Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database

Zheng-Jiang Zhu[1], Andrew W Schultz[1], Junhua Wang[1], Caroline H Johnson[1], Steven M Yannone[2], Gary J Patti[3–5] & Gary Siuzdak[1]

[1]Scripps Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, La Jolla, California, USA. [2]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. [3]Department of Chemistry, Washington University, St. Louis, Missouri, USA. [4]Department of Genetics, Washington University, St. Louis, Missouri, USA. [5]Department of Medicine, Washington University, St. Louis, Missouri, USA. Correspondence should be addressed to G.J.P. (gjpattij@wustl.edu) or G.S. (siuzdak@scripps.edu).

Published online 7 February 2013; doi:10.1038/nprot.2013.004

Untargeted metabolomics provides a comprehensive platform for identifying metabolites whose levels are altered between two or more populations. By using liquid chromatography quadrupole time-of-flight mass spectrometry (LC-Q-TOF-MS), hundreds to thousands of peaks with a unique *m/z* ratio and retention time are routinely detected from most biological samples in an untargeted profiling experiment. Each peak, termed a metabolomic feature, can be characterized on the basis of its accurate mass, retention time and tandem mass spectral fragmentation pattern. Here a seven-step protocol is suggested for such a characterization by using the METLIN metabolite database. The protocol starts from untargeted metabolomic LC-Q-TOF-MS data that have been analyzed with the bioinformatics program XCMS, and it describes a strategy for selecting interesting features as well as performing subsequent targeted tandem MS. The seven steps described will require 2–4 h to complete per feature, depending on the compound.

INTRODUCTION

Metabolomics has emerged as a powerful technique for understanding the small-molecule basis of biological processes such as those associated with disease pathogenesis[1,2], interactions of microbial communities[3], microbial biochemistry[4,5], plant physiology[6], drug mode of action[7] and metabolism[8]. In general, there are two technological platforms used to perform metabolomics, which involve either nuclear magnetic resonance (NMR) spectroscopy[9,10] or MS[11,12]. Although NMR provides unique structural information about metabolites, it suffers from limitations in sensitivity and chemical resolution. In contrast, MS provides less-conclusive structural information, but given its sensitivity and large dynamic range, it allows for the detection of many more chemical species in a single experiment. Each of these technologies has been successfully applied to systematically studying metabolites; however, MS methods are more commonly used for comprehensive investigations that are global in scope. The strength of MS-based metabolomics is best realized when coupled to a chromatographic technique, such as capillary electrophoresis, gas chromatography (GC) or LC, the latter two being the most popular. GC/MS-based metabolomics is a robust, well-established technique[13–15]. Because of the reproducibility of the chromatography, retention time can be paired with the electron impact–derived fragmentation spectra[16] and compared against the National Institute of Standards and Technology (NIST)[17] or Fiehn metabolomic[18] databases to make identifications. However, the majority of metabolites must be derivatized to make them more volatile and more thermally stable, which introduces a source of error and complicates identification[19].

In the past decade, LC/MS-based analysis has moved to the forefront because of its ability to analyze and identify underivatized and thermally labile metabolites. In contrast with electron impact (EI), electrospray ionization (ESI)[20] (and, to a lesser extent, atmospheric pressure chemical ionization[20]) is a soft mechanism for ionizing molecules, leaving the molecular ion intact. There are two major approaches to LC/MS-based metabolomic experiments: the targeted[21–23] and untargeted[24–28] analysis. In untargeted metabolomics, one tries to observe as many unknown and known metabolic peaks as possible, comparing the ion intensity between the same peaks present in two or more groups of samples. The disadvantage of this technique is that it is not optimized for a specific metabolite and is less quantitative. The advantage is that it provides an opportunity to observe a large number of known and unknown metabolites, which may provide novel insights into a biological system[3,5,29]. Coupled to a high-resolution mass spectrometer[30], such as a TOF[22,31], Orbitrap[32,33] or a Fourier transform–ion cyclotron resonance (FT-ICR)[34] instrument, high mass accuracy can be obtained. This can greatly reduce the number of potential molecular formulas corresponding to one metabolic peak, but there may still be several possible molecular formulas that are appropriate for the accurate mass data (depending on the resolution of the instrument), and numerous potential isomers for each molecular formula. More structural information can be obtained by examining the fragmentation pattern. Combining the high-resolution precursor ion with data from a fragmentation mechanism (obtained by MS/MS) reduces the number of possible metabolites to a single structure or a narrow set of structures (see limitations below). When searching against a metabolite database—in the case of this protocol, the METLIN database[35] (http://metlin.scripps.edu)—it is therefore best to match both the accurate mass and the fragmentation data (MS/MS spectra) for each metabolite peak. Retention times, relative to other metabolites of known identity and similar structural class, also support the structural determination. This protocol describes an

approach to provide rigorous characterization of metabolites from LC/MS-based metabolomic data.

Q-TOF–based characterization of metabolites

In this protocol, metabolites are characterized using an LC-Q-TOF instrument in combination with the METLIN database[35] (http://metlin.scripps.edu). The Q-TOF provides the ability to collect both high-resolution precursor and fragmentation data, facilitating the characterization of metabolites. When used in conjunction with the METLIN database, which provides the user with the ability to search for the precursor ion, its fragments and neutral losses, the characterization of metabolites is highly augmented. METLIN is the largest curated database of high-resolution tandem mass spectra, covering over 10,000 metabolites. The fragmentation spectra are essential for the elucidation and confirmation of metabolites. Matching the retention time and fragmentation of a metabolite with those of an authentic standard can confirm its identity. One of the advantages of tandem MS in the Q-TOF is that collision energies can be adjusted to enhance or decrease the degree of fragmentation, thereby revealing more information about the metabolite. Some metabolites, however, do not fragment well or fragment poorly when an adduct (e.g., Na^+) is present. The adduct stabilizes the ion and can give limited fragmentation, but trying different ionization strategies or solvent mixtures can ameliorate this.

Untargeted metabolomics begins with an initial profiling experiment, often in which two or more sample groups are profiled via LC-MS and statistically compared, with only the dysregulated metabolites being characterized[2–4,6,29,36]. There are a few exceptions in which only one sample group is analyzed in studies characterizing as many metabolites as possible in one biofluid[37,38]. Two excellent protocols are available for LC-MS profiling experiments in urine[39] and in plasma and serum[14]. These protocols can be easily adjusted to other sample types. The key to obtaining good results is to carefully design the experiments so that there are enough biological replicates to make the results statistically significant (i.e., they must not be underpowered). Appropriate power calculations must be carried out first to determine the sample size that will have a statistically significant effect[40]. There are a number of factors that need to be considered, such as biological variation, sample preparation and others; these are discussed in more detail by Brown *et al.*[41]. Depending on the biological variability of the system, we recommend that the minimal numbers of each sample group be four to six for cell culture, six to eight for animals and ten or more for humans. After analysis of the initial profiling data by using a peak alignment and statistical analysis package, such as XCMS[42] or XCMSOnline[43], a list of dysregulated metabolic peaks with a retention time and *m/z* will be generated. The protocol reported here is for the systematic analysis of the dysregulated features on this list. The stages of the procedure are as follows: (i) determine the adduct and charge of a metabolite feature of interest; (ii) inspect MS data to determine whether a peak is real and of sufficient intensity for MS/MS; (iii) perform targeted MS/MS; (iv) search precursor in METLIN; (v) search MS/MS in METLIN; (vi) compare experimental MS/MS with METLIN; and (vii) verify that the characterization is correct using a standard.

Limitations of this approach

Many of the limitations listed below can be mitigated using specialized MS techniques, and thus may not impose real challenges. It is, however, very important to consider these points when carrying out general metabolomic approaches and before optimizing methods for specific chemical species or in response to specific problems.

First, low-abundance ions can be hard to identify if the precursor ion intensity is low (generally below 5,000 counts for an Agilent Q-TOF), making it difficult to obtain the high-quality fragment spectra needed to support a structural assignment. This is not, however, a problem for many peaks, and examples of high-sensitivity MS-based metabolite identifications include 3.5 fmol of dimethylsphingosine (DMS) per mg of dorsal horn[36], or an upper-attomolar range in the analysis of *Methylobacterium extroquens* AM1 (ref. 44).

Second, MS-based analysis provides little, if any, information about the stereochemistry of the metabolites identified and is often insufficient to determine the positions of double bonds in acyl tails. Some specialized techniques have been developed to overcome this problem and have involved the use of ion mobility[45], the addition of Li^+ with multiple rounds of fragmentation[46] and ozone-induced dissociation[47]. The location of these bonds may be important; for example, isobaric ω-3 or ω-6 isomers of a lipid can have markedly different biological roles[48].

Third, isobaric species that co-elute will provide a convoluted mass spectrum, making it difficult to characterize either species. MS is prone to ion suppression[49]; therefore, co-eluted species also affect the quantification of molecules and reduce the ability to observe ions that are less capable of ionization in the presence of an interfering metabolite. Furthermore, isobaric and other species with very similar masses could be fragmented together if not well isolated, thus introducing contamination into the MS/MS spectra and hindering characterization, possibly leading to false negatives. Appropriate chromatographic methods can be developed, which can help resolve different species and reduce some issues with ion suppression. Ion mobility can also aid in the separation of isobaric species in gas phase, which reduces contamination of MS/MS spectra.

Fourth, in-source fragmentation is sometimes observed for species containing a labile group. It can generate one or more abundant fragments that show a similar level of dysregulation compared with other peaks at the same retention time[44]. If two or more dysregulated peaks co-elute, one must ensure that the peaks are not fragments from the same molecule. In **Supplementary Figure 1**, an example of this is shown in which two species (*m/z* of 339.2892 and *m/z* 480.3084) with the same retention time are observed to be dysregulated. The peak 480.3085 corresponds to a lysoPE(18:1/0:0), whereas 339.2892 is a major fragment of this lysoPE, a dehydrated oleoyl (18:1) glycerol. Without recognizing that the lysoPE is the dysregulated metabolite, one may falsely identify the in-source fragment, oleoyl glycerol, as a dysregulated metabolite.

In addition, new tandem MS techniques, such as MS^E (from Waters) and SWATH (from AB Sciex), have recently emerged. MS/MS data acquired from MS^E and SWATH techniques have not yet been tested with METLIN MS/MS spectral comparison.

Finally, this approach does not provide an unequivocal identification of a metabolite. It does, however, provide a higher level of confidence than high-resolution mass alone. To quantitatively evaluate the confidence of metabolite identification, a scoring system is under development. For better confidence, standards should be acquired and run on the same instrument

with the same instrument parameters. The retention time and fragmentation patterns must then match between the sample and the standard to extend the Q-TOF–based characterization to identification, and if the retention time does not match it implies that the characterization is incorrect. For metabolites in which a higher level of confidence is needed, an orthogonal method should also be used to validate the metabolite structure. NMR, for example, has the benefit of structural identification and accurate characterization; furthermore, when coupled to LC, it can be highly effective for metabolite elucidation[50]. Metabolites lacking commercial standards should be chemically synthesized and compared as above[5,51]. For some experiments, this level of rigor may be unnecessary, depending on the scope of the biological question[2].

MATERIALS

REAGENTS

- Acetonitrile with 0.1% (vol/vol) formic acid (Honeywell B&J brand, LC-MS grade) **! CAUTION** Acetonitrile is highly flammable.
- Water with 0.1% (vol/vol) formic acid (Honeywell B&J brand, LC-MS grade)
- Extracted samples from biofluids, yeast, cells or animal tissues in autosampler vials (Sample extraction methods have been extensively reported in the previous literature[28,36,52])

EQUIPMENT

- LC-Q-TOF system: ultraperformance liquid chromatography (UPLC) or LC system; Q-TOF mass spectrometer; column (C18, HILIC and so on) used in initial profiling experiment
- Instrument method from MS-profiling experiment
- A personal computer with an Internet connection and a web browser
- XCMS output spreadsheet from an MS profiling experiment (extracted sample analyzed using the LC-Q-TOF system; see Equipment Setup for more detail)
- Spectral files from the original profiling experiment
- Software for mass spectral analysis, provided by instrument vendor (e.g., Agilent MassHunter, AB Sciex PeakView, Bruker Compass, Waters MassLynx)

EQUIPMENT SETUP

LC-MS instrument setup This protocol is mainly based on using an Agilent 1200 series HPLC system coupled to an Agilent 6538 Q-TOF-MS with Agilent MassHunter (Version B.04.00) and XCMSOnline software (version 1.21.1). There are many other hardware and software combinations that can be used with METLIN; check the instrumentation and software documentation for assistance. To ensure a high level of mass accuracy, the instrument should be calibrated before running the samples according to the manufacturer's guidelines. Ensure that samples are properly mixed and thawed before placing them in an autosampler tray. Install mobile phases, prime system pump and tubing. Install the column and ensure that it is properly equilibrated before injecting the samples.

XCMS output spreadsheet For the analysis of untargeted mass spectrometric data, we recommended using XCMSOnline software (https://xcmsonline.scripps.edu), which can process and analyze data from Agilent, AB Sciex, Bruker, Thermo Fisher and Waters hardware. The file formats of these platforms can be seen at https://xcmsonline.scripps.edu/docs/fileformats.html, along with notes on how to convert the files into the appropriate formats. The user manual for XCMSOnline can be found at https://xcmsonline.scripps.edu/docs/usermanual.pdf, and related information can also be found in a recent publication[43].

PROCEDURE

Stage 1: Determine adduct and charge of a metabolite feature of interest

▲ **CRITICAL** The total ion chromatogram (TIC) and extracted ion chromatograms (EIC or XIC) should be retrieved from the spectral files from the original profiling experiment. This can be done through the data analysis software provided by the instrument vendor. Each instrument vendor has its own software, and each offers similar functions for retrieving the TIC and EICs. Here we used Agilent MassHunter as an example to demonstrate this stage of the procedure.

1| Pick peaks of interest from the XCMS output spreadsheet (see Equipment Setup).

2| By using MassHunter, open the spectral file for a sample and search for the peak of interest by retention time and accurate mass.

3| In MassHunter, select *File* → *Open Data File* to select the data to analyze. The TIC should be displayed as in **Figure 1a**.

4| Select *Chromatograms* → *Extract Chromatograms*. In Type, select EIC.

5| On the MS Chromatogram tab, set the MS level to MS; for *m/z* value(s), type in your value.

6| On the Advanced tab, define the single *m/z* expansion to a symmetric parts per million (p.p.m.) value. For this example, 496.3409, ±20 ppm was used. Click OK. The EIC should appear as in **Figure 1b**, and a peak with an appropriate retention time (RT) for your peak of interest should be visible. The EIC will also display other species with very similar *m/z*, indicating isobaric species that may be present.

7| With the Walk Chromatogram cursor selected, click on the EIC at the retention time of your peak of interest. The MS spectrum will appear.

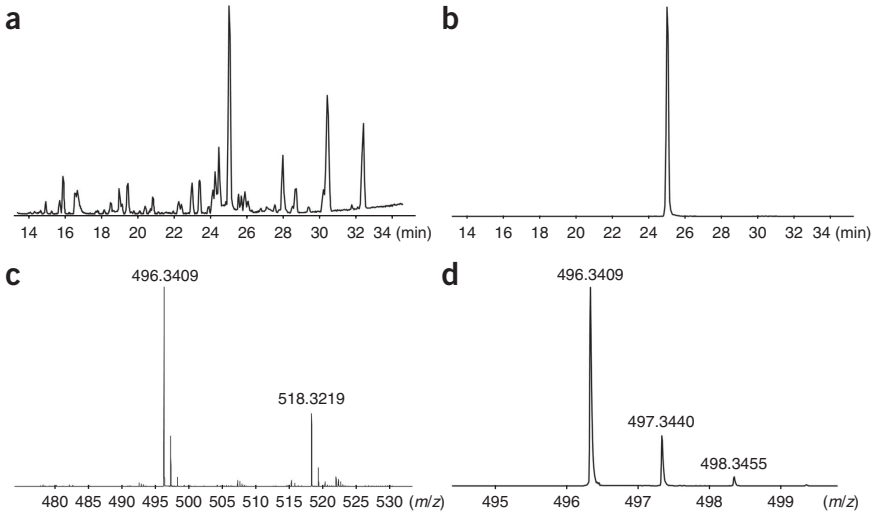


Figure 1 | Determination of monoisotopic peak, charge state and adduct of the precursor ion. (**a**) The TIC for a represenative sample. (**b**) The EIC showing one peak at *m/z* 496.3409. (**c**) The mass spectrum at 24.5 min, scaled to highlight the peaks at *m/z* 496.3409 and 518.3219, which represents the protiated ($[M + H]^+$) and sodiated ($[M + Na]^+$) species for the same metabolomic feature. (**d**) Zooming in further on the peak at 496.3409 reveals a series of isotope peaks of M + 1 and M + 2.

8| By using the Range Select cursor, zoom in on the MS spectra as in **Figure 1c**. Determine the adduct of your peak. In this case, 496.3409 is likely $[M + H]^+$, as a peak of ~22 Da (518.3219) is present, which would correspond to the $[M + Na]^+$.

9| Zoom in further on the MS spectrum (**Fig. 1d**) and determine the charge for the peak. As there is a series of isotope peaks ~1 Da larger after the most intense peak, it is singly charged. Subtracting the proton provides the neutral mass for this species of 495.3336.

? TROUBLESHOOTING

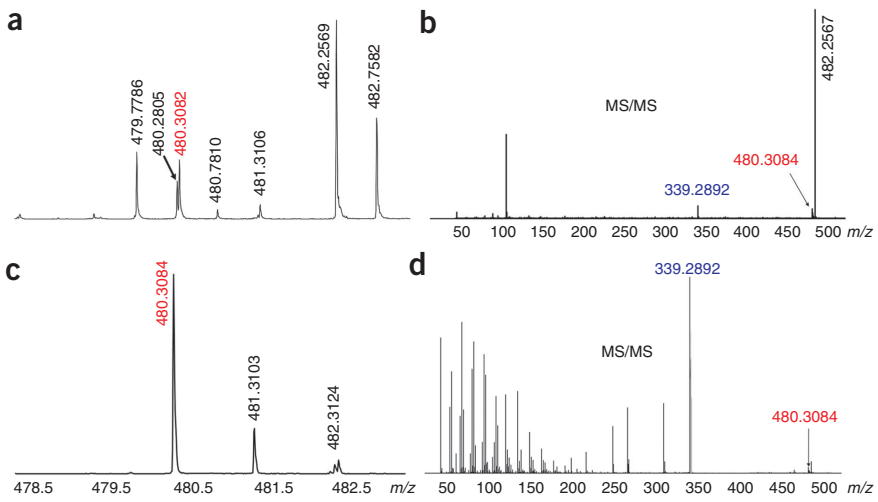
Stage 2: Inspect the MS data to determine whether the peak is real and of sufficient intensity

10| Look for co-eluting ions within 1–2 *m/z* of the peak of interest in the MS spectra, as these may have convoluted the fragment spectra. In **Figure 2**, a group of peaks is observed in which the separation is insufficient. Several species, such as *m/z* 480.2805, *m/z* 480.3082 and *m/z* 482.2569, are not resolved and will fragment together, creating convoluted MS/MS spectra (**Fig. 2b**). Once the species *m/z* 480.3082 is fully resolved by chromatography (**Fig. 2c**), the generated MS/MS spectrum shows good spectral purity. In addition to achieving high-quality MS/MS spectra, the feature of interest should have an intensity greater than 5,000 (for an Agilent Q-TOF). The intensity requirement is empirical. Other Q-TOF instruments from different vendors may have different intensity requirements. The parent ion intensity is required to ensure that the MS/MS spectra have sufficient signal-to-noise ratios (S/N). If the peak is not pure (i.e., with co-eluting species within 1–2 *m/z*) or intense enough, it will be difficult to obtain good MS/MS spectra and thus a meaningful characterization. All examined features with good chromatographic resolution and peak intensities can be grouped for the MS/MS experiments in Stage 3.

? TROUBLESHOOTING

Stage 3: Perform targeted MS/MS

▲ CRITICAL The purpose of this section is to perform targeted MS/MS for the list of features with acceptable chromatographic resolution and peak intensity as discussed in Stage 2. Various instruments have different ways to perform targeted MS/MS experiments. Here we used the Agilent Q-TOF as an example.



11| In MassHunter software, open the instrument method used to collect the original MS profiling data.

Figure 2 | Insufficient chromatographic resolution of a species can lead to overlapping peaks that produce convoluted MS/MS spectra. (**a**) Insufficent resolution of the species *m/z* 480.3082 from other components in a sample provides several overlapping peaks. (**b**) Fragmentation of the unresolved species (*m/z* 480.3082) from panel **a** results in a convoluted spectrum containing at least two species, *m/z* 480.3084 and *m/z* 482.2567. (**c**) Chromatographic resolution of the species *m/z* 480.3084. (**d**) Fragmentation of a resolved spectrum of *m/z* 480.3084 from panel **c** allows for the characterization of lysoPE(18:1/0:0).

12| Under the Q-TOF tab, click on the tab for targeted MS/MS.

13| Input the *m/z* value of the feature, set an RT window of at least 1 min and set isolation to medium, unless co-eluting species dictate a narrower window. More than one feature may be programmed as needed.

14| Save this method, and then inject and analyze the sample with the new method. The collected data will be used in Stage 5.

Stage 4: Search precursor in METLIN

15| In your web browser, open METLIN (http://metlin.scripps.edu). In Search, select Simple.

16| In the mass widow, input the accurate mass value of the parent ion (**Fig. 3**).

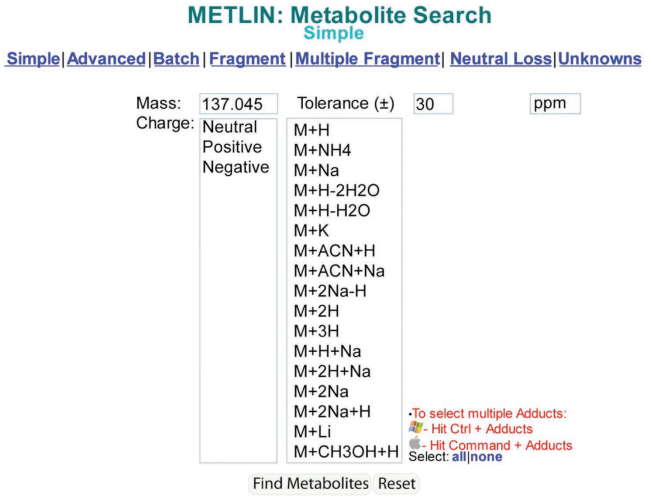


Figure 3 | Screenshot of metabolite search in METLIN. The simple metabolite search panel, with 137.045 inputed and M + H selected as the adduct.

17| Select the charge and adducts determined in Stage 1 (Steps 8 and 9).

18| The default and maximum tolerance of 30 ppm is generally acceptable for Q-TOF experiments; adjust the parameters as appropriate for your specific mass spectrometer. Generally, it is best to use a slightly wider window than the theoretical tolerance for an instrument.

19| Click on the 'Find Metabolites' button.

? TROUBLESHOOTING

Stage 5: Search MS/MS in METLIN

20| Open the newly created MS/MS data file in Agilent MassHunter. To examine the MS/MS spectra, select *Chromatogram* → *Extract Chromatograms*; for Type, select TIC, and in the MS Chromatogram tab select MS level: MS/MS and select the precursor ion for the peak of interest.

21| Use the Walk Chromatogram cursor to click on individual scans at and near your peak of interest.

22| Inspect the individual MS/MS scans at and around this RT to assess spectral purity. Often a portion of the precursor ion will remain intact, making it easier to identify the spectrum of interest and assess spectral purity. Generally, if a similar fragmentation pattern is consistently seen across a few scans, and the MS spectrum lacks co-eluting species within a few *m/z*, then the spectra can be considered pure and sufficiently intense to identify the peak of interest.

? TROUBLESHOOTING

23| Scroll through the metabolites returned by the METLIN search in Stage 4 to find ones with MS/MS data (indicated by a 'View' button) (**Fig. 4**).

24| Click on 'View'. The spectrum will appear (**Fig. 5**).

25| Click on individual lines in the spectral table to select a specific precursor and voltage; the appropriate spectrum will appear. You can right-click and drag a box to zoom in. Roll your cursor over a spectral peak and the exact mass for the fragment will be displayed along with a predicted structure for that fragment if available. Click 'Reset zoom' in the upper left to zoom back out. Right-click and hold 'move' to move the spectral window around the page. To close, click on 'close' in the upper right corner.

Stage 6: Compare experimental MS/MS with METLIN

26| Compare your experimental spectra with the spectra in METLIN by visual inspection. If the same fragment ions are present in the experimental spectra and the METLIN spectra with very similar intensity ratios, you have a match, as seen for phenylalanine (**Fig. 6a**), arachidonic acid (**Fig. 6b**) and hypoxanthine in positive and negative modes

(**Fig. 6c**,**d**). Hypoxanthine in positive mode (**Fig. 6c**) is a good match, as the major experimental fragments are of similar intensity as the standard, although there is some low-intensity contamination. If you find an acceptable match, you can go to Stage 7. If several high-intensity ions are missing or the ratios are markedly different (as seen in **Fig. 7**, in which the intensity ratios between the experimental spectra in black are different from the standard spectra in red), you have not found a match.

? TROUBLESHOOTING

Stage 7: Verify that the characterization is correct using a standard

27| If you found an exact match between your experimental spectra at both the precursor and fragment levels, then you have characterized the metabolite. Depending on the level of confidence needed in your analysis, you should follow up with additional techniques to support your identification. Techniques such as FTICR-MS or NMR can give you an additional level of confidence, although metabolite concentrations often prevent the use of NMR to characterize metabolites. The highest level of confidence is obtained when standards are synthesized or purchased, and compared by LC-MS/MS to confirm retention time and MS/MS with the same parameters.

? TROUBLESHOOTING

Stage 1: If it is determined that your metabolic peak of interest is an isotope peak, one must be cautious that this may be a false positive. If your peak is an adduct other than M + H or M − H, one should look back at the original profiling experiment to see whether the monoisotopic peak or M + H or M − H is also dysregulated. If this is the case, complete this protocol with the M + H or M−H ion. If it is not dysregulated, do another simple search in METLIN with the correct adduct selected. As we discussed above (**Supplementary Fig. 1**), in-source fragments should also be checked. These in-source fragments always co-elute with their parent ions. If the in-source fragment ion is identified, one should look for the parent ion at the same retention time. If the parent ion is also dysregulated, complete this protocol with the parent ion.

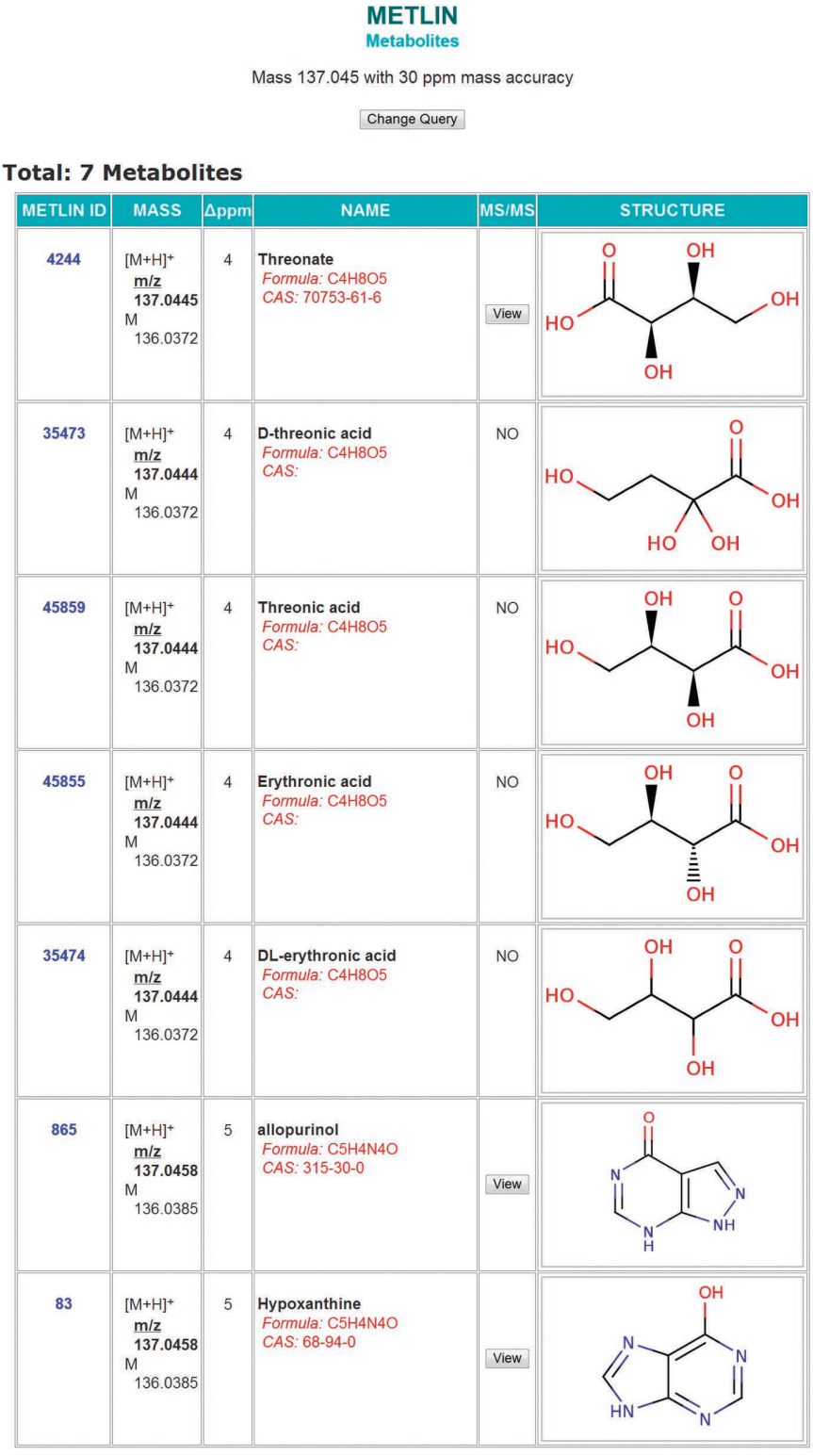


Figure 4 | Screenshot of the returned metabolites from the search for 137.045 in METLIN, with structural and mass spectral information.

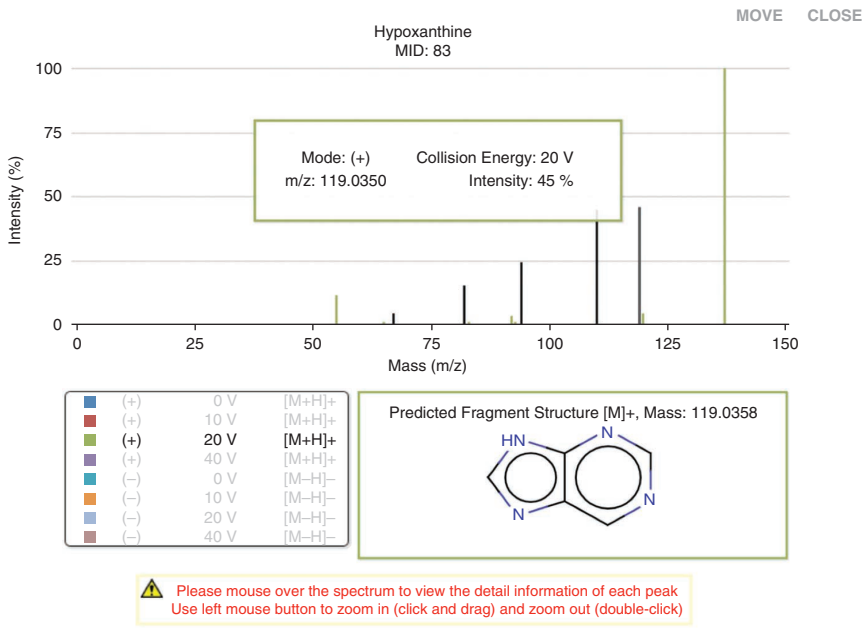


Figure 5 | Screenshot of the spectrum of hypoxanthine. The fragmentation spectrum is shown. Clicking on the other voltages in the black-bounded box displays the appropriate spectrum. The act of rolling over a fragment peak (such as 119) with your mouse reveals a predicted fragment structure and details about the exact mass and relative intensity of that fragment.

Stage 2: If a co-eluting metabolic peak is within 1–2 *m/z* of your ion of interest, it may provide a convoluted spectrum. If you suspect that this is the case, you should refragment this species with a narrower isolation window. If it is within 1 Da, this may not be sufficient to isolate the species, and you may need to use another approach to identify this peak. If two ions are co-eluting, different chromatographic conditions may allow these two species to be separated as in **Figure 2**.

Stage 4: If no metabolites are returned from the search, you can increase the tolerance value, or add additional adducts if appropriate. For the ionic metabolites, when searching the METLIN database, the 'neutral' should be chosen for the 'charge' setting. In addition, the isotopic pattern distribution also helps predict the empirical formula of unknown compounds. Most data analysis tools provided by instrument vendors have this function.

Stage 5: If you cannot identify the precursor ion in Step 22, you may want to rerun the sample, performing fragmentation at a lower energy in Stage 3. If the precursor is identified, but there is insufficient fragmentation, you may want to rerun the sample, fragmenting at a higher energy in Stage 3.

Stage 6: Note that MS/MS spectra in the METLIN database are acquired on Agilent Q-TOF mass spectrometers. Although we have demonstrated that other Q-TOF mass spectrometers have similar MS/MS spectra to those in the METLIN database[53], the relative intensities of fragment ions in MS/MS spectra may be slightly different, depending on the instrument settings. In addition, MS/MS spectra in METLIN database are acquired with an isolation window of 1.3 Da, and thus there is no isotopic peak for fragment ions. When MS/MS spectra in Stage 3 are acquired with a wider isolation window (e.g., 4 Da), one should expect that isotopic peaks will be shown in the MS/MS spectra.

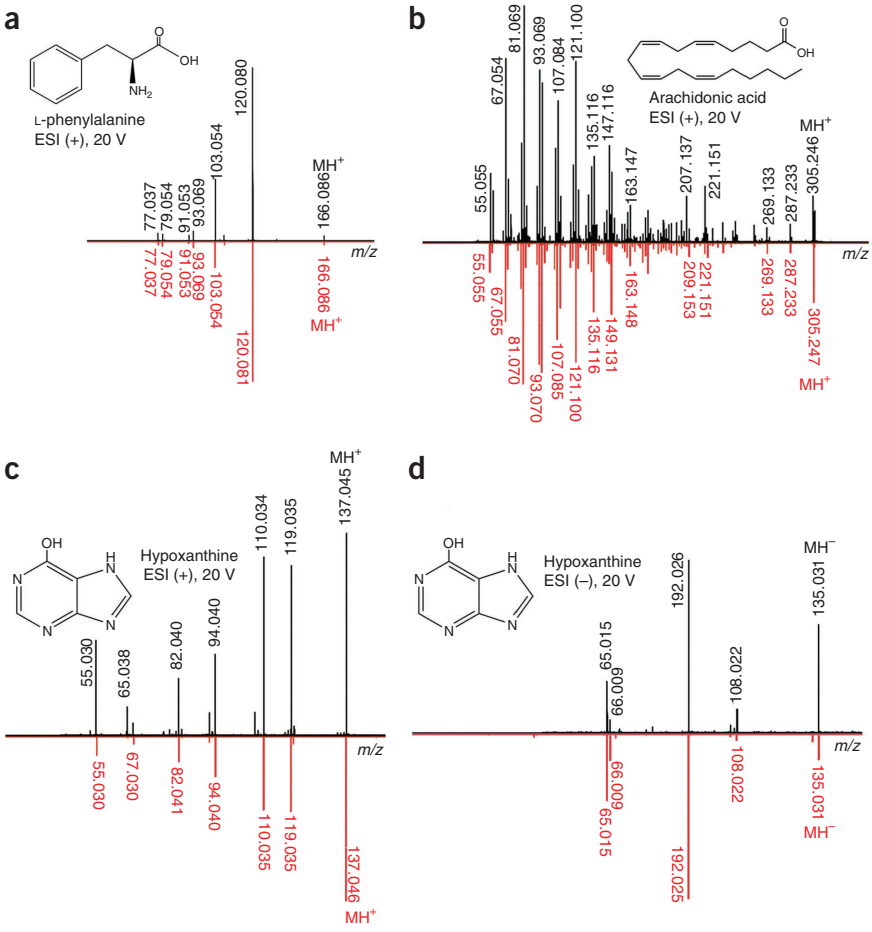


Figure 6 | A comparison of experimental (black) and METLIN standard (red) spectra for three metabolites. (**a**–**d**) Phenylalanine (**a**), arachidonic acid (**b**) and hypoxanthine in positive (**c**) and negative (**d**) mode.

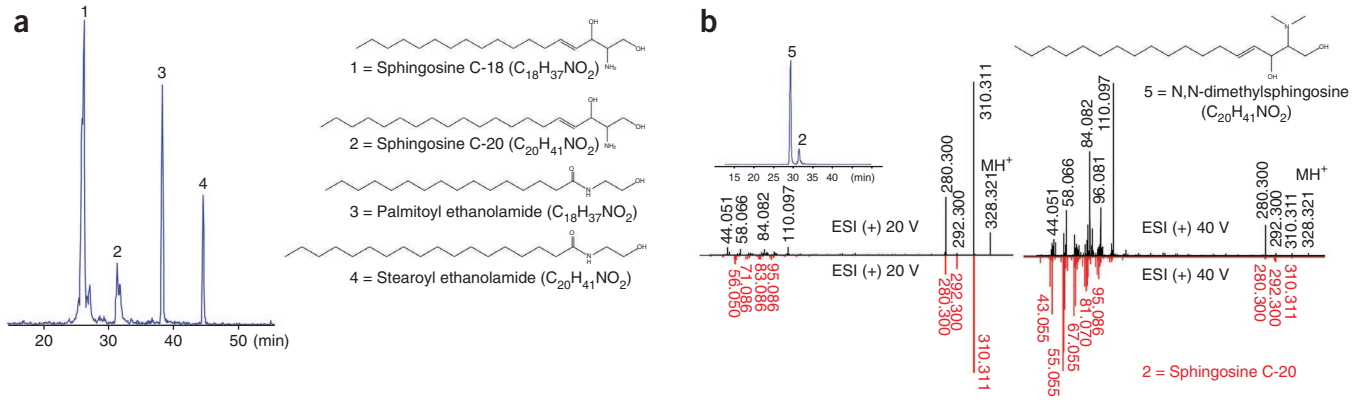


Figure 7 | The importance of retention time, accurate mass and fragmentation for identification. (**a**) Separation of sphingosine C-18 (peak 1), sphingosine C-20 (peak 2), palmitoyl ethanolamide (peak 3) and stearoyl ethanolamide (peak 4) from a tissue extract analyzed by reversed-phase–HPLC-MS/MS. Note that the isobaric species 1 and 3 are well separated and can be identified. Without separation, a convoluted spectrum would be produced. (**b**) Characterization of *N*,*N*-dimethylsphingosine (DMS). In a separate analysis, DMS (peak 5) was observed, which is isobaric with sphingosine C-20 (peak 2) and stearoyl ethanolamide (not observed in this analysis). Top, the MS/MS spectra of DMS (black) acquired at the collision energy of 20 V and 40 V, respectively. Bottom, the MS/MS spectra of sphingosine C-20 (red) in METLIN database with the collision energy of 20 V and 40 V, respectively. Comparison of the experimental spectra of DMS against sphingosine C-20 reveals a poor match because of different ratios between the higher intensity species at 20 V and a poor correlation in the lower mass species at 40 V.

● TIMING

This protocol should take 2–4 h, depending on the metabolite.

ANTICIPATED RESULTS

This protocol allows one to characterize a peak of interest in an untargeted metabolomic experiment if it is a metabolite found in METLIN, or is an analog of a metabolite in METLIN. Metabolites that are not in METLIN or not analogs of known metabolites are difficult to identify with this technique, although this protocol will provide information that would be valuable when used in combination with other analytical techniques. Some cases that have proved challenging when attempting to identify unknown metabolites are discussed below; they include examples of metabolites that have no exact match in METLIN and metabolites that co-elute with other metabolic peaks of similar *m/z*.

For our first example, the metabolic peak of interest has an *m/z* of 496.3409 and an RT of 24.5. The ion spectrum is extracted (**Fig. 1c**) from the TIC, and upon inspection of the spectrum at *m/z* 496.3409 another peak is observed at *m/z* 518.3219, which is 21.981 amu larger. This is characteristic of the $[M + Na]^+$ peak and supports the fact that *m/z* 496.3409 is the $[M + H]^+$ peak ($Na^+ - H^+ = 21.9820$). As also noted (**Fig. 1d**), two isotope peaks for the *m/z* 496.3409 peak can be seen, *m/z* 497.3440 and *m/z* 498.3455. As these peaks are approximately +1 and +2 from the $[M + H]^+$ peak, it adds validation that this is a singly charged ion and that *m/z* 496.3409 is indeed the protonated monoisotopic mass of the molecule.

To determine the structure of the species at 480.3082 in **Figure 2**, caution must be taken to avoid potential contamination from the species at *m/z* 479.7786 $[M + 2H]^{2+}$, *m/z* 480.2805 and *m/z* 482.2569 $[M + 2H]^{2+}$. Indeed, when *m/z* 480.3082 is isolated and fragmented, the spectrum in **Figure 2b** is obtained, which contains both *m/z* 480.2805 (isotope of *m/z* 479.7786) and *m/z* 482.2567 species. In this situation, *m/z* 480.3082 cannot be identified, as the MS/MS spectrum is suppressed and contaminated. If chromatography is used to separate these species, as shown in **Figure 2c**, a pure MS/MS spectrum can be obtained for *m/z* 480.3084 (**Fig. 2d**), which is characterized as lysoPE(18:1/0:0). The use of a narrow isolation window may also be useful to prevent contamination by other species if the mass difference of two species is sufficient.

The characterization of three metabolites, phenylalanine, arachidonic acid and hypoxanthine, is depicted in **Figure 6**. The simple fragmentation of the experimental phenylalanine (**Fig. 6a**) and the more complex arachidonic acid (**Fig. 6b**) match the standard METLIN spectra in both intensity ratio and accurate mass of the fragments, supporting their identification. The experimental spectrum for hypoxanthine in negative mode (**Fig. 6d**) matches well with the METLIN spectrum, although there is substantially more contamination in the experimental sample than observed in positive mode (**Fig. 6c**). The observation that hypoxanthine is dysregulated in both positive and negative modes also supports the characterization of this peak.

In addition to the MS/MS pattern, the retention time is another key parameter to consider. As seen in **Figure 4**, a search for *m/z* 137.0450 returns seven hits. The first five hits (such as threonate) are organic acids, and the remaining two hits (allopurinol and hypoxanthine) are more basic metabolites. The two types of metabolites could be differentiated by their retention time and ionization efficiency using positive-mode ESI. This helps narrow down the candidates before comparing MS/MS spectra. However, to further differentiate allopurinol and hypoxanthine, MS/MS matching is necessary.

Another example for the importance of retention time is shown in **Figure 7**. The precursor ion *m/z* 300.2889 is appropriate for both sphingosine C-18 and palmitoylethanolamide, which have the same formula of $C_{18}H_{37}NO_2$ (**Fig. 7a**). These molecules are indistinguishable by accurate mass alone. If these molecules were not resolved by chromatography, both species would be selected to fragment at the same time, generating a convoluted spectrum that would hinder the identification of either species. When resolved, the individual species can be analyzed and structures can be assigned to each peak, as represented by peaks 1 and 3 in **Figure 7a**. The relative retention time can support a structural assignment. In **Figure 7**, two additional peaks, 2 and 4, can be seen, which are analogs of 1 and 3 but are an additional two carbon units long. In general, on C18-based columns, increasing chain number and increasing saturation increases the retention time for a group of molecules with the same functional group. Observing a later retention time for sphingosine C-20 over sphingosine C-18 and stearoyl ethanolamide over palmitoylethanolamide is consistent with their characterization. **Figure 7b** shows the importance of MS/MS spectral matching to differentiate the *N*,*N*-dimethylsphingosine (DMS, peak 5) and its isobaric species sphingosine C-20 (peak 2).

Investigators who have access to pure standards of compounds that are not currently characterized in METLIN can email metlin@scripps.edu to arrange for these to be added to the database.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS This work was supported by the California Institute of Regenerative Medicine (no. TR1-01219) (G.S.), the US National Institutes of Health (nos. R01 CA170737 (G.S.), R24 EY017540 (G.S.), P30 MH062261 (G.S.), RC1 HL101034(G.S.), P01 DA026146 (G.S.), and 1R01 ES022181-01) (G.J.P.) and the US National Institutes of Health-National Institute on Aging (no. L30 AG0 038036) (G.J.P.). Financial support was also received from the US Department of Energy (grant nos. FG02-07ER64325 and DE-AC0205CH11231) (G.S.).

AUTHOR CONTRIBUTIONS Z.-J.Z., A.W.S. and J.W. and contributed equally to the work described. G.J.P. and G.S. supervised the work. A.W.S., J.W. and G.J.P. performed the experiments. Z.-J.Z., A.W.S., J.W. and C.H.J. wrote the manuscript. Z.-J.Z., S.M.Y., G.J.P. and G.S. read and revised the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nprot.2013.004. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
2. Wikoff, W.R., Gangoiti, J.A., Barshop, B.A. & Siuzdak, G. Metabolomics identifies perturbations in human disorders of propionate metabolism. *Clin. Chem.* **53**, 2169–2176 (2007).
3. 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).
4. Vinayavekhin, N. & Saghatelian, A. Regulation of alkyl-dihydrothiazole-carboxylates (ATCs) by iron and the Pyochelin gene cluster in *Pseudomonas aeruginosa*. *ACS Chem. Biol.* **4**, 617–623 (2009).
5. Kalisiak, J. *et al.* Identification of a new endogenous metabolite and the characterization of its protein interactions through an immobilization approach. *J. Am. Chem. Soc.* **131**, 378–386 (2008).
6. Leiss, K.A., Maltese, F., Choi, Y.H., Verpoorte, R. & Klinkhamer, P.G.L. Identification of chlorogenic acid as a resistance factor for thrips in *Chrysanthemum*. *Plant Physiol.* **150**, 1567–1575 (2009).
7. Allen, J. *et al.* Discrimination of modes of action of antifungal substances by use of metabolic footprinting. *Appl. Environ. Microbiol.* **70**, 6157–6165 (2004).
8. Clayton, T.A., Baker, D., Lindon, J.C., Everett, J.R. & Nicholson, J.K. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl Acad. Sci. USA* **106**, 14728–14733 (2009).
9. Ludwig, C. & Viant, M.R. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochem. Anal.* **21**, 22–32 (2010).
10. Powers, R. NMR metabolomics and drug discovery. *Magn. Reson. Chem.* **47**, S2–S11 (2009).
11. Dettmer, K., Aronov, P.A. & Hammock, B.D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51–78 (2007).
12. Lei, Z., Huhman, D. & Sumner, L.W. Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* **286**, 25435–25442 (2011).
13. Smart, K.F., Aggio, R.B.M., Van Houtte, J.R. & Villas-Boas, S.G. Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography-mass spectrometry. *Nat. Protoc.* **5**, 1709–1729 (2010).
14. Dunn, W.B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **6**, 1060–1083 (2011).
15. Chan, E.C.Y., Pasikanti, K.K. & Nicholson, J.K. Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat. Protoc.* **6**, 1483–1499 (2011).
16. Fiehn, O. *et al.* Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–1161 (2000).
17. Babushok, V.I. *et al.* Development of a database of gas chromatographic retention properties of organic compounds. *J. Chromatogr. A* **1157**, 414–421 (2007).
18. Kind, T. *et al.* FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* **81**, 10038–10048 (2009).
19. Xu, F., Zou, L. & Ong, C.N. Multiorigination of chromatographic peaks in derivatized GC/MS metabolomics: a confounder that influences metabolic pathway interpretation. *J. Proteome Res.* **8**, 5657–5665 (2009).
20. 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 (2007).
21. Wishart, D.S. *et al.* The human cerebrospinal fluid metabolome. *J. Chromatogr. B* **871**, 164–173 (2008).
22. Lu, W., Bennett, B.D. & Rabinowitz, J.D. Analytical strategies for LC–MS-based targeted metabolomics. *J. Chromatogr. B* **871**, 236–242 (2008).
23. Kaddurah-Daouk, R. *et al.* Lipidomic analysis of variation in response to simvastatin in the Cholesterol and Pharmacogenetics Study. *Metabolomics* **6**, 191–201 (2010).
24. Vinayavekhin, N. & Saghatelian, A. Untargeted metabolomics. *Curr. Protoc. Mol. Biol.* **90**, 30.1.1–30.1.24 (2001).
25. Johnson, C.H. *et al.* Radiation metabolomics. 4. UPLC-ESI-QTOFMS–based metabolomics for urinary biomarker discovery in γ-irradiated rats. *Radiat. Res.* **175**, 473–484 (2011).
26. Trupp, M. *et al.* Metabolomics reveals amino acids contribute to variation in response to simvastatin treatment. *PLoS ONE* **7**, e38386 (2012).
27. Wikoff, W.R., Kalisak, E., Trauger, S., Manchester, M. & Siuzdak, G. Response and recovery in the plasma metabolome tracks the acute LCMV-induced immune response. *J. Proteome Res.* **8**, 3578–3587 (2009).
28. Panopoulos, A.D. *et al.* The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res.* **22**, 168–177 (2012).
29. Yanes, O. *et al.* Metabolic oxidation regulates embryonic stem cell differentiation. *Nat. Chem. Biol.* **6**, 411–417 (2010).
30. Marshall, A.G. & Hendrickson, C.L. High-resolution mass spectrometers. *Annu. Rev. Anal. Chem.* **1**, 579–599 (2008).

31. Verhoeven, H.A., Ric de Vos, C.H., Bino, R.J. & Hall, R.D. Plant metabolomics strategies based upon quadrupole time-of-flight mass spectrometry (QTOF-MS). *Plant Metabolomics*, **57**, 33–48 (2006).
32. Kamleh, A. *et al.* Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules. *Rapid Commun. Mass Spectrom.* **22**, 1912–1918 (2008).
33. Breitling, R., Pitt, A.R. & Barrett, M.P. Precision mapping of the metabolome. *Trends Biotechnol.* **24**, 543–548 (2006).
34. Brown, S.C., Kruppa, G. & Dasseux, J.-L. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom. Rev.* **24**, 223–231 (2005).
35. Smith, C.A. *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747–751 (2005).
36. Patti, G.J. *et al.* Metabolomics implicates altered sphingolipids in chronic pain of neuropathic origin. *Nature Chem. Biol.* **8**, 232–234 (2012).
37. Psychogios, N. *et al.* The human serum metabolome. *PLoS One* **6**, e16957 (2011).
38. Chen, L., Zhou, L., Chan, E.C.Y., Neo, J. & Beuerman, R.W. Characterization of the human tear metabolome by LC–MS/MS. *J. Proteome Res.* **10**, 4876–4882 (2011).
39. Want, E.J. *et al.* Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* **5**, 1005–1018 (2010).
40. Nebert, D.W., Zhang, G. & Vesell, E.S. From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. *Drug Metab. Rev.* **40**, 187–224 (2008).
41. Brown, M. *et al.* A metabolome pipeline: from concept to data to knowledge. *Metabolomics* **1**, 39–51 (2005).
42. 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).
43. Tautenhahn, R., Patti, G.J., Rinehart, D. & Siuzdak, G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* **84**, 5035–5039 (2012).
44. Kiefer, P., Delmotte, N.l. & Vorholt, J.A. Nanoscale ion-pair reversed-phase HPLC-MS for sensitive metabolome analysis. *Anal. Chem.* **83**, 850–855 (2010).
45. Castro-Perez, J. *et al.* Localization of fatty acyl and double bond positions in phosphatidylcholines using a dual-stage CID fragmentation coupled with ion mobility mass spectrometry. *J. Am. Soc. Mass Spectrom.* **22**, 1552–1567 (2011).
46. Hsu, F.-F. & Turk, J. Elucidation of the double-bond position of long-chain unsaturated fatty acids by multiple-stage linear ion-trap mass spectrometry with electrospray ionization. *J. Am. Soc. Mass Spectrom.* **19**, 1673–1680 (2008).
47. Thomas, M.C. *et al.* Ozone-induced dissociation: elucidation of double bond position within mass-selected lipid ions. *Anal. Chem.* **80**, 303–311 (2007).
48. Gian Luigi, R. Dietary *n*-6 and *n*-3 polyunsaturated fatty acids: From biochemistry to clinical implications in cardiovascular prevention. *Biochem. Pharmacol.* **77**, 937–946 (2009).
49. Ding, J. *et al.* Capillary LC coupled with high-mass measurement accuracy mass spectrometry for metabolic profiling. *Anal. Chem.* **79**, 6081–6093 (2007).
50. Lindon, J.C. & Nicholson, J.K. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu. Rev. Anal. Chem.* **1**, 45–69 (2008).
51. Cravatt, B. *et al.* Chemical characterization of a family of brain lipids that induce sleep. *Science* **268**, 1506–1509 (1995).
52. Yanes, O., Tautenhahn, R., Patti, G.J. & Siuzdak, G. Expanding coverage of the metabolome for global metabolite profiling. *Anal. Chem.* **83**, 2152–2161 (2011).
53. Tautenhahn, R. *et al.* An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat. Biotechnol.* **30**, 826–828 (2012).