PERSPECTIVES

INNOVATION

Metabolomics: the apogee of the omics trilogy

Gary J. Patti, Oscar Yanes and Gary Siuzdak

Abstract | Metabolites, the chemical entities that are transformed during metabolism, provide a functional readout of cellular biochemistry. With emerging technologies in mass spectrometry, thousands of metabolites can now be quantitatively measured from minimal amounts of biological material, which has thereby enabled systems-level analyses. By performing global metabolite profiling, also known as untargeted metabolomics, new discoveries linking cellular pathways to biological mechanism are being revealed and are shaping our understanding of cell biology, physiology and medicine.

Metabolites are small molecules that are chemically transformed during metabolism and, as such, they provide a functional readout of cellular state. Unlike genes and proteins, the functions of which are subject to epigenetic regulation and post-translational modifications, respectively, metabolites serve as direct signatures of biochemical activity and are therefore easier to correlate with phenotype. In this context, metabolite profiling, or metabolomics, has become a powerful approach that has been widely adopted for clinical diagnostics.

The metabolome — typically defined as the collection of small molecules produced by cells — offers a window for interrogating how mechanistic biochemistry relates to cellular phenotype. With developments in mass spectrometry, it is now possible to rapidly measure thousands of metabolites simultaneously from only minimal amounts of sample[1]. In particular, recent innovations in instrumentation, bioinformatic tools and software enable the comprehensive analysis of cellular metabolites without bias. In many instances, these metabolites can be spatially localized within biological specimens with imaging mass spectrometry[2,3].

The application of these technologies has revealed system-wide alterations of unexpected metabolic pathways related to phenotypic perturbations. Moreover, many of the molecules detected are currently not included in databases and metabolite repositories, indicating the extent to which our picture of cellular metabolism is incomplete[4,5]. Nonetheless, the field of metabolomics has made remarkable progress within the past decade and has implemented new tools that have offered mechanistic insights by allowing for the correlation of biochemical changes with phenotype.

In this Innovation article, we first define and differentiate between the targeted and untargeted approaches to metabolomics. We then highlight the value of untargeted metabolomics in particular and outline a guide to performing such studies. Finally, we describe selected applications of untargeted metabolomics and discuss their potential in cell biology.

> “metabolites serve as direct signatures of biochemical activity”

Designing a metabolomic experiment

The first step in performing metabolomics is to determine the number of metabolites to be measured. In some instances, it may be of interest to examine a defined set of metabolites by using a targeted approach. In other cases, an untargeted or global approach may be taken in which as many metabolites as possible are measured and compared between samples without bias. Ultimately, the number and chemical composition of metabolites to be studied is a defining attribute of any metabolomic experiment and shapes experimental design with respect to sample preparation and choice of instrumentation.

Targeted metabolomics. This approach refers to a method in which a specified list of metabolites is measured, typically focusing on one or more related pathways of interest[6]. Targeted metabolomic approaches are commonly driven by a specific biochemical question or hypothesis that motivates the investigation of a particular pathway (FIG. 1a). This approach can be effective for pharmacokinetic studies of drug metabolism as well as for measuring the influence of therapeutics or genetic modifications on a specific enzyme[7]. Developments in mass spectrometry and nuclear magnetic resonance (NMR) offer distinct advantages for performing targeted metabolomic studies because of their specificity and quantitative reproducibility; however, there are many analytical tools available for measuring metabolites that could in principle be considered, such as ultraviolet-visible spectroscopy and flame ionization. Although the term ‘metabolomics’ was only recently coined, examples of targeted studies of metabolites date back to the earliest of scientific inquiries[8–12]. Therefore, there is a wealth of literature investigating optimal protocols for the sample preparation and analysis of specific classes of metabolites that has been discussed extensively elsewhere[13–17].

Not to diminish their significance, targeted approaches have undoubtedly played an important part in the development of the field of metabolomics. In particular, advances have been made in using triple quadrupole (QqQ) mass spectrometry to perform selected reaction monitoring experiments such that routine methods are now available for analysing most of the metabolites in central carbon metabolism, as well as amino acids and nucleotides at

© 2012 Macmillan Publishers Limited. All rights reserved

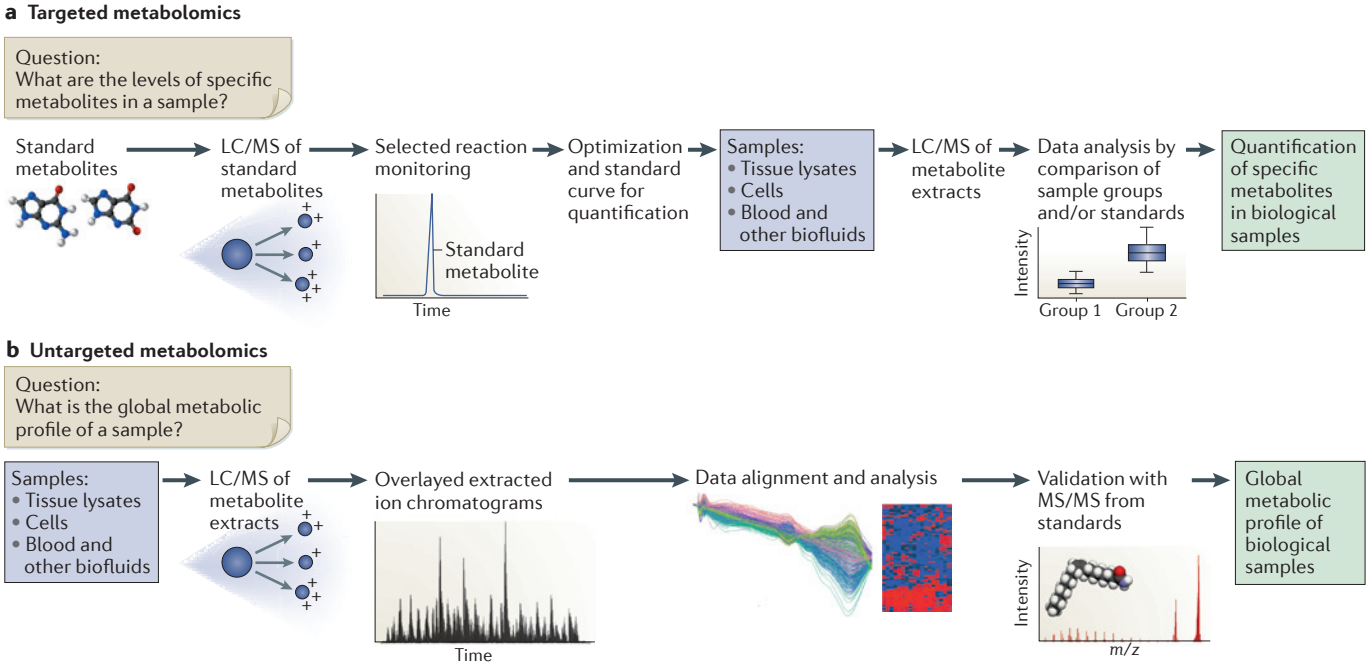

Figure 1 | **The targeted and untargeted workflow for LC/MS-based metabolomics. a** | In the triple quadrupole (QqQ)-based targeted metabolomic workflow, standard compounds for the metabolites of interest are first used to set up selected reaction monitoring methods. Here, optimal instrument voltages are determined and response curves are generated for absolute quantification. After the targeted methods have been established on the basis of standard metabolites, metabolites are extracted from tissues, biofluids or cell cultures and analysed. The data output provides quantification only of those metabolites for which standard methods have been built. **b** | In the untargeted metabolomic workflow, metabolites are first isolated from biological samples and subsequently analysed by liquid chromatography followed by mass spectrometry (LC/MS). After data acquisition, the results are processed by using bioinformatic software such as XCMS to perform nonlinear retention time alignment and identify peaks that are changing between the groups of samples measured. The *m/z* values for the peaks of interest are searched in metabolite databases to obtain putative identifications. Putative identifications are then confirmed by comparing tandem mass spectrometry (MS/MS) data and retention time data to that of standard compounds. The untargeted workflow is global in scope and outputs data related to comprehensive cellular metabolism.

their naturally occurring physiological concentrations[18–20]. These developments provide a highly sensitive and robust method for measuring a significant number of biologically important metabolites with relatively high throughput. Additionally, QqQ mass spectrometry methods are quantitatively reliable and therefore offer opportunities to achieve absolute quantification of low-concentration metabolites that are difficult to detect with less sensitive methods such as NMR (FIG. 1a). By applying QqQ mass spectrometry-based methods to human plasma, targeted lists of metabolites can be screened as potential metabolic signatures for disease. For example, targeted screening recently revealed citric acid metabolites and a small group of essential amino acids as metabolic signatures of myocardial ischaemia and diabetes, respectively[21,22]. In another diabetes-related study, targeted metabolomic methods were used to investigate patient response to glucose challenge[23]. Here, the levels of specific plasma metabolites were measured after glucose ingestion to determine insulin response in patients.

Untargeted metabolomics. Untargeted metabolomic methods are global in scope and have the aim of simultaneously measuring as many metabolites as possible from biological samples without bias (FIG. 1b). Although untargeted metabolomics can be performed by using either NMR or mass spectrometry technologies, liquid chromatography followed by mass spectrometry (LC/MS) enables the detection of the most metabolites and has therefore been the technique of choice for global metabolite profiling efforts[24–27]. By using LC/MS-based metabolomic methods, thousands of peaks can be routinely detected from biological samples[14,28,29] (FIG. 1b). Each of these peaks is referred to as a metabolite feature and corresponds to a detected ion with a unique mass-to-charge ratio and a unique retention time (it should be noted that some metabolites may produce more than one peak).

> “The untargeted [metabolomic] workflow is global in scope and outputs data related to comprehensive cellular metabolism.”

In contrast to targeted metabolomic results, untargeted metabolomic data sets are exceedingly complex, with file sizes on the order of gigabytes per sample for some new high-resolution mass spectrometry instruments. Manual inspection of the thousands of peaks detected is impractical and is complicated by experimental drifts in instrumentation. In LC/MS experiments, for example, there are deviations in retention time from sample to sample as a consequence of column degradation, sample carryover, small fluctuations in room temperature and mobile phase pH, as well as other variations. Although these challenges initially presented substantial obstacles for interpreting untargeted profiling data, major progress has been made in the past decade such that the ability to measure dysregulated peaks in global metabolomic data sets has now become routine with the introduction of metabolomic software such as MathDAMP, MetAlign, MZMine and XCMS[1,30–34]. These accomplishments have already revealed that an astounding number

© 2012 Macmillan Publishers Limited. All rights reserved

of metabolites remain uncharacterized with respect to their structure and function and, furthermore, that many of these uncharacterized metabolites change as a function of health and disease[4]. It is in this area that untargeted metabolomics has great potential to provide insights into fundamental biological processes. The remainder of this article will focus on the untargeted metabolomic approach.

Impetus for untargeted metabolomics. In 1941, G. Beadle and E. L. Tatum proposed the one gene–one enzyme hypothesis[35]. This hypothesis was based on their experimental results showing that X-ray-induced mutant strains of the fungus *Neurospora crassa* were unable to carry out specific biochemical reactions[35,36]. By systematically adding individual compounds to minimal *N. crassa* media and screening for those that rescued the growth of mutant strains, Beadle and Tatum identified metabolites whose biosynthesis had been affected by genetic mutation. In doing so they were the first to directly connect genotype to phenotype at the molecular level. From their results they purported that a single gene serves as the primary control of a single function, in this case a specific biosynthetic reaction.

In many ways, modern day metabolomic experiments are similar in that they seek to connect genotype and phenotype by metabolite screening (FIG. 2a). The experimental screening methods used today, however, are much advanced and allow us to study many more compounds simultaneously. Additionally, contemporary metabolic profiling experiments have the advantage of being complemented by genomic sequencing and proteomic screening[37–40]. The field of systems biology has emerged from the combination of these global analyses and has shown us that the effects of a single, non-lethal gene mutation can be dauntingly large[41]. Indeed, single gene mutations can affect a considerable number of metabolic pathways, thereby complicating the hypothesis that a single gene controls a single function (FIG. 2b). Moreover, mutations in some unique genes have unexpected phenotypic effects. As an example, consider the abnormal dauer formation (*daf-2*) gene, which encodes an insulin-like receptor in the nematode worm *Caenorhabditis elegans*. Mutations in *daf-2* cause *C. elegans* to live more than twice as long as its wild-type counterpart and result in alterations in the abundance of at least 86 identified proteins[42,43]. Or, as another example, consider genes that encode for enzymes of the phosphoinositide 3-kinase family. The protein products of these genes function in cell growth, proliferation, differentiation, motility and signal transduction, and mutations in these genes are thought to have an oncogenic role in some cancers[44].

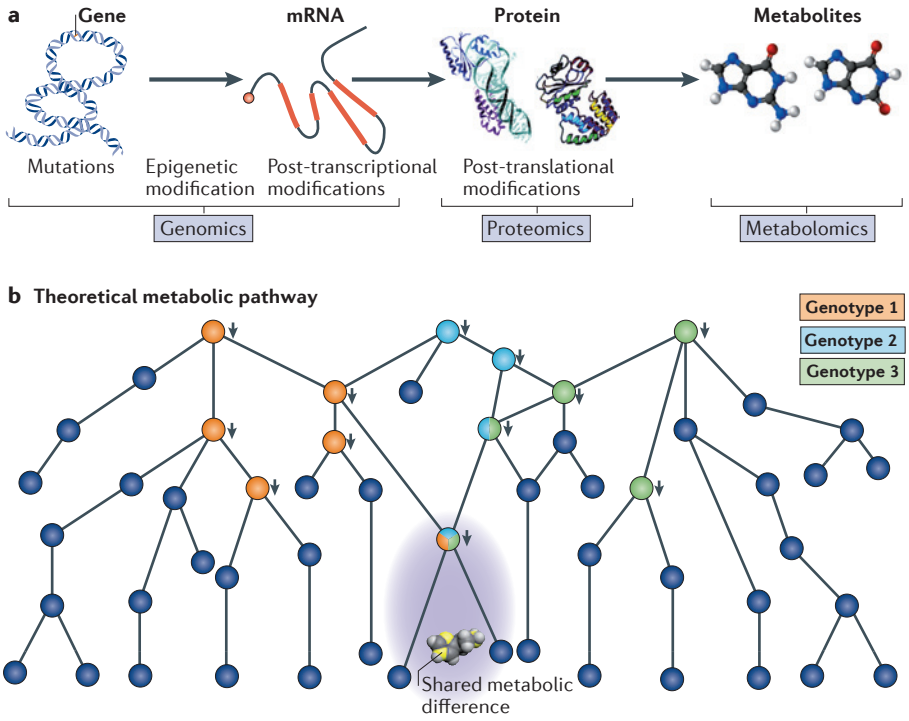

Figure 2 | **The central dogma of biology and the omic cascade. a** | Whereas genes and proteins are subject to regulatory epigenetic processes and post-translational modifications, respectively, metabolites represent downstream biochemical end products that are closer to the phenotype. Accordingly, it is easier to correlate metabolomic profiles with phenotype compared to genomic, transcriptomic and proteomic profiles. **b** | A schematic of a metabolic pathway. Metabolites are represented as circles, the levels of which can be measured by mass spectrometry. An alteration in a single enzyme can lead to a cascade of metabolic perturbations. Here, metabolites whose levels are altered in each of three theoretical genotypes are shown by coloured circles. Orange, light blue, and green circles represent metabolites altered in genotype 1, genotype 2 and genotype 3, respectively. Metabolites whose levels are similarly altered in multiple genotypes are represented by multi-coloured circles. These shared alterations may be useful in identifying phenotypically important biochemical perturbations.

As these examples highlight, one gene can influence a multitude of metabolic pathways and thereby have a functional role in many cellular processes. Even knowledge of encoded protein structure is often insufficient to infer function at the whole-organism level. Such functions can have intricate regulatory mechanisms involving epigenetic control, post-translational modifications and feedback loops that enable context-dependent activation or deactivation (FIG. 2a). Thus, investigations aimed at detangling the role of any one specific gene benefit from systems-level analyses. These types of global studies were once limited to genes, transcripts and proteins, but technological developments over the past decade now allow for the untargeted profiling of metabolites and provide opportunities to comprehensively track metabolic reactions directly for the first time (FIG. 2b).

Untargeted metabolomic workflow

Although untargeted metabolomic experiments are often hypothesis generating rather than hypothesis driven, it is important to carefully construct an experimental design that maximizes the number of metabolites detected and their quantitative reproducibility. With the workflow that is described below, metabolite identification is a manual and time-intensive process. Thus, the choice of sample type, preparation, chromatographic separation and analytical instrumentation should be considered and the choice that is most likely to yield high-quality data used for analysis. Here, we focus on an LC/MS-based workflow because this technique enables the detection of the highest number of metabolites and requires only

© 2012 Macmillan Publishers Limited. All rights reserved

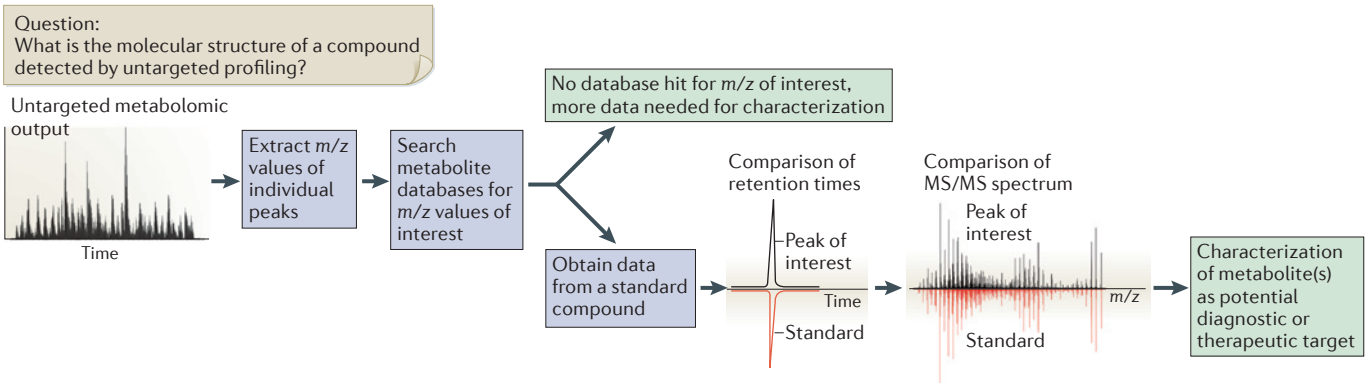

Figure 3 | **Metabolite characterization in the untargeted metabolomic workflow.** In liquid chromatography followed by mass spectrometry (LC/MS)-based untargeted metabolomics, metabolites are structurally characterized on the basis of accurate mass, retention time and tandem mass spectrometry (MS/MS) data according to the workflow shown. First, *m/z* values of interest are searched in metabolite databases. When a hit is returned within the expected error of the mass spectrometer, the retention time and MS/MS data of a standard compound is compared to that from the biological sample. Standard data may be available in a metabolite database such as the Human Metabolome Database and METLIN or generated experimentally.

minimal amounts of sample (for example, typically less than 25 mg of tissue, around 1 million cells or approximately 50 µL of biofluids such as plasma and urine).

Sample preparation and data acquisition. The first step in the untargeted metabolomic workflow is to isolate metabolites from biological samples (FIG. 1b). Several approaches involving sample homogenization and protein precipitation have been used, and these are described in detail elsewhere[14,15,45,46]. Prior to mass spectrometry analysis, isolated metabolites are separated chromatographically by using relatively short solvent gradients (on the order of minutes) that allow for high-throughput analysis of large numbers of samples. The physiochemical landscape of the metabolome is highly heterogeneous, so to increase the number of compounds detected, multiplexed methods for the extraction and separation of metabolites are used[47]. For example, extracting the same cells with both organic and aqueous solvents increases the number of hydrophobic and hydrophilic compounds observed, respectively. Similarly, reversed-phase chromatography is better suited for the separation of hydrophobic metabolites, whereas hydrophilic-interaction chromatography generally separates hydrophilic compounds more effectively. Most frequently, data is collected on a quadrupole time-of-flight (QTOF) mass spectrometer or an Orbitrap mass spectrometer, but other time-of-flight and ion trap instruments can also be used[28,29,48]. Given the challenge of predicting tandem mass spectrometry (MS/MS) fragmentation patterns for most metabolites, untargeted metabolomic profiling data are typically acquired in MS^1 mode (that is, only the mass-to-charge ratio (*m/z*) of the intact metabolite is measured)[49,50], unlike the alternating MS^1 and MS/MS mode used in shotgun omic approaches.

> "our understanding of metabolism is evolving much like our notion of physics evolved in the early twentieth century"

Data analysis. With recent developments in bioinformatic tools, identification of metabolite peaks that are differentially altered between sample groups has become a relatively automated process. Several metabolomic software programs that provide a method for peak picking, nonlinear retention time alignment, visualization, relative quantification and statistical analysis are available[1,51]. The most widely used metabolomic software is XCMS, which is freely available online and allows users to upload data, perform data processing and browse results within a web-based interface[1].

Metabolite identification. It is important to note that the metabolomic software currently available does not output metabolite identifications. Rather, it provides a table of features with *p*-values and fold changes related to their difference in relative intensity between samples. To determine the identity of a feature of interest, the accurate mass of the compound is first searched in metabolite databases such as the Human Metabolome Database and METLIN[52–54]. A database match represents only a putative metabolite assignment that must be confirmed by comparing the retention time and MS/MS data of a model compound to that from the feature of interest in the research sample (FIG. 3). Currently, MS/MS data for features selected from the profiling results are obtained from additional experiments, and matching of MS/MS fragmentation patterns is performed manually by inspection. These additional analyses are time intensive and represent the rate-limiting step of the untargeted metabolomic workflow. Additionally, although metabolite databases have grown considerably over the past decade, a substantial number of metabolite features detected from biological samples do not return any matches. Identification of these unknown features requires *de novo* characterization with traditional methods. Taken together, it should be recognized that comprehensive identification of all metabolite features detected by LC/MS is currently impractical for most samples analysed.

Addressing the challenges

Untargeted metabolomics has revealed that the number of endogenous metabolites in biological systems is larger than anticipated and cannot be accounted for merely by canonical biochemical pathways. That is, the masses of a significant fraction of compounds detected in global analyses do not match any of the masses included in metabolite databases. Therefore, given that the metabolome is not encoded in the genome in the same way as proteins and transcripts, systems-level studies of metabolites are complicated by attempting to analyse an undefined set of molecules. In response to this challenge, metabolite

© 2012 Macmillan Publishers Limited. All rights reserved

databases have expanded rapidly over the past decade. Although database expansion has facilitated untargeted studies, there are still many metabolites for which the chemical structure, cellular function, biochemical pathway and anatomical location remain uncharacterized. Here, innovative technologies and experimental strategies that can be coupled with untargeted profiling are driving progress in the field.

Improving metabolite databases. Over the past decade, the information catalogued in metabolite databases has evolved beyond lists of one-dimensional data that are traditionally acquired by mass spectrometry- and NMR-based screens. The Human Metabolome Database, for example, includes a 'MetaboCard' for each of its included metabolites (~8,550)[52,53]. In addition to having molecular weights and experimental NMR spectra, the MetaboCards list information on each compound's biochemical pathway, concentration, anatomical location, metabolizing enzymes and related disorders when available. Currently, the Human Metabolome Database and METLIN are among the most widely used metabolite databases publicly available[54]. Like the Human Metabolome Database, METLIN contains experimental data for a subset of the total number of compounds included (~45,000), having MS/MS data available for more than 10,000 metabolites. For each of these metabolites, MS/MS data were experimentally generated from model compounds analysed at four different collision energies in both positive and negative mode. When used together with other publicly available tools, the Human Metabolome and METLIN databases can facilitate both metabolite identification and data interpretation.

Meta-analysis: prioritizing unknowns. Alterations in a single enzyme can lead to a cascade of metabolic perturbations that are functionally unrelated to the phenotype of interest. Untargeted metabolomic profiling of a particular disease or mutant can therefore reveal hundreds of alterations that are unlikely to have mechanistic implications. Given the resources needed to identify both known and unknown compounds, strategies to reduce lists of potentially interesting features before committing to identifying them are of great utility. One such strategy is meta-analysis, in which untargeted profiling data from multiple studies are compared (FIG. 2b). By comparing multiple models of a disease, for example, features that are not similarly altered in each of the comparisons may be de-prioritized as being less likely to be related to the shared phenotypic pathology. To automate the comparison of untargeted metabolomic data, freely available software called metaXCMS has been recently developed[55]. As proof of concept, metaXCMS was applied to investigate three pain models of different pathogenic aetiologies: inflammation, acute heat and spontaneous arthritis[56]. Although hundreds of metabolite features were found to be altered in each model, only three were similarly dysregulated among all the groups. One of the shared metabolites was identified as histamine, a well-characterized mediator of pain that works by several mechanisms. The application of similar data-reduction strategies to other biological systems may justify aggressive analytical investigations of unknown features that are likely to be physiologically relevant.

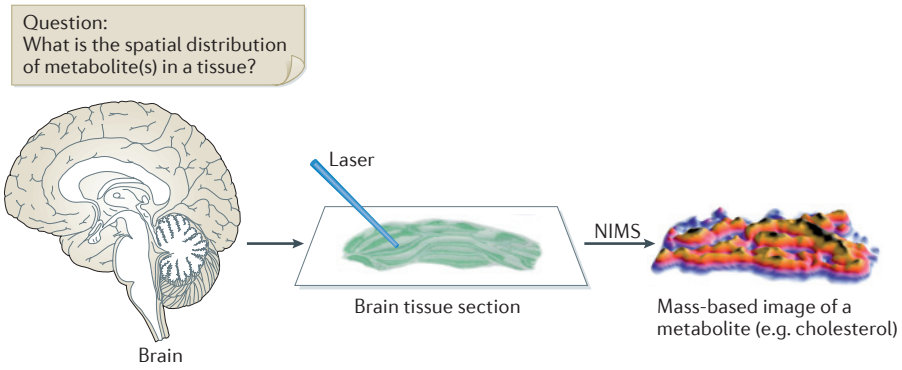

Figure 4 | **Spatial localization of metabolites in tissue by mass spectrometry-based imaging.** An example of a surface-based image of cholesterol from mouse brain obtained by using nanostructure-initiator mass spectrometry (NIMS)[2]. NIMS is well suited for metabolite imaging because it is highly sensitive and does not suffer from matrix interference in the low-mass range. Sections of frozen tissue are first transferred to a NIMS chip that is subsequently analysed by using a laser-induced desorption and ionization approach. By systematically rastering the laser across the tissue, a mass spectrum is generated from each point. The mass spectral intensity of the metabolite of interest is plotted spatially to generate images, as shown here for cholesterol.

Imaging approaches for localizing metabolites. One of the first steps in the untargeted metabolomic workflow applied to biological tissue is metabolite isolation by sample homogenization. Thus, standard metabolic profiling techniques do not permit high-resolution spatial localization of metabolites within samples. Investigations of heterogeneous tissues such as the brain are therefore complicated by the averaging of various cell types, each with a potentially unique metabolome. Given these limitations, correlating a dysregulated metabolite with a specific region of tissue or cell type can be challenging.

NMR-based imaging technologies have been applied to spatially localize metabolites in intact samples, but these methods have limited chemical specificity and sensitivity[57–59]. By contrast, mass spectrometry-based approaches relying on matrix-assisted laser desorption ionization (MALDI) offer improved chemical specificity and sensitivity, but they are limited in their application to metabolites owing to background interference caused by the matrix in the low-mass region that is characteristic of metabolites[60]. As an alternative, a matrix-free technique called nanostructure-initiator mass spectrometry (NIMS) has been developed for the analysis of metabolites with high sensitivity and spatial resolution[61,62] (FIG. 4). By using NIMS to analyse 3 μm sections of brain tissue from mice with impaired cholesterol biosynthesis, metabolic precursors of cholesterol were found to localize to the cerebellum and brainstem[2]. These types of NIMS imaging applications coupled with histology will allow metabolite localization patterns to be correlated with tissue pathology and drive developments in our understanding of chemical physiology.

Untargeted metabolomics applied

Given its sensitivity, high throughput and minimal sample requirements, untargeted metabolomics has wide applicability across a myriad of biological questions. Despite its relatively recent emergence as a global profiling technology, untargeted metabolomics has already increased our understanding of comprehensive cellular metabolism and has been used to address a number of

© 2012 Macmillan Publishers Limited. All rights reserved

Glossary

Imaging mass spectrometry
A surface-based approach in which molecules such as metabolites and proteins are spatially analysed in biological specimens. Common imaging mass spectrometry techniques are matrix-assisted laser desorption ionization (MALDI), secondary ion mass spectrometry (SIMS) and nanostructure-initiator mass spectrometry (NIMS).

Ion trap
A type of mass spectrometer that traps ions by using electric or magnetic fields. Once trapped, the ions are analysed to determine their mass-to-charge ratios. Tandem mass spectrometry can also be performed on selected ions by isolating them in the trap and then subjecting them to dissociation.

Matrix-assisted laser desorption ionization
(MALDI). A surface-based mass spectrometry approach in which analytes are embedded in a chemical matrix that absorbs energy from an ultraviolet laser, resulting in analyte desorption and ionization.

MS^1 mode
The mode of a mass spectrometer in which only the mass-to-charge ratio of the intact ion is measured. In these experiments no tandem mass spectrometry is performed.

Nanostructure-initiator mass spectrometry
(NIMS). A nanostructure surface-based mass spectrometry approach that does not require a matrix. NIMS is commonly used for metabolomic studies and metabolite imaging.

Nuclear magnetic resonance
(NMR). An analytical technique that exploits the different responses to radiofrequency stimuli by chemically distinct atomic nuclei in an external magnetic field to provide information about the structure and dynamics of molecules.

Quadrupole time-of-flight
(QTOF). A mass spectrometer commonly used to perform untargeted metabolomics. By using the quadrupole as a focusing lens, the time-of-flight mass analyser can be used to acquire profiling data. Alternatively, the quadrupole can be used to select ions for tandem mass spectrometry experiments.

Tandem mass spectrometry
(MS/MS). A type of mass spectrometry in which ions are selectively isolated and then fragmented. The mass-to-charge ratio of each molecular fragment is measured and used for structural characterization.

Triple quadrupole
(QqQ). A type of mass spectrometer that is often used for targeted studies owing to its sensitivity and specificity. The QqQ analyser consists of a quadrupole, a quadrupole collision cell and a second quadrupole, respectively. The first selects and analyses ions of interest, the second is used as a collision cell for fragmentation and the third analyses fragments.

biomedical issues. In particular, it has been useful in identifying altered metabolic pathways in disease that represent novel drug targets: an evolving biomedical application referred to as 'therapeutic metabolomics' (REF. 63). An example of this application is the discovery of increased levels of the metabolite 2-hydroxyglutarate in cancer cells with isocitrate dehydrogenase 1 mutations, which are a common feature of a major subset of primary human brain cancers[64]. These results suggest that inhibition of 2-hydroxyglutarate production may be an effective therapeutic approach to slow or halt conversion of a low-grade glioma into lethal secondary glioblastoma. In another example, levels of the sphingolipid dimethylsphingosine were found to be increased in the spinal cords of rats suffering from neuropathic pain[65]. Increased levels of dimethylsphingosine were determined to induce pain-like behaviour *in vivo* and point to the inhibition of methyltransferase or ceramidase as potential therapeutic approaches for treating chronic pain by blocking dimethylsphingosine production.

Another area in which untargeted metabolomics has been successfully applied is in characterizing gene and protein function. In addition to successfully identifying the function of unknown genes and proteins, untargeted profiling has been applied to discover new functions for known genes and proteins. By screening for metabolites that accumulate after gene mutation or enzyme inhibition, unanticipated connections between the proteome and metabolome have been established that were not accurately predicted from *in vitro* activity measurements[66]. For example, to characterize a yeast gene of unknown function (*YKL215C*), untargeted methods were applied to organisms harbouring a mutation in *YKL215C*. Increased levels of 5-oxoproline were detected in these organisms, allowing the assignment of YKL215C as an oxoprolinase[48]. In another example, an untargeted screen identified a previously unidentified activity for the yeast enzyme sedoheptulose-1,7-bisphosphatase. The finding that sedoheptulose-1,7-biosphosphatase hydrolyses sedoheptulose-1,7-bisphosphate to sedoheptulose-7-phosphate identified a thermodynamically driven route from trioses produced by glycolysis to the synthesis of ribose[67]. A similar type of enzyme-activity characterization was also accomplished for *Mycobacterium tuberculosis* by incubating a purified recombinant enzyme with a mycobacterial small molecule extract. The small molecule extract was analysed by LC/MS for altered substrate and product, resulting in the incompletely characterized protein Rv1248c being assigned as a 2-hydroxy-3-oxoadipate synthase[68]. As these examples highlight, untargeted metabolomics has implications not only for therapeutic screening but also for providing chemical insight across a broad area of mechanistic cell biology.

Concluding remarks

Although there has been a long-standing interest in metabolic profiling, only recently have technologies emerged that enable the global analysis of metabolites at a systems level, comparable to its omic predecessors. Unlike genomics, transcriptomics and proteomics, however, metabolomics provides a tool for measuring biochemical activity directly by monitoring the substrates and products transformed during cellular metabolism. Untargeted profiling of these chemical transformations at a global level serves as a phenotypic readout that can be used effectively in diagnosing pathologies, identifying therapeutic targets of disease and investigating the mechanisms of fundamental biological processes.

Although untargeted metabolomics is still in its infancy, early studies have shown that the complexity of cellular metabolism exceeds that expected on the basis of classical biochemical pathways. In this sense, our understanding of metabolism is evolving much like our notion of physics evolved in the early twentieth century with the emergence of experimental results such as the photoelectric effect, which could not be explained by Newtonian laws[69,70]. Ultimately, the ideas that emerged from this disparity resulted in a new set of principles for understanding physical phenomena known as quantum mechanics. As metabolomic technologies continue to advance and facilitate the characterization of unknown pathways, the potential of untargeted metabolomics to shape our understanding of global metabolism is yet to be fully realized.

Gary J. Patti is at the Departments of Chemistry, Genetics and Medicine, Washington University School of Medicine, St. Louis, Missouri 63110, USA.

Oscar Yanes is at The Metabolomics Platform of the Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders, University of Rovira i Virgili, Tarragona, Spain.

Gary Siuzdak is at The Center for Metabolomics and Mass Spectrometry, Departments of Chemistry and Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

Correspondence to G.J.P. and G.S.
e-mails: gjpattij@wustl.edu; siuzdak@scripps.edu

doi:10.1038/nrm3314

1. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
2. Patti, G. J. *et al.* Nanostructure-initiator mass spectrometry (NIMS) imaging of brain cholesterol metabolites in Smith-Lemli-Opitz syndrome. *Neuroscience* **170**, 858–864 (2010).

© 2012 Macmillan Publishers Limited. All rights reserved

3. Koizumi, S. *et al.* Imaging mass spectrometry revealed the production of lyso-phosphatidylcholine in the injured ischemic rat brain. *Neuroscience* **168**, 219–225 (2010).
4. Baker, M. Metabolomics: from small molecules to big ideas. *Nature Methods* **8**, 117–121 (2011).
5. Kind, T., Scholz, M. & Fiehn, O. How large is the metabolome? A critical analysis of data exchange practices in chemistry. *PLoS ONE* **4**, e5440 (2009).
6. Dudley, E., Yousef, M., Wang, Y. & Griffiths, W. J. Targeted metabolomics and mass spectrometry. *Adv. Protein Chem. Struct. Biol.* **80**, 45–83 (2010).
7. Nicholson, J. K., Connelly, J., Lindon, J. C. & Holmes, E. Metabonomics: a platform for studying drug toxicity and gene function. *Nature Rev. Drug Discov.* **1**, 153–161 (2002).
8. Dirren, H., Robinson, A. B. & Pauling, L. Sex-related patterns in the profiles of human urinary amino acids. *Clin. Chem.* **21**, 1970–1975 (1975).
9. Davison, J. M. & Cheyne, G. A. History of the measurement of glucose in urine: a cautionary tale. *Med. Hist.* **18**, 194–197 (1974).
10. Robinson, A. B. & Robinson, N. E. Origins of metabolic profiling. *Methods Mol. Biol.* **708**, 1–23 (2011).
11. Dalgliesh, C. E., Horning, E. C., Horning, M. G., Knox, K. L. & Yarger, K. A gas-liquid-chromatographic procedure for separating a wide range of metabolites occuring in urine or tissue extracts. *Biochem. J.* **101**, 792–810 (1966).
12. Eknoyan, G. Santorio Sanctorius (1561–1636) — founding father of metabolic balance studies. *Am. J. Nephrol.* **19**, 226–233 (1999).
13. Astarita, G., Ahmed, F. & Piomelli, D. Lipidomic analysis of biological samples by liquid chromatography coupled to mass spectrometry. *Methods Mol. Biol.* **579**, 201–219 (2009).
14. Want, E. J. *et al.* Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal. Chem.* **78**, 743–752 (2006).
15. Yanes, O., Tautenhahn, R., Patti, G. J. & Siuzdak, G. Expanding coverage of the metabolome for global metabolite profiling. *Anal. Chem.* **83**, 2152–2161 (2011).
16. Pike, L. S., Smift, A. L., Croteau, N. J., Ferrick, D. A. & Wu, M. Inhibition of fatty acid oxidation by etomoxir impairs NADPH production and increases reactive oxygen species resulting in ATP depletion and cell death in human glioblastoma cells. *Biochim. Biophys. Acta* **1807**, 726–734 (2011).
17. Buxton, D. B., Schwaiger, M., Nguyen, A., Phelps, M. E. & Schelbert, H. R. Radiolabeled acetate as a tracer of myocardial tricarboxylic acid cycle flux. *Circ. Res.* **63**, 628–634 (1988).
18. Bajad, S. U. *et al.* Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* **1125**, 76–88 (2006).
19. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chem. Biol.* **5**, 593–599 (2009).
20. Buescher, J. M., Moco, S., Sauer, U. & Zamboni, N. Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Anal. Chem.* **82**, 4403–4412 (2010).
21. Sabatine, M. S. *et al.* Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* **112**, 3868–3875 (2005).
22. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nature Med.* **17**, 448–453 (2011).
23. Shaham, O. *et al.* Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. *Mol. Syst. Biol.* **4**, 214 (2008).
24. Barton, R. H., Nicholson, J. K., Elliott, P. & Holmes, E. High-throughput ^{1}H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *Int. J. Epidemiol.* **37** (Suppl. 1), I31–I40 (2008).
25. Buscher, J. M., Czernik, D., Ewald, J. C., Sauer, U. & Zamboni, N. Cross-platform comparison of methods for quantitative metabolomics of primary metabolism. *Anal. Chem.* **81**, 2135–2143 (2009).
26. Jonsson, P. *et al.* Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. *Analyst* **130**, 701–707 (2005).
27. Nordstrom, A., Want, E., Northen, T., Lehtio, J. & Siuzdak, G. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. *Anal. Chem.* **80**, 421–429 (2008).
28. Wikoff, W. R. *et al.* Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl Acad. Sci. USA* **106**, 3698–3703 (2009).
29. Yanes, O. *et al.* Metabolic oxidation regulates embryonic stem cell differentiation. *Nature Chem. Biol.* **6**, 411–417 (2010).
30. Baran, R. *et al.* MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics* **7**, 530 (2006).
31. Lommen, A. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal. Chem.* **81**, 3079–3086 (2009).
32. Katajamaa, M., Miettinen, J. & Oresic, M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **22**, 634–636 (2006).
33. Katajamaa, M. & Oresic, M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* **6**, 179 (2005).
34. Tautenhahn, R., Bottcher, C. & Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**, 504 (2008).
35. Beadle, G. W. & Tatum, E. L. Genetic control of biochemical reactions in neurospora. *Proc. Natl Acad. Sci. USA* **27**, 499–506 (1941).
36. Tatum, E. L. & Beadle, G. W. Genetic control of biochemical reactions in neurospora: an "Aminobenzoicless" mutant. *Proc. Natl Acad. Sci. USA* **28**, 234–243 (1942).
37. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54–60 (2011).
38. Nicholson, G. *et al.* Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol. Syst. Biol.* **7**, 525 (2011).
39. Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics* **2**, 155–168 (2001).
40. Trauger, S. A. *et al.* Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. *J. Proteome Res.* **7**, 1027–1035 (2008).
41. Kirschner, M. W. The meaning of systems biology. *Cell* **121**, 503–504 (2005).
42. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
43. Dong, M. Q. *et al.* Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*. *Science* **317**, 660–663 (2007).
44. Kang, S., Bader, A. G. & Vogt, P. K. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proc. Natl Acad. Sci. USA* **102**, 802–807 (2005).
45. Villas-Boas, S. G., Hojer-Pedersen, J., Akesson, M., Smedsgaard, J. & Nielsen, J. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **22**, 1155–1169 (2005).
46. Wittmann, C., Kromer, J. O., Kiefer, P., Binz, T. & Heinzle, E. Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. *Anal. Biochem.* **327**, 135–139 (2004).
47. Patti, G. J. Separation strategies for untargeted metabolomics. *J. Sep. Sci.* **34**, 3460–3469 (2011).
48. Lu, W. *et al.* Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Anal. Chem.* **82**, 3212–3221 (2010).
49. Han, X. & Gross, R. W. 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 (2005).
50. Wolters, D. A., Washburn, M. P. & Yates, J. R. 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**, 5683–5690 (2001).
51. Melamud, E., Vastag, L. & Rabinowitz, J. D. Metabolomic analysis and visualization engine for LC-MS data. *Anal. Chem.* **82**, 9818–9826 (2010).
52. Wishart, D. S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37**, D603–D610 (2009).
53. Forsythe, I. J. & Wishart, D. S. Exploring human metabolites using the human metabolome database. *Curr. Protoc. Bioinformatics* **Ch. 14**, Unit 14.8 (2009).
54. Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747–751 (2005).
55. Patti, G. J., Tautenhahn, R. & Siuzdak, G. Meta-analysis of untargeted metabolomic data from multiple profiling experiments. *Nature Protoc.* **7**, 508–516 (2012).
56. Tautenhahn, R. *et al.* metaXCMS: second-order analysis of untargeted metabolomics data. *Anal. Chem.* **83**, 696–700 (2011).
57. Bottomley, P. A., Hardy, C. J. & Roemer, P. B. Phosphate metabolite imaging and concentration measurements in human heart by nuclear magnetic resonance. *Magn. Reson. Med.* **14**, 425–434 (1990).
58. Golman, K., Zandt, R. I., Lerche, M., Pehrson, R. & Ardenkjaer-Larsen, J. H. Metabolic imaging by hyperpolarized 13C magnetic resonance imaging for *in vivo* tumor diagnosis. *Cancer Res.* **66**, 10855–10860 (2006).
59. Cakmakci, H., Pekcevik, Y., Yis, U., Unalp, A. & Kurul, S. Diagnostic value of proton MR spectroscopy and diffusion-weighted MR imaging in childhood inherited neurometabolic brain diseases and review of the literature. *Eur. J. Radiol.* **74**, e161–e171 (2010).
60. Greving, M. P., Patti, G. J. & Siuzdak, G. Nanostructure-initiator mass spectrometry metabolite analysis and imaging. *Anal. Chem.* **83**, 2–7 (2011).
61. Northen, T. R. *et al.* Clathrate nanostructures for mass spectrometry. *Nature* **449**, 1033–1036 (2007).
62. Woo, H. K., Northen, T. R., Yanes, O. & Siuzdak, G. Nanostructure-initiator mass spectrometry: a protocol for preparing and applying NIMS surfaces for high-sensitivity mass analysis. *Nature Protoc.* **3**, 1341–1349 (2008).
63. Rabinowitz, J. D., Purdy, J. G., Vastag, L., Shenk, T. & Koyuncu, E. Metabolomics in drug target discovery. *Cold Spring Harb. Symp. Quant. Biol.* 23 Nov 2011 (doi:10.1101/sqb.2011.76.010694).
64. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
65. Patti, G. J. *et al.* Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin. *Nature Chem. Biol.* **8**, 232–234 (2012).
66. Saghatelian, A. *et al.* Assignment of endogenous substrates to enzymes by global metabolite profiling. *Biochemistry* **43**, 14332–14339 (2004).
67. Clasquin, M. F. *et al.* Riboneogenesis in yeast. *Cell* **145**, 969–980 (2011).
68. de Carvalho, L. P. *et al.* Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase. *Chem. Biol.* **17**, 323–332 (2010).
69. Feynman, R. P., Leighton, R. B. & Sands, M. L. *The Feynman Lectures on Physics* (Addison-Wesley Publishing Company, Boston, 1963).
70. Segre, E. *From X-rays to Quarks: Modern Physicists and Their Discoveries* (Dover Publications, New York, 2007).

Acknowledgements
The authors gratefully acknowledge that their work is supported by the California Institute of Regenerative Medicine (grant TR1-01219; G.S.), the US National Institutes of Health (grants R24 EY017540-04, P30 MH062261-10 and P01 DA026146-02; G.S.), the US Department of Energy (grants FG02-07ER64325 and DE-AC0205CH11231; G.S.), the US National Institutes of Health/National Institute on Aging (L30 AG0 038036; G.J.P.) and start-up funds from Washington University (G.J.P.).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Gary J. Patti's homepage: http://pattilab.wustl.edu/
Oscar Yanes' homepage: http://www.yaneslab.com/
Gary Siuzdak's homepage: http://masspec.scripps.edu/
Human Metabolome Database: http://www.hmdb.ca/
METLIN: http://metlin.scripps.edu/
XCMS (bioinformatic software for analysing untargeted LC/MS-based metabolomic data): https://xcmsonline.scripps.edu/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2012 Macmillan Publishers Limited. All rights reserved