Global metabolic profiling procedures for urine using UPLC–MS

Elizabeth J Want[1], Ian D Wilson[2], Helen Gika[3], Georgios Theodoridis[3], Robert S Plumb[4], John Shockcor[4], Elaine Holmes[1] & Jeremy K Nicholson[1]

[1]Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK. [2]Department of Clinical Pharmacology, Drug Metabolism and Pharmacokinetics, AstraZeneca, Macclesfield, Cheshire, UK. [3]Laboratory of Analytical Chemistry, Department of Chemistry, Aristotle University of Thessaloniki, Thessaloniki, Greece. [4]Pharmaceutical Business Operations, Waters Corporation, Milford, Massachusetts, USA. Correspondence should be addressed to E.J.W. (e.want@imperial.ac.uk).

Published online 6 May 2010; doi:10.1038/nprot.2010.50

The production of 'global' metabolite profiles involves measuring low molecular-weight metabolites (< 1 kDa) in complex biofluids/tissues to study perturbations in response to physiological challenges, toxic insults or disease processes. Information-rich analytical platforms, such as mass spectrometry (MS), are needed. Here we describe the application of ultra-performance liquid chromatography–MS (UPLC–MS) to urinary metabolite profiling, including sample preparation, stability/storage and the selection of chromatographic conditions that balance metabolome coverage, chromatographic resolution and throughput. We discuss quality control and metabolite identification, as well as provide details of multivariate data analysis approaches for analyzing such MS data. Using this protocol, the analysis of a sample set in 96-well plate format, would take ca. 30 h, including 1 h for system setup, 1–2 h for sample preparation, 24 h for UPLC–MS analysis and 1–2 h for initial data processing. The use of UPLC–MS for metabolic profiling in this way is not faster than the conventional HPLC-based methods but, because of improved chromatographic performance, provides superior metabolome coverage.

INTRODUCTION

The measurement of metabolite levels and variation in biofluids can offer many insights into disease processes, drug toxicity and response to therapeutic intervention, as well as providing information regarding the effects of growth and aging, diurnal variation, nutrition and exercise on metabolism[1–15]. ^{1}H NMR spectroscopy and mass spectrometry (MS) are the two main analytical spectroscopic approaches in metabolic profiling, usually offering complementary information, but with different operational performance characteristics[1–9]. The protocols for the efficient use of NMR spectroscopy in biofluids and tissues have been presented recently[16]. Here we consider the particular protocols and approach needed to efficiently analyze urine samples by MS, which pose a particular analytical challenge in metabolic profiling. Metabolite separation techniques before MS analysis, such as liquid chromatography (LC), can both reduce mass spectral complexity and provide additional information on metabolite physicochemical properties, which may help in metabolite identification. It is appreciated that the parallel application of different MS-based techniques, e.g., GC–MS and LC–MS, combined with NMR spectroscopy, may be required for comprehensive metabolic profiling.

MS-based metabolite profiling

In the context of metabolic profiling, MS, when carried out carefully, can afford sensitive, accurate and reproducible measurements of the metabolites present in biofluids, tissues or organisms, with the ability to cover a wide dynamic range [17–20]. These attributes are essential for addressing the challenges of biomarker discovery, as the range of metabolite concentrations easily exceeds nine orders of magnitude in many biofluids, and the diversity of molecular species encompasses simple amino and organic acids, hormones, neurotransmitters, vitamins, peptides, lipids and complex carbohydrates. MS provides a sensitive and reproducible approach to metabolic profiling. Detailed information can be obtained regarding the metabolic state of a biofluid and structural information can be obtained on a wide range of important metabolites. The role of MS in metabolic profiling is evolving constantly, as both instrumentation and software becomes more sophisticated and researchers realize current technological capabilities. Additional challenges arise in generating a comprehensive metabolite profile, downstream data processing and analysis, and structural characterization/elucidation of important metabolites. A typical MS-based metabolic profiling workflow is shown in **Figure 1**.

Urine samples for metabolite profiling

Urine poses several analytical challenges for metabolic profiling, due to large variations in ionic strength, pH and osmolarity, particularly under conditions of physiological stress. Further, urine possesses a huge dynamic range of metabolite concentrations, as well as being subject to variable and unpredictable dilution. It is important to note that there is an extreme diversity of chemical classes in urine, encompassing microbial cometabolites as well as mammalian metabolites. In the case of humans, drugs, pollutants and industrial chemicals may also be present in urine. In summary, there are more possible compounds present in urine than any other biological matrix, making data analysis and biological interpretation challenging. However, urine is a key biological matrix in metabolic profiling studies, as its collection is noninvasive (and therefore simple), and urine samples are less likely to be volume-limited, although this is dependent on both the animal and collection times[21,22]. Further, urine can easily be sampled in a serial manner, allowing temporal metabolic changes to be studied. As urine is not under homeostatic regulation, being a waste product, it can reflect metabolic disregulation, thus providing insights into system-wide changes in response to physiological challenges or disease processes.

Sample preparation

Urine, particularly from healthy human individuals, contains relatively little protein (or other high molecular mass compounds)

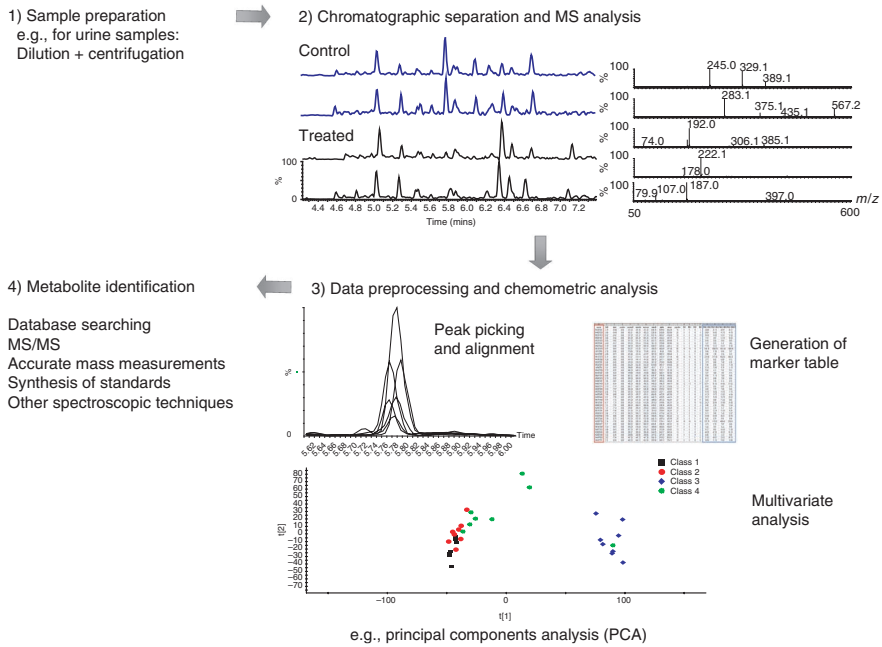


Figure 1 | Flow diagram of a typical MS-based metabolite profiling workflow. Step 1 is sample preparation, followed by MS analysis, usually coupled to a LC or GC separation step. A key component is data analysis, which can be divided into data preprocessing and chemometric analysis. This is then followed by the identification of important metabolites.

because of filtration through the renal tubules. Obviously this may not be the case in certain renal diseases, or in experimental animal studies, as rodents are physiologically proteinuric. The sample preparation approaches for MS urine analyses are much simpler than for biofluids such as serum or plasma (or example, refs. 18–20), or tissues. Often, centrifugation to remove particulates followed by dilution with water (1:1 to 1:3 vol/vol depending on origin of urine, that is human or rat) is all that is required, and such an approach clearly minimizes the potential for analyte losses. An alternative approach for the removal of urinary proteins and particulates is the use of molecular weight cut-off filters, though care must be taken in preparing and handling these filters to minimize the risk of sample contamination, particularly if glycerol has been used in their manufacture or storage.

After the removal of particulates, samples are then transferred to appropriate LC vials (typically maximum recovery vials) or, more commonly in large-scale metabolic profiling studies, 96-well plates with cap mats. Prepared samples should be kept on ice or in the fridge at 0–4 °C before transferring to the autosampler, where they should also be kept at 0–4 °C throughout the analysis. If necessary, prepared samples can be stored frozen at the lowest available temperature (at least −20 °C) before analysis.

Separation techniques

Although direct infusion of samples into a mass spectrometer can provide a rapid method for obtaining metabolic fingerprints, this can lead to the loss of signals for particular analytes as a result of ion suppression due to competing analytes entering the MS simultaneously[23–25]. Therefore, for metabolic profiling studies, it is better practice to carry out a separation, which is not only typically chromatographic but also using capillary electrophoresis (CE) before MS analysis to reduce the potential for ion suppression. Gas-chromatography–MS (GC–MS) was one of the first MS-based metabolic profiling techniques[26] and is still widely used today for both global metabolite analysis and targeted applications such as urinary steroid profiling[27]. However, the use of GC in this role has limitations, especially for urine, as by definition separation takes place in the gas phase and analytes must therefore be volatile. Many of the analytes present in urine are polar, ionic and relatively involatile and, as a result, require complex and lengthy sample preparation and derivatization to obtain samples that are both suitable for injection onto the GC column and that contain volatile analytes (a useful review of derivatization procedures is to be found in ref. 28).

Liquid chromatography, especially reversed-phase (RP)LC, is well suited to the analysis of the types of polar, water soluble molecules typically encountered in urine, with the ability to measure a wide range of chemical classes of molecules, enabling the generation of complex metabolic profiles without the need for prior derivatization. With a properly optimized separation, the number of coeluting analytes entering the mass spectrometer ion source at any one time can be reduced and so ion suppression is decreased. An efficient analytical separation will result in improved detection limits and therefore better mass spectral data quality because of reduced background noise.

Column choice

The choice of column for LC analysis is dependent on the matrix of interest, and so by default, the compounds being analyzed. In urine, the components of interest are predominantly of low molecular mass and are generally hydrophilic. Hence, RP columns, typically C18-bonded silicas, with good retention power, provide a good general system for the metabolite profiling of urine. High-strength silica (HSS) columns for ultra-performance LC (UPLC)–MS systems show improved retention of certain polar metabolites and thus are a possible alternative to traditional C18 columns[29]. To ensure retention, the samples are loaded onto the column under conditions where the mobile phase is predominantly aqueous (i.e., 99–100% water), and therefore of low elutropic strength, after which the analytes are eluted from the column and into the MS using a gradient with increasing organic solvent content (most often acetonitrile or methanol). Mobile phase additives, e.g., formic acid, are often added to reduce the pH of the mobile phase, to suppress the ionization of weak organic acids, and thereby improve retention.

Although conventional HPLC is well suited to urine analysis, the introduction of UPLC, with its greatly enhanced chromatographic efficiency, has improved sensitivity, resolution and analysis time, resulting in the detection of an even greater number of metabolites[30].

Although RP chromatography is the standard approach for separating medium polar and nonpolar analytes, highly polar metabolites will not be retained and so elute with the void volume, thus hindering unambiguous identification and accurate quantification. For these applications, hydrophilic interaction chromatography

(HILIC), using either silica or derivatized silica columns, provides complementary information to that obtained using RP chromatography[31–33]. HILIC approaches combined with electrospray ionization (ESI)–MS techniques have already been applied to the analysis of urine[33,34] and show different selectivity compared with conventional RP separations[34]. HILIC provides an improved means of profiling certain classes of polar analytes, thereby giving a different view of the composition of the urine samples compared with the RP mode. Together, the application of RP and HILIC columns for urine analysis by UPLC has been shown to provide complementary metabolite information, and thus enhanced metabolome coverage[34]. However, for general profiling applications conventional RP-based methods provide a good starting point for the analysis of urine.

Considerations for optimizing UPLC conditions

As urinary metabolites cover a wide range of polarities, isocratic solvent systems are not suitable for comprehensive analysis and so gradient elution is used, where the eluotropic strength of the solvent is increased as the analysis progresses. The choice of gradient is sample dependent, but also relates to the question being asked, i.e., whether the analysis will be untargeted or targeted. UPLC gradients for urine samples can range from as short as 5 min up to 30 min for more complete metabolome coverage[34–37]. If the goal of a researcher's study is to detect the greatest number of metabolites in their samples, they are then advised to try different column dimensions and gradients to determine which is most suitable for their sample set. However, there is a choice to be made between throughput and metabolome coverage/peak resolution as, in general, short runs will detect fewer ions than long ones for a variety of reasons, including ion suppression. Empirically, a reasonable compromise between speed of analysis and metabolome coverage for urine is provided by a 10–12 min UPLC–MS gradient, which will afford good metabolite separation combined with moderate throughput (**Fig. 2**). Column temperature is another, often neglected, parameter that should be actively controlled to ensure reproducible chromatography. Generally separations are carried out at controlled temperatures, typically up to 60 °C[38], though higher temperatures have been effectively used in a number of cases[38,39]. For reversed-phase analyses, the gradient will often start at high aqueous (99–100% water), and ramp up to high organic content (95–100%; acetonitrile/methanol). The gradient we recommend here has a hold step from 0 to 1 min at 99% aqueous mobile phase (water) to allow the salts to be washed from the column. The advantage of going to such high proportions of organic modifier during the analytical run is that unwanted contaminants are eluted from the column, preventing deterioration in performance through the run. After the wash period, the mobile phase composition then reverts to starting conditions, where it is held for an appropriate period (approximately five column volumes) to ensure column re-equilibration before the next sample is injected. With 10–12 min runs and a 2–3 min re-equilibration period, ca. 100 samples can be run per day, but by using shorter (2–5 min) analysis times throughput can be increased to many hundreds of samples in a 24 h period.

Choice of ionization and mass analyzer

For urine, ESI is typically employed for metabolite analysis as it is well suited to the polar/ionic nature of the analytes. Ions are generated directly from the liquid phase into the gas phase, establishing this technique as a convenient mass analysis platform for both liquid chromatography and automated sample analysis. This relatively 'soft' ionization technique results in minimal fragmentation and thereby enables the detection of a wide range of molecules with excellent quantitative analysis and good (though analyte-dependent) sensitivity. Multiply charged ions can be generated, thus enabling the analysis of both small and large molecules. Metabolites containing only C, H and O may be expected to be detected by LC–MS in negative ion mode, whereas metabolites, which in addition contain N, would be expected to be preferentially ionized in positive-ion mode. Therefore, to enhance metabolome coverage, ionization should be carried out in both positive and negative mode, enabling the detection of two sets of analytes which may differ significantly[40,41]. Depending on the mass analyzer, detection in positive and negative mode can be carried out simultaneously in a single run[42], reducing both analysis time and injection variability. However, in some cases this can reduce sensitivity due to loss of data during polarity switching.

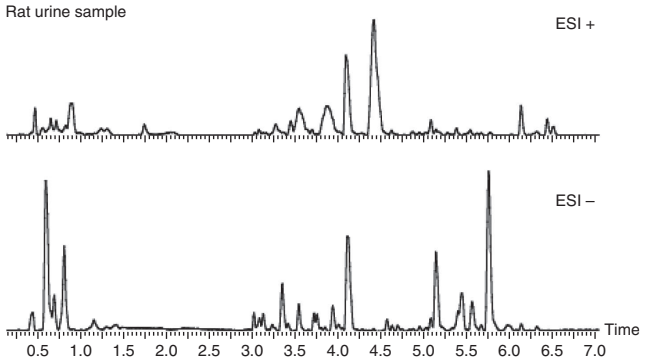


Figure 2 | A BPI UPLC–MS chromatogram of a urine sample run in positive ESI mode (top) and negative ESI mode (below). Positive and negative ESI modes can offer complementary information regarding the metabolite profile.

Mass spectrometry can be carried out using mass analyzers with a range of mass resolution, from low (1,000) to very high (>100,000) as discussed below. Low mass-resolution MS include single and triple quadrupoles and quadrupole ion-traps, which are capable of measuring metabolite masses with unit resolution. Triple quadrupole instruments can be used to carry out tandem mass analysis (MS/MS). Here, each quadrupole has a separate function; the first quadrupole (Q1) scans across a preset *m/z* range for selection of one or more ions of interest, with fragmentation in the second quadrupole (Q2, or collision cell) using a collision gas (argon or helium). Q2 is typically an octapole in modern triple quadrupole instruments. Fragment ions generated in Q2 can either be analyzed in the third quadrupole (Q3) or subjected to further selection, in a subsequent selected reaction monitoring (SRM) experiment. This SRM capability of triple quadrupole instruments provides a highly sensitive approach for quantifying known metabolites and is best applied to targeted metabolic profiling experiments. Quadrupole ion-traps can be used for both MS scanning and MS/MS studies. A notable feature of the quadrupole ion-trap is the ability to carry out MS^n experiments, where multiple collision-induced dissociation experiments can be carried out serially. However, a particular disadvantage to quadrupole ion-traps is the upper limit on the ratio between precursor *m/z* and the lowest trapped fragment ion, which is ~0.3 (the 'one-third rule') and is a disadvantage for small molecule analysis[43]. In general, low-mass resolution mass analyzers are not desirable for global, untargeted metabolic profiling studies,

as they lack sufficient resolution to resolve coeluting metabolites having the same nominal mass.

High-mass resolution is a definite advantage when measuring complex biofluid samples such as urine, both in terms of the detection of distinct species and in structural elucidation of unknown compounds. Time-of-flight (ToF) instruments and Q-ToFs have virtually unlimited mass ranges and offer fast scanning capabilities and mass accuracies on the order of 5 p.p.m. Owing to their fast scanning speeds, ToFs and Q-ToFs are widely used in GC- UPLC- and MALDI-based metabolic profiling studies. Further, Q-ToF mass analyzers are highly suited for obtaining metabolite fragmentation data. Accurate mass measurements can be carried out on both precursor and product ions simultaneously (MS^E), thereby providing further structural information and aiding metabolite identification[44–46]. MS/MS experiments can also be used to probe the metabolome for specific compound classes by screening for characteristic ions and neutral losses. Fourier transform (FT) MS offers the highest mass resolution and mass measurement accuracy (sub-p.p.m.) of all analyzers, as well as very high sensitivity (attomole) and advanced structural and thermodynamic elucidation of molecules, with the ability to perform high mass-measurement accuracy MS^n experiments, often at p.p.m. level. The 'Orbitrap' MS[47] is an electrostatic ion-trap using fast FT (FFT). It provides high mass accuracy (1–2 p.p.m.) and resolution (up to 100,000), as well as a dynamic range of 5,000, and is usually operated together with a linear ion-trap as a hybrid instrument.

MS data

For a UPLC–MS-based metabolic profiling study, data will usually be collected in both positive and negative ESI modes, often using a mass range of 50–1,000 *m/z*. Typically, the sample is run separately in each mode, producing two data files for each sample for each run, although some instruments permit rapid polarity switching. Data can be displayed as a total ion chromatogram (TIC), which is the total ion signal versus time (or scan number). Alternatively, the base peak chromatogram (BPI) can be displayed, which is similar to the TIC but monitors a small window around only the most intense peak at any one time, thus representing the intensity of the most intense peak at every moment in the analysis. BPIs are often cleaner in appearance than TICs because many of the smaller peaks that sum together in a TIC to produce a large background are ignored. Typical UPLC–MS BPI chromatograms of urine are shown in **Figure 2** for positive (upper) and negative (lower) ESI, respectively. UPLC coupled with MS^E technology can provide both parent and fragment mass information of metabolites in one chromatographic run, illustrated here with a urine sample (**Fig. 3**) for *p*-cresol sulfate with *m/z* 187.

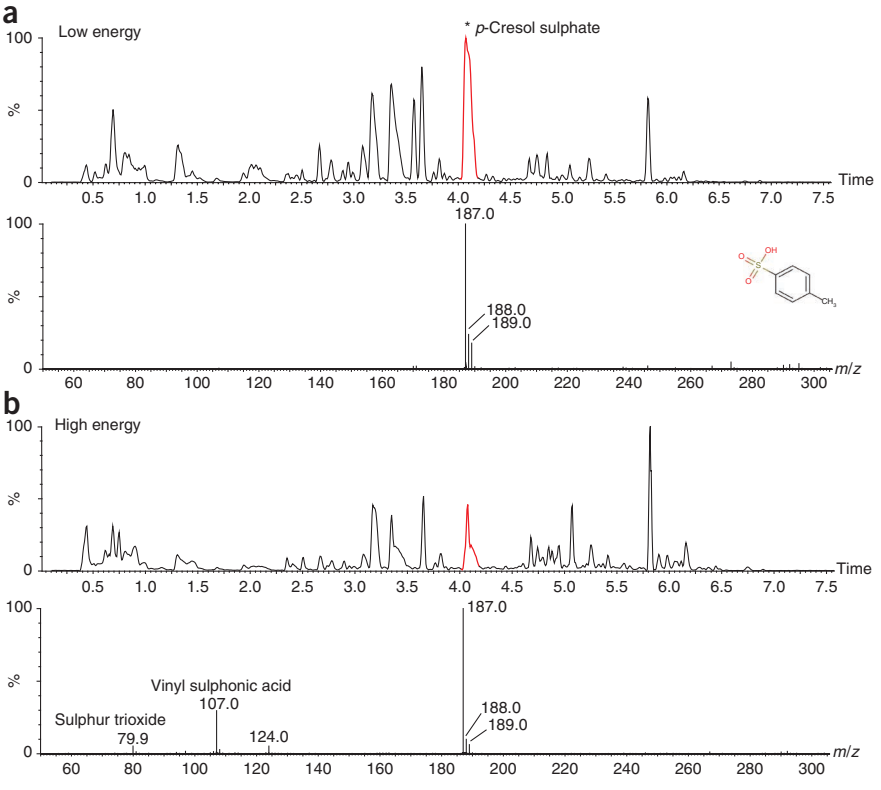


Figure 3 | MS^E data from a urine sample. (**a**) The top chromatogram shows the low energy data from a human urine sample, whereas the lower chromatogram shows the high energy data. (**b**) The corresponding low and high-energy mass spectra data are shown for *p*-cresol sulphate (*m/z* 187), with characteristic fragmentation information revealed through MS^E.

Experimental design

Key factors to consider when designing UPLC–MS metabolite profiling experiments include (1) sample collection and storage; (2) column conditioning; (3) the composition of test mixtures used; (4) the type and number of quality control samples and any 'blank' samples; (5) run order, i.e., how to carry out sample randomization; (6) the number and type of replicates to be analyzed; (7) the total run length; and (8) total number of samples and batch size.

Sample collection and storage. Urine is a convenient, minimally invasive biofluid for 'global' metabolite profiling. However, for useful data to be obtained it must be carefully collected and stored[17,21–22]. In the case of humans, both timed and 24 h collections can be obtained. With timed (or 'spot') collections, a mid-stream sample should be collected into a suitable container, sub-aliquoted into the sample containers to be used for storage, and frozen immediately with subsequent storage at the lowest available temperature (usually −20 °C or lower) to avoid metabolite decay and thus changes in metabolic profiles. This practice of sub-aliquoting the samples before storage will minimize subsequent freeze–thaw cycles. In the case of 24 h collections, it would seem to be good practice to store the sample in the refrigerator (4 °C) between collections. Few studies have been carried out to determine the best conditions for sample storage and those that have been carried out show little difference in metabolic profiles between human urine samples stored at either −20 or −80 °C when analyzed by LC–MS[17]. However, pragmatically, the lowest available temperature should be used for sample storage and, despite a lack of evidence for large effects on metabolic profiles[17], the number of freeze–thaw cycles should be minimized.

As urine provides an excellent bacterial growth medium it may also be advisable to add an antibacterial agent such as sodium azide (0.05–0.1% wt/vol) to the sample to prevent, or inhibit, the microbial degradation of the sample. For animal samples, similar considerations

apply, and in the case of the collection of, e.g., 24 h rodent samples from animals housed in 'metabolism cages', specifically designed for such purposes, collection over ice (preferably dry ice) into sodium azide-containing collection vessels is preferred[16].

The containers used for both collection and storage (if different) should be carefully screened by MS (using the analytical method that will be used for the sample set) to ensure that they do not provide a source of unwanted contaminants (such as polyethylene glycol or plasticizers and so on). These contaminants can coelute with metabolites of interest, thus causing differential ion suppression, but they can also compromise the integrity of the obtained spectra of peaks of interest.

Column conditioning. Repeatable results are clearly key in order to obtain useful metabolic profile data. In considering any LC–MS analysis, there are three features of the analytical system that are required to be stable to achieve this; namely, retention time (RT), signal intensity and mass accuracy. A common observation when carrying out LC and UPLC–MS for global metabolic analysis has been that the first few injections of sample provide unrepresentative results, mainly because of small changes in both the chromatographic RT and signal intensity[48,49]. Usually after 5–10 injections of the matrix (in this case, urine), RTs stabilize as the column becomes 'conditioned' and the system then shows little variability through the remainder of the run. It is therefore good practice to run at least five pooled 'quality control' (QC) samples (see below) at the beginning of the run and use the data derived from these samples to demonstrate, postrun, system suitability. There are also good arguments to run suitable test mixtures prior to the run (see below), and also at the end of it (e.g., refs. 48,49), as these will give a rapid indication of (a) system suitability and (b) system stability, thereby providing an early alert to problems resulting from system contamination or instrument failure such as a decline in sensitivity, RT shifts or changes in mass accuracy. At the end of each run the column should be washed thoroughly with a strongly eluotropic solvent, e.g., methanol or acetonitrile, and the MS inlet and source meticulously cleaned before the next run to prevent the build up of contaminants and ensure continuing good performance. The run length permitted before cleaning will depend on the nature of the samples being analyzed, as this will have an effect on the contamination of the source.

Run order. When experimental samples are run in a time sequence, responses can depend on the run order as the source of the MS can become contaminated, leading to gradual changes in instrument sensitivity over time. Providing that these changes are not major, the subsequent data treatment will not be too adversely affected provided that careful randomization of the samples has been carried out to ensure that all of the experimental groups are affected to the same extent. This means that any subsequent statistical analysis of the data remains unbiased, thereby helping to ensure validity of the experiment. Ideally, the sample run order should be orthogonal to the samples to eliminate bias. One way to randomize the samples is to use a randomized block design, constructed to reduce noise or variance in the data. The samples are divided into relatively homogeneous subgroups or blocks and the desired experimental design is implemented within each of these blocks or subgroups[50]. The variability within each block should be less than the variability of the entire sample set and thus each estimate of the treatment effect within a block is more efficient than estimates across the entire sample.

Replicates. Replicate measurements can be included to enable good statistics at the end of the experiment to be demonstrated. Both technical (repeat injections of the same sample) and biological replicates (different samples measured under the same conditions) can be included. Technical replicates are useful for determining whether an outlier sample is actually biologically different or just part of the usual system variability. However, for a robust method, the QC samples (see below) should behave in exactly the same way as the test samples and, providing that the analytical data for these are satisfactory, the need for technical replicates is reduced.

Test mixtures and QC samples. A major challenge with LC–MS-based methods is, as indicated above, the potential for the characteristics of the analytical system to change with time during the analysis and it is incumbent on the investigator to put in place systems that enable data quality to be assessed. A variety of approaches have been advocated for ensuring that the results obtained from global metabolic profiles studies are valid, including the use of internal standards, test mixtures and QC samples[48,49,51–53]. Test mixtures, comprising a limited number of components and prepared from commercially available standards provide a rapid means of assessing gross performance characteristics (RT stability, peak shape, detector response and mass accuracy). These test mixture components can also be spiked into samples of the matrix of interest to avoid injections onto the column of a solution, which may 'wash' the column and thus remove the effects of the QC equilibration. A typical QC sample, for a small sample set, would be a pooled urine sample, prepared by mixing aliquots of the samples to be analyzed and therefore broadly representative of the whole sample set[41,48,49]. For larger, epidemiological or large multicenter clinical studies where thousands of samples collected over many months are involved, the approach of using a pooled sample made from the test samples themselves is clearly impractical. For these longer-term investigations, a bulk QC sample can be prepared from a representative subset of subjects, subaliquoted to minimize freeze-thaw cycle effects and stored frozen (at the lowest available temperature) until required. The QCs are then injected at regular intervals (i.e., every ten samples) throughout the analytical run to provide a set of data from which repeatability can be assessed as described in **Figure 4**.

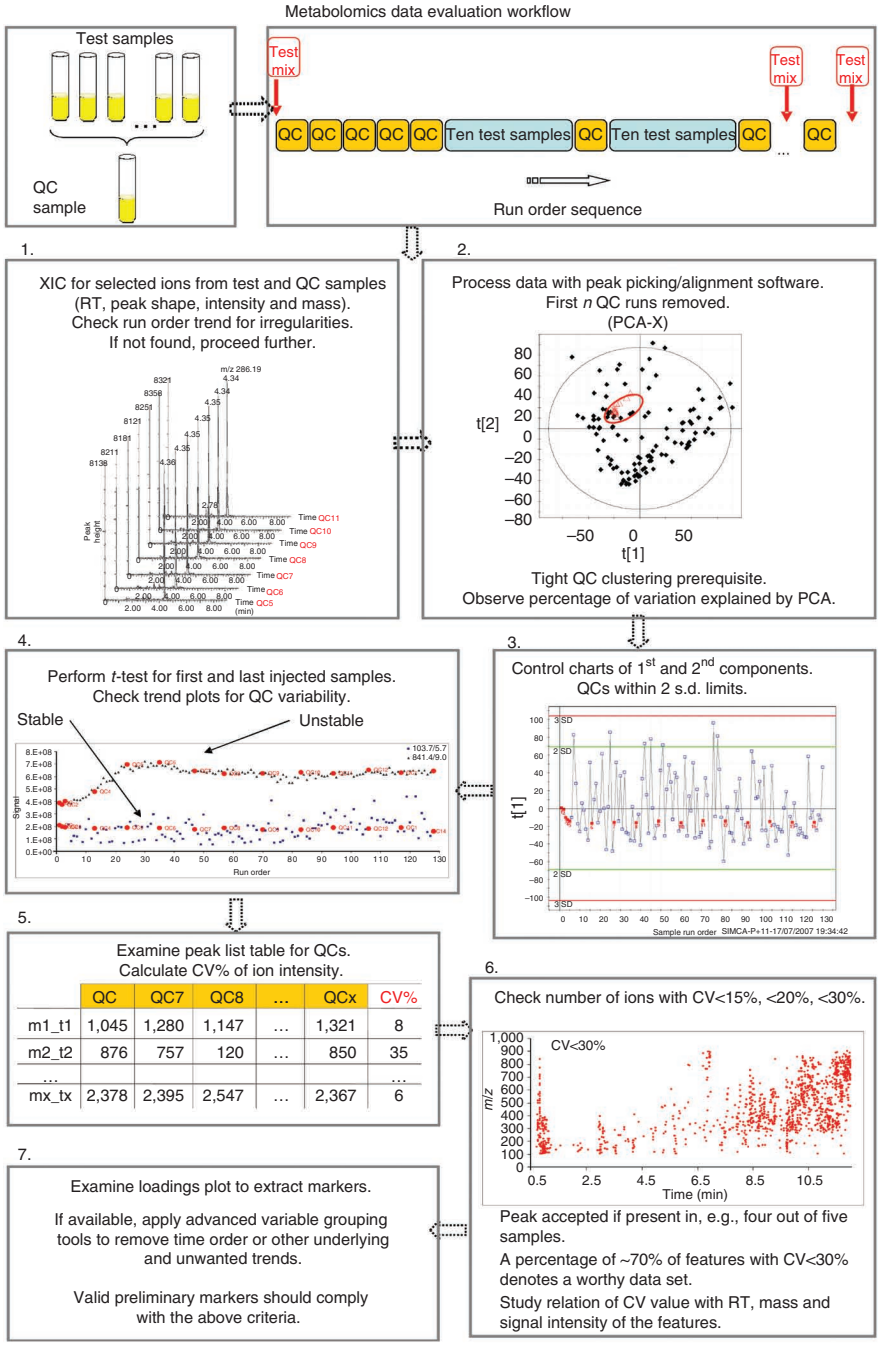
A typical run would be constructed as follows:

Test mix
↓
Conditioning QCs (5 – 10)
↓
QC
↓
Test samples (10)
↓
QC
↓
[Repeat × X]
↓
Test mix

For details regarding how to assess data quality see the ANTICIPATED RESULTS section.

Data mining. Chromatography–MS platforms can produce vast volumes of data for every study. Typically, raw data files range from 100 MB for each sample. LC–MS datasets are complex and thus require extensive preprocessing before statistical analysis. Peaks/metabolites need to be detected in the samples, matched or aligned across samples and then compared between samples. Peak alignment is a crucial step. Metabolite profiling studies may include many samples. Over the course of a run, peaks may shift because of factors such as changes in temperature (hence the need for temperature control of the column), mobile phase composition (it has been reported that peak shifts can be more pronounced as the proportion of organic solvent increases in the mobile phase) and sample pH in addition to column contamination[48,49,54].

Software for data preprocessing. Nowadays, data preprocessing software comes with the instrument software. These packages will incorporate the main aspects of preprocessing detailed above, as well as multivariate statistical capabilities in some cases. There are also several freeware packages available that are instrument independent and may be useful when comparing data from several platforms. These include XCMS[55], MZMine[56], MSFACTS[57], MATHDAMP[58] and MET-IDEA[59]. Some of these packages may be modified by the user, giving even greater flexibility in data analysis. The ultimate output can be termed either a 'metabolite' or 'marker' or 'feature' table. This can then be exported into software such as SIMCA or MATLAB for multivariate analysis such as principal components analysis (PCA) and partial least squares–discrimination analysis (PLS-DA). Important variable lists/loadings plots can give potential biomarkers, usually as *m/z*_RT pairs (**Fig. 5**). The time needed for data preprocessing depends on the size of the sample set being analyzed (both the number of samples and the size of the data files), the computational power and the software employed. Key issues when processing mass spectrometry data such as those obtained using this protocol include peak detection, alignment and normalization. An in-depth review of these data analysis challenges is beyond the scope of this protocol, but good reviews can be found in references 60,61.



	QC	QC7	QC8	...	QCx	CV%
m1_t1	1,045	1,280	1,147	...	1,321	8
m2_t2	876	757	120	...	850	35
...						...
mx_tx	2,378	2,395	2,547	...	2,367	6

Figure 4 | Flow chart of the validation guidelines followed.

Biomarker characterization using UPLC–MS. Once particular ions have been highlighted as being potential biomarkers, they must then be structurally characterized and identified so that they can be put into biological context (**Fig. 6**). The identification of biomarkers can be a significant challenge in MS-based metabolic profiling. The application of MS/MS can be used to provide structural information based on fragmentation, and accurate mass measurements can be used to generate probable empirical formulae. If carried out on a Q-ToF, high mass-accuracy measurements can be obtained for both parent and daughter ions, whereas the use of 'Orbitrap' or Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers[47,62–64] can provide even higher levels of mass accuracy. The application of MS^E, in which fragmentation information is obtained in the same run as scan data, requires less analysis time and smaller sample sizes. Such data can greatly reduce the metabolic 'search space' for unknowns,

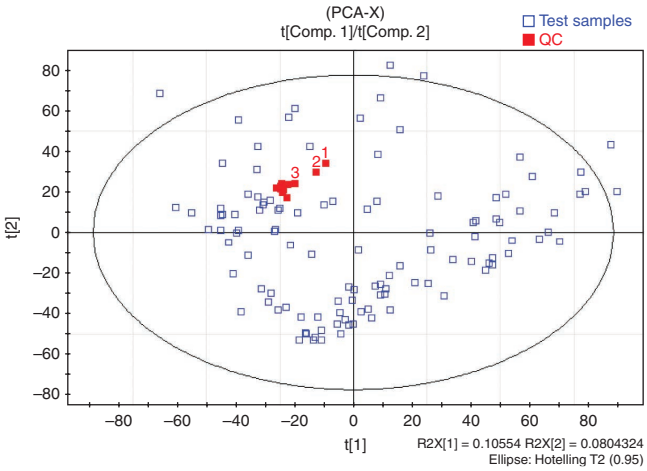


Figure 5 | A two-dimensional PCA scores plot (PC1 versus PC2) of human urine samples (blue) and QCs (red) obtained by UPLC–MS in positive ESI. The first three conditioning injections of the QC are numbered. The quality of the QC data can then also be estimated by looking at their variability with respect to run order as shown in **Figure 6**, which displays the first component t[1] as a plot versus the samples in run order, thereby showing t[1] as it evolves in time (with the 2 and 3 σ limits also shown). This clearly shows the essential stability of the QC samples through the whole of the run (29 h). This type of result also provides some assurance that there are no major run-order-related changes occurring as the analysis proceeds (reproduced with permission from ref. 49).

but are not guaranteed to provide unequivocal structure identification. For confirmation of identity, a comparison of RT and MS/MS fragmentation patterns with an authentic standard remains the 'gold standard'. The recently developed approach of Statistical HeterospectroscopY (SHY)[65] can combine efficiently (UPLC)–MS and NMR data (collected in series or parallel) to improve the discovery of robust markers[46,65]. Through SHY, both structural information and biological information can be obtained regarding metabolic pathway activities, as well as information on connectivities between pathways.

Databases for metabolite identification. Once potential biomarker candidates have been determined from the data analysis, they need to be identified. Researchers can consult online databases such as Chemspider, HMDB[66,67], METLIN[68] and MZedDB[69] as well as the KEGG database.

LC–NMR–MS for structural elucidation. Where a potential biomarker cannot be simply identified on the basis of the MS data acquired during the initial sample profiling, recourse to more detailed studies must be made. A typical structure elucidation protocol in clinical, biological and natural product research involves MS rapid screening and preliminary structure investigation, followed by supplementary NMR structure determination. If the target metabolite can be isolated by simple methods such as solid phase extraction (SPE)/chromatography (SPEC)[70,71], or 'preparative' HPLC, then NMR spectroscopy can further structural information that may enable identification. Generally, even if target compound isolation is not carried out, samples will need to be concentrated before performing the appropriate NMR spectroscopic analyses to compensate for the relative insensitivity of the technique. Alternatively, LC–NMR (or LC–SPE–NMR) may be employed to circumvent the need for previous isolation[72–75]. However, data correlation based on independent LC–MS and LC–NMR results from the same sample is sometimes difficult because of almost inevitable minor differences in the chromatographic separation obtained by the two systems. An obvious solution is the combination of MS and NMR into one integrated LC system as online-coupled LC–(SPE)–NMR–MS. This combination has been shown to be a powerful tool for the detection and identification of both known and, importantly, unknown compounds in complex samples, including urine[76–79].

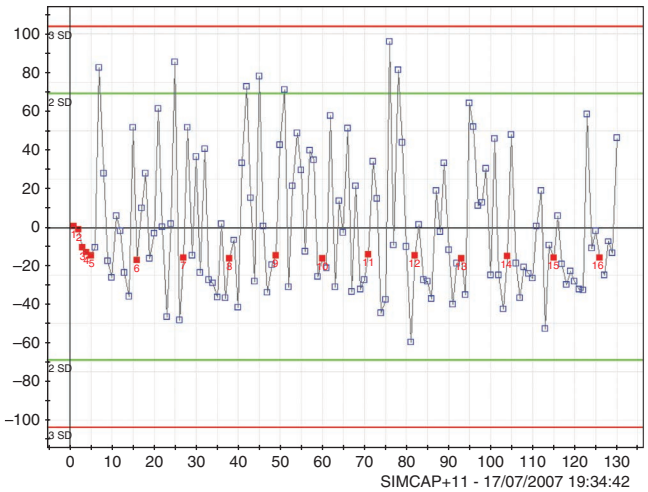


Figure 6 | PCA time series plot showing the first PC component (t[1] versus samples in run order). QCs are colored as red squares and test samples are colored in blue. *X* axis numbers represent sample number: 130 injections. *Y* axis is arbitrary (3 s.d.) (reproduced with permission from ref. 49).

Biomarker validation. The protocol described above is designed to enable the discovery and identification of potential biomarkers in urine. However, as with any method developed to study a very wide range of analytes, it cannot be expected to be optimized for any of them. Once compounds have been identified as potential biomarkers, then further investigations should be undertaken to develop fully validated analytical procedures to confirm that these analytes do indeed accurately reflect differences between control and test populations, including appropriate quantitative methods.

The following protocol details the metabolite profiling of urine samples by UPLC–MS. This protocol covers specific aspects of sample collection, storage and preparation. Details are provided on sample analysis, but reference to manufacturer's guidelines for instrument setup and operation are recommended at all times.

MATERIALS

REAGENTS

- Water (Sigma-Aldrich; LC–MS CHROMASOLV, FLUKA, cat. no. 39253-1L)
- Acetonitrile (Sigma-Aldrich) ! CAUTION Acetonitrile is highly flammable.
- Methanol (Sigma-Aldrich) ! CAUTION Methanol is highly flammable.
- Formic acid (Sigma-Aldrich; Fluka, cat. no. 94318-50ML-F) ! CAUTION Formic acid is corrosive and volatile.
- Isopropanol (Sigma-Aldrich) ! CAUTION Isopropanol is highly flammable.
- Leucine enkephalin acetate salt hydrate (Sigma-Aldrich, cat. no. L9133-25MG) (or alternative lock mass compound)
- Sodium formate (or alternative calibration solution)
- UPLC mobile phases (see REAGENT SETUP)
- Argon for applying gas to mass spectrometer collision cell
- Sodium azide ! CAUTION Sodium azide is highly toxic and is a heat-sensitive explosive in the solid state.

EQUIPMENT

- Acquity UPLC system (Waters) or similar (e.g., Agilent)
- Q-ToF mass spectrometer (Micromass) equipped with an ESI source and lockspray or similar (Agilent)
- Peek tubing
- Analytical columns (e.g., Acquity C18, HSS, HILIC or similar)
- Precolumn filters (Waters)
- Sep-Pak SPE C18 cartridges (Waters) or similar
- 96-well plates (350 µl volume) (Waters)
- Sealing cap mats (VWR)
- Maximum recovery vials with caps (Waters) or similar
- 1.5 and 2 ml Eppendorf tubes
- Solvent evaporator
- Ultrasonic bath
- Storage tubes
- Plastic bags
- Glass bottles
- Pipettes and pipette tips
- Software: Masslynx data management software 4.0 (Waters) or similar
  Microsoft Excel
  SIMCA-P or MATLAB software
  R and associated software packages

REAGENT SETUP

Urine samples Collect urine (either timed or 24 h collection) into suitable container and then subaliquot into labeled tubes/containers and store at the lowest available temperature (minimum −20 °C) until prepared for analysis as in Sample preparation section.

Preservatives For human and animal urine samples, add sodium azide to the sample after collection (to result in a total concentration of azide of min 0.05% wt/vol). For 24-h rodent samples from animals housed in metabolism cages, collect over ice/dry ice into vessels containing sodium azide.

Note: Timed collections: a midstream sample should be collected. 24 h collections: store the sample in the fridge between collections.

UPLC–MS mobile phases Prepare mobile phase A: 100% high-grade water with 0.1% formic acid and mobile phase B: 100% acetonitrile with 0.1% formic acid or 100% methanol with 0.1% formic acid. Prepare sufficient solutions to enable analysis of whole sample set. ! CAUTION All solutions should be prepared in a fume hood.

Leucine enkephalin lock mass solution Prepare a solution of leucine enkephalin in water:acetonitrile 50:50 to obtain a final concentration of 200 pg μl^{-1} or according to the manufacturer's specifications. Dilute appropriately for positive mode and negative ionization modes, as a more concentrated solution will be needed for negative mode. Prepare sufficient solutions to enable analysis of whole sample set. Store solution at 4 °C until use.

Sodium formate calibration solution To carry out instrument calibration, prepare a 0.1 mg ml^{-1} stock solution in water. Add 1 ml of stock solution to 9 ml isopropanol to give 0.01 mg ml^{-1} solution in 90% isopropanol and 10% water. Store at 4 °C. Or use alternative calibration solution at appropriate concentration.

Note: Alternative compounds can be used for the lock mass and/or calibration solution. Please follow the manufacturer's guidelines for concentration and storage conditions.

EQUIPMENT SETUP

UPLC–MS instrument setup Mass accuracy work on a ToF or Q-ToF MS requires calibration to be carried out before the instrument is used. Therefore, at the beginning of each sample set (or at alternative specified times), the instrument should be calibrated according to manufacturer guidelines. On the Q-Tof Premier and LCT Premier, a set-up wizard is used, but this procedure can also be performed manually. See UPLC-MS Data Acquisition. Perform additional instrument system checks if required, according to manufacturer guidelines.

Note: Prepare the mobile phases as in REAGENT SETUP. Prime system pump and tubing.

General maintenance of the system Cleaning: The source of MS can become contaminated during sample analysis, leading to changes in instrument sensitivity over time. Cleaning should be carried out according to manufacturer guidelines. The user may decide to clean the instrument at specific time points, i.e., after a well-plate or appropriate sample batch (we would recommend that the instrument was cleaned at the end of each batch).

Note: The cleaning regimen may be matrix dependent. The researcher should be aware that more concentrated urine samples may cause the source to become dirtier more quickly and so the cleaning regimen may need to be more stringent and/or frequent (see ? TROUBLESHOOTING).

Calibration: See manufacturer's guidelines.

PROCEDURE

Urine sample preparation ● TIMING 1–2 h

1| Prepare urine samples using option A or B as in the guidelines described below. Quantities apply to the above-described UPLC–MS conditions. Adjust the quantities accordingly, depending on different vendor requirements.

! CAUTION Take appropriate precautions when handling samples from diseased individuals.

(A) Centrifugation and dilution

(i) Centrifuge 60 µl urine at 10,000*g* for 10 min to remove particulates (4 °C).
(ii) Remove 50 µl and add to 100 µl water. Mix well.
(iii) Prepare samples into either 96-well plates or glass LC vials.
(iv) Proceed to Step 2, UPLC–MS data acquisition and preprocessing.

■ PAUSE POINT Prepared sample can be stored at −20 °C or below.

(B) SPE

(i) Centrifuge urine sample at 10,000*g* for 10 min to remove particulates (4 °C).
(ii) Condition and equilibrate sorbent according to the manufacturer's instructions (e.g., condition with 500 µl of MeOH and equilibrate with 500 µl of water).
(iii) Acidify sample if necessary (following the manufacturer's instructions).
(iv) Load sample onto sorbent (following the manufacturer's instructions).
(v) Wash according to manufacturer's instructions (e.g., 2% HCOOH or 5% NH_4OH in water).

(vi) Elute according to the manufacturer's instructions (e.g., MeOH followed by 2% HCOOH in MeOH or 5% NH_4OH in MeOH).
(vii) Evaporate both elution samples to dryness.
(viii) Reconstitute sample in water.
(ix) Prepare samples into either 96-well plates or glass LC vials. Proceed to Step 2 of UPLC–MS data acquisition and preprocessing.

Considerations for sample handling
Collection of urine—time points per 24 h collection
Type of collection and storage containers
Possible sources of contamination
Preservative used, e.g., sodium azide
Filtration step
Freeze–thaw cycles
Temperature of sample storage and autosampler temperature
Length of storage time and time in autosampler

■ **PAUSE POINT** Prepared sample can be stored at −20 °C or below.

UPLC–MS data acquisition and preprocessing ● **TIMING 12 min**

2| Centrifuge 96-well plate or vials at 10,000*g* for 5 min (4 °C).

3| Load 96-well plate or vials into autosampler maintained at 4 °C.

4| Select ESI ionization mode (positive or negative).

5| Carry out instrument setup i.e., (A) accurate mass and (B) calibration.

(A) Accurate mass

(i) Infuse appropriate concentration of leucine enkephalin (or alternative lockmass solution) into instrument. Follow set-up procedures.

(B) Calibration

(i) Infuse sodium formate solution (or alternative calibration solution) into the instrument. Follow setup procedures. As a general rule, the residual (in mDa) on each individual calibration point should be <1.5 mDa. Ideally, the majority of calibration points will have residuals of <0.5 mDa. A measure of the 'fit' of the calibration line to the experimental data is given in the error of the residual.

▲ **CRITICAL STEP** Ion counts must be below 200 counts per second in continuum mode for both options (A) and (B) to ensure proper accurate mass and calibration calculations to be performed. Adjust capillary voltage and cone voltage until criteria are filled. *Note*: With a setup wizard, this may be done automatically.

6| Select suitable gradient for sample, e.g., urine UPLC–MS 12 min run. See **Boxes 1** and **2**.
? TROUBLESHOOTING

7| Select suitable MS experimental parameters. See **Box 3**.
? TROUBLESHOOTING

8| Acquire data. Experimental parameters for reverse phase UPLC are given in **Box 1**; for HILIC in **Box 2**; and for mass spectrometry in **Box 3**.
? TROUBLESHOOTING

UPLC–MS data processing

Data preprocessing and peak alignment ● **TIMING ~2 h for a 96-well plate**

9| Use appropriate software to extract and align all mass signals above a defined threshold. A signal-to-noise threshold of 3 is typically used in analytical chemistry. Often there are regions of the chromatogram that do not contain useful data, such as the solvent front at the very start and also the last portion of the chromatogram (where the column is re-equilibrating). Most software will enable the user to choose the region of the chromatogram to process. For example, Markerlynx software within the Masslynx data management software will output the results as a markers table. This contains *m/z*, RT and intensity information for all detected features. There is also the capability to carry out multivariate analysis, such as PCA. The outputted data from most software can be transferred into MS Excel for basic statistical analysis or SIMCA-P/MATLAB for multivariate statistical analysis.

BOX 1 | REVERSED PHASE ULTRA-PERFORMANCE LIQUID CHROMATOGRAPHY (UPLC)–MASS SPECTROMETRY (MS) GRADIENT (12 MIN RUN)

Column: 2.1 × 100 mm (1.7 µm) HSS T3 Acquity (Waters)
Injection volume: 5 µl
Flow rate: 0.5 ml min^{-1}
Sample temperature, 4 °C; column temperature, 40 °C
Mobile phases:
A: 0.1% formic acid in water
B: 0.1% formic acid in acetonitrile

Time (min)	A (%)	B (%)
0	99	1
1	99	1
3	85	15
6	50	50
9	5	95
10	5	95
10.1	99	1

BOX 2 | HILIC ULTRA-PERFORMANCE LIQUID CHROMATOGRAPHY (UPLC)–MASS SPECTROMETRY (MS) GRADIENT (12 MIN RUN)

Column: 2.1 × 100 mm (1.7 µm) BEH HILIC Acquity (Waters)
Injection volume: 5 µl
Flow rate: 0.4 ml min^{-1}
Sample temperature, 4 °C; column temperature, 40 °C
Mobile phases:
A: 95% acetonitrile, 5% ammonium acetate (10 mM final concentration)
B: 50% acetonitrile, 50% ammonium acetate (10 mM final concentration)

Time (min)	A (%)	B (%)
0.0	99	1
1.0	99	1
12.0	0	100
12.1	99	1
15	99	1

BOX 3 | MASS SPECTROMETRY (MS) SETUP

Perform system set-up and calibration as described. The procedures may vary depending on instrument type.
Set parameters including:
Capillary voltage, e.g., 3.2 kV electrospray ionization (ESI)+, 2.4 kV ESI–
Source temperature, e.g., 120 °C
Desolvation temperature, e.g., 350 °C
Cone gas flow, e.g., 25 liter h^{-1}
Desolvation gas flow, e.g., 900 liter h^{-1}
Note: These parameters are guidelines only, based on an ultra-performance liquid chromatography (UPLC) flow rate of 500 µl min^{-1}.

● TIMING

Step 1, Urine sample preparation: 1–2 h
Steps 2–8, UPLC–MS data acquisition and preprocessing: 12 min
Step 9, UPLC–MS data processing: ~2 h for a 96-well plate

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
Chromatography 6,8	High back-pressure	Blockage in capillary transfer line/injection loop/column frit, due to particulate matter from sample	Remove the probe from source and flush (neat formic acid may aid in clearing the blockage) Clean or replace column/loop
	Poor peak shape	Column degradation Overloading of sample	Clean or replace column Dilute sample/improve sample preparation
	No/few peaks	Failed injection/needle blockage Sample concentration too low	Flush needle Reinject sample Reprepare/concentrate sample
	Drop in baseline	Ion suppression, perhaps due to high salt levels in sample	Improve sample preparation (e.g., perform solid phase extraction (SPE)) Optimize chromatographic gradient to minimize coelution of peaks if possible
	Carry-over	Appropriate wash solvents not selected Chromatography not optimized	Choose suitable wash solvents Optimize chromatographic gradient
	Loss of sensitivity	Matrix suppression Poor recovery	Improve sample preparation (e.g., perform SPE)
Mass spectrometry 7,8	Unsteady beam	Capillary/sample cone voltages not optimal Capillary is protruding too far from end of probe Probe is too far into source Liquid chromatography (LC) solvent flow is not correct/steady Solvents have been adequately degassed Desolvation/nebulizer gas flow is not steady Desolvation temperature is not set correctly for liquid flow rate used	Tune sample cone and capillary Change length of capillary protruding from probe Move probe away from source Degas solvent, reset and remeasure the flow rate Check and adjust nitrogen supply pressure Check manual for guidelines Check and adjust desolvation temperature Check manual for guidelines
	Loss of sensitivity	Ion source is dirty	Clean the source according to manufacturer guidelines
	High chemical or electronic noise levels	Signal threshold set too low Detector damaged and producing micro discharges	Reduce detector voltage

ANTICIPATED RESULTS

Criteria for assessing data quality

Test mixture assessment. The data obtained from the test mixtures (RT, peak shape, signal intensity, mass accuracy and so on) can be used to rapidly determine if the instrumental setup is suitable for the analysis (through the first injection) and to determine if major changes have occurred during the analysis (through the last injection). If major changes have occurred, this would automatically invalidate the results obtained for the samples. However, even supposing that the test mixture data are found to be acceptable, these results do not validate those for the samples themselves that, because of their much greater complexity, may be more variable. What the test mixture data provide is an assurance that it is worth proceeding to evaluate the results from the QCs.

QC sample assessment. There are several steps in the analysis of the QC data, beginning with the simple multivariate approach of PCA. Ideally, if the analysis has been carried out well, the PCA should show that the first 'conditioning' injections of the QC sample 'track' towards the main group of QC samples as the analytical system equilibrates. After confirming that the system has been adequately stabilized, the data for these initial 'blank' injections of the QC samples are discarded and the data derived from the in-run QC samples can then be scrutinized in detail. Typical UPLC–MS data for human urine (positive ESI) are shown in **Figure 5,** where the first few 'blank' or conditioning QC injections are highlighted. The relatively tight clustering of the main group of QC samples suggests that the data are worth further study. It is worth mentioning that the number of QC samples required to condition the column is highly dependent on the matrix being analyzed, e.g., this number would be much higher for serum samples[19]. Where problems have occurred during the run, this results in the distribution of QC over a large area such that it is quite clear that no usable data can be derived from this analysis.

Armed with the information from the test mixture and PCA of the QC samples, a more detailed assessment of data quality can be carried out (**Fig. 6**), such as examination of RT stability for selected ions present in the QCs covering a range of RTs. Typically for UPLC, RT variation is usually negligible, with coefficient of variation (CV)% values <1%. Similar examination of peak height/areas for these ions should also show good repeatability (although for peak height/area there is always a dependence on ion intensity with the more intense ions giving generally better repeatability). Mass accuracy should also show lower variability for these ions. At this point the analytical variability in the processed data from the whole QC dataset can be examined for the evidence of good overall repeatability with a view to then moving into the test set data for biomarker detection.

In terms of accepting individual ions as potential markers there is, as yet, no consensus as to what criteria should apply. However, for conventional bioanalysis, the Food and Drug Administration (FDA) recommends that a CV of 15% of the nominal value be applied (except for concentrations close to the limit of quantification, where 20% is considered to be adequate)[52] and for biomarkers an upper limit of 30% can be accepted[48]. We therefore recommend that potential marker ions be assessed using this approach and that highly variable ions (CV of greater than 30%) should be rejected as unacceptable for the purpose of biomarker discovery. A suggested workflow for accepting LC–MS-generated metabolic profiling data as fit for in-depth, statistical analysis as part of biomarker discovery is shown in **Figure 4**. In our view, failure to pass any of these stages should trigger a reanalysis of the sample set.

Sample stability during analysis

In any metabolic profiling study containing more than a handful of samples, it is likely that the time from the first to the last analysis will be 24 h or longer, meaning that samples will be present in the autosampler (albeit at <4 °C) for some time before analysis. Clearly, sample degradation over this period would adversely affect the subsequent interpretation of the data. In addition, if for any reason the run should fail, a decision may need to be taken to either reanalyze the samples or prepare a new batch (which may be difficult in the case of limited samples). The short-term stability of prepared urine samples in an autosampler at 4 °C was investigated by the daily reanalysis of the aliquots of the urine QC sample over 6 d. This showed that the QCs appeared to be stable for up to 48 h, after which changes were noted suggesting that this was the maximum time that samples should be kept under such conditions[17].

ACKNOWLEDGMENTS The author E. Want acknowledges funding support from Waters Corporation. H. Gika acknowledges support from the European Committee through a European Reintegration Grant (ERG 202132).

AUTHOR CONTRIBUTIONS All authors contributed extensively to the work presented in this paper.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Published online at http://www.natureprotocols.com/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Nicholson, J.K., Lindon, J.C. & Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189 (1999).

2. Nicholson, J.K., Connelly, J., Lindon, J.C. & Holmes, E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **1**, 153–161 (2002).
3. Fiehn, O. Metabolomics—the link between genotypes and phenotypes. *Plant. Mol. Biol.* **48**, 155–171 (2002).
4. Nicholson, J.K. & Lindon, J.C. Systems biology: metabonomics. *Nature* **455**, 1054–1056 (2008).
5. Wishart, D.S. Applications of metabolomics in drug discovery and development. *Drugs R. D.* **9**, 307–322 (2008).
6. Clarke, C.J. & Haselden, J.N. Metabolic profiling as a tool for understanding mechanisms of toxicity. *Toxicol. Pathol.* **36**, 140–147 (2008).
7. Bollard, M.E. *et al.* Comparative metabonomics of differential hydrazine toxicity in the rat and mouse. *Toxicol. Appl. Pharmacol.* **204**, 135–151 (2005).
8. Lindon, J.C., Holmes, E. & Nicholson, J.K. Metabonomics in pharmaceutical R&D. *FEBS J.* **274**, 1140–1151 (2007).
9. Coen, M., Holmes, E., Lindon, J.C. & Nicholson, J.K. NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology. *Chem. Res. Toxicol.* **21**, 9–27 (2008).
10. Hinkelbein, J. *et al.* Alterations in cerebral metabolomics and proteomic expression during sepsis. *Curr. Neurovasc. Res.* **4**, 280–288 (2007).
11. Bertini, I. *et al.* The metabonomic signature of celiac disease. *J. Proteome. Res.* **8**, 170–177 (2009).
12. Gowda, G.A. *et al.* Metabolomics-based methods for early disease diagnostics. *Expert. Rev. Mol. Diagn.* **8**, 617–633 (2008).
13. Lenz, E.M. *et al.* Metabonomics, dietary influences and cultural differences: a ^{1}H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *J. Pharm. Biomed. Anal.* **36**, 841–849 (2004).
14. Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **4**, e1000282 (2008).
15. Holmes, E. *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453**, 396–400 (2008).
16. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2**, 2692–2703 (2007).
17. Gika, H.G., Theodoridis, G. & Wilson, I.D. Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine. Sample stability under different handling and storage conditions for metabonomics studies. *J. Chromatogr. A* **1189**, 314–322 (2008).
18. 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).
19. Michopoulos, F., Lai, L., Gika, H., Theodoridis, G. & Wilson, I.D. UPLC-MS-based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. *J. Proteome. Res.* **8**, 2114–2121 (2009).
20. Zelena, E. *et al.* Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal. Chem.* **81**, 1357–1364 (2009).
21. Barton, R.H., Nicholson, J.K., Elliott, P. & Holmes, E. High-throughput 1H 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).
22. Lauridsen, M., Hansen, S.H., Jaroszewski, J.W. & Cornett, C. Human urine as test material in 1H NMR-based metabonomics: recommendations for sample preparation and storage. *Anal. Chem.* **79**, 1181–1186 (2007).
23. Matuszewski, B.K., Constanzer, M.L. & Chavez-Eng, C.M. Matrix effect in quantitative LC/MS/MS analyses of biological fluids: a method for determination of finasteride in human plasma at picogram per milliliter concentrations. *Anal. Chem.* **70**, 882–889 (1998).
24. Gangl, E.T., Annan, M.M., Spooner, N. & Vouros, P. Reduction of signal suppression effects in ESI-MS using a nanosplitting device. *Anal. Chem.* **73**, 5635–5644 (2001).
25. Gustavsson, S.A., Samskog, J., Markides, K.E. & Långström, B. Studies of signal suppression in liquid chromatography-electrospray ionization mass spectrometry using volatile ion-pairing reagents. *J. Chromatogr. A* **937**, 41–47 (2001).
26. Jellum, E. Profiling of human body fluids in healthy and diseased states using gas chromatography and mass spectrometry, with special reference to organic acids. *J. Chromatogr.* **143**, 427–462 (1977).
27. Taylor, N.F. Urinary steroid profiling. *Methods Mol. Biol.* **324**, 159–175 (2006).
28. Halket, J.M *et al.* Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Botany* **56**, 219–243 (2005).
29. New, L.S. & Chan, E.C. Evaluation of BEH C18, BEH HILIC, and HSS T3 (C18) column chemistries for the UPLC-MS-MS analysis of glutathione, glutathione disulfide, and ophthalmic acid in mouse liver and human plasma. *J. Chromatogr. Sci.* **46**, 209–214 (2008).
30. Wilson, I.D. *et al.* High resolution 'ultra performance' liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome. Res.* **4**, 591–598 (2005).
31. Tolstikov, V.V. & Fiehn, O. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* **301**, 298–307 (2002).
32. Idborg, H., Zamani, L., Schuppe-Koistinen, I. & Jacobsson, S. Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry. *J. Chromatogr. B* **828**, 9–13 (2005).
33. Cubbon, S., Bradbury, T., Wilson, J. & Thomas-Oates, J. Hydrophilic interaction chromatography for mass spectrometric metabonomic studies of urine. *Anal. Chem.* **79**, 8911–8918 (2007).
34. Gika, H.G., Theodoridis, G.A. & Wilson, I.D. Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine. *J. Sep. Sci.* **31**, 1598–1608 (2008).
35. Plumb, R.S. *et al.* A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. *Analyst* **130**, 844–849 (2005).
36. Kind, T., Tolstikov, V., Fiehn, O. & Weiss, R.H. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal. Biochem.* **363**, 185–195 (2007).
37. Guy, P.A., Tavazzi, I., Bruce, S.J., Ramadan, Z. & Kochhar, S. Global metabolic profiling analysis on human urine by UPLC-TOFMS: issues and method validation in nutritional metabolomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **871**, 253–260 (2008).
38. Plumb, R.S. *et al.* Generation of ultrahigh peak capacity LC separations via elevated temperatures and high linear mobile-phase velocities. *Anal. Chem.* **78**, 7278–7283 (2006).
39. Gika, H.G., Theodoridis, G., Extance, J., Edge, A.M. & Wilson, I.D. High temperature-ultraperformance liquid chromatography–mass spectrometry for the metabonomic analysis of Zucker rat urine. *J. Chrom. B.* **871**, 279–287 (2008).
40. Lenz, E.M., Bright, J., Knight, R., Wilson, I.D. & Major, H. A metabonomic investigation of the biochemical effects of mercuric chloride in the rat using ^{1}H NMR and HPLC-TOF/MS: time dependant changes in the urinary profile of endogenous metabolites as a result of nephrotoxicity. *Analyst* **129**, 535–541 (2004).
41. Nordström, A., Want, E., Northen, T., Lehtiö, J. & Siuzdak, G. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. *Anal. Chem.* **80**, 421–429 (2008).
42. Leandro, C.C., Hancock, P., Fussell, R.J. & Keely, B.J. Ultra-performance liquid chromatography for the determination of pesticide residues in foods by tandem quadrupole mass spectrometry with polarity switching. *J. Chromatogr. A* **1144**, 161–169 (2007).
43. Want, E.J., Nordström, A., Morita, H. & Siuzdak, G. From exogenous to endogenous: the inevitable imprint of mass spectrometry in metabolomics. *J. Proteome. Res.* **6**, 459–468 (2007).
44. Plumb, R.S. *et al.* UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* **20**, 1989–1994 (2006).
45. Bateman, K.P. *et al.* MSE with mass defect filtering for *in vitro* and *in vivo* metabolite identification. *Rapid Commun. Mass Spectrom.* **21**, 1485–1496 (2007).
46. Crockford, D.J. ^{1}H NMR and UPLC-MS(E) statistical heterospectroscopy: characterization of drug metabolites (xenometabolome) in epidemiological studies. *Anal. Chem.* **80**, 6835–6844 (2008).
47. Perry, R.H., Cooks, R.G. & Noll, R.J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom. Rev.* **27**, 661–699 (2008).
48. Gika, H.G., Theodoridis, G.A., Wingate, J.E. & Wilson, I.D. Within day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *J. Proteome. Res.* **6**, 3291–3303 (2007).
49. Gika, H.G., Macpherson, E., Theodoridis, G. & Wilson, I.D. Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *J. Chromatogr. B* **871**, 299–305 (2008).

50. Baker, J.M. *et al.* A metabolomic study of substantial equivalence of field-grown genetically modified wheat. *Plant Biotechnol. J.* **4**, 381–392 (2006).
51. Sangster, T., Major, H., Plumb, R., Wilson, A.J. & Wilson, I.D. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. *Analyst* **131**, 1075–1078 (2006).
52. *FDA Guidance for Industry, Bioanalytical method Validation*, Food and Drug Administration, Centre for Drug Evaluation and Research (CDER), May 2001.
53. Viswanathan, C.T. *et al.* Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *Pharm. Res.* **24**, 1962–1973 (2007).
54. Pham-Tuan, H., Kaskavelis, L., Daykin, C.A. & Janssen, H. Method development in high-performance liquid chromatography for high-throughput profiling and metabonomic studies of biofluid samples. *J. Chromatogr. B* **789**, 283–301 (2003).
55. 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).
56. Katajamaa, M., Miettinen, J. & Oresic, M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **22**, 634–636 (2006).
57. Duran, A.L., Yang, J., Wang, L.J. & Sumner, L.W. Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). *Bioinformatics* **19**, 2283–2293 (2003).
58. Baran, R. *et al.* MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics* **7**, 530–538 (2006).
59. Broeckling, C.D., Reddy, I.R., Duran, A.L., Zhao, X.C. & Sumner, L.W. MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. *Anal. Chem.* **78**, 4334–4341 (2006).
60. Katajamaa, M. & Oresic, M.J. Data processing for mass spectrometry-based metabolomics. *Chromatogr. A* **1158**, 318–328 (2007).
61. Sumner, L.W., Urbanczyk-Wochniak, E. & Broeckling, C.D. Metabolomics data analysis, visualization, and integration. *Methods Mol. Biol.* **406**, 409–436 (2007).
62. Ruan, Q. *et al.* An integrated method for metabolite detection and identification using a linear ion trap/Orbitrap mass spectrometer and multiple data processing techniques: application to indinavir metabolite detection. *J. Mass. Spectrom.* **43**, 251–261 (2008).
63. Zhang, N.R. *et al.* Quantitation of small molecules using high-resolution accurate mass spectrometers—a different approach for analysis of biological samples. *Rapid Commun. Mass Spectrom.* **23**, 1085–1094 (2009).
64. Ohta, D., Shibata, D. & Kanaya, S. Metabolic profiling using Fourier-transform ion-cyclotron-resonance mass spectrometry. *Anal. Bioanal. Chem.* **389**, 1469–1475 (2007).
65. Crockford, D.J. *et al.* Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies. *Anal. Chem.* **78**, 363–371 (2006).
66. Wishart, D.S. *et al.* HMDB: the human metabolome database. *Nucleic Acids Res.* **35** (Database issue): D521–D526 (2007).
67. Wishart, D.S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37** (Database issue): D603–D610 (2009).
68. Smith, C.A. *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747–751 (2005).
69. Draper, J. *et al.* Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour 'rules'. *BMC Bioinformatics* **10**, 227 (2009).
70. Wilson, I.D. & Nicholson, J.K. Solid-phase extraction chromatography and nuclear magnetic resonance spectroscopy for the identification and isolation of drug metabolites in urine. *Anal. Chem.* **59**, 2830–2832 (1987).
71. Baranyi, M., Milusheva, E., Vizi, E.S. & Sperlágh, B. Chromatographic analysis of dopamine metabolism in a Parkinsonian model. *J. Chromatogr. A* **1120**, 13–20 (2006).
72. Gavaghan, C.L. *et al.* Directly coupled high-performance liquid chromatography and nuclear magnetic resonance spectroscopy with chemometric studies on metabolic variation in Sprague-Dawley rats. *Anal. Biochem.* **291**, 245–252 (2001).
73. Spraul, M., Hofmann, M., Dvortsack, P., Nicholson, J.K. & Wilson, I.D. High-performance liquid chromatography coupled to high-field proton nuclear magnetic resonance spectroscopy: application to the urinary metabolites of ibuprofen. *Anal. Chem.* **65**, 327–330 (1993).
74. Dear, G.J. *et al.* Mass directed peak selection, an efficient method of drug metabolite identification using directly coupled liquid chromatography-mass spectrometry-nuclear magnetic resonance spectroscopy. *J. Chromatogr. B Biomed. Sci. Appl.* **748**, 281–293 (2000).
75. Lin, Y., Schiavo, S., Orjala, J., Vouros, P. & Kautz, R. Microscale LC-MS-NMR platform applied to the identification of active cyanobacterial metabolites. *Anal. Chem.* **80**, 8045–8054 (2008).
76. Shockcor, J.P. *et al.* Combined HPLC, NMR spectroscopy, and Ion-trap mass spectrometry with application to the detection and characterisation of xenobiotic and endogenous metabolites in human urine. *Anal. Chem.* **68**, 4431–4435 (1996).
77. Nicholls, A.W, Wilson, I.D., Godejohann, M., Nicholson, J.K. & Shockcor, J.P. Identification of phenacetin metabolites in human urine after administration of phenacetin-C2H3: measurement of futile metabolic deacetylation via HPLC/MS-SPE-NMR and HPLC-ToF MS. *Xenobiotica* **36**, 615–629 (2006).
78. Spraul, M. *et al.* Advancing NMR sensitivity for LC-NMR-MS using a cryoflow probe: application to the analysis of acetaminophen metabolites in urine. *Anal. Chem.* **75**, 1536–1541 (2003).
79. Shockcor, J.P., Unger, S.E., Savina, P., Nicholson, J.K. & Lindon, J.C. Application of directly coupled LC-NMR-MS to the structural elucidation of metabolites of the HIV-1 reverse-transcriptase inhibitor BW935U83. *J. Chromatogr. B Biomed. Sci. Appl.* **748**, 269–279 (2000).