Published in final edited form as:

Anal Chem. 2018 July 17; 90(14): 8538–8545. doi:10.1021/acs.analchem.8b01591.

Isotope Labeling-assisted Evaluation of Hydrophilic and Hydrophobic Liquid Chromatograph-Mass Spectrometry for Metabolomics Profiling

Boer Xie^{#†}, Yuanyuan Wang^{#†}, Drew R. Jones^{#†,⊥}, Kaushik Kumar Dey[†], Xusheng Wang[‡], Yuxin Li[‡], Ji-Hoon Cho[‡], Timothy I. Shaw^{‡,§}, Haiyan Tan[‡], and Junmin Peng^{†,‡,*}

[†]Departments of Structural Biology and Developmental Neurobiology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA

[‡]St. Jude Proteomics Facility, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA

[§]Department of Computational Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA

[#] These authors contributed equally to this work.

Abstract

High throughput untargeted metabolomics usually relies on complementary liquid chromatography-mass spectrometry (LC-MS) methods to expand the coverage of diverse metabolites, but the integration of those methods is not fully characterized. We systematically investigated the performance of hydrophilic interaction liquid chromatography (HILIC)-MS and nano-flow reverse phase liquid chromatography (nRPLC)-MS under 8 LC-MS settings, varying stationary phases (HILIC and C18), mobile phases (acidic and basic pH), and MS ionization modes (positive and negative). Whereas nRPLC-MS optimization was previously reported, we found in HILIC-MS (2.1 mm × 150 mm) that the optimal performance was achieved in a 60 min gradient with 100 µl/min flow rate by loading metabolite extracts from 2 mg of cell/tissue samples. Since peak features were highly compromised by contaminants, we used stable isotope labeled yeast to enhance formula identification for comparing different LC-MS conditions. The 8 LC-MS settings enabled the detection of a total of 1,050 formulas, among which 78%, 73%, and 62% formulas were recovered by the best combination of 4, 3, and 2 LC-MS settings, respectively. Moreover, these yeast samples were harvested in the presence or absence of nitrogen starvation, enabling quantitative comparisons of altered formulas and metabolite structures, followed by validation with selected synthetic metabolites. The results revealed that nitrogen starvation

*Corresponding Author: junmin.peng@stjude.org.

[⊥]Current address: Department of Biochemistry and Molecular Pharmacology, Langone Medical Center, New York University, NY, 10016, USA

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

Supporting Information Available: peak capacity calculation for label free samples, evaluation of quantitative data, theoretical evaluation of mass accuracy and isotope labeling on formula identification, and additional figures and tables.

The authors declare no competing financial interest.

downregulated amino acid components but upregulated uridine-related metabolism. In summary, this study introduces a thorough evaluation of hydrophilicity and hydrophobicity-based LC-MS, and provides information for selecting complementary settings to balance throughput and efficiency during metabolomics experiments.

Graphical Abstract

Keywords

metabolomics; metabolome; mass spectrometry; HILIC; RPLC; LC-MS; stable isotope labeling; multiplexing

INTRODUCTION

Metabolomics, viewed as the end point of the “omics” central dogma, is the systematic analysis of small molecules to provide unique insights into fundamental biological research.

¹ Combined with genomics, transcriptomics, and proteomics, an integrative “omics” study can be conducted for a holistic understanding of biological mechanisms in complex systems.

² Metabolites directly reflect biochemical activities, and are highly diverse in chemical and physical properties.^{2–5} Liquid chromatography (LC)-mass spectrometry (MS) based global metabolomics is a powerful analytical tool enabling detection, quantitation, and structure elucidation of hundreds of metabolites with high sensitivity, but fully recovering and identifying the complete metabolome from a complex biological sample is still a challenge.^{1,6–9}

Efficient LC separation prior to MS detection is a key step to resolve metabolites, reducing ion suppression, and thus improving sensitivity and metabolome coverage.^{3,10} Because of its stability and versatility, reverse phase liquid chromatography (RPLC) is one of the most

implemented LC systems for global metabolomics. We and other groups reported the systematic optimization of RPLC-high resolution MS for metabolomics studies, suggesting that in nRPLC-MS (75 $\mu\text{m} \times 100\text{ mm}$), the best performance was obtained in a 60 min gradient with 0.25 $\mu\text{L}/\text{min}$ flow rate by injecting metabolites from 2 mg of cells.^{11–13} However, highly polar and ionic metabolites retain poorly on RPLC columns, preventing accurate identification and quantification.^{3,14,15} Ion-pair reagents may be introduced into the RPLC mobile phase to improve retention and separation, but they also lead to MS signal suppression, and thus are not preferable for LC-MS based metabolomics.^{3,16,17} To overcome the limitations of RPLC, hydrophilic interaction chromatography (HILIC) has gained popularity for analyzing polar metabolites.^{15,18–20} The majority of the HILIC-MS methods were specifically optimized for certain chemical classes.^{21–24} A comprehensive analysis covering both hydrophilic and hydrophobic analytes can be achieved by combining HILIC and RPLC, but the assessment of these studies were mainly based on the number of detected peak features, not the identifiable metabolites.^{25–27}

In addition to stationary phase, detection of diverse metabolites by LC-MS also depends on mobile phase and MS ionization mode.^{13,20,28} Mobile phase pH is another key factor for metabolite detection due to its influence on chemical properties and functional groups of the analytes, which affects retention time, LC peak width, and MS ionization.²⁹ As a consequence of this high diversity of metabolite properties, a single mobile phase pH cannot retain and separate different classes of metabolites. Besides commonly used positive ion mode, a large number of small molecules are easily ionized in negative ion mode especially true for organic acids and alcohols.^{5,30} Thus, both MS ionization modes should be employed to improve metabolome coverage.

Peak features are defined by unique m/z and retention time, and are usually used as the main criterion for untargeted metabolomics because of the challenges in reliable identification of metabolites.^{10,26,27,31} But peak features are not an ideal measurement for comparing runs using different LC-MS settings, because the peak features are highly confounded by contaminated, non-metabolite MS peaks.³² Indeed, Mahieu et al. reported that less than 10% of the peak features from an LC-MS analysis are from unique metabolites.³² Moreover, it is difficult to compare peak features under different LC-MS settings, due to changes in retention time. Chemical formulas serve as a better criterion for evaluation and comparison of different LC-MS settings. With the assistance of stable isotope labeling technique, a commonly used method for dynamics metabolic flux study, enhanced formula assignment and the ability to discriminate between contamination and metabolite peaks can be achieved.^{33–37}

In this study, we report a comprehensive optimization of HILIC-MS method and its integration with previously optimized nRPLC-MS¹¹ in 8 different LC-MS settings. To facilitate the comparative analysis, we used stable isotope labeled yeast samples for elucidating reliable chemical formulas by isotope labeled pairs. Finally, we applied this HILIC and nRPLC-MS based metabolomics pipeline to a study of nitrogen starved yeast cells to simultaneously identify and quantify metabolites response to nitrogen starvation.

EXPERIMENTAL SECTION

Stable Isotope Labeling of Yeast

Saccharomyces cerevisiae (Fleischmann) were cultured under four different minimal media. The natural isotopic abundance media was prepared using yeast nitrogen base without amino acids and ammonium sulfate (BD Biosciences), with the addition of 5 g/L ammonium sulfate and 20 g/L glucose (Sigma). In the other three media, ^{13}C -6 glucose and ^{15}N -2 ammonium sulfate were used separately or in combination. Each culture was maintained for at least 30 generations to ensure fully labeling before nitrogen starvation. During nitrogen starvation, ammonium sulfate was decreased to 0.5 g/L. All cultures were seeded to an OD_{600} of 0.1, grew to 0.5, and then harvested by centrifuging at 3,000 g for 5 min.

Metabolite Extraction and Pooling

Metabolites were extracted from rat brain tissues or yeast cells essentially following our reported protocol.¹¹ Rat brain tissues (Pel Freez Biologicals) were pulverized under liquid nitrogen, and extracted by freezing 80% (v/v) acetonitrile (ACN, 10 μL per mg tissue) along with 1.0 mm glass beads (Next Advance, 2 μL beads per mg tissue), by vortexing (3 min, 3,000 rpm in a 1 on/1 off pattern to prevent overheating). The lysate was centrifuged at 21,000 g for 5 min and the supernatant was dried for storage. Yeast cells were extracted using the similar protocol. For equal pooling, metabolite concentrations in the lysates were analyzed by UV absorbance at 300 nm as previously reported.³⁸ The pooled mixture was aliquoted and dried for further analysis.

Acidic pH and basic pH nRPLC-MS/MS analysis

The acidic buffers consisted of buffer A (0.2% formic acid) and B (0.2% formic acid in 100% ACN), while the basic pH buffers were buffer A (0.5 mM ammonium fluoride) and B (100% ACN). The dried samples were dissolved with buffer A, and analyzed by a self-packed column (75 $\mu\text{m} \times 100 \text{ mm}$, 1.9 μm C18 beads, Dr. Maisch GmbH) connected to Waters nanoAcquity UPLC and Orbitrap Q Exactive HF (Thermo Scientific). LC-MS settings included 0.25 $\mu\text{L}/\text{min}$ flow rate, biphasic gradient in 60 min as previously reported,¹¹ mass range (50–750 m/z or 100–1,500 m/z), MS1 parameters (120K resolution, 1E6 AGC, and 50 ms maximal injection time), 1.6 Da isolation window, TOP20 MS2 parameters (30K resolution, 1E5 AGC, and ~100 ms maximal injection time), and dynamic exclusion of 20 sec. Eluted metabolites were ionized in positive (2.5 kV) or negative ion mode (–2 kV).

Acidic pH and basic pH HILIC-MS/MS analysis

Buffer A (20 mM ammonium acetate in 90% ACN) and B (20 mM ammonium acetate in water) were adjusted by acetic acid (to pH 3) or ammonium hydroxide (to pH 8). The dried samples were solubilized in 80% ACN, and analyzed by a SeQuant ZIC-HILIC column (2.1 mm $\times$ 150 mm, 3.5 μm resin, EMD Millipore) with Waters Acquity UPLC and Orbitrap Q Exactive HF. LC-MS settings included 100 $\mu\text{L}/\text{min}$ flow rate, 90 min triphasic gradient (1–5% B in 20 min, 5–30% B in 60 min, 30–50% B in 10 min), the same MS1 and MS2 parameters as in the RPLC-MS/MS analysis. Eluted metabolites were ionized in positive (3.8 kV) or negative ion mode (–3 kV).

Metabolite identification and quantification for labeled samples

The metabolite analysis was performed by a newly developed software suite JUMPm³⁹. The comparison between JUMPm and XCMS on peak feature detection can be found in Table S1. RAW files were converted to mzXML format, followed by peak feature extraction. Stable isotope labeled pairs were detected to derive nitrogen and carbon numbers in the metabolites for enhanced formula identification. MS2 spectra were searched against the yeast metabolome database 1.0 (YMDB, 2,007 entries)⁴⁰ to match tentative metabolites. The assigned metabolites were filtered by matching scores to reduce false discovery rate (FDR) to less than 1%, and were further manually examined by matching experimental MS2 spectra with published spectra libraries, MetFrag⁴¹ and mzCloud⁴².

Metabolite qualification was first performed individually for each LC-MS setting by calculating fold changes between normal and starved conditions for each pair in MS scans followed by summarization. If one formula was identified by multiple LC-MS settings, the fold changes from each setting were compared. If the difference is within 2 standard deviation ($SD = 0.5$, $p > 0.05$), these measurements were considered for the same component and the data were averaged. Otherwise, these measurements might be derived from different isomers with the same formula, and the data were not merged.

RESULTS AND DISCUSSION

In untargeted metabolomics, maximizing metabolome coverage by the combination of different LC-MS settings is the key to improve the performance of the pipeline. For this purpose, we performed a systematic investigation of the hydrophilicity and hydrophobicity-based metabolomics pipeline under 8 LC-MS conditions (Figure 1). To expand the metabolome coverage, HILIC and RPLC were used to retain polar and non-polar metabolites, respectively, in combination with different pH mobile phases (acidic and basic), as well as positive and negative ionization mode.

Systematic Optimization of HILIC-MS Parameters

We selected a Zwitterionic stationary phase in this study, as this resin balances permanent surface charges to provide both positive and negative functional groups, enabling efficient separation of acidic, basic, and neutral polar compounds, and was reported to yield the best performance in detecting metabolite features.^{26,43} The reproducibility of HILIC-MS was evaluated by three repeated LC-MS runs loaded with a rat brain metabolite sample (Figure 2A). Base peak profiles for these replicates were almost identical, and the retention time of the major peaks were essentially the same (< 1 min shift). The number of peak features detected in each replicate were highly consistent ($6,163 \pm 53$ with the coefficient of variation of 1.5%), demonstrating high reproducibility of the ZIC-HILIC-MS system.

Loading capacity of the ZIC-HILIC column was examined for its influence on peak feature detection. The loading amount was estimated by the brain tissue amount for extraction. When increasing the loading amount from 0.1 to 2 mg tissue equivalent, we detected 2.6 fold more peak features (Figure 2B). Further addition of the loading, however, did not profit peak feature detection, suggesting that the HILIC-MS system was saturated at

approximately 2 mg. Consistently, we also performed a loading titration of yeast cells and obtained the similar result (equivalent to ~2 mg yeast cells, Figure S1). To further understand the relationship of sample loading amount and detected peak features, we examined the ion intensity and peak widths during the titration. In the example of tyrosine, the ion intensity (i.e., peak height) was increased ~73 fold from 0.1 mg to 17 mg, and the peak width was also changed from 0.17 to 0.62 min (Figure 2C), indicating the benefit of increased ion intensity is offset by peak broadening. In addition, with increased loading amounts, the ion intensity could be compromised by LC-MS matrix effect led by co-eluting ions.⁴⁴ This phenomenon was also observed in the optimization of nRPLC-MS analysis.⁴⁵ Thus, the 2 mg loading amount was selected as a standard for the HILIC-MS method.

Flow rate is another important parameter affecting column efficiency. The HILIC-MS system was tested with flow rates ranged from 400 μ l/min down to 20 μ l/min with 1–50% buffer B gradient profile (Figure 3A). When the flow rate decreased from 400 to 100 μ l/min, more peak features were detected, because the slower flow rate resulted in more concentrated eluates to increase sensitivity. Further decrease of flow rate to 50 μ l/min did no benefit on peak feature detection, and 21.1% fewer peak features were observed for 20 μ l/min flow rate, which may be associated with longer elution time and peak broadening at lower flow rate. Indeed, the peak widths for tyrosine was changed from 0.11 min at 400 μ l/min to 0.83 min at 20 μ l/min (Figure 3B). Therefore, we fixed the flow rate at 100 μ l/min for this HILIC-MS system.

Optimal elution time was also determined for this HILIC-MS method by using various LC gradients from 10 to 160 min. The resulting curve exhibited a plateau at around 80-min time point with 6,900 peak features, representing a 55.8% increase as compared to the 10-min time point (Figure 3C). Longer elution time allows better separation to reduce MS ion suppression but increase the diffusion of metabolites to broaden peaks. As expected, 160-min time point only gave a 2.1% increase of peak features. For instance, the peak widths of tyrosine was raised from 0.14 to 0.65 min from 10 to 160 min elution (Figure 3D).

In addition, we calculated peak capacity for the HILIC-MS runs with various elution time. Peak capacity, defined as the number of peaks that can be separated during a gradient run with a certain resolution, is a reliable measurement of LC performance. The peak width used for peak capacity calculation was defined as 4 standard deviation (SD) of the detected peaks (Figure 3E). The average peak width for all detected features across the entire elution was computed and the corresponding peak capacity for each gradient was derived. A positive correlation was observed when peak capacities were plotted as a function of elution time. The peak capacity rises from ~25 to 60 as the elution time increase from 10 to 80 min (Figure 3F), which is consistent with the observation of increased peak features. Lengthening gradient to 160 min increase peak capacity to ~70 whereas detected peak features improved slightly. To balance the column performance and experiment efficiency, a 90 min elution time was recommended to this HILIC-MS method.

Evaluation of Combined HILIC/RPLC-MS Metabolomics Pipeline

The optimized dual LC-MS metabolomics pipeline was applied to stable isotope labeled yeast samples for performance evaluation. Stable isotope labeling is a credentialing

technology, which enables removal of contamination peaks from data to enhance biological peak feature detection, and derives labeled atom numbers (e.g. nitrogen and carbon) for accurate identification (Figure S2).^{33–37} Results from theoretical evaluation of isotope labeling on formula identification also indicated that this strategy has the potential to increase the confidence of metabolite identifications in untargeted experiments by dramatically decreasing the number of possible formulas that could match to one specific mass (Figure S3). Thus, chemical formulas were determined with the assistance of stable isotope labeling for rational assessment of this pipeline.

Evaluation was conducted in the following steps. First, elution profiles were compared for each baseline chromatogram under different LC-MS settings (Figure 4A and 4B). Under the 8 different LC-MS conditions (HILIC/nRPLC-MS, acidic/basic pH, and positive/negative ion mode), chromatographic patterns were changed as a result of various separation chemistries induced by stationary phase, mobile phase and analytes. Next, peak features and metabolite formulas were identified. The numbers of peak features and formulas in each LC-MS condition are listed (Figure 4C). Compared to the unlabeled samples (Figure 3), the isotope labeled samples yielded more peak features, because of the addition of multiple isotopically labeled peaks (Figure 1). Furthermore, though thousands of peak features were determined under each condition, less than 10% of these peak features could be assigned with formulas, consistent with recently reported analyses,³² in which a high level of contaminated ion signals, artifacts and degenerate peaks were suggested. Only biological compounds identified as labeled metabolite formulas, were used for comparison between different LC-MS settings. Additionally, 62–93% of identified formulas overlapped from duplicate runs for each LC-MS setting tested (Table S2), which indicates good reproducibility for formula detection while the small percentage of unique formulas obtained from each duplicated run could be caused by under-sampling issues.

We then plotted the total and unique formulas detected under each LC-MS setting (Figure 5A). Positive-ion mode gave more formulas regardless of the stationary phase, probably because there are more metabolites that favor positive ionization condition. Under positive ESI, acidic pH RPLC-MS had the most identified formulas ($n = 517$), whereas under negative ESI, basic pH HILIC-MS obtained the best result ($n = 243$). If evaluating the effect of mobile phase pH alone, acidic pH provided slightly better results than basic pH in positive ion mode, consistent with the notion that acidic pH promotes protonation compatible with formation of positive ions, while a contrary result was observed for negative ion mode. Despite the majority of formulas being shared between different LC-MS conditions, around 10–26% were unique to one specific condition, supporting the conclusion that combination of multiple LC-MS conditions are important to cover the whole metabolome. This conclusion can also be illustrated by the Venn diagrams that comparing and contrasting formulas detected under different LC-MS settings for HILIC and RP columns separately (Figure S4A, S4B) as well as summed formulas obtained from each column (Figure S4C).

The overlap of different LC-MS conditions was further assessed by plotting the percentage of formulas detected in 1 to 8 conditions (Figure 5B). A total of 1,050 formulas were identified from all 8 conditions. Approximately 40% of the formulas were detected only in

one LC-MS condition; less than 1% of formulas were simultaneously detectable in more than 7 conditions. The overlapped formulas decreased with the addition of more LC-MS conditions, further suggesting the complementarity of these LC-MS settings.

Different arrangements of LC-MS conditions were also studied, and the best combinations of 4, 3, and 2 LC-MS conditions were analyzed (Figure 5C), which recovered 78%, 73%, and 62% of the total identified formulas, respectively. An example is illustrated in Figure 5D, where acidic pH RPLC-MS (ESI+) was merged with basic pH HILIC-MS (ESI-) to show high complementarity. Among the 655 formulas identified, only 16% of the formulas were shared between these 2 conditions demonstrating their orthogonal coverage. In summary, these results have important implications for selecting complementary LC-MS settings to balance throughput and efficiency during metabolomics experiments.

Application to Nitrogen Starved Yeast Cells: Metabolite Quantification, Identification and Validation

During the systematic optimization of the HILIC/RPLC LC-MS metabolomics pipeline, we used yeast samples cultured under normal and nitrogen starvation conditions. Natural isotopic ($^{12}\text{C}^{14}\text{N}$) and $^{12}\text{C}^{15}\text{N}$ labeled yeast cells were designed as duplicates for the normal condition. Similarly, $^{13}\text{C}^{14}\text{N}$ and $^{13}\text{C}^{15}\text{N}$ labeled yeast cells were used as duplicates for nitrogen starvation condition.³⁸ Metabolites from each yeast culture were extracted and mixed in a 1:1:1:1 ratio for 4-plex metabolomics profiling using the 8 LC-MS settings. We then combined the data together and focused on 730 formulas, in which both carbon and nitrogen atom numbers were confirmed by fully isotopic labeling. While the null comparison ($\log_2$ ratio) within replicates represented experimental variation, poor correlation expected ($r = 0.18$), the authentic comparison between the normal and starvation conditions was highly reproducible ($r = 0.95$) (Figure 6A), indicating accurate quantification of metabolites in this study. The null and authentic comparison also allowed the estimation of false discovery rate (FDR) with adjustable cutoff of $\log_2$ ratio. For example, when the cutoff was selected to be 1.5, we accepted 3 and 325 formulas in the null and authentic comparisons, respectively. Thus, the FDR was estimated at less than 1% (0.92%, 3/325, Figure 6B), resulting in 325 accepted formulas. Of these hits, 27 formulas were quantified consistently in at least 3 LC-MS settings (see Supporting Information for details). These 27 formulas were further identified, and manually validated by matching experimental MS2 spectra with published spectra libraries (Figure 6C).

Definitive assignment of metabolites to a formula is often a challenge in large-scale metabolomics,^{4,46,47} since multiple metabolites may share the same formula. For instance, we found that the formula $\text{C}_5\text{H}_{10}\text{N}_2\text{O}_3$ was matched to either glutamine (Figure S5A) or ureidoisobutyric acid (Figure S5B) in the same LC-MS run. Interestingly, the two metabolites exhibited different retention times (56 min and 29 min), different MS2 spectra, and even different $\log_2$ ratio results (-7.9 and 2.8). Taking account of these possibilities, we examined and confirmed the data by retention time, MS2 pattern, and quantification data, to further minimize false identification (Figure 6C).

The metabolic profile alterations upon nitrogen starvation may be explained by direct depletion of those nitrogen-related compounds, and indirect accumulation of metabolites due

to the stress response. Amino acids are logical candidates for growth limitation during ammonium starvation, and are linked to ammonium via glutamine or glutamate.^{48,49} Indeed, we observed significant decreases of glutamine and histidine, a proteinogenic amino acid receiving nitrogen from glutamine (Figure 6C, Figure S6), in agreement with previous publication.^{48–50} We also identified uridine and uridine diphosphate (UDP)-glucose upregulation under starvation as previously reported.^{48–50} Moreover, we identified several novel metabolite changes in this study, such as downregulation of uridine monophosphate (UMP), and upregulation of glycerophosphocholine (GPC).

To validate these novel identifications, we also performed the analysis under the same LC-MS conditions with 8 synthetic metabolite standards (Figure S7). We matched of the m/z, retention time and MS2 spectra providing additional evidence for the correct identification of these metabolites. These findings indicate that our HILIC/RPLC LC-MS metabolomics pipeline was able to profile the yeast cell metabolome with improved confidence in metabolite identification and increased accuracy of data quantification, providing a promising method for metabolomics study.

CONCLUSION

In this study, we systematically investigated the effect of different LC-MS conditions and combinations on the accuracy of metabolite detection at the formula level. Numerous LC-MS parameters were optimized leading to a robust and reproducible global HILIC-MS method. Together with previously optimized nanoscale nRPLC-MS, we evaluated the performance of 8 LC-MS conditions and proposed the best combinations for global metabolomics analysis. The stable isotope labeling technique also allows simultaneous identification and quantification for up to four sets of yeast experimental conditions with enhanced formula identification, representing a multiplexed global metabolomics method for general application.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank all other lab and facility members for helpful discussion. This work was partially supported by National Institutes of Health Grants R01GM114260, R01AG047928, R01AG053987, and ALSAC (American Lebanese Syrian Associated Charities). The MS analysis was performed in the St. Jude Children's Research Hospital Proteomics Facility, partially supported by NIH Cancer Center Support Grant P30CA021765.

ABBREVIATIONS:

HILIC	hydrophilic interaction liquid chromatography
nRPLC	nano-flow reverse-phase liquid chromatography
MS/MS	tandem mass spectrometry

REFERENCES

- (1). Patti GJ; Yanes O; Siuzdak G *Nat Rev Mol Cell Biol* 2012, 13, 263–269. [PubMed: 22436749]
- (2). Patel VR; Eckel-Mahan K; Sassone-Corsi P; Baldi P *Nat Methods* 2012, 9, 772–773.
- (3). Patti GJ *J Sep Sci* 2011, 34, 3460–3469. [PubMed: 21972197]
- (4). Doerr A *Nature Methods* 2017, 14, 32–32.
- (5). Nordstrom A; Want E; Northen T; Lehtio J; Siuzdak G *Anal Chem* 2008, 80, 421–429. [PubMed: 18085752]
- (6). Gowda GA; Djukovic D *Methods in molecular biology* 2014, 1198, 3–12. [PubMed: 25270919]
- (7). Johnson CH; Ivanisevic J; Siuzdak G *Nat Rev Mol Cell Biol* 2016, 17, 451–459. [PubMed: 26979502]
- (8). Wishart DS *Nat Rev Drug Discov* 2016, 15, 473–484. [PubMed: 26965202]
- (9). Lei Z; Huhman DV; Sumner LW *J Biol Chem* 2011, 286, 25435–25442. [PubMed: 21632543]
- (10). Yanes O; Tautenhahn R; Patti GJ; Siuzdak G *Anal Chem* 2011, 83, 2152–2161. [PubMed: 21329365]
- (11). Jones DR; Wu Z; Chauhan D; Anderson KC; Peng J *Anal Chem* 2014, 86, 3667–3675. [PubMed: 24611431]
- (12). Li Z; Tatlay J; Li L *Anal Chem* 2015, 87, 11468–11474. [PubMed: 26482335]
- (13). Waybright TJ; Van QN; Muschik GM; Conrads TP; Veenstra TD; Issaq HJ *J Liq Chromatogr R T* 2006, 29, 2475–2497.
- (14). Lu W; Bennett BD; Rabinowitz JD *J Chromatogr B Analyt Technol Biomed Life Sci* 2008, 871, 236–242.
- (15). Spagou K; Tsoukali H; Raikos N; Gika H; Wilson ID; Theodoridis G *J Sep Sci* 2010, 33, 716–727. [PubMed: 20187037]
- (16). Lu W; Clasquin MF; Melamud E; Amador-Noguez D; Caudy AA; Rabinowitz JD *Anal Chem* 2010, 82, 3212–3221. [PubMed: 20349993]
- (17). Knee JM; Rzezniczak TZ; Barsch A; Guo KZ; Merritt TJ *J Chromatogr B Analyt Technol Biomed Life Sci* 2013, 936, 63–73.
- (18). Buszewski B; Noga S *Anal Bioanal Chem* 2012, 402, 231–247. [PubMed: 21879300]
- (19). Tang DQ; Zou L; Yin XX; Ong CN *Mass Spectrom Rev* 2016, 35, 574–600. [PubMed: 25284160]
- (20). Vorkas PA; Isaac G; Anwar MA; Davies AH; Want EJ; Nicholson JK; Holmes E *Anal Chem* 2015, 87, 4184–4193. [PubMed: 25664760]
- (21). Du Y; Li YJ; Hu XX; Deng X; Qian ZT; Li Z; Guo MZ; Tang DQ *Biomed Chromatogr* 2017, 31.
- (22). Sriboonvorakul N; Leepipatpiboon N; Dondorp AM; Pouplin T; White NJ; Tarning J; Lindegardh N *J Chromatogr B Analyt Technol Biomed Life Sci* 2013, 941, 116–122.
- (23). Martens-Lobenhoffer J; Surdacki A; Bode-Boger SM *Chromatographia* 2013, 76, 1755–1759.
- (24). Gopu CL; Hari PR; George R; Harikrishnan S; Sreenivasan K *J Chromatogr B Analyt Technol Biomed Life Sci* 2013, 939, 32–37.
- (25). Cai XM; Li RB *Sci Rep-Uk* 2016, 6, 1–10.
- (26). Contrepolis K; Jiang L; Snyder M *Mol Cell Proteomics* 2015, 14, 1684–1695. [PubMed: 25787789]
- (27). Ivanisevic J; Zhu ZJ; Plate L; Tautenhahn R; Chen S; O'Brien PJ; Johnson CH; Marletta MA; Patti GJ; Siuzdak G *Anal Chem* 2013, 85, 6876–6884. [PubMed: 23781873]
- (28). Huhman DV; Sumner LW *Phytochemistry* 2002, 59, 347–360. [PubMed: 11830145]
- (29). Dolan JW *Lc Gc N Am* 2017, 35, 22–28.
- (30). Wu Z; Gao W; Phelps MA; Wu D; Miller DD; Dalton JT *Anal Chem* 2004, 76, 839–847. [PubMed: 14750883]
- (31). Tautenhahn R; Patti GJ; Rinehart D; Siuzdak G *Anal Chem* 2012, 84, 5035–5039. [PubMed: 22533540]
- (32). Mahieu NG; Patti GJ *Anal Chem* 2017, 89, 10397–10406. [PubMed: 28914531]

- (33). Giavalisco P; Kohl K; Hummel J; Seiwert B; Willmitzer L *Anal Chem* 2009, 81, 6546–6551. [PubMed: 19588932]
- (34). Giavalisco P; Li Y; Matthes A; Eckhardt A; Hubberten HM; Hesse H; Segu S; Hummel J; Kohl K; Willmitzer L *Plant J* 2011, 68, 364–376. [PubMed: 21699588]
- (35). Zhou R; Tseng CL; Huan T; Li L *Anal Chem* 2014, 86, 4675–4679. [PubMed: 24766305]
- (36). Bueschl C; Krska R; Kluger B; Schuhmacher R *Anal Bioanal Chem* 2013, 405, 27–33. [PubMed: 23010843]
- (37). Chokkathukalam A; Kim DH; Barrett MP; Breitling R; Creek DJ *Bioanalysis* 2014, 6, 511–524. [PubMed: 24568354]
- (38). Jones DR, W. X, Shaw T, Cho JH, Peng J. *Metabolomics (Los Angel)* 2017, 7, 1–6.
- (39). Jones DR; Wang X; Shaw T; Cho J-H; Chen P-C; Dey KK; Zhou S; Li Y; Kim NC; Taylor JP; Kolli U; Li J; Peng J *bioRxiv* 2016, 1–24.
- (40). Jewison T; Knox C; Neveu V; Djoumbou Y; Guo AC; Lee J; Liu P; Mandal R; Krishnamurthy R; Sinelnikov I; Wilson M; Wishart DS *Nucleic Acids Res* 2012, 40, D815–D820. [PubMed: 22064855]
- (41). MetFrag. <https://msbi.ipb-halle.de/MetFragBeta/>, 2010.
- (42). CLOUD, m. z: <https://www.mzcloud.org/>, 2013.
- (43). Lindegardh N; Hanpithakpong W; Phakdeeraj A; Singhasivanon R; Farrar J; Hien TT; White NJ; Day NP J. *Journal of Chromatography A* 2008, 1215, 145–151.
- (44). Trufelli H; Palma P; Famiglini G; Cappiello A *Mass Spectrom Rev* 2011, 30, 491–509. [PubMed: 21500246]
- (45). Xu P; Duong DM; Peng JM J *Proteome Res* 2009, 8, 3944–3950. [PubMed: 19566079]
- (46). Dunn WB; Erban A; Weber RJM; Creek DJ; Brown M; Breitling R; Hankemeier T; Goodacre R; Neumann S; Kopka J; Viant MR *Metabolomics* 2013, 9, S44–S66.
- (47). Sumner LW; Amberg A; Barrett D; Beale MH; Beger R; Daykin CA; Fan TW; Fiehn O; Goodacre R; Griffin JL; Hankemeier T; Hardy N; Harnly J; Higashi R; Kopka J; Lane AN; Lindon JC; Marriott P; Nicholls AW; Reily MD; Thaden JJ; Viant MR *Metabolomics* 2007, 3, 211–221. [PubMed: 24039616]
- (48). Brauer MJ; Yuan J; Bennett BD; Lu W; Kimball E; Botstein D; Rabinowitz JD *Proc Natl Acad Sci U S A* 2006, 103, 19302–19307. [PubMed: 17159141]
- (49). Boer VM; Crutchfield CA; Bradley PH; Botstein D; Rabinowitz JD *Mol Biol Cell* 2010, 21, 198–211. [PubMed: 19889834]
- (50). Klosinska MM; Crutchfield CA; Bradley PH; Rabinowitz JD; Broach JR *Gene Dev* 2011, 25, 336–349. [PubMed: 21289062]

FIGURE 1. Experimental design and procedures for global metabolomics analysis

Yeast cells were stable isotope labeled, extracted, mixed at an equal ratio, and analyzed under 8 different LC-MS settings (HILIC/nRPLC-MS, acidic/basic mobile phase pH, and ESI+/-). Metabolites were identified and quantified for further statistical and biological analysis.

FIGURE 2. Evaluation of HILIC-MS system reproducibility and optimization of loading amount

(A) Base peak chromatograms of three technically repeated runs loaded with metabolites extracted from 2 mg rat brain tissue and eluted in a 1–50% buffer B over 60 min. (B) Number of detected peak features with different loading levels. Data points shown were collected from duplicate runs for each loading amount tested. (C) Effect of loading amounts on the ion intensity and peak width of tyrosine, one of the identified metabolite.

FIGURE 3. Optimization of flow rate and gradient length and evaluation of peak capacity for HILIC-MS system

(A) Number of detected peak features with different flow rates. Data points shown were collected from duplicate runs for each flow rate and the numbers were normalized to the maximal data point. (B) Effect of flow rates on the peak width of tyrosine. (C) Number of detected peak features with different gradient lengths. Data points shown were collected from duplicate runs for each gradient length tested. (D) Effect of gradient time on the peak width of tyrosine. (E) Peak width used for peak capacity calculation was identified as four times of standard deviation for each peak. (F) Peak capacity plotted against gradient length. Peak capacities were calculated as gradient time (min) divided by average peak width.

FIGURE 4. Elution profiles of the global metabolomics pipeline in different LC-MS settings
(A) Base peak chromatograms of HILIC-MS metabolomics analysis of stable isotope labeled yeast cells. (B) Base peak chromatograms of nRPLC-MS metabolomics analysis of stable isotope labeled yeast cells under acid/basic mobile phase pH and ESI+/- with 100–1,500 m/z . (C) Detailed table contains number of detected peak features and metabolite formulas for each of the 8 LC-MS conditions obtained from duplicated runs. Repeated metabolite formulas were counted once for this analysis without considering the existence of possible isomers.

FIGURE 5. Statistical profiles of combined HILIC and RPLC global metabolomics pipeline
 (A) The distribution of unique (light gray) and shared (dark gray) metabolite formulas detected in each LC-MS condition. The total number of detected formulas (combined from two replicates) for each LC-MS condition was shown above each bar. (B) Percentage of formulas detected in 1–8 conditions. (C) Details on combination of complimentary conditions. Total number of formulas obtained from 8 conditions combination was used as a standard to calculate the recovery rate for 4, 3, and 2 conditions combinations. (D) Percentage of formulas detected in 2 conditions combination (acidic pH RPLC-MS in ESI+ and basic pH HILIC-MS in ESI-). Repeated metabolite formulas were counted once for this analysis without considering the existence of possible isomers.

FIGURE 6. Statistical analysis of nitrogen starved yeast cells

(A) Null (intragroup) and real (intergroup) comparison. Each data point represents one identified metabolite formula, and red cycle represents the FDR threshold. (B) FDR vs. log₂ ratio curve used to select FDR cutoff (0.92%) in this study. (C) Details on 27 most significantly changed metabolites detected in at least three different LC-MS conditions with similar fold change values. The metabolites were grouped based on chemical properties.