Published in final edited form as:

Anal Chem. 2014 April 01; 86(7): 3667–3675. doi:10.1021/ac500476a.

A nano Ultra-Performance Liquid Chromatography – High Resolution Mass Spectrometry Approach for Global Metabolomic Profiling and Case Study on Drug-Resistant Multiple Myeloma

Drew R. Jones¹, Zhiping Wu¹, Dharminder Chauhan², Kenneth C. Anderson², and Junmin Peng^{1,3,*}

¹Departments of Structural Biology and Developmental Neurobiology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

²Dana-Farber Cancer Institute, 44 Binney Street, Dana 1B02, Boston, MA 02115

³St. Jude Proteomics Facility, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

Abstract

Global metabolomics relies on highly reproducible and sensitive detection of a wide range of metabolites in biological samples. Here we report the optimization of metabolome analysis by nanoflow ultra-performance liquid chromatography coupled to high resolution orbitrap mass spectrometry. Reliable peak features were extracted from the LC-MS run based on mandatory detection in duplicates and additional noise filtering according to blank injections. The run-to-run variation in peak area showed a median of 14%, and the false discovery rate during a mock comparison was evaluated. To maximize the number of peak features identified, we systematically characterized the effect of sample loading amount, gradient length, and MS resolution. The number of features initially rose and later reached a plateau as a function of sample amount, fitting a hyperbolic curve. Longer gradients improved unique feature detection in part by time-resolving isobaric species. Increasing the MS resolution up to 120,000 also aided in the differentiation of near isobaric metabolites, but higher MS resolution reduced the data acquisition rate and conferred no benefits, as predicted from a theoretical simulation of possible metabolites. Moreover, a biphasic LC gradient allowed even distribution of peak features across the elution, yielding markedly more peak features than the linear gradient. Using this robust nUPLC-HRMS platform, we were able to consistently analyze ~6,500 metabolite features in a single 60 min gradient from 2 milligram of yeast, equivalent to ~50 million cells. We applied this optimized method in a case study of drug (bortezomib) resistant and drug-sensitive multiple myeloma cells. Overall, 18% of metabolite features were matched to KEGG identifiers enabling pathway enrichment analysis.

*Corresponding Author Junmin Peng (junmin.peng@stjude.org).

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

SUPPORTING INFORMATION AVAILABLE

This information is available free of charge via the Internet at <http://pubs.acs.org/>.

Principal component analysis and heat map data correctly clustered isogenic phenotypes, highlighting the potential for hundreds of small molecule biomarkers of cancer drug resistance.

Keywords

Metabolomics; nLC; High Resolution; Mass Spectrometry; Yeast; Multiple Myeloma; Drug Resistant

INTRODUCTION

Metabolomics is becoming increasingly important for the analysis of basic biochemical mechanisms as well as in the search for diagnostic, prognostic, and predictive markers of disease. Research efforts are typically categorized as either targeted or global, depending on the type of information sought and the instrumentation used¹. For global studies, the structural assignment of metabolite features remains the rate-limiting step², but fruitful conclusions can nevertheless be extracted by correlating metabolite features with disease state or treatment level. Preliminary metabolite identification can be achieved by matching ions to comprehensive metabolite databases such as HMDB 3.0³ or METLIN⁴. Therefore, such discovery based experiments rely on casting as wide a net as possible with respect to metabolite detection in order to generate hypotheses for further examination. Targeted metabolic analyses gain specificity and quantitative information by focusing on a subset of known metabolites⁵. Recent landmark metabolomics studies have used both targeted and global methods either alone or in concert to evaluate broad hypotheses and discover important mechanisms of disease. For instance, Zheng et al. developed a chemical isotope labeling strategy for profiling ~1,000 human salivary metabolites and found 18 species associated with mild cognitive impairment⁶. Further, Jain et al. targeted 219 metabolites and were able to identify mitochondrial glycine as a key mediator of cancer cell proliferation by assessing the consumption and release profiles in the NCI-60 panel of cancer cell lines⁷. In contrast, Panopoulos et al. performed the first global metabolomics study (~5,000 features) of embryonic stem cells and induced pluripotent stem cells, and discovered a metabolomic signature which was correlated with the efficiency of cellular reprogramming⁸. Further work by Hou et al. has demonstrated that pluripotent stem cells can now be chemically induced using only small molecules⁹.

Discovery metabolomics can be performed using a variety of complementary analytical tools with nuclear magnetic resonance (NMR) and mass spectrometry (MS) being the most common platforms. Modern NMR systems are capable of rapidly delivering a metabolomic profile containing several hundred features. These analyses are highly reproducible, providing a reliable tool for metabolomic classification studies in tissue biopsies or various biofluids¹⁰. The current obstacles in NMR studies are related to instrument sensitivity and deconvolution of the spectra into discrete information on metabolite levels. While NMR acquires an aggregate spectrum, MS can be coupled to either gas chromatography (GC-MS) or liquid chromatography systems (LC-MS) providing a second dimension of resolution to enable better detection.

GC-MS and LC-MS are complementary tools for profiling different types of metabolites. GC-MS is favorable in the analysis of volatile metabolites or derivatized metabolites (e.g. lipidomics)¹¹. For example, a GC-MS platform was successfully used to link metabolite markers (e.g. citrate) with glioblastoma tumor grade and prognosis¹². LC-MS is a powerful platform for analyzing soluble metabolites that can be separated on selected columns (e.g. reverse phase or HILIC columns)¹³. The availability of different MS instruments also diversifies the technology, including triple quadrupole, quadrupole-time of flight, linear ion trap, high resolution orbitrap and Fourier transform ion cyclotron resonance mass spectrometers. Tandem mass spectrometry (MS/MS) has also been implemented to acquire product ions upon fragmentation, greatly enhancing the identification and quantification of metabolite features^{3,4}. Ion mobility mass spectrometry provides additional gas-phase separation power to distinguish and characterize unknown metabolites¹⁴. Moreover, capillary electrophoresis-mass spectrometry (CE-MS) is an alternative platform that offers high resolution and high sensitivity for detecting metabolites. When CE-MS was applied to metabolic profiling of oral, breast, and pancreatic cancer by non-invasive saliva analysis, approximately 3,050 features were detected which led to the identification of 57 potential cancer biomarkers. Anionic metabolites are well suited to nano flow methodologies and some studies suggest that sensitivity can be increased by the addition of metal chelating agents to the mobile phase¹⁵. Semi-targeted metabolomics studies have also benefited from capillary flow systems, especially with respect to chromatographic resolution of di- and tri-phosphates with HILIC resins¹⁶. The high sensitivity of capillary LC has even been applied to the analysis of single cells, for instance in the detection of 200 metabolites by Ni et. al. in a single islets of Langerhans¹⁷. Nevertheless, nanoflow LC has yet to be applied to large-scale untargeted metabolomics experiments.

In this study, we aimed to apply the latest LC-MS technology to increase the coverage of a complex metabolome. A nanoflow ultra-performance liquid chromatography system coupled to a high resolution (up to 480,000) mass spectrometer (nUPLC-HRMS) has been widely adopted by the proteomics community because of high sensitivity and resolution¹⁸. To determine the practical utility of this system with respect to metabolite feature detection and discrimination, we systematically characterized our sample processing from metabolite extraction to statistical filtering and feature reporting, while exploring a variety of experimental conditions such as sample loading, gradient length, gradient composition, and MS resolution. This optimization was performed with the metabolome extracted from yeast cells. Using the optimized platform, we were able to detect ~6,500 features from 2 milligrams of yeast. Finally, we applied this methodology to a case study of drug-resistant multiple myeloma and used gene ontology approaches to identify novel regulatory pathways involved in bortezomib resistance.

EXPERIMENTAL SECTION

Materials –

LCMS grade acetonitrile (ACN), water, and formic acid (Sigma), glass beads (Next Advance), YPD broth, nanoLC column and Oasis solid phase extraction kit (Waters).

Yeast Culture –

A single colony of yeast By4742 (Open Biosystems) was inoculated into a 2 mL starter culture of YPD broth (50g/L) and then transferred to 100 mL flasks and grown to an optical density of 1 OD₆₀₀.

Culture of Multiple Myeloma Cells –

The ANBL-6 bortezomib sensitive (BS) multiple myeloma cell line and its isogenic bortezomib resistant (BR) counterpart were cultured in RPMI 1640 (Corning). Media was also supplemented with 10% fetal bovine serum (Sigma), 100 U/mL penicillin, and 100 µg/mL streptomycin, 10 ng/mL of IL-6 (R&D Systems) and maintained in 5% CO₂ at 37°C. BR cells were constantly exposed to 5 nM bortezomib (Santa Cruz) during cell culture and storage.

Metabolite Extraction –

All extractions followed the generalized sample workflow. Yeast cultures were centrifuged and the media supernatant was removed and stored. All cells were washed twice with a 2-fold volume of water to remove YPD broth and 100 mg of packed cells were aliquoted for each replicate extraction in 1.5 mL microfuge tubes. Metabolites were extracted by glass bead vortexing¹⁹ in the presence of freezing 80% (vol/vol) LCMS grade ACN similar to widely used metabolomics sample preparation procedures²⁰. Each vial contained 100 µL of 0.5 mm glass beads and 50 µL of 1 mm beads. Vials were vortexed for a total of 3 min at 3,000 rpm in a 1 on/1 off pattern in order to keep samples cold. The resulting lysate was then centrifuged at 21,000 x g for 5 min to clarify the ACN phase, which was transferred to a fresh tube. This solution was then aliquoted in 100 µL volumes for storage at –80°C. For analysis, samples were dried using a centrifugal vacuum concentrator. For non-desalted samples, a 100 µL aliquot was dried, and resolubilized in 10 µL of 0.2% formic acid.

Desalting –

The 100 µL aliquot was dried and then resolubilized in 100 µL of 0.1% trifluoroacetic acid (TFA) and loaded onto the solid-phase extraction column (Oasis HLB, 10 mg sorbent) after being sequentially conditioned with 1 mL of 100% ACN and 1 mL of 0.1% TFA. The column was then washed three times with 1 mL of 0.1% TFA, and finally eluted with 1 mL of 100% ACN. This eluent was dried by speed vac and resolubilized in 10 µL of 0.2% formic acid.

LC-MS Instrumentation –

Metabolomics data were acquired using a waters nanoAcquity UPLC system, coupled to an Orbitrap Elite (Thermo Scientific) in profile mode. Instruments were controlled by XCalibur (version 2.2 SP1.48, Thermo). Injection volume was held at 2 µL for all analyses. The sample loading volume was set to twice the injection volume at a flow rate of 0.3 µL/min using the partial loop injection method. The analytical flow rate was 0.2 µL/min unless otherwise specified. Metabolites were separated by a Waters nanoAcquity UPLC column (75 µm × 100 mm with 1.7 µm BEH C18 particles). Eluting metabolites were ionized by ESI with the use of a PicoTip emitter, 10 ± 1 µm (New Objective) operating at 3kV in the

positive ion mode. The MS resolution was set to 240,000 unless otherwise specified. The 480,000 resolution setting was available through a non-standard acquisition mode (Developer's Kit, ThermoFisher). Mobile phase A and B consisted of LCMS grade water or acetonitrile, respectively, each with 0.2% formic acid. Various gradient lengths are described throughout the text, but all ranged from 1%B to 75%B over the allotted time unless otherwise described. This gradient was followed by a 1 min jump to 99% B which was held for 5 min to ensure all sample material was eluted from the column. Finally, the column was equilibrated for 20 min at 1%B during data acquisition to compensate for the delay time of the nano-LC system.

Data Analysis –

All data points represent duplicate LC-MS injections. The resulting high resolution LC-MS metabolomic data was analyzed by XCMS in the 32-bit version 2.15.3 of R (The R Project for Statistical Computing)²¹. The .raw files were converted to the .mzXML format using ProteoWizard²² (32-bit) MSConvert with a peakPicking MS level 1 filter to generate centroid data. The resulting files were then used for feature detection which was performed using the “centWave”²³ method and global retention time correction was performed using the “obiwarp”²⁴ method. Observed features were then grouped across samples using one round of the generic “group” function²⁵. The bandwidth parameter was set to “5” based on peak width and parameter search optimization. The “mzwid” parameter was dependent on instrument resolution settings and was determined by measuring the full peak width at m/z 300. For technical duplicates, the “minfrac” parameter was set to 1 such that reported features had to be present in both replicates. This stringent requirement significantly decreased the amount of features due to noise. Aligned peak data were exported and further analyzed in excel as described. An additional peak noise threshold parameter was identified for each gradient system (i.e. different length, flow rate, etc.) based on technical duplicates of blank injections. The 90th percentile feature peak area was calculated and this value was then set as the peak noise threshold. For non-blank samples, any features with peak area values below this threshold were discarded as noise.

Metabolites from multiple myeloma cells were identified by matching high resolution m/z values to the Human Metabolome Database and Metlin, enabling KEGG²⁶ identifications to be assigned. Heat map data were generated with the “gplots” library while multidimensional analysis was performed using the “MASS” library in R. Supplemental metabolite ontology data were generated through KegArray version 1.2.3.

RESULTS AND DISCUSSION

Reproducibility of the nUPLC-HRMS System –

We first demonstrated the reproducibility of the nUPLC-HRMS system (Figure 1A) by analyzing three aliquots of yeast metabolite extracts in duplicate across two weeks. The base peak chromatograms show the overall similarity of the data (Figure 1B). We then aimed to assess the inter-run, and quantitative reproducibility of the system. The yeast metabolite extract was analyzed in quadruplicate ($n = 4$) and the replicates were arbitrarily assigned into mock-control and mock-treatment groups. After performing feature detection in XCMS

(n=10,898 without additional noise filtering) we examined the technical reproducibility of the replicates with respect to feature peak area (Figure 1C) similar to previous metabolomics studies²⁷. The resulting scatter plot shows that there is no clear bias associated with small or large features being more or less reproducible. The respective histogram shows that more than half of the reported metabolite features were reproducible with a coefficient of variation less than 14% (Figure 1D). However, the distribution is right-tailed showing that a small fraction of reported features may have a large variation in peak area between replicates. There was no association between feature retention time or intensity with respect to reproducibility ($R^2=0.049$). We next performed a t-test to look for mock-regulated features which would provide an estimate of the false discovery rate (FDR) for global metabolomics studies. There was no clear relationship between small or large features being more likely to give false differences between groups (Figure 1E). We observed that only 3.8% of reported features appeared to be differentially regulated ($p < 0.05$) between mock-control and mock-treatment groups in this null experiment (Figure 1F). Theoretically, a 5% false positive rate would be expected with this cutoff ($p < 0.05$) due to multiple hypothesis testing. These consistent results strongly indicate that under the described conditions, global metabolomics data are not prone to report falsely regulated features between groups.

An ongoing challenge in our data analysis is the origin of noise within the system. Despite the orbitrap platform being very “clean” compared to other detectors, several hundred chromatographic features were observed during an analysis of blank samples. Further inspection of these features suggests that many of the observed peaks are real analytes as those features are highly reproducible in our runs. They represent the background contamination present in the solvents, plumbing, and the atmosphere²⁸. In contrast, analysis of an authentic injection typically yielded more than 15,000 features. Noise filtering protocols (e.g. features required to be detected in replicates) reduced the number of features by ~33%. Other stringent filtering options (see Experimental Section) could further reduce this number to several thousand highly reproducible metabolite features. Therefore, it is critical to achieve a balance between maximum feature detection and noise suppression. Carry over between samples is also a potentially confounding issue, especially in nanoflow LC analyses. We mitigated this factor by performing blank injections both before and after analytical sequences. Solid-phase extraction/desalting is known to play an important role in removing resilient hydrophobic species which can cause carryover. Comparison of these blank sample groups showed that carryover was not a problem in the analysis of yeast metabolites using our extraction protocol.

Optimization of Metabolite Extraction Methods –

Metabolites were extracted from yeast using 50 & 80% organic extraction^{20,29} (Figure 2A). We tested both concentrations of acetonitrile to examine if protein is completely precipitated, which can otherwise compromise method reproducibility or column life-span. Briefly, samples were extracted, concentrated by drying, and then desalted for LC-MS analysis. We found that the samples extracted with 50% acetonitrile contained 1.5 $\mu\text{g}/\mu\text{L}$ of protein. This level is at least 30-fold higher than when tissues were extracted with 80% ACN. The protein level in the 80% extract was below the minimum detection limit of the protein assay (Figure 2B). As a control, the total protein was extracted by 8M urea using an

optimized extraction method as previously published¹⁹. We further compared the glass bead metabolite extraction method versus the Barocycler NEP2320® with each using 80% ACN as the extraction solution. Comparison of the chromatograms showed that the barocycler was less efficient, though the overall profile appeared similar (data not shown). We also set out to determine if either condition adversely impacted the amount of features that could be detected in downstream analyses. Under the same analytical conditions, duplicate analyses of the two conditions showed that a similar number of metabolite features were detected with either 50% or 80% acetonitrile (Figure 2C). Overall these results corroborate other detailed studies on the impact of solvent composition³⁰.

Feature Detection as a Function of Sample Loading –

We next assessed the binding and peak capacity of the nano LC column using a 30 min gradient from 1–75 % buffer B. Yeast samples were extracted and serially diluted to maintain a constant injection volume for analysis. This array of diluted yeast metabolite extracts was analyzed in duplicate and each injection level was analyzed by XCMS. Features were discarded if their peak area was less than the noise threshold. The data fit a hyperbolic curve ($R^2=0.99$), with half-maximal features identified at a loading of 6.1×10^6 yeast cells (Figure 3A). Under these chromatographic conditions, there was almost no benefit to loading more than 5×10^7 cells in terms of observed features. A logarithmic plot of the same loading curve shows the linear and plateau regions (Figure 3A inset). We next examined whether the failure to observe more features at higher sample loading amounts was due to column saturation. We observed that the amount of sample loaded was linearly correlated with peak area (average $R^2 = 0.94$) as seen in a representative set of features (Figure 3B), indicating that the column was not saturated by sample loading, and that feature detection may be limited by the separation power of LC and/or MS. These data also demonstrate that the peak area of reported features is proportional to the amount of sample loaded over a dynamic range of at least 3 orders of magnitude. We then manually examined a large number of the reported features to determine if high sample loading affects chromatographic performance. The extracted ion chromatogram of 325.246 m/z shows that the peak shape remained unchanged through low (6×10^5 cells) and high (6×10^7 cells) sample loading (Figure 3C), suggesting that LC performance did not deteriorate with high loading.

These experiments show that the nUPLC-HRMS system can detect several thousand authentic features with a linear response across 3–4 orders of magnitude. The instrument's dynamic range is likely greater than this, but attempts at higher sample loading levels are hindered by the solubility of the metabolites and the 5 μ L sample loading loop in our configuration. Additionally, larger injection volumes would compromise throughput due to longer loading times on the nano-LC system. Human blood metabolites are known to span at least 9 orders of magnitude³¹ from picomolar (e.g. estradiol³²) to millimolar (e.g. glucose³³) concentrations. Many metabolites are only conditionally present because the metabolome is closely correlated with the biological phenotype, which is inherently dynamic. Therefore, considerable effort will be required to fully catalogue the metabolome of a given species or cell-type³. Finally, our method may not capture highly hydrophilic or unstable metabolites.

Distribution of Metabolite m/z and Effect of MS Resolution –

We also aimed to characterize some basic attributes of the yeast metabolome. A histogram of m/z versus proportion of features detected shows that the central tendency of the population is around 300 m/z (Figure 4A). About 97% of all detected features were between 150 and 600 m/z , though some features up to ~900 m/z were detected. The number of detected features might increase with better MS resolution (from 15,000 to 480,000, Figure 4B). To evaluate the effect of MS resolution, we calculated the theoretical benefit of high resolution using an accurate mass molecular formula generator. We generated all possible molecular formulas ($n = 15,652$) with a monoisotopic mass between 50 and ~600 units which fit the pattern $C_{1-30}H_{0-42}N_{0-6}O_{0-6}$, based on the observed mass range of metabolites and known biological metabolite formulas. The resulting table of theoretical formulas was sorted by mass and plotted in a histogram (Figure 4C). For example, at lower resolution (15,000) many metabolite features became indistinguishable, confounding the interpretation of results. A resolution of 120,000 could enable 3.6-fold more features to be distinguished compared to a resolution of 15,000. We then empirically determined the impact of MS resolution on sensitivity and feature detection using the 30 minute LC gradient. As expected, higher instrument resolution increased feature detection up to a resolution of 120,000 and the ascending trend was highly consistent with the theoretical simulation (3.9-fold more features with 120,000 versus 15,000, Figure 4D). Surprisingly, if MS was set higher than 120,000 it led to a decrease in the total number of observed features. This loss of features is likely due to the longer transient time required for higher resolution, and therefore a lower sampling frequency. For instance, at 120,000 the MS1 cycle time is ~0.43 seconds versus ~1.6 seconds at 480,000. Very narrow and/or low abundant peaks may not be acquired or have fewer points per peak leading to inconsistent detection by automated peak algorithms. Moreover, at high resolution (>100,000), carbon, nitrogen, and other isotopes were detectable for many isotopic clusters³⁴. This rich source of isotopic information may prove useful for putative metabolite identification. Our data indicate that a resolution of 120,000 is optimal for balancing sensitivity, discrimination power, and sampling rate, in global metabolomics studies in this system.

The total number of empirical features was much lower than the theoretical number of formulas, but the theoretical distribution still provided an excellent estimate of the relative benefit of high resolution instrumentation. As a general rule, the number of discriminated metabolites appears to increase 10% for every 10k increase in MS resolution. These data suggest that there may be very little benefit to using an MS resolution higher than 120,000 with respect to feature discrimination. However, future instruments may still improve scan speed resulting in more frequent MS scans and more data points for each peak, potentially improving the quantitative reproducibility of the system. No matter how high the MS resolution, isomeric metabolites will remain indistinguishable in standard MS scans. In this case, differing LC retention time and MS/MS data may provide assistance. Further advances in distinguishing formula and structural isomers will likely rely on different LC systems or ion mobility mass spectrometry³⁵.

Effect of Gradient Length on Feature Detection –

We next tested four different gradient lengths (15, 30, 60, and 90 min) to determine if a longer gradient may afford better chromatographic resolution and more features. Each of the four gradients were tested at three different sample levels in duplicate to determine which conditions provided the best compromise between feature detection and sensitivity. The resulting chromatograms show that longer gradients result in more resolution between metabolite features (Figure 5A). Slower gradients were critical for achieving baseline resolution between isobars. For instance, the extracted ion current of 440.273 m/z shows that two isobaric compounds, which cannot be distinguished at 0.5 ppm, can be resolved with sufficient gradient time (Figure 5B). Such separations are common in our data, and partly explain why longer gradients afford much higher numbers of metabolite features. For low and medium amounts of sample, there is a maximum benefit to running a longer gradient, around 60 min (Figure 5C). Indeed, longer gradients can even decrease the number of observed features by decreasing sensitivity through peak broadening, a phenomenon that is also observed during LC-MS analysis of complex peptide mixtures³⁶. At the highest level of sample we injected, the 90 minute gradient identified 16% more features than the 60 min gradient. However, the 60 minute gradient proved to be more economical for high-throughput analysis and performed better at low and medium sample loadings, making it the primary choice for global analyses.

Gradient Optimization –

Peak density plots revealed that the simple linear gradient from 1–75 %B led to a surge of metabolite features early in the gradient followed by a trailing-off of detected features. We set out to improve the gradient program to increase the efficiency of instrument use and gain more detected features. Using the optimal resolution (120,000) and 60 minute gradient, we compared the simple linear gradient to a biphasic gradient at a sample loading of 6×10^7 yeast cells for each method (Figure 6). The biphasic gradient gave a 60% increase in the number of observed features ($n = 6,472$) compared to the linear gradient. This large improvement is explained by the feature density plot (Figure 6B) which shows that with the biphasic gradient, features are more evenly distributed throughout the run. It may be of interest that the peak number of features in either density plot is ~250/min. This value may represent a physical limitation such as ion suppression and/or dynamic range limitations of the orbitrap.

Application to Multiple Myeloma Cells –

To determine the applicability of the method for testing metabolomics hypotheses in mammalian systems, we cultured and analyzed bortezomib resistant (BR) and bortezomib sensitive (BS) cells in quadruplicates (Figure 7A). We observed 8,776 metabolite features across the 8 samples analyzed (minfrac=1, each group). After further noise filtering and accurate mass database search, we were able to detect 5,799 features, in which 1,060 features were tentatively assigned a KEGG ID (Figure 7B). We next examined the fold-change ratios of the features between BR and BS cell lines. A histogram of their distribution (Figure 7C) shows a normally distributed population with the majority of metabolites varying by less than 4-fold. In contrast, among features with KEGG IDs, the most down-

and up-regulated metabolites varied by 86 and 34-fold respectively. The metabolite data were visualized using a heat map to determine if such data might distinguish between sample groups (Figure 7D). Stark differences were observed between the BR and BS groups, with good reproducibility among replicates. These data indicate that there may be broad metabolic dysregulation due to background bortezomib toxicity and/or metabolic alterations contributing to drug resistance. To further examine the potential role of metabolites as markers of the BR phenotype, we performed a Principal Components Analysis (PCA) and visualized the data with Multi-Dimensional Scaling (MDS) (Figure 7E). The tentatively identified metabolites were assessed by metabolite ontology analysis to identify potential pathways to target in future validation experiments. Overall there were significant alterations in the anabolism/catabolism of purines, pyrimidines (Supplemental S1), various CoAs, among others (Table S1).

CONCLUSIONS

Our newly described nUPLC-HRMS method makes major advances in the sensitivity and specificity of global metabolomics studies. These data demonstrate that up to ~6,500 authentic metabolite features can be detected from a couple milligrams of sample material. Inadequate noise filtering can lead to spurious and fruitless feature identifications, which can be mitigated by incorporating blank samples into feature searching. We also found that there is no theoretical or empirical benefit to MS resolutions higher than 120,000 in general metabolomics analyses. A case study of bortezomib resistant myeloma cells revealed a broad range of metabolic disturbances which may contribute to the drug resistant phenotype. Both PCA and heatmap techniques demonstrated the high information content of metabolite profiles. This method is also amenable to clinical biopsy samples where tissue material may be limited. These studies provide a framework for performing global metabolomics studies with a platform that is accessible to a broad range of proteomics scientists³⁷.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank for Y. Yang for yeast culture, Z. Wu for laboratory assistance, H. Tan, A. High and V. Pagala for MS instrument guidance, Y. Xu and X. Wang for software installation, and other lab and facility members for helpful discussion.

Funding Sources

This work was partially supported by the American Cancer Society grant (RSG-09-181) and ALSAC (American Lebanese Syrian Associated Charities).

ABBREVIATIONS

nUPLC-HRMS	nano ultra-performance liquid chromatography – high resolution mass spectrometry
FDR	false discovery rate

EIC	extracted ion chromatogram
ACN	acetonitrile
CE	capillary electrophoresis
NMR	nuclear magnetic resonance
GC-MS	gas chromatography – mass spectrometry
PCA	principal components analysis
MDS	multi-dimensional scaling
BR/BS	bortezomib resistant/sensitive

REFERENCES

- (1). Roberts LD; Souza AL; Gerszten RE; Clish CB *Curr Protoc Mol Biol* 2012, Chapter 30, Unit 30 2 1–24.
- (2). Wishart DS *Bioanalysis* 2011, 3, 1769–82; Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat Rev Mol Cell Biol* 2012, 13, 263–9. [PubMed: 21827274]
- (3). Wishart DS; Jewison T; Guo AC; Wilson M; Knox C; Liu Y; Djoumbou Y; Mandal R; Aziat F; Dong E; Bouatra S; Sinelnikov I; Arndt D; Xia J; Liu P; Yallou F; Bjorn Dahl T; Perez-Pineiro R; Eisner R; Allen F; Neveu V; Greiner R; Scalbert A *Nucleic Acids Res* 2013, 41, D801–7. [PubMed: 23161693]
- (4). Zhu ZJ; Schultz AW; Wang J; Johnson CH; Yannone SM; Patti GJ; Siuzdak G *Nat Protoc* 2013, 8, 451–60. [PubMed: 23391889]
- (5). Prasain JK; Arabshahi A; Taub PR; Sweeney S; Moore R; Sharer JD; Barnes S J *Chromatogr B Analyt Technol Biomed Life Sci* 2013, 913–914, 161–8.
- (6). Zheng J; Dixon RA; Li L *Anal Chem* 2012, 84, 10802–11. [PubMed: 23150892]
- (7). Jain M; Nilsson R; Sharma S; Madhusudhan N; Kitami T; Souza AL; Kafri R; Kirschner MW; Clish CB; Mootha VK *Science* 2012, 336, 1040–4. [PubMed: 22628656]
- (8). Panopoulos AD; Yanes O; Ruiz S; Kida YS; Diep D; Tautenhahn R; Herreras A; Batchelder EM; Plongthongkum N; Lutz M; Berggren WT; Zhang K; Evans RM; Siuzdak G; Izpisua Belmonte J. C. *Cell Res* 2012, 22, 168–77. [PubMed: 22064701]
- (9). Hou P; Li Y; Zhang X; Liu C; Guan J; Li H; Zhao T; Ye J; Yang W; Liu K; Ge J; Xu J; Zhang Q; Zhao Y; Deng H *Science* 2013, 341, 651–4. [PubMed: 23868920]
- (10). Nicholson JK; Holmes E; Kinross JM; Darzi AW; Takats Z; Lindon JC *Nature* 2012, 491, 384–92. [PubMed: 23151581]
- (11). Chan EC; Pasikanti KK; Nicholson JK *Nat Protoc* 2011, 6, 1483–99. [PubMed: 21959233]
- (12). Nakamizo S; Sasayama T; Shinohara M; Irino Y; Nishiumi S; Nishihara M; Tanaka H; Tanaka K; Mizukawa K; Itoh T; Taniguchi M; Hosoda K; Yoshida M; Kohmura E *J Neurooncol* 2013, 113, 65–74. [PubMed: 23456655]
- (13). Sana TR; Gordon DB; Fischer SM; Tichy SE; Kitagawa N; Lai C; Gosnell WL; Chang SP *PLoS one* 2013, 8, e60840; Kim, B.; Moon, J. Y.; Choi, M. H.; Yang, H. H.; Lee, S.; Lim, K. S.; Yoon, S. H.; Yu, K. S.; Jang, I. J.; Cho, J. Y. *J Proteome Res* 2013, 12, 1359–68; Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D.; Lehmann, R.; Schmitt-Kopplin, P. *J Chromatogr A* 2013, 1292, 51–65. [PubMed: 23593322]
- (14). Kaplan KA; Chiu VM; Lukus PA; Zhang X; Siems WF; Schenk JO; Hill HH, Jr. *Anal Bioanal Chem* 2013, 405, 1959–68; Derewacz, D. K.; Goodwin, C. R.; McNees, C. R.; McLean, J. A.; Bachmann, B. O. *Proc Natl Acad Sci U S A* 2013, 110, 2336–41. [PubMed: 23314481]
- (15). Myint KT; Uehara T; Aoshima K; Oda Y *Anal Chem* 2009, 81, 7766–72. [PubMed: 19702289]

- (16). Burgess K; Creek D; Dewsbury P; Cook K; Barrett MP *Rapid Commun Mass Spectrom* 2011, 25, 3447–52. [PubMed: 22002700]
- (17). Ni Q; Reid KR; Burant CF; Kennedy RT *Anal Chem* 2008, 80, 3539–46. [PubMed: 18399659]
- (18). Yates JR; Ruse CI; Nakorchevsky A *Annu Rev Biomed Eng* 2009, 11, 49–79. [PubMed: 19400705]
- (19). Na CH; Jones DR; Yang Y; Wang X; Xu Y; Peng J *Journal of proteome research* 2012, 11, 4722–32. [PubMed: 22871113]
- (20). Yuan M; Breitkopf SB; Yang X; Asara JM *Nat Protoc* 2012, 7, 872–81. [PubMed: 22498707]
- (21). Team RC; R Foundation for Statistical Computing: Vienna, Austria, 2013.
- (22). Kessner D; Chambers M; Burke R; Agus D; Mallick P *Bioinformatics* 2008, 24, 2534–6. [PubMed: 18606607]
- (23). Tautenhahn R; Bottcher C; Neumann S *BMC Bioinformatics* 2008, 9, 504. [PubMed: 19040729]
- (24). Prince JT; Marcotte EM *Anal Chem* 2006, 78, 6140–52. [PubMed: 16944896]
- (25). Smith CA; Want EJ; O'Maille G; Abagyan R; Siuzdak G *Anal Chem* 2006, 78, 779–87. [PubMed: 16448051]
- (26). Arakawa K; Kono N; Yamada Y; Mori H; Tomita M *In Silico Biol* 2005, 5, 419–23. [PubMed: 16268787]
- (27). Crews B; Wikoff WR; Patti GJ; Woo HK; Kalisiak E; Heideker J; Siuzdak G *Anal Chem* 2009, 81, 8538–44. [PubMed: 19764780]
- (28). Keller BO; Sui J; Young AB; Whittall RM *Anal Chim Acta* 2008, 627, 71–81. [PubMed: 18790129]
- (29). Want EJ; Masson P; Michopoulos F; Wilson ID; Theodoridis G; Plumb RS; Shockcor J; Loftus N; Holmes E; Nicholson JK *Nat Protoc* 2013, 8, 17–32. [PubMed: 23222455]
- (30). Want EJ; O'Maille G; Smith CA; Brandon TR; Uritboonthai W; Qin C; Trauger SA; Siuzdak G *Anal Chem* 2006, 78, 743–52. [PubMed: 16448047]
- (31). Want EJ; Cravatt BF; Siuzdak G *Chembiochem* 2005, 6, 1941–51. [PubMed: 16206229]
- (32). Lee JS; Yaffe K; Lui LY; Cauley J; Taylor B; Browner W; Cummings S *Arch Neurol* 2010, 67, 195–201. [PubMed: 20142527]
- (33). Maeda E; Kataoka M; Hino M; Kajimoto K; Kaji N; Tokeshi M; Kido J; Shinohara Y; Baba Y *Electrophoresis* 2007, 28, 2927–33. [PubMed: 17640093]
- (34). Thurman EM; Ferrer I *Anal Bioanal Chem* 2010, 397, 2807–16. [PubMed: 20229008]
- (35). Merenbloom SI; Glaskin RS; Henson ZB; Clemmer DE *Anal Chem* 2009, 81, 1482–7; Glaskin, R. S.; Ewing, M. A.; Clemmer, D. E. *Anal Chem* 2013. [PubMed: 19143495]
- (36). Xu P; Duong DM; Peng JJ *Proteome Res.* 2009, 8, 3944–50.
- (37). Fischer R; Bowness P; Kessler BM *Proteomics* 2013, 201300192, [Epub ahead of print].

Figure 1.

(A) nUPLC column set-up. (B) Base-peak chromatograms demonstrating interday reproducibility (across 2-weeks). (C) Scatter plot and (D) histogram showing the reproducibility of replicate analyses ($n = 4$) of yeast metabolite extracts. (E) Scatter plot and (F) histogram estimating the false discovery rate of pairwise metabolomics discovery experiments in a null experiment.

Figure 2.

(A) Sample workflow. (B) The level of protein in metabolite extracts was assessed by BCA assay in independent triplicates. For the samples extracted by 80% acetonitrile, the minimal detection amount of the method was used to represent an upper limit of protein in the sample. (C) Yeast metabolite extracts were analyzed in duplicate to determine how lysis conditions impacted feature detection.

Figure 3.

(A) Hyperbolic fit ($R^2 = 0.99$) showing the saturation of detectable metabolite features with respect to amount of sample loaded with semi-log inset plot. (B) Selected 25 features showing the linear response of feature peak area with respect to amount of sample loaded. (C) Extracted ion chromatogram of the 325.246 m/z ion demonstrating peak symmetry at low (6e5 cells) and high (6e7 cells) levels of sample loading; scientific notation indicates peak area as detected using the Genesis algorithm.

Figure 4.

(A) Distribution of feature mass-to-charge ratios in yeast metabolite extracts. (B) Overlay of MS1 monoisotopic peak (403.279 m/z) at various MS resolutions. (C) Theoretical impact of MS resolution on feature discrimination. (D) Empirical impact of MS resolution on feature detection vs. predicted.

Figure 5.

(A) Base-peak chromatograms showing the impact of gradient length on chromatographic resolution. (B) Extracted ion currents (EICs) of 440.273 m/z under different gradient lengths showing increasing temporal resolution of isobaric metabolite features with longer gradient methods. (C) Duplicate analyses exploring feature detection as a function of gradient length and sample loading.

Figure 6.

(A) Gradient composition profiles for the linear (blue) and biphasic (red) methods. (B) Peak density plot through time for the linear and biphasic gradients. (C) Impact of gradient design on metabolite feature detection in yeast extracts.

Figure 7.

(A) Workflow for pairwise analysis of drug-sensitive and drug-resistant multiple myeloma cell lines. (B) Metabolite features detected and identified in myeloma cells. (C) Distribution of metabolite feature ratios between drug resistant and sensitive myeloma cells. Up and down-regulated metabolites are defined as being in the lower or top 5% as indicated by the dashed red lines. (D) Heat map of unfiltered metabolite features showing distinct clustering of BS and BR sample groups. (E) Multi-dimensional scaling plot showing distinct grouping between BR (red) and BS (blue) samples groups based on metabolite features.