Review Article **Open Access**

The Future of Metabolomic Profiling in Population-Based Research: Opportunities and Challenges

Krista A Zanetti[1], Eliza Mette[1], Padma Maruvada[2], John Milner[3], Steven C Moore[4], Holly L Nicastro[5], Sharon A Ross[5], Joshua N Sampson[4], Mukesh Verma[1] and Joseph Su L[1*]

[1]*Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, USA*
[2]*Division of Digestive Diseases and Nutrition, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, USA*
[3]*Agricultural Research Service, United States Department of Agriculture, Beltsville, USA*
[4]*Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, USA*
[5]*Divison of Cancer Prevention, National Cancer Institute, Rockville, USA*

Abstract

Metabolomics is an approach that employs technologies to measure small molecule metabolites in biological samples, thus providing epidemiologists and other investigators with a means to discover biomarkers of disease risk, diagnosis, and prognosis. To advance the field of metabolomics, the National Institutes of Health (NIH) is investing $65 million through the NIH Common Fund's Metabolomics Program, which will support comprehensive metabolomics resource cores, training in metabolomics, metabolomics technology development, metabolomics reference standards synthesis, and data sharing and international collaboration. While this infrastructure will be essential, there remain several challenges to broad implementation of metabolomics in population-based studies. To facilitate the use of metabolomics in population-based studies, the NIH-sponsored 'Think Tank on the Use of Metabolomics in Population-Based Research' was held to discuss the current opportunities and challenges of the field and identify potential solutions and/or strategies to address challenges. Insights and conclusions gained from the Think Tank are summarized here.

Keywords: Metabolomics; Epidemiology; Population; Biomarkers; Profiling

Introduction

The suffix "-omics" refers to the collective technologies that measure the components of a large family of cellular molecules, such as genes (genomics) or proteins (proteomics), and examine the function of and communication between the various types of cellular molecules. Metabolomics is the study of small molecules of both endogenous and exogenous origin, such as metabolic substrates and their products, lipids, small peptides, vitamins and other protein cofactors, generated by metabolism [1,2]. This approach, which uses the analytical techniques of mass spectrometry and nuclear magnetic resonance, has received more attention in recent years as an ideal methodology to unravel signals closer to the culmination of the disease process. The compounds identified through metabolomic profiling represent a range of intermediate metabolic pathways that may serve as biomarkers of exposure, susceptibility or disease [3-6]. Therefore, it is a valuable approach for deciphering metabolic outcomes with a phenotypic change.

The monitoring of relative changes in metabolomic profiles in predisposed versus healthy individuals may help identify unique metabolites involved in disease processes [3]. These profiles have been used to predict the risk of diabetes [7-9], cardiovascular disease [4], and lung cancer [10], diagnose prostate cancer [11], differentiate benign and malignant ovarian tumors [12], and identify biomarkers of Crohn's disease [13]. Such shifts also may identify diagnostic biomarkers, which could provide insights into strategies for disease prevention and be used to monitor the response to treatment. In recent years, metabolomics and other post-genome-wide association study (GWAS) platforms, such as proteomics and transcriptomics, have undergone rapid improvement in both their reliability and throughput; as such, it may be an opportune time for their use in epidemiologic studies [14-17]. Although metabolomic profiling has been used in some larger-scale population studies [18-21], the number of published reports to date remains small. If the successes of genomics and transcriptomics in epidemiology are reliable indicators, there is a large, yet untapped, potential for metabolomics to contribute to public health research.

The National Institutes of Health (NIH) invested $14.3 million in fiscal year 2013, and will potentially invest more than $51.4 million over five years, to accelerate this field. This investment was established through the NIH Common Fund's Metabolomics Program and will support the following core areas: comprehensive metabolomics resource cores, technology development, reference standards synthesis, and training and educational activities in metabolomics [22]. To date, six Regional Comprehensive Metabolomics Resource Cores have been established across the United States at University of Michigan, Research Triangle Institute, University of California, Davis, University of Florida, University of Kentucky and Mayo Clinic. Additionally, the Metabolomics Data Repository and Coordination Center at the University of California, San Diego (UCSD) serves as a national data repository that will coordinate data sharing and dissemination activities from national and international metabolomics centers. Other components include technology development, training and development of courses and workshops. In 2012, the National Heart, Lung, and Blood Institute also funded the Anchoring Metabolomic

Changes to Phenotype P20 Project, which has the mission to use metabolomics to capture the molecular information that is most proximal to a cardiovascular or lung phenotype [23]. These investments should spark strategies to advance the field of metabolomics for broader biomedical and public health research, as well as provide additional capacity to support metabolomics analyses in population-based studies.

The 'Think Tank on the Use of Metabolomics in Population-Based Research' meeting, held at the NIH in September of 2012, intended to address critical issues in preparation for the effective use of metabolomic profiling in large population studies. Approximately 100 scientists from academia, government, non-profit organizations, and industry were invited to discuss the current state of the field and identify strategies to advance the use of this approach in population-based research. This review provides a brief synopsis of this workshop, discusses the opportunities and challenges for moving this technology into epidemiologic studies, and highlights what epidemiologists should consider for metabolomics being successfully incorporated in future population-based studies.

Think Tank on the Use of Metabolomics in Population-Based Research

The foci of the 'Think Tank on the Use of Metabolomics in Population-Based Research' were: 1.) goals and opportunities for moving metabolomic technology to population-based research, 2.) integration of data for metabolomics studies, and 3.) resources and infrastructure needed to support metabolomics studies in population-based research. Presentations by Drs. Svati Shah (Duke University), Oliver Fiehn (University of California, Davis), Bruce Kristal (Brigham and Women's Hospital and Harvard Medical School), Robert Gerszten (Harvard Medical School), John Milner (USDA), and Mukesh Verma (NCI) described the current state of the field, as well as how the NIH is currently supporting metabolomics research. These presentations were followed by facilitator-led breakout sessions, in which the participants were charged with answering three key questions in each area of focus outlined above, and are briefly summarized below.

Opportunities and Challenges for the Use of Metabolomics in Population-Based Research

Integration of metabolomics data with other "omics" data: Metabolomic profiling is of particular interest in population-based research due to its potential to evaluate the effects of environmental exposures, conduct risk assessments, predict disease development, diagnose disease, and monitor disease progression. The proximity of metabolomic signals to disease processes when compared with other "omics" technologies offer the possibility to better identify disease mechanisms, thereby increasing a functional understanding. Yet, a barrier to the application of this approach is the lack of well-characterized intermediate phenotypes for most diseases. The integration of metabolomic data with other "omics" data, including data that are publically available, will enable novel interrogation approaches and discovery of functional implications [24]. Therefore, methods development for integration of metabolomics data with other "omics" data must be made a priority.

Study design and sample collection: Although opportunities exist for the use of metabolomic profiling in population-based research, multiple challenges prevent the proper integration of these data into population-based studies for meaningful interpretation. Therefore, investigators should strive to understand the principles of metabolomics to better determine and apply their appropriate uses. In addition, it is essential for metabolite profiles to be validated both for analytical purposes and clinical use as biomarkers. Some of the primary challenges to population-based studies include the incorporation of this technology in the initial study design, identifying appropriate sample collection protocols and quality control methods, and the selection of analytical approaches and quantification techniques. New strategies will likely be necessary for combining data from different analytical platforms to allow generalizability of data and interpretation. Additionally, pilot and feasibility studies will need to be conducted to assess intra-individual metabolite stability, technical precision, and the effect of sample storage and processing. Examining the effect of sample storage on metabolite variability is especially important given the number of prospective studies that have archived biological samples. Furthermore, as the number of samples assayed increases, the volume of data generated through the use epidemiology studies has the potential to become difficult to manage. Therefore, the need to develop better methods for analyzing large amounts of data remains a critical barrier, including improvement of statistical and bioinformatic methods for data analysis.

Unfortunately, metabolomic profiling is used far too frequently post-hoc in epidemiology studies. When the opportunity arises, investigators should strive for well-designed, prospective studies that would establish causal effect, as well as temporal changes. In any case, causality is difficult to establish with a clear biological explanation in association studies, even with particularly well-designed studies. Therefore, it is critical to improve methodologies for integration of metabolomics data with other data, such as genomic and proteomic, to help parse out the functional and temporal relationship between a biomarker and an effect.

A multitude of factors should be considered in selecting the specific analytical technique to be adopted for metabolomics studies. Although not specific to population-based studies, it is important to understand, and in turn, identify the most appropriate analytical technique to use for the goals of the study or hypothesis being tested. Two major technologies, mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR), are generally considered as they can measure hundreds to thousands of unique chemical entities. MS is highly sensitive and has the capacity to detect metabolites with concentrations in the picomolar range and above, requires small biospecimen volumes, and enables metabolites to be individually identified and quantified. However, MS has relatively lower analytical reproducibility, requires more complex software and algorithms for routine data analysis, and poorly represents highly polar metabolites when using standard chromatography protocols. Although there are well-established techniques to analyze polar metabolites by MS, the use of multiple techniques for each sample to generate data increases costs and diminishes productivity. In contrast, NMR allows for the comprehensive generation of metabolite profiles by a single nondestructive method, is fully automated, inherently quantitative, and has very high analytical reproducibility with a well-established mathematical and statistical tool box. The disadvantages to using NMR include its relative insensitivity in detecting metabolites with concentrations in the micromolar range and below. Furthermore, the validity of both NMR and MS is dependent on the quality of sample collection and handling, as well as the available metadata. Before applying either of these technologies to population-based studies, investigators must consider the advantages and disadvantages in relation to the design and aims of the study.

Availability of standards: Another barrier is a paucity of commercially available standards for identification and quantification of metabolites for humans, as well as data comparisons across studies. Standard reference materials should be developed that can be used broadly to examine a wide range of metabolites and are suitable for applications that interrogate complex human matrices. Particular consideration should be given to the quantity needed and which compounds should be included. Additionally, the lack of carefully selected, well-annotated, and easily accessible reference samples greatly limits investigations. Although the standard National Institute of Standards and Technology (NIST) pooled plasma reference set is available [25], there needs to be further effort put toward the development of additional reference standards. For example, standards for other biological media, such as urine, are needed for investigating the most physiologically plausible pathways that best reflect the etiology of disease. In the case of epidemiologic consortia, different laboratories may be involved in sample analyses; and in turn, inter-laboratory comparisons are problematic without reference samples.

Quality control: An opportunity surrounds the use of archived samples from population-based studies to gain insights into optimizing collection and storage protocols for different media for sample integrity. In addition to high quality samples, quantitative robustness requires provision for quality controls, pooled references, and standard reference materials to control for instrumentation variability drift and allow for comparisons across laboratories. Well-standardized protocols for sample collection, storage, and analysis are also needed.

Central database and data sharing: One major challenge is the lack of standardized data sharing methods. Undeniably, the barriers to data sharing are complex, including large storage requirements, conflicting formats, and the high cost of storing large quantities of data. These issues are less attributed to data size, although still a factor, than to the lack of storage structure, inherent differences in the output from differing technologies, and requirement for standardized essential metadata. An opportunity exists to address data sharing through the new Data Coordinating Centers established by the NIH Common Fund and is a priority for the Data Repository and Coordination Center at UCSD. Additionally, there are efforts of this nature taking place in Europe, specifically The European Bioinformatics Institute's (EBI) MetaboLights [27]. Furthermore, these issues will be addressed through the NIH Big Data to Knowledge (BD2K) initiative, which is focused on how to best address the management and use of the large amounts of biomedical data generated by new technologies [28]. Currently, data storage and sharing responsibilities remain with individual laboratories, and will continue as such, until better data sharing protocols are developed. The field may consider following in the footsteps of genomics and focus on the establishment of a central database, similar to dbGap, which is used for data sharing of genomic data [29], to archive and distribute the results of metabolomics studies. The framework for establishing a database of this nature is already in place, and there are clear benefits to data sharing, including the use of shared data for replication, verification, and secondary data analysis studies, as well as the cost savings associated with using shared data for these analyses. An essential step will be to establish a mechanism for data sharing that builds on the assignment of unique accession/identification numbers to analytes, and would need to be addressed in data sharing protocols and metadata requirements. It will not only allow for data sharing, but general data mining and pathway analyses, as well as literature searches using citation databases, including PubMed.

Future Directions

Consideration should initially be given to easily accessible short-term goals, such as planning for metabolomic analyses during study design, the establishment of a mechanism for data sharing, and capitalizing on archived samples from population-based studies to optimize collection and storage protocols. Once the short-term goals are addressed, the more complex barriers can then be tackled. As a number of population-based studies using metabolomic profiling have already emerged, now is the time to be vigilant to assure continued advancement in the field. There is a significant need for increased communication and collaboration among epidemiologists, biochemists, statisticians and other stakeholders to tackle the issues outlined in this commentary. The 'Think Tank on the Use of Metabolomics in Population-Based Research' was the initial step in opening the lines of communication amongst these groups; and the provocative conversation needs to continue. Although there are currently many identified challenges for the application of metabolomic profiling in population-based studies, recent advancements continue to transform metabolomics technologies into more reliable and sought after platforms.

References

1. Claudino WM, Quattrone A, Biganzoli L, Pestrin M, Bertini I, et al. (2007) Metabolomics: available results, current research projects in breast cancer, and future applications. J Clinical Oncol 25: 2840-2846.

2. Wang H, Tso VK, Slupsky CM, Fedorak RN (2010) Metabolomics and detection of colorectal cancer in humans: a systematic review. Future Oncol 6: 1395-1406.

3. Sampson JN, Boca SM, Shu XO, Stolzenberg-Solomon RZ, Matthews CE, et al. (2013) Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. Cancer Epidemiol Biomarkers Prev 22: 631-640.

4. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, et al. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472: 57-63.

5. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, et al. (2014) An atlas of genetic influences on human blood metabolites. Nat Genet 46: 543-550.

6. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, et al. (2014) The food metabolome: a window over dietary exposure. Am J Clin Nutr 99: 1286-1308.

7. Zhao X, Fritsche J, Wang J, Chen J, Rittig K, et al. (2010) Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. Metabolomics 6: 362-374.

8. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, et al. (2011) Metabolite profiles and the risk of developing diabetes. Nat Med 17: 448-453.

9. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, et al. (2012) Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol 8: 615.

10. Mathé EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, et al. (2014) Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. Cancer Res 74: 3259-3270.

11. Abate-Shen C, Shen MM (2009) Diagnostics: The prostate-cancer metabolome. Nature 457: 799-800.

12. Zhang T, Wu X, Ke C, Yin M, Li Z, et al. (2013) Identification of potential biomarkers for ovarian cancer by urinary metabolomic profiling. J Proteome Res 12: 505-512.

13. Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, et al. (2009) Metabolomics reveals metabolic biomarkers of Crohn's disease. PLoS One 4: e6386.

14. Chadeau-Hyam M, Ebbels TM, Brown IJ, Chan Q, Stamler J, et al. (2010) Metabolic profiling and the metabolome-wide association study: significance level for biomarker identification. J Proteome Res 9: 4620-4627.

15. Dumas ME, Maibaum EC, Teague C, Ueshima H, Zhou B, et al. (2006) Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. Anal Chem 78: 2199-2208.

16. Loo RL, Chan Q, Brown IJ, Robertson CE, Stamler J, et al. (2012) A comparison of self-reported analgesic use and detection of urinary ibuprofen and acetaminophen metabolites by means of metabonomics: the INTERMAP Study. Am J Epidemiol 175: 348-358.

17. Yap IK, Brown IJ, Chan Q, Wijeyesekera A, Garcia-Perez I, et al. (2010) Metabolome-wide association study identifies multiple biomarkers that discriminate north and south Chinese populations at differing risks of cardiovascular disease: INTERMAP study. J Proteome Res 9: 6647-6654.

18. Gieger C, Geistlinger L, Altmaier E, Hrabé de Angelis M, Kronenberg F, et al. (2008) Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet 4: e1000282.

19. Gaudet MM, Falk RT, Stevens RD, Gunter MJ, Bain JR, et al. (2012) Analysis of serum metabolic profiles in women with endometrial cancer and controls in a population-based case-control study. J Clin Endocrinol Metab 97: 3216-3223.

20. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, et al. (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 457: 910-914.

21. Zeleniuch-Jacquotte A, Shore RE, Afanasyeva Y, Lukanova A, Sieri S, et al. (2011) Postmenopausal circulating levels of 2- and 16α-hydroxyestrone and risk of endometrial cancer. Br J Cancer 105: 1458-1464.

22. National Institutes of Health (2014) The Common Fund; Metabolomics. National Institutes of Health, Office of Strategic Coordination.

23. Washington University School of Medicine in St. Louis (2012) Anchoring Metabolomic Changes to Phenotype P20 Project.

24. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, et al. (2010) A genome-wide perspective of genetic variation in human metabolism. Nat Genet 42: 137-141.

25. Phinney KW, Ballihaut G, Bedner M, Benford BS, Camara JE, et al. (2013) Development of a Standard Reference Material for metabolomics research. Anal Chem 85: 11732-11738.

26. National Institutes of Health (2014) Metabolomics Workbench. The NIH Common Fund and UC San Diego.

27. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, et al. (2013) MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res 41: D781-D786.

28. National Institutes of Health (2013) Request for Information (RFI): Training Needs in Response to Big Data to Knowledge (BD2K) Initiative. National Institutes of Health Office of Extramural Research.

29. NCBI Resource Coordinators (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 41: D8-D20.