A practical approach to detect unique metabolic patterns for personalized medicine†

Jennifer M. Johnson,[abc] Tianwei Yu,[d] Frederick H. Strobel[e] and Dean P. Jones*[ac]

Received 20th May 2010, Accepted 30th July 2010
DOI: 10.1039/c0an00333f

Information-rich technologies have advanced personalized medicine, yet obstacles limit measurement of large numbers of chemicals in human samples. Current laboratory tests measure hundreds of chemicals based upon existing knowledge of exposures, metabolism and disease mechanisms. Practical issues of cost and throughput preclude measurement of thousands of chemicals. Additionally, individuals are genetically diverse and have different exposures and response characteristics; some have disease mechanisms that have not yet been elucidated. Consequently, methods are needed to detect unique metabolic characteristics without presumption of known pathways, exposures or disease mechanisms, *i.e.*, using a top-down approach. In this report, we describe profiling of human plasma with liquid chromatography (LC) coupled to Fourier-transform mass spectrometry (FTMS). FTMS is a high-resolution mass spectrometer providing mass accuracy and resolution to discriminate thousands of *m*/*z* features, which are peaks defined by *m*/*z*, retention time and intensity. We demonstrate that LC-FTMS detects 2000 *m*/*z* features in 10 min. These features include known and unidentified chemicals with *m*/*z* between 85 and 850, most with <10% coefficient of variation. Comparison of metabolic profiles for 4 healthy individuals showed that 62% of the *m*/*z* features were common while 10% were unique and 770 discriminated the individuals. Because the simple one-step extraction and automated analysis is rapid and cost-effective, the approach is practical for personalized medicine. This provides a basis to rapidly characterize novel metabolic patterns which can be linked to genetics, environment and/or lifestyle.

Introduction

Environmental influences, including diet, chemical exposures and health behavior, are highly variable among individuals and contribute significantly to human health risks.[1,2] Wild[3] emphasized the need to complement the human genome project with an effort to systematically characterize the lifetime spectrum of exposures (the "Exposome") which impact individual disease risk. Practical metabolic profiling methods for such characterization of exposures and consequent metabolic responses are not available. This limits progress in understanding individual health risks and, importantly, also limits the promises of personalized medicine.[4]

Several characteristics are needed for metabolic profiling in personalized medicine. An important need is rapid analysis, similar to the one minute per sample goal of genomics.[5] Routine clinical use also requires sampling to be simple and processing to be automatable. Methods need to be sensitive enough to detect thousands of diverse chemicals over a broad range of concentrations, and measurements must also be reproducible in quantification. A cost-effective and affordable platform employing these capabilities would provide a high-performance metabolic profiling method that is practical for routine use in human health assessment.

The profile of metabolites in human plasma reflects the product of all environmental influences, in addition to genetics, epigenetics and proteomics, affecting health outcome and disease state.[3,6] Many platforms are available to provide targeted analyses of hundreds of biomolecules, and these have been used extensively to characterize differences associated with exposures and disease.[7,8] Current profiling methods are powerful but are limiting in one or more of the key needs for high-performance metabolic profiling. Methods using nuclear magnetic resonance spectroscopy are often simple, quantitative and have a reasonable throughput, but are limited by sensitivity. On the other hand, mass spectrometry offers measurements with increased sensitivity and broader dynamic range but usually needs to be coupled to a separation technique such as chromatography or electrophoresis.

The most common method for metabolic profiling uses liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS).[9] Chromatography is used to fractionate complex biological mixtures, ion fragmentation is used for identification of individual chemicals, and stable isotopic dilution is used for quantification.[9] This approach can provide accurate quantification of hundreds of known chemicals in a single biological sample. However, LC-MS/MS is not practical for analysis of thousands of chemicals in a short analysis time. Instrument

[a]*Department of Medicine, Division of Pulmonary, Allergy and Critical Care, Emory University, Atlanta, GA, 30322, USA. E-mail: dpjones@emory.edu; Fax: +1 404 712 2974; Tel: +1 404 727 5970*
[b]*Nutrition and Health Sciences Graduate Program, Emory University, Atlanta, GA, 30322, USA*
[c]*Clinical Biomarkers Laboratory, Emory University, Atlanta, GA, 30322, USA*
[d]*Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, 30322, USA*
[e]*Mass Spectrometry Center, Emory University, Atlanta, GA, 30322, USA*
† Electronic supplementary information (ESI) available: Additional figures showing analytic and biologic variability; table of 1464 endogenous human metabolites and 1769 dipeptides and tripeptides. See DOI: 10.1039/c0an00333f

sensitivity is inversely proportional to the number of compounds that need identification and quantification.[10] Additionally, simultaneous fragmentation of chemicals with similar nominal masses produces a complex mixture of product ions, which can be difficult to interpret. With lower mass resolution instruments, the only effective solution is to improve separation prior to mass spectrometry. While ultra-high performance liquid chromatography (UPLC) has provided a substantial improvement in capabilities,[11] the time required to separate thousands of chemicals for MS/MS remains a limit to this approach.

Quantification by stable isotopic dilution is also a limitation to increasing the number of chemicals measured in a biological sample by MS/MS. This internal standardization approach quantifies a chemical in terms of ion intensity of a product ion relative to the ion intensity of that product ion produced from a known amount of a stable isotopic form of the same chemical. However, stable isotopic forms are not readily available for all metabolites. When available, stable isotopic chemicals are expensive so that increasing the number of internal standards to thousands becomes cost prohibitive. Furthermore, in personalized medicine, there is an expectation that uncharacterized or unanticipated chemicals will be present in some individuals, creating an analytic need not addressed by targeted analysis. Thus, for high performance metabolic profiling in personalized medicine, new approaches are needed which avoid or minimize the limitations associated with separation, fragmentation and standardization procedures of LC-MS/MS.

During the past several years, high resolution mass spectrometers have become available[12] which can overcome the requirements for lengthy separation and fragmentation. These instruments provide sufficient mass accuracy to allow prediction of elemental composition without fragmentation for small molecules.[12] Consequently, coeluting chemicals can be distinguished so long as they are not isomers, thereby minimizing separation requirements. Three high-sensitivity instruments with resolving power greater than 10 000 are available for use with atmospheric ionization techniques, time-of-flight (TOF) and Fourier-transform (FT) [ion cyclotron resonance (ICR) and Orbitrap].[12] These instruments are commonly used for proteomics, including studies of covalent modifications of peptides and proteins, imaging, and mechanistic studies involving isotope exchange. The high resolution instruments are also being used for small molecule studies such as analysis of drug metabolism, metabolic disease, and environmental chemistry.[13–15] In such applications, the characteristics of the instruments determine the extent of chromatographic separation and time needed to allow detection of thousands of chemicals in a single analysis. The TOF instruments scan rapidly but have less resolving power than the FT-ICR and Orbitrap instruments.[12] This places a greater demand on the chromatographic separation for TOF instruments. Resolving power for the FT-ICR and Orbitrap is increased by increasing scan time, but an increase in scan time decreases the number of spectra over a chromatographic peak. In principle, either of these instruments could be optimized to detect thousands of chemicals, with the expectation that a commercially available Orbitrap would take approximately twice as long to achieve the same resolution as an FT-ICR. A new Orbitrap has also been described that has capabilities similar to the FT-ICR.[16] Routine application of such a profiling approach could provide knowledge of the relative abundance of chemicals in biological samples, such as plasma, defined by accurate *m*/*z*. While not achieving the same rigor in quantification as obtained by LC-MS/MS, the method could be adapted with surrogate standardization[17] or other normalization procedure to generate systematic description of chemical profiles for use in predictive and personalized medicine.

In the present study, we developed a rapid LC-FTMS method that relies on the high mass accuracy and mass resolution of the FTMS to generate a metabolic profile of human plasma. The method was designed to use a simple, one-step sample preparation and automated analysis. A self-adjusting peak identification and quantification method[18] was used to extract information on *m*/*z* features, which are peaks defined by *m*/*z*, intensity, and retention time. The results show that the LC-FTMS method is suitable to measure approximately 2000 known and unidentified chemicals in a 10 μL extract of human plasma in 10 min. This high-performance metabolic profiling method provides a basis to characterize novel metabolic patterns and individual exposures which can be linked to genetics, epigenetics, environment and/or lifestyle for personalized medicine.

Experimental

Materials

Sodium bicarbonate, heparin, acetonitrile (HPLC grade), formic acid (puriss. p.a. 98%), water (HPLC grade) and amino acid standard mix (A-6407) were from Sigma-Aldrich (St Louis).

Human subjects

This study was approved by the Emory University Institutional Review Board (Protocol #581-2006). Four subjects were recruited from Emory University and were self-described as healthy. There were no inclusion or exclusion criteria for participation other than voluntary participation, informed consent and self-described as healthy. Demographic information and personal health information were not collected. Each participant gave his/her informed consent prior to inclusion in the study. ESI† relevant to the question of variation within an individual over time were from a separate study of sulfur amino metabolism approved by the Emory University Institutional Review Board (Protocol #1098-2003); details of this protocol have been published.[19]

Sample collection and preparation

Blood was collected with a butterfly needle and syringe and dripped into a heparinized microcentrifuge tube; samples were centrifuged for 2 min at 13 000*g* and plasma aliquots were separated and stored at −80 °C.[20] For analysis, samples were thawed and acetonitrile was added (2 : 1, v/v) to precipitate protein;[21] samples were centrifuged at 13 000*g* for 5 min and supernatant was maintained at 4 °C until injection onto the LC-FTMS (Ultimate®, LC Packings coupled to Thermo LTQ-FT mass spectrometer, Thermo Fisher).

High performance liquid chromatography

Plasma extracts were loaded onto a Famos® (LC Packings) autosampler and 10 μL of sample were injected onto the LC-FTMS system. Analyte separation was done with a Hamilton PRPX-110S (2.1 × 10 cm) anion exchange column using a formic acid gradient. Preliminary studies showed that inclusion of a short, end-capped C18 pre-column (Higgins Analytical Targa guard) was necessary for desalting and optimal separation. The chromatography was a minor modification of the methods in the article by Johnson *et al.*;[20] the flow rate was 0.35 mL min^{-1}. Solution A was 0.1% (v/v) formic acid in a 1 : 1 mix of water : acetonitrile. Solution B was 1% (v/v) formic acid in a 1 : 1 mix of water : acetonitrile. Chromatography consisted of a 2 min period with 100% A, followed by a 6 min linear gradient from 100% A to 100% B and 2 min at 100% B. A Switchos (LC Packings) control valve was used to alternate analysis of samples on two columns, allowing one column to undergo a wash cycle while separation was performed on the other column. For the off-line wash, 2% (v/v) formic acid in acetonitrile was used at 0.5 mL min^{-1}. Columns were equilibrated to initial conditions of 100% A for 2 min at 0.5 mL min^{-1} prior to the next injection.

Mass spectrometry

A Thermo LTQ-FT mass spectrometer (Thermo Fisher, San Diego, CA) was set to collect data from *m*/*z* 85 to 850. Following optimization (see Results and discussion) the final operating conditions used a spray voltage of 5.5 kV, sheath gas of 40 (arbitrary units), capillary temperature of 275 °C, capillary voltage of 44 V and tube lens of 120 V. Ion transfer optics were optimized automatically. Maximum injection time was 500 ms, and the maximum number of ions collected for each scan was 3×10^6. A wide range scan was used for the FT-ICR with mass resolution of 50 000.

Data collection and processing

Data were collected continuously over the 10 min chromatographic separation and stored as.raw files. These files were converted using Xcalibur file converter software (Thermo Fisher, San Diego, CA) to.cdf files for further data processing. An adaptive processing software package (apLCMS, http://www.sph.emory.edu/apLCMS) designed for use with LC-FTMS data was used for peak extraction and quantification of ion intensities.[18] This software provided *m*/*z* feature tables containing *m*/*z*, retention time and integrated ion intensity for each *m*/*z* feature, obtained through 5 major processing steps: (1) noise filter, (2) peak identification in terms of peak location (*m*/*z* and retention time), peak width and intensity, (3) retention time correction, (4) *m*/*z* peak alignment across multiple spectra, and (5) re-analysis to capture peaks originally missed because of weak signal relative to the signal to noise filter.

Three dimensional (3D) projections, coefficients of variation (CV), principal component analysis (PCA), statistical analysis of microarrays (SAM) and false discovery rate (FDR) analyses were performed in R 2.8.0 (http://www.r-project.org/).

Results and discussion

Protocol development

Method development involved systematic evaluation of operating conditions of the LTQ-FT mass spectrometer along with tests of sample extraction and variations in chromatography. To deproteinate samples, extractions with different volume equivalents of acetonitrile and methanol were compared. Protein removal was evaluated by the Lowry method and by SDS-PAGE with Coomassie blue staining and densitometry. Extraction with acetonitrile 2 : 1 showed 98% protein removal, slightly better than with methanol 2 : 1 (96% protein removal), and was used for subsequent studies.

The LTQ-FT mass spectrometer is optimized to analyze samples over a 10-fold range of *m*/*z*. In practical terms, this means that the best a single analysis can do is capture a segment of the metabolome. We initially performed analyses with *m*/*z* range from 75 to 750, designed to include all of the common amino acids and dipeptides. Because the dipeptides were not in metabolic databases and metabolic databases include drugs which will not be common to most samples, we compiled a list of 1464 human metabolites from the literature, plus 1769 dipeptides and tripeptides, to use as a basis to estimate the metabolic coverage likely to be obtained with the *m*/*z* range (Table S1†). Based upon this table, the *m*/*z* range from 75 to 750 included 90% of the metabolites, suggesting suitability to obtain extensive characterization of chemicals in human plasma. However, a solvent peak at *m*/*z* 82 caused a significant dynamic range limitation in the LTQ-FT mass spectrometer. To avoid this problem, the analysis range was changed to 85 to 850. This range does not include glycine or other small metabolites in the *m*/*z* range from 75 to 85, but includes 91% of the compiled metabolite table. Consequently, while the selected range is suitable for profiling of human plasma, variations of the *m*/*z* range or combination of analyses with different *m*/*z* ranges can be used to optimize analyses for different purposes. Initial tests of reproducibility showed that successive analyses of a single sample resulted in increasing ion suppression due to accumulation of lysophosphatidylcholine species (C16 : 0, C18 : 1 and C18 : 2 being the most abundant) on the column. This ion suppression effect was eliminated by introduction of a wash cycle as described in the Experimental section.

FTMS parameters were tested to minimize the scan time while maximizing dynamic range. Mass resolution of 50 000 was chosen to allow the scan speed to be increased to 500 ms while still maintaining adequate mass resolution to resolve co-eluting compounds. The maximum number of ions collected in the ion cyclotron resonance cell for each scan was varied to test sensitivity *vs.* loss of mass accuracy due to space-charge effect and found to give optimal sensitivity without significant loss of mass accuracy at 3×10^6 ions per scan. This number is higher than the 1×10^6 that is recommended for the instrument but enhanced the number of *m*/*z* features detected and still provided a mass accuracy within 5 ppm. For 17 common amino acids, all were within 5 ppm and the average mass accuracy was 3.5 ppm.

Visualization of plasma metabolic profiles by LC-FTMS

Data were visualized in a series of 3D plots representing relative abundance as a function of *m*/*z* and run time (Fig. 1). Results

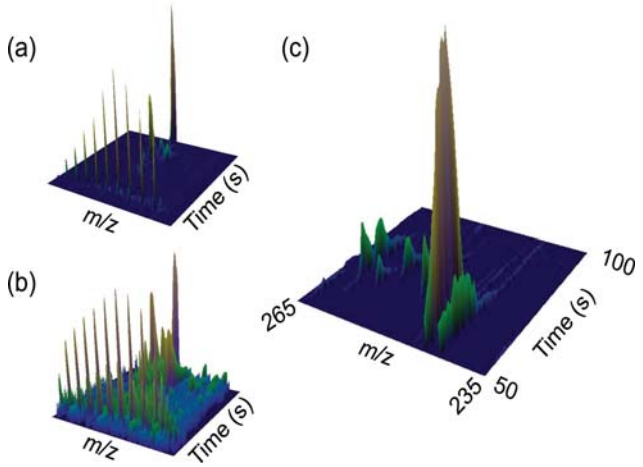

Fig. 1 Feature profile of human plasma obtained by 10 min LC-FTMS analysis. Eluate was ionised by ESI in the positive ion mode. Data were visualised using R software as three-dimensional projections in which the surface plane represents the *m*/*z* and elution time, while the projection (z-axis) represents the ion signal intensity. A. Raw data. B. Cubed root-transformation of data to enhance visualisation of lower abundance peaks. C. Magnification of *m*/*z* 235–265 and RT 50–100 s to enhance visualisation of individual peaks. The largest peak corresponded to the [$^{13}C_2$]-cystine which was spiked into the plasma prior to deproteination and the peak at 241.0311 corresponds to endogenous cystine. Additional analytes detected represent unidentified chemicals in human plasma.

showed large peaks corresponding to a series of sodium formate clusters at the beginning of the run (desalting of the plasma) and lysophosphatidylcholine species (C16 : 0, C18 : 1 and C18 : 2) eluting toward the end of the run (Fig. 1a). Because of the high abundance of these features, a cube root transformation provided a better visualization of the distribution of lower intensity peaks (Fig. 1b) and showed the complexity of the data in terms of the number of features detected. Expanded plots showed that individual *m*/*z* features eluted with time as expected for discrete analytes (Fig. 1c). In the region shown, cystine (*m*/*z* 241.031) coeluted with [$^{13}C_2$]-cystine (*m*/*z* 243.034), which was added to the plasma as an internal standard. The smaller peaks in the background are unidentified chemicals. Targeted searches for 19 of the common amino acids showed that corresponding *m*/*z* features were present for 17, and these coeluted with authentic standards (data not shown). Two of the amino acids (valine, histidine) were not detected.

Analysis of systematic variation within a sample set

To determine if systematic changes occur with the LC-FTMS analysis from run to run, plasma extracts were analyzed repetitively, processed by apLCMS to obtain an *m*/*z* feature list, and the features were analyzed by principal component analysis (PCA). Results showed a change in trajectory in the 3-dimensional (3-D) plot of the first 3 principal components with injection number for the first 4 to 5 runs before becoming reproducible (Fig. S1A†). Thus, a 5-run pre-sequence was needed to obtain reproducible elution characteristics. Subsequent studies showed that this pre-sequence was required for equilibration of the column to the matrix and gradient conditions rather than due to differences between plasma samples. After the pre-sequence, no further equilibration was needed for analysis of sample sets up to 80. Because analyses were performed in a shared-use facility which limited access to the FTMS, sample sets greater than 80 were not tested.

Analysis of variation between analytical columns

Because column retention characteristics vary as a function of the prior use of a column, we performed an experiment to determine possible effects of differences in column history on *m*/*z* feature detection. Two columns were used with identical lot number but with a different history in terms of number of analytical runs, one was new and the other had >200 analyses over a 4 month period. A series of samples were run concurrently, alternating analyses on the two columns. The same number of *m*/*z* features was obtained for the columns, but PCA of the *m*/*z* features showed separation of profiles by column (Fig. S1B†). This problem can be minimized by standard operating procedures for column use based upon retention characteristics of standards. Importantly, the data confirm the expectation that analyses run on different columns cannot be directly combined into a common dataset without appropriate normalization procedures.

Because metabolic profiling for personalized medicine requires extraction of comparable information from different samples, an experiment was designed using cross-correlation analysis to determine whether the same type of information was obtained from samples analyzed on different columns. For this purpose, a set of 34 samples collected over a 10 d period from an individual[19] was analyzed concurrently, in duplicate, on two columns. The samples were run in reverse order on the columns to maximize variation due to run order. Correlation analysis of the *m*/*z* feature data showed common patterns (Fig. S2†). The results suggest that the same metabolic information is extracted from the LC-FTMS analysis when samples are analyzed on different columns and that response characteristics were not affected by the order of analyses.

Qualitative analysis of metabolic profiles

To explore the potential utility for personalized medicine, studies were conducted to determine the qualitative and quantitative characteristics of the *m*/*z* features obtained by the LC-FTMS method. To address the qualitative nature of the metabolic profiles, plasma samples were collected from 4 individuals without selection criteria other than self-described as healthy, *i.e.*, there was no standardization concerning meals, time of day, age, *etc.* To determine how many replicates were needed to obtain information for an individual, criteria for *m*/*z* feature extraction by the apLCMS routine were systematically varied. With parameters set to report a feature if present in only one analysis per individual, a total of 2124 features was detected. The average percentage of features present in duplicate samples from each individual was then calculated to be 88.8% for Subject 1, 88.6% for Subject 2, 87.4% for Subject 3 and 86.2% for Subject 4. These data show that duplicate analyses are sufficient to capture most of the metabolites. More replicates provided limited improvement, but this improvement was offset by longer analysis time.

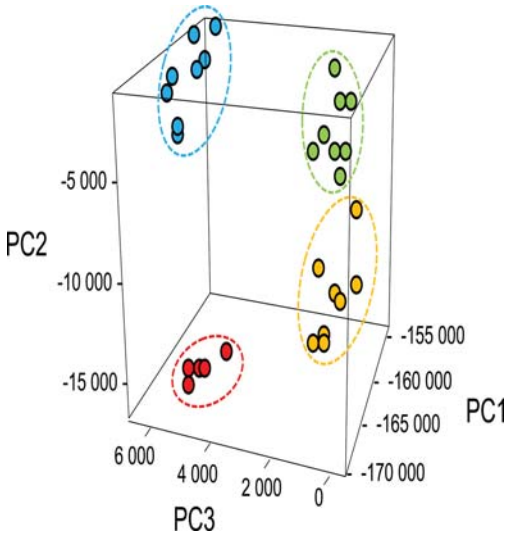

Fig. 2 Three-dimensional plot of unsupervised Principal Component Analysis (PCA) of LC-FTMS data for replicate plasma samples from 4 healthy individuals. PCA was performed to determine whether variation in the metabolic profiles of individuals was greater than the metabolic variation due to the variability of detection by LC-FTMS. Results show that replicate profiles separate according to individual despite the cumulative variability of thousands of *m*/*z* features.

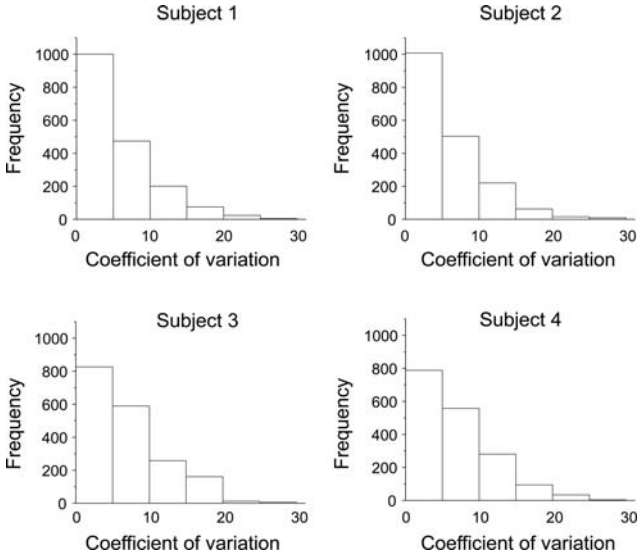

Fig. 4 Histograms showing number of peaks with ranges of signal variation from LC-FTMS analysis. Variation is expressed as coefficient of variation (CV) obtained from analysis using automated noise level selection and separation, and smoothing-based feature alignment for peak selection.[6] Ten microliter samples from 4 individuals were analyzed by LC-FTMS with either 6 replicates (Subject 1) or 8 replicates (Subjects 2-4). Results show that most features have CV below 10%.

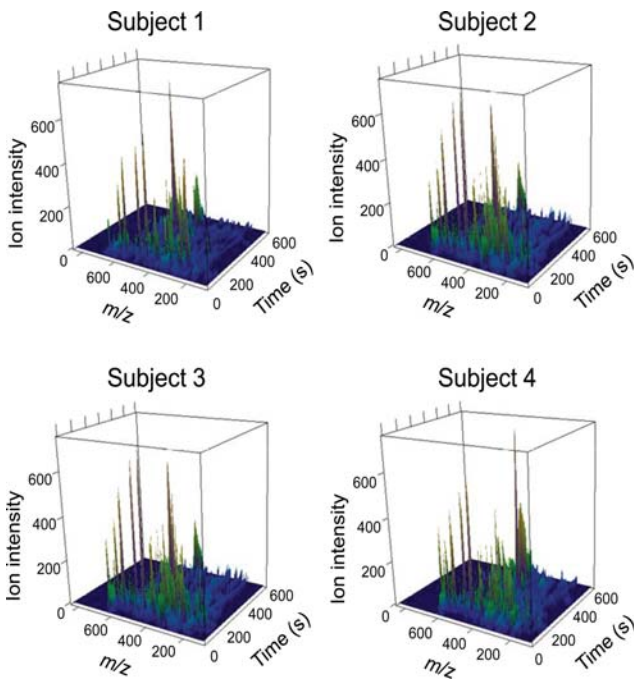

Fig. 3 Three-dimensional landscape projection of *m*/*z* features which distinguish individuals. Statistical Analysis of Microarrays (SAM) using False Discovery Rate (FDR) of 0.1% was applied to the spectra of four individuals to determine features that differ significantly between individuals. This method removes all features that do not differ between individuals. Results showed that 770 features with unique *m*/*z* were significantly different between individuals. Projections of these 770 features reveal unique individual metabolic landscapes of the individuals.

To evaluate differences between the four individuals, the data were re-processed with apLCMS software with the criterion that an *m*/*z* feature had to be present in 25% of the 30 profiles analyzed. Analysis by PCA showed that metabolic patterns separated individuals and were consistent from run to run (Fig. 2). Analysis of data collected from a sulfur amino acid study with chemically defined diets conducted over a 10 d time course[19] further showed that intra-individual variation due to change of diet was smaller than inter-individual differences (Fig. S3†). Thus, metabolic profiles are qualitatively similar on a run-to-run basis. Landscape projections of features responsible for individual differences among the 4 subjects showed that some distinguishing features are unique to an individual but most are common to multiple individuals and vary in intensity (Fig. 3).

Quantitative reproducibility

To evaluate quantitative reproducibility for individual features, coefficients of variation (CV) were calculated for the *m*/*z* features that were detected in the above analyses. For each individual, at least 80% of the features had a CV of less than 10%, and more than 90% of features had a CV less than 20% (Fig. 4). Thus, the quantitative reproducibility is suitable for most features, and even with a CV cutoff of 10%, quantitative information is available for over 1700 *m*/*z* features.

Metabolic characteristics of individuals

To evaluate potential suitability of this method to detect differences which might be useful as biomarkers of health or disease, we determined how many of the *m*/*z* features detected in the 4 individuals were unique to an individual and how many were

common to multiple individuals. Results showed that 183 of the features (10%) were unique to an individual, while 1167 features (62%) were common to all 4 individuals (Fig. S4†). Of the remaining 524 features, 269 were found in 2 individuals and 255 were found in 3 of the individuals. The percentage of features (10%) identified as unique is probably an over-estimate because this comparison included only four individuals; each individual compared to a larger number of individuals would be likely to have a smaller percentage of unique features. In addition, samples were from individuals who were not fasted, and the differences in food intake could also have contributed to differences in *m*/*z* features between subjects.

To further examine differences, Statistical Analysis of Microarrays (SAM) was used with a false discovery rate (FDR) set at 0.1%. Results showed 770 features were significant in distinguishing individuals. The data show that the LC-FTMS method is sensitive enough to provide metabolic profiles which can distinguish individuals in terms of differences in plasma metabolites. The presence of some unique features in individuals also suggests that the approach could be used to discover qualitative metabolic differences which contribute to health risks.

Characterization of *m*/*z* features

As indicated in the Introduction, an attractive feature of the high-resolution mass spectrometers is that the high mass accuracy and resolution can be used as a basis to predict elemental compositions of small molecules,[12] and this can be used to obtain tentative identities of common metabolites. We used Table S1† with exact mass for 1464 endogenously produced metabolites and 1769 dipeptides and tripeptides for comparison to the features observed in human plasma by LC-FTMS. Results showed 209 (10%) matches of high mass accuracy *m*/*z* values. These matches included amino acids, carbohydrates, phospholipids and energy intermediates. In comparison, a search against the web-based Madison Metabolomics Consortium Database (MMCD), which includes plant metabolites, drugs and common environmental metabolites,[22] restricting mass accuracy to a 10 ppm range, revealed 445 (22%) matches. Thus, the data show that many of the *m*/*z* features represent unidentified chemicals in human plasma. This highlights a need to integrate *m*/*z* feature analysis with correlation[23] and pathway[24] analysis to guide research in metabolic profiling for personalized medicine.

Comparison of top-down profiling to targeted analysis

In a previous study with a targeted metabolomics approach,[20] we found that cystine was significantly higher in individuals over the age of 60 compared to individuals under the age of 30. We reexamined these samples to see if our top-down method could detect this difference. The results showed that cystine was significantly higher ($p = 0.002$) in the older group, consistent with the results of the targeted study. The top-down approach also yielded 32 other *m*/*z* features that differed between groups with a *p*-value less than 0.01. These included the amino acids glutamate and methionine, which were confirmed using authentic standards and MS/MS; glutamate was higher in the older subjects and methionine was lower in the older subjects consistent with previous results.[25,26] Other tentatively identified chemicals which increased in the older subjects were endogenously derived cystinyl-bis-glycine (metabolite of GSH) and 1-hydroxy-2-methyl-2-butenyl-4-diphosphate (intermediate in isoprene biosynthetic pathway) and exogenously derived sofalcone (anti-ulcer medication) and xanthohumol E (flavonoid found in hops). These results suggest that the top-down concept for metabolic profiling can be used to simultaneously provide both targeted and untargeted capabilities.

Transition in analytic technologies to support personalized medicine

Targeted analysis of known chemicals is a critical component of diagnostic medicine and, in plant research, is often the only way of studying the phenotypic changes brought on by genetic mutations.[27] Pharmacology and toxicology extensively use targeted analysis to study metabolic fates of small molecule organic compounds.[27–31] The ability to collect, process and store large amounts of data further allows detailed metabolic information to be used in clinical and epidemiological research.[32–37] Advances in separation sciences and mass spectrometry have led to substantial improvement in accuracy and sensitivity of these assays, and currently used approaches generally serve needs well. However, as concepts of personalized and predictive medicine have developed, needs have been recognized which are not readily addressed by these methods. In particular, genomics research has shown that many genes contribute in small ways to risk of common diseases. Similarly, metabolic patterns associated with disease are likely to be complex with multiple metabolites and metabolic pathways being important in specific disease processes. While currently available methods, including powerful LC-MS/MS platforms, provide useful capabilities, they are limited in speed, sensitivity and/or number of analytes measured. The moderately high throughput method described here measures 2000 *m*/*z* features in a small volume of plasma with a single extraction step. This can be compared to a recent study which measured 300 targeted chemicals.[25] The latter involved multiple extraction steps, centrifugation, drying and reconstitution and combination of LC-MS and GC-MS platforms. Thus, application of high resolution mass spectrometry allows for minimal sample processing and chromatographic separation, providing a substantial advance in ability to provide high performance metabolic profiling.

An important observation from the current and previous MS studies is that 50–80% of the *m*/*z* features detected do not correspond to known metabolites.[38] Fragmentation of *m*/*z* features not found in databases shows product ion spectra that are common to organic biomolecules suggesting these unknowns are uncharacterized environmental and biological chemicals.[4] This highlights a need for a top-down approach in which chemicals are measured in biological systems without regard to known metabolic pathways or chemical identities. Such an approach could be used with a cumulative metabolic database consisting of *m*/*z* features associated with health and disease phenotypes. This could result in identification of new chemical associations with disease and further elucidate disease mechanisms. Chemical identities for uncharacterized *m*/*z* features could be added to the database as chemical structures are established. Similarly, quantitative calibration could be incorporated as

authentic standards become available. In this way, high performance metabolic profiling utilizing the capabilities of high resolution mass spectrometry could accelerate development of personalized medicine.

Conclusion

This report describes a method to obtain extensive and reproducible metabolic profiling for routine clinical use. The method uses the capability of high resolution mass spectrometry to rapidly measure 2000 plasma metabolites without regard to known chemical identity. This provides a practical approach to detect unique metabolic patterns which would be cost-effective for personalized medicine.

Acknowledgements

This research was supported by NIH ES012870, DK066008, ES016731, DK007734, ES011195 and the Emory-Georgia Tech Predictive Health Institute.

References

1 W. C. Willett, *Science*, 2002, **296**, 695–698.
2 M. T. Smith and S. M. Rappaport, *Environ. Health Perspect.*, 2009, **117**, A334–A335.
3 C. P. Wild, *Cancer Epidemiol., Biomarkers Prev.*, 2005, **14**, 1847–1850.
4 M. J. Gibney, M. Walsh, L. Brennan, H. M. Roche, B. German and B. van Ommen, *Am. J. Clin. Nutr.*, 2005, **82**, 497–503.
5 S. C. Schuster, *Nat. Methods*, 2008, **5**, 16–18.
6 W. Weckwerth, *Annu. Rev. Plant Biol.*, 2003, **54**, 669–689.
7 T. Kuhara, *Mass Spectrom. Rev.*, 2005, **24**, 814–827.
8 H. G. Gika, G. A. Theodoridis and I. D. Wilson, *J. Chromatogr., A*, 2008, **1189**, 314–322.
9 A. M. Evans, C. D. DeHaven, T. Barrett, M. Mitchell and E. Milgram, *Anal. Chem.*, 2009, **81**, 6656–6667.
10 A. A. Ramos, H. Yang, L. E. Rosen and X. Yao, *Anal. Chem.*, 2006, **78**, 6391–6397.
11 L. Novakova, L. Matysova and P. Solich, *Talanta*, 2006, **68**, 908–918.
12 A. G. Marshall and C. L. Hendrickson, *Annu. Rev. Anal. Chem.*, 2008, **1**, 579–599.
13 O. Shaham, N. G. Slate, O. Goldberger, Q. Xu, A. Ramanathan, A. L. Souza, C. B. Clish, K. B. Sims and V. K. Mootha, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 1571–1575.
14 C. J. Clarke and J. N. Haselden, *Toxicol. Pathol.*, 2008, **36**, 140–147.
15 A. C. Hogenboom, J. A. van Leerdam and P. de Voogt, *J. Chromatogr., A*, 2009, **1216**, 510–519.
16 J. V. Olsen, J. C. Schwartz, J. Griep-Raming, M. L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E. R. Wouters, M. Senko, A. Makarov, M. Mann and S. Horning, *Mol. Cell. Proteomics*, 2009, **8**, 2759–2769.
17 H. B. Greizerstein, P. Gigliotti, J. Vena, J. Freudenheim and P. J. Kostyniak, *J. Anal. Toxicol.*, 1997, **21**, 558–566.
18 T. Yu, Y. Park, J. M. Johnson and D. P. Jones, *Bioinformatics*, 2009, **25**, 1930–1936.
19 H. W. Cho, S. B. Kim, M. K. Jeong, Y. Park, N. G. Miller, T. R. Ziegler and D. P. Jones, *International Journal of Data Mining and Bioinformatics*, 2008, **2**, 176–192.
20 J. M. Johnson, F. H. Strobel, M. Reed, J. Pohl and D. P. Jones, *Clin. Chim. Acta*, 2008, **396**, 43–48.
21 E. J. Want, G. O'Maille, C. A. Smith, T. R. Brandon, W. Uritboonthai, C. Qin, S. A. Trauger and G. Siuzdak, *Anal. Chem.*, 2006, **78**, 743–752.
22 Q. Cui, I. A. Lewis, A. D. Hegeman, M. E. Anderson, J. Li, C. F. Schulte, W. M. Westler, H. R. Eghbalnia, M. R. Sussman and J. L. Markley, *Nat. Biotechnol.*, 2008, **26**, 162–164.
23 V. P. Makinen, P. Soininen, C. Forsblom, M. Parkkonen, P. Ingman, K. Kaski, P. H. Groop and M. Ala-Korpela, *Mol. Syst. Biol.*, 2008, **4**, 167.
24 D. Huster, T. D. Purnat, J. L. Burkhead, M. Ralle, O. Fiehn, F. Stuckert, N. E. Olson, D. Teupser and S. Lutsenko, *J. Biol. Chem.*, 2007, **282**, 8343–8355.
25 K. A. Lawton, A. Berger, M. Mitchell, K. E. Milgram, A. M. Evans, L. Guo, R. W. Hanson, S. C. Kalhan, J. A. Ryals and M. V. Milburn, *Pharmacogenomics*, 2008, **9**, 383–397.
26 M. D. Armstrong and U. Stave, *Metab., Clin. Exp.*, 1973, **22**, 561–569.
27 O. Fiehn, *Plant Mol. Biol.*, 2002, **48**, 155–171.
28 J. C. Lindon, H. C. Keun, T. M. Ebbels, J. M. Pearce, E. Holmes and J. K. Nicholson, *Pharmacogenomics*, 2005, **6**, 691–699.
29 E. Bohus, M. Coen, H. C. Keun, T. M. Ebbels, O. Beckonert, J. C. Lindon, E. Holmes, B. Noszal and J. K. Nicholson, *J. Proteome Res.*, 2008, **7**, 4435–4445.
30 M. Coen, Y. S. Hong, T. A. Clayton, C. M. Rohde, J. T. Pearce, M. D. Reily, D. G. Robertson, E. Holmes, J. C. Lindon and J. K. Nicholson, *J. Proteome Res.*, 2007, **6**, 2711–2719.
31 H. Li, Y. Ni, M. Su, Y. Qiu, M. Zhou, M. Qiu, A. Zhao, L. Zhao and W. Jia, *J. Proteome Res.*, 2007, **6**, 1364–1370.
32 E. K. Kemsley, G. Le Gall, J. R. Dainty, A. D. Watson, L. J. Harvey, H. S. Tapp and I. J. Colquhoun, *Br. J. Nutr.*, 2007, **98**, 1–14.
33 T. Kuhara, *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.*, 2007, **855**, 42–50.
34 N. Raikos, G. Theodoridis, E. Alexiadou, H. Gika, H. Argiriadou, H. Parlapani and H. Tsoukali, *J. Sep. Sci.*, 2009, **32**, 1018–1026.
35 S. Rezzi, Z. Ramadan, F. P. Martin, L. B. Fay, P. van Bladeren, J. C. Lindon, J. K. Nicholson and S. Kochhar, *J. Proteome Res.*, 2007, **6**, 4469–4477.
36 M. C. Walsh, L. Brennan, J. P. Malthouse, H. M. Roche and M. J. Gibney, *Am. J. Clin. Nutr.*, 2006, **84**, 531–539.
37 M. C. Walsh, L. Brennan, E. Pujos-Guillot, J. L. Sebedio, A. Scalbert, A. Fagan, D. G. Higgins and M. J. Gibney, *Am. J. Clin. Nutr.*, 2007, **86**, 1687–1693.
38 M. G. Miller, *J. Proteome Res.*, 2007, **6**, 540–545.

