REVIEW

ACS chemical biology

Exploring Disease through Metabolomics

Nawaporn Vinayavekhin[†], Edwin A. Homan[†], and Alan Saghatelian*

Harvard University, 12 Oxford St., Cambridge, Massachusetts 02138. [†]These authors contributed equally to this work.

ABSTRACT Metabolomics approaches provide an analysis of changing metabolite levels in biological samples. In the past decade, technical advances have spurred the application of metabolomics in a variety of diverse research areas spanning basic, biomedical, and clinical sciences. In particular, improvements in instrumentation, data analysis software, and the development of metabolite databases have accelerated the measurement and identification of metabolites. Metabolomics approaches have been applied to a number of important problems, which include the discovery of biomarkers as well as mechanistic studies aimed at discovering metabolites or metabolic pathways that regulate cellular and physiological processes. By providing access to a portion of biomolecular space not covered by other profiling approaches (*e.g.*, proteomics and genomics), metabolomics offers unique insights into small molecule regulation and signaling in biology. In the following review, we look at the integration of metabolomics approaches in different areas of basic and biomedical research, and try to point out the areas in which these approaches have enriched our understanding of cellular and physiological biology, especially within the context of pathways linked to disease.

Metabolites serve critical roles in biology due to their involvement in cellular and physiological energetics, structure, and signaling. Fatty acids contribute, for instance, to all three of these central processes through fatty acid metabolism, cellular membrane formation, and nuclear receptor activation, respectively. Similar functions can be delineated for different types of metabolites, such as amino acids, sugars, and many other metabolite classes. Unlike RNA and proteins, metabolites are not directly coded into the genome. Instead, metabolites are products of cellular and physiological metabolism, a product of the biochemistry and biochemical pathways used by a cell or tissue to promote survival. As a result, analysis of the metabolome—the total collection of metabolites in a cell, tissue, or organism—reveals changes in specific metabolites, and also provides a read-out of differences in the underlying metabolic pathways, which can be used to better understand biological mechanisms.

At the turn of the 20th century, Garrod developed his famous "one gene, one enzyme" hypothesis while trying to understand changes in urine metabolite levels associated with the inborn error of metabolism, alkaptonuria. Garrod was able to detect metabolic differences before the advent of modern analytical techniques because alkaptonuria results in dysregulated tyrosine metabolism causing the urine of patients to turn black upon standing (*1, 2*). A number of inborn errors are now known which, if detected early enough, can sometimes be treated. For example, phenylketonuria causes dysregulated phenylalanine metabolism and the symptoms of this disease can be blunted by a diet low in phenylalanine; this motivated the development of a mass spectrometry (MS)-based screening test for newborns that is broadly used (*3*). These examples highlight the importance of being able to measure metabolites levels *in vivo* for understanding and, eventually, treating disease. The need for broad metabolome analysis inspired the development and application of metabolom-

*Corresponding author, saghatelian@chemistry.harvard.edu.

Received for review October 30, 2009 and accepted December 19, 2009.

Published online December 20, 2009

10.1021/cb900271r CCC: $40.75

ics approaches, which aimed to be rapid methods for the quantitation of metabolites directly from cells and tissues.

Modern metabolomics methods attempt to extend the breadth and depth of metabolome analysis to yield insights into biological mechanism and disease (*4–6*). An early example of a metabolomics experiment—albeit many years prior to the invention of the term—can be traced to the work of Linus Pauling in the late 1960s and early 1970s (*7–9*). In this research the stated goal was to use metabolic measurements as a window into underlying differences in human genetics related to disease. It was thought that measurements of changes in metabolite levels associated with any disease would reveal the molecular etiology of the disease, as exemplified by Garrod's work. In these early studies, metabolomics measurements were restricted to breath and urine vapor because they relied on using gas−liquid partition chromatography (GLPC) as the analytical platform (*8*). While GLPC was able to characterize 250 and 280 peaks in a chromatogram from breath and urine, respectively, the metabolome coverage was insufficient to obtain any meaningful data about disease. Thus, one of the powerful driving forces in the field of metabolomics has been the development of new technologies, primarily in mass spectrometry and data analysis, that provide more depth and breadth than has ever been available to the scientific community.

While the biologist has long had access to general tools to study biopolymers (DNA, RNA, protein) within the cell, metabolites have proved counterintuitively to be more challenging class of biomolecules to analyze broadly (*10*). Importantly, the same twenty amino acids make up the vast majority of proteins, and these shared building blocks provide proteins with related physicochemical properties that enable general analytical tools (*e.g.*, SDS-PAGE) to be developed. By contrast, the simplicity of metabolites masks the vast differences in their chemical properties, which contribute to the difficulty of building general tools analogous to the methods used for biopolymers. Consequently, analytical approaches for metabolite analysis have tended to divide the metabolome into pools of molecules with similar properties, such as hydrophobic or hydrophilic molecules, that enable a single analytical approach to be used (*11–13*). In the past few years, improvements in chromatography, mass spectrometry, and small molecule identification have propelled metabolomics forward as a general approach, alongside proteomics and genomics, in the biological sciences (*10, 14, 15*).

Modern Approaches to Metabolomics. The typical workflow for any metabolomics experiment requires three key steps: metabolite isolation, metabolite detection, and data analysis (Figure 1). As mentioned, metabolite isolation is often accomplished while simultaneously fractionating the metabolome into pools of molecules with similar properties. For example, the analysis of lipids commences with the organic extraction of tissues, which acts to isolate these metabolites for analysis and to separate hydrophobic molecules away from hydrophilic metabolites. The metabolites of interest will determine the exact conditions used for the metabolite extraction step, as well as the conditions for the analysis of the metabolites. The two predominant analytical methods that have been employed for metabolite analysis are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy.

NMR provides a rapid, nondestructive method to look broadly at changes in metabolite levels *in vitro* or *in vivo*. For example, NMR metabolomics was used to profile urine from humans to study the secondary metabolism of drugs. The goal was to identify patterns within the urine prior to acetaminophen administration that might be predictive of the manner in which that individual would metabolize the drug (*16*). Urine analysis by NMR revealed that the levels of sulfated acetaminophen were inversely correlated with the levels of sulfated paracresol (*p*-cresol). These results suggested a mechanism whereby *p*-cresol and acetaminophen would compete for sulfation and, as a result, higher *p*-cresol levels would lead to lower levels of sulfated acetaminophen. More importantly, since *p*-cresol is a product of bacterial metabolism of tyrosine in the gut, these experiments reveal the interplay between gut bacteria and secondary drug metabolism, a result that would have only been evident through metabolome analysis. In another interesting example, *T. cylindrosporum*, an insect pathogen, was grown under different culture conditions and analyzed by 2D-NMR. Researchers were able to detect and structurally characterize a new class of fungal metabolites, taking advantage of the ability of NMR to produce detailed structural information (*17*). In addition, strides have been made to couple NMR to liquid chromatography (LC) and the sensitivity of NMR can also be improved through the implementation of capillary NMR methods (*18–22*).

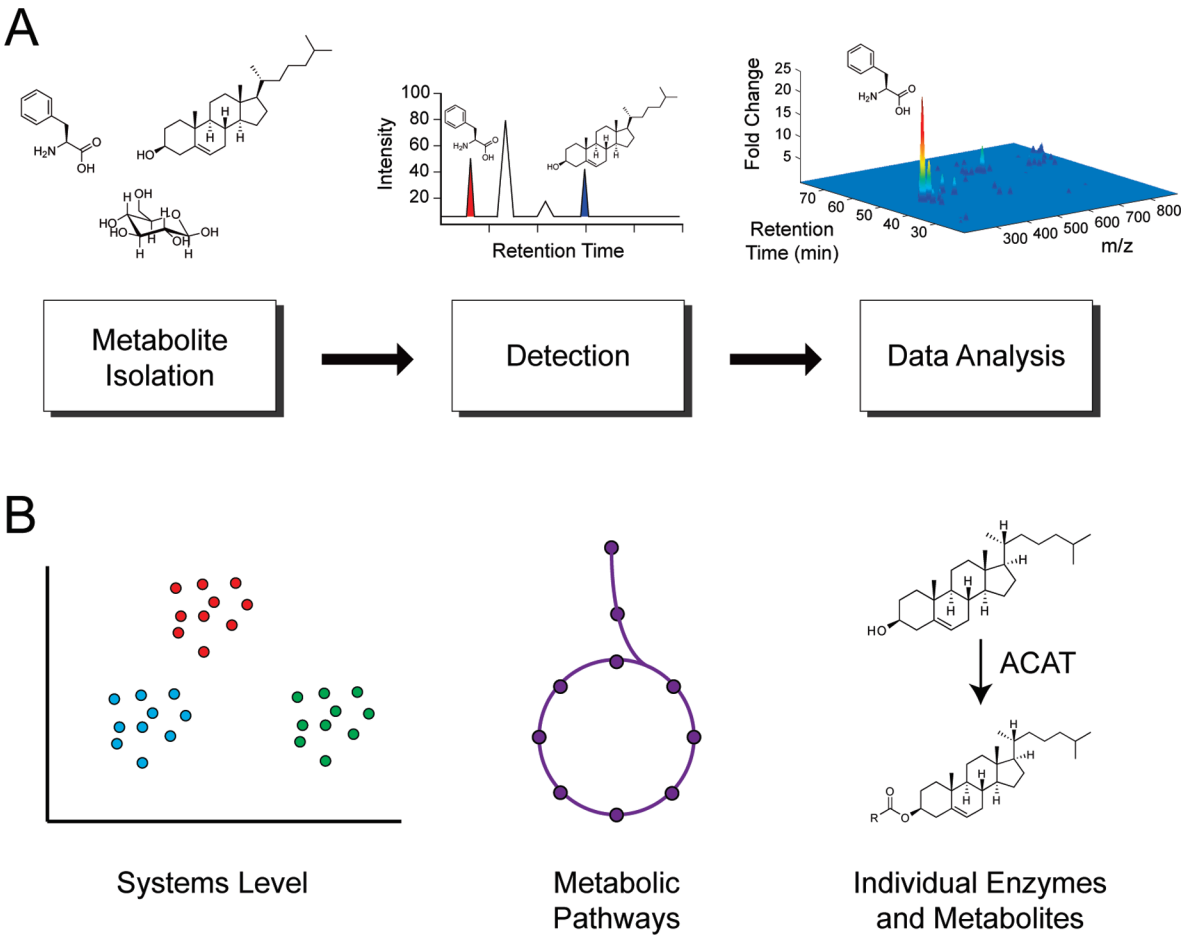

Figure 1. Overview of metabolomics approaches. A) The typical metabolomics workflow has three key steps: metabolite isolation, metabolite detection, and data analysis. B) Depending on the type of data analysis, metabolomics data can be used to phenotype samples (*i.e.*, cluster groups of samples based on similar metabolic profiles), to reveal differences in specific metabolic pathways, and to identify differences in specific metabolites or enzymes between samples.

While the speed and ability of NMR structural characterization is unmatched, the improved sensitivity and resolution of mass spectrometry (MS) permits greater metabolome coverage and has resulted in a higher usage of MS-based metabolomics techniques. In some cases, MS-based metabolomics is performed without chromatography. Instead, samples are directly infused into the mass spectrometer and the resulting spectra can be used to identify individual metabolites or overall patterns associated with a particular sample. For example, an approach developed by Gross and colleagues, termed shotgun lipidomics, uses direct infusion of lipid extracts from biological samples to measure the abundance and acyl chain distribution of a number of important lipid classes, including phospholipids and triglycerides (*23, 24*). To complement these extremely high-throughput direct infusion methods, hyphenated MS-based metabolomics approaches, such as gas chromatography (GC)-MS and LC-MS, have been developed. By resolving complex samples chromatographically, ion suppression by competing metabolite ions can be minimized, leading to improved coverage of the metabolome.

In this vein, targeted GC-MS analysis of wild-type and mutant strains of potato plants identified numerous metabolites and was able to detect and quantify changes in amino acids, carbohydrates, and organic acids (*25*). Similarly, LC-MS methods also permit the de-

tection and quantification of metabolites from biological samples, including larger, less volatile metabolites that would be difficult to analyze by GC-MS. Methods based on LC-MS aim to measure hydrophilic or hydrophobic metabolites, since the physicochemical properties of these pools differ substantially. Indeed, Rabinowitz and others have developed methods for the analysis of hydrophilic metabolites using LC-MS (*26*), while a number of other laboratories, including a large group associated with the LIPID MAPS initiative (*27*), have developed methods to target hydrophobic metabolites.

Broadly speaking, MS based metabolomics experiments are performed in either the targeted or untargeted mode, depending on what type of information is required (Figure 2). Targeted analysis refers to metabolome analysis that focuses on specific metabolites and typically uses internal standards and multiple reaction monitoring (MRM) for absolute quantitation. By contrast, untargeted analysis scans across the entire mass range and measures the amount of a metabolite based on the ion intensity in the mass spectrometer (*i.e.*, the area under the curve of the peak in the LC-MS chromatogram) to enable relative quantitation between two compared sets of samples. To be sure, this quantitation is subject to certain limits since phenomena such as ion suppression can disrupt accurate quantitation. Moreover, adducts of certain metabolites, such as sodium and ammonium salts, can increase the complexity of the data set, but these adducts can typically be filtered relatively easily. Thus, while an untargeted metabolomics experiment should contain much of the information found in a targeted experiment, the use of an internal standard during targeted measurements typically leads to better quantitation. As a general rule, then, one should follow up any untargeted profiling measurement with a targeted quantitative analysis in order to corroborate the observed difference in ion signal intensity. Additionally, targeted approaches increases their sensitivity and enables the detection of ions that are missed by untargeted analyses. Ultimately, the application will dictate which approach is better suited to the final goals of a project.

Another important technical advance has been the development of software tools for the analysis of metabolomics data to enable the quantitation and identification of ions within LC-MS data sets. For example, the program XCMS overcame several hurdles by its ability to align, quantify, and statistically rank ions between sets of LC-MS metabolomics data sets (*28, 29*). Thus, the development of XCMS enabled the speedy detection and relative quantitation of ions between numerous samples. After detection, the next step in a metabolomics experiment requires the identification of metabolites, especially in an untargeted experiment. In terms of metabolite identification, the recent development and improvement of metabolite databases such as METLIN (*30*), LIPID MAPS (*31*), and others have improved metabolite identification from its ion signature. Metabolites are typically identified by matching the ion signature to an entry in the database (Figure 2). In cases where no match can be found, suggesting that the ion may correspond to a novel metabolite, additional strategies for structural elucidation become necessary (Figure 2).

For example, metabolomics analysis of mice lacking the enzyme fatty acid amide hydrolase (FAAH) identified a set of ions regulated by this enzyme in the nervous system and peripheral tissues that did not correspond to any known metabolite. The combination of accurate mass and tandem mass spectrometry led to the hypothesis that these compounds were N-acyl taurines (NATs), whose structures were confirmed by chemical synthesis of a standard (*32*). Improvements in MS technology, especially instruments capable of accurate mass and tandem MS experiments (*33–35*), are making structure elucidation easier every day. Moreover, as novel metabolites are discovered and introduced into the currently available databases, the automated conversion of LC-MS data into lists of metabolites will become easier.

In total, these technical advances have now extended the coverage of metabolomics approaches to enable the use of these methods in basic, biomedical, and clinical research.

KEYWORDS

Ions: Here, referring to charged metabolites (*e.g.*, $[M+H]^+$ and $[M-H]^-$) detected by mass spectrometry.

Ion suppression: A phenomenon in electrospray ionization mass spectrometry (ESI-MS) where metabolites are competing for ionization. As a result, lowering amount of signals is observed for the competing metabolites.

Metabolome: The total collection of metabolites in a cell, tissue, or organism.

Metabolomics: A comparative analysis of metabolome levels under different conditions.

Targeted metabolomics: Quantitative studies of a defined set of known metabolites. In this type of analysis, experiment is generally performed using multiple reaction monitoring (MRM) or internal standards for absolute quantification.

Untargeted metabolomics: Relative quantitation of broad range of metabolites, both knowns and unknowns, in different sample sets. Generally, information about metabolite levels in a sample is reported in terms of fold changes relative to the levels in other samples.

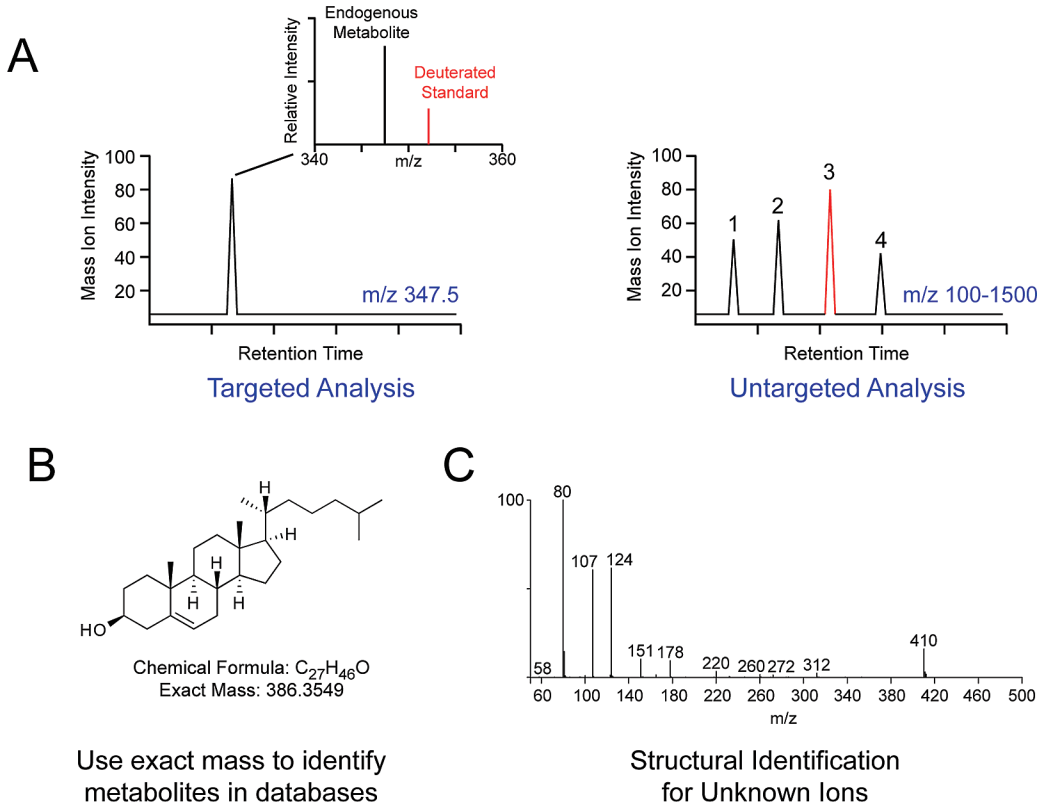

Figure 2. Mass spectrometry based metabolomics approaches and metabolite identification. A) In a targeted metabolomics experiment isotopic standards are added for the absolute quantitation of metabolites, while in an untargeted analysis changes in all metabolites are measured by scanning a broad mass range (*e.g.*, *m*/*z* 100−1500) and relative metabolite levels are quantified by ion intensity. B) Accurate mass can provide molecular formula candidates for compound identification using metabolite databases. C) When the ion is believed to correspond to a novel metabolite, additional analytical experiments, such as mass spectrometry (tandem MS shown) and NMR, must be performed to characterize its structure.

Application of Metabolomics. Metabolomics, like other global profiling approaches (genomics and proteomics), enables researchers to study biology by a hypothesis-generating approach that relies on an initial observation obtained via profiling (*10*). Metabolomics approaches complement standard biochemistry, molecular biology, and physiology to reveal new insights into the molecular pathways underlying biological processes. For example, metabolomics approaches have been applied to the discovery of new biomarkers and, more importantly, to try to gain a better understanding of the molecular etiology of disease (*36, 37*). The data from a metabolomics experiment can be analyzed in a variety of ways depending on the type of information desired (Figure 1). The overall pattern of changes in the metabolome can be interpreted as a molecular phenotype with which to categorize samples (*38*). In addition, changes in specific metabolites can be used as diagnostic markers or to reveal a dysregulated pathway for further study. To date, there have been a number of studies, applied to a variety of organisms, that have used metabolomics data to understand the role of pathways and metabolites in cells, tissues, and organisms. For the purpose of this review, we have focused on recent work studying problems in biomedical research and have divided studies based on these research areas related to a specific disease: metabolic disease, cancer, infectious disease, and regenerative medicine.

Metabolic disease. Metabolic disease is an obvious choice for the application of metabolomics because many of the underlying causes of these diseases are thought to result from dysregulation in small molecule

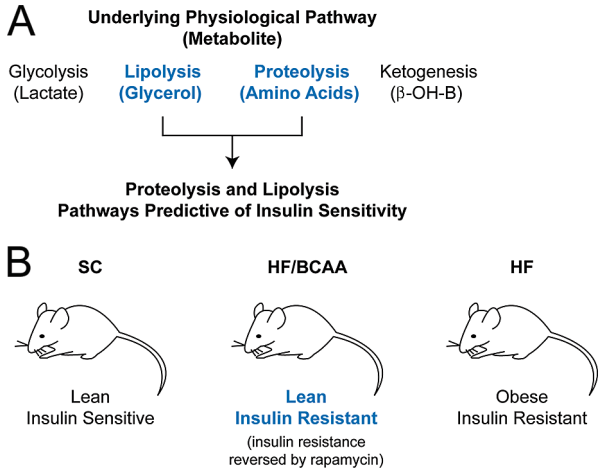

Figure 3. Metabolomics in the study of metabolic disease. A) Metabolomics analysis was used to establish a metabolic signature for glucose response in humans. Initial analysis of plasma from healthy subjects undergoing glucose challenge revealed diagnostic metabolite changes in four insulin-dependent pathways: lactate in glycolysis, glycerol in lipolysis, amino acids in proteolysis, and β-hydroxybutyrate (β-OHB) in ketogenesis. When this insight was investigated in insulin-resistant subjects, Leu/Ile levels and glycerol levels were confirmed to be highly predictive markers of fasting insulin levels and insulin sensitivity. B) Molecular profiling of lean and obese human subjects revealed that differences in branched chain amino acid (BCAA) levels correlated with insulin resistance. The finding was later tested in rats fed standard chow (SC), a high fat diet supplemented with BCAA (HF/BCAA), or a high fat (HF) diet. As expected, animals on SC were lean and insulin-sensitive, while animals on the HF diet were obese and insulin-resistant. Interestingly, rats on the HF/BCAA diet were lean but insulin-resistant. Mechanistic studies suggested that BCAAs were acting through the mammalian target of rapamycin (mTOR).

metabolism. In one elegant example, Shaham and colleagues performed an LC-MS-based metabolomic analysis of plasma from two independent cohorts of subjects undergoing glucose challenge to discover reproducible changes in 18 plasma metabolites, some expected and others unexpected (*39*) (Figure 3). Changes in the metabolome revealed differences in four physiological pathways, including enhanced glucose metabolism, decreased lipolysis, ketogenesis, and proteolysis, which are all insulin-regulated pathways.

This link to insulin action prompted subsequent experiments that measured changes in these 18 metabolites in subjects with impaired glucose tolerance. Univariate and multivariate data analysis revealed that changes in Leu/Ile levels, markers for proteolysis, and glycerol, a marker for lipogenesis, were highly predictive of fasting insulin levels. In doing so, this data suggests insulin-dependent pathways-other than glucose uptake and gluconeogenesis-are also important in insulin resistance, providing a much broader view of the changes associated with insulin resistance *in vivo*. Future studies based on these results will provide greater detail into the mechanisms that connect these metabolic changes to insulin resistance and, in doing so, might reveal additional courses for therapeutic intervention. More generally, this example highlights the value of metabolomics analysis in improving our understanding of the underlying molecular pathways of disease.

Additionally, results garnered from metabolomics experiments can also facilitate subsequent mechanistic studies to identify a causative relationship, if any, between changes in metabolite levels and disease. An example of this overall approach was recently provided by Newgard and colleagues who used a broad molecular profiling approach, which included LC-MS-based metabolomics, to identify differences in branched chain amino acid (BCAA) metabolism that correlated with insulin resistance (*40*) (Figure 3). The molecular profiling was performed using 74 obese and 67 lean human subjects and many of the expected differences were seen, such as elevated leptin levels in obese samples, which confirmed the quality of the samples and robustness of their analytical methods. Multivariate analysis of the data, using principal component analysis (PCA), identified molecules that could best distinguish between lean and obese subjects. PCA analysis revealed that BCAAs, aromatic amino acids, and short chain acyl carnitines (C3 and C5) provided metabolic signatures that could distinguish lean and obese samples.

To test whether any of these metabolites directly contribute to insulin resistance, experiments were later designed to study the effect of BCAAs, introduced through diet, on insulin resistance in rats. Rats were placed on three different diets in these studies: standard chow (SC), high fat (HF), and high fat-BCAA (HF/BCAA). In terms of weight gain, the HF/BCAA rats were more similar in weight to the SC rats than they were to the HF group, which was attributed to lower caloric intake for the HF/BCAA group. However, the HF/BCAA group was as insulin resistant as the HF group. Moreover, when caloric intake between HF and HF/BCAA groups was normalized (*i.e.*, less food for the HF rats), the HF group was no longer insulin resistant, which suggests that the

BCAAs in the diet were contributing to the insulin resistance phenotype. The mechanistic hypothesis developed from these observations was that changes in BCAA levels were contributing to differences in insulin sensitivity through mTOR signaling, a known regulator of insulin resistance (*41*). Indeed, treatment of rats with the mTOR inhibitor, rapamycin, was able to improve glucose tolerance in HF/BCAA rats but had no effect in the HF group. In total, this study demonstrates the contribution of BCAAs, especially on a HF diet background, to the development of insulin resistance. Furthermore, this example highlights the value of metabolomics approaches in developing and testing new mechanistic models of disease through the discovery of signaling molecules and pathways-in this case BCAA signaling through mTOR.

Cancer. One of the advantages of metabolomics is its ability to analyze metabolites from any source, regardless of origin. Ramanathan and colleagues applied a targeted LC-MS-based metabolomics analysis to a series of engineered cancer cell lines that were progressively more tumorigenic due to the introduction of 1, 2, 3, or 4 oncogenes (telomerase, small and large T antigen, and oncogenic H-ras) (*42*) (Figure 4). In addition, these studies measured metabolic states in the presence of small molecules, selected to inhibit specific metabolic and nutrient-sensing pathways, to create a multidimensional matrix that could interpret the changes in cellular metabolism as a function of gene expression or small molecule levels. Changes in glucose uptake and oxygen consumption in these cells could be assessed by changes in metabolite levels and by treatment with small molecule pathway inhibitors, revealing the dependence of a cell line on a particular metabolic pathway. In total, the work of Ramanathan correlated the emergence of aerobic glycolysis, the Warburg effect, with the expression of known oncogenes and the emergence of the tumorigenic state in these cells (*43*). In doing so, they have established that changes in cancer cell metabolism are downstream of these oncogenes and, as a result, might be potential targets for the development of novel therapeutics.

Evidence that aerobic glycolysis is a potential therapeutic target in cancer has been demonstrated in a recent study which revealed that the M2 isoform of pyruvate kinase is responsible for the switch to aerobic glycolysis (*44*). Indeed, inhibition of the M2 isoform reduces aerobic glycolysis and inhibits tumorigenic capability of cancer cells, as evidenced by reduced tumor formation during xenograft studies. These results establish the importance of aerobic glycolysis in sustaining tumor growth and demonstrate that the M2 isoform of the pyruvate kinase is a potential target for the development of novel chemotherapeutic agents. More generally, metabolic pathways are gaining increased attention for their role in cancer as recently demonstrated by the description of the *R*-(2)-hydroxyglutarate as an oncometabolite (*45*).

Lipid profiling can also reveal pathways capable of regulating cancer cell phenotypes. Chiang and coworkers utilized an untargeted LC-MS-based metabolomics platform to measure changes in the lipidome of cells treated with an inhibitor against the human lipase KIAA1363 (*46*) (Figure 4). This enzyme was originally identified as an invasiveness-related serine hydrolase due to its marked upregulation, measured by activity-based proteomics, in a series of invasive cancer cell lines derived from multiple tissues. Metabolomic analysis of KIAA1363-inhibited cells revealed that this enzyme regulates cellular levels of monoalkylglycerol ether (MAGE) levels. Prior work had identified MAGE lipid levels as being strongly elevated in neoplastic tumors (*47*), but this was the first report of an enzyme that regulated MAGE levels in cancer cells. Connecting these lipids to an enzymatic activity-especially one that could be inhibited-enabled testing of the hypothesis that this pathway might be involved in a signaling pathway that caused invasiveness.

Treatment of SKOV-3 cells, an invasive ovarian cancer cell line, with virus containing an shRNA against KIAA1363 resulted in a cell line with appreciably lower levels of KIAA1363, and correspondingly lower levels of MAGEs, which indicated that KIAA1363 is the rate-determining step in the production of MAGE in these cell lines. Subsequent experiments revealed that cell lines with lower KIAA1363 activity were less invasive *in vitro* and, more importantly, showed attenuated growth *in vivo* during a xenograft experiment in mice. These experiments established the importance of the KIAA1363-mediated metabolic pathway in cancer cell phenotypes, including growth *in vivo*. This study highlights the value of untargeted metabolomics to discover biologically important metabolic pathways, and identify natural small molecules, MAGEs, that can regulate cellular phenotypes.

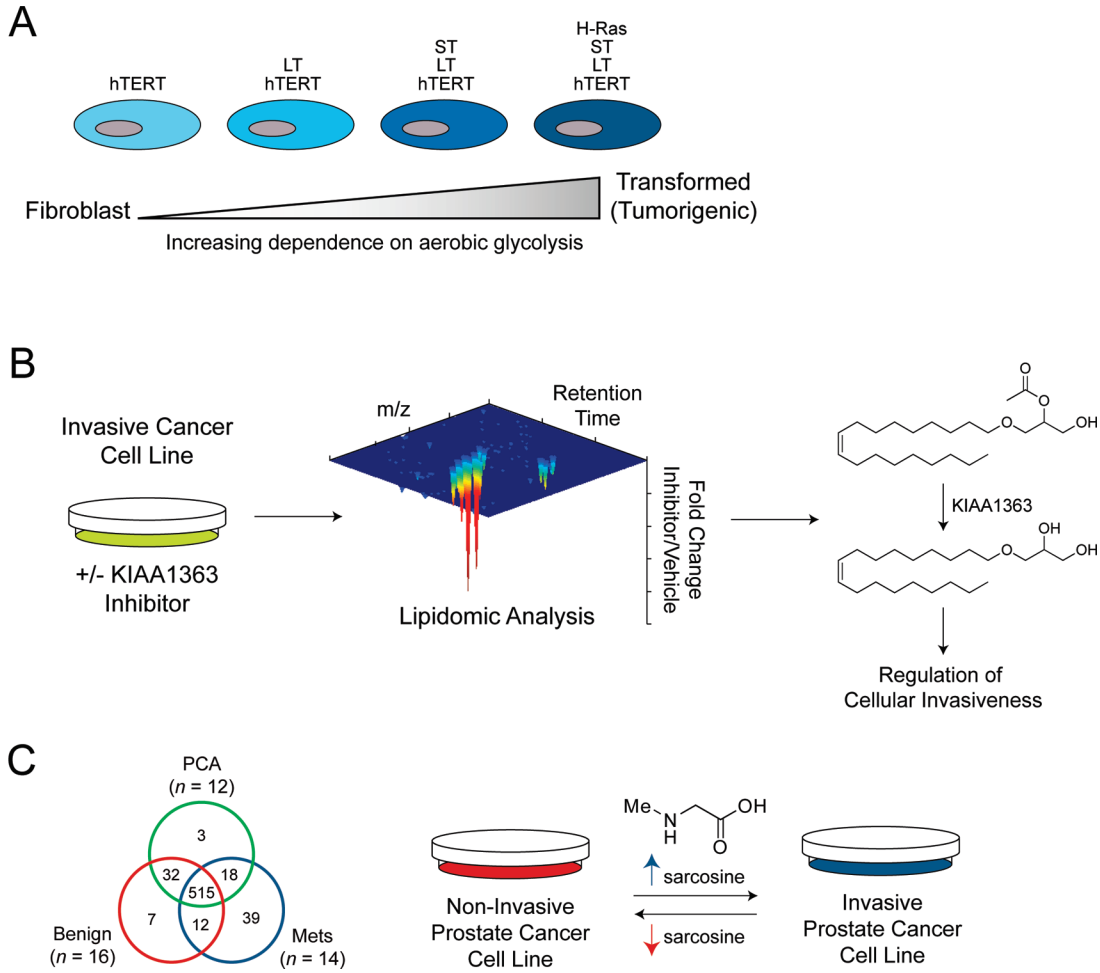

Figure 4. Application of metabolomics approaches to cancer research. A) The addition of oncogenes, such as telomerase (hTERT), large T antigen (LT), small T antigen (ST), and H-Ras transform a fibroblast cell into a tumorigenic state. Ramanathan and colleagues analyzed these engineered cells using an LC-MS based metabolomics approach to reveal that cells start to depend more on aerobic glycolysis as they become more tumorigenic. B) Chiang and co-workers utilized an untargeted LC-MS-based metabolomics approach to establish a link between the monoalkylglycerol ether (MAGE) pathway and KIAA1363, an uncharacterized invasiveness-related hydrolase. KIAA1363 regulates the production of MAGEs, and inhibition of this enzyme attenuates invasiveness *in vitro* and slows tumor growth *in vivo*. C) Metabolomics analysis of clinical samples from patients with prostate tumors by Sreekumar and colleagues revealed numerous metabolites associated with benign, prostate cancer, and metastatic prostate cancer. Further investigation showed that addition of sarcosine to benign prostate cell lines could promote invasiveness and inhibition of sarcosine production in invasive cell lines would reduce invasiveness.

In our final example of metabolomics applied to cancer, metabolomic analysis of clinical prostate cancer samples by Sreekumar and colleagues identified sarcosine as a biomarker and potential contributor to prostate cancer progression (*48*) (Figure 4). This study began by looking at metabolites in 262 clinical samples from patients with prostate cancer. The overall metabolite profile provided a fingerprint that could distinguish between benign prostate, cancerous tissue, and prostate cancer that had already metastasized. More specifically, metabolomic analysis identified six metabolites, such as sarcosine, uracil, and proline, that increased in the progression from benign tissue to a metastatic tumor. Sarcosine was selected for further study because

its structure, *N*-methylglycine, integrated information from amino acid metabolism and single carbon metabolism (methylation), which were thought to be important in cancer cell metabolism. Detailed quantitation confirmed the original measurements and showed a dramatic increase in sarcosine levels when comparing benign to cancerous and metastatic tissues. Moreover, sarcosine levels from urine performed better than prostate specific antigen (PSA) in predicting cancer in the examples used in this paper.

Changes in sarcosine metabolism were also apparent in cancer cell lines, as elevated levels of sarcosine were measured in invasive prostate cancer lines compared to noninvasive counterparts. These findings led to the hypothesis that sarcosine levels might contribute to cellular invasiveness. Indeed, addition of sarcosine to noninvasive benign prostate epithelial cells led to a dose-dependent increase in the invasiveness of the cells. Similarly, alterations in the biosynthesis or catabolism of sarcosine also caused differences in invasiveness, with higher sarcosine levels leading to increased invasiveness. Lastly, analysis of gene transcription changes upon treatment with an androgen, a known factor in prostate cancer progression, revealed changes in the expression of enzymes in the sarcosine pathway that could predict increased invasiveness. This androgen-dependent regulation of the sarcosine-regulating enzymes suggests that the sarcosine pathway might be part of the natural progression of prostate cancer *in vivo* and could potentially be a target for the development of novel therapeutics. In total, these representative examples highlight the utility of metabolomics in understanding cellular energetics and metabolic signaling in the regulation of cancer.

Infectious Disease. Metabolomics has also made an impact in infectious disease. For example, metabolomics has been used to study global metabolic responses of mice to parasites (*Trypanosoma brucei* brucei) (*49*) and viruses, such as simian immunodeficiency virus (SIV) (*50*), cytolomegalovirus (CMV) (*51*), and hepatitis B virus (HBV) (*52*). In one example, changes in metabolic flux between uninfected cells and cells infected with human cytolomegalovirus (HCMV) was used by Munger and colleagues to assess pathways that are upregulated by the viral infection (*51*) (Figure 5). Rather than measuring the steady state levels of metabolites, a metabolic labeling strategy was used to measure flux through metabolic pathways using a targeted LC-MS/MS metabolomics platform (*53*). This approach provides more information than steady state measurements because it can reveal whether the changes in the level of a given metabolite are a function of differences in the biosynthesis or the catabolism of a metabolite. Since these experiments were carried out in cell culture, the authors were able to utilize a number of different sources of labeled carbon, including ^{13}C-labeled glucose and ^{13}C-labeled glutamine. There were numerous changes in flux during HCMV infection, including large increases into the citric acid (TCA) cycle. One interesting finding was a large increase in flux between citrate and malonyl-CoA, which is used in fatty acid synthesis, suggesting that free fatty acid biosynthesis is upregulated during HCMV infection. Experiments using radiolabeled glucose confirmed the upregulation of fatty acid biosynthesis upon HCMV infection, which represented a novel biological finding and warranted further study into the role of this pathway in HCMV infection and replication.

Since HCMV is an enveloped virus the authors suspected that changes in fatty acid metabolism might be directly involved in viral replication. To test this hypothesis chemical inhibition of the *de novo* fatty acid synthesis pathway was used to assess the role of fatty acid biosynthesis in viral replication. Two enzymes in the fatty acid biosynthetic pathway acetyl-CoA carboxylase (ACC) and fatty acid synthase (FAS) were targeted using known inhibitors 5-tetradecyloxy-2-furoic acid (TOFA) and trans-4-carboxy-5-octyl-3-methylene-butyrolactone (C75), respectively. Importantly, the addition of either of these inhibitors resulted in a dose-dependent decrease in the viral titers of HCMV-infected cells, indicating that fatty acid biosynthesis is critical for the efficient replication of HCMV. More generally, the importance of *de novo* fatty acid synthesis in enveloped virus production was extended to another virus, influenza A. Influenza A production was also limited by the addition of TOFA or C75, suggesting that the need for *de novo* fatty acid synthesis is conserved among these two enveloped viruses; this finding points to the possible use of inhibitors of fatty acid synthesis as treatments for enveloped virus-mediated infections in general. This work highlights the ability of metabolomics approaches to identify new pathways linked to the regulation of an infectious disease (*e.g.*, viral replication), leading to the discovery of new targets for therapeutic intervention.

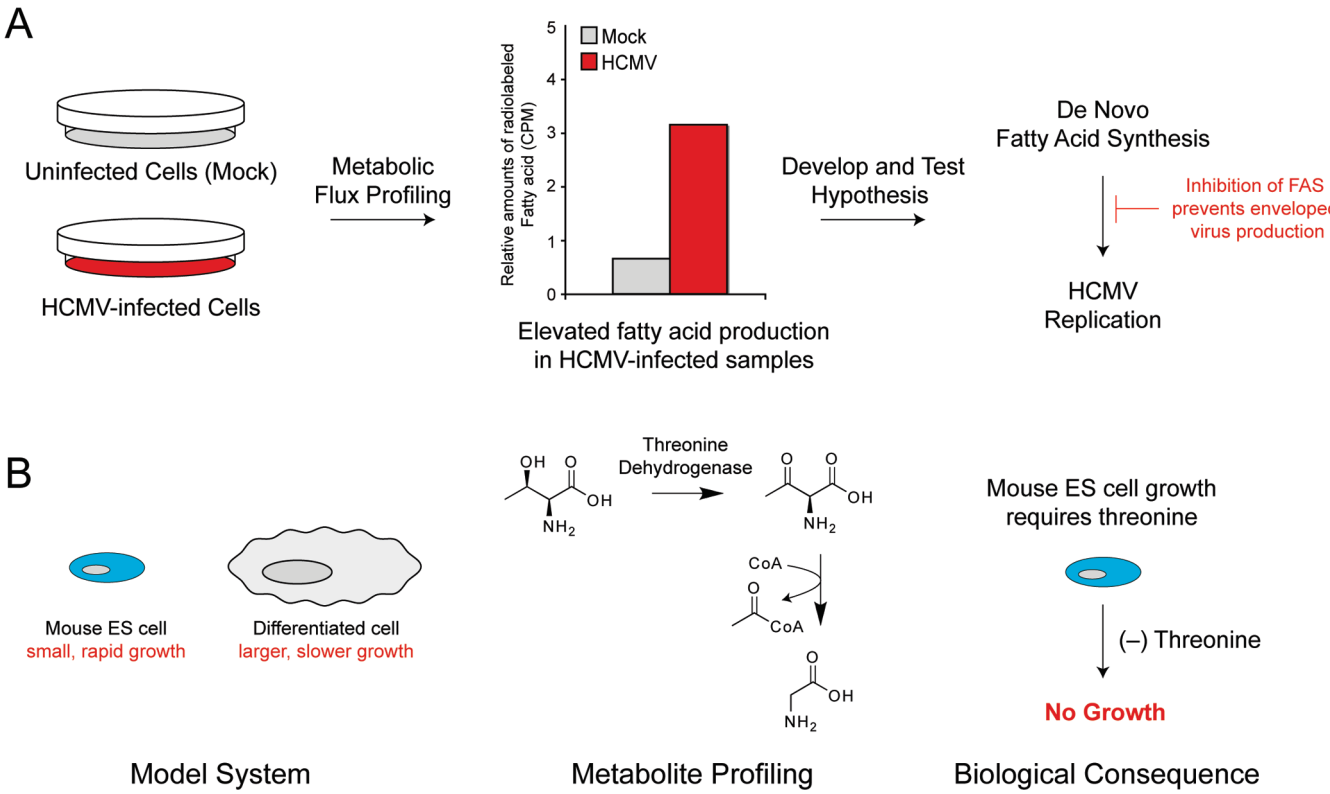

Figure 5. Metabolomics in infectious disease and regenerative medicine. A) Metabolic flux analysis of cells infected with human cytolomegalovirus (HCMV) by Munger and co-workers revealed that virus-infected cells produce more fatty acids. Subsequent experiments targeted fatty acid synthase and demonstrated that *de novo* fatty acid synthesis is necessary for efficient HCMV replication. B) Wang and colleagues reasoned that the smaller size and rapid growth of mouse ES cells, in comparison to differentiated cells, was likely due to a large difference in cellular metabolism. Metabolomic analysis of mouse ES cells revealed the importance of threonine metabolism in these cells through dramatic upregulation of the enzyme threonine dehydrogenase. While they are able to grow in the absence of all other amino acids, mouse ES cells are uniquely dependent on threonine metabolism and do not grow in the absence of threonine.

Regenerative Medicine. Stem cells have the ability to undergo self-renewal and the capacity to differentiate into a variety of cell types. As interest in stem cells has grown, a number of studies have looked at a variety of factors involved in stem cell biology, including genes and proteins that regulate the production, replication, and differentiation of stem cells. However, it was not until very recently that the metabolism of stem cells was examined by Wang and co-workers, who focused their studies on mouse embryonic stem (ES) cells (*54*) (Figure 5). Inspired by the small size and rapid cell division associated with mouse ES cells, the authors reasoned that the metabolism of stem cells was likely altered in comparison to differentiated cells. This idea was tested by analyzing mouse ES cells using a targeted LC-MS-based metabolomics platform to identify metabolites specifically associated with the stem cell state. In these experiments, mouse ES cells cultured under standard conditions were compared to mouse ES cells cultured under conditions lacking leukemia inhibiting factor (LIF), which leads to changes in growth rate and cell morphology of the stem cells, before ultimately inducing the cells to differentiate. One of the observations in these experiments was an elevation in a number of metabolites involved in purine biosynthesis in the ES cells grown under standard conditions, which was not surprising due to the rapid growth rate of the mouse ES cells that would require additional DNA production. Since purine metabolism requires sources of one-carbon metabolism these results also suggested that ES cells require a large flux through one-carbon metabolism.

Subsequent gene expression analysis for proteins known to be involved in one-carbon metabolism revealed that the enzyme threonine dehydrogenase (TDH) was elevated 1000-fold in mouse ES cells compared to seven adult tissues. TDH oxidizes threonine to afford L-2-Amino-3-oxobutanoic acid and this metabolite is

then further catabolized to afford one molecule of acetyl-CoA and one molecule of glycine (Figure 5). The importance of the TDH pathway in ES cell growth was also supported by experiments where ES cells were grown in the absence of a single amino acid, termed "drop-out" experiments by the authors. Remarkably, ES cells were able to grow in the absence of all amino acids except threonine. This effect was specific to ES cells as other cell lines, such as Hela cells, were not sensitive to threonine "drop-out". Moreover, addition of excess glycine to the media was not able to rescue the diminished growth, nor was growth rescued by the presence of a threonine derivative that was metabolized by TDH to provide glycine but not acetyl-CoA, indicating that both glycine and acetyl-CoA produced by TDH are necessary for ES cell growth. Lastly, these results extended to mouse embryos, which were also shown to be reliant on threonine for growth.

Interestingly, while mouse ES cell and embryos require threonine to grow, humans possess an inactive variant of the TDH enzyme. In contrast to the mouse ES cells, human ES cells grow at a much slower rate (35 h vs 5 h doubling time), leading the authors to speculate that the introduction of an active form of TDH into these human ES cells might accelerate human ES cell growth and improve their ease of use. Thus, the discovery of the central role that TDH plays in mouse ES cell metabolism has answered important questions about the metabolism of mouse ES cells, but has also raised many important lines of inquiry regarding the inner workings of human ES cells. Again, these important biological insights were predicated on observations and hypotheses derived from metabolomics experiments.

Future Directions. These examples demonstrate the value of metabolomics as a global profiling tool in the discovery of metabolites and metabolic pathways central to biological regulation, including the regulation of disease. As mentioned, technological advances have brought metabolomics to the point where these techniques can find general application in the biological and biomedical sciences. Looking ahead, critical hurdles must be overcome to drive the next phase of metabolomics development and applications. In particular, improvements in automated metabolite identification from LC-MS data sets will expand the number of potential users of metabolomics approaches. Once these automated approaches are improved it will also become easier to integrate metabolomics data into systems biology workflows, facilitating new experiments that integrate metabolomics with other global profiling experiments. Moreover, a systems-level treatment of metabolomics data sets with pathway analysis software will begin to reveal the interplay between genes, proteins, and metabolites in cellular regulation and signaling.

Lastly, it is evident that metabolomics methods reside squarely at the interface of chemistry and biology. The discovery of new metabolites will require the chemical synthesis of these compounds for structural confirmation and continued biological study. The development and application of metabolomics methods may still be in its adolescence; however, the examples above demonstrate that the integration of metabolomics approaches into basic and biomedical research is already improving our understanding of biological regulation and signaling at the cellular and organismal level, with important applications in the study of disease.

REFERENCES

1. Garrod, A. E. (1899) A contribution to the study of alkaptonuria, *Med. Chir. Trans. 82*, 367–394.
2. Garrod, A. E. (1923) *Inborn Errors of Metabolism*, 2nd ed., Oxford University Press, London.
3. Chace, D. H., Sherwin, J. E., Hillman, S. L., Lorey, F., and Cunningham, G. C. (1998) Use of phenylalanine-to-tyrosine ratio determined by tandem mass spectrometry to improve newborn screening for phenylketonuria of early discharge specimens collected in the first 24 h, *Clin. Chem. 44*, 2405–2409.
4. Watkins, S. M., and German, J. B. (2002) Metabolomics and biochemical profiling in drug discovery and development, *Curr. Opin. Mol. Ther. 4*, 224–228.
5. Kaddurah-Daouk, R., Kristal, B. S., and Weinshilboum, R. M. (2008) Metabolomics: a global biochemical approach to drug response and disease, *Ann. Rev. Pharmacol. Toxicol. 48*, 653–683.
6. Nicholson, J. K., Connelly, J., Lindon, J. C., and Holmes, E. (2002) Metabonomics: a platform for studying drug toxicity and gene function, *Nat. Rev. Drug Discovery 1*, 153–161.
7. Pauling, L. (1968) Orthomolecular psychiatry, *Science 160*, 265.
8. Pauling, L., Robinson, A. B., Teranish, R., and Cary, P. (1971) Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography, *Proc. Natl. Acad. Sci. U.S.A. 68*, 2374–2376.
9. Robinson, A. B., and Pauling, L. (1974) Techniques of orthomolecular diagnosis, *Clin. Chem. 20*, 961–965.
10. Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., and Kell, D. B. (2004) Metabolomics by numbers: acquiring and understanding global metabolite data, *Trends Biotechnol. 22*, 245–252.
11. Kind, T., Tolstikov, V., Fiehn, O., and Weiss, R. H. (2007) A comprehensive urinary metabolomic approach for identifying kidney cancer, *Anal. Biochem. 363*, 185–195.

12. Watkins, S. M., Reifsnyder, P. R., Pan, H. J., German, J. B., and Leiter, E. H. (2002) Lipid metabolome-wide effects of the PPARγ agonist rosiglitazone, *J. Lipid Res. 43*, 1809–1817.
13. Tolstikov, V. V., and Fiehn, O. (2002) Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry, *Anal. Biochem. 301*, 298–307.
14. Rochfort, S. (2005) Metabolomics reviewed: A new "Omics" platform technology for systems biology and implications for natural products research, *J. Nat. Prod. 68*, 1813–1820.
15. Dettmer, K., Aronov, P. A., and Hammock, B. D. (2007) Mass spectrometry-based metabolomics, *Mass Spectrom. Rev. 26*, 51–78.
16. Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R., and Nicholson, J. K. (2009) Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism, *Proc. Natl. Acad. Sci. U.S.A. 106*, 14728–14733.
17. Schroeder, F. C., Gibson, D. M., Churchill, A. C. L., Sojikul, P., Wursthorn, E. J., Krasnoff, S. B., and Clardy, J. (2007) Differential analysis of 2D NMR spectra: New natural products from a pilot-scale fungal extract library, *Angew. Chem., Int. Ed. 46*, 901–904.
18. Watanabe, N., and Niki, E. (1978) Direct-coupling of FT-NMR to high-performance liquid-chromatography, *Proc. Jpn. Acad. Ser. B 54*, 194–199.
19. Exarchou, V., Krucker, M., van Beek, T. A., Vervoort, J., Gerothanassis, I. P., and Albert, K. (2005) LC-NMR coupling technology: recent advancements and applications in natural products analysis, *Magn. Reson. Chem. 43*, 681–687.
20. Bayer, E., Albert, K., Nieder, M., and Grom, E. (1979) Online coupling of high-performance liquid-chromatography and nuclear magnetic-resonance, *J. Chromatogr. 186*, 497–507.
21. Schroeder, F. C., and Gronquist, M. (2006) Extending the scope of NMR spectroscopy with microcoil probes, *Angew. Chem., Int. Ed. 45*, 7122–7131.
22. Olson, D. L., Peck, T. L., Webb, A. G., Magin, R. L., and Sweedler, J. V. (1995) High-resolution microcoil H-1-NMR for mass-limited, nanoliter-volume samples, *Science 270*, 1967–1970.
23. Han, X. L., Yang, J. Y., Cheng, H., Ye, H. P., and Gross, R. W. (2004) Toward fingerprinting cellular lipidomes directly from biological samples by two-dimensional electrospray ionization mass spectrometry, *Anal. Biochem. 330*, 317–331.
24. Han, X. L., and Gross, R. W. (2005) Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples, *Mass Spectrom. Rev. 24*, 367–412.
25. Weckwerth, W., Loureiro, M. E., Wenzel, K., and Fiehn, O. (2004) Differential metabolic networks unravel the effects of silent plant phenotypes, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7809–7814.
26. Bajad, S. U., Lu, W. Y., Kimball, E. H., Yuan, J., Peterson, C., and Rabinowitz, J. D. (2006) Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry, *J. Chromatogr. A 1125*, 76–88.
27. Schmelzer, K., Fahy, E., Subramaniam, S., Dennis, E. A. (2007) The lipid maps initiative in lipidomics, in *Lipidomics and Bioactive Lipids: Mass-Spectrometry-Based Lipid Analysis*, pp 171–183, Elsevier Academic Press Inc., San Diego.
28. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal. Chem. 78*, 779–787.
29. Benton, H. P., Wong, D. M., Trauger, S. A., and Siuzdak, G. (2008) XCMS2: Processing tandem mass spectrometry data for metabolite identification and structural characterization, *Anal. Chem. 80*, 6382–6389.
30. Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., Siuzdak, G. (2005) *METLIN–A Metabolite Mass Spectral Database*, pp 747–751, Lippincott Williams & Wilkins, Baltimore.
31. Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., Merrill, A. H., Murphy, R. C., Raetz, C. R. H., Russell, D. W., and Subramaniam, S. (2007) LMSD: lipid maps structure database, *Nucleic Acids Res. 35*, D527–D532.
32. Saghatelian, A., Trauger, S. A., Want, E. J., Hawkins, E. G., Siuzdak, G., and Cravatt, B. F. (2004) Assignment of endogenous substrates to enzymes by global metabolite profiling, *Biochemistry 43*, 14332–14339.
33. Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis, *Anal. Chem. 72*, 1156–1162.
34. Shevchenko, A., Chernushevich, I., Ens, W., Standing, K. G., Thomson, B., Wilm, M., and Mann, M. (1997) Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer, *Rapid Commun. Mass Spectrom. 11*, 1015–1024.
35. Morris, H. R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R. S., Hoyes, J., and Bateman, R. H. (1996) High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer, *Rapid Commun. Mass Spectrom. 10*, 889–896.
36. Schlotterbeck, G., Ross, A., Dieterle, F., and Senn, H. (2006) Metabolic profiling technologies for biomarker discovery in biomedicine and drug development, *Pharmacogenomics 7*, 1055–1075.
37. Quinones, M. P., and Kaddurah-Daouk, R. (2009) Metabolomics tools for identifying biomarkers for neuropsychiatric diseases, *Neurobiol. Dis. 35*, 165–176.
38. Clayton, T. A., Lindon, J. C., Cloarec, O., Antti, H., Charuel, C., Hanton, G., Provost, J. P., Le Net, J. L., Baker, D., Walley, R. J., Everett, J. R., and Nicholson, J. K. (2006) Pharmaco-metabonomic phenotyping and personalized drug treatment, *Nature 440*, 1073–1077.
39. Shaham, O., Wei, R., Wang, T. J., Ricciardi, C., Lewis, G. D., Vasan, R. S., Carr, S. A., Thadhani, R., Gerszten, R. E., and Mootha, V. K. (2008) Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity, *Mol. Syst. Biol. 4*, 214.
40. Newgard, C. B., An, J., Bain, J. R., Muehlbauer, M. J., Stevens, R. D., Lien, L. F., Haqq, A. M., Shah, S. H., Arlotto, M., Slentz, C. A., Rochon, J., Gallup, D., Ilkayeva, O., Wenner, B. R., Yancy, W. S., Jr., Eisenson, H., Musante, G., Surwit, R. S., Millington, D. S., Butler, M. D., and Svetkey, L. P. (2009) A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance, *Cell Metab. 9*, 311–326.
41. Um, S. H., Frigerio, F., Watanabe, M., Picard, F., Joaquin, M., Sticker, M., Fumagalli, S., Allegrini, P. R., Kozma, S. C., Auwerx, J., and Thomas, G. (2004) Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity, *Nature 431*, 200–205.
42. Ramanathan, A., Wang, C., and Schreiber, S. L. (2005) Perturbational profiling of a cell-line model of tumorigenesis by using metabolic measurements, *Proc. Natl. Acad. Sci. U.S.A. 102*, 5992–5997.
43. Warburg, O., Posener, K., and Negelein, E. (1924) Über den Stoffwechsel der Carcinomzelle, *Biochem. Z. 152*, 309–344.
44. Christofk, H. R., Vander Heiden, M. G., Harris, M. H., Ramanathan, A., Gerszten, R. E., Wei, R., Fleming, M. D., Schreiber, S. L., and Cantley, L. C. (2008) The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth, *Nature 452*, 230–U274.

45. Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liau, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate, *Nature 462*, 739.
46. Chiang, K. P., Niessen, S., Saghatelian, A., and Cravatt, B. F. (2006) An enzyme that regulates ether lipid signaling pathways in cancer annotated by multidimensional profiling, *Chem. Biol. 13*, 1041–1050.
47. Snyder, F., and Wood, R. (1969) Alkyl and alk-1-enyl ethers of glycerol in lipids from normal and neoplastic human tissues, *Cancer Res. 29*, 251–257.
48. Sreekumar, A., Poisson, L. M., Rajendiran, T. M., Khan, A. P., Cao, Q., Yu, J. D., Laxman, B., Mehra, R., Lonigro, R. J., Li, Y., Nyati, M. K., Ahsan, A., Kalyana-Sundaram, S., Han, B., Cao, X. H., Byun, J., Omenn, G. S., Ghosh, D., Pennathur, S., Alexander, D. C., Berger, A., Shuster, J. R., Wei, J. T., Varambally, S., Beecher, C., and Chinnaiyan, A. M. (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression, *Nature 457*, 910–914.
49. Wang, Y., Utzinger, J., Saric, J., Li, J. V., Burckhardt, J., Dirnhofer, S., Nicholson, J. K., Singer, B. H., Brun, R., and Holmes, E. (2008) Global metabolic responses of mice to *Trypanosoma brucei brucei* infection, *Proc. Natl. Acad. Sci. U.S.A. 105*, 6127–6132.
50. Wikoff, W. R., Pendyala, G., Siuzdak, G., and Fox, H. S. (2008) Metabolomic analysis of the cerebrospinal fluid reveals changes in phospholipase expression in the CNS of SIV-infected macaques, *J. Clin. Invest. 118*, 2661–2669.
51. Munger, J., Bennett, B. D., Parikh, A., Feng, X. J., McArdle, J., Rabitz, H. A., Shenk, T., and Rabinowitz, J. D. (2008) Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy, *Nat. Biotechnol. 26*, 1179–1186.
52. Rodgers, M. A., Saghatelian, A., and Yang, P. L. (2009) Identification of an overabundant cholesterol precursor in hepatitis B virus replicating cells by untargeted lipid metabolite profiling, *J. Am. Chem. Soc. 131*, 5030–5031.
53. Yuan, J., Fowler, W. U., Kimball, E., Lu, W., and Rabinowitz, J. D. (2006) Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*, *Nat. Chem. Biol. 2*, 529–530.
54. Wang, J., Alexander, P., Wu, L. J., Hammer, R., Cleaver, O., and McKnight, S. L. (2009) Dependence of mouse embryonic stem cells on threonine catabolism, *Science 325*, 435–439.