Published in final edited form as:

Mol Cell. 2015 May 21; 58(4): 699–706. doi:10.1016/j.molcel.2015.04.021.

Defining the Metabolome: Size, Flux, and Regulation

Nicola Zamboni[1], **Alan Saghatelian**[2], and **Gary J Patti**[3]

[1]Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland [2]Clayton Foundation Laboratories for Peptide Biology, Salk Institute for Biological Studies, La Jolla, California 92037, United States [3]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA Department of Chemistry, Washington University, St. Louis, MO 63130, USA

Renewed interest in metabolic research over the last two decades has inspired an explosion of technological developments for studying metabolism. At the forefront of methodological innovation is an approach referred to as “untargeted” or “discovery” metabolomics. The experimental objective of this technique is to comprehensively measure the entire metabolome, which constitutes a largely undefined set of molecules. Given its potential comprehensive coverage, untargeted metabolomics is often the first choice of experiments for investigators pursuing a metabolic research question. It is important to recognize, however, that untargeted metabolomics may not always be the optimal experimental approach. Conventionally, untargeted metabolomics only provides information about relative differences in metabolite pool sizes. Therefore, depending on the specific scientific question at hand, a complementary approach involving stable isotopes (such as metabolic flux analysis) may be better suited to provide biological insights. Unlike untargeted metabolomics, stable-isotope methods can provide information about differences in reaction rates.

So which metabolic research questions are best tailored for each of the various experimental approaches? What are the limitations of untargeted metabolomics? What are the challenges of metabolic flux analysis? How have each been successfully applied? Research investigators Gary Patti, Nicola Zamboni, and Alan Saghatelian address these questions in the vignettes that follow. First, Gary Patti discusses opportunities and challenges of untargeted metabolomics within the framework of potential unknown metabolites in “How Big is the Metabolome? Opportunities and Challenges”. Next, Nicola Zamboni describes the power of metabolic flux analysis and explores its most frequent limitations and pitfalls in “Modern Stable-Isotope Metabolic Flux Analysis”. Finally, Alan Saghatelian contrasts different instrumentation platforms available for metabolite measurements and highlights

Correspondence to: Nicola Zamboni; Gary J Patti.

some successful applications of both metabolomics and metabolic flux analysis in "Biological Lessons from Metabolite Profiling".

How Big Is the Metabolome? Opportunities and Challenges

Gary J. Patti

In 1955, Donald Nicholson compiled all of the metabolic reactions known at that time into a single chart. This chart, which he drew by hand with stencils, provided the first perspective of a comprehensive cellular metabolome. The rendering displayed only about 20 metabolic pathways.(Nicholson, 1970)

The majority of metabolic pathways taught in today's undergraduate biochemistry curriculum had been discovered and mapped onto comprehensive charts by the 1960's. There was a growing perception that the picture of the cellular metabolome was complete. In 1964, Nobel laureate Ernst Boris Chain offered his perspective on the greatest landmarks in biochemical research. He listed one major achievement as the elucidation of biochemical pathways.(Chain, 1965) Chain categorized the success of elucidating pathways into three historical eras: (i) the "pre-isotope era" in which the enzyme activities of pathways were determined in cell-free extracts, (ii) the "isotope era" in which metabolite transformations were mapped with tracers, and (iii) the "era of biochemical genetics" where the expression of biosynthetic enzymes were manipulated to establish reaction sequences.(Chain, 1965)

This perspective that the complete cellular metabolome had been elucidated did not evolve greatly until recently when researchers began applying cutting-edge mass spectrometry to study metabolism comprehensively. The results have been unexpected. Thousands of signals can be detected from the metabolic extracts of biological samples whose masses do not match any of those predicted based on the conventional biochemical charts. Data from this new experimental approach, termed "metabolomics", have challenged the idea that the picture of cellular metabolism is complete. While it is unclear at this time precisely how many unknown metabolites are represented in metabolomic data, already metabolomic technologies have been applied to discover new metabolites and unexpected pathway fluxes that have important physiological relevance.(Dang et al., 2009; Kalisiak et al., 2009; Mathe et al., 2014; Patti et al., 2012a; Yore et al., 2014)

With the rise of metabolomics, a new era of biochemical discovery has begun. For the first time in several decades, excitement to discover new metabolites and pathways is at the forefront of biological research.(McKnight, 2010)

The unknown metabolome: a love and hate relationship—There are several technological platforms available for performing metabolomics, but liquid chromatography/mass spectrometry (LC/MS) is most commonly used for discovery (i.e., untargeted) experiments because thousands of signals are routinely detected from the metabolic extract of a biological sample.(Patti et al., 2012b) The biggest challenge in performing discovery metabolomic experiments is translating these signals, termed features, into metabolite identities. While there are certainly more resources available now compared to ten years ago, the process of making metabolite identifications is still low throughput. It is common to

spend weeks to months analyzing each metabolomic dataset and, even then, only a relatively small number (<20) of metabolite identifications are often made.

The possibility that hundreds or thousands of LC/MS signals might correspond to unknown metabolites that have yet to be described is one of the most exciting aspects of performing discovery metabolomics. This possibility is also the single reason that performing metabolomics is so challenging. For example, consider a situation where there is interest in identifying an ion with a mass-to-charge value of 808.118 as detected in negative-ionization mode. The first step would be searching the value of 808.118 in metabolomic databases such as METLIN and HMDB.(Tautenhahn et al., 2012a; Wishart et al., 2009) Given that modern mass spectrometers can routinely measure mass-to-charge ratios of metabolites with an error of 25 ppm or less, the databases would be searched for candidate compounds over the mass-to-charge interval 808.098–808.138. Currently, the only hit returned from this search is acetyl-CoA. Yet, because there may be unknown metabolites with mass-to-charge values within this interval that have not been input into metabolomic databases, this result would not substantiate identifying the ion with a mass-to-charge value of 808.118 as acetyl-CoA. Without a complete parts list of the metabolome, accurate mass is never sufficient to confidently identify a metabolite.

Instead, to reliably identify a metabolite, its mass-to-charge value, chromatographic retention time, isotopic pattern, and fragmentation data are generally used together in combination.(Patti et al., 2012b) The sum of this information is highly specific to a compound and therefore it is unlikely that any two metabolites, known or unknown, produce the same set of data. Using retention time and fragmentation data to identify metabolites, however, leads to some major experimental challenges. The remainder of my discussion is dedicated to briefly describing some bioinformatic resources that have been developed to help with the process.

Metabolomics on the cloud—One of the most widely used software packages for processing raw LC/MS-based metabolomic data is a program called XCMS. Just five years ago, installation and operation of XCMS required familiarity with the R programing language. Recently, however, a cloud-based version of XCMS was developed called XCMS Online that uses an intuitive graphical interface.(Tautenhahn et al., 2012b) Metabolomic data are uploaded to the XCMS Online server much like files are uploaded as an attachment to an email. After the data are processed, the results are integrated with the METLIN database so that candidates for the metabolite identities of each signal (i.e., feature) are listed. The candidate list is produced on the basis of each feature's mass-to-charge value. As discussed above, the mass-to-charge value of a feature alone is insufficient to substantiate a metabolite identification. To facilitate metabolite assignments based on fragmentation data, a number of metabolomic databases such as METLIN, HMDB, and MassBank have begun incorporating experimental fragmentation data acquired from model standards.(Horai et al., 2010; Tautenhahn et al., 2012a; Wishart et al., 2009) At this time, the METLIN database contains fragmentation data for more than 12,000 model standards. Current efforts are focused on further integrating XCMS Online and METLIN so that metabolite assignments can be made during data processing that are based on both mass-to-charge values and fragmentation data,

an approach that has been referred to as autonomous metabolomics.(Benton et al., 2014; Tautenhahn et al., 2012a)

Credentialing features: honey, I shrunk the metabolome—In a typical discovery profiling experiment performed with LC/MS-based technologies, thousands of features are routinely detected from most biological samples. When the mass-to-charge ratios of each detected feature are searched against metabolomic databases, a surprisingly small fraction provide database hits. The challenge with interpreting this result with respect to the size of the unknown metabolome is that each feature detected does not necessarily correspond to a different metabolite. As a consequence, the number of "unknown" compounds detected is overestimated upon simple inspection. For example, naturally occurring isotopes can result in one metabolite species being detected as multiple features. Mostly, this results from the relatively high concentration of ^{13}C that occurs naturally and results in a mass shift. Another reason a metabolite can produce multiple features is because metabolites can ionize as several adducts. That is, in addition to being detected as $[M+H]^+$, metabolites might also be detected as $[M+Na]^+$, $[M+NH_4]^+$, etc. Further, metabolites can sometimes fragment or form non-covalent interactions with other metabolites when entering the mass spectrometer. This too inflates the feature count independent of the number of metabolites present.

Most features corresponding to isotopes, adducts, and fragments can be readily identified by using software programs such as CAMERA.(Kuhl et al., 2012) But even after these processing steps, the number of features detected can still be larger than the number of metabolites present due to artifacts associated with contaminants, chemical noise, and bioinformatic noise. To identify and remove these artifactual features, isotopic labeling methods such as credentialing and isotope-ratio outlier analysis have been developed. (Mahieu et al., 2014; Stupp et al., 2013) In brief, identical samples with and without ^{13}C labels are mixed. Features of biological origin consequently produce a unique isotopic pattern in the mass spectra, whereas artifactual features do not. Features with the appropriate isotopic patterns are said to be "credentialed".

It is important to highlight that even when these extensive filtering strategies are applied to *Escherichia coli*, there remain hundreds of credentialed features whose mass-to-charge values do not return any matches in current metabolomic databases.(Mahieu et al., 2014) Of course, it is possible that some credentialed features returning database hits are themselves unknowns that have matching mass-to-charge values but do not have matching structures. Alternatively, it may be that some credentialed features not returning database hits are known metabolites that have undergone an extracellular transformation (e.g., during the extraction process). Only after every credentialed feature is structurally characterized will a definitive calculation of the number of unknowns detected be possible. Certainly the results are expected to be at least partially organism specific. The credentialing experiments performed thus far, however, suggest that the number of unknown metabolites detected by LC/MS-based metabolomics is much smaller than the thousands of features without database hits appearing in the raw data prior to filtering.

The "metabolomics era" of biochemical pathway elucidation—In 1955, Nicholson provided the first picture of the comprehensive cellular metabolome by

constructing a chart connecting the ~20 biochemical pathways that had then been characterized. By 1970, Nicholson's chart contained just over 400 metabolites. Today, comprehensive metabolic charts have expanded to now include more than four times as many metabolites.(Ogata et al., 1999; Thiele et al., 2013) The introduction of new technologies for studying metabolism has raised the question of how complete even that picture is.

Undoubtedly, advances in mass spectrometry-based metabolomics have reinvigorated research in metabolite discovery. The extent to which metabolomic technologies will impact the canonical view of comprehensive metabolism in textbooks remains to be seen, but it is interesting to consider that we have entered a new period of biochemical research. Recall that Chain organized the history of pathway discovery into three periods: the pre-isotope era, the isotope era, and the era of biochemical genetics. Building upon Chain's perspective, we might say that we have entered a fourth era of elucidating biochemical pathways: the "metabolomics era". In most cases, understanding how newly discovered metabolites integrate into our current comprehensive chart of metabolism will require using stable isotopes and metabolic flux analysis as described by Nicola Zamboni below.

Modern Stable-Isotope Metabolic Flux Analysis

Nicola Zamboni

While untargeted metabolite profiling approaches have yielded many important insights into a multitude of biological problems and may potentially lead to an update of our perspective of the number of metabolites present in a cell, conventional untargeted metabolomic experiments generally only provide information about metabolite concentration. This is a static snapshot that cannot be translated into a dynamic map of metabolite traffic on biochemical routes. A more complete understanding of biochemical pathways can be obtained by using metabolic flux analysis. Metabolic fluxes are the in vivo velocities of metabolic reactions. These include the rate of transformation of intermediates by enzymes and of the transport of metabolites between compartments. In single cells or multicellular organisms, metabolic fluxes are an emerging property as they depend on the systemic organization and interplay of enzymes, carriers, and substrates in the environment. Knowledge of metabolic fluxes is essential to unravel sites and mechanisms of metabolic regulation, and thus augment our understanding on how metabolism is embedded in cellular decisions. In diseased cells, information on fluxes paves the road for identification of selective therapeutic targets.

The frequent questions related to metabolic fluxes are (i) discovery of the catabolic fate of a given nutrient, (ii) discovery of the biosynthetic origin of a given intracellular compound, (iii) targeted analysis of fluxes at a specific metabolic node or reaction, or (iv) determination of cellular balances for redox (i.e., NADPH and NADH) and energy (ATP) carriers. In all these quests, the method of choice is metabolic flux analysis with isotopic tracers, e.g., ^{13}C and ^{2}H.

Basic principles of flux experiments—As any other reaction rate, metabolic fluxes cannot be measured directly. In cell cultures, but not in vivo, it is possible to measure the

consumption rate of nutrients and secretion of metabolic byproducts by monitoring the time-dependent depletion or production, respectively, in spent medium. Intracellularly, however, metabolite levels remain constant in a dynamic equilibrium between producing and consuming reactions and don't inform on molecular flow. Instead, intracellular fluxes can be accessed by administering nutrients labeled with stable isotopes, typically ^{13}C or ^{2}H. Uptake and enzymatic transformation of labeled substrates propagates heavy isotopes through the metabolic network in a flux-dependent fashion. In some cases – but not always – the resulting labeling pattern of metabolites can be used to infer from which nutrient it originated or, if alternative pathways exist, through which enzymatic route.

Two fundamental designs exist for metabolic flux analysis studies: stationary and non-stationary experiments. In stationary experiments, labeling patterns are evaluated at the end-point when they become invariant over time, i.e. an isotopic equilibrium was reached (Zamboni et al., 2009). The main advantage of stationary experiments is that at isotopic steady-state the measured labeling patterns are fully independent of metabolite levels. The latter can be safely neglected in the calculation of fluxes from labeling data facilitating data acquisition and interpretation. On the downside, stationary isotopic experiments are only useful to resolve the relative contribution of different pathways to the synthesis of a common derivative, but are not suited to quantify the flux in a simple linear pathway. Stationary labeling experiments must last sufficiently long to attain isotopic stationarity. In peripheral pathways, it can take days. During this time, conditions and fluxes must be constant. In cell biology, stationary metabolic flux analysis is best suited to determine what nutrients contributed to the biosynthesis of intracellular metabolites. For this purpose, single nutrients are uniformly ^{13}C-labeled while all other substrates are provided with natural labeling. At isotopic steady state, the ^{13}C-enrichment of each metabolite (i.e., its *fractional labeling*) indicates what ratio originated from the labeled substrate.

Non-stationary labeling experiments, in contrast, focus on the kinetics of tracer propagation before isotopic equilibrium (Wiechert and Nöh, 2013). This relaxes most limitations of stationary experiments: labeling experiments can be as short as a few minutes and it becomes possible to estimate fluxes within linear pathways. This increased power over stationary analyses comes with extra costs. First, many more time-points, and thus more samples and more instrument time, are necessary to assemble the data set. Second, data analysis is substantially more complex. Third and most important, non-stationary labeling transients depend both on metabolic fluxes and on intracellular metabolite concentration. Therefore, the latter have to be measured by canonical quantitative metabolomics and included in the interpretation.

Tracer choice—The labeling configuration of the tracer is a pivotal component of the labeling experiment. As mentioned before, 100% uniformly ^{13}C-labeled substrates are the preferred for identifying their intracellular fate in any cellular system. In microorganisms, the combination of uniformly ^{13}C-labeled and unlabeled substrates has been largely used to assess multiple fluxes in central metabolism with a single experiment. The mix between labeled and unlabeled forms of the same substrate allows to resolve alternative pathways that use different enzymes. The classical example is glucose catabolism through glycolysis or the pentose-phosphate pathway (PPP). With 100% ^{13}C-glucose, the two pathways are

indistinguishable because they both produce completely ^{13}C-labeled pyruvate. If 50% uniformly ^{13}C and 50% unlabeled glucose are mixed, catabolism through the transaldolase and transketolase in the PPP will combine labeled and unlabeled carbon backbones and lead to partly labeled pyruvate forms which differentiate from the either completely ^{13}C-labeled or unlabeled pyruvate molecules produced by glycolysis. In mammalian cells, however, there are many more interfering pathways and nutrients that prevent resolving several fluxes at once with generalist tracers. In contrast, positionally enriched tracers harboring stable isotopes only at specific positions are used to resolve important reactions. This strategy is particularly suited to assay decarboxylating enzymes if it is possible to label its substrate such as all label is lost as $^{13}CO_2$. This ad-hoc strategy is frequently used in the analysis of the oxidative PPP, pyruvate dehydrogenase, or oxidative TCA cycle.

Measurement of labeling patterns—Non-radioactive isotopes are amenable with instrumentation platforms such as mass spectrometry and nuclear magnetic resonance, which provide powerful tools for determining the isotopic patterns of numerous metabolites or macromolecules. The community is dominated by two types of instruments. Gas chromatography/mass spectrometry (GC/MS) is the most affordable technology and perfectly suited to analyze isotopic patterns in amino and organic acids, sugars, and fatty acids. GC/MS often adopts electron impact ionization, which intrinsically fragments analytes and therefore provides additional information on the localization of the isotopic label within the metabolite structure. In contrast, LC/MS uses a soft ionization technique which preserve molecular ions. Compared to GC/MS, it delivers less positional information but offers better coverage of endogenous metabolites. Therefore, LC/MS is preferred in non-stationary labeling experiments to monitor the isotopic transients in proximity of the important reactions and also quantify the levels of metabolites required for non-stationary data analysis.

Interpretation of labeling patterns—An ideally designed tracer experiment allows for verification of the working hypothesis through direct observation of the labeling data or the fractional labeling of a metabolite. Statistical significance is checked by simple univariate testing. This is a practical strategy when searching for the origin of a given metabolite, and the possible substrates or routes are characterized by different label content. It is also used in combination with selective inhibition of alternative pathways (i.e., isoenzymes) by gene knockout and knockdown or pharmacological inhibition, for which qualitative comparison of the resulting labeling patterns is sufficient to assess their relative contribution. An illustrative example is given by (Son et al., 2013).

Unfortunately, it is not always feasible to design such an explicit labeling experiment. Interpretation is complicated by the propagation and scrambling of tracer over the entire network following carbon fluxes. A metabolic alteration occurring in the early steps of tracer catabolism may affect the labeling patterns of all - close and distant - metabolites. The logic consequence is that a difference in labeling patterns could have been caused by upstream alterations in metabolism. An unbiased analysis of labeling patterns must consider the potential effect of distant pathways. This is particularly challenging for metabolic cycles, reversible reactions, and extensive compartmentalization that introduces numerous degrees

of freedom. Already in mid-sized systems such as central carbon metabolism, this task exceeds human intuition and calls for formal analysis.

Mathematical inference of flux data from labeling patterns can tackle the problem from essentially two complementary angles: globally or locally. In the global approach, all cellular fluxes are sought simultaneously to find the set of flux solutions that best matches measured labeling patterns. This approach builds on constraint-based modeling and therefore requires possibly accurate measurement of the uptake and secretion rate of all cellular substrates and products, respectively, to balance all species in the system. Intracellular fluxes are fitted iteratively in a procedure that becomes computationally hard with increasing number of model size and dimensions. Although powerful implementations exist both for the stationary and non-stationary case (e.g. 13CFLUX2 (Weitzel et al., 2013) and INCA (Young, 2014)), this approach quickly hits its limits when solving flux in central carbon metabolism of mammalian cell lines grown with a variety of carbon sources. In practice, precise flux estimations are obtained only by combining multiple labeling experiments with different tracers and by fitting fluxes within amended models in which reactions or compartments were neglected. Unless justified by independent data, e.g., on protein expression, the model simplifications increase the risk of returning overly precise but inaccurate flux values.

The local approach focuses only on a small part of the metabolic network, typically not more than a few reactions sharing a common metabolite, and is geared towards estimating local fluxes based solely on local data. In the stationary case, a local analysis only provides relative fluxes, i.e., so-called flux ratios. Absolute flux rates can be estimated with nonstationary labeling experiments and local metabolome concentrations. Compared to the global approach, the local analysis doesn't require any uptake and secretion measurements and is computationally much simpler. There is no special software to recommend for local flux analysis, but the principles are described in (Hörl et al., 2013; Yuan et al., 2008; Zamboni et al., 2009). The local analysis is an efficient tactic to cope with complex flux analysis problems, i.e., resolving fluxes in mammalian cells (Zamboni, 2011), by decomposing the daunting challenge of resolving many fluxes at once in tractable tasks to be solved with ad-hoc designs. This is best exemplified in the study by Fan et al. (Fan et al., 2014), in which numerous experiments with ^{2}H, ^{13}C, and ^{14}C tracers and local and quantitative analyses were assembled into a cell-wide and yet accurate picture of NADPH metabolism.

Frequent limitations and pitfalls—Calculability of fluxes in real systems depends on multiple factors (Zamboni, 2011; Zamboni et al., 2009). Some factors are structural, e.g., the mapping of atoms in metabolic reactions or the availability of mass spectrometry data. These are known a priori and can thus be included in experimental design before the labeling experiment is performed. Several factors are more fluid and difficult to capture in advance. These include the mixing of label caused by fast reversible reactions, the uncertainty in determining labeling patterns for metabolites that can exist in different cellular compartments, the actual speed of label propagation far away from substrate entry, etc. The extent by which these latter factors affect and jeopardize flux measurement depends on the actual fluxes and metabolic state of the cell and can only be assessed *a posteriori*.

Experience teaches us that biased flux analyses are mostly caused by two issues: over-generalization and over-simplification. Generalization refers to the transfer of knowledge or assumptions from a previous study to a new study. This is commonly done to avoid measuring biomass composition or to claim that the effect of enzyme reversibility, or compartmentalization, is negligible. Although these aspects were carefully ascertained in the previous study, it is risky to generalize across different cell types, different environments, or different proliferation rates. Simplification of the model was discussed before as a habit to enable flux calculation in underdetermined systems. It is common practice to disregard compartmentalization of enzymes and intermediates or reaction/pathways that are meant to have only a secondary role. It must be stressed that data are only good to prove models wrong. Hence, a good fit of measurements to a given model are not proof that the underlying fluxes reflect reality unless all plausible models were exhaustively tested.

Experience has also thought us that a purist approach without some simplifications and generalization drastically prevents the generation of testable hypotheses on metabolic fluxes. Finding the right equilibrium between calculation and assumptions depends on critical assessment. Results should be openly challenged by the scientist. Positive and negative controls should be included in the analysis as for any other analytical technique.

Biological Lessons from Metabolite Profiling

Alan Saghatelian

How do we detect and quantify metabolites?—There are many ways to detect and quantify metabolite levels or to measure labeling patterns in compounds. Early methods to measure metabolites relied on thin layer chromatography (TLC) of abundant compounds or radiolabeling followed by TLC for less abundant compounds. The emergence of gas chromatography and GC/MS proved incredibly important to improving our understanding of cellular metabolism. Still, today, GC/MS is an important and widely used method for detecting metabolites, as well as a technology often used for performing metabolic flux analysis as described by Nicola Zamboni above. Indeed, GC/MS was used in much of the early metabolomics work, which was done in plants (Weckwerth et al., 2004). Plants were an excellent choice because the combination of metabolomics and genetics enabled the metabolic characterization of many genes. Weckwerth, Fiehn and colleagues used GC/MS-based metabolomics of potatoes to identify metabolic functions of genes with silent phenotypes, demonstrating the power of metabolomics in gene characterization (Weckwerth et al., 2004).

Nuclear magnetic resonance (NMR) spectroscopy has also been used for metabolomics, provides rapid metabolome analysis, and can sometimes be used to resolve the location of stable isotopes within a molecular structure.(Cheng et al., 2011) In an interesting demonstration of NMR metabolomics, Dumas, et. al. described changes in choline metabolism that are linked to fatty liver through measurements of plasma and urine metabolites (Dumas et al., 2006).

The most popular method for metabolomics, however, has been mass spectrometry-based metabolomics. Shotgun metabolomics refers to methods that infuse extracted metabolome

samples without any chromatography (Adamovich et al., 2014; Han and Gross, 2005). Like NMR, these methods are extremely fast and enable rapid detection and quantitation of metabolites. Adamovich and co-workers utilized shotgun lipidomics to quantify levels of acyl glycerols, sphingolipids, and phospholipids in mice that lack *per1* and *per2*, two key circadian rhythm genes, and observed the regulation of trigylceride levels by the circadian clock (Adamovich et al., 2014).

Greater coverage of the metabolome is achieved by combining chromatography with mass spectrometry. Different types of chromatography have been successfully employed for greater metabolome coverage, including the aforementioned GC, liquid chromatography, and capillary electrophoresis (CE) (Patti, 2011; Soga et al., 2003). The choice of chromatography methods depends on metabolites of interest. GC/MS has been used extensively, and successfully, to analyze the metabolome and flux distributions, but metabolites must be chemically derivatized so they are volatile (Weckwerth et al., 2004). Larger molecules or molecules that are heat labile are therefore not compatible with GC. Capillary electrophoresis-mass spectrometry (CE-MS) is excellent for charged molecules such as small organic acids, amino acids, and charged (phosphorylated) sugars (Soga et al., 2003). Lastly, of the methods currently in use, LC/MS is the most widely used technique for metabolomics because it provides the greatest coverage of the metabolome as discussed by Gary Patti in preceding sections (Want et al., 2007).

The first step in any metabolomics or stable-isotope tracer experiment is the isolation of the metabolites from cells or tissues. The extraction method can vary depending on the metabolites of interest, with robust protocols for hydrophilic (Bennett et al., 2008) and hydrophobic metabolites having been reported (Inloes et al., 2014). Isolated metabolites are then analyzed by LC/MS. Many types of LC columns have been developed for hydrophilic and hydrophobic analytes. In a metabolomic profiling experiment, the mass spectrometer is set to scan a large mass range (i.e. 75–1500 m/z) such that any ionizable metabolite of sufficient concentration can be detected and quantified. By scanning a broad mass range, metabolomic profiling experiments are less sensitive and can miss lower abundant metabolites. Therefore, targeted methods have been developed for a number of important metabolites, such as oxysterols (McDonald et al., 2012), which provide the sensitivity to analyze the entire metabolite family, but not the entire metabolome. Similarly targeted methods are often used in metabolic flux analysis (Yuan et al., 2008).

In addition to using stable isotopes for flux analysis, stable isotope standards are also used to validate changing metabolite levels identified during a metabolomics or lipidomics experiment. The use of stable isotope labeled standards (^{2}H, ^{13}C or ^{15}N) for absolute quantitation is referred to as isotope dilution mass spectrometry (IDMS) (Bennett et al., 2008). Stable isotope standards are added into the sample during extraction and the ratio of these molecules to the endogenous metabolite enables absolute quantitation of the endogenous metabolite. These experiments are performed in combination with targeted LC/MS, which provides improved signal-to-noise and sensitivity, resulting in a more accurate quantitation. An analogy in genomics is the use of quantitative PCR to validate the results from an RNA-Seq experiment. Validation by IDMS provides the highest confidence

in fold changes that were observed during the profiling experiment and also confirms the structural assignment.

Biological lessons from metabolomics and metabolic flux analysis—As metabolomics and metabolic flux experiments have become more prevalent, their use in biology and biomedical research has expanded and led to fundamental insights in several fields. The cell biologists Atilla-Gokcumen, and colleagues, for example, combined lipidomics with cell biology to identify the role of lipids, and lipid modifying enzymes, in cytokinesis, a physical process in cell division that results in the splitting of the cytoplasm into newly divided cells (Atilla-Gokcumen et al., 2014). In this study, the group utilized untargeted lipidomics to identify several lipids that were changing (accumulating) during cytokinesis and/or in the midbody, a transient structure that appears during the end of cytokinesis and is required for the completion of cell division. Knockdown of the enzymes that metabolize key midbody lipids affected cytokinesis, revealing new metabolic regulators of this fundamental process.

A creative use of metabolomics has also revealed new RNA modifications. Chen, Kowtoniuk and co-workers began by purifying cellular RNAs to remove small molecules (Chen et al., 2009). These RNAs are then digested with nuclease P1, which breaks down the RNAs into nucleosides monophosphates, which can then be analyzed using LC/MS-based metabolomics. Comparison of digested RNA to a sample treated with an inactive nuclease P1 (heat denatured) provided the first evidence for an NAD-linked RNA in *E.coli* and *S.venezuelae*. The high levels of this modification (3000 copies/cell) suggested that NAD-modified RNA served a cellular function. Indeed, subsequent work by Cahová, et. al. demonstrated that 5'-NAD serves as a bacterial RNA cap to regulate RNA degradation (Cahova et al., 2014). This example highlights the power of metabolomics to make new discoveries in well-studied systems such as *E.coli*, which bodes well for the continued impact of metabolomics in biological and biomedical research.

Cancer biologists have used metabolomics and metabolic flux analysis to better understand the contribution of aerobic glycolysis, the "Warburg effect", to the cancer phenotype (Lunt and Vander Heiden, 2011). Initially, Warburg hypothesized that increased glycolysis was associated with mitochondrial dysfunction, but newer studies have demonstrated that cancer-causing genes drive the transition to increased glycolytic rates (Ramanathan et al., 2005). More recently, Loscale, et. al. used metabolic flux analysis to understand glucose utilization in cancer cells (Locasale et al., 2011). They observed that a substantial fraction of glucose carbon in cancer cells is diverted into amino acid metabolism (serine and glycine) through the overexpression of phosphoglycerate dehydrogenase (PHGDH). Moreover, a reduction of PHGDG expression impaired proliferation, indicating the importance of this pathway in driving cell growth. Presumably, increased amino acid production is needed to support protein translation. This is one of several examples that indicate a role for aerobic glycolysis in producing biomass for cellular proliferation (Lunt et al., 2015).

Metabolomics has also been used to identify the functions of genes associated with cancer. Mutations in the isocitrate dehydrogenase 1 (IDH1), for example, had been linked with human brain cancer, but the biochemical role of the mutant IDH1 was unclear. The classical

function of IDH1 is to convert isocitrate to alpha-ketoglutarate, but cancer-associated IDH1 mutations confer a new activity onto this enzyme. Untargeted metabolomics of mutant $IDH1_{Arg132His}$ by Dang and colleagues revealed that this mutant enzyme is able to convert alpha-ketoglutarate into 2-hydroxyglutarate (2-HG) (Dang et al., 2009). 2-HG was referred to as an oncometabolite since elevation of 2-HG levels has been shown to increase the risk of brain tumors. Inhibitors of mutant IDH1 that prevent 2-HG production slow tumor growth (Rohle et al., 2013), demonstrating the value of these metabolomics studies in revealing a new therapeutic target.

The combination of genetics, metabolomics and chemical biology has also proven useful in studying mutations in DDHD2, a poorly characterized enzyme linked to hereditary spastic paraplegia (HSP). Inoles and colleagues created a mouse model that lacked DDHD2 ($DDHD2^{-/-}$ mice) to determine the physiological consequences of loss of DDHD2 activity, and used this model to better understand the biochemical function of this enzyme using lipidomics (Inloes et al., 2014). $DDHD2^{-/-}$ mice showed age-related dysfunction in motor function and cognitive function. Lipidomic comparison of $DDHD2^{-/-}$ to WT mice revealed a marked accumulation of tryglycerides in the brains of $DDHD2^{-/-}$ mice, which was accompanied by lipid droplet accumulation. DDHD2 had been reported to have phospholipase activity *in vitro* (Higgs et al., 1998), but the analysis of the $DDHD2^{-/-}$ mice clearly showed that this enzyme is a brain triglyceride hydrolase with no detectable impact on phospholipids. This DDHD2 analysis underscores the importance of characterizing enzymes in their native environments where the contributions of the entire proteome can be taken into account. Furthermore, the correct characterization of the metabolic defects associated with $DDHD2^{-/-}$ provides insights that can be utilized in the eventual development of therapeutics to treat HSP.

In closing, metabolomics and metabolic flux analysis are providing new insights into numerous biological problems. In addition to the aforementioned studies, additional work has utilized metabolomics to study metabolic disease, such as metabolic risk factors for type 2 diabetes (Rhee et al., 2011; Wang et al., 2013) and the identification of metabolically beneficial lipids (Cao et al., 2008; Oh et al., 2010), including novel lipids (Yore et al., 2014). Metabolomics is also being used by immunologists and microbiologists to study the impact of the gut microbiome on physiology (Ridaura et al., 2013). As metabolomics and metabolic flux analysis continues to be utilized in the biological and biomedical sciences, we will learn more about the role of metabolism in regulating important biological processes, and in the case of disease biology these studies have the potential to reveal new therapeutic opportunities (Rohle et al., 2013).

References

Adamovich Y, Rousso-Noori L, Zwighaft Z, Neufeld-Cohen A, Golik M, Kraut-Cohen J, Wang M, Han X, Asher G. Circadian clocks and feeding time regulate the oscillations and levels of hepatic triglycerides. Cell metabolism. 2014; 19:319–330. [PubMed: 24506873]

Atilla-Gokcumen GE, Muro E, Relat-Goberna J, Sasse S, Bedigian A, Coughlin ML, Garcia-Manyes S, Eggert US. Dividing cells regulate their lipid composition and localization. Cell. 2014; 156:428–439. [PubMed: 24462247]

Bennett BD, Yuan J, Kimball EH, Rabinowitz JD. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat Protoc. 2008; 3:1299–1311. [PubMed: 18714298]

Benton HP, Ivanisevic J, Mahieu NG, Kurczy ME, Johnson CH, Franco L, Rinehart D, Valentine E, Gowda H, Ubhi BK, et al. An autonomous metabolomic workflow for untargeted profiling. Anal Chem. 2014

Cahova H, Winz ML, Hofer K, Nubel G, Jaschke A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. Nature. 2014

Cao H, Gerhold K, Mayers JR, Wiest MM, Watkins SM, Hotamisligil GS. Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism. Cell. 2008; 134:933–944. [PubMed: 18805087]

Chain EB. LANDMARKS AND PERSPECTIVES IN BIOCHEMICAL RESEARCH. Br Med J. 1965; 1:209–220. [PubMed: 14228153]

CONCL. Chen YG, Kowtoniuk WE, Agarwal I, Shen Y, Liu DR. LC/MS analysis of cellular RNA reveals NAD-linked RNA. Nature chemical biology. 2009; 5:879–881. [PubMed: 19820715]

Cheng T, Sudderth J, Yang C, Mullen AR, Jin ES, Mates JM, DeBerardinis RJ. Pyruvate carboxylase is required for glutamine-independent growth of tumor cells. Proc Natl Acad Sci U S A. 2011; 108:8674–8679. [PubMed: 21555572]

Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature. 2009; 462:739–744. [PubMed: 19935646]

Dumas ME, Barton RH, Toye A, Cloarec O, Blancher C, Rothwell A, Fearnside J, Tatoud R, Blanc V, Lindon JC, et al. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:12511–12516. [PubMed: 16895997]

Fan J, Ye J, Kamphorst JJ, Shlomi T, Thompson CB, Rabinowitz JD. Quantitative flux analysis reveals folate-dependent NADPH production. Nature. 2014; 510:298–302. [PubMed: 24805240]

Han X, Gross RW. Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. Mass spectrometry reviews. 2005; 24:367–412. [PubMed: 15389848]

Higgs HN, Han MH, Johnson GE, Glomset JA. Cloning of a phosphatidic acid-preferring phospholipase A1 from bovine testis. Journal of Biological Chemistry. 1998; 273:5468–5477. [PubMed: 9488669]

Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, et al. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010; 45:703–714. [PubMed: 20623627]

Hörl M, Schnidder J, Sauer U, Zamboni N. Non-stationary ^{13}C-metabolic flux ratio analysis. Biotechnol Bioeng. 2013; 110:3164–3176. [PubMed: 23860906]

Inloes JM, Hsu KL, Dix MM, Viader A, Masuda K, Takei T, Wood MR, Cravatt BF. The hereditary spastic paraplegia-related enzyme DDHD2 is a principal brain triglyceride lipase. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:14924–14929. [PubMed: 25267624]

Kalisiak J, Trauger SA, Kalisiak E, Morita H, Fokin VV, Adams MW, Sharpless KB, Siuzdak G. Identification of a new endogenous metabolite and the characterization of its protein interactions through an immobilization approach. J Am Chem Soc. 2009; 131:378–386. [PubMed: 19055353]

Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. Anal Chem. 2012; 84:283–289. [PubMed: 22111785]

Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, Bass AJ, Heffron G, Metallo CM, Muranen T, Sharfi H, et al. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nature genetics. 2011; 43:869–874. [PubMed: 21804546]

Lunt SY, Muralidhar V, Hosios AM, Israelsen WJ, Gui DY, Newhouse L, Ogrodzinski M, Hecht V, Xu K, Acevedo PN, et al. Pyruvate kinase isoform expression alters nucleotide synthesis to impact cell proliferation. Molecular cell. 2015; 57:95–107. [PubMed: 25482511]

Lunt SY, Vander Heiden MG. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. Annu Rev Cell Dev Biol. 2011; 27:441–464. [PubMed: 21985671]

Mahieu NG, Huang X, Chen YJ, Patti GJ. Credentialed Features: A Platform to Benchmark and Optimize Untargeted Metabolomic Methods. Anal Chem. 2014

Mathe EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, Bowman ED, Shields PG, Idle JR, Smith PB, Anami K, et al. Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. Cancer Res. 2014; 74:3259–3270. [PubMed: 24736543]

McDonald JG, Smith DD, Stiles AR, Russell DW. A comprehensive method for extraction and quantitative analysis of sterols and secosteroids from human plasma. Journal of lipid research. 2012; 53:1399–1409. [PubMed: 22517925]

McKnight SL. On Getting There from Here. Science. 2010; 330:1338–1339. [PubMed: 21127243]

Nicholson, SDaDE. An Introduction to Metabolic Pathways. Oxford and Edinburgh: Blackwell Scientific Publications; 1970.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 1999; 27:29–34. [PubMed: 9847135]

Oh DY, Talukdar S, Bae EJ, Imamura T, Morinaga H, Fan W, Li P, Lu WJ, Watkins SM, Olefsky JM. GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. Cell. 2010; 142:687–698. [PubMed: 20813258]

Patti GJ. Separation strategies for untargeted metabolomics. J Sep Sci. 2011; 34:3460–3469. [PubMed: 21972197]

Patti GJ, Yanes O, Shriver LP, Courade JP, Tautenhahn R, Manchester M, Siuzdak G. Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin. Nat Chem Biol. 2012a; 8:232–234. [PubMed: 22267119]

Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012b; 13:263–269. [PubMed: 22436749]

Ramanathan A, Wang C, Schreiber SL. Perturbational profiling of a cell-line model of tumorigenesis by using metabolic measurements. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:5992–5997. [PubMed: 15840712]

Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, Yang E, Farrell L, Fox CS, O'Donnell CJ, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. The Journal of clinical investigation. 2011; 121:1402–1411. [PubMed: 21403394]

Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science. 2013; 341:1241214. [PubMed: 24009397]

Rohle D, Popovici-Muller J, Palaskas N, Turcan S, Grommes C, Campos C, Tsoi J, Clark O, Oldrini B, Komisopoulou E. An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. Science. 2013; 340:626–630. [PubMed: 23558169]

Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. Journal of proteome research. 2003; 2:488–494. [PubMed: 14582645]

Son J, Lyssiotis CA, Ying H, Wang X, Hua S, Ligorio M, Perera RM, Ferrone CR, Mullarky E, Shyh-Chang N, et al. Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. Nature. 2013; 496:101–105. [PubMed: 23535601]

Stupp GS, Clendinen CS, Ajredini R, Szewc MA, Garrett T, Menger RF, Yost RA, Beecher C, Edison AS. Isotopic ratio outlier analysis global metabolomics of Caenorhabditis elegans. Anal Chem. 2013; 85:11858–11865. [PubMed: 24274725]

Tautenhahn R, Cho K, Uritboonthai W, Zhu Z, Patti GJ, Siuzdak G. An accelerated workflow for untargeted metabolomics using the METLIN database. Nat Biotechnol. 2012a; 30:826–828. [PubMed: 22965049]

Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. Anal Chem. 2012b; 84:5035–5039. [PubMed: 22533540]

Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, et al. A community-driven global reconstruction of human metabolism. Nat Biotechnol. 2013; 31:419–425. [PubMed: 23455439]

Wang TJ, Ngo D, Psychogios N, Dejam A, Larson MG, Vasan RS, Ghorbani A, O'Sullivan J, Cheng S, Rhee EP, et al. 2-Aminoadipic acid is a biomarker for diabetes risk. The Journal of clinical investigation. 2013; 123:4309–4317. [PubMed: 24091325]

Want EJ, Nordstrom A, Morita H, Siuzdak G. From exogenous to endogenous: the inevitable imprint of mass spectrometry in metabolomics. J Proteome Res. 2007; 6:459–468. [PubMed: 17269703]

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O. Differential metabolic networks unravel the effects of silent plant phenotypes. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:7809–7814. [PubMed: 15136733]

Weitzel M, Noh K, Dalman T, Niedenfuhr S, Stute B, Wiechert W. 13CFLUX2--high-performance software suite for (13)C-metabolic flux analysis. Bioinformatics. 2013; 29:143–145. [PubMed: 23110970]

Wiechert W, Nöh K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. Curr Opin Biotechnol. 2013; 24:979–986. [PubMed: 23623747]

Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, et al. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 2009; 37:D603–610. [PubMed: 18953024]

Yore MM, Syed I, Moraes-Vieira PM, Zhang T, Herman MA, Homan EA, Patel RT, Lee J, Chen S, Peroni OD, et al. Discovery of a class of endogenous mammalian lipids with antidiabetic and anti-inflammatory effects. Cell. 2014; 159:318–332. [PubMed: 25303528]

Young JD. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. Bioinformatics. 2014; 30:1333–1335. [PubMed: 24413674]

Yuan J, Bennett BD, Rabinowitz JD. Kinetic flux profiling for quantitation of cellular metabolic fluxes. Nat Protoc. 2008; 3:1328–1340. [PubMed: 18714301]

Zamboni N. 13C metabolic flux analysis in complex systems. Curr Opin Biotechnol. 2011; 22:103–108. [PubMed: 20833526]

Zamboni N, Fendt SM, Rühl M, Sauer U. ^{13}C-based metabolic flux analysis. Nat Protoc . 2009; 4:878–892. [PubMed: 19478804]