



## REVIEW

# New frontiers in metabolomics: from measurement to insight

## [version 1; referees: 3 approved]

Eli Riekeberg, Robert Powers

Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE, USA

**v1** **First published:** 19 Jul 2017, 6(F1000 Faculty Rev):1148 (doi: 10.12688/f1000research.11495.1)

**Latest published:** 19 Jul 2017, 6(F1000 Faculty Rev):1148 (doi: 10.12688/f1000research.11495.1)

### Abstract

Metabolomics is the newest addition to the “omics” disciplines and has shown rapid growth in its application to human health research because of fundamental advancements in measurement and analysis techniques. Metabolomics has unique and proven advantages in systems biology and biomarker discovery. The next generation of analysis techniques promises even richer and more complete analysis capabilities that will enable earlier clinical diagnosis, drug refinement, and personalized medicine. A review of current advancements in methodologies and statistical analysis that are enhancing and improving the performance of metabolomics is presented along with highlights of some recent successful applications.

### Open Peer Review

**Referee Status:** 

	Invited Referees		
	1	2	3
<b>version 1</b> published 19 Jul 2017			

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty. In order to make these reviews as comprehensive and accessible as possible, peer review takes place before publication; the referees are listed below, but their reports are not formally published.

- 1 **Brent Winston**, University of Calgary, Canada
- 2 **Uwe Christians**, University of Colorado, USA
- 3 **Christoph Steinbeck**, Friedrich-Schiller-University Jena, Germany

### Discuss this article

Comments (0)

**Corresponding author:** Robert Powers ([rpowers3@unl.edu](mailto:rpowers3@unl.edu))

**Competing interests:** The authors declare that they have no competing interests.

**How to cite this article:** Riekeberg E and Powers R. **New frontiers in metabolomics: from measurement to insight [version 1; referees: 3 approved]** *F1000Research* 2017, **6**(F1000 Faculty Rev):1148 (doi: [10.12688/f1000research.11495.1](https://doi.org/10.12688/f1000research.11495.1))

**Copyright:** © 2017 Riekeberg E and Powers R. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Grant information:** This manuscript was supported in part by funds from the National Institutes of Health (NIH) (R01 AI087668, R21 AI087561, R01 CA163649, P20 RR-17675, and P30 GM103335), the Nebraska Tobacco Settlement Biomedical Research Development Fund, the Nebraska Research Council, and the Nebraska Center for Integrated Biomolecular Communication Systems Biology Core (NIH National Institute of General Medical Sciences; P20-GM113126). This material is based on work supported by the National Science Foundation under award number DBI-1660921. The research was performed in facilities renovated with support from the NIH (RR015468-01).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**First published:** 19 Jul 2017, **6**(F1000 Faculty Rev):1148 (doi: [10.12688/f1000research.11495.1](https://doi.org/10.12688/f1000research.11495.1))

## Introduction

Metabolomics is a rapidly growing field of study that endeavors to measure the complete set of metabolites (generally considered to be the intermediates and products of cellular metabolism less than 1 kDa in size) within a biological sample (that is, the metabolome) in order to achieve a global view of the state of the system<sup>1</sup>. Typically, metabolomics is focused only on characterizing the water-soluble metabolites, whereas lipidomics is a specialized discipline that investigates only lipids<sup>2</sup>. Water-soluble metabolites are part of a mobile, open biological system and as a result can readily interact and communicate with the environment, including the microbiome<sup>3</sup>. This is also true for some lipids but to a much lesser extent. Consequently, metabolomics has become an essential resource for systems biology because of its unique perspective relative to genomics and proteomics. Numerous studies have measured the relative upregulation or downregulation of genes or proteins to infer changes in biological function. However, it has been shown that even for common metabolic processes, such as glycolysis, a change in the cellular concentration of an enzyme does not necessarily lead to a proportional change in metabolic flux<sup>4</sup>. Thus, whereas genomics and proteomics identify what *might* happen, metabolomics identifies what *is* actually happening in the system. This realization demands a different perspective and requires the measurement of transcriptional, proteomic, and metabolomic data in order to obtain a complete picture of the system's response to environmental or genetic stress.

As another illustration, a "silent mutation" does not produce an observable change in phenotype despite an alteration in a gene or protein product. Therefore, metabolomics profiling can be used to decode the function of silent mutations, such as the *Pfk26* and *Pfk27* genes in *Saccharomyces cerevisiae* that both encode the glycolytic/gluconeogenesis regulator phosphofructokinase 2<sup>5</sup>. Via co-response and cluster analysis, these genes were observed to exhibit similar metabolite profiles which differed from other genes impacting energy metabolism. For these reasons, methods to directly measure metabolite concentrations within cells, tissues, organs, or other biological samples are crucial for fully understanding a system when traditional omics studies (for example, genomics, proteomics, and transcriptomics) are deemed insufficient. To date, nuclear magnetic resonance (NMR)<sup>6</sup> and mass spectrometry (MS)<sup>7</sup> have been the primary analytical techniques used to characterize a metabolome. NMR and MS are typically combined with univariate and multivariate statistical methods to identify major metabolite changes and to identify potential biomarkers<sup>8</sup>. Nevertheless, despite the tremendous growth in the field, critical protocols and techniques are still under development. Herein, we present recent advances in methodologies and statistical analysis that are enhancing and improving the performance of metabolomics while extending the applications in which metabolomics can play a significant role.

## Essential components of a metabolomics study

Conceptually, an untargeted metabolomics study is quite simple. Biological samples are obtained from two or more experimental groups to be compared (healthy versus diseased, wild-type versus gene knockout, and so on) and the metabolites are extracted. These metabolic extracts then are measured by using numerous instrumental techniques, of which NMR and liquid chromatography

(LC)-MS are the most common. The resulting spectra then are subjected to statistical analysis techniques such as principal component analysis (PCA) and orthogonal projection onto latent structures (OPLS) to determine the most significant spectral features that define each group<sup>9,10</sup>. Finally, these spectral features then can be assigned to distinct metabolites and metabolic pathways by using spectral libraries of known metabolites<sup>11,12</sup>. In this manner, untargeted metabolomics is discovery-based since it reveals previously unknown information about how a system responds to environmental or genetic stress. Conversely, targeted metabolomics focuses on analyzing a specific set of metabolites on the basis of some prior knowledge about the system. As a result, targeted metabolomics studies tend to be more sensitive and quantitative and have a higher reproducibility and a lower false-positive rate relative to untargeted metabolomics.

Protocols for obtaining and extracting the metabolome have been well developed and exhaustively reviewed for a wide range of biological samples, including cell cultures, urine, blood/serum, and both animal- and plant-derived tissues<sup>13-16</sup>. Although these protocols are readily available, the variable stability of metabolites means that even minor changes in procedure can have a major impact on the observed metabolome. The fast turnover rate of enzymes and the variable temperature and chemical stability of metabolites require that metabolomics samples be collected quickly and handled uniformly and that all enzymatic activity be rapidly quenched in order to minimize biologically irrelevant deviations between samples that may result from the processing protocol<sup>16,17</sup>. Thus, the optimization of the sample preparation protocol is essential to a successful metabolomics study. Conversely, the most likely source of bias is improper handling of the metabolomics samples. An important consideration is that the complete metabolome cannot be captured in a single extraction protocol. This is in stark contrast to modern genomics, which can reliably cover the entire genome of an organism. A metabolomics extraction protocol will usually focus on only a subset of metabolites (for example, water-soluble metabolites or lipids). Furthermore, an extraction protocol may focus on either a highly reproducible and quantitative extraction of a restricted set of metabolites (that is, targeted metabolomics) or the global collection of all possible metabolites (that is, untargeted metabolomics) with a possible reduction in precision. In general, 200 to 500 metabolites may be observed by targeted metabolomics, whereas upwards of 1,500 metabolites have been detected in untargeted metabolomics studies<sup>18</sup>.

Following extraction and subsequent data collection, the final and perhaps most crucial step is metabolite assignment, which typically is accomplished by a comparison with spectral libraries of known metabolites. This is not a trivial task since the number of possible metabolites can be prohibitively large, and a large segment of the metabolome is either unknown or lacking a reference spectrum. For example, the human metabolome is estimated to contain around 150,000 metabolites<sup>18</sup>, but the Human Metabolome Database<sup>11</sup> contains only around 74,000 metabolites (as of 1 June 2017). Thus, there are still many unknown metabolites and a true estimate of the size of the human metabolome is challenging<sup>19</sup>. Another complicating factor is that different organisms may have completely unique metabolomes. For instance, plants have over 45,000 known secondary metabolites<sup>20</sup>. Finally, there may be ambiguities in making

a metabolite assignment because of chemical shift overlap or identical masses (for example, isomers). As a result, the assignment of metabolites to spectral features may be as low as 4 to 5%<sup>21</sup>.

### A tale of two methods: mass spectrometry or nuclear magnetic resonance?

Perhaps the most important choice that can be made in a metabolomics study is which instrumental platform is used. Although a wide range of instruments have been used for metabolomics, including capillary electrophoresis, infrared spectroscopy, and Raman spectroscopy, only NMR and MS are routinely used for metabolomics. NMR and MS are often applied in metabolomics investigations because of their inherent complementarity, which results from their distinct advantages and disadvantages<sup>22</sup>. NMR is highly reproducible and quantitative, has simple sample preparation protocols, and is able to measure analytes over a wide range of solvent conditions<sup>23</sup>. Despite these advantages, the main limitation of NMR is its low sensitivity, which restricts its application to measuring the most abundant metabolites in the sample (micromolar to millimolar range). This has been noted as a significant hurdle that has slowed the widespread adoption of NMR by the metabolomics community<sup>22,24</sup>. Conversely, the high sensitivity and low detection limits of MS enable the detection of subtle metabolic changes that are invisible by NMR. With this increase in sensitivity, the detection of thousands of peaks is relatively common<sup>25</sup>, but untargeted MS metabolomics studies often are not quantitative in nature. Since MS detectors rely on ionization processes, MS is restricted to detecting metabolites that readily ionize. Correspondingly, a significant reduction in observable metabolites may occur depending on the specifics of the sample being considered<sup>26</sup>. For a detailed overview of the utility of various MS detectors to metabolomics, see the review article by Dunn *et al.*<sup>27</sup> (2005).

MS also suffers from reproducibility problems since contaminants within the sample can change the ionization efficiency of metabolites<sup>2</sup>. Specifically, quantitation is challenging in untargeted MS since peak intensity is dependent on ionization efficiency, which varies between metabolites and also is strongly dependent on experimental conditions that may result in varying ion suppression<sup>28</sup>. One issue of particular relevance to MS is the relatively narrow nominal mass and mass defect distribution of the metabolome which results in significant peak overlap<sup>29</sup>. This can be resolved by coupling MS to a chromatographic method, most commonly LC or gas chromatography (GC), to resolve overlapping peaks and to aid in the metabolite identification based on retention time and the properties of the stationary phase.

GC was the first separation technique applied to the analysis of metabolic mixtures; for example, GC-MS was used to identify biomarkers for diagnosing phenylketonuria in 1970<sup>30</sup>. GC-MS is particularly beneficial for the analysis of volatile metabolite mixtures since minimal sample preparation is required; in some cases, samples can be directly analyzed. Furthermore, a number of applications of GC-MS uniquely involve detecting volatile metabolites; two examples are the measurement of exhaled breath condensates for diagnosing lung cancer<sup>31</sup> and the monitoring of volatile paper degradation products from historic books<sup>32</sup>. An obvious disadvantage of GC-MS is its reliance on analyte volatility, where metabolites of low volatility or low temperature

stability may be modified or destroyed<sup>17</sup>. Limited metabolite volatility can be overcome through the use of derivatization schemes, but derivatization is time-consuming. More importantly, differences in the efficiency of the derivatization<sup>33</sup> and differences in the stability of the derivatized metabolites<sup>17</sup> may dramatically perturb the apparent concentrations of the metabolites, possibly leading to an erroneous biological conclusion.

LC was not widely used for metabolomics until the 1980s<sup>27</sup> and this was due to technical limitations with interfacing LC and mass spectrometers. A main advantage of LC over GC is that most metabolites can be detected intact and without modification from a derivatizing agent. Additionally, LC provides an accurate analysis of thermally unstable or reactive metabolites since the separation typically occurs at room temperature. However, the introduction of a liquid phase does introduce a higher variability in retention times<sup>34</sup>, an increase in ion suppression due to matrix effects<sup>35</sup>, and a lower resolution relative to GC.

NMR and MS tend to observe a distinct set of metabolites from the same metabolomics sample. Consequently, there is a growing trend in metabolomics to perform tandem studies in which the same sample is analyzed by both NMR and MS<sup>36–39</sup>. In this manner, the coverage of the metabolome is significantly increased by taking advantage of the strengths of both methods. NMR identifies trends in metabolic alteration along core metabolic pathways and provides a context for the interpretation of the low-abundance metabolites identified by MS. Of course, the combined use of NMR and MS leads to a proportional increase in data set size with the added complexity of the simultaneous processing, analysis, and interpretation of two dissimilar data types.

### Data processing and interpretation

Metabolomics experiments generate large data sets that require specialized tools for analysis. Numerous software packages for data pre-processing and statistical analysis are available and have been reviewed elsewhere<sup>40,41</sup>. Unfortunately, no single software exists that can simultaneously perform all of the critical steps needed for an analysis of a combined NMR and MS data set. Although the statistical techniques applied to NMR and MS data sets are largely the same, each technique requires a unique set of pre-processing tools and algorithms prior to modelling. For example, an NMR spectrum has to be Fourier-transformed and phased, whereas centroiding and de-isotoping are required in MS. Owing largely to these data type-specific processing requirements, newly developed software is almost exclusively restricted to one method or the other. Conversely, there has been minimal effort in developing tools capable of working with both NMR and MS data sets<sup>40</sup>.

There are two general approaches to integrating NMR and MS data sets into a single coherent study. The first involves samples simply being independently analyzed by each method. The separate data sets then are compared in order to identify consistencies in the metabolic alterations observed by each technique. The main advantage of the approach is simplicity since it does not require any significant protocol changes. Also, the confidence of a metabolite assignment may be significantly increased if it is identified by both methods. Furthermore, a measure of internal consistency may be achieved if metabolite concentrations can be estimated by both methods.

However, significant information can be lost during this process since, for example, ambiguity in peak assignments sometimes can be resolved by information from the other method. There is also a lack of statistical correlation since the data sets are independently analyzed. Although the manual curation of independent data sets is the dominant method currently used by metabolomics investigators, it also suffers from reproducibility problems due to potential biases in data interpretation (for example, metabolite assignment methods) among other issues.

The second approach to combining NMR and MS data sets is to simultaneously integrate each data set into a *single* statistical model using a multiblock analysis. Multiblock analysis encompasses a variety of methods that combine multiple data sets prior to conventional multivariate analysis. In addition to combining multiple instrumental data sources<sup>42</sup>, multiblock analysis has been successfully employed to combine data sets from different omics disciplines<sup>43</sup>. Multiblock methods are preferable to independent analysis since the relative contributions of each data set still can be quantified, but importantly the larger combined data set is likely to result in models with greater predictive ability and resolving power than either method alone<sup>44</sup>. However, the software tools to perform multiblock analyses are crude and often rely on custom sets of pre-processing routines using multiple software packages. The lack of integrated analysis tools and software is a major roadblock in metabolomics, especially in light of the growing interest in combining NMR and MS data sets.

## Recent advances in metabolomics

### Dynamic nuclear polarization

NMR metabolomics investigations, especially those concerned with achieving a high confidence in metabolite identification, require two-dimensional NMR methods to resolve the overlap present in one-dimensional spectra. In general, this requires isotopic labeling with NMR-active nuclei like <sup>13</sup>C and <sup>15</sup>N because of their low natural abundance. In the last few years, dynamic nuclear polarization (DNP) has evolved from a structural biology tool in the area of solid-state NMR to have potential applications in solution-state metabolomics<sup>45</sup>. In DNP, a solid, frozen metabolomics sample at about 1.5 K is polarized in the presence of microwave-irradiated free-radicals, which induces a temporary hyperpolarization in spin-active nuclei through a transfer of polarization from electrons to nuclei. The sample then needs to be rapidly melted and transferred to an NMR spectrometer to take advantage of the greatly enhanced sensitivity (>10,000-fold)<sup>46</sup>. The dramatic increase in sensitivity avoids the need for isotopic labelling, especially for *in vivo* samples, and may permit the detection of low-abundance metabolites. Nevertheless, DNP experiments are limited by T<sub>1</sub> relaxation rates, resulting in a short measurement window of the dynamically polarized samples. DNP also requires substantial hardware modifications and accessories (for example, microwave generator) to rapidly thaw and shuttle samples back and forth from the NMR spectrometer. DNP has also been applied to <sup>13</sup>C-labeled metabolites that then are used as a tracer compound for *in vivo* imaging<sup>47</sup>. This requires close proximity of the polarizer and magnetic resonance imaging spectrometer to allow for rapid transfer, dissolution, and injection of the <sup>13</sup>C-labeled metabolite given the relatively short T<sub>1</sub> of 30 to 40 seconds for a <sup>13</sup>C-labeled carboxyl group. Despite these technical obstacles, DNP has been successfully used to monitor a single

metabolite (for example, pyruvate) in living tissue (for example, heart) by magnetic resonance imaging<sup>48</sup>. Besides the short measurement time, another challenge with the application of DNP to *in vivo* imaging is the limited number of <sup>13</sup>C-labeled metabolites that can be polarized and tolerated at the concentrations needed for imaging (25 to 80 mM) and that are also a useful biological probe. In addition to pyruvate, bicarbonate, fumarate, urea, glutamine, and dehydroascorbate have been used for *in vivo* imaging<sup>47</sup>. Despite fundamental issues of reproducibility and limitations in sample preparation, DNP protocols and technology are rapidly advancing and one day could become a routine tool for metabolomics<sup>49</sup>.

### Disease profiling and personalized medicine

Metabolomics can be used to profile an individual's responses to a drug treatment or other medical therapy by monitoring metabolite changes in readily obtainable biofluids (for example, blood and urine). A unique advantage of metabolites as biomarkers is the likely occurrence of observing a set of multiple metabolites with distinctly different concentration changes that are correlated with a disease state or treatment response. Correspondingly, multiple metabolites, instead of a single biomarker, are expected to yield a higher sensitivity and selectivity. For example, plasma baseline levels of xanthine, 2-hydroxyvaleric acid, succinic acid, stearic acid, and fructose prior to simvastatin treatment were observed to reliably predict a good or poor response in reducing low-density lipoprotein cholesterol<sup>50</sup>. The OPLS model yielded a 70% sensitivity and 79% specificity with a corresponding area under the receiver operating characteristic curve of 0.84. Thus, metabolomics can be used to predict whether a patient will respond to a drug in addition to being used as a semi-quantitative prognosis of disease progression.

For example, in a recent study of patients with tuberculosis (TB), urine samples that were collected over the course of a 6-month period became more similar to those of a non-TB control group during the course of first-line anti-TB therapy (for example, isoniazid, ethambutol, or pyrazinamide). Metabolomics has also been successfully employed to identify serum metabolic alterations associated with psoriasis<sup>51</sup>. Importantly, the metabolomics results were consistent with trends previously observed in genomics and proteomics studies. The metabolome changes were observed to reverse following successful corticosteroid treatment<sup>52</sup>. Interestingly, the authors identified an increased demand for glutamine, which had not been previously reported in psoriasis<sup>52,53</sup>. Glutamine demand is directly associated with diseases characterized by increased cellular proliferation, such as in cancers. A significant alteration in β-isosterol, which is a commonly employed herbal remedy, was also observed. Thus, metabolomics may also be used to identify a patient's use of alternative treatments outside of his or her physician's knowledge or recommendation. In this manner, metabolomics may assist in determining whether co-administration of a complementary treatment was beneficial or detrimental to a patient's therapeutic outcome.

### New trends in data analysis

Much of the data analysis approach in metabolomics has been largely borrowed from the field of chemometrics, which pioneered the application of PCA and PLS to chemical systems<sup>54</sup>. Although these are powerful statistical tools, the current trend in metabolomics data analysis is evaluating the efficacy of new algorithms and statistical methods to improve group separation and



metabolite identification. PCA, PLS, and OPLS are all commonly employed by metabolomics investigators, but newer approaches, including support vector machine (SVM)<sup>55</sup>, random forest (RF)<sup>56</sup>, and self-organizing map (SOM)<sup>57</sup> algorithms, have all been recently applied to metabolomics data sets.

Despite having been formalized since 1992<sup>58</sup>, SVM has been used extensively only in the analysis of gene microarray data, particularly due to its performance on data sets characterized by a large number of variables and few samples<sup>59,60</sup>. SVM is also able to identify non-linear relationships that violate the linearity assumptions of PCA and OPLS, making it easily generalizable. SVM has been recently applied in biomarker discovery for ovarian cancer, and a model using serum-derived LC-MS spectra was able to predict disease onset with higher accuracy than the currently accepted method of CA-125 serum monitoring<sup>61</sup>. A major caveat of SVM is its restriction to binary classification problems: it is able to discriminate between only two experimental groups. Simply, spectra belonging to two experimental groups are represented as points in  $n$ -dimensional space, where  $n$  corresponds to the number of observed metabolites. A hyperplane then is calculated that best separates the points from the two groups. The coefficients of the calculated hyperplane are used to determine which metabolites are most important for discriminating between the two groups. Although methods have been proposed to extend SVM to multi-class problems, they are often done by breaking down the data set into an ensemble of binary groups that oversimplifies the problem and leads to uninformative models<sup>62</sup>.

The RF algorithm is a decision tree-based method that uses random subsets of the data to construct multiple models, which then are combined to create an average model in a process known as bootstrap aggregation. In the decision tree method, samples are mapped to a target value (that is, which experimental class the sample belongs to) using a set of variable-based decision rules that separate the samples into groups corresponding to the target value. These newly formed groups can be further subdivided according to new variables, and each “branch” of separation is repeated until the samples can be fully differentiated. The major advantage of the decision tree is its imperviousness to scaling and variable normalization, an extremely common problem in metabolomics data<sup>63</sup>. The disadvantages include an extreme propensity for overfitting and having extremely poor generalizability that severely limits its utility. RF addresses this limitation by creating an ensemble of partial decision trees that, when combined into an overall model, reduces variance and overfitting<sup>64</sup>. In particular, the RF algorithm, being relatively unaffected by scaling and normalization and easily handling both large data sets and missing values, is highly adaptable to the realities of real-world data sets. A major disadvantage of RF is that the method requires extensive “tuning” of default parameters by the investigator in order to obtain the best model. Also, the resulting decision tree can be hard to visualize for large data sets<sup>65</sup>. RFs have shown clinical value: they have been used to determine a set of serum protein and metabolic biomarkers in prostate cancer with higher predictive accuracy than the current prostate-specific antigen biomarker<sup>66,67</sup>. See Gromski *et al.* (2015) for an excellent review of the SVM and RF algorithms that also includes comparisons with other mainstream techniques<sup>65</sup>.

SOM is an approach similar to PCA that reduces multi-dimensional problems to a more easily interpretable low-dimensional grid to visualize natural clustering trends and groupings within a data set. SOM can be applied to the same tasks as PCA but without the biases toward high-variance metabolites. SOMs, like SVM, have the ability to detect non-linear relationships between detected metabolites<sup>68</sup>. SOMs have been successfully applied to develop biomarkers for early-stage renal cell carcinoma as well as to predict patient response to surgical intervention with a predictive accuracy of 94.74%<sup>69</sup>. In comparison with the other statistical methods, SOM has been severely limited in metabolomics because of a computationally intensive algorithm and the lack of a pre-packaged software, which has significantly diminished its accessibility to the wider research community<sup>70</sup>. Nevertheless, the usage of SOMs in metabolomics is steadily rising, and comparative analyses are beginning to demonstrate that SOMs are an acceptable alternative to more traditional clustering algorithms<sup>71</sup>.

In addition to statistical methods applied directly to spectral profiles, identified metabolites can be used with pathway analysis<sup>72</sup> to understand metabolite interactions with known pathways or to discover mechanisms of action in pharmaceutical natural product research<sup>73</sup>. Metabolomics data sets can generate an overwhelming and seemingly disjointed list of metabolites, which pathway analysis aims to place into a broader biological context by assigning metabolites to relevant metabolic pathways. This is done through a number of software tools that integrate putatively identified metabolites with pathway information from various databases. For example, MetaboAnalyst 3.0 is a suite of metabolomics tools (<http://www.metaboanalyst.ca>), which includes modules for metabolite enrichment analysis (MSEA), metabolite pathway analysis (MetPA), and an integrated pathway analysis. The user input is typically a list of metabolites (with or without concentrations) or genes or both. MSEA provides a ranked list of potentially key metabolic pathways based on the observed number of metabolites associated with that pathway (that is, metabolite set enrichment)<sup>74</sup>. MetPA combines MSEA with a pathway topology analysis to provide an overall pathway analysis to identify the metabolic pathways primarily impacted in the study<sup>72</sup>. The integrated pathway analysis combines both metabolomics and genomics data with enrichment analysis and topology analysis to again identify the pathways (in rank order) that were primarily impacted in the study<sup>75</sup>. MetaboSignal (<https://bioconductor.org/>) is an alternative approach to pathway analysis which employs directed graphs with network topology approaches to compute centrality measures to correlate gene-metabolite relationships through shortest-path distances<sup>76</sup>. Thus, unlike the MetaboAnalyst 3.0 tools, the output of MetaboSignal is a network map of gene-metabolite connectivities. Cytoscape (<http://www.cytoscape.org/>) is a generalized network interaction and visualization tool that works with a variety of data sets, including metabolomics data. Cytoscape combined with MetScape 3 (<http://metscape.ncibi.org/>)<sup>77</sup> can generate network maps similar to those of MetaboSignal from metabolomics or genomics data or both<sup>78</sup>. MetScape uses known pathways from Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>79</sup> and Edinburgh Human Metabolic Network (EHMN)<sup>80</sup> databases and gene set enrichment analysis to generate these networks in order to visualize the impacted metabolic pathways. In essence, there is

some significant overlap in the capabilities of MetaboAnalyst 3.0, MetaboSignal, and Cytoscape/MetScape 3. Importantly, pathway analysis provides an interaction network that may identify centralized hubs where metabolic pathways coincide or where bottlenecks may occur.

The limited connectivity or altered flow through specific metabolic nodes (that is, change in metabolic flux) may identify functionally essential biological processes<sup>81</sup>. These essential pathways then can be selectively targeted. By genetically or chemically restricting a potentially essential metabolic pathway, it is possible to ascertain the relevance of the pathway to a systems response to an environmental stress (that is, drug resistance) and potentially reverse or negate the response<sup>82</sup>.

Pathway analysis also allows for the integration of multi-modal omics data, such as combining gene-expression and metabolomics data to uncover gene and protein functions. For example, metabolite profiles were integrated with genome-wide screening of single-nucleotide polymorphisms (SNPs) to identify the molecular mechanism of the *NAT8* and *PYROXD2* genes. Briefly, SNPs were ranked according to the strength of an association with observed metabolites. The regions where these SNPs occur on the chromosomes then were screened to determine at what position in the genome the gene/protein product responsible for the mediation is stored. With this approach, it was suggested that the *NAT8* and *PYROXD2* genes were responsible for mediation of serum diethylamine levels<sup>83</sup>, a novel insight for these previously under-annotated genes.

As another illustration, transcriptomic and metabolomic data from *Arabidopsis thaliana* were integrated to characterize the biological response resulting from the over-expression of *PAP1*, a gene known to cause profound accumulation of anthocyanins and to encode a MYB transcription factor regulating flavonoid biosynthesis. The authors were able to correlate the biosynthesis of cyanidin and quercetin derivatives with a specific set of upregulated genes that enabled them to identify the function of two uncharacterized proteins: a flavonoid 3-O-glucosyltransferase and anthocyanin 5-O-glucosyltransferase<sup>84</sup>. Numerous tools are now available for pathway analysis of metabolomics data<sup>79,85,86</sup>, which will significantly improve data interpretation and simplify our understanding of biological relevance. Thus, pathway analysis is becoming a routine component of a detailed metabolomics analysis.

### Conclusions: What does the future hold?

The recent advancements in metabolomics outlined herein have been shown to enhance its utility in systems biology research and to have a beneficial impact on medical research and personalized medicine. The measurement of metabolomics profiles has been shown to be useful for monitoring treatment efficacies from both pharmaceutical and surgical interventions. As our understanding of the relationship between disease state and the chemical profile of biofluids grows, metabolomics is expected to become a routine approach for monitoring disease development and progression, as a tool for disease diagnosis, and for understanding the underlying molecular mechanisms of drug resistance. Metabolite profiles could be obtained at regular intervals and screened for

changes over a patient's lifetime as a diagnostic tool and a means to monitor a patient's overall health status. Some of this work has already begun; an ongoing Alphabet (parent company of Google) "moon-shot project" is a baseline study attempting to determine the inherent level of variability in human medical data that is not associated with a disease. Though still in its infancy, a similar approach using metabolomic profiles may be used to determine the inherent variability in biofluid profiles for healthy individuals. In this manner, metabolic profiles associated with disease onset and progression can be easily distinguished from the known variance in healthy individuals.

Some of the biggest challenges remaining in the field of metabolomics involve fundamental limits in experimental methodology. Metabolomics requires relatively high-cost instrumentation and complex data analysis and still suffers from issues of sample-to-sample variability. Although great strides in each of these areas have been made, there is still more work to be done before metabolomics can become a key and routine part of a clinical practice. Nevertheless, metabolomics continues to make important contributions to both medical research and general systems biology studies. In fact, the ability to directly measure metabolite concentration changes by using a targeted NMR or MS approach would greatly benefit investigations into a range of research areas that often are overlooked by other methods. In this manner, a metabolomics assay that targets a select and specific set of metabolites can be used to develop a highly reproducible and quantifiable assay that can be translated into a validated clinical assay.

### Abbreviations

DNP, dynamic nuclear polarization; GC, gas chromatography; LC, liquid chromatography; MetPA, metabolite pathway analysis; MS, mass spectrometry; MSEA, metabolite enrichment analysis; NMR, nuclear magnetic resonance; OPLS, orthogonal projection onto latent structures; PCA, principal component analysis; PLS, projection onto latent structures; RF, random forest; SNP, single-nucleotide polymorphism; SOM, self-organizing map; SVM, support vector machine; TB, tuberculosis.

### Competing interests

The authors declare that they have no competing interests.

### Grant information

This manuscript was supported in part by funds from the National Institutes of Health (NIH) (R01 AI087668, R21 AI087561, R01 CA163649, P20 RR-17675, and P30 GM103335), the Nebraska Tobacco Settlement Biomedical Research Development Fund, the Nebraska Research Council, and the Nebraska Center for Integrated Biomolecular Communication Systems Biology Core (NIH National Institute of General Medical Sciences; P20-GM113126). This material is based on work supported by the National Science Foundation under award number DBI-1660921. The research was performed in facilities renovated with support from the NIH (RR015468-01).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## References



1. Johnson CH, Ivanisevic J, Siuzdak G: **Metabolomics: beyond biomarkers and towards mechanisms.** *Nat Rev Mol Cell Biol.* 2016; **17**(7): 451–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Antignac J, De Wasch K, Monteau F, et al.: **The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis.** *Anal Chim Acta.* 2005; **529**(1–2): 129–36.  
[Publisher Full Text](#)
3. Turnbaugh PJ, Ley RE, Hamady M, et al.: **The human microbiome project.** *Nature.* 2007; **449**(7164): 804–10.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. **F**ter Kuile BH, Westerhoff HV: **Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway.** *FEBS Lett.* 2001; **500**(3): 169–71.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
5. Raamsdonk LM, Teusink B, Broadhurst D, et al.: **A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations.** *Nat Biotechnol.* 2001; **19**(1): 45–50.  
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Nagana Gowda GA, Raftery D: **Recent Advances in NMR-Based Metabolomics.** *Anal Chem.* 2017; **89**(1): 490–510.  
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Theodoridis GA, Gika HG, Want EJ, et al.: **Liquid chromatography-mass spectrometry based global metabolite profiling: a review.** *Anal Chim Acta.* 2012; **711**: 7–16.  
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Griffiths WJ, Koal T, Wang Y, et al.: **Targeted metabolomics for biomarker discovery.** *Angew Chem Int Ed Engl.* 2010; **49**(32): 5426–45.  
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Dai H, Xiao C, Liu H, et al.: **Combined NMR and LC-DAD-MS analysis reveals comprehensive metabolomic variations for three phenotypic cultivars of *Salvia miltiorrhiza* Bunge.** *J Proteome Res.* 2010; **9**(3): 1565–78.  
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Das MK, Bishwal SC, Das A, et al.: **Deregulated tyrosine-phenylalanine metabolism in pulmonary tuberculosis patients.** *J Proteome Res.* 2015; **14**(4): 1947–56.  
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Wishart DS, Jewison T, Guo AC, et al.: **HMDB 3.0—The Human Metabolome Database in 2013.** *Nucleic Acids Res.* 2013; **41**(Database issue): D801–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Go EP: **Database resources in metabolomics: an overview.** *J Neuroimmune Pharmacol.* 2010; **5**(1): 18–30.  
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Krishnan P, Kruger NJ, Ratcliffe RG: **Metabolite fingerprinting and profiling in plants using NMR.** *J Exp Bot.* 2005; **56**(410): 255–65.  
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Kind T, Tolstikov V, Fiehn O, et al.: **A comprehensive urinary metabolomic approach for identifying kidney cancer.** *Anal Biochem.* 2007; **363**(2): 185–95.  
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Wu H, Southam AD, Hines A, et al.: **High-throughput tissue extraction protocol for NMR- and MS-based metabolomics.** *Anal Biochem.* 2008; **372**(2): 204–12.  
[PubMed Abstract](#) | [Publisher Full Text](#)
16. **F** Breier M, Wahl S, Prehn C, et al.: **Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples.** *PLoS One.* 2014; **9**(2): e89728.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
17. **F** Fang M, Ivanisevic J, Benton HP, et al.: **Thermal Degradation of Small Molecules: A Global Metabolomic Investigation.** *Anal Chem.* 2015; **87**(21): 10935–41.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
18. Markley JL, Brüschweiler R, Edison AS, et al.: **The future of NMR-based metabolomics.** *Curr Opin Biotechnol.* 2017; **43**: 34–40.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Wishart DS, Lewis MJ, Morrissey JA, et al.: **The human cerebrospinal fluid metabolome.** *J Chromatogr B Anal Technol Biomed Life Sci.* 2008; **871**(2): 164–73.  
[PubMed Abstract](#) | [Publisher Full Text](#)
20. De Luca V, St Pierre B: **The cell and developmental biology of alkaloid biosynthesis.** *Trends Plant Sci.* 2000; **5**(4): 168–73.  
[PubMed Abstract](#) | [Publisher Full Text](#)
21. **F** Dias DA, Jones OA, Beale DJ, et al.: **Current and Future Perspectives on the Structural Identification of Small Molecules in Biological Systems.** *Metabolites.* 2016; **6**(4): pii: E46.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
22. Marshall DD, Powers R: **Beyond the paradigm: Combining mass spectrometry and nuclear magnetic resonance for metabolomics.** *Prog Nucl Magn Reson Spectrosc.* 2017; **100**: 1–16.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Pan Z, Raftery D: **Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics.** *Anal Bioanal Chem.* 2007; **387**(2): 525–7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Gowda GA, Zhang S, Gu H, et al.: **Metabolomics-based methods for early disease diagnostics.** *Expert Rev Mol Diagn.* 2008; **8**(5): 617–33.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Dettmer K, Aronov PA, Hammock BD: **Mass spectrometry-based metabolomics.** *Mass Spectrom Rev.* 2007; **26**(1): 51–78.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Copeland JC, Zehr LJ, Cerny RL, et al.: **The applicability of molecular descriptors for designing an electrospray ionization mass spectrometry compatible library for drug discovery.** *Comb Chem High Throughput Screen.* 2012; **15**(10): 806–15.  
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Dunn WB, Bailey NJ, Johnson HE: **Measuring the metabolome: current analytical technologies.** *Analyst.* 2005; **130**(5): 606–25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Annesley TM: **Ion suppression in mass spectrometry.** *Clin Chem.* 2003; **49**(7): 1041–4.  
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Kell DB: **Metabolomics and systems biology: making sense of the soup.** *Curr Opin Microbiol.* 2004; **7**(3): 296–307.  
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Blau K, Cameron HH, Summer GK: **Diagnosis of phenylketonuria by gas chromatography.** *Methods Med Res.* 1970; **12**: 100–5.  
[PubMed Abstract](#)
31. Fuchs P, Loeseken C, Schubert JK, et al.: **Breath gas aldehydes as biomarkers of lung cancer.** *Int J Cancer.* 2010; **126**(11): 2663–70.  
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Strlic M, Thomas J, Trafela T, et al.: **Material degradomics: on the smell of old books.** *Anal Chem.* 2009; **81**(20): 8617–22.  
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Villas-Bôas SG, Smart KF, Sivakumaran S, et al.: **Alkylation or Silylation for Analysis of Amino and Non-Amino Organic Acids by GC-MS?** *Metabolites.* 2011; **1**(1): 3–20.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Barwick VJ: **Sources of uncertainty in gas chromatography and high-performance liquid chromatography.** *J Chromatogr A.* 1999; **849**(1): 13–33.  
[Publisher Full Text](#)
35. Matuszewski BK, Constanzer ML, Chavez-Eng CM: **Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS.** *Anal Chem.* 2003; **75**(13): 3019–30.  
[PubMed Abstract](#) | [Publisher Full Text](#)
36. Yanshole VV, Snytnikova OA, Kiryutin AS, et al.: **Metabolomics of the rat lens: a combined LC-MS and NMR study.** *Exp Eye Res.* 2014; **125**: 71–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
37. **F** Bingol K, Brüschweiler-Li L, Yu C, et al.: **Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid identification of new metabolites in complex mixtures.** *Anal Chem.* 2015; **87**(7): 3864–70.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
38. **F** Bingol K, Brüschweiler R: **NMR/MS Translator for the Enhanced Simultaneous Analysis of Metabolomics Mixtures by NMR Spectroscopy and Mass Spectrometry: Application to Human Urine.** *J Proteome Res.* 2015; **14**(6): 2642–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
39. Bingol K, Brüschweiler R: **Two elephants in the room: new hybrid nuclear magnetic resonance and mass spectrometry approaches for metabolomics.** *Curr Opin Clin Nutr Metab Care.* 2015; **18**(5): 471–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Misra BB, van der Hoof JJ: **Updates in metabolomics tools and resources: 2014–2015.** *Electrophoresis.* 2016; **37**(1): 86–110.  
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Castillo S, Gopalacharyulu P, Yetukuri L, et al.: **Algorithms and tools for the preprocessing of LC–MS metabolomics data.** *Chemometr Intell Lab Syst.* 2011; **108**(1): 23–32.  
[Publisher Full Text](#)
42. **F** Karaman I, Nørskov NP, Yde CC, et al.: **Sparse multi-block PLSR for biomarker discovery when integrating data from LC–MS and NMR metabolomics.** *Metabolomics.* 2015; **11**(2): 367–79.  
[Publisher Full Text](#) | [F1000 Recommendation](#)
43. **F** Boccari J, Rutledge DN: **A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion.** *Anal Chim Acta.* 2013; **769**: 30–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
44. Marshall DD, Lei S, Worley B, et al.: **Combining DI-ESI-MS and NMR datasets for metabolic profiling.** *Metabolomics.* 2015; **11**(2): 391–402.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)



45. **F** Dumez JN, Milani J, Vuichoud B, *et al.*: **Hyperpolarized NMR of plant and cancer cell extracts at natural abundance.** *Analyst*. 2015; **140**(17): 5860–3. [PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
46. Ardenjaer-Larsen JH, Fridlund B, Gram A, *et al.*: **Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR.** *Proc Natl Acad Sci U S A*. 2003; **100**(18): 10158–63. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Dutta P, Martinez GV, Gillies RJ: **A New Horizon of DNP technology: Application to In-vivo <sup>13</sup>C Magnetic Resonance Spectroscopy and Imaging.** *Biophys Rev*. 2013; **5**(3): 271–81. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Schroeder MA, Cochlin LE, Heather LC, *et al.*: **In vivo assessment of pyruvate dehydrogenase flux in the heart using hyperpolarized carbon-13 magnetic resonance.** *Proc Natl Acad Sci U S A*. 2008; **105**(33): 12051–6. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. **F** Bornet A, Maucourt M, Deborde C, *et al.*: **Highly Repeatable Dissolution Dynamic Nuclear Polarization for Heteronuclear NMR Metabolomics.** *Anal Chem*. 2016; **88**(12): 6179–83. [PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
50. **F** Trupp M, Zhu H, Wikoff WR, *et al.*: **Metabolomics reveals amino acids contribute to variation in response to simvastatin treatment.** *PLoS One*. 2012; **7**(7): e38386. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
51. Armstrong AW, Wu J, Johnson MA, *et al.*: **Metabolomics in psoriatic disease: pilot study reveals metabolite differences in psoriasis and psoriatic arthritis [version 1; referees: 2 approved].** *F1000Res*. 2014; **3**: 248. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
52. Sitter B, Johnsson MK, Halgunset J, *et al.*: **Metabolic changes in psoriatic skin under topical corticosteroid treatment.** *BMC Dermatol*. 2013; **13**: 8. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Medina MA: **Glutamine and cancer.** *J Nutr*. 2001; **131**(9 Suppl): 2539S–42S; discussion 2550S–1S. [PubMed Abstract](#)
54. Wold S, Sjöström M, Eriksson L: **PLS-regression: A basic tool of chemometrics.** *Chemometr Intell Lab Syst*. 2001; **58**: 109–30. [Publisher Full Text](#)
55. **F** Heinemann J, Mazurie A, Tokmina-Lukaszewska M, *et al.*: **Application of support vector machines to metabolomics experiments with limited replicates.** *Metabolomics*. 2014; **10**(6): 1121–8. [Publisher Full Text](#) | [F1000 Recommendation](#)
56. Chen T, Cao Y, Zhang Y, *et al.*: **Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection.** *Evid Based Complement Alternat Med*. 2013; **2013**: 298183. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Patterson AD, Li H, Eichler GS, *et al.*: **UPLC-ESI-TOFMS-based metabolomics and gene expression dynamics inspector self-organizing metabolomic maps as tools for understanding the cellular response to ionizing radiation.** *Anal Chem*. 2008; **80**(3): 665–74. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
58. Boser BE, Guyon IM, Vapnik VN: **A training algorithm for optimal margin classifiers.** In: Haussler D, editor; Pittsburgh, Pennsylvania, United States. 1992; 144–152. [Publisher Full Text](#)
59. Furey TS, Cristianini N, Duffy N, *et al.*: **Support vector machine classification and validation of cancer tissue samples using microarray expression data.** *Bioinformatics*. 2000; **16**(10): 906–14. [PubMed Abstract](#) | [Publisher Full Text](#)
60. Mukherjee S, Tamayo P, Slonim D, *et al.*: **Support vector machine classification of microarray data.** *Tech Rep.* AI Memo; 1999. [Reference Source](#)
61. **F** Gaul DA, Mezencev R, Long TQ, *et al.*: **Highly-accurate metabolomic detection of early-stage ovarian cancer.** *Sci Rep*. 2015; **5**: 16351. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
62. Hsu CW, Lin CJ: **A comparison of methods for multiclass support vector machines.** *IEEE Trans Neural Netw*. 2002; **13**(2): 415–25. [PubMed Abstract](#) | [Publisher Full Text](#)
63. Bryan K, Brennan L, Cunningham P: **MetaFIND: a feature analysis tool for metabolomics data.** *BMC Bioinformatics*. 2008; **9**: 470. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
64. Breiman L: **Random Forests.** *Mach Learn*. 2001; **45**(1): 5–32. [Publisher Full Text](#)
65. **F** Gromski PS, Muhamadali H, Ellis DI, *et al.*: **A tutorial review: Metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding.** *Anal Chim Acta*. 2015; **879**: 10–23. [PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
66. Fan Y, Murphy TB, Byrne JC, *et al.*: **Applying random forests to identify biomarker panels in serum 2D-DIGE data for the detection and staging of prostate cancer.** *J Proteome Res*. 2011; **10**(3): 1361–73. [PubMed Abstract](#) | [Publisher Full Text](#)
67. Thompson IM, Pauler DK, Goodman PJ, *et al.*: **Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter.** *N Engl J Med*. 2004; **350**(22): 2239–46. [PubMed Abstract](#) | [Publisher Full Text](#)
68. Blekherman G, Laubenbacher R, Cortes DF, *et al.*: **Bioinformatics tools for cancer metabolomics.** *Metabolomics*. 2011; **7**(3): 329–43. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
69. **F** Zheng H, Ji J, Zhao L, *et al.*: **Prediction and diagnosis of renal cell carcinoma using nuclear magnetic resonance-based serum metabolomics and self-organizing maps.** *Oncotarget*. 2016; **7**(37): 59189–98. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
70. Lloyd GR, Wongravee K, Silwood CJ, *et al.*: **Self Organising Maps for variable selection: Application to human saliva analysed by nuclear magnetic resonance spectroscopy to investigate the effect of an oral healthcare product.** *Chemometr Intell Lab Syst*. 2009; **98**(2): 149–61. [Publisher Full Text](#)
71. **F** Wijetunge CD, Li Z, Saeed I, *et al.*: **Exploratory analysis of high-throughput metabolomic data.** *Metabolomics*. 2013; **9**(6): 1311–20. [Publisher Full Text](#) | [F1000 Recommendation](#)
72. Xia J, Wishart DS: **MetPA: a web-based metabolomics tool for pathway analysis and visualization.** *Bioinformatics*. 2010; **26**(18): 2342–4. [PubMed Abstract](#) | [Publisher Full Text](#)
73. **F** Ma X, Chi YH, Niu M, *et al.*: **Metabolomics Coupled with Multivariate Data and Pathway Analysis on Potential Biomarkers in Cholestasis and Intervention Effect of *Paonia lactiflora* Pall.** *Front Pharmacol*. 2016; **7**: 14. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
74. 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 Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
75. **F** Xia J, Sinelnikov IV, Han B, *et al.*: **MetaboAnalyst 3.0—making metabolomics more meaningful.** *Nucleic Acids Res*. 2015; **43**(W1): W251–7. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
76. **F** Rodriguez-Martinez A, Ayala R, Posma JM, *et al.*: **MetaboSignal: a network-based approach for topological analysis of metabolite regulation via metabolic and signaling pathways.** *Bioinformatics*. 2010; **26**(5): 773–5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
77. **F** Basu S, Duren W, Evans CR, *et al.*: **Sparse network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data.** *Bioinformatics*. 2017; **33**(10): 1545–53. [PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
78. Christmas R, Avila-Campillo I, Bolouri H, *et al.*: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Am Assoc Cancer Res Educ B*. 2005; **12**–16. [Reference Source](#)
79. Kanehisa M, Goto S, Sato Y, *et al.*: **Data, information, knowledge and principle: back to metabolism in KEGG.** *Nucleic Acids Res*. 2014; **42**(Database issue): D199–205. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
80. Ma H, Sorokin A, Mazein A, *et al.*: **The Edinburgh human metabolic network reconstruction and its functional analysis.** *Mol Syst Biol*. 2007; **3**: 135. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
81. Jeong H, Mason SP, Barabási AL, *et al.*: **Lethality and centrality in protein networks.** *Nature*. 2001; **411**(6833): 41–2. [PubMed Abstract](#) | [Publisher Full Text](#)
82. Gaupp R, Lei S, Reed JM, *et al.*: ***Staphylococcus aureus* metabolic adaptations during the transition from a daptomycin susceptibility phenotype to a daptomycin nonsusceptibility phenotype.** *Antimicrob Agents Chemother*. 2015; **59**(7): 4226–38. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
83. Nicholson G, Rantalainen M, Li JV, *et al.*: **A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection.** *PLoS Genet*. 2011; **7**(9): e1002270. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
84. **F** Tohge T, Nishiyama Y, Hirai MY, *et al.*: **Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor.** *Plant J*. 2005; **42**(2): 218–35. [PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
85. Vastrik I, D'Eustachio P, Schmidt E, *et al.*: **Reactome: a knowledge base of biologic pathways and processes.** *Genome Biol*. 2007; **8**(3): R39. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
86. **F** Caspi R, Billington R, Ferrer L, *et al.*: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.** *Nucleic Acids Res*. 2016; **44**(D1): D471–80. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)

## Open Peer Review

Current Referee Status:



---

### Editorial Note on the Review Process

**F1000 Faculty Reviews** are commissioned from members of the prestigious **F1000 Faculty** and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

---

### The referees who approved this article are:

#### Version 1

- 1 **Christoph Steinbeck** Friedrich-Schiller-University Jena, Jena, Germany  
**Competing Interests:** No competing interests were disclosed.
- 1 **Uwe Christians** University of Colorado, Aurora, Colorado, USA  
**Competing Interests:** No competing interests were disclosed.
- 1 **Brent Winston** University of Calgary, Calgary, AB, Canada  
**Competing Interests:** No competing interests were disclosed.