

Establishing Reporting Standards for Metabolomic and Metabonomic Studies: A Call for Participation

OLIVER FIEHN,¹ BRUCE KRISTAL,² BEN VAN OMMEN,³ LLOYD W. SUMNER,⁴
SUSANNA-ASSUNTA SANSONE,⁵ CHRIS TAYLOR,⁵ NIGEL HARDY,⁶
and RIMA KADDURAH-DAOUK⁷

ABSTRACT

Metabolite concentrations in cellular systems are very much dependent on the physiological, environmental, and genetic status of an organism and are regarded as the ultimate result of cellular regulation, resulting in the visible phenotypes. Therefore, the comprehensive analysis of metabolite levels and fluxes renders a suitable tool for assessing the degree of perturbation in biological systems. Lessons derived from development of other OMICS areas (genomics, proteomics, and transcriptomics) have shown that large-scale comparisons and interpretations will require the re-use of data over long periods of time and by multiple laboratories with different expertise and backgrounds. Reaching this goal will require standardization of reporting structures of metabolomic studies for journal publication purposes, for regulatory deposition, and for database dissemination. An initiative by the Metabolomics Society is presented that aims to define important aspects of metabolomic workflows. These include biological study designs, chemical analysis, and data processing, as well as the ontologies that are necessary in this framework.

This paper is part of the special issue of OMICS on data standards.

INTRODUCTION

METABOLOMICS aims at quantifying and identifying all metabolites in a given biological context (Fiehn, 2001; Wikipedia, 2005)—an aim that admittedly is as challenging as full-scale proteomics. The objective of metabolomics targets at assessing metabolic changes in a comprehensive and global manner in order to infer biological functions or provide detailed biochemical responses of cellular systems. Accordingly, the mission of the Metabolomics Society (2005) is to promote the international growth and devel-

¹UC Davis Genome Center, Davis, California.

²Burke Rehabilitation Hospital, White Plains, New York.

³TNO Food and Nutrition Research, AJ Zeist, The Netherlands.

⁴Noble Foundation, Ardmore, Oklahoma.

⁵EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom.

⁶Department of Computer Science, University of Wales, Aberystwyth, United Kingdom.

⁷Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina.

opment of the field of metabolomics, to provide the opportunity for collaboration among researchers in metabolomics, including connections between academia, government and industry, regulators and vendors in the field of metabolomics, and to provide opportunities for dissemination of research achievements in workshops, conferences, and journal publications.

GROWTH OF STANDARDIZATION

It was soon realized that the very name of the society—“metabolomics”—has led to a certain level of misunderstanding, because an important part of the work on globally assessing metabolism has originated from biomedical focused research, with large emphasis on toxicity and drug-related metabolic responses in mammalian organisms. This field had been termed “metabonomics” (Lindon et al., 2003; Wikipedia, 2005). The difference in the meaning between these two words might be valued by insiders, but could be more confusing for biologists who just want to understand metabolic responses in their favorite organism. Still, these two terms and associated (and again slightly different) areas of “metabolite profiling,” “metabolic fingerprinting,” “lipidomics,” and others provided the motivation for the Metabolomics Society to initiate an effort to standardize the reporting of experiments, parameters of the studies, projects, and associated result data. The basic rationale behind the approach to standardize reporting of metabol/nomics data is the understanding, that metabolism readily responds to both minor and major systems perturbations in a highly flexible manner. The value of standardized reporting of OMICS data has generally been appreciated (Quackenbush, 2004), and for metabolomics, it is critical in three different views:

1. When reporting (summarized) data and interpretations in classical journals, a detailed description of biological and experimental conditions is needed to follow the author’s arguments, and, if desired, to repeat the study or to design careful variations of it.
2. When reporting individual (processed, but not summarized) data in journals or databases, even more metadata are needed to enable researcher to reproduce results.
3. When reporting metabolomic data in larger databases, a large variety of metadata needs to be amassed not only to enable reproduction of general findings, but indeed to facilitate re-using data to enable discovery of new (unforeseen) relationships between metabolic events and associated metadata in order to unravel novel biological functions and relationships.

There is a broad consensus that the outcome of a metabolomics study (comprising biological perturbations and measurements of the corresponding metabolic responses) very much depends on the conditions of two major parameters—“biology” and “experimental measurement.” The ability to quickly and reliably compile metabolic data on many organisms and perturbations hence calls for reporting standards that equally call for both “study design” and “method” metadata. Consequently, the discussion on standardization of metabolomic reporting needs to distinguish these parameters, each of which may further be broken down into subparameters such as “method(s) for data acquisition,” “method for data processing,” and “method for data reporting.” Detailed reporting of all aspects of the pipeline is also expected to support development, dissemination, and validation of best practice.

There are a couple of advantages (and some disadvantages) of metabolomic approaches when comparing to other OMICS techniques. Among the advantages, there are two striking features. First, the quantitative and structural analysis of metabolites leans on technologies that have been developed over more than 50 years, some of which being mature enough to have evolved standard operating procedures and validated benchmark databases of metabolite concentrations (Ridley et al., 2004). Second, metabolomics compares more favourably to proteomics or transcriptomics with respect to the price-per-sample for data acquisition. This enables researchers to acquire more data at lower cost, which fosters statistical analyses of studies which may involve potentially considerable inherent biological variation within populations.

The advent of modern technologies, most notably high-speed computers, has allowed sample analyses with continually improving limits of detection with increased sample turnover and greater metabolome coverage. Consequently, a number of analytical approaches have been developed for metabolomics and metabonomics research that are novel or at least unusual with respect to classical analytical chemistry of metabo-

lites. Some platforms and data export formats are shared with proteomics (such as mass spectrometry), and some aspects of data transformation and processing are similar to proteomics and transcriptomics. When the Metabolomics Society (2005) launched its own initiative on standardization efforts in October 2004, it consequently sought the advice from and interact with initiatives that have already paved the way in other OMICS fields, such as the Microarray Gene Expression Data (MGED) society (Ball and Brazma, *this issue*) and the HUPO Proteomics Standards Initiative (PSI) (Taylor et al., *this issue*). The objective of these interactions is to develop a draft document within the year 2006 that may guide further discussions and improvements before it might be implemented into standard reporting prerequisites required by journals, scientific bodies, regulatory offices, or funding agencies. Although the initiative is in its early beginning phase, a number of workshops, discussions, and reports have preceded and qualified the route of work (NIH, 2005; Jenkins et al., 2004; Lindon et al., 2005; MetaboMeeting, 2005). All consortia and individuals involved in discussing “metabolomics reporting standards” agree that consensus is sought in an inclusive manner, bringing in the various backgrounds of expertise and research objectives. Researchers aiming at plant metabolomics have started discussions on standardizations in this field using the term MIAMET (Minimum Information About a METabolomics experiment) and proposed a way to annotate unidentified metabolic signals (Bino et al., 2004). The Standard Metabolic Reporting Structure (SMRS; www.smrsgroup.org/) discussion group started from a pharmacological and toxicological background involving major industry representatives, resulting in important reports (Lindon et al., 2005) and detailed recommendations (SMRS, 2005). Consequently, it has been recognized as a timely task to combine and strengthen these ongoing efforts in a concerted effort, coordinated by the Metabolomics Society. An Oversight Committee on Reporting Standards has been formed, and five working groups (some of which are divided into further subgroups) have been established:

1. Biological context metadata (Working Group A)
2. Chemical analysis (Working Group B)
3. Data processing (Working Group C)
4. Ontology (Working Group D)
5. Data exchange (Working Group E)

The structure of the working groups thus follows the general workflow model in metabolomics: from a description of the study design to sample workup, data acquisition, processing and export, bound together by controlled vocabularies and relationships between the terms used. The general aim of all five working groups is to describe, but not prescribe, the study itself and the technical details in an effort to reach a broad consensus in the community on which minimal requirements need to be given.

Due to the dependence of metabolic biosynthesis and turnover rates on the actual physiological and environmental conditions, a large emphasis is given to the group dealing with biological context (Working Group A) which is subdivided into (1) mammalian studies, including clinical research and nutrigenomics, (2) plant studies, (3) cell cultures and microbiology, and (4) environmental studies. For the case of nutrition metabolomics is assumed to become an approach of major importance (as nutrition is all about metabolites, metabolism and its relation to health). Consequently, a very accurate description of the biological setting (e.g., study design, period of fasting, pre- or post-prandial sampling, run-in period) is necessary. Especially the issue of a quantified food intake prior and during the study is highly relevant, as this strongly influences the metabolome in organs, plasma and urine. This area is covered by a joint effort of the European Nutrigenomics Organisation (NuGO, 2005) and the American Society of Nutrition Sciences (ASNS). The discussion in environmental studies area is lead by the NERC Environmental Bioinformatics Centre (NEBC) (Field et al., 2005) and aims to fulfil the diverse needs of those working on the metabolomics of environmentally relevant organisms, which are not covered by the model organism community. NuGO and the environmental communities also synergize under the Reporting Structure for Biological Investigation (RSBI) working group (Sansone et al., *this issue*). Although incubated under the MGED Society umbrella, it is the intent of RSBI to work in the wider functional genomics context, engaging in extensive outreach and liaison activity with several efforts. “Multi-OMICS” approaches to biological research (e.g., nutrition, environment and toxicology) will heavily impact the structure and content of reporting structures, as well as ontologies and exchange formats being developed by technology-driven efforts.

REPORTING STANDARDS FOR METABOLOMIC AND METABONOMIC STUDIES

The challenge for each of these groups is to distinguish minimal descriptors of a study from the current best practice which may change according to new insights. For example, different types of light bulbs in a plant climate chamber will affect the metabolome depending on both the age of the bulbs and the light emission spectra. However, reporting such details cannot be made mandatory because such detailed descriptions go far beyond that required from renowned journal editors, societies or agencies working in plant sciences. One way to approach this problem is to outline the framework for study descriptions (Jenkins et al., 2004), but allow actual each individual research group to go beyond minimal requirements in order to lead by example how more detailed reporting structures may be implemented. This way, case studies may be helpful for guiding and successively improving reporting standards in metabolomics.

Examples for the other four working groups may highlight the intentions and paths that are taken to develop reporting standards. For example, a variety of Standard Operating Procedures for sample preparation and technology platforms is used for quantification of metabolites, each associated to a different set of descriptors. Consequently, Working Group B has initiated discussions on minimal requirements to describe such technical details, that may reach out for accurate quantifications using stable isotope labelling or even flux measurements (the turnover of metabolites through a given metabolic pathway). For example, use of different types of columns in liquid chromatography will result in exclusion of either highly polar compounds (in “reversed phase” mode) or of highly lipophilic compounds (in “normal phase” mode). Therefore, a description of the details of separation techniques used prior to mass spectrometric detection is mandatory. Working Group C takes on data that originate from data acquisition platforms in order to outline criteria that are needed to describe raw data processing schemas, including normalization and data transformation methods. The data processing group (Working Group C) will also aim at recommendations to report results gained by statistical methods. The ontology group (Working Group D) is tasked to tackle the semantics issue. The first phase of the work focuses on the collection of controlled vocabularies to describe the experimental components, including design, materials, instruments, and data types. These vocabularies will be defined and—where required—a list of synonyms will be created. Controlled vocabularies aim to provide a set of descriptors for the consistent semantic understanding of data across these disparate data sources. Conforming to general accepted view that duplication and incompatibility should be avoided where possible, the ontology working group will reach out, evaluate and leverage previous and relevant work done by the MGED Society (Ball and Brazma, *this issue*) and HUPO-PSI (Taylor et al., *this issue*), as well as existing and emerging works in the metabolomics field. The ultimate goal is to develop ontology-based knowledge representations that have proved to be successful in providing the semantics for standardised annotation, integration and exchange of data. In this second phase of the work, the group will engage in the Functional Genomics Ontology (FuGO) project, a wider collaborative effort, bringing together transcriptomics and proteomics communities (Whetzel et al., *this issue*). The larger scientific community will best be served if the resulting ontology overcomes duplications across omics studies where commonality of the concepts exists.

Working Group E is concerned with data exchange formats. Data may be generated and stored using a variety of technologies, some open and some proprietary. A data exchange standard provides a means by which data can be transmitted between producers and storage, with guarantees of common semantics and estimable validity. Any format should consider the relevant reporting requirement specifications to be key use cases, and should therefore as a minimum support (but not require) the capture of all the metadata, analytical data (spectroscopic and chromatographic) and analyses specified therein. A fundamental view of this group is that the reuse of existing standards from this and other fields is extremely desirable, as is collaboratively working on new standards where none yet exist. Initiatives from this and other fields will be carefully considered for their applicability to modern metabolomics. This includes, in the analytical data area for example, the mzData standard from the Mass Spectrometry Working Group of HUPO-PSI (Taylor et al., *this issue*), the Collaborative Computing Project for the NMR Community (CCPN, 2006) and NMR-STAR (2006).

CONCLUSION

The aim is to develop a draft document outlining the basic principles of such reporting standards during the year 2006. The authors wish to encourage readers to participate in this process on any level and regis-

ter to the mailing lists hosted by the Metabolomics Society website (<http://msi-workgroups.sourceforge.net>). The process itself gears to achieve a community consensus which obviously cannot be reached without heavy interactions and debates. An important corner stone will be the upcoming annual meeting of the Metabolomics Society in June 2006 in Boston (Metabolomics Society, 2005) where progress will be reported to the larger metabolomics community, with further discussions within the different fields, for example, during the 2006 Fourth International Plant Metabolomics Conference (Plant Metabolomics, 2005). Eventually, it is important to understand that –omics research, especially metabolomics, is not about applying fancy technologies or any particular type of instrument or approach, but rather about gaining biological information in a comprehensive way using a workflow that ties sophisticated study designs to global analysis of cellular molecules and appropriate data processing and interpretation of resulting structured data. Metabolomics is not about the latest turn in NMR or mass spectrometry, but about gaining insight into metabolic regulation on a global scale.

ACKNOWLEDGMENTS

We acknowledge the contributions of the participants and co-authors of work related to standardization and reporting efforts in metabolomics. Specifically, we are thankful for the researchers who agreed to chair the working groups in the recent initiative on Metabolomics Standards, and the members of the Metabolomics Society’s oversight committee on Reporting Standards (see the Metabolomics Society website for a complete list of Reporting Standards chairs and committee members). We are particularly thankful for Arthur Castle who organized the Metabolomics Standards workshop for the NIH (NIDDK, NIEHS, NIGMS) in conjunction with the Metabolomics Society, presided by Dr. Rima Kaddurah-Daouk.

REFERENCES

- BINO, R.J., HALL, R.D., FIEHN, O., et al. (2004). Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* **9**, 418–425.
- CCPN. (2006). Collaborative Computing Project for the NMR Community. Available at: www.ccpn.ac.uk/datamodel/datamodel.html.
- Nugo. (2005). European Nutrigenomics Organisation. Available at: www.nugo.org/metabolomics
- FIEHN, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics* **2**, 155–168.
- FIELD et al. (2005).
- JENKINS, H., HARDY, N., BECKMANN, M., et al. (2004). A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* **22**, 1601–1605.
- LINDON, J.C., HOLMES, E., and NICHOLSON, J.K. (2003). So what’s the deal with metabonomics? Metabonomics measures the fingerprint of biochemical perturbations caused by disease, drugs, and toxins. *Analyt Chem* **75**, 384A–391A.
- LINDON, J.C., NICHOLSON, J.K., HOLMES, E., et al. (2005). Summary recommendations for standardization and reporting of metabolic analyses. *Nat Biotechnol* **23**, 833–838.
- METABOLOMICS SOCIETY. (2005). Mission statement. Available at: <http://metabolomicssociety.org/mission.html>.
- METABOLOMICS SOCIETY. (2005). Oversight Committee on “Reporting Standards.” Available at: <http://metabolomicssociety.org/mstandards.html>.
- METABOMEETING. (2005). MetaboMeeting 1.0 conference. Available at: http://smrsgroup.sourceforge.net/mm_report.html.
- METABOLOMICS SOCIETY. (2005). Available at: <http://metabolomicssociety.org/boston2006/>.
- WIKIPEDIA. (2005). Available at: <http://en.wikipedia.org/wiki/Metabonomics>.
- WIKIPEDIA. (2005). Available at: <http://en.wikipedia.org/wiki/Metabolomics>.
- NIH. (2005). National Institutes of Health metabolomics workshop. Available at: www.nidkk.nih.gov/fund/other/metabolomics2005/.
- NMR-STAR. (2006). www.bmrwisc.edu/.
- PLANT METABOLOMICS. (2005). Plant Metabolomics Conference. Available at: www.pmet06.org/index.html.

REPORTING STANDARDS FOR METABOLOMIC AND METABONOMIC STUDIES

- QUACKENBUSH, J. (2004). Data standards for “omic” science. *Nat Biotechnol* **22**, 613–614.
- RIDLEY, W.P., SHILLITO, R.D., COATS, I., et al. (2004). Development of the international life sciences institute crop composition database. *J Food Compos Analy*, **17**, 423–438.
- SMRS. (2005). Standard Metabolic Reporting Structure. Available at: www.smrsgroup.org/.

Address reprint requests to:

Dr. Oliver Fiehn
UC Davis Genome Center
451 E. Health Sci. Dr.
Davis, CA 95616

E-mail: ofiehn@ucdavis.edu