Published in final edited form as:

Curr Opin Biotechnol. 2017 February ; 43: 56–61. doi:10.1016/j.copbio.2016.08.005.

Increasing rigor in NMR-based metabolomics through validated and open source tools

Hamid R Eghbalnia, Pedro R Romero, William M Westler, Kumaran Baskaran, Eldon L Ulrich, and John L Markley

Biochemistry Department, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA

Abstract

The metabolome, the collection of small molecules associated with an organism, is a growing subject of inquiry, with the data utilized for data-intensive systems biology, disease diagnostics, biomarker discovery, and the broader characterization of small molecules in mixtures. Owing to their close proximity to the functional endpoints that govern an organism's phenotype, metabolites are highly informative about functional states. The field of metabolomics identifies and quantifies endogenous and exogenous metabolites in biological samples. Information acquired from nuclear magnetic spectroscopy (NMR), mass spectrometry (MS), and the published literature, as processed by statistical approaches, are driving increasingly wider applications of metabolomics. This review focuses on the role of databases and software tools in advancing the rigor, robustness, reproducibility, and validation of metabolomics studies.

Introduction

Within the discipline of systems biology, metabolomics complements other 'omics' (genomics, transcriptomics, and proteomics) and is particularly important because small molecules are more proximate to the networks that regulate the phenotype [1]. Mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy are complementary techniques often used for the detection and identification of metabolites [2,3]. More recently, NMR spectroscopy has been used extensively in evaluating and exploring the human metabolome for markers of disease states and in understanding the diversity of metabolic pathways in a variety of organisms. The strengths of NMR spectroscopy in metabolic studies include the variety of compounds that can be detected, the ability to quantify the concentrations of individual metabolites, and the flexibility to handle many different sample types. The major limitations of NMR spectroscopy are its low sensitivity and spectral resolution. The sensitivity issue can be addressed through sample concentration and/or the use of cryogenic probes [4], and the resolution issue can be addressed through the use of two-dimensional NMR approaches [5,6], increases in field strength [7], and new experimental techniques [8]. The utility of NMR as a rapid and non-destructive method has been bolstered by its decades-long successful use in key medicinal,

natural product, and organic chemistry application to identify and quantitate both major products and impurities in complex reaction mixtures [9].

Figure 1 outlines the main technological steps that enable NMR metabolomics. Metabolomics, and other omics investigations, produce vast amounts of both product-oriented and process-oriented data; thus their sole description as text in manuscripts is insufficient [10¹¹]. The process-oriented data collected for a specific research goal is likely to be useful in guiding decisions in the larger research enterprise by making information about best-practices and optimal analysis tools widely accessible. Downstream steps involving functional analysis of biological systems [11¹²], biomarker discovery [12], and medical diagnostics are no exception. Databases and computational tools play a key role in enabling the transfer of experimental information within and between research groups. Moreover, improved databases and tools can go a long way toward addressing the scientific requirements of reproducibility common to all omics sciences—a subject that has taken center stage in recent years. An external motivating factor, in the wake of the recent scientific malpractice scandals [13], is the increasing mandate by governments and funding agencies for reproducible research. There is tacit acknowledgement that the scientific impact of data-intensive omics research benefits significantly from improved data preservation, data sharing, and reproducible computation. This review addresses recent efforts toward the creation of databases, open software tools, and standardization, and summarizes the challenges these developments must meet in order to advance rigor in metabolomics.

The role of NMR metabolomics is not limited to the studies exemplified by the workflow in Figure 1. Metabolites are known to control material and energy flux in biochemical reactions and regulate biological processes through signaling cascades [14]. Metabolites interact with proteins enzymatically as substrates and products, or allosterically as cofactors or ligands [15,16]. Another use of NMR in metabolomics is in the emerging field of stable isotope tracers used for accurate pathway reconstruction [11¹⁷,17].

It is difficult to establish sharp boundaries between the processing, analysis, and interpretation steps, as intimated by Figure 1. The landscape of technology and tools is in constant flux, and a drive for integration and assimilation of multiple steps in the process creates shifting boundaries. The databases and software tools presented in this review have been selected on the basis of active development and update history—a tool or software has been included if it is available for download, if it has received an update in the past three years, and if most of its functions are not contained in another related package already included in the tables. Additional databases not included in this review are ones that primarily focus on mass spectrometry. A number of commercial technologies are also available, and additional details can be obtained from the vendors' web sites (ACD/NMR from ACD/Labs, AMIX from Bruker, Chenomx NMR from Chenomx, KnowItAll from Bio-Rad, DataChord/Spectrum Miner from One Moon Scientific, Mnova NMR from Mestrelab Research, TopSpin from Bruker, VnmrJ from Varian, CMC-assist from Bruker, NUTS from Acorn NMR). While these commercial tools satisfy some needs of the market, the key issue of openness, which impacts the requirements of reproducibility, still remains to be addressed.

Metabolomics databases

A large number of publicly-available databases contain information relevant to metabolomics studies (Table 1). Of these the two most heavily used in NMR-based metabolomics are the Human Metabolome Data Bank (HMDB) [18] and the small molecule database at the Biological Magnetic Resonance data Bank (BMRB) [19]. New repositories, for example the Metabolome Exchange, aggregate experimental data from multiple repositories. The National Institutes of Health (NIH) Common Fund Metabolomics Program has supported the establishment of a Data Repository and Coordinating Center (DRCC). The DRCC initiative is a response to the burgeoning development of databases. More process-oriented databases are likely to emerge that will make the sharing of data about the findings of studies, the detailed workflow, and the computational processes, more practical. The combination of product and process databases is expected to have a lasting impact on the requirement in metabolomics for scientific reproducibility. Because each database is being developed to address a specialized research need, an increasing number of data formats and content will continue to be added to the list of existing databases. While format differences raise issues of accessibility as well as interoperability, the continuing expansion of databases brings forward issues related to data integrity and validity. The provenance and priority of data are often hard to trace, and complex questions of data discrepancy can lead to significant issues regarding information integrity.

Software tools for NMR-based metabolomics

Various methods and strategies are used in NMR-based metabolomics (Table 2) [for a review, see [27]]. While tools for data processing and analysis have received more development attention (Figure 1), significant opportunities exist for advancing tools relevant to experimental design and data interpretation. For example, software for the creation and comparative analysis of results from alternative computational approaches and parametric settings can be helpful for evaluating apparent and actual differences that may arise as a result of options selected for data processing and analysis. Chemometric analysis is a popular approach with a history of use in other fields. In NMR, chemically related spectral pattern resolution may be sufficient for establishing the presence of differences in sample chemistry. The tool chest of chemometrics is diverse, and the development of new pattern-based approaches remains active (Table 2). PCA is one of the most popular approaches in the chemometrics tool chest. The practice often involves the preprocessing of NMR spectra in order to ‘align’ peaks and a binning process to ‘coarse grain’ the spectra. These steps may influence the outcome of analysis, and it is often left to the investigator to evaluate the evidence.

The quantitative nature of NMR spectroscopy makes it an ideal analytical method for comprehensive metabolite profiling. Beyond ^1H 1D NMR, new software and increasingly rich NMR metabolite spectral databases are advancing the unique abilities of NMR spectroscopy to identify and quantify concentrations and fluxes of small molecules [8]. Public and commercial databases now contain experimental 1D and 2D spectra and extracted spectral parameters for over a thousand compounds [28]. Public databases containing experimental NMR data from complex metabolic studies are emerging, but the

task of identification and quantification remains challenging. One promising approach involves the combination of MS and NMR data [29]. Software and resources to support quantitative NMR metabolomics have had an impact in opening the field to non-specialists with interest in quantification by streamlining the correct execution of several analytically challenging steps (Table 2). The analysis of 1D and 2D NMR spectra of complex biological mixtures continues to be performed mostly through manual processes that are likely to introduce bias. In cases involving large data sets, the task of manually processing hundreds, or perhaps thousands of spectra, may be impractical. Formal schemas, defined interfaces, and interoperability protocols are sorely needed.

Similar to many emerging technological fields, the specialization of tools in NMR metabolomics has proceeded along paths that include both increasing integration as well as discretization. Tools such as PhenoMeter [30] package several functionalities into a single integrated tool, while discrete tools, for example SpeckTackle [31] and jsNMR, address a discrete metabolomics software need — namely visualization. The ongoing effort toward standardization (e.g., nmrML) may enhance interoperability, reuse, and expansion of open source software by encouraging the development of discrete units of software with well-defined interfaces. The interpretation of data obtained from comprehensive profiling approaches requires the knowledge of biological pathways. A number of databases and visualization tools have been adapted for metabolomics use (Table 2). Despite advances offered by these packages, exciting and challenging research questions remain to be addressed. For example, more powerful data processing methods can provide robust measures of inferential support that can be integrated into the pathway analysis step. The integration of multiple steps, while convenient, can result in masking important analytic details. The use of standardized formats to create discrete pipelines has the advantage of functioning in many integrated workflows, as well as offering improvements in reproducibility.

Future developments

A number of initiatives to standardize metabolomics data formats are underway, and these efforts are expected to face the same growing pains as those in other omics fields. In the EU, the ‘coordination of standards in metabolomics’ (COSMOS) is developing infrastructure and exchange standards for metabolomics within the European metabolomics community [48]. Along with standardization, national and international funding agencies are increasingly requesting publicly funded research data and software to become *open access*. In practice, change in standards are the result of social, technical, economic, political, and legal activities, which include strategic decisions that can have a dramatic effect on day to day operations of the research enterprise. Moreover, the impact felt by changing practices is not homogeneous. For example, defining a new data exchange protocol has a significant impact on technology developers, but may have little or no impact on technology users. On the other hand, the use of standard vocabularies and protocols within defined laboratory management systems has its highest impact on technology users. Nonetheless, apt technological innovations can find balanced solutions that address the tension between standardization and the scientific endeavor.

For example, decoupling data exchange, data storage, and compute-cycle delivery would enable continued scientific innovation at the computational level, while, at the same time, the goals of standardization and data preservation are met. One approach to decoupling data exchange and storage is to adopt a flexible data model. The STAR archival data format of BMRB, adopted as NMR-STAR, [49], is a widely accepted model that could serve the role of a decoupled data exchange protocol [50]. It is a backward compatible self-defining format in which every data item has attributes that describe its features, including explicit definitions of relationships among data items. It is human-readable, and no domain knowledge is required to read the files. The format is easily converted to XML, it supports the requisite RDF (Resource Description Framework), and its flexibility provides a strong hedge against the potential of future disruptions due to format. The decoupling of compute-cycle delivery can be achieved by using a virtual machine as the technology delivery vehicle. Moreover, the use of virtual machines simplifies the long-term preservation of data and procedures. More broadly, we can improve the odds for success by adopting newer design methodologies and architectures.

Conclusions

To enable the precise reproduction of results, it is important to define reporting requirements associated with experimental design, data acquisition and variable manipulation during data processing, downstream statistical analysis, and data interpretation. For reproducible and extensible research, minimum additional requirements include the description of processes, software used, and support for data and software versioning. The scientific enterprise is an increasingly interconnected activity in which data exchange and data preservation are both essential requirements.

Acknowledgments

Research of the authors received support from NIH R01 GM109046 and P41GM103399 (to JLM), P41GM111135 (to HRE).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Kim J, Reed JL. Refining metabolic models and accounting for regulatory effects. *Curr Opin Biotechnol.* 2014; 29:34–38. [PubMed: 24632483]
2. Aretz I, Meierhofer D. Advantages and pitfalls of mass spectrometry based metabolome profiling in systems biology. *Int J Mol Sci.* 2016:17.
3. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods Mol Biol.* 2015; 1277:161–193. [PubMed: 25677154]
4. Ramaswamy V, et al. Development of a (1)(3)C-optimized 1. 5-mm high temperature superconducting NMR probe. *J Magn Reson.* 2013; 235:58–65. [PubMed: 23969086]
5. Bingol K, et al. Quantitative analysis of metabolic mixtures by two-dimensional ¹³C constant-time TOCSY NMR spectroscopy. *Anal Chem.* 2013; 85:6414–6420. [PubMed: 23773204]

6. Clendinen CS, et al. (13)C NMR metabolomics: INADEQUATE network analysis. *Anal Chem*. 2015; 87:5698–5706. [PubMed: 25932900]
7. Ardenkjaer-Larsen JH, et al. Increase in signal-to-noise ratio of >10,000 times in liquid-state NMR. *Proc Natl Acad Sci U S A*. 2003; 100:10158–10163. [PubMed: 12930897]
8. Bingol K, et al. Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis*. 2016; 8:557–573. [PubMed: 26915807]
9. Halabalaki M, et al. Recent advances and new strategies in the NMR-based identification of natural products. *Curr Opin Biotechnol*. 2014; 25:1–7. [PubMed: 24484874]
- 10. Rocca-Serra P, et al. Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics*. 2016; 12:14. This reference provides an overview of the current state of standards for data sharing in metabolomics, and investigates the broader impact of standardization in some detail. In this context, the authors observe that, with the increasing amounts of data being produced, mere description of investigations as text in manuscripts is no longer sufficient to enable reuse. In addition to discussions related to some of the recent developments in standardization, and the links between standardization and the objectives of reproducible research, other important, but less commonly discussed topics associated to social and political perspectives of standardization are touched upon. The manuscript offers a strong rationale for standardization, and notes that further incentives are needed to ensure the production of standards-compliant data sets. As examples, these incentives may include new opportunities to publish data sets, additional requirements in the form of author guidelines in scientific journals, and guidance for submission of data sets to public repositories. While, the social, economic, and political aspects of standardization and data sharing could be further elaborated, the technical and scientific arguments present a valuable contribution. This work does not address challenges faced by traditional standardization models as a result of the current high velocity production of distributed, heterogeneous, and overlapping data. [PubMed: 26612985]
- 11. Fan TW, Lane AN. Applications of NMR spectroscopy to systems biochemistry. *Prog Nucl Magn Reson Spectrosc*. 2016; 92–93:18–53. The utility of NMR spectroscopy in metabolomics is discussed in detail. The material related to NMR applications in stable isotope-enriched tracer research and its use for elucidating biochemical pathways and networks is comprehensive and well-elucidated. In addition to a detailed discussion of several areas in which NMR has unparalleled advantages for metabolic studies, the review illustrates the uses of NMR for spatially selective *in vivo* imaging and dynamical analysis of metabolism in tissues of living organisms. Various direct and isotope-edited 1D and 2D NMR methods suitable for profiling metabolites and their isotopomer distributions by stable isotope-resolved metabolomics (SIRM) analysis are described. The importance of sample preparation (rapid cryoquenching, efficient extraction, and chemoselective derivatization) is highlighted. A section on future developments in NMR technologies and reagents identifies potential tools that could further empower NMR in systems biochemical research. Overall, the review makes a strong case for why NMR is uniquely suited for performing quantitative and rigorous analysis, especially as it relates to isotopomer analysis.
12. Emwas AH, et al. Recommendations and standardization of biomarker quantification using NMR-based metabolomics with particular focus on urinary analysis. *J Proteome Res*. 2016; 15:360–373. [PubMed: 26745651]
13. Stern AM, et al. Financial costs and personal consequences of research misconduct resulting in retracted publications. *Elife*. 2014; 3:pe02956.
14. Kimmelman AC. Metabolic Dependencies in RAS-Driven Cancers. *Clin Cancer Res*. 2015; 21:1828–1834. [PubMed: 25878364]
15. Menzies KJ, et al. Protein acetylation in metabolism —metabolites and cofactors. *Nat Rev Endocrinol*. 2016; 12:43–60. [PubMed: 26503676]
16. Stark JL, et al. NMRmix: a tool for the optimization of compound mixtures in 1D (1)H NMR ligand affinity screens. *J Proteome Res*. 2016; 15:1360–1368. [PubMed: 26965640]
17. Zamboni N, Saghatelian A, Patti GJ. Defining the metabolome: size, flux, and regulation. *Mol Cell*. 2015; 58:699–706. [PubMed: 26000853]
18. Wishart DS, et al. HMDB 3.0 — the human metabolome database in 2013. *Nucleic Acids Res*. 2013; 41(Database issue):D801–D807. [PubMed: 23161693]

19. Ulrich EL, et al. BioMagResBank. *Nucleic Acids Res.* 2008; 36(Database issue):D402–D408. [PubMed: 17984079]
- 20. Sud M, et al. Metabolomics Workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 2016; 44:D463–D470. The Metabolomics Workbench, which is part of the data coordinating effort of the National Institute of Health (NIH) Common Fund's Metabolomics Program, is discussed. It is a public repository for metabolomics metadata and experimental data. The scope is extensive and includes experimental platforms, metabolite standards, metabolite structures, protocols, tutorials, and training material and other educational resources. As currently described, the precise integration model of several prominent NMR databases, and numerous expanding metabolomics data repositories, remains unclear. For example, two of these databases, the HMDB and the BMRB, are freely available and highly utilized electronic databases containing information about metabolites. HMDB uses the BMRB small molecule archive extensively for primary (time-domain) and processed (images of spectra and peak lists) NMR data for over 1200 compounds, including most metabolites found at high abundance in biological fluids and tissue extracts. The combination of HMDB and BMRB contains spectra available for most compounds including ¹H 1D, ¹³C 1D, ¹³C 90° DEPT, ¹³C 135° DEPT, ¹H–¹H COSY, ¹H–¹H TOCSY, ¹H–¹³C HSQC, and ¹H–¹³C HMBC. Any integration must address potentially duplicate information — for example, the InChI string, the canonical and isomeric SMILES representation, and Beilstein name, PubChem Substance (SID), PubChem Compound (CID), KEGG: Compound ID, CAS Registry IDs, PDB Chemical Component. [PubMed: 26467476]
21. Kale NS, et al. MetaboLights: an open-access database repository for metabolomics data. *Curr Protoc Bioinformatics.* 2016; 53:14-13-1–14-13-18.
22. Steinbeck C, Krause S, Kuhn S. NMRShiftDB-constructing a free chemical information system with open-source components. *J Chem Inf Comput Sci.* 2003; 43:1733–1739. [PubMed: 14632418]
23. Yue Y, et al. TMDb: a literature-curated database for small molecular compounds found from tea. *BMC Plant Biol.* 2014; 14:p243.
24. Kelder T, et al. WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.* 2012; 40(Database issue):D1301–D1307. [PubMed: 22096230]
25. Jewison T, et al. SMPDB 2. 0: big improvements to the Small Molecule Pathway Database. *Nucleic Acids Res.* 2014; 42(Database issue):D478–D484. [PubMed: 24203708]
- 26. Fabregat A, et al. The Reactome pathway Knowledgebase. *Nucleic Acids Res.* 2016; 44:D481–D487. The most recent update of Reactome Knowledgebase is an extended version of a classic metabolic map, encapsulated in a single consistent data model. Reactome provides molecular details of signal transduction, transport, DNA replication, metabolism and other cellular processes as an ordered network of molecular transformations. In addition to its functions as an archive of biological processes, Reactome contains tools for discovering unexpected functional relationships in data. Along with WikiPathways, the two are the most popular freely available databases for biological pathways. Reactome pathways are centrally curated by selected domain experts while WikiPathways are created and continually curated by any interested party. There is a nascent collaboration between WikiPathways and Reactome because of the mutual benefits of combining these two approaches. The new release represents the ongoing expansion of the Reactome database in the past three years as it continues to integrate new information. The new information has overlaps with a number of existing databases, and is likely to overlap in information content with metabolomics database efforts in progress. The continued expansion and specialization of databases, such as Reactome, offers an illustration of the challenge faced by traditional models of standardization. [PubMed: 26656494]
27. Bharti SK, Roy R. Metabolite identification in NMR-based metabolomics. *Curr Metabolomics.* 2014; 2:163–173.
28. Bingol K, et al. Customized metabolomics database for the analysis of NMR (1)H–(1)H TOCSY and (1)(3)C–(1)H HSQC-TOCSY spectra of complex mixtures. *Anal Chem.* 2014; 86:5494–5501. [PubMed: 24773139]
29. Bingol K, et al. Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid identification of new metabolites in complex mixtures. *Anal Chem.* 2015; 87:3864–3870. [PubMed: 25674812]

30. Carroll AJ, et al. PhenoMeter: a metabolome database search tool using statistical similarity matching of metabolic phenotypes for high-confidence detection of functional links. *Front Bioeng Biotechnol.* 2015; 3:106. [PubMed: 26284240]
31. Beisken S, et al. SpeckTackle: JavaScript charts for spectroscopy. *J Cheminform.* 2015; 7:17. [PubMed: 25984241]
- 32. Mohamed A, Nguyen CH, Mamitsuka H. NMRPro: an integrated web component for interactive processing and visualization of NMR spectra. *Bioinformatics.* 2016; 32(13):2067–2068. NMRPro is a web-based, interactive, client-server model for NMR data processing. This is one of the first web-based NMR data processing systems that is provided in open-source format. The open source format is helpful for future repeatability and reproducibility of results. NMRPro also addresses the lack of necessary processing and interactive visualization functionalities in current NMR metabolomics applications by providing a web component that can be easily incorporated into existing web applications. Interactive online processing and visualization of data using a client server model is implemented by integrating the server-side processing with client-side interactive visualization through three parts: a Python package to efficiently process large NMR datasets on the server-side, a Django App managing server-client interaction, and SpecdrawJS for client-side interactive visualization. Although a number of web-based tools are available, and new processing tools such as Bayesil address the need for more sophisticated statistical processing, the open source release of NMRPro is significant in that it deals with a key element of reproducibility requirements that other web-based tools will need to address. NMRPro's level of robustness and functionality has yet to match that of more sophisticated tools, for example, MVPACK, which provides an open-source platform for performing standardized protocols on one-dimensional ¹H NMR metabolomics data. [PubMed: 27153725]
33. Hedjazi L, et al. mQTL. NMR: an integrated suite for genetic mapping of quantitative variations of (1)H NMR-based metabolic profiles. *Anal Chem.* 2015; 87:4377–4384. [PubMed: 25803548]
34. Ravanbakhsh S, et al. Accurate, fully-automated NMR spectral profiling for metabolomics. *PLoS One.* 2015; 10:e0124219. [PubMed: 26017271]
35. Beisken S, et al. KNIME-CDK: workflow-driven cheminformatics. *BMC Bioinformatics.* 2013; 14:257. [PubMed: 24103053]
36. Chignola F, et al. The CCPN Metabolomics Project: a fast protocol for metabolite identification by 2D-NMR. *Bioinformatics.* 2011; 27:885–886. [PubMed: 21216771]
37. Worley B, Powers R. MVAPACK: a complete data handling package for NMR metabolomics. *ACS Chem Biol.* 2014; 9:1138–1144. [PubMed: 24576144]
38. Xia J, et al. MetaboAnalyst 3. 0 — making metabolomics more meaningful. *Nucleic Acids Res.* 2015; 43:W251–W257. [PubMed: 25897128]
39. Lewis IA, Schommer SC, Markley JL. rNMR: open source software for identifying and quantifying metabolites in NMR spectra. *Magn Reson Chem.* 2009; 47(Suppl 1):S123–S126. [PubMed: 19821464]
40. Klein MS, Oefner PJ, Gronwald W. MetaboQuant: a tool combining individual peak calibration and outlier detection for accurate metabolite quantification in 1D (1)H and (1)H–(13)C HSQC NMR spectra. *Biotechniques.* 2013; 54:251–256. [PubMed: 23662895]
41. Delaglio F, et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* 1995; 6:277–293. [PubMed: 8520220]
42. Chylla RA, et al. Deconvolution of two-dimensional NMR spectra by fast maximum likelihood reconstruction: application to quantitative metabolomics. *Anal Chem.* 2011; 83:4871–4880. [PubMed: 21526800]
43. Sansone SA, et al. The first RSBI (ISA-TAB) workshop: “can a simple format work for complex studies? OMICS. 2008; 12:143–149. [PubMed: 18447634]
44. Karnovsky A, et al. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics.* 2012; 28:373–380. [PubMed: 22135418]
45. Jauhiainen A, et al. Normalization of metabolomics data with applications to correlation maps. *Bioinformatics.* 2014; 30:2155–2161. [PubMed: 24711654]
46. Karp PD, et al. Pathway Tools version 13. 0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform.* 2010; 11:40–79. [PubMed: 19955237]

47. Savorani F, Tomasi G, Engelsen SB. icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. *J Magn Reson.* 2010; 202:190–202. [PubMed: 20004603]
48. Salek RM, et al. COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access. *Metabolomics.* 2015; 11:1587–1597. [PubMed: 26491418]
49. Ulrich EL, et al. STAR/CIF macromolecular NMR data dictionaries and data file formats. *Acta Crystallogr.* 1996; A52:C577–C577.
50. Fitzgerald, PMD., et al. 4.5 Macromolecular dictionary (mmCIF). In: Hall, SR., McMahon, B., editors. *International Tables for Crystallography G. Definition and Exchange of Crystallographic Data.* Springer; 2005. p. 295-443.

Figure 1.

Process, decisions, and outcomes in a metabolomics study. A study is defined by its goals and the decisions in each step along the path from experimental design to data interpretation. Process-oriented information documents the ‘how’ and ‘why’ — the specific actions taken, and relevant reasons for the actions, in order to create an outcome. Product-oriented information identifies the ‘what’ — the product. Additional information, related to who and where, for example, are also sometimes stored. Access to goal-specific and method-specific information at every step of the process improves the productivity of the research. Databases can also preserve and share key information regarding sample collection and storage practices as well as sample preparation methods.

Table 1

Collated list of recently updated metabolomics databases. Additional annotation for items marked with ‘*’ is provided in the references.

Databases
<i>Name</i>
BMRB [19]
HMDB [18]*
Metabolomics Workbench [20]*
MetaboLights [21]
NMRShiftDB [22]
TMDB [23]
WikiPathways [24]
SMPDB [25]
Reactome [26*]

Table 2

Software tools that perform a range of functions as noted by the headers in the table. Additional annotation for items marked with ‘*’ is provided in the references.

NMR data processing Spectral processing	Pattern analysis Chemometrics	Profiling Identification/quantification
NMRPro [32]*	mQTL [33]*	COLMAR/TOCCATA [8]*
Bayesil [34]*	KNIME [35]	CCPN [36]
MVAPACK [37]*	PhenoMeter [30]*	MetaboAnalyst 3.0 [38]*
rNMR [39]		MetaboQuant [40]
NMRPipe [41]		Newton [42]
NMR data processing		Omic integration
Specialized resources		Network analysis
ISA-TAB [43]		Grinn
SpeckTackle [31]		MetScape [44]
Metabnorm [45]		Bio-Cyc [46]
Icoshift [47]		rWikipathways