

Individual metabolism should guide agriculture toward foods for improved health and nutrition¹⁻³

Steven M Watkins, Bruce D Hammock, John W Newman, and J Bruce German

ABSTRACT Genomics and bioinformatics have the vast potential to identify genes that cause disease by investigating whole-genome databases. Comparison of an individual's genotype with a genomic database will allow the prescription of drugs to be tailored to an individual's genotype. This same bioinformatic approach, applied to the study of human metabolites, has the potential to identify and validate targets to improve personalized nutritional health and thus serve to define the added value for the next generation of foods and crops. Advances in high-throughput analytic chemistry and computing technologies make the creation of a vast database of metabolites possible for several subsets of metabolites, including lipids and organic acids. In creating integrative databases of metabolites for bioinformatic investigation, the current concept of measuring single biomarkers must be expanded to 3 dimensions to 1) include a highly comprehensive set of metabolite measurements (a profile) by multiparallel analyses, 2) measure the metabolic profile of individuals over time rather than simply in the fasted state, and 3) integrate these metabolic profiles with genomic, expression, and proteomic databases. Application of the knowledge of individual metabolism will revolutionize the ability of nutrition to deliver health benefits through food in the same way that knowledge of genomics will revolutionize individual treatment of disease with pharmaceuticals. *Am J Clin Nutr* 2001;74:283-6.

KEY WORDS Metabolomics, lipids, metabolites, genomics, bioinformatics, agriculture, medicine, nutrition

INTRODUCTION

There is little doubt that improving the nutritional properties of food will be of considerable value to both the public and the agricultural industry; however, there is no consensus on how to achieve this goal, particularly on how to identify which specific target molecules to change. The uncertainty arises, ironically enough, not from a limited knowledge of either plant or human genetics, but from the older and much considered question, "What is nutritious?" Currently, the goal of nutritional improvement of agriculture is to produce changes in crops and foods that provide health benefits to all people. Few modifications to existing commodities that meet such a standard can be imagined because of the complex and heterogeneous makeup of humans. Recent mistakes made while attempting to improve the food supply highlight this conundrum. β -Carotene was interpreted to be a highly beneficial

ingredient, aiding in the protection against certain cancers, especially lung cancer. In response to this interpretation, a major shift in the carotenoid content of the food supply was underway when 2 large intervention trials—the Carotene and Retinol Efficacy Trial in the United States (1) and the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study in Finland (2)—discovered that high intakes of β -carotene as a supplement actually increased the incidence of lung cancer in smokers. Such unexpected responses among specific populations exemplify why the food and agriculture industries cannot afford to target improvements in health with single modifications in food composition based on conclusions inferred from the measurement of single biomarkers. Thus, while agriculture necessarily waits for nutrition researchers to identify the next biomarker of global health, it may be more important to develop alternative strategies to assess the meaning of health in more individualistic terms. To meet this challenge, it is reasonable to look to the success of genomics for guidance.

AN INFORMATIC APPROACH

Genomics was developed to simultaneously identify the elements of heredity and to assign biological function to these elements. Despite the inherent complexity of the genome, the invention of just a few molecular tools enabled genomics to flourish into the science we know today. It is reasonable to assume that in the near future, most common genetic diseases will be identified. The power of the knowledge emerging from the genome is that the identification of the genetic basis of an inherited disease provides logical strategies to treat those afflicted on an individual basis. Beyond its application to diseases with demonstrably genetic causes, however, the direct utility of genomics by itself

¹From the Departments of Food Science and Technology and of Entomology and the Cancer Research Center, University of California, Davis; Lipomics Technologies, Inc, West Sacramento, CA; and the Nestle Research Center, Lausanne, Switzerland.

²Supported in part by grants R01 ES02710, P42 ES04699, and P30 ES05707 from the National Institute of Environmental Health Sciences and grant 97-35302-4406 from the US Department of Agriculture.

³Reprints not available. Address correspondence to SM Watkins, Lipomics Technologies, Inc, 2545 Boatman Avenue, West Sacramento, CA 95691. E-mail: steve.watkins@lipomics.com.

Received November 27, 2000.

Accepted for publication March 19, 2001.

TABLE 1
Nomenclature of “omics”

Subject	Inclusive set (for an individual)	Statistical study (many individuals)
Genes	Genome ¹	Genomics
Messenger RNA	Expressed genome ²	Expression analysis
Proteins	Proteome ²	Proteomics
Metabolites	Metabolome ³	Metabolomics
Phenotype	Phenome ³	Phenomics

¹Inclusive set finite and not responsive to environment.

²Inclusive set responsive to environment, but limited by genotype.

³Inclusive set responsive to environment and not limited by genotype.

diminishes. Ultimately, changes in phenotype and not changes in genes are of interest to nutrition and health. The gap between genes and phenotype is spanned by many biochemical steps, each with individual specificities and a sensitivity to diet and the environment. Metabolites are the quantifiable molecules that best reflect phenotype; however, modern biological informatics has yet to embrace the study of metabolites as aggressively as genomics. An understanding of the many factors that contribute to health is necessary and will increase as technologies that quantify metabolites within individuals improve. Technologies that allow the creation of and interaction with accessible annotated databases of metabolite concentrations that reflect individuals with various phenotypes are also needed; fortunately, such technologies are available. In this context, substantial databases of metabolite concentrations will be predictive resources to quantify the relation between metabolites and health. Therefore, metabolite informatics, or metabolomics (Table 1), is the logical next step in understanding the role of nutrition in modifying metabolism and ultimately in promoting health.

Individual metabolite measurements as assessments of biology are by no means new. The application of an informatic approach to the study of metabolites in individuals, however, represents an important advance. Currently, many scientists still view their goal as ultimately reductionist and strive to identify the single best biomarker that reflects phenotype. The inadequate success of single biomarkers in predicting chronic disease attests to the need for more global and integrated approaches for assessing metabolism. Thus, the study of metabolites must be redefined in parallel with genomic and proteomic analyses as the means to allow researchers to measure a large number or even an entire set of metabolites. The entire metabolome, with all of its individual concentrations and quantitative intrarelations, forms the metabolic basis of a phenotype. Therefore, only a metabolomic approach can accurately assess the complex role of metabolites when defining individual health. In part, the reluctance to study metabolism within the framework of informatics arises from the inherent complexity of metabolite profiling. Although expression analysis and proteomics (Table 1) are responsive to the environment and are thus more complex than genomics, they are constrained, at least in theory, by the number of genes present in an organism. The overall metabolome is not confined to the products of genes, and thus, the metabolome represents a potentially massive, inclusive set of compounds. Furthermore, a metabolite profile for a single individual is neither constant among individual cells, nor is it stable over time. Implementing a metabolomic research strategy involves planning for considerable complexity.

LIPID METABOLOMICS

The technologies to quantify every metabolite in a system simultaneously have not emerged, making metabolomics currently impractical in a global sense. However, for certain subsets of metabolites, especially lipids, present analytic methods exist that produce a spectrum of data easily developed into a metabolomic database. Currently, fatty acids, glycerolipids, sterols, and numerous bioactive lipid mediators, including products of epoxygenase, lipoxygenase, and cyclooxygenase pathways, are quantifiable in biological samples. Thus, a few parallel analyses are capable of defining a nearly complete lipid profile of a sample.

Lipids are an attractive subset of metabolites for the first practical steps in metabolomic applications. In addition to their ubiquitous cellular functions as structural, energetic, and bioactive signaling molecules, lipids reflect both diet and metabolism. Fatty acids and sterols, the major components of most lipids, survive digestion intact and enter cellular metabolic pathways retaining distinct intramolecular features of their biological origin. Thus, a unique aspect of lipid metabolomic analysis is that the information yielded by an experiment reflects the ultimate expression of genomics, proteomics, and environment as a lipid metabolome. Metabolomics, if approached as the establishment of a substantial database containing quantitative information on the lipid metabolites within a tissue or blood compartment, is not an implausible ambition. To the contrary, the technology to define the lipid metabolome is currently available.

Early attempts to use a metabolomic strategy for investigating phenotype have proven valuable across a broad spectrum of biological research. In microbiology, changes in metabolite profiles were used to describe the global metabolic response and variable glucose metabolism of *Escherichia coli* under different growth conditions (3). Metabolome analyses were also used to identify the global changes in *E. coli* metabolism caused by changes in population density (4). Raamsdonk et al (5) used metabolomic analyses of yeast to identify the metabolic function of deleted genes for which there was no observable phenotypic consequence of their deletion. Use of metabolomics to identify the function of genes shows the versatility and power of metabolomics. Unlike genomics and proteomics, metabolomics can be used to identify changes that occur at all levels of biology, from genetics to the environment. The direct results of nutritional, genomic, or expression differences can be observed in a metabolite profile. This strategy is also widely accepted in plant research as a method of screening for desirable traits and of understanding the phenotypic expression of genes (6, 7). In humans, metabolomic analyses of amino acid profiles have been used to screen newborns for inborn errors of metabolism (8). The success of metabolomics in this wide variety of applications shows that human metabolomics can be expanded beyond screening for overt genetic defects to a system that identifies the optimal nutritional needs of an individual.

The utility of metabolite profiling is not limited to making assessments about the status of individuals; in fact, one advantage of metabolomics is the potential to use the database as a tool for *in silico* investigations. The availability of such databases will be particularly helpful for applying bioinformatic approaches to nutrition because once a metabolic profile is developed for a specific nutritional state, it can be compared with the metabolomic database to determine the relations between diet and phenotype. The ability to mine large databases

in silico will be an advantage of metabolomics to nutrition because testing every conceivable nutrient by single clinical trials is not possible.

APPLICATION OF METABOLOMICS TO NUTRITION AND AGRICULTURE

Integrating metabolomics with genomic and proteomic strategies will be useful in traditional medical applications, eg, discovering drug targets, defining drug efficacy, and preventing side effects. However, individual metabolomic assessment is absolutely essential to nutritional applications initiated at the level of agricultural improvement. Accurate health predictors are needed to translate health recommendations ranging from nutritional opinions to the production of agricultural commodities and food in general because of the costs and time involved in implementing such changes. The first generation of attempts to apply genetic modification to improving agricultural traits, such as pest resistance, showed that the production of new agricultural crops requires a substantial international effort at all levels of the agribusiness community. Production of new crops for nutritional reasons will be even more difficult. Knowing that a nutritional advantage is genuine and either valuable to a large segment of the population or clearly beneficial to a specific population is indispensable before improvements can succeed. Such confidence cannot be generated by a single isolated biomarker, but can only be generated by a more global approach.

An example of the value of combining the use of accurate metabolite assessments with quantitative databases is exemplified by the potential to apply the knowledge gained to redesigning the lipid composition of oil seed crops. Fatty acid composition has a major effect on the value of crops. The difference in price between coconut oil and canola oil relates in part to the composition of fatty acids in each and their putative roles in altering the risk of cardiovascular disease. The significant potential of genomics or cultivar selection to modify the fatty acid composition of oil seed crops to improve their health value is currently untapped because it is not yet clear what composition would be most beneficial for even a significant subset of the population. In effect, it is necessary to map the overall metabolomic expression of lipid metabolism to ensure that a particular metabolite composition truly improves overall health. Such a comprehensive approach is necessary to ensure that a change in lipid composition that is deemed healthy when viewed as lowering the risk of one disease does not simultaneously raise the risk of developing another. For example, blood biomarkers that assess the effects of dietary oils on health include plasma triacylglycerols and cholesterol. These biomarkers are useful in predicting cardiovascular disease, but not for predicting other disease endpoints. Beyond their use as phenotypic predictors, these markers are not especially informative about the metabolic pathways that create or modify them because they are derived from multiple pathways and locations. Thus, total triacylglycerol or cholesterol measurements do not indicate how and why individuals differ in their response to dietary oils. Historically, the choice to limit analyses of blood to a few select biomarkers was based on limitations in technologies capable of generating more comprehensive analyses and on the informatic tools capable of interpreting them. Neither of these limitations persists. Technologic tools to allow researchers a comprehensive metabolomic approach to study individual metabolism are now available.

DEVELOPING METABOLOMICS

Integrated databases

For metabolomics to develop a global knowledge base analogous to the genome knowledge, it is imperative that data be produced and reported in quantitative terms. Although it is typical for metabolite data to be reported as a percentage of the total or nonquantitative format, these data are influenced by the number of analytes and covariation, are not comparable between experiments, and provide little basis for interpreting how metabolites interact among themselves and with other biomolecules. Truly quantitative data can be integrated from multiple sources into a single seamless database, regardless of the number of metabolites measured in each discrete analysis. Thus, abandoning rigorously quantitative methodology in return for high-throughput analyses will yield fragmented and non-integratable data and certainly nonusable databases.


Biochemical and structural metabolomics

Although researchers can begin profiling metabolite compositions in earnest using current technology, there are analytic hurdles to developing a complete knowledge of metabolites and their roles in biology. Particularly difficult to assess by current methods is the information inherent in the structure of a cell or compartment that cannot be observed from a strictly biochemical analysis. Examples of these important structural domains include subcellular sphingomyelin-cholesterol raft complexes and subsets of metabolites that associate or act at specific locations. Because this structural information is not easily assessed by present analytic technologies, it is necessary to parse metabolomics into 2 distinct phases: biochemical metabolomics and structural metabolomics. As biochemical profiling proceeds, it will be important to recognize that structural and functional domains exist and that they are essential aspects of the function of metabolites in biology. The technologies to define structural interaction are not, however, far behind.

SUMMARY

The knowledge emerging from genomics will revolutionize our view of human health. Determining how food interacts with individual genotypes, however, requires approaches other than those typically associated with functional genomics. A metabolomic strategy provides the potential to go beyond improvements in medicine to the development of an understanding of how diet can modulate health. Once metabolic catalogs of individual phenotypic states are generated, this knowledge will be used as the means to provide therapies or to define optimal metabolism specific to an individual. Health and performance will go beyond the simple definition of health as “freedom from disease” and will redefine metabolism for optimum protection, prevention, and performance. Moving from the use of single biomarkers to a metabolomic analysis is a necessary step because many approaches to lowering the unilateral risk of one disease in an individual simply increase the risk of another disease in the same individual. A pertinent example of this problem is the change in nutritional recommendations from the consumption of high-fat to high-carbohydrate diets. It is widely understood that high-fat diets increase serum LDLs and thus the risk of cardiovascular disease in most individuals; however, high-carbohydrate diets increase serum triacylglycerols (9) and the risk of cardiovascular disease in a subset of the population, particularly in some women (10). The measurement of every metabolite

involved in lipid metabolism allows for subtle differences in the predisposition or progression of disease among individuals to be elucidated. The broader and much more exciting aspect of this technology is thus the generation of metabolic profiles that are not simply markers of disease but are metabolic maps that can be used to identify specific genes or activities influential in the progression of disease or the maintenance of overall health. In these applications, metabolomics is a subset of functional genomics and should be pursued as vigorously as should expression and proteomic strategies. The value of genomic, expression, proteomic, and metabolomic databases in predicting phenotype will be enhanced dramatically by the next logical step—their horizontal integration into global bioinformatic databases.

The ultimate application of these approaches, of course, is to generate knowledge of metabolism as it relates to health, predispositions to disease, or other health outcomes. In nutritional terms, for example, understanding the variation in metabolic responses to diet is the goal of the science of nutrition. Before embarking on wholesale renovations of agricultural products for nutritional improvement, metabolomics is uniquely qualified to address the questions that must be answered to succeed. Although the development of this technology is likely to be driven by human health concerns, rapid identification of lipids and other metabolite classes will be useful in a variety of applications, such as plant and animal breeding, the characterization of transgenic crops, and basic research. Metabolomics will be a part of the future of nutrition and agriculture and is currently practical by means of profiling lipid metabolites. 

We gratefully acknowledge the writing and editing support of Cora Dillard.

REFERENCES

1. Omenn GS, Goodman GE, Thornquist MD, et al. Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. *J Natl Cancer Inst* 1996;88:1550–9.
2. Albanes D, Heinonen OP, Huttunen JK, et al. Effects of alpha-tocopherol and beta-carotene supplements on cancer incidence in the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study. *Am J Clin Nutr* 1995;62(suppl):1427S–30S.
3. Tweeddale H, Notley-McRobb L, Ferenci T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool (“metabolome”) analysis. *J Bacteriol* 1998;180:5109–16.
4. Liu X, Ng C, Ferenci T. Global adaptations resulting from high population densities in *Escherichia coli* cultures. *J Bacteriol* 2000;182:4158–64.
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;19:45–50.
6. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey R, Willmitzer L. Metabolite profiling for plant functional genomics. *Nat Biotechnol* 2000;18:1157–61.
7. Glassbrook N, Beecher C, Ryals J. Metabolic profiling on the right path. *Nat Biotechnol* 2000;18:1157–61.
8. Rashed MS, Bucknall MP, Little D, et al. Screening blood spots for inborn errors of metabolism by electrospray tandem mass spectrometry with a microplate batch process and a computer algorithm for automated flagging of abnormal profiles. *Clin Chem* 1997;43:1129–41.
9. Kasim-Karakas SE, Almario RU, Mueller WM, Peerson J. Changes in plasma lipoproteins during low-fat, high-carbohydrate diets: effects of energy intake. *Am J Clin Nutr* 2000;71:1439–47.
10. Liu S, Willett WC, Stampfer MJ, et al. A prospective study of dietary glycemic load, carbohydrate intake, and risk of coronary heart disease in US women. *Am J Clin Nutr* 2000;71:1455–61.