HHS Public Access

Author manuscript

Analyst. Author manuscript; available in PMC 2020 July 21.

Published in final edited form as:

Analyst. 2019 July 21; 144(14): 4331–4341. doi:10.1039/c9an00560a.

Integrating Comprehensive Two-dimensional Gas Chromatography Mass Spectrometry and Parallel Two-dimensional Liquid Chromatography Mass Spectrometry for Untargeted Metabolomics

Md Aminul Islam Prodhan[1,2,3,4], **Biyun Shi**[1,4], **Ming Song**[2,3,6], **Liqing He**[1,2,3,4], **Fang Yuan**[1,2,3,4], **Xinmin Yin**[1,4], **Patrick Bohman**[8], **Craig McClain**[2,3,5,6,7], **Xiang Zhang**[1,2,3,4,5]

[1]Department of Chemistry, University of Louisville, Louisville, KY 40208, USA

[2]University of Louisville Alcohol Research Center, University of Louisville, Louisville, KY 40208, USA

[3]University of Louisville Hepatobiology & Toxicology Program, University of Louisville, Louisville, KY 40208, USA

[4]Center for Regulatory and Environmental Analytical Metabolomics, University of Louisville, Louisville, KY 40208, USA

[5]Department of Pharmacology & Toxicology, University of Louisville, Louisville, KY 40208, USA

[6]Department of Medicine, University of Louisville, Louisville, KY 40208, USA,

[7]Robley Rex Louisville VAMC, Louisville, Kentucky 40292, USA,

[8]Thermo Fisher Scientific International Inc., 3000 Lakeside Dr., Bannockburn, IL, 60015, USA

Abstract

The diverse characteristics and large number of entities make metabolite separation challenging in metabolomics. To date, there is not a singular instrument capable of analyzing all types of metabolites. In order to achieve a better separation for higher peak capacity and accurate metabolite identification and quantification, we integrated GC×GC-MS and parallel 2DLC-MS for analysis of polar metabolites. To test the performance of the developed system, 13 rats were fed different diets to form two animal groups. Polar metabolites extracted from rat livers were analyzed by GC×GC-MS, parallel 2DLC-MS (−) and parallel 2DLC-MS (+), respectively. By integrating all data together, 58 metabolites were detected with significant change in their abundance levels between groups ($p \leq 0.05$). Of the 58 metabolites, three metabolites were detected in two platforms and two in all three platforms. Manual examination showed that discrepancy of metabolite regulation measured by different platforms was mainly caused by the poor shape of chromatographic peaks resulted from low instrument response. Pathway analysis demonstrated that integrating the results from multiple platforms increased the confidence of metabolic pathway assignment.

*CORRESPONDING AUTHOR: Prof. Xiang Zhang, Department of Chemistry, University of Louisville, **2210 South Brook Street, Louisville, KY 40208, USA. Phone:** +01 502 852 8878. Fax: +01 502 852 8149. xiang.zhang@louisville.edu.

Keywords

GC×GC-MS; 2DLC-MS; untargeted metabolomics; multiple analytical platforms; integration

1. Introduction

Metabolites are the intermediates and products of all biological processes that take place in a biological system. Metabolites can be polar or non-polar, as well as organic or inorganic compounds[1,2]. The diverse chemical characteristics and huge number of entities make metabolite separation challenging in metabolomics. Separation methods such as liquid chromatography (LC) and gas chromatography (GC) are usually coupled with mass spectrometry (MS) to increase the metabolite coverage[3]. Due to the limited peak capacity of a single column, one dimensional separation method using LC or GC can only resolve a limited number of metabolites in a biological sample[4]. In order to obtain better separation and higher peak capacity, multi-dimensional separation methods have been developed and applied in metabolomics even though the multi-dimensional separation usually needs long separation time and reduces the sample throughput[5–7].

Comprehensive two-dimensional gas chromatography mass spectrometry (GC×GC-MS) uses two capillary GC columns with different stationary phases for separation of metabolites[8–12]. The two columns are usually connected via a thermal modulator. The second column is typically much shorter than the first (i.e., 1−2 m as opposed to 30−60 m for the first column) and is generally operated at a higher temperature. In the case of metabolites that co-elute from the first column, the second column may allow for further separation due to the different stationary phases as well as the different column temperatures. Thus, GC×GC-MS provides superior chromatographic peak capacity, selectivity, and lower detection limit for analysis of metabolites.

Two-dimensional liquid chromatography (2DLC) is usually configured in either a heart-cutting mode or a comprehensive mode[13, 14]. The heart-cutting analysis involves the collection of a few peaks of interest from the elution of the first dimension column, and then subjecting only those peaks onto the second dimension column for further separation. This configuration can increase signal-to-noise ratio and improve sensitivity for targeted metabolites[15]. In the comprehensive configuration, eluate from the first dimension column is collected into multiple fractions and each fraction is subjected to the second dimension column for further separation. The comprehensive 2DLC offers an increased resolution[16]. Most of the 2DLC systems are operated in online mode, where the second dimension separation is carried out simultaneously with the first dimension separation[17]. While the online mode of 2DLC has many advantages such as improved reliability and sample throughput, shortened analysis time and minimum sample loss, it requires the second dimension analysis to be completed during the time needed to collect and transfer a fraction from the first dimension column except that the 2DLC was configured in a stop-flow mode[18]. Another limitation of the online 2DLC technique is that the mobile phases used in the two columns must be compatible in both miscibility and solvent strength. In addition, some metabolites may partition between the fractions collected from the first dimension

column, resulting in large variation in metabolite quantification and even metabolite identification.

Klavins, *et al.* developed a parallel 2DLC system to perform orthogonal hydrophilic interaction chromatography (HILIC) and reverse phase chromatography (RPC) in one analytical run, where a sample was first delivered to two sample loops during sample loading[19]. The two sample aliquots were then simultaneously injected onto a dual column setup, and parallel separations were performed on the two columns. The eluates from the two columns were then merged and subjected to a mass spectrometer for further analysis. This strategy is simple yet effective for coupling HILIC and RPC for the purpose of decreasing analysis time and increasing throughput. Furthermore, the parallel 2DLC-MS configuration allows the use of two long columns and gradient time to increase separation power. However, parallel 2DLC-MS has the potential of peak overlapping which incurs resolution problem. On the other hand, unlike the comprehensive 2DLC-MD, parallel 2DLC-MS does not suffer from the peak partition and solvent miscibility and solvent strength issues.

While the effectiveness of both GC×GC-MS and 2DLC-MS for metabolomics has been separately demonstrated in multiple studies[9, 20–23], analysis of biological samples on the two platforms has not yet been explored. In the current study, we aimed to integrate the GC×GC-MS and the parallel 2DLC-MS for wider metabolite coverage, high confidence of metabolite identification and quantification, as well as high confidence of metabolic pathway assignment. The performance of developed system was tested by analyzing polar metabolites extracted from rat livers, where each metabolite extract was analyzed by GC×GC-MS and parallel 2DLC-MS, respectively. After metabolite identification and quantification, the performance of GC×GC-MS and parallel 2DLC-MS was assessed based on the number of identified metabolites, the accuracy of metabolite quantification, and the extent of their pathway coverage.

2. Experimental

2.1. Animal treatment

Thirteen male weanling Sprague–Dawley rats (35–45 g) from the Harlan Laboratories (Indianapolis, IN) were fed (*ad lib*) a purified AIN-76 diet with a defined copper content in form of cupric carbonate. The animals were housed in stainless steel cages in a temperature and humidity controlled room with a 12:12 h light–dark cycle. The 13 rats formed two groups, Group 1 (G1, n=6) and Group 2 (G2, n=7). The rats in G1 received adequate dose of copper (6.0 ppm) with free access to deionized water. The rats in G2 received supplemental dose of copper (20 ppm) with free access to deionized water containing 30% fructose (w/v). Fructose enriched drinking water was changed twice each week. All animals were fed for 4 weeks. At the end of the feeding period, each rat was killed under anesthesia with pentobarbital (50 mg/kg I.P. injection) after overnight fasting. Portions of rat liver were fixed with 10% formalin for subsequent sectioning, while others were snap-frozen with liquid nitrogen. All animal procedures were performed in accordance with the Guidelines for Care and Use of Laboratory Animals of the University of Louisville and approved by the American Association of Accreditation of Laboratory Animal Care.

2.2. Metabolite sample preparation

All samples were processed in random order to avoid systemic bias. After placing about 100 mg of liver tissue in a 1.5-mL Eppendorf tube, water was added at a ratio of 100 mg liver/mL water. After that, glass beads were added to the tube and the sample was homogenized using a Retsch MM 200 model mixer mill (Fisher Scientific, Hampton, NH, USA). To extract polar metabolites for GC×GC-MS analysis, 800 μL methanol was added to 200 μL homogenized liver. The mixture was vortex-mixed for 2 min and then placed on ice for 10 min. After another 2 min of vortex mixing, the sample was centrifuged at 15,000 rpm for 20 min at 4 °C. Seven hundred micro-liters of supernatant was transferred into a glass vial and dried in a SpeedVac evaporator to remove methanol, followed by lyophilization to remove water. The dried metabolite extract was then dissolved with 30 μL of 20 mg/mL methoxyamine hydrochloride pyridine solution followed by vigorous vortex mixing for 1 min. Methoxymation was carried out by sonicating the sample for 20 min and incubating it at 60 °C for 1 h. Derivatization was conducted by adding 30 μL of N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide with 1% tert-butyldimethylchlorosilane to the glass vial. After 1 h incubation at 60 °C, the mixture was transferred to a GC vial for analysis. A pooled sample was prepared by mixing 30 μL of derivatized metabolite extract from each sample.

To extract polar metabolites for parallel 2DLC-MS analysis, 400 μL methanol was added to 100 μL homogenized liver. The mixtures was vortex-mixed and centrifuged using the same extraction protocol used for GC×GC-MS analysis. Three hundred micro-liters of supernatant was transferred into a glass vial. Methanol in the sample was removed using a SpeedVac evaporator and water was removed by lyophilization. The dried metabolite extract was reconstituted with 100 μL 20% acetonitrile. The reconstitution was immediately preceded the parallel 2DLC-MS analysis. A pooled sample was prepared by mixing 50 μL of metabolite extract from every sample of the same group.

2.3. GC×GC-MS and its data analysis

A LECO (St. Joseph, MI, USA) Pegasus GC×GC-MS instrument was coupled with an Agilent 6890 gas chromatography and a Gerstel MPS2 auto-sampler (GERSTEL Inc., Linthicum, MD, USA), featuring a LECO two-stage cryogenic modulator and a secondary oven. The primary column was a 60 m × 0.25 mm $^{1}d_{c}$ × 0.25 μm $^{1}d_{f}$ DB-5 ms capillary column (phenyl arylene polymer virtually equivalent to (5%-phenyl)-methylpolysiloxane). The secondary GC column was a 1 m × 0.25 mm $^{2}d_{c}$ × 0.25 μm $^{2}d_{f}$ DB-17 ms column ((50% phenyl)-methylpolysiloxane) that was placed inside the secondary GC oven following the thermal modulator. Both columns were obtained from Agilent Technologies (Agilent Technologies J&W, Santa Clara, CA, USA). The helium carrier gas (99.999% purity) flow rate was set to 2.0 mL/min at a corrected constant flow with pressure ramps. The inlet temperature was set to 280 °C. The primary column temperature was programmed with an initial temperature of 60 °C for 0.5 min, then ramped at 5 °C/min to 270 °C, and maintained at 270 °C for 15 min. The secondary column temperature program was set to an initial temperature of 70 °C for 0.5 min and then ramped at the same temperature gradient used in the first column to 280 °C. The thermal modulator was +15 °C compared with the primary oven. The other instrument parameters were as: modulation period 2 s, mass range 29–800 *m/z*, spectrum acquisition rate 200 mass spectra per second, ion source chamber temperature

230 °C, transfer line temperature 280 °C, detector voltage 1420 V, electron energy 70 eV, and split ratio 20:1. The acceleration voltage was turned on after a solvent delay of 640 s.

The pooled sample was analyzed by GC×GC-MS eight times. The experiment data of the pooled sample were used to monitor the instrument variation. In addition, an aliquot of C_7−C_{30} *n*-alkane series was analyzed for retention index calculation.

To analyze the GC×GC-MS data, LECO's instrument control software, ChromaTOF (version 4.21), was used for peak picking and tentative metabolite identification. The threshold of spectral similarity score was set as ≥ 500 with a maximum value of 1000. MetPP software was used for retention index matching, peak merging, peak list alignment, normalization and statistical significance test[24, 25]. The *p*-value threshold was set as $p \leq 0.001$ for retention index matching.

To verify the tentative identification of metabolites detected with significant abundance difference between groups, commercially available authentic standards of those compounds were analyzed by GC×GC-MS under the identical experimental conditions as those used for analyses of biological samples. A tentative metabolite assignment was considered as a correct identification only if the experimental information of the authentic metabolite agreed with the corresponding information of a chromatographic peak in the biological samples, i.e., difference of the first dimension retention time ≤ 10 s, difference of the second dimension retention time ≤ 0.06 s, and the mass spectral similarity ≥ 500.

2.4. Parallel 2DLC-MS and its data analysis

All samples were randomly analyzed on a Thermo Q Exactive HF Hybrid Quadrupole-Orbitrap Mass Spectrometer coupled with a Thermo UltiMate 3000 HPLC system (Thermo Fisher Scientific, Inc., Germany). The UltiMate 3000 HPLC system was equipped with a hydrophilic interaction chromatography (HILIC) column and a reverse phase chromatography (RPC) column that were configured in parallel mode[19]. The HILIC column (SeQuant® ZIC®-cHILIC, 150×2.1 mm i.d., 3 μm) was purchased from EMD Millipore (Darmstadt, Germany). The RPC column (ACQUITY UPLC HSS T3, 150×2.1 mm i.d., 1.8 μm) was purchased from Waters Corp. (Milford, MA, USA). The temperature of those two columns was each set to 40 °C. The HILIC column was operated as follows: mobile phase A was 10 mM ammonium acetate (pH adjusted to 3.25 with acetate) in water and mobile phase B was 100% acetonitrile with 0.1% formic acid. The gradient was: 0.0 min, 95% B; 0.0 to 5.0 min, 95% B to 35% B; 5.0 to 6.0 min, 35% B; 6.0 to 6.1 min, 35% B to 5% B; 6.1 to 23.0 min, 5% B; 23.0 to 23.1 min, 5% B to 95% B; 23.1 to 40.0 min, 95% B. The flow rate was set to 0.3 mL/min. For the RPC column, the mobile phase A was water with 0.1% formic acid and mobile phase B was 100% acetonitrile with 0.1% formic acid. The gradient was as follows: 0.0 min, 5% B; 0.0 to 5.0 min, 5% B; 5.0 to 6.1 min, 5% B to 15% B; 6.1 to 10.0 min, 15% B to 60% B; 10.0 to 12.0 min, 60% B; 12.0 to 14.0 min, 60% B to 100% B; 14.0 to 27.0 min, 100% B; 27.0 to 27.1 min, 100% B to 5% B; 27.1 to 40.0 min, 5% B. The flow rate was 0.4 mL/min.

The metabolite extract of each biological sample or the pooled sample was analyzed by 2DLC-MS in positive mode (+) and negative mode (−), respectively. The electrospray

ionization probe was fixed at level C. The parameters for the probe were set as follows: sheath gas 55 arbitrary unit, auxiliary gas 15 arbitrary unit, sweep gas 3 arbitrary unit, spray voltage 3.5 kV, capillary temperature 320 °C, S-lens RF level 65.0, and auxiliary gas heater temperature 450 °C. The method of mass spectrometer was set as follows: full scan range 50–750 *m/z*, resolution 30,000, maximum injection time 50 ms, and automatic gain control (AGC) target 1×10^6 ions for both positive and negative modes.

The pooled sample was also analyzed by 2DLC-MS/MS (−) and 2DLC-MS/MS (+), respectively, to acquire the MS/MS spectra of metabolites. The 2DLC-MS/MS method and electrospray ionization condition were the same as those used in parallel 2DLC-MS analyses. The parameters used for mass spectrometry were as follows: for full-MS scan, scan range 50–750 m/z, resolution 30,000, maximum injection time 50 ms, and AGC target 1×10^6 ions; for dd-MS^2 scan, resolution 15,000, maximum injection time 100 ms, AGC target 5×10^4 ions, loop count 6, isolation window 1.3 m/z, and dynamic exclusion time 1.2 s. Each pooled sample was analyzed using 3 collision energies, i.e., 20, 40 and 60 eV, respectively.

To analyze the experimental data, all 2DLC-MS (−) and 2DLC-MS (+) data were processed using MetSign software for spectrum deconvolution, metabolite assignment, cross-sample peak list alignment, normalization, pattern recognition, and statistical significance test[26–29]. Metabolite identification was achieved in MetSign using the 2DLC-MS (−) and 2DLC-MS (+) data of biological samples and the 2DLC-MS/MS (−) and 2DLC-MS/MS (+) data of the pooled sample. MetSign respectively aligned the 2DLC-MS/MS (−) and 2DLC-MS/MS (+) data to the 2DLC-MS (−) and 2DLC-MS (+) data by retention time and parent ion *m/z* values with following thresholds: retention time variation ≤ 0.2 min and parent ion *m/z* variation ≤ 4 ppm. To identify the metabolites in the pooled sample, the parent ion *m/z*, retention time, and MS/MS spectra of a metabolite were matched to the corresponding information of 205 metabolite standards recorded in an in-house database, where the matching thresholds were set as follows: MS/MS spectral similarity ≥ 0.4, retention time difference ≤ 0.15 min, and *m/z* variation ≤ 4 ppm. The 2DLC-MS/MS (−) and 2DLC-MS/MS (+) data without a match in the in-house database were further analyzed using Compound Discoverer 2.0 software (Thermo Fisher Scientific, Inc., Germany) to match the remaining MS/MS spectra to the MS/MS spectra recorded in the Compound Discoverer database with a threshold of MS/MS spectral similarity score ≥ 20. For the peaks detected in the 2DLC-MS (−) or 2DLC-MS (+) data of biological samples that were not matched to any metabolites by MS/MS spectral matching, the *m/z* values of those parent ions were then matched to the compounds recorded in the Human Metabolome Database (HMDB) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. The *m/z* variation window was set ≤ 4 ppm.

3. Results and Discussion

The 2DLC-MS can be operated in either polarity switching mode or in (+) and (−) runs separately. In the polarity switching mode, the number of MS/MS scans across chromatographic peaks is reduced owing to polarity switching, resulting in some low abundance metabolites without MS/MS spectra. In order to maximize the chance of

detecting more metabolites, we chose to analyze each sample in the two different ionization modes in two separate runs even though this approach reduced the sample throughput.

3.1. Metabolite coverage by GC×GC-MS

Metabolite identification in analysis of GC×GC-MS data was done using the GC×GC-MS data of the pooled sample and biological samples by spectral similarity matching followed by retention index matching[25, 30]. Table 1 lists the numbers of detected chromatographic peaks and their identification results from the pooled sample and the rat liver samples. Out of 13 biological samples, one sample from G1 was detected as an outlier during the peaks alignment step due to very smaller number of peaks detected compared to the other samples. We believe this was caused by the injection in GC×GC-MS. Therefore, this sample was excluded in the subsequent analysis.

While about 3,138 ± 153 chromatographic peaks were detected in the eight injections of the pooled sample, 830 ± 91 metabolites were identified. Details of data processing parameters are listed in Supplementary Table S1. Using the same set of parameters, 3,314 ± 301 chromatographic peaks were detected from the biological samples, from which 874 ± 65 metabolites were identified. A metabolite might be identified from multiple distinct chromatographic peaks owing to incomplete derivatization, presence of isomers, or false identifications. After removing the redundant metabolite identifications, 434 ± 35 unique metabolites were identified from the pooled sample and 461 ± 32 unique metabolites were identified from the biological samples.

In GC×GC-MS, metabolites co-eluted from the first dimension GC column might be separated by the second dimension GC column (Supplementary Figure S1), which can increase the chance of acquiring high quality EI mass spectrum for metabolite identification. Despite the excellent instrumental capability, the identification accuracy of current mass spectrum matching method is only about 79.6%[31–33]. Therefore, additional information such as retention index must be used to reduce the rate of false identifications[34, 35]. On average, retention index matching removed about 8.9% of false identifications generated by mass spectrum matching (Table 1). Overall, only about 13.7% chromatographic peaks detected by GC×GC-MS were assigned to metabolites in this study.

3.2. Metabolite coverage by parallel 2DLC-MS

To identify metabolites from the 2DLC-MS data, 2DLC-MS/MS (−) and 2DLC-MS/MS (+) data of the pooled sample were first respectively aligned with the 2DLC-MS (−) and 2DLC-MS (+) data of the biological samples based on the retention time and parent ion *m/z* of each metabolite. The aligned data were then used for metabolite identification. Table 2 summarizes the identification results of the 2DLC-MS data. We made fresh solvent of mobile phase for analysis of three samples by 2DLC-MS, one sample from G1 and two samples from G2. In order to avoid the potential problems of solvent change, the experiment data of these three samples were not included in Table 2. Details of metabolites identified in this study by MS/MS matching are listed in Supplementary Tables S2 and S3.

A total of 15,108 ± 399 and 15,509 ± 681 features were detected from the pooled sample by 2DLC-MS (−) and 2DLC-MS (+), respectively. Here, a feature in 2DLC-MS data was

defined by retention time and isotopic peak *m/z* value. By parent ion *m/z* matching, 2,467 ± 144 features in 2DLC-MS (−) and 4,422 ± 153 features in 2DLC-MS (+) of the pooled sample were assigned to at least one metabolite in KEGG or HMDB databases. By MS/MS spectrum matching, 264 ± 18 metabolites were identified from the 2DLC-MS/MS (−) data and 227 ± 7 metabolites were identified from the 2DLC-MS/MS (+) data.

Table 2 also shows that 14,939 ± 269 and 17,617 ± 410 features were detected from the biological samples by 2DLC-MS (−) and 2DLC-MS (+), respectively. Among those features, 3,551 ± 85 and 5,560 ± 180 metabolites were respectively assigned to the 2DLC-MS (−) and 2DLC-MS (+) data based on parent ion *m/z* matching. By MS/MS spectrum matching, 262 ± 15 and 200 ± 9 metabolites were identified from the 2DLC-MS/MS (−) and 2DLC-MS/MS (+), respectively.

By parent ion *m/z* matching, about 23.8% to 31.6% of metabolites giving rise to the features in 2DLC-MS (−) and 2DLC-MS (+) were assigned to at least one metabolite. However, the percentage of assignment was dramatically reduced to only 1.1% to 1.8% when the MS/MS spectra were used for metabolite identification, even though we maximized the chance of acquiring high quality MS/MS spectra for each metabolite by fragmenting each parent ion using three collision energies, i.e., 20, 40 and 60 eV, respectively. The extremely low percentage of metabolite identification agrees with Silva, *et al.*'s observation that 98% of the instrumental data were not used in metabolomics[36]. Multiple factors contributed to those results. For instance, a fraction of the isotopic peaks detected by MS are not monoisotopic peaks and therefore cannot be matched to the metabolites. During the experiment, the top six abundant ions were subjected to MS/MS data acquisition in the dd-MS^2 mode. A number of metabolites with low instrument response were not selected for MS/MS spectra acquisition even though they were detected in the full MS mode. Furthermore, a metabolite in the sample might not be present in our in-house database and the Compound Discoverer database, and therefore could not have an identification result by MS/MS spectrum matching.

3.3. Integrating GC×GC-MS and 2DLC-MS data for high metabolite coverage

Figure 1A depicts the identification results of the three analytical platforms GC×GC-MS, 2DLC-MS (−), and 2DLC-MS (+) in analyzing the polar metabolites extracted from the biological samples. A total of 3,965 peaks were assigned to metabolites from the GC×GC-MS, 2DLC-MS (−), and 2DLC-MS (+) data. A majority of metabolites were assigned by parent ion *m/z* in the 2DLC-MS (−) and 2DLC-MS (+) data, while all metabolites were identified from the GC×GC-MS data by mass spectrum matching. Therefore, the number of metabolites assigned using the 2DLC-MS (−) and 2DLC-MS (+) data are much larger than that using the GC×GC-MS data. Owing to the diverse chemical properties of the GC and LC columns, each analytical platform favors detection of different metabolites in the biological samples. As expected, the GC×GC-MS and 2DLC-MS have much different metabolite coverage. Only 37 metabolites were commonly assigned to the GC×GC-MS and 2DLC-MS data. Of those 37 metabolites, 32 metabolites were assigned to the GC×GC-MS and the 2DLC-MS (−) data, and 28 metabolites were assigned to the GC×GC-MS and 2DLC-MS (+) data. There were only 23 metabolites detected by all three analytical platforms.

Assigning a metabolite to a peak by parent ion *m/z* only in the analysis of 2DLC-MS data generates a very high rate of false assignments. Figure 1B depicts the overlap of metabolite identification among GC×GC-MS, 2DLC-MS/MS (−), and 2DLC-MS/MS (+) data by MS/MS spectrum matching and other constrains, i.e., retention index filtering in analysis of GC×GC-MS data, parent ion *m/z* matching in analysis of 2DLC-MS/MS (−) and 2DLC-MS/MS (+) data. A total of 326 metabolites were identified by the three platforms, of which 205, 120 and 69 metabolites were identified from the GC×GC-MS, 2DLC-MS/MS (−) and 2DLC-MS/MS (+) data, respectively. Only 22 metabolites were commonly identified in all three platforms, which is only 7.0 % of the total metabolites identified by the three platforms.

In this study, we employed a parallel 2DLC-MS platform that was configured with a HILIC column and a RPC column. The downside of such a configuration is that the metabolite coverage was not dramatically increased. Furthermore, one metabolite might be detected twice and therefore increased the chance of metabolite overlapping in mass spectrometry. However, the parallel 2DLC-MS configuration does not have the problem of metabolite partition between two or more fractions that occurs in the comprehensive 2DLC configuration. In addition, we can use long columns and long gradient times to improve the separation. Overall, each platform, i.e., GC×GC-MS, 2DLC-MS/MS (−), or 2DLC-MS/MS (+), has limited metabolite coverage. The number of metabolites commonly detected by all those platforms is very small. Therefore, it is necessary to analyze the biological samples on different platforms to increase metabolite coverage.

3.4. Integrating GC×GC-MS and parallel 2DLC-MS for accurate metabolite quantification

During the metabolite identification, metabolites assigned by parent ion *m/z* matching have a high ratio of false identifications compared with the identification by MS/MS spectrum matching. In order to ensure the high degree of confidence in biomarker discovery, only metabolites identified by MS/MS spectrum matching were used for the metabolite quantification. A pairwise two-tail *t*-test with equal variance was used to study the abundance change of each metabolite between G1 and G2, during which sample labels were permutated up to 1000 times. Supplementary Table S4 shows that 41, 13 and 11 metabolites were detected with significant changes in their abundance levels between G1 and G2 in the GC×GC-MS, 2DLC-MS (−) and 2DLC-MS (+) data, respectively. Figure 2 shows the overlap of those metabolites. It is clear that each platform detected different sets of metabolites that had significant changes in their abundance levels. Among those metabolites, two were detected by all three platforms and three were detected by two of the three platforms. Integrating the results of the three platforms generated a comprehensive set of metabolites, i.e., 58 metabolites.

Table 3 lists the details of the abundance information of those five metabolites that were detected by more than one platform. The regulation directions of those metabolites detected in different platforms agree to each other, i.e., the fold-change of a metabolite detected in different platforms are all either larger than 1.0 or less than 1.0. Furthermore, except taurine, the fold-changes of those metabolites are almost identical, indicating the robustness and accuracy of the three platforms. Figures 3A, 3B, and 3C depict the detection of L-ornithine

in different platforms. This metabolite had relatively large instrument response and good chromatographic peak shape in all three platforms, and its chromatographic peak did not overlap with other compounds. Therefore, the quantification results of this metabolites were almost identical among the three platforms, with fold-changes of 0.6, 0.7 and 0.9 in GC×GC-MS, 2DLC-MS (−) and 2DLC-MS (+), respectively. However, the fold-change of taurine detected by GC×GC-MS had a larger variation compared to those detected by 2DLC-MS (−) and 2DLC-MS (+). Figures 3E and 3F show that taurine had a good instrument response and good peak shape in 2DLC-MS (−) and 2DLC-MS (+), respectively. Therefore, the changes of its abundance levels between G1 and G2 detected by 2DLC-MS (−) and 2DLC-MS (+) were very similar with fold-change of 1.8 and 1.7, respectively. However, the instrument response of taurine in GC×GC-MS was low and its chromatographic peak in the second dimension GC was very poor (Figure 3D). Furthermore, the chromatographic peak of taurine overlapped with an abundant peak. For those reasons, the data analysis software ChromaTOF could not accurately quantify the abundance of taurine, and resulted in a large variation in its fold-change compared with the fold-change calculated from the 2DLC-MS (−) and 2DLC-MS (+) data.

Overall, the GC×GC-MS, 2DLC-MS (−), and 2DLC-MS (+) platforms implemented in this study are robust for metabolite quantification. Manual analysis of the data reveals that the variation in metabolite quantification was mainly induced by the poor instrument response and the limited accuracy of data analysis software.

3.5. Biomarker discovery and pathway analysis

For metabolic pathway analysis, we first used the metabolites detected by each platform that had significant changes in their abundance levels between groups as the input of MetaboAnalyst software[37], to recognize the pathways that were affected by the treatments of rats in G1 and G2. From the 41 significant metabolites found from the GC×GC-MS data, the MetaboAnalyst software produced five pathways with $p \leq 0.05$ (Table 4). Likewise, MetaboAnalyst showed that seven pathways were statistically significant using the 13 significant metabolites detected from the 2DLC-MS (−) data, and 5 pathways using the 11 significant metabolites detected from the 2DLC-MS (+) data. We then combined all 58 significant metabolites detected from the three platforms and used them as the input of MetaboAnalyst software. By doing so, eight pathways were considered as significantly impacted pathways (Table 4).

Figure 4 shows the overlap of the pathways using different sets of significant metabolites for metabolic pathway analysis. Compared with the results of pathway analysis using the data acquired by individual platform, the confidence of pathway assignment using the integrated data was significantly increased. For instance, 8 metabolites associated with aminoacyl-tRNA biosynthesis pathway were detected by integrating the data of all three platforms. However, 0, 3 and 6 metabolites were detected by 2DLC-MS (−), 2DLC-MS (+) and GC×GC-MS, respectively. Using the integrated data acquired from multiple platforms clearly increase the confidence of metabolic pathway analysis. The confidence of assigning others pathways was also increased, including arginine and proline metabolism; valine,

leucine and isoleucine biosynthesis; alanine, aspartate and glutamate metabolism; D-glutamine and D-glutamate metabolism; and butanoate metabolism.

We also performed another analysis by matching all detected metabolites to the metabolic pathway regardless whether those metabolites had significant changes in their abundance levels between groups. For example, biosynthesis of unsaturated fatty acid pathway was a significantly impacted pathway as suggested by the 58 significant metabolites detected from the three platforms. Supplementary Table S5 lists the metabolites in this pathway that were detected by different platforms. Docosahexaenoic acid (DHA), linoleic acid (LA), and gamma-linolenic acid (GLA), were only detected by GC×GC-MS. Prostaglandin G2, leukotriene B4, LXA_4/LXB_4 were detected by 2DLC-MS (−), and stearidonic acid was only detected by 2DLC-MS (+). Detecting more metabolites in a pathway and knowing whether their abundance levels were affected by the biomedical treatments clearly can narrow down to the specific steps in the pathways that were affected by the treatments.

In summary, one analytical platform is not enough to give high metabolite coverage in a metabolic pathway, and the use of only one platform reduces the confidence of metabolic pathway assignment. Integrating the data acquired from multiple platforms not only provides high metabolite coverage, but also increases the confidence of pathway assignment and the confidence of biomarker discovery. The details of the biomarkers and pathways discovered in this study will be described in a separate report.

5. Conclusions

Polar metabolites extracted from rat livers were analyzed by GC×GC-MS and parallel 2DLC-MS, respectively. 903 ± 96 chromatographic peaks were detected by GC×GC-MS, and 2,467 ± 144 and 4,422 ± 153 features were respectively detected by 2DLC-MS (−) and 2DLC-MS (+). Integrating the experimental data acquired from the three platforms clearly increased the metabolite coverage. A total of 58 metabolites were detected with significant changes in their abundance levels between groups. Three of the 58 metabolites were detected in two platforms and two in all three platforms. The agreement of metabolite regulation detected by different platforms demonstrated the robustness and accuracy of the three platforms used in the current study. Manual examination showed that the discrepancy of metabolite regulation measured by different platforms was mainly caused by the poor shape of chromatographic peaks. Pathway analysis demonstrated that integrating the results from multiple platforms increased the confidence of metabolic pathway assignment. While the developed method showed excellent results in terms of metabolites coverage and increased confidence of metabolic pathway assignment, however, several factors still need to be worked on for better identification and quantification. Metabolite identification still remains as a major challenge in untargeted metabolomics. While a huge number of isotopic peaks (raw data) were generated from the instrument, most of them remained unidentified. Incompleteness of the database contributed to this problem. Future endeavor would be the combination of all available MS/MS databases to maximize the metabolite assignment and develop robust bioinformatics tools for accurate metabolite identification. Furthermore, a comparison of the parallel 2DLC-MS results with the comprehensive 2DLC-MS results

would enlighten us the overall capability of the multidimensional analytical technique for the untargeted metabolomics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

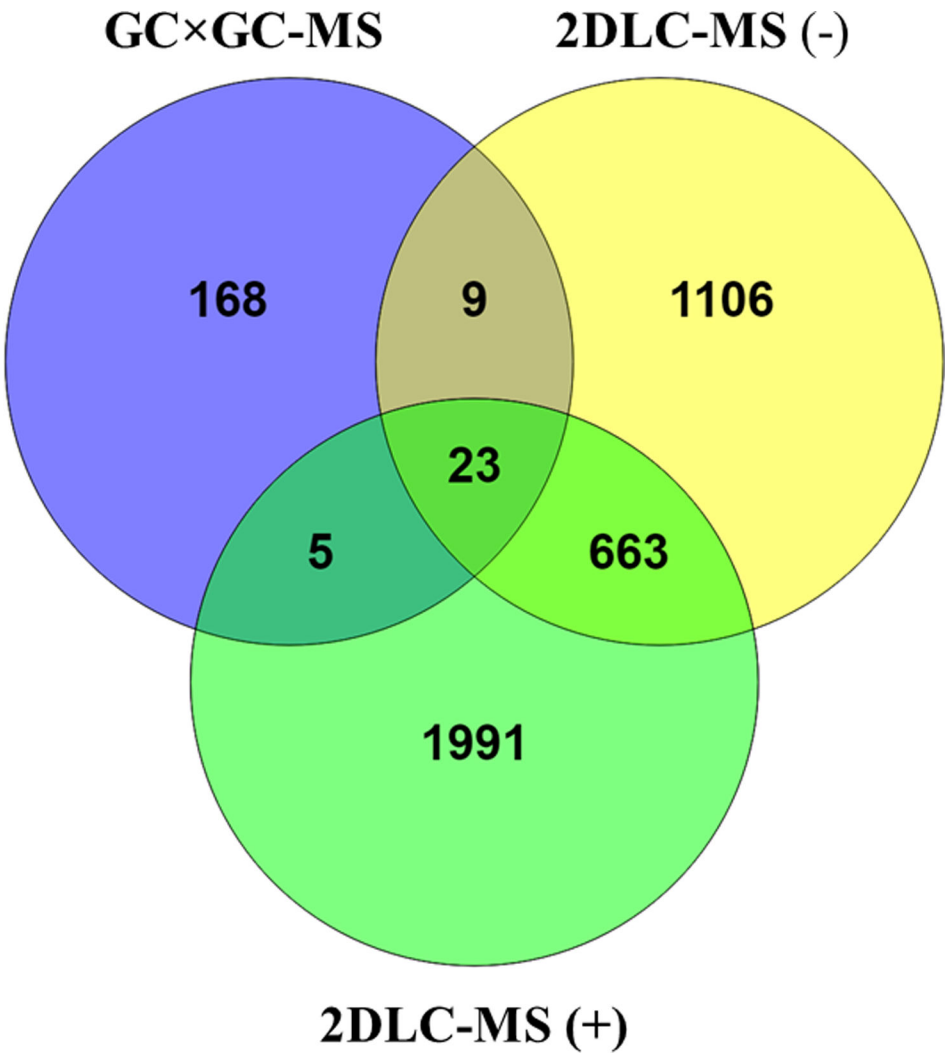
ACKNOWLEDGEMENTS

The authors thank Mrs. Marion McClain for review of this manuscript. This work was supported by NIH grant nos. 1S10OD020106–01 (XZ), 1P20GM113226 (CJM), 1P50AA024337 (CJM), 1U01AA021893 (CJM), 1U01AA021901 (CJM), 1U01AA022489–01A1 (CJM) and 1R01AA023681 (CJM), the Department of Veterans Affairs 1I01BX002996–01A2 (CJM).

REFERENCE

1. Sugimoto M, Wong DT, Hirayama A, Soga T and Tomita M, Metabolomics: Official journal of the Metabolomic Society, 2010, 6, 78–95. [PubMed: 20300169]
2. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B, Pujos-Guillot E, Verheij E, Wishart D and Wopereis S, Metabolomics: Official journal of the Metabolomic Society, 2009, 5, 435–458. [PubMed: 20046865]
3. Kuehnbaum NL and Britz-McKibbin P, Chemical reviews, 2013, 113, 2437–2468. [PubMed: 23506082]
4. Stoll DR and Carr PW, Anal. Chem, 2017, 89, 519–531. [PubMed: 27935671]
5. Almstetter MF, Oefner PJ and Dettmer K, Analytical and bioanalytical chemistry, 2012, 402, 1993–2013. [PubMed: 22249417]
6. Higgins Keppler EA, Jenkins CL, Davis TJ and Bean HD, TrAC Trends in Analytical Chemistry, 2018, 109, 275–286.
7. Kim S, Yin X, Prodhan MAI, Zhang X, Zhong Z and Kato I, Journal of Chromatographic Science, 2019, 57, 385–396. [PubMed: 30796770]
8. Shi X, Wei X, Koo I, Schmidt RH, Yin X, Kim SH, Vaughn A, McClain CJ, Arteel GE and Zhang X, J Proteome Res, 2013, 13, 547–554. [PubMed: 24328084]
9. Wei X, Song M, Yin X, Schuschke DA, Koo I, McClain CJ and Zhang X, J Proteome Res, 2015, 14, 4050–4058. [PubMed: 26216400]
10. Shi X, Wei X, Koo I, Schmidt RH, Yin X, Kim SH, Vaughn A, McClain CJ, Arteel GE and Zhang X, J Proteome Res, 2013, 13, 547–554. [PubMed: 24328084]
11. Shi X, Wei X, Yin X, Wang Y, Zhang M, Zhao C, Zhao H, McClain CJ, Feng W and Zhang X, J Proteome Res, 2015, 14, 1174–1182. [PubMed: 25592873]
12. Winnike JH, Wei X, Knagge KJ, Colman SD, Gregory SG and Zhang X, J Proteome Res, 2015, 14, 1810–1817. [PubMed: 25735966]
13. Lu W, Bennett BD and Rabinowitz JD, Journal of Chromatography B, 2008, 871, 236–242.
14. Cacciola F, Jandera P and Mondello L, Chromatographia, 2007, 66, 661–667.
15. Breidbach A and Ulberth F, Analytical and bioanalytical chemistry, 2015, 407, 3159–3167. [PubMed: 25015044]
16. Stoll DR, Li X, Wang X, Carr PW, Porter SE and Rutan SC, J Chromatogr A, 2007, 1168, 3–43. [PubMed: 17888443]
17. Erni F and Frei R, J Chromatogr A, 1978, 149, 561–569.
18. Wang SY, Li J, Shi XZ, Qiao LZ, Lu X and Xu GW, J Chromatogr A, 2013, 1321, 65–72. [PubMed: 24238711]
19. Klavins K, Drexler H, Hann S and Koellensperger G, Anal. Chem, 2014, 86, 4145–4150. [PubMed: 24678888]
20. Shi X, Wei X, Koo I, Schmidt RH, Yin X, Kim SH, Vaughn A, McClain CJ, Arteel GE, Zhang X and Watson WH, J Proteome Res, 2014, 13, 547–554. [PubMed: 24328084]

21. Warner DR, Liu H, Ghosh Dastidar S, Warner JB, Prodhan MAI, Yin X, Zhang X, Feldstein AE, Gao B, Prough RA, McClain CJ and Kirpich IA, Plos One, 2018, 13, e0204119. [PubMed: 30256818]
22. Fairchild JN, Horvath K, Gooding JR, Campagna SR and Guiochon G, J Chromatogr A, 2010, 1217, 8161–8166. [PubMed: 21094946]
23. Chung D-H, Golden JE, Adcock RS, Schroeder CE, Chu Y-K, Sotsky JB, Cramer DE, Chilton PM, Song C, Anantpadma M, Davey RA, Prodhan AI, Yin X and Zhang X, Antimicrobial Agents and Chemotherapy, 2016, 60, 4552–4562. [PubMed: 27185801]
24. Wei X, Shi X, Koo I, Kim S, Schmidt RH, Arteel GE, Watson WH, McClain C and Zhang X, Bioinformatics, 2013, 29, 1786–1792. [PubMed: 23665844]
25. Koo I, Shi X, Kim S and Zhang X, J Chromatogr A, 2014, 1337, 202–210. [PubMed: 24630063]
26. Wei X, Shi X, Kim S, Zhang L, Patrick JS, Binkley J, McClain C and Zhang X, Anal. Chem, 2012, 84, 7963–7971. [PubMed: 22931487]
27. Wei X, Lorkiewicz P, Salabei JK, Shi B, Hill BG, Kim S, McClain CJ and Zhang X, Anal. Methods, 2017, 9, 2275–2283. [PubMed: 28674558]
28. Wei X, Shi X, Kim S, S K, Patrick JS, Binkley J, Kong M, McClain C and Zhang X, Anal. Chem, 2014, 86, 2156–2165. [PubMed: 24533635]
29. Wei X, Sun W, Shi X, Koo I, Wang B, Zhang J, Yin X, Tang Y, Bogdanov B, Kim S, Zhou Z, McClain C and Zhang X, Anal. Chem, 2011, 83, 7668–7675. [PubMed: 21932828]
30. Zhang J, Fang A, Wang B, Kim SH, Bogdanov B, Zhou Z, McClain C and Zhang X, J Chromatogr A, 2011, 1218, 6522–6530. [PubMed: 21813131]
31. Kim S, Koo I, Fang AQ and Zhang X, BMC Bioinformatics, 2011, 12.
32. Koo I, Kim S and Zhang X, J Chromatogr A, 2013, 1298, 132–138. [PubMed: 23726352]
33. Wei XL, Koo I, Kim S and Zhang X, Analyst, 2014, 139, 2507–2514. [PubMed: 24665464]
34. Prodhan MAI, Yin X, Kim S, McClain C and Zhang X, J Chromatogr A, 2018, 1539, 62–70. [PubMed: 29395161]
35. Prodhan MAI, Sleman AA, Kim S, McClain C and Zhang X, Journal of Analysis and Testing, 2018, 2, 263–273.
36. da Silva RR, Dorrestein PC and Quinn RA, Proceedings of the National Academy of Sciences, 2015, 112, 12549–12550.
37. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS and Xia J, Nucleic Acids Research, 2018, 46, W486–W494. [PubMed: 29762782]



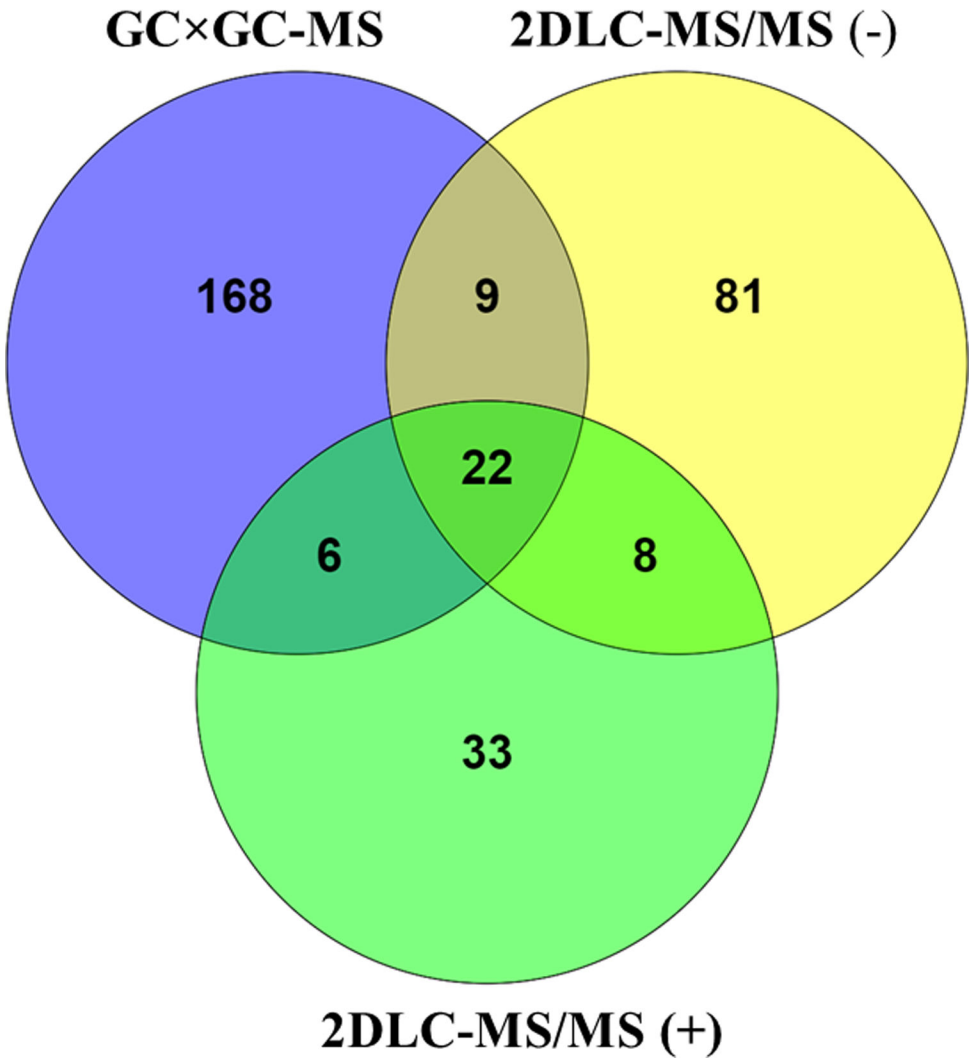


Figure 1.

Overlap of metabolite identification. Metabolites were identified from GC×GC-MS by EI mass spectrum matching and retention index matching. (A) The metabolite identification in analysis of 2DLC-MS data was done by parent ion m/z matching. (B) The metabolite identification in analysis of 2DLC-MS/MS data by MS/MS spectrum matching with or without retention time match.

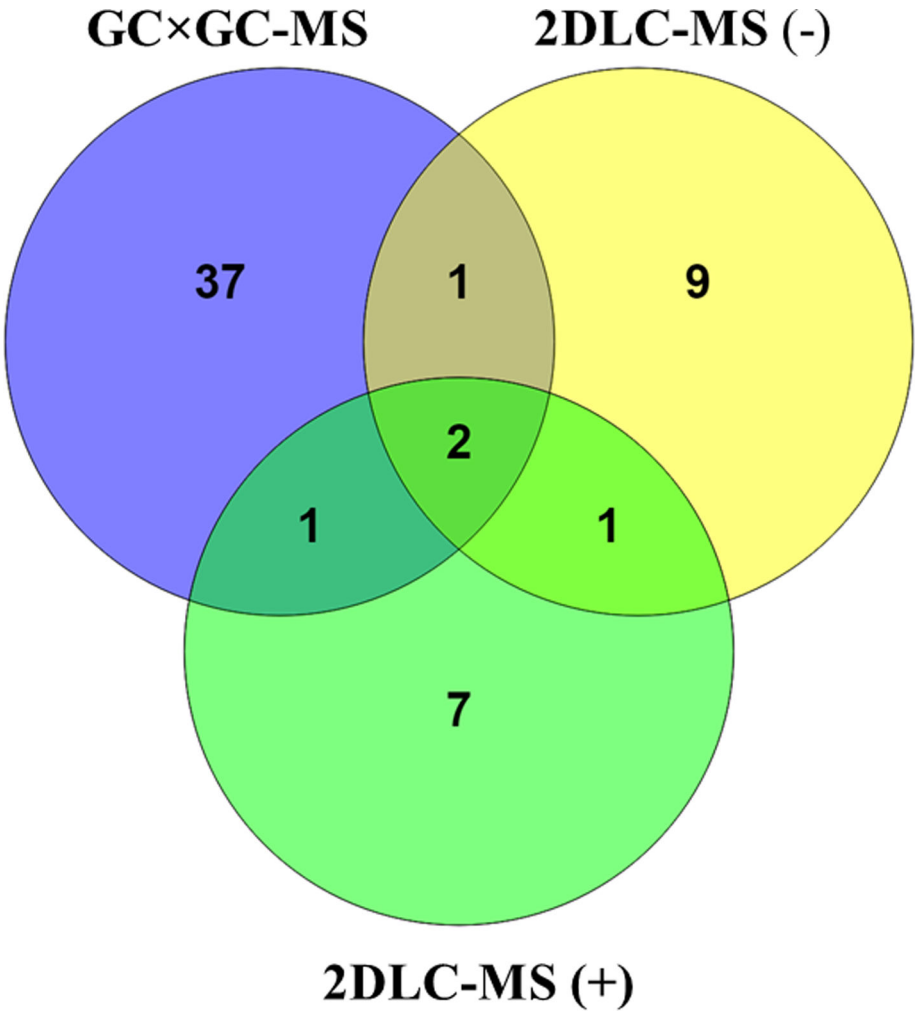


Figure 2.
Overlap of metabolites detected with significant changes in their abundance levels between groups by three platforms.

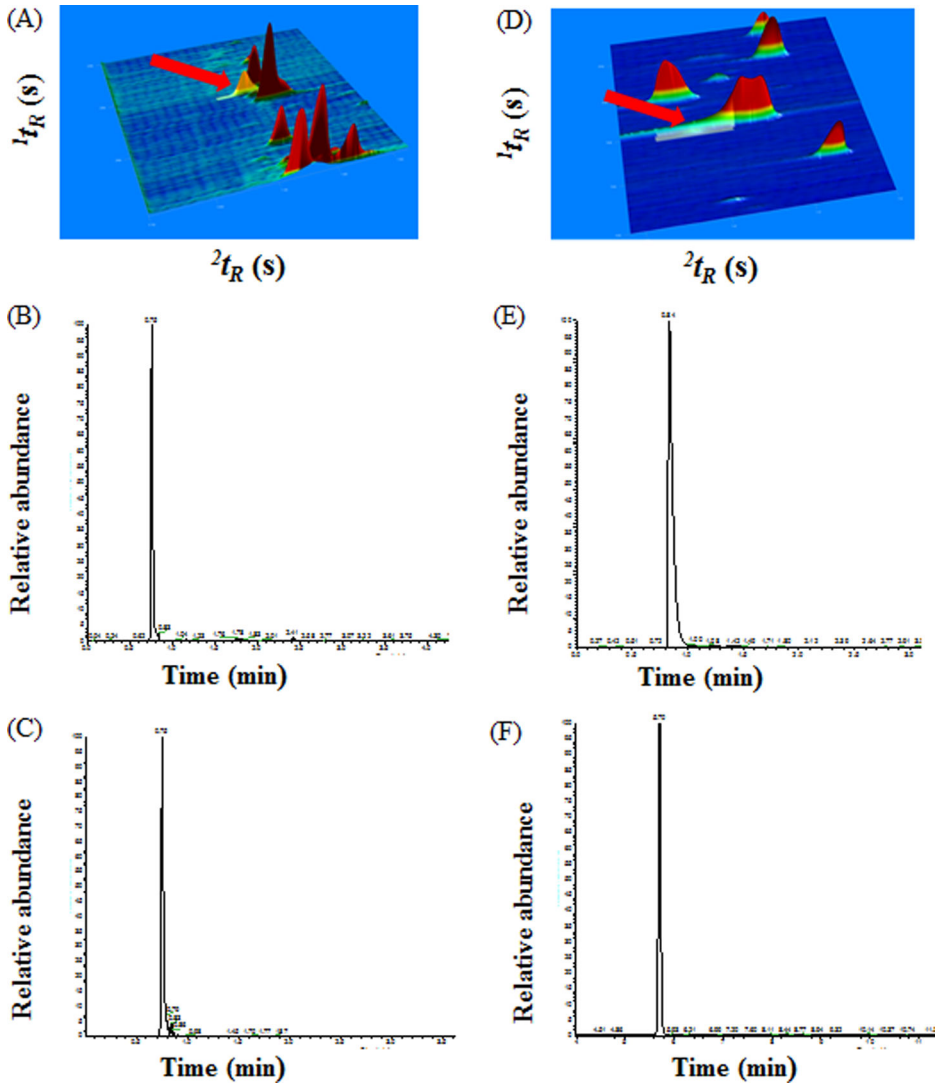


Figure 3.

Samples of instrument response of a metabolite affecting the quantification accuracy of that metabolite. (A) three dimensional chromatographic peak of L-ornithine detected by GC×GC-MS. (B) extracted ion chromatograms of L-ornithine in a randomly selected biological samples detected by 2DLC-MS (−). (C) Extracted ion chromatograms of L-ornithine in a randomly selected biological samples detected by 2DLC-MS (+). (D) three dimensional chromatographic peak of taurine detected by GC×GC-MS. (E) extracted ion chromatograms of taurine in a randomly selected biological samples detected by 2DLC-MS (−). (F) Extracted ion chromatograms of taurine in a randomly selected biological samples detected by 2DLC-MS (+).

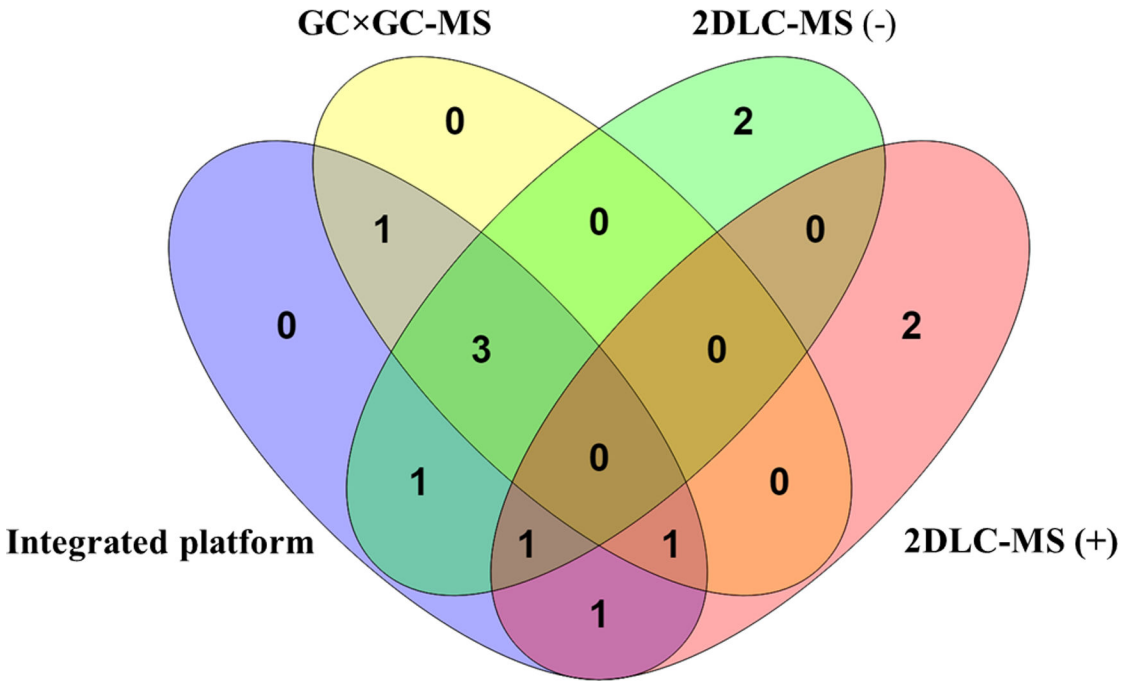


Figure 4.
Overlap of the pathways that were affected using different set of significant metabolites for metabolic pathway analysis.

Table 1.

Numbers of detected peaks from pooled sample and biological samples by GC×GC-MS and the results of metabolite identification using different matching methods

Sample	Analysis ID	Chromatographic peaks	Metabolite identification		
			Similarity score threshold ≥ 500	After RI matching	Unique metabolites
Pooled sample	Inj-1	3280	992	911	476
	Inj-2	3004	836	778	423
	Inj-3	3443	842	778	404
	Inj-4	3082	936	862	429
	Inj-5	3029	927	864	450
	Inj-6	3106	941	637	367
	Inj-7	3210	1037	947	447
	Inj-8	2952	712	865	477
	Average	3138±153	903±96	830±91	434±35
Biological sample	S-1	3389	1017	902	484
	S-2	3371	953	842	462
	S-3	3304	984	845	469
	S-4	3531	1003	918	507
	S-5	3873	1080	925	506
	S-6	3263	974	881	443
	S-7	3062	990	898	458
	S-8	3288	1075	954	480
	S-9	2559	750	693	399
	S-10	3508	961	859	443
	S-11	3154	964	846	413
	S-12	3463	1048	929	470
	Average	3314±301	983±82	874±65	461±32

Table 2.

Numbers of detected peaks from pooled sample and biological samples by 2DLC-MS and the results of metabolite identification using different matching methods

Sample	Analysis ID	2DLC-MS (−)			2DLC-MS (+)		
		Isotopic peaks	Public DB	MS/MS DB	Isotopic peaks	Public DB	MS/MS DB
Pooled sample	Inj-1	15586	2706	278	14617	4241	217
	Inj-2	14903	2526	276	14834	4215	224
	Inj-3	15381	2494	275	14823	4265	217
	Inj-4	15196	2564	267	15109	4405	229
	Inj-5	14951	2440	274	16033	4590	231
	Inj-6	14302	2201	220	16238	4621	223
	Inj-7	14959	2313	255	16341	4546	236
	Inj-8	15585	2494	270	16079	4489	236
	Average	15108±399	2467±144	264±18	15509±681	4422±153	227±7
Biological sample	S-1	15112	3590	264	18128	5791	214
	S-2	14753	3363	266	17164	5561	197
	S-3	15184	3625	274	17875	5572	201
	S-4	14699	3560	262	18193	5503	205
	S-5	15075	3673	276	17922	5371	186
	S-6	14776	3514	278	17082	5457	189
	S-7	14844	3463	242	17806	5806	196
	S-8	15514	3553	279	17450	5519	217
	S-9	14888	3623	237	17010	5235	196
	S-10	14546	3550	243	17541	5788	202
	Average	14939±269	3551±85	262±15	17617±410	5560±180	200±9

Table 3.

Quantification information of metabolites that were detected by more than one platform with significant changes in their abundance levels between groups

Compound	*p*-value			Fold-change		
	GC×GC-MS	2DLC-MS(−)	2DLC-MS(+)	GC×GC-MS	2DLC-MS(−)	2DLC-MS(+)
Taurine	4.6E-04	1.7E-03	3.2E-02	5.6	1.8	1.7
Ornithine	4.1E-02	4.1E-02	3.2E-02	0.6	0.7	0.9
Phenylalanine	2.8E-02	5.7E-01	2.4E-02	0.6	0.6	0.7
Malic acid	5.0E-02	4.9E-02	-	1.6	1.3	-
Hypotaurine	-	1.4E-03	1.6E-03	-	2.2	2.3

Table 4.

Pathways affected by the treatment difference between groups

Significantly impacted pathways	*p*-value	Match status
Integrated platforms		
Aminoacyl-tRNA biosynthesis	1.1E-03	8/69
Arginine and proline metabolism	2.1E-03	6/44
Valine, leucine and isoleucine biosynthesis	4.1E-03	3/11
Butanoate metabolism	4.3E-03	4/22
Alanine, aspartate and glutamate metabolism	5.9E-03	4/24
Biosynthesis of unsaturated fatty acids	9.1E-03	5/42
D-Glutamine and D-glutamate metabolism	9.3E-03	2/5
Taurine and hypotaurine metabolism	2.4E-02	2/8
GC×GC-MS		
Biosynthesis of unsaturated fatty acids	2.3E-03	5/42
Aminoacyl-tRNA biosynthesis	4.2E-03	6/69
Butanoate metabolism	1.3E-02	3/22
Alanine, aspartate and glutamate metabolism	1.7E-02	3/24
Valine, leucine and isoleucine biosynthesis	2.5E-02	2/11
2DLC-MS (−)		
Taurine and hypotaurine metabolism	1.2E-03	2/8
Valine, leucine and isoleucine biosynthesis	2.4E-03	2/11
Citrate cycle (TCA cycle)	8.0E-03	2/20
Butanoate metabolism	9.6E-03	2/22
Alanine, aspartate and glutamate metabolism	1.1E-02	2/24
D-Glutamine and D-glutamate metabolism	3.5E-02	1/5
Primary bile acid biosynthesis	3.9E-02	2/46
2DLC-MS (+)		
Arginine and proline metabolism	2.3E-04	4/44
Taurine and hypotaurine metabolism	1.5E-03	2/8
Aminoacyl-tRNA biosynthesis	1.4E-02	3/69
Phenylalanine, tyrosine and tryptophan biosynthesis	3.1E-02	1/4
Cyanoamino acid metabolism	4.6E-02	1/6