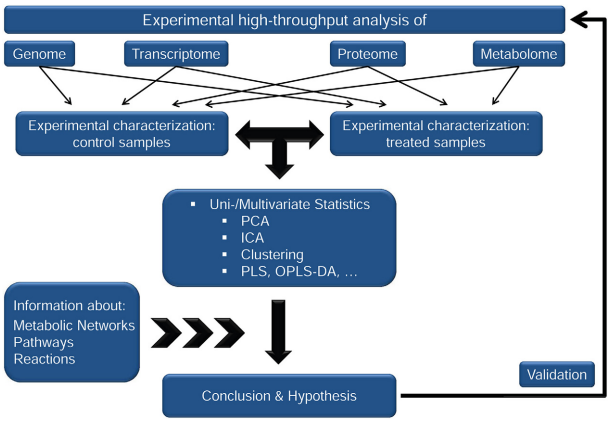

FIGURE 1 | A typical systems biology workflow to derive and test hypotheses from experimental high-throughput analysis.

LINEARIZATION OF METABOLIC FUNCTIONS

The linearization of non-linear functions around a certain state is a frequently applied approach in mathematics, physics or control

theory to approximate solutions of complex and chaotic systems (Strogatz, 1994). In a very figurative context one could think of a polynomial function $y(x)$, containing several minima and maxima. Yet, instead of considering and analysing the function $y(x)$ for all x, we are only interested in the behavior of $y(x)$ in one certain point x_0. In this point, the function is described by the coordinates $(x_0, y(x_0))$. We can approximate the function $y(x)$ by drawing the tangent in $(x_0, y(x_0)$. With this, we have approximated the original function by a linear function. In a metabolic context, x_0 would represent a certain steady-state metabolite concentration while $y(x)$ represents the so-called metabolite function taking the steady-state value $y(x_0)$. In the following paragraphs, the time-dependent metabolic function will be re-written as $f_M(t)$.

Comparing the linearized approximation with the original function it becomes obvious that these solutions are only valid in a very narrow and predefined interval of the fundamental function. However, they still provide a highly informative insight and can help to trace back basic principles of the original function—particularly if this function is highly complex. Ultimately, they may allow for the simulation and prediction of the systems behavior, thereby broadening the current knowledge about the origin of the system's complexity.

To exemplarily transfer this strategy to a systems biology approach, we consider a time-dependent change in the concentration of a metabolite M. Mathematically this can be expressed by an ordinary differential equation (ODE). While the left side of the equation describes the changes in M with changes in time t, the right side of the equation describes all (reaction) rates affecting the concentration of M. This can be summarized by the metabolic function $f_M(t)$. The reaction rates again depend on various parameters, variables and functions, such as inhibitor/activator concentrations, thermodynamic constraints, posttranslational modification, protein levels, and so on. This automatically connects the regulation of metabolite levels to all other levels of molecular organization (**Figure 2**). Each of the functions contains various non-linear elements, e.g., enzyme kinetics (Michaelis-Menten/Hill/...) or thermodynamic equations (Arrhenius/...), resulting in a highly non-linear description of the biological system. Applying the principle of linearization allows for the characterization of such a non-linear system around an experimentally analyzed metabolic (steady) state by replacing the non-linear with linear functions. In a simple two-dimensional example this can be illustrated by the tangent in one point of the original metabolic function (**Figure 3A**). Mathematically this is performed by differentiating the metabolic function with respect to a reaction variable, such as time, metabolites or other parts of the abovementioned molecular organization (see **Figure 2**). If this is performed for all considered system variables at the same metabolic steady-state, the partial derivatives are combined in the so-called Jacobian matrix (**Figure 3B**). By inducing perturbations and analysing the deflection of the metabolic system, characteristic features, like stability of stead-states or sensitivity of fluxes, can be estimated and summarized in the Jacobian matrix (Steuer et al., 2006; Chen and Chen, 2009; Reznik and Segre, 2010). While the mathematical theory behind these approaches is fully established (Strogatz, 1994) and is commonly applied in engineering sciences (Föllinger and Konigorski, 2013), the direct integration of experimental (metabolomics) data is still challenging. In recent work, we have focused on the development of methods for a direct integration of metabolomics data, merging the covariance information with a genome-wide metabolic network structure. Thereby, we could successfully link experimental metabolomics data to strategies of subcellular compartmentation (Nägele and Weckwerth, 2013) and regulation of enzyme activity (Nägele et al., 2014). Listing all the presented approaches together with many others from the research fields of biomathematics,

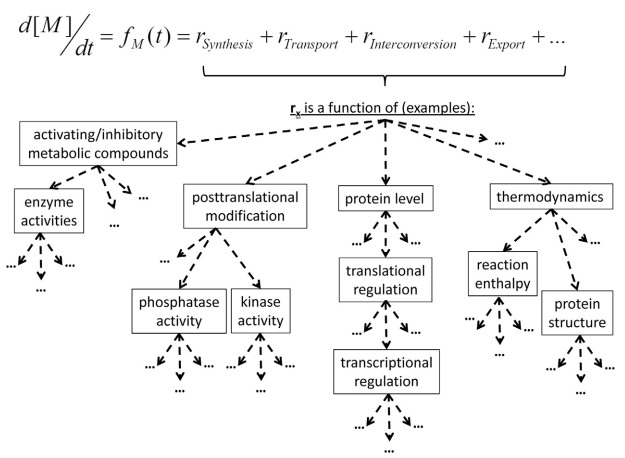

FIGURE 2 | Definition of metabolic functions, $f_M(t)$, and examples for regulatory processes. The dashed arrows indicate functional dependencies and can be read like "is a function of."

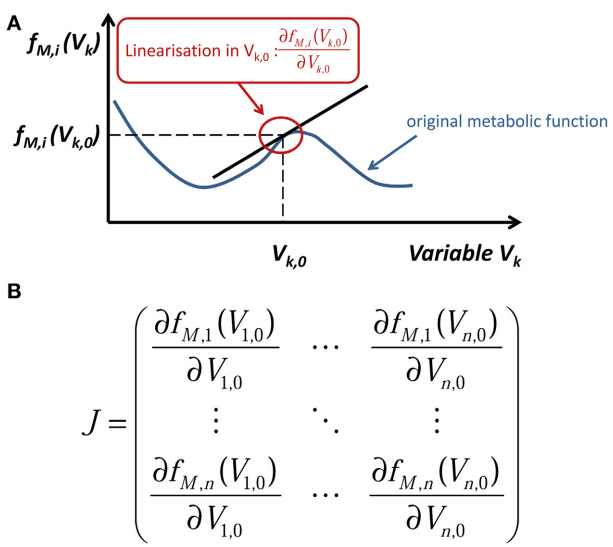

FIGURE 3 | Schematic representation of the linearization process of a metabolic function. (A) Graphical draft of the linearization procedure of a metabolic function, $f_{M,i}(V_k)$, at a certain metabolic state, $V_{k,0}$. **(B)** The Jacobian matrix J comprising all results of the linearization process, i.e., partial derivatives.

theoretical biology, bioinformatics and cybernetics, will yield an astonishing diversity of comprehensive models and strategies, which have been developed during the last decades. One central challenge of the next decades' metabolomics research will be the integration and application of many of these theoretical platforms, to exploit the experimental high-throughput data as efficient as possible.

CONCLUSION AND OUTLOOK

Summarizing the above mentioned findings, open questions and approaches, metabolomics plays a central role in current systems biology research. Future work on the integration of different experimental metabolomics techniques will broaden the coverage of metabolomes. The interpretation of resulting multidimensional data arrays in context of metabolic network information at genome-scale will significantly promote our understanding of complex metabolic networks. Combining and integrating statistical methods and strategies of mathematical modeling is a promising approach to improve our skills in construing comprehensive metabolomics data with respect to biochemical regulation in multi-layered and highly compartmentalized biological systems. This will essentially contribute to a very profound knowledge across all domains of life and provide us with new ideas and perspectives to solve upcoming questions of global concern.

ACKNOWLEDGMENTS

I would like to thank the whole MoSys-Team for great support. Particularly, I would like to thank Wolfram Weckwerth, Matthias Nagler, David Lyon, Ella Nukarinen, Lisa Haberl and Lena Fragner for many fruitful discussions and critical comments. I also thank the reviewers for their valuable contributions and suggestions. I apologize to all authors in this research field whose work I could not mention due to limitations in topics and space. This work was supported by the EU-Marie-Curie ITN MERIT (GA 2010-264474) and the Austrian Science Fund (FWF, project P 26342-B21).

REFERENCES

Aldridge, B. B., and Rhee, K. Y. (2014). Microbial metabolomics: innovation, application, insight. *Curr. Opin. Microbiol.* 19C, 90–96. doi: 10.1016/j.mib.2014.06.009

Almstetter, M. F., Oefner, P. J., and Dettmer, K. (2012). Comprehensive two-dimensional gas chromatography in metabolomics. *Anal. Bioanal. Chem.* 402, 1993–2013. doi: 10.1007/s00216-011-5630-y

Altaf-Ul-Amin, M., Afendi, F. M., Kiboi, S. K., and Kanaya, S. (2014). Systems biology in the context of big data and networks. *Biomed. Res. Int.* 2014:428570. doi: 10.1155/2014/428570

Arrivault, S., Guenther, M., Florian, A., Encke, B., Feil, R., Vosloh, D., et al. (2014). Dissecting the subcellular compartmentation of proteins and metabolites in Arabidopsis leaves using non-aqueous fractionation. *Mol. Cell. Proteomics* 13, 2246–2259. doi: 10.1074/mcp.M114.038190

Bellaire, A., Ischebeck, T., Staedler, Y., Weinhaeuser, I., Mair, A., Parameswaran, S., et al. (2014). Metabolism and development—integration of micro computed tomography data and metabolite profiling reveals metabolic reprogramming from floral initiation to silique development. *New Phytol.* 202, 322–335. doi: 10.1111/nph.12631

Carrera, J., Estrela, R., Luo, J., Rai, N., Tsoukalas, A., and Tagkopoulos, I. (2014). An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of *Escherichia coli*. *Mol. Syst. Biol.* 10, 735. doi: 10.15252/msb.20145108

Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F., Balaji, S., Fu, W., et al. (2011). Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* 7, 518. doi: 10.1038/msb.2011.52

Chen, B. S., and Chen, P. W. (2009). On the estimation of robustness and filtering ability of dynamic biochemical networks under process delays, internal parametric perturbations and external disturbances. *Math. Biosci.* 222, 92–108. doi: 10.1016/j.mbs.2009.09.004

Creek, D. J., Dunn, W. B., Fiehn, O., Griffin, J. L., Hall, R. D., Lei, Z., et al. (2014). Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics* 10, 350–353. doi: 10.1007/s11306-014-0656-8

Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., et al. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1777–1782. doi: 10.1073/pnas.0610772104

Föllinger, O., and Konigorski, U. (2013). *Regelungstechnik Einführung in die Methoden und ihre Anwendung*. Berlin: VDE.

Ghosh, S., Matsuoka, Y., Asai, Y., Hsin, K.-Y., and Kitano, H. (2011). Software for systems biology: from tools to integrated platforms. *Nat. Rev. Genet.* 12, 821–832. doi: 10.1038/nrg3096

Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L., and Bino, R. (2002). Plant metabolomics: the missing link in functional genomics strategies. *Plant Cell* 14, 1437–1440. doi: 10.1105/tpc.140720

Herrgard, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., et al. (2008). A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* 26, 1155–1160. doi: 10.1038/nbt1492

Hill, C. B., and Roessner, U. (2013). "Metabolic profiling of plants by GC–MS," in *The Handbook of Plant Metabolomics*, eds W. Weckwerth and G. Kahl (Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA), 1–23.

Hill, K., Porco, S., Lobet, G., Zappala, S., Mooney, S., Draye, X., et al. (2013). Root systems biology: integrative modeling across scales, from gene regulatory networks to the rhizosphere. *Plant Physiol.* 163, 1487–1503. doi: 10.1104/pp.113.227215

Hopfgartner, G., and Varesio, E. (2013). "Tandem mass spectrometry hyphenated with HPLC and UHPLC for targeted metabolomics," in *Metabolomics in Practice*, eds M. Lämmerhofer and W. Weckwerth (Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA), 21–37.

Jennings, K. R. (2000). The changing impact of the collision-induced decomposition of ions on mass spectrometry. *Int. J. Mass Spectrom.* 200, 479–493. doi: 10.1016/S1387-3806(00)00325-0

Kaplan, F., Kopka, J., Haskell, D. W., Zhao, W., Schiller, K. C., Gatzke, N., et al. (2004). Exploring the temperature-stress metabolome of Arabidopsis. *Plant Physiol.* 136, 4159–4168. doi: 10.1104/pp.104.052142

Klie, S., Krueger, S., Krall, L., Giavalisco, P., Flugge, U. I., Willmitzer, L., et al. (2011). Analysis of the compartmentalized metabolome—a validation of the non-aqueous fractionation technique. *Front. Plant Sci.* 2:55. doi: 10.3389/fpls.2011.00055

Korman, A., Oh, A., Raskind, A., and Banks, D. (2012). Statistical methods in metabolomics. *Methods Mol. Biol.* 856, 381–413. doi: 10.1007/978-1-61779-585-5_16

Krueger, S., Giavalisco, P., Krall, L., Steinhauser, M.-C., Büssis, D., Usadel, B., et al. (2011). A topological map of the compartmentalized *Arabidopsis thaliana* leaf metabolome. *PLoS ONE* 6:e17806. doi: 10.1371/journal.pone.0017806

Kuehnbaum, N. L., and Britz-Mckibbin, P. (2013). New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chem. Rev.* 113, 2437–2468. doi: 10.1021/cr300484s

Le Cao, K. A., Boitard, S., and Besse, P. (2011). Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* 12:253. doi: 10.1186/1471-2105-12-253

Leon, Z., Garcia-Canaveras, J. C., Donato, M. T., and Lahoz, A. (2013). Mammalian cell metabolomics: experimental design and sample preparation. *Electrophoresis* 34, 2762–2775. doi: 10.1002/elps.201200605

Madala, N. E., Piater, L. A., Steenkamp, P. A., and Dubery, I. A. (2014). Multivariate statistical models of metabolomic data reveals different metabolite distribution patterns in isonitrosoacetophenone-elicited *Nicotiana tabacum* and Sorghum bicolor cells. *Springerplus* 3:254. doi: 10.1186/2193-1801-3-254

Mardinoglu, A., Gatto, F., and Nielsen, J. (2013). Genome-scale modeling of human metabolism—a systems biology approach. *Biotechnol. J.* 8, 985–996. doi: 10.1002/biot.201200275

Masakapalli, S. K., Kruger, N. J., and Ratcliffe, R. G. (2013). The metabolic flux phenotype of heterotrophic Arabidopsis cells reveals a complex response to changes in nitrogen supply. *Plant J.* 74, 569–582. doi: 10.1111/tpj.12142

Masakapalli, S. K., Le Lay, P., Huddleston, J. E., Pollock, N. L., Kruger, N. J., and Ratcliffe, R. G. (2010). Subcellular flux analysis of central metabolism in a heterotrophic Arabidopsis cell suspension using steady-state stable isotope labeling. *Plant Physiol.* 152, 602–619. doi: 10.1104/pp.109.151316

Mei-Ling, L., Hai-Ling, X., Da-Nian, X., and Zhi-Sheng, C. (2006). Identification of certain chemical agents in complex organic solutions by gas chromatography/tandem mass spectrometry. *J. Mass Spectrom.* 41, 1453–1458. doi: 10.1002/jms.1116

Meinert, C., and Meierhenrich, U. J. (2012). A new dimension in separation science: comprehensive two-dimensional gas chromatography. *Angew. Chem. Int. Ed. Engl.* 51, 10460–10470. doi: 10.1002/anie.201200842

Mintz-Oron, S., Meir, S., Malitsky, S., Ruppin, E., Aharoni, A., and Shlomi, T. (2012). Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 339–344. doi: 10.1073/pnas.1100358109

Nägele, T., and Heyer, A. G. (2013). Approximating subcellular organisation of carbohydrate metabolism during cold acclimation in different natural accessions of *Arabidopsis thaliana*. *New Phytol.* 198, 777–787. doi: 10.1111/nph.12201

Nägele, T., Mair, A., Sun, X., Fragner, L., Teige, M., and Weckwerth, W. (2014). Solving the differential biochemical Jacobian from metabolomics covariance data. *PLoS ONE* 9:e92299. doi: 10.1371/journal.pone.0092299

Nägele, T., and Weckwerth, W. (2013). A workflow for mathematical modeling of subcellular metabolic pathways in leaf metabolism of *Arabidopsis thaliana*. *Front. Plant Sci.* 4:541. doi: 10.3389/fpls.2013.00541

Nobeli, I., Ponstingl, H., Krissinel, E. B., and Thornton, J. M. (2003). A structure-based anatomy of the *E.coli* metabolome. *J. Mol. Biol.* 334, 697–719. doi: 10.1016/j.jmb.2003.10.008

Oikawa, A., Matsuda, F., Kikuyama, M., Mimura, T., and Saito, K. (2011). Metabolomics of a single vacuole reveals metabolic dynamism in an *Alga Chara* Australis. *Plant Physiol.* 157, 544–551. doi: 10.1104/pp.111.183772

Okada, T., Afendi, F. M., Altaf-Ul-Amin, M., Takahashi, H., Nakamura, K., and Kanaya, S. (2010). Metabolomics of medicinal plants: the importance of multivariate analysis of analytical chemistry data. *Curr. Comput. Aided Drug Des.* 6, 179–196. doi: 10.2174/157340910791760055

Reznik, E., and Segre, D. (2010). On the stability of metabolic cycles. *J. Theor. Biol.* 266, 536–549. doi: 10.1016/j.jtbi.2010.07.023

Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N., and Willmitzer, L. (2000). Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.* 23, 131–142. doi: 10.1046/j.1365-313x.2000.00774.x

Sanchez, B. J., Perez-Correa, J. R., and Agosin, E. (2014). Construction of robust dynamic genome-scale metabolic model structures of *Saccharomyces cerevisiae* through iterative re-parameterization. *Metab. Eng.* 25C, 159–173. doi: 10.1016/j.ymben.2014.07.004

Sheth, B., and Thaker, V. (2014). Plant systems biology: insights, advances and challenges. *Planta* 240, 33–54. doi: 10.1007/s00425-014-2059-5

Simmler, C., Napolitano, J. G., Mcalpine, J. B., Chen, S. N., and Pauli, G. F. (2014). Universal quantitative NMR analysis of complex natural samples. *Curr. Opin. Biotechnol.* 25, 51–59. doi: 10.1016/j.copbio.2013.08.004

Steuer, R., Gross, T., Selbig, J., and Blasius, B. (2006). Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11868–11873. doi: 10.1073/pnas.0600013103

Steuer, R., Morgenthal, K., Weckwerth, W., and Selbig, J. (2007). A gentle guide to the analysis of metabolomic data. *Methods Mol. Biol.* 358, 105–126. doi: 10.1007/978-1-59745-244-1_7

Strogatz, S. H. (1994). *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering*. Reading, MA: Addison-Wesley.

Sugimoto, M., Kawakami, M., Robert, M., Soga, T., and Tomita, M. (2012). Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. *Curr. Bioinform.* 7, 96–108. doi: 10.2174/157489312799304431

Sun, X., and Weckwerth, W. (2012). COVAIN: a toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. *Metabolomics* 8, S81–S93. doi: 10.1007/s11306-012-0399-3

Szecowka, M., Heise, R., Tohge, T., Nunes-Nesi, A., Vosloh, D., Huege, J., et al. (2013). Metabolic fluxes in an illuminated Arabidopsis rosette. *Plant Cell* 25, 694–714. doi: 10.1105/tpc.112.106989

Uarrota, V. G., Moresco, R., Coelho, B., Nunes Eda, C., Peruch, L. A., Neubert Ede, O., et al. (2014). Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration. *Food Chem.* 161, 67–78. doi: 10.1016/j.foodchem.2014.03.110

Van Agthoven, M. A., Delsuc, M. A., Bodenhausen, G., and Rolando, C. (2013). Towards analytically useful two-dimensional Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Bioanal. Chem.* 405, 51–61. doi: 10.1007/s00216-012-6422-8

Vandamme, D., Fitzmaurice, W., Kholodenko, B., and Kolch, W. (2013). Systems medicine: helping us understand the complexity of disease. *QJM* 106, 891–895. doi: 10.1093/qjmed/hct163

Weckwerth, W. (2008). Integration of metabolomics and proteomics in molecular plant physiology-coping with the complexity by data-dimensionality reduction. *Physiol. Plant.* 132, 176–189. doi: 10.1111/j.1399-3054.2007.01011.x

Weckwerth, W. (2011). Unpredictability of metabolism-the key role of metabolomics science in combination with next-generation genome sequencing. *Anal. Bioanal. Chem.* 400, 1967–1978. doi: 10.1007/s00216-011-4948-9

Westerhuis, J. A., Van Velzen, E. J., Hoefsloot, H. C., and Smilde, A. K. (2010). Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics* 6, 119–128. doi: 10.1007/s11306-009-0185-z

Yao, F., Coquery, J., and Le Cao, K. A. (2012). Independent principal component analysis for biologically meaningful dimension reduction of large biological data sets. *BMC Bioinformatics* 13:24. doi: 10.1186/1471-2105-13-24

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 August 2014; accepted: 23 October 2014; published online: 06 November 2014.

Citation: Nägele T (2014) Linking metabolomics data to underlying metabolic regulation. Front. Mol. Biosci. ***1****:22. doi: 10.3389/fmolb.2014.00022*

This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences.

Copyright © 2014 Nägele. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.