metabolites

MDPI

Article

A Strategy for Uncovering the Serum Metabolome by Direct-Infusion High-Resolution Mass Spectrometry

Xiaoshan Sun [1,2,3], Zhen Jia [1,3,4], Yuqing Zhang [1,3,5], Xinjie Zhao [1,2,3], Chunxia Zhao [1,2,3], Xin Lu [1,2,3,*] and Guowang Xu [1,2,3,*]

1 CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China
2 University of Chinese Academy of Sciences, Beijing 100049, China
3 Liaoning Province Key Laboratory of Metabolomics, Dalian 116023, China
4 Department of Cell Biology, College of Life Sciences, China Medical University, Shenyang 110122, China
5 Zhang Dayu School of Chemistry, Dalian University of Technology, Dalian 116024, China
* Correspondence: luxin001@dicp.ac.cn (X.L.); xugw@dicp.ac.cn (G.X.)

Abstract: Direct infusion nanoelectrospray high-resolution mass spectrometry (DI-nESI-HRMS) is a promising tool for high-throughput metabolomics analysis. However, metabolite assignment is limited by the inadequate mass accuracy and chemical space of the metabolome database. Here, a serum metabolome characterization method was proposed to make full use of the potential of DI-nESI-HRMS. Different from the widely used database search approach, unambiguous formula assignments were achieved by a reaction network combined with mass accuracy and isotopic patterns filter. To provide enough initial known nodes, an initial network was directly constructed by known metabolite formulas. Then experimental formula candidates were screened by the predefined reaction with the network. The effects of sources and scales of networks on assignment performance were investigated. Further, a scoring rule for filtering unambiguous formula candidates was proposed. The developed approach was validated by a pooled serum sample spiked with reference standards. The coverage and accuracy rates for the spiked standards were 98.9% and 93.6%, respectively. A total of 1958 monoisotopic features were assigned with unique formula candidates for the pooled serum, which is twice more than the database search. Finally, a case study of serum metabolomics in diabetes was carried out using the developed method.

Keywords: metabolomics; direct-infusion; high-resolution mass spectrometry; formula assignment; reaction network

check for updates

Citation: Sun, X.; Jia, Z.; Zhang, Y.; Zhao, X.; Zhao, C.; Lu, X.; Xu, G. A Strategy for Uncovering the Serum Metabolome by Direct-Infusion High-Resolution Mass Spectrometry. *Metabolites* **2023**, *13*, 460. https://doi.org/10.3390/metabo13030460

Academic Editor: Petr G. Lokhov

Received: 4 March 2023
Revised: 18 March 2023
Accepted: 20 March 2023
Published: 22 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Direct infusion combined with high-resolution mass spectrometry (DI-HRMS) is an attractive alternative for high-throughput metabolomics analysis due to its simplicity, high resolution, high mass accuracy, and time-saving [1,2]. In metabolomics analysis, nanoelectrospray ionization (nESI) DI-HRMS provides similar discrimination capabilities to conventional liquid chromatography-mass spectrometry (LC-MS) while greatly reducing the total analysis time [3,4]. DI-HRMS approach has been widely applied in various metabolomics studies [5–7], such as urinary metabolic profiling in human epidemiological investigations [8], metabolite analysis for the diagnosis of inborn errors of metabolism [9–11], and serum metabolomics analysis in diabetes [12].

Metabolite annotation is widely recognized as a critical part of metabolomics analysis. Accurate molecular formula assignment is a prerequisite for identifying unknown metabolites [13]. Typically, metabolite annotation relies on metabolome database searching using accurate mass, where the extent of database coverage determines the annotation performance [14–16]. Sarvin et al. [17] recently analyzed the m/z value distribution of ions

and identified optimal scan ranges, resulting in a ~50% increase in feature detection compared to conventional spectral-stitching FI-MS [18]. However, this study employed publicly available databases, such as the LIPID MAPS Structure Database (LMSD) for lipidomics data [19] and the Human Metabolome Database (HMDB) for metabolomics data [20,21], and only 564 out of 3233 features in metabolomics data and 401 out of 3339 features in lipidomics data obtained putative annotations with a 5 ppm mass tolerance, respectively. Alternatively, the stochastic molecular generator method, which incorporates chemical constraints (e.g., high mass accuracy, limited element number/species, and high-fidelity isotopic pattern), allows reliable formula assignment without database limitations [22–24]. But this approach is not suitable for high-resolution mass spectrometry analysis, especially for low-abundance and high-mass MS signals, for which multiple formula candidates may match [25–27].

Metabolites serve as substrates and products of vital biological reactions [28,29], so it is very promising to exploit the reaction relationships among them for metabolome annotation. The approach of utilizing mass differences has been applied for molecular formula assignment in ultra-high-resolution mass spectrometry [30–32]. In this method, spectral features are represented as nodes, while pre-defined metabolic reactions, as reflected by mass differences, serve as edges connecting these nodes [33,34]. High mass accuracy (<1 ppm) is crucial for minimizing the number of candidate molecular formulas and possible reaction types. Nevertheless, it is difficult to achieve in high-resolution mass spectrometry methods [35]. To address this challenge, Chen et al. [36] proposed a global network optimization approach, NetID, which annotates formulas for as many MS features as possible, including adducts, fragments and isotopes of metabolites in untargeted LC-HRMS data. In NetID seed nodes were limited to the detected features and assigned by matching peaks to the HMDB with a mass tolerance of 10 ppm, it is especially important to use highly reliable metabolite formulas as seed nodes [37].

In this work, a serum metabolome characterization method was proposed to fully explore the potential of DI-nESI-HRMS. To enhance the coverage and accuracy of unique formula assignment, a reaction network was constructed and endogenous metabolites from the HMDB served as initial seed nodes. A scoring system was established to further screen reliable molecular formulas from the possible candidates based on the topological relationship in the reaction network, mass accuracy, and isotopic fine structure. Finally, the developed method was applied to a high-throughput study of serum metabolomics in diabetes.

2. Materials and Methods

2.1. Chemicals

LC-MS grade formic acid was purchased from J&K Scientific Ltd. (Beijing, China). HPLC grade acetonitrile, methanol, and chloroform were purchased from Merck (Darmstadt, Germany). Ultrapure water was prepared by a Milli-Q Ultrapure water system (Millipore, Billerica, MA, USA). Seventy-eight metabolites used for evaluating annotation accuracy were supplied by Sigma-Aldrich (St. Louis, MO, USA) (Table S1). Nine stable isotope labeled internal standards (ISs) were used to normalize MS features in DI-nESI-HRMS. (Table S2). Carnitine C2:0-d_3, carnitine C12:0-d_3, and carnitine C16:0-d_3 were purchased from International Laboratory (South San Francisco, CA, USA). Choline-d_4, phenylalanine-d_5 (Phe-d5), tryptophan-(indole-d_5) (Trp-d_5), leucine-d_3 (Leu-d_3), glutamine-d_5, cholic acid-d_4 (CA-d_4), chenodeoxycholic acid-d_4 (CDCA-d_4) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2. Sample Information and Preparation

The serum samples were collected from 38 healthy control individuals and 31 type 2 diabetes (T2D) patients at The Second Hospital of Dalian Medical University. All enrolled participants had written consents and the research protocol (No. 2019(124)) was approved

by the Ethics and Human Subjects Committee of the Second Hospital of Dalian Medical University. Detailed clinical data of the samples are summarized in Table S3.

A pooled sample was prepared by combining 10 µL aliquots from each serum sample as a quality control (QC) sample. The working solution of IS was mixed in methanol for subsequent protein precipitation of serum samples. Method linearity was assessed with the mixture of 10 metabolite standards at 6 different concentrate levels (Table S4).

For DI-nESI HRMS analysis, routine protein precipitation was first performed. Then, 200 µL of ice-cold methanol containing IS was added into a 1.5 mL Eppendorf tube containing 50 µL of a serum sample, followed by thorough vortexing for 1 min, and centrifuged for 10 min at 4 °C and 15,000× *g*. 220 µL of the supernatant was vacuum concentrated at 4 °C by CentriVap Centrifugal Vacuum Concentrators (Labconco, Kansas City, MO, USA). Then, the freeze-dried residue was diluted 20 times with methanol/water 2:1 (*v*/*v*) containing 0.1% formic acid. After vortexing for 1 min, the sample solution was centrifuged at 4 °C and 15,000× *g* for 10 min. Finally, the supernatant was transferred into 96-well plates for DI-nESI HRMS analysis.

2.3. DI-nESI HRMS Analysis

In DI-nESI HRMS analysis, TriVersa Nanomate chip electrospray system (Advion BioSciences, Ithaca, NY, USA) was coupled to a Q Exactive HF (Thermo Fisher Scientific, Rockford, IL, USA). For the NanoMate system, the voltage and gas pressure were set as 1.7 kV and 0.6 psi, respectively. For mass spectrometry acquisition, the capillary temperature was set at 270 °C and S-lens level was 50. The data of the full scan based on the spectral-stitching acquisition were acquired in positive ionization mode and the mass resolution at m/z 200 was 240,000, the scan windows were set as follows: m/z 65–235, m/z 225–315, m/z 305–355, m/z 345–395, m/z 385–435, m/z 425–475, m/z 465–515, m/z 505–562, m/z 552–609. Maximum IT was 200 ms and Microscan was set as 3. The total analysis time was 0.6 min per sample. The spectrum data type above all acquisition methods was set as centroid mode.

2.4. Data Processing

For DI-nESI-Orbitrap MS analysis, Xcalibur software (Thermo Fisher Scientific, Rockford, IL, USA) was used to visualize and process rawdata files. Theoretical formulas for MS features were generated according to the rules of elemental type (C, H, O, N, P, S), hydrogen/carbon ratios (0.4~5.1) and isotopic fine structure, and further filtration restricting elemental composition and reaction network analysis was performed using in-house Python scripts. The subsequent steps were processed including peak lists export, noise filtration (S/N > 10), peaks alignment (5 ppm), and blank reduction.

The MS feature intensities were normalized by ISs and only those features with a relative standard deviation (RSD) less than 30% in QC were retained for further statistical analysis. Finally, the peak table was used for statistical analyses, and the partial least squares-discriminant analysis (PLS-DA) was performed. Significantly differential metabolites between health and T2D diabetes were screened out by nonparametric tests ($p < 0.05$) and VIP > 1.

An initial reaction network was constructed using unique molecular formulas (seed nodes) extracted from the Human Metabolome Database (HMDB, https://hmdb.ca/metabolites) (accessed on 16 December 2022). Three filtering criteria were employed for the selection of seed metabolites. Criterion 1 included metabolites filtered by the biospecimen of "Blood" and the origin of "Endogenous", criterion 2 included metabolites filtered by the origin of "Endogenous", and criterion 3 included all metabolites in the HMDB without any filtering. After restricting the elemental composition to C, H, O, N, P, and S, non-redundant formulas within a mass range of 50~800 were retained. To establish edges between nodes, the differences between any two formulas were obtained. If they matched the pre-defined reactions listed in Table S5, edges between nodes were established. Next, all possible formula candidates of experimental features were generated based on their

accurate mass. The formula differences between the candidates and the nodes in the initial reaction network were calculated using custom Python scripts. If the formula differences matched the pre-defined reactions, edges were connected between the possible formula candidates and the nodes in the initial reaction network. The topological parameters were obtained by network analysis. The molecular network was visualized using Cytoscape (version 3.8.0).

3. Results

3.1. Workflow of the Developed Method

The entire workflow of the developed comprehensive characterization of serum metabolome is displayed in Figure 1. First, untargeted serum metabolome analysis was performed using spectral-stitching acquisition based on DI-nESI HRMS. Possible formula candidates for all the detected monoisotopic MS features were generated according to the rules outlined in the Data processing section with a mass accuracy of 2 ppm. A reaction network approach was then used to screen reliable formula candidates, using 76 predefined biological reactions as edge species (Table S5). The initial reaction network was constructed using unique molecular formula records in the metabolome database as seed nodes. The nodes were connected when their formula difference met with the predefined reactions. Then, the formula difference between all possible candidates and nodes in the initial reaction network was calculated. If the formula difference satisfies the predefined reaction, a connection is established between the candidate formula and seed nodes. For the monoisotopic MS features with multiple formula candidates connected with the initial reaction network, a scoring system was applied to select the top-ranked candidate as the unique formula. These unique formula candidates were newly-seeded nodes to integrate with the initial seeds and reconstructed the reaction network. For the monoisotopic MS features without any formula candidates connected to the initial reaction network, the formula candidates were screened using the reconstructed network again. This filtering process was repeated until no additional features were assigned with unique formulas. Finally, all the unique formula assignments were summarized.

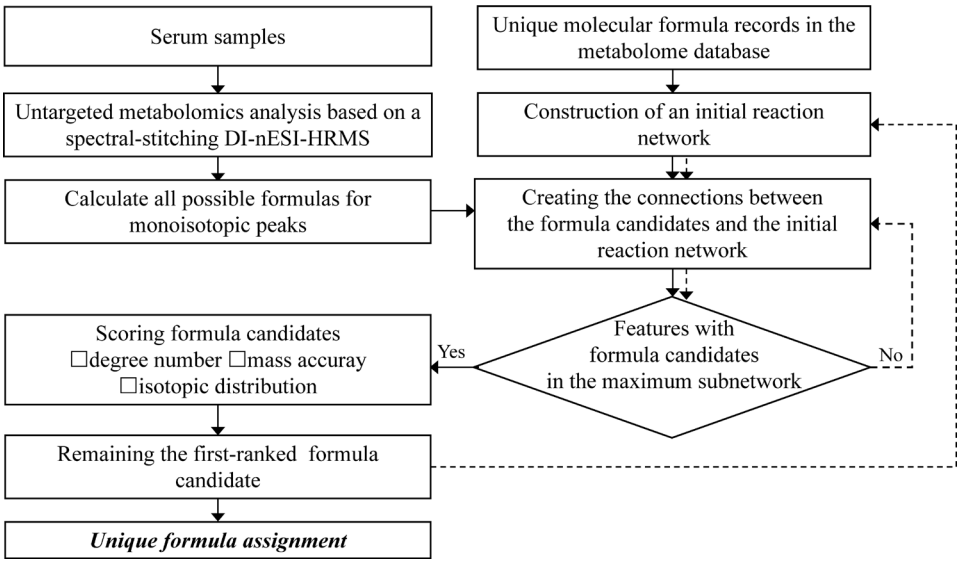

Figure 1. The workflow of a comprehensive serum metabolome analysis based on DI-nESI HRMS.

3.2. The Method Establishment

To evaluate the mass accuracy of spectral-stitching DI-nESI HRMS, 78 reference metabolites dissolved in pure solvent (standard mixture) were analyzed (Table S1). A total of 101 monoisotopic MS features related to the standard mixture, including three adduct types ($[M+H]^+$, $[M+Na]^+$, and $[M+K]^+$) were detected. As shown in Figure S1, the

mass error distribution of the metabolites was mainly clustered within 2 ppm and near zero. Therefore, a mass error of 2 ppm was used in the following formula assignment.

Metabolism in a biological system is highly interconnected through metabolic reactions. Thus, the relevance between potential formula candidates and known endogenous metabolites in a reaction network can be leveraged to estimate the reliability of the formula assignment. The HMDB is a comprehensive database that contains known human metabolites [11]. To improve the accuracy and coverage of formula assignments, metabolites were selected from the HMDB and served as seed nodes for constructing an initial reaction network. To investigate the impact of seed nodes on formula assignment, three filtering criteria were employed for obtaining "endogenous metabolites from blood" (criterion 1), "endogenous metabolites" (criterion 2), and "all metabolites" (criterion 3) in the HMDB database. A total of 1650, 5620, and 12,351 non-redundant formulas were obtained for criteria 1, 2, and 3, respectively. Three initial metabolic reaction networks were constructed, in which 1457/4536, 5324/27,598, and 11,597/72,383 nodes/edges were obtained using criteria 1, 2, and 3 (Figure 2).

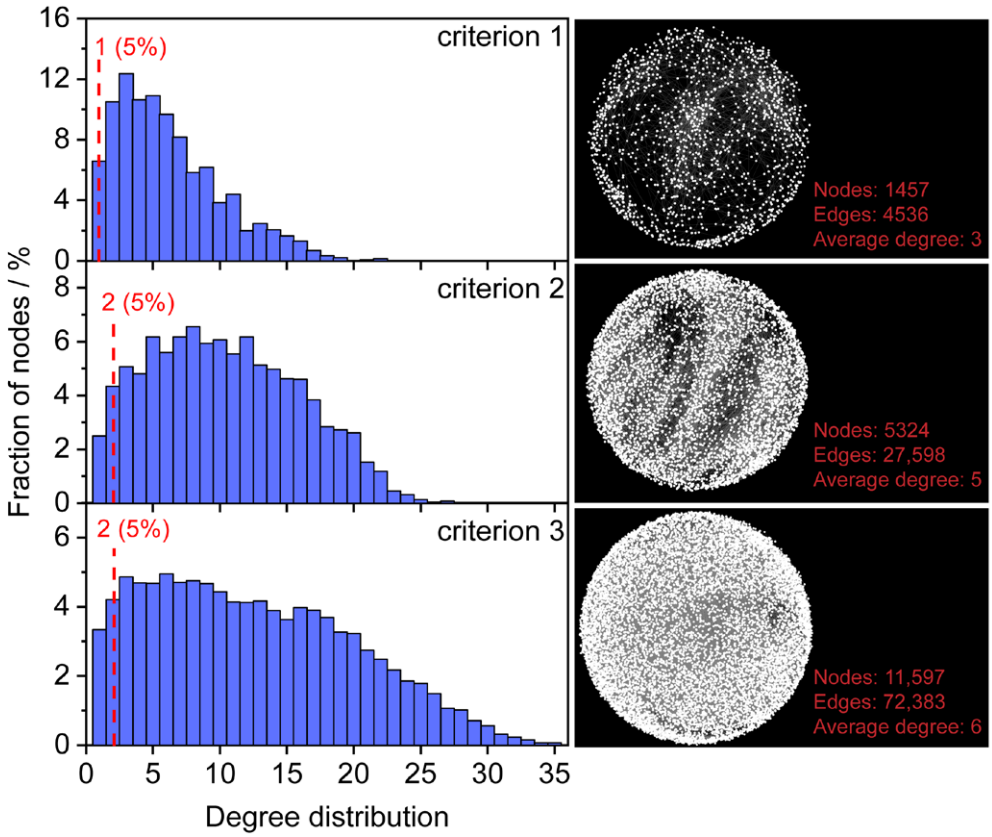

Figure 2. Degree distributions of seed nodes in the initial network. Criterion 1 represents endogenous metabolites from blood (1650), criterion 2 represents endogenous (5620) and criterion 3 represents all metabolites in the HMDB (12,351).

Degree distributions of the three initial metabolic reaction networks were evaluated (Figure 2). With the increase of non-redundant formulas from 1650 to 12,351, the nodes display closer connections of the average degree from 3 to 6. The fractions of nodes with a degree value less than three were 17.1% (criterion 1), 6.8% (criterion 2), and 7.5% (criterion 3), respectively. The initial network constructed by endogenous metabolites (criterion 2) had the lowest proportion of low-degree nodes. Considering formula candidates with a high degree may present higher confidence, the cut-off values of the degree to the three initial reaction networks were set 2, 3, and 3, respectively.

The above standard mixture was used to assess three initial reaction networks. The overlap of the molecular formula between the standard mixture and metabolites in criteria 1, 2, and 3 was 58, 72, and 77, respectively (Figure S2). Using a mass accuracy of 2 ppm, 457 possible formula candidates for 101 features in the standard mixture were generated. All the possible formula candidates were further connected with three initial reaction networks. Among these possible formula candidates, 140 possible formula candidates corresponding to 98 features (criterion 1), 166 formula candidates corresponding to 101 features (criterion 2), and 198 possible formula candidates corresponding to 101 features have a network connection. A total of 119/140, 127/166, and 171/198 possible formula candidates had one more (criterion 1) and two more neighbor nodes (criteria 2 and 3). Through the network filter, the number of MS features with multiple formula candidates increased from 19 (criterion 1) to 23 (criterion 2) and 32 (criterion 3), respectively.

After filtering by the reaction network, there were still many monoisotopic MS features with multiple formula candidates. A scoring rule was established to further rank formula candidates using three alternative scoring criteria, including mass accuracy, isotope pattern, and degree of formula candidates (Equation (1)).

$$\text{Score} = W_{\text{degree}}\text{Score}_{\text{degree}} - W_{m/z}\text{Score}_{m/z} + W_{\text{iso}}\,\text{Score}_{\text{iso}} \quad (1)$$

The W_{degree}, $W_{m/z}$ and W_{iso} represent the weight coefficients of each criterion. $\text{Score}_{m/z}$ represents the mass accuracy score between experimental m/z (m/z_{E}) and theoretical m/z (m/z_{T}) values (Equation (2)):

$$\text{Score}_{m/z} = \frac{|(m/z_{\text{E}} - m/z_{\text{T}})|}{m/z_{\text{E}} \times 2} \times 10^6 \quad (2)$$

$\text{Score}_{\text{iso}}$ represents the similarity of isotopic distribution between experimental and theory isotopic patterns (Equation (3)):

$$\text{Score}_{\text{iso}} = 0.5 \times \text{Score}_{mz} + 0.5 \times \text{Score}_{\text{int}} \quad (3)$$

In which $\text{Score}_{mz} = 1 - \frac{|(m/z_{\text{E}} - m/z_{\text{T}})|/(m/z_{\text{T}}) \times 10^6}{\text{Tolerance}_{m/z}}$ and $\text{Score}_{\text{int}} = 1 - \frac{|(\text{int}_{\text{E}} - \text{int}_{\text{T}})|/(\text{int}_{\text{T}}) \times 10^6}{\text{Tolerance}_{\text{int}}}$. Where int_{E} and int_{T} are the experimental and theoretical relative intensity, respectively. $\text{Tolerance}_{m/z}$ and $\text{Tolerance}_{\text{int}}$ represent the mass tolerance of 2 ppm and the relative intensity tolerance of 500%, respectively [38].

The effects of the weight coefficients on formula assignments were evaluated using the above standard mixture (Figure S3). If only accurate mass filtering (W_{degree} = 0, $W_{m/z}$ = 1, W_{iso} = 0) was used, the correct assignment rate for standards was 73.3~81.2%. If degree filtering was considered alone (with W_{degree} = 1, $W_{m/z}$ = 0, W_{iso} = 0), the correct assignment rate for standards increased to 85.1~87.1%. The combination of all three scoring criteria resulted in a correct assignment rate of approximately 90%. The optimal weight coefficients were determined to be W_{degree} = 0.5, $W_{m/z}$ = 0.3, and W_{iso} = 0.2.

The assignment performance of three initial reaction networks for the standard mixture was compared using the optimal weight coefficients (Figure 3). With the increase of non-redundant formulas from 1650 to 12,351, the correct assignment rate slightly increased from 89.1% to 93.1%. Using criteria 1 and 2, 4% of monoisotopic MS features were not retained any candidates because they lacked connections with the initial network. Furthermore, to assess the performance of the three initial networks, a target-decoy strategy was implemented. A set of 1000 known formulas presented in the HMDB database (blood matrix) and 1000 decoy formulas absent in the HMDB database (token from the ChEMBL database) were used. The line charts in Figure 3 demonstrate a similar filtering trend between the known/decoy formulas and the standard mixture. The percentage of formula assignment rates for known formulas increased slightly from 86.3%, 91.0% to 97.9% (solid line), while the percentage of formula assignment rates for decoy formulas significantly increased from 14.9%, 22.6% to 61.5% (dash line). The results implied that initial networks

under criteria 1 and 2 exhibited more specific assignments for the biosample compared to criterion 3. As a result, the initial network consisting of 5620 endogenous metabolites (criterion 2) demonstrated superior assignment performance and was used for the subsequent characterization.

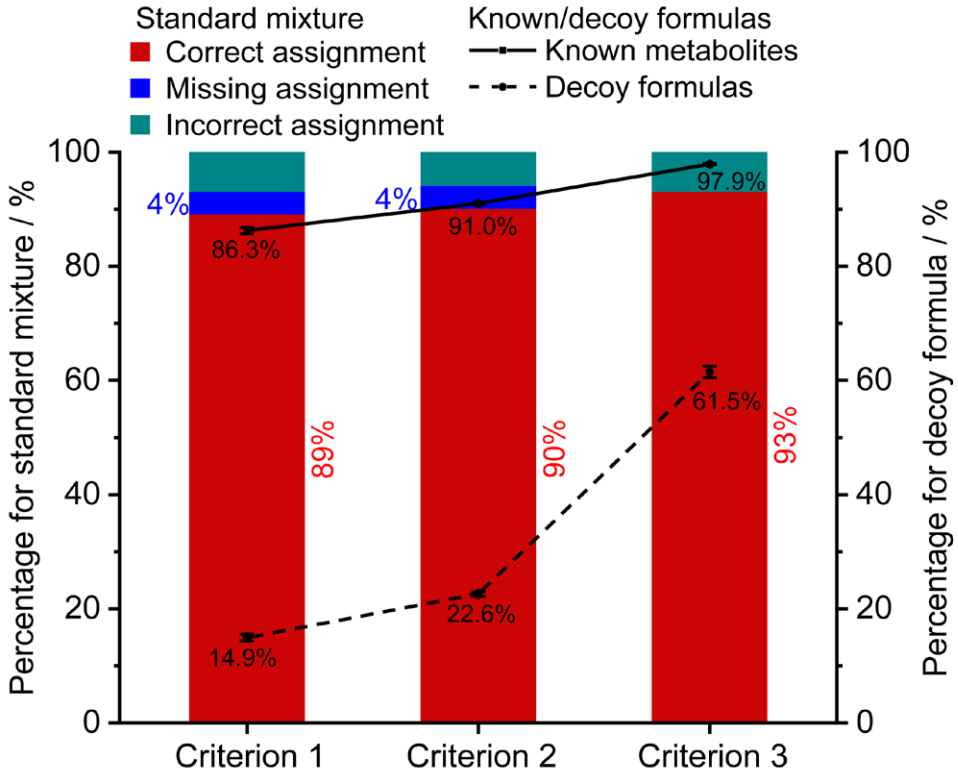

Figure 3. The effects of the initial networks on the assignment performance. The formula assignment results of the standard mixture (stacked column), 1000 blood metabolites (solid line), and 1000 decoy formulas (dash line). Criterion 1, 2, and 3 represent the initial network constructed by 1650 endogenous formulas from blood, 5620 endogenous formulas, and all the 12,351 formulas in the HMDB database, respectively. Error bars represent ±SE (n = 3).

3.3. Method Validation

A pooled serum sample spiked 78 reference standards was used for method validation. As shown in Figure 4A, 96 monoisotopic MS features were detected to be associated with 68 reference standards, which were assigned 353 potential formula candidates. Using the workflow described above, 95 out of the 96 monoisotopic MS features were successfully assigned with unique formula candidates through the two rounds of assignments (Figure 4A, stacked column). A total of 89 out of 95 assigned features were correctly allocated unique formulas. As a result, the coverage and accuracy rates were 98.9% (red solid line) and 93.6% (red dash line), respectively. For the spiked pooled serum sample, 3140 monoisotopic MS features were assigned with 19,474 possible formula candidates. Among these features, 1958 features were assigned with unique formula candidates after seven rounds of assignments (grey columns). The coverage rate was up to 62.3% (grey solid line). It is noteworthy that 18.1% of features were assigned between the second and seventh rounds.

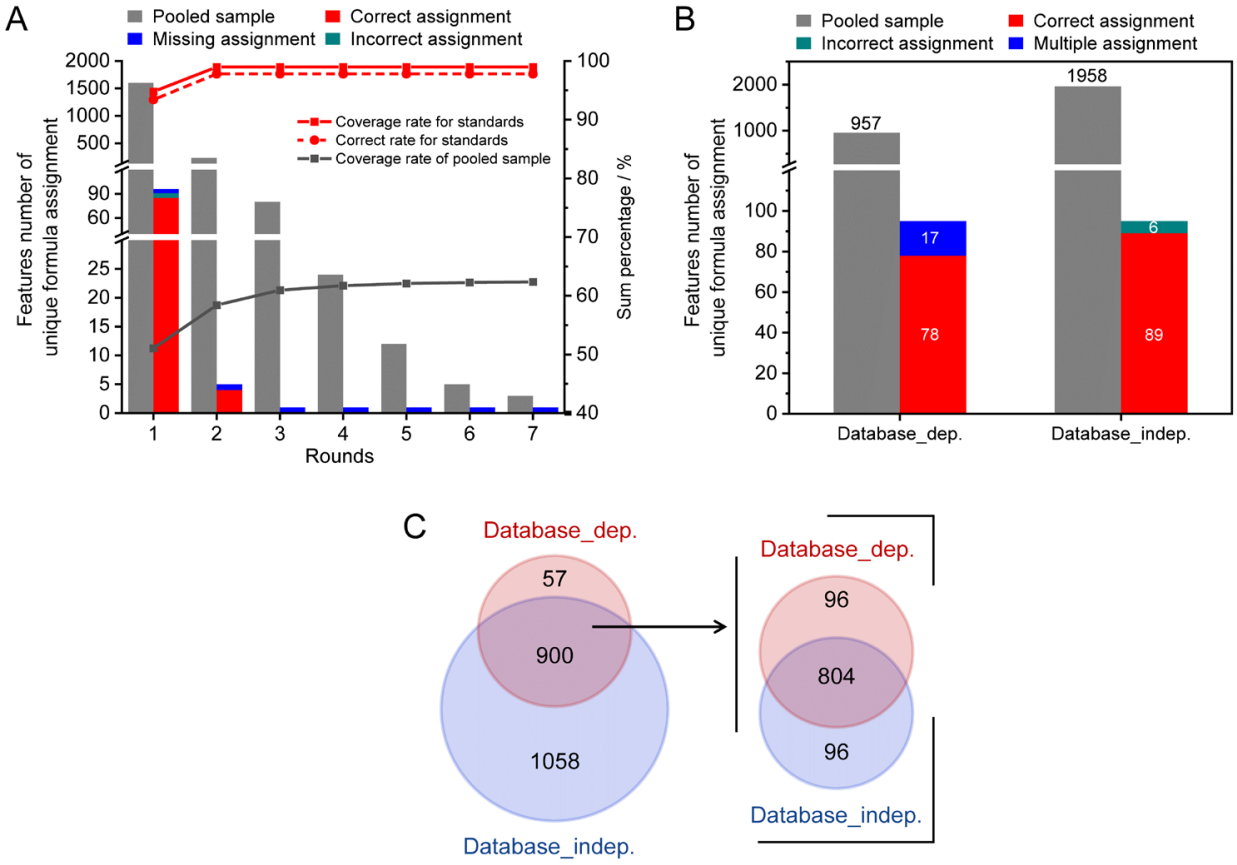

Figure 4. The method validation results using a spiked serum sample. (**A**) The coverage and accuracy rates of the developed method. (**B**) Comparison of assignment performance between the developed method and the database search. (**C**) The Venn diagram of assignment results of both methods for the spiked pooled serum sample.

The assignment performance of the developed method was compared with the conventional formula assignments method of searching m/z against the metabolome database (database-dependent method). The results are shown in Figure 4B. For the 68 spiked reference standards, the coverage rates of the two methods were similar (99%). However, the accuracy rates of unambiguous formula candidates were improved from 81.3% (78/96) to 92.7% (89/96) by the developed method. A total of 17.7 % (17/96) of MS features obtained multiple formula candidates by searching m/z against the HMDB (stacked column). For the spiked pooled serum sample, only 957 MS features were able to obtain unambiguous formula candidates by HMDB database search using a mass tolerance of 2 ppm, which yielded less than half the number of formulas compared to the developed method. The Venn diagram of the assigned monoisotopic MS features by two methods is shown in Figure 4C. It can be observed that the assigned monoisotopic MS features by the database search had good agreement with the developed method (900 out of 957). Among these 900 shared assigned monoisotopic MS features, about 89% (804/900) of MS features annotated consistent molecular formulas. The results indicated that the developed method showed high coverage and accuracy for formula assignment in the characterization of complex biosamples.

3.4. Application for Serum Metabolomic Analysis in Diabetes

To verify the practicability of the developed method, it was used to investigate the metabolic alterations in diabetes. Table S3 lists the detailed clinical information of participants corresponding to two groups of healthy control and diabetes patients. No significant differences except for blood glucose concentration were observed between the two groups.

The repeatability of the developed acquisition method was measured using QC samples. The RSD values were estimated (Figure S4A). It shows that 86% of the detected MS features had RSD less than 30%, accounting for 96% of the sum peak area. Then, only those assigned features with an RSD of less than 30% in QC were subjected to further statistical analysis. The method linearities were evaluated using 10 reference metabolites (Table S4). The method exhibited good linearities with linear correlation coefficients (R^2) of 0.9975–1, and linear ranges spanning two to four orders of magnitude. The results demonstrated that the current acquisition method was appropriate for serum metabolomics analysis.

Figure S4B represents a PLS-DA score plot of two groups. The separated clusters were observed between healthy control and patients. To further validate the classification models, the permutation test was performed (M = 200). The intercept values of R^2 and Q^2 were below 0.4 and −0.05, respectively, which indicated no overfitting of the PLS-DA model (Figure S4C).

A total of 57 significantly differential formulas were obtained between healthy control and T2D patients with VIP > 1 and $p < 0.05$ (Table S6). A heat map of the significantly changed metabolic features revealed good clustering between the control and T2D groups (Figure 5A). Four typical metabolites ($[C_6H_{12}O_6+Na]^+$, $[C_6H_{13}NO_2+H]^+$, $[C_6H_{11}NO_2+Na]^+$ and $[C_7H_{15}NO_3+H]^+$) are displayed in Figure 5B. Compared with the controls, significantly increased $[C_6H_{12}O_6+Na]^+$ (glucose), $[C_6H_{13}NO_2+H]^+$ (Leucine/Isoleucine) and decreased $[C_6H_{11}NO_2+Na]^+$ (pipecolinic acid), $[C_7H_{15}NO_3+H]^+$ (carnitine) in diabetes were observed. As shown in the clinical information (Table S3), blood glucose existed a significant difference between healthy and T2D groups, it was consistent with our result as $[C_6H_{12}O_6+Na]^+$. In the previous study, the abnormal concentrations of carnitines also appeared in the T2D groups owing to the fatty acid oxidation dysregulation [39,40], our result showed a decrease of $[C_7H_{15}NO_3+H]^+$ in T2D group. A positive association of branch-chain amino acids in diabetes had been concluded by modulating insulin secretion and leading to the pancreaticb-cell exhaustion [41,42]. Leucine/isoleucine ($C_6H_{13}NO_2$) recognized as branch chain amino acids had a robust association with the risk of T2D, a similar trend with $[C_6H_{13}NO_2+H]^+$ is shown in Figure 5B. In addition, Ouyang et al. observed that pipecolinic acid ($C_6H_{11}NO_2$) was decreased in T2D patients in a Chinese prospective cohort study [43].

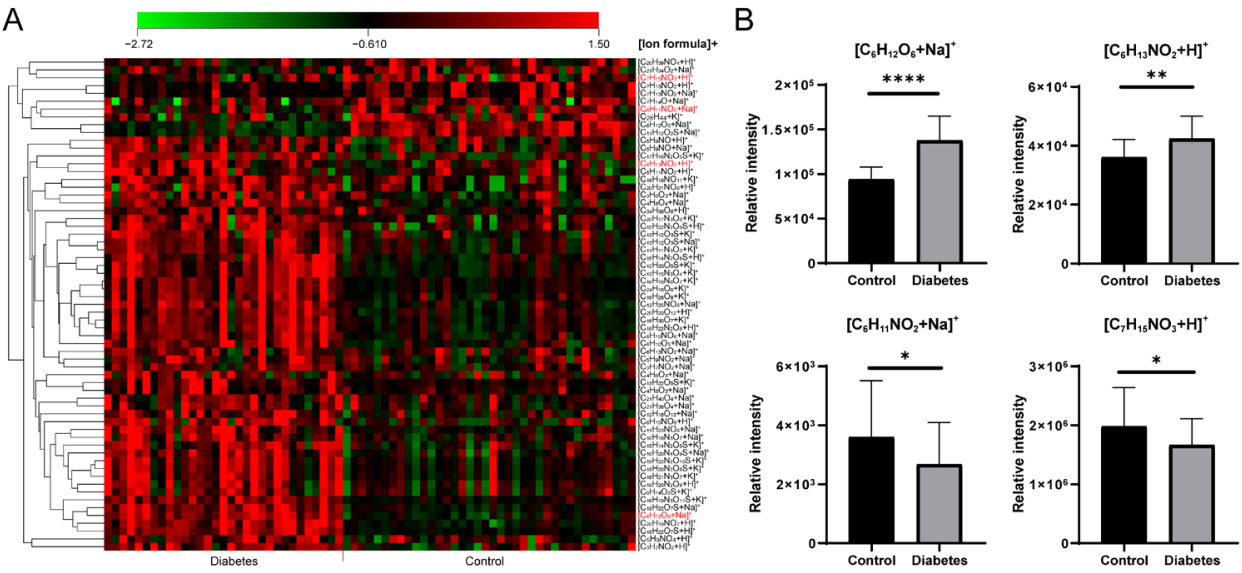

Figure 5. (**A**) The heat map of significantly changed ion formulas between control and diabetes groups, (**B**) Changes of $[C_6H_{12}O_6+Na]^+$, $[C_6H_{13}NO_2+H]^+$, $[C_6H_{11}NO_2+Na]^+$ and $[C_7H_{15}NO_3+H]^+$ between control and diabetes groups, respectively. **** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$.

4. Discussion

DI-nESI HRMS is a promising high-throughput analytical method for large-scale cohort metabolomic study [44,45]. High coverage and reliable assignment of MS features is a critical prerequisite for comprehensive metabolomics analysis [46]. In the present study, we aimed to develop a comprehensive metabolomics analysis method in direct-infusion high-resolution mass spectrometry. Our result revealed that based on a reaction network combing with mass accuracy and isotopic distribution, more MS peaks could be assigned with a confident unique formula.

For untargeted metabolomics analysis based on DI-HRMS, the most widely used approach for formula assignments was searching m/z against metabolome databases. According to the performance of mass resolution (Rs) and mass accuracy in the high-resolution mass spectrometer, appropriate mass tolerance parameter was obtained, such as 10~20 ppm for time-of-flight (Rs on the order of 35,000), 5~15 ppm for orbitrap (Rs on the order of 200,000 at m/z 200) and 0.5 ppm for Fourier transform ion cyclotron mass spectrometers (Rs 900,000 at m/z 200) [47,48]. However, even with tremendous advances in database size and scape, the identification of metabolites still is a primary challenge in the field [49,50].

The relationship between metabolites attracted wide attention for formula identification in untargeted DI ultrahigh resolution MS. Reaction network approach based on mass difference had been used for MS feature assignments, in which experimental features as nodes and chemical reactions as edges [19,22]. However, reliable and enough experimental features need to be first assigned as initial seed nodes. It is difficult to achieve for the untargeted metabolomics analysis using HRMS. Instead of only experimental features used for the reaction network, the formula records from the metabolome database were used to construct the initial reaction network in this study. The connections between formula candidates and the initial reaction network were established for filtering multiple candidates. The assignment performance was improved due to sufficient and accurate known formulas from the initial network. Furthermore, using assigned unambiguous formula candidates as newly added seeds to reconstruct networks, multiple rounds of annotation were designed to further improve annotation efficiency (1.22-fold increase after 7 rounds in this study). Integration of mass accuracy and isotopic pattern filter resulted in a further improvement in assignment accuracy.

Although the developed method demonstrated more than twice the number of unique assigned formulas for a pooled serum sample compared to the database search, still about 38% of MS features could not obtain unambiguous formula candidates. The false negatives are possibly caused by the initial network filter. To avoid false positives, the initial network in this study was constructed using only the most common metabolic reactions and endogenous formulas from the HMDB database. Furthermore, the excellent assignment performance for serum untargeted metabolic features is partly attributed to the comprehensive HMDB database for the construction of the initial reaction network. Nevertheless, for the less studied objects, the method performance may be suboptimal because the initial network can't be established or the initial network size is too small. Furthermore, it is essential to accurately confirm the structures of differential metabolites and then determine key differential metabolites from them in subsequent studies.

5. Conclusions

A comprehensive serum metabolome characterization method in DI-nESI HRMS was proposed in this study. High coverage, high confidence, and database-independent formula assignments were achieved by reaction network combined with mass accuracy and isotopic pattern filter. A total of 1958 monoisotopic features were assigned the unique formula in a pooled serum sample, while only 957 were annotated by searching m/z against HMDB. This method has the advantages of comprehensive formula assignment and high-speed acquisition, it has great application potential in large-scale cohort metabolomic studies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/metabo13030460/s1, Figure S1: Mass error distribution of metabolites standards; Figure S2: The overlap between the standards formulas and seed nodes in three initial reaction networks; Figure S3: Weight coefficients optimization for the degree, mass error and isotopic distribution; Figure S4: (A) Estimation of detection repeatability. (B) PLS-DA scores plot of healthy and diabetes groups. (C) Cross-validation plots of the PLS-DA model; Table S1: Information of metabolites standards; Table S2: Information of internal standards; Table S3: Clinical information of participants; Table S4: Information of 10 standards for linearity; Table S5. Reactions types; Table S6. Significantly changed features between healthy and T2D groups.

Author Contributions: Conceptualization, X.S., C.Z., X.Z., X.L. and G.X.; methodology, X.S. and Z.J.; software, Z.J.; validation, X.S.; formal analysis, X.S.; investigation, X.S.; resources, Y.Z.; data curation, X.S. and Z.J.; writing—original draft preparation, X.S.; writing—review and editing, C.Z., X.Z., X.L. and G.X.; supervision, X.Z., X.L. and G.X.; funding acquisition, X.L. and G.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (21934006, 22274153, 22174141, 22274151, and 22074145), the Innovation Program (DICP I202004) of Science and Research from Dalian Institute of Chemical Physics, Chinese Academy of Sciences, and the AI S&T Program (Grant No. DNL-YL A202202) from Yulin Branch, Dalian National Laboratory for Clean Energy, CAS, China.

Institutional Review Board Statement: The study was approved by the Ethics Committee of the Second Hospital of Dalian Medical University (2019(124) and October 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data will be made available upon request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Southam, A.D.; Weber, R.J.M.; Engel, J.; Jones, M.R.; Viant, M.R. A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics. *Nat. Protoc.* **2016**, *12*, 310–328. [CrossRef] [PubMed]
2. Habchi, B.; Alves, S.; Jouan-Rimbaud Bouveresse, D.; Moslah, B.; Paris, A.; Lécluse, Y.; Gauduchon, P.; Lebailly, P.; Rutledge, D.N.; Rathahao-Paris, E. An innovative chemometric method for processing direct introduction high resolution mass spectrometry metabolomic data: Independent component–discriminant analysis (IC–DA). *Metabolomics* **2017**, *13*, 45. [CrossRef]
3. Chekmeneva, E.; Dos Santos Correia, G.; Gómez-Romero, M.; Stamler, J.; Chan, Q.; Elliott, P.; Nicholson, J.K.; Holmes, E. Ultra-Performance Liquid Chromatography–High-Resolution Mass Spectrometry and Direct Infusion–High-Resolution Mass Spectrometry for Combined Exploratory and Targeted Metabolic Profiling of Human Urine. *J. Proteome Res.* **2018**, *17*, 3492–3502. [CrossRef]
4. Pöhö, P.; Lipponen, K.; Bespalov, M.M.; Sikanen, T.; Kotiaho, T.; Kostiainen, R. Comparison of liquid chromatography-mass spectrometry and direct infusion microchip electrospray ionization mass spectrometry in global metabolomics of cell samples. *Eur. J. Pharm. Sci.* **2019**, *138*, 104991. [CrossRef] [PubMed]
5. Malinowska, J.M.; Palosaari, T.; Sund, J.; Carpi, D.; Bouhifd, M.; Weber, R.J.M.; Whelan, M.; Viant, M.R. Integrating in vitro metabolomics with a 96-well high-throughput screening platform. *Metabolomics* **2022**, *18*, 11. [CrossRef] [PubMed]
6. Bowen, T.J.; Hall, A.R.; Lloyd, G.R.; Weber, R.J.M.; Wilson, A.; Pointon, A.; Viant, M.R. An Extensive Metabolomics Workflow to Discover Cardiotoxin-Induced Molecular Perturbations in Microtissues. *Metabolites* **2021**, *11*, 644. [CrossRef]
7. González-Domínguez, R.; Sayago, A.; Fernández-Recamales, Á. Direct infusion mass spectrometry for metabolomic phenotyping of diseases. *Bioanalysis* **2017**, *9*, 131–148. [CrossRef]
8. Chekmeneva, E.; Dos Santos Correia, G.; Chan, Q.; Wijeyesekera, A.; Tin, A.; Young, J.H.; Elliott, P.; Nicholson, J.K.; Holmes, E. Optimization and Application of Direct Infusion Nanoelectrospray HRMS Method for Large-Scale Urinary Metabolic Phenotyping in Molecular Epidemiology. *J. Proteome Res.* **2017**, *16*, 1646–1658. [CrossRef]
9. de Sain-van der Velden, M.G.M.; van der Ham, M.; Gerrits, J.; Prinsen, H.C.M.T.; Willemsen, M.; Pras-Raves, M.L.; Jans, J.J.; Verhoeven-Duif, N.M. Quantification of metabolites in dried blood spots by direct infusion high resolution mass spectrometry. *Anal. Chim. Acta* **2017**, *979*, 45–50. [CrossRef]
10. Haijes, H.A.; van der Ham, M.; Gerrits, J.; van Hasselt, P.M.; Prinsen, H.C.M.T.; de Sain-van der Velden, M.G.M.; Verhoeven-Duif, N.M.; Jans, J.J.M. Direct-infusion based metabolomics unveils biochemical profiles of inborn errors of metabolism in cerebrospinal fluid. *Mol. Genet. Metab.* **2019**, *127*, 51–57. [CrossRef] [PubMed]

11. Haijes, H.A.; Willemsen, M.; Van der Ham, M.; Gerrits, J.; Pras-Raves, M.L.; Prinsen, H.C.M.T.; Van Hasselt, P.M.; De Sain-van der Velden, M.G.M.; Verhoeven-Duif, N.M.; Jans, J.J.M. Direct Infusion Based Metabolomics Identifies Metabolic Disease in Patients' Dried Blood Spots and Plasma. *Metabolites* **2019**, *9*, 12. [CrossRef] [PubMed]
12. Wang, L.; Lv, W.; Sun, X.; Zheng, F.; Xu, T.; Liu, X.; Li, H.; Lu, X.; Peng, X.; Hu, C.; et al. Strategy for Nontargeted Metabolomic Annotation and Quantitation Using a High-Resolution Spectral-Stitching Nanoelectrospray Direct-Infusion Mass Spectrometry with Data-Independent Acquisition. *Anal. Chem.* **2021**, *93*, 10528–10537. [CrossRef] [PubMed]
13. Ludwig, M.; Nothias, L.-F.; Dührkop, K.; Koester, I.; Fleischauer, M.; Hoffmann, M.A.; Petras, D.; Vargas, F.; Morsy, M.; Aluwihare, L.; et al. Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nat. Mach. Intell.* **2020**, *2*, 629–641. [CrossRef]
14. Wolthuis, J.C.; Magnusdottir, S.; Pras-Raves, M.; Moshiri, M.; Jans, J.J.M.; Burgering, B.; van Mil, S.; de Ridder, J. MetaboShiny: Interactive analysis and metabolite annotation of mass spectrometry-based metabolomics data. *Metabolomics* **2020**, *16*, 99. [CrossRef]
15. Mitchell, J.M.; Flight, R.M.; Moseley, H.N.B. Small Molecule Isotope Resolved Formula Enumeration: A Methodology for Assigning Isotopologues and Metabolite Formulas in Fourier Transform Mass Spectra. *Anal. Chem.* **2019**, *91*, 8933–8940. [CrossRef] [PubMed]
16. Haijes, H.A.; Willemse, E.A.; Gerrits, J.; van der Flier, W.M.; Teunissen, C.E.; Verhoeven-Duif, N.M.; Jans, J.J.M. Assessing the Pre-Analytical Stability of Small-Molecule Metabolites in Cerebrospinal Fluid Using Direct-Infusion Metabolomics. *Metabolites* **2019**, *9*, 236. [CrossRef] [PubMed]
17. Sarvin, B.; Lagziel, S.; Sarvin, N.; Mukha, D.; Kumar, P.; Aizenshtein, E.; Shlomi, T. Fast and sensitive flow-injection mass spectrometry metabolomics by analyzing sample-specific ion distributions. *Nat. Commun.* **2020**, *11*, 3186. [CrossRef]
18. Southam, A.D.; Payne, T.G.; Cooper, H.J.; Arvanitis, T.N.; Viant, M.R. Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method. *Anal. Chem.* **2007**, *79*, 4595–4602. [CrossRef]
19. Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E.A.; Glass, C.K.; Merrill, A.H.; Murphy, R.C.; Raetz, C.R.H.; Russell, D.; et al. LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007**, *35*, D527–D532. [CrossRef]
20. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [CrossRef]
21. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. [CrossRef] [PubMed]
22. Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinform.* **2007**, *8*, 105. [CrossRef] [PubMed]
23. Huang, D.; Bouza, M.; Gaul, D.A.; Leach, F.E.; Amster, I.J.; Schroeder, F.C.; Edison, A.S.; Fernández, F.M. Comparison of High-Resolution Fourier Transform Mass Spectrometry Platforms for Putative Metabolite Annotation. *Anal. Chem.* **2021**, *93*, 12374–12382. [CrossRef] [PubMed]
24. Zielinski, A.T.; Kourtchev, I.; Bortolini, C.; Fuller, S.J.; Giorio, C.; Popoola, O.A.M.; Bogialli, S.; Tapparo, A.; Jones, R.L.; Kalberer, M. A new processing scheme for ultra-high resolution direct infusion mass spectrometry data. *Atmospheric Environ.* **2018**, *178*, 129–139. [CrossRef]
25. Cao, D.; Hao, Z.; Hu, M.; Geng, F.; Rao, Z.; Niu, H.; Shi, Y.; Cai, Y.; Zhou, Y.; Liu, J.; et al. A feasible strategy to improve confident elemental composition determination of compounds in complex organic mixture such as natural organic matter by FTICR-MS without internal calibration. *Sci. Total. Environ.* **2021**, *751*, 142255. [CrossRef] [PubMed]
26. Fiehn, O.; Robertson, D.; Griffin, J.; van der Werf, M.; Nikolau, B.; Morrison, N.; Sumner, L.W.; Goodacre, R.; Hardy, N.W.; Taylor, C.; et al. The metabolomics standards initiative (MSI). *Metabolomics* **2007**, *3*, 175–178. [CrossRef]
27. Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinform.* **2006**, *7*, 234. [CrossRef] [PubMed]
28. Alden, N.; Krishnan, S.; Porokhin, V.; Raju, R.; McElearney, K.; Gilbert, A.; Lee, K. Biologically Consistent Annotation of Metabolomics Data. *Anal. Chem.* **2017**, *89*, 13097–13104. [CrossRef] [PubMed]
29. Nielsen, J. Systems biology of metabolism. *Annu. Rev. Biochem.* **2017**, *86*, 245–275. [CrossRef]
30. Moritz, F.; Kaling, M.; Schnitzler, J.-P.; Schmitt-Kopplin, P. Characterization of poplar metabotypes via mass difference enrichment analysis. *Plant, Cell Environ.* **2017**, *40*, 1057–1073. [CrossRef]
31. Tziotis, D.; Hertkorn, N.; Schmitt-Kopplin, P. Kendrick-Analogous Network Visualisation of Ion Cyclotron Resonance Fourier Transform Mass Spectra: Improved Options for the Assignment of Elemental Compositions and the Classification of Organic Molecular Complexity. *Eur. J. Mass Spectrom.* **2011**, *17*, 415–421. [CrossRef]
32. Witting, M.; Lucio, M.; Tziotis, D.; Wägele, B.; Suhre, K.; Voulhoux, R.; Garvis, S.; Schmitt-Kopplin, P. DI-ICR-FT-MS-based high-throughput deep metabotyping: A case study of the Caenorhabditis elegans–Pseudomonas aeruginosa infection model. *Anal. Bioanal. Chem.* **2015**, *407*, 1059–1073. [CrossRef]
33. Amara, A.; Frainay, C.; Jourdan, F.; Naake, T.; Neumann, S.; Novoa-Del-Toro, E.M.; Salek, R.M.; Salzer, L.; Scharfenberg, S.; Witting, M. Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation. *Front. Mol. Biosci.* **2022**, *9*, 841373. [CrossRef] [PubMed]

34. Schmitt-Kopplin, P.; Hemmler, D.; Moritz, F.; Gougeon, R.D.; Lucio, M.; Meringer, M.; Müller, C.; Harir, M.; Hertkorn, N. Systems chemical analytics: Introduction to the challenges of chemical complexity analysis. *Faraday Discuss.* **2019**, *218*, 9–28. [CrossRef] [PubMed]
35. Traquete, F.; Luz, J.; Cordeiro, C.; Silva, M.S.; Ferreira, A.E.N. Graph Properties of Mass-Difference Networks for Profiling and Discrimination in Untargeted Metabolomics. *Front. Mol. Biosci.* **2022**, *9*, 917911. [CrossRef]
36. Chen, L.; Lu, W.; Wang, L.; Xing, X.; Chen, Z.; Teng, X.; Zeng, X.; Muscarella, A.D.; Shen, Y.; Cowan, A.; et al. Metabolite discovery through global annotation of untargeted metabolomics data. *Nat. Methods* **2021**, *18*, 1377–1385. [CrossRef] [PubMed]
37. Forcisi, S.; Moritz, F.; Lucio, M.; Lehmann, R.; Stefan, N.; Schmitt-Kopplin, P. Solutions for Low and High Accuracy Mass Spectrometric Data Matching: A Data-Driven Annotation Strategy in Nontargeted Metabolomics. *Anal. Chem.* **2015**, *87*, 8917–8924. [CrossRef] [PubMed]
38. Shen, X.; Wang, R.; Xiong, X.; Yin, Y.; Cai, Y.; Ma, Z.; Liu, N.; Zhu, Z.-J. Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. *Nat. Commun.* **2019**, *10*, 1516. [CrossRef]
39. Sun, L.; Liang, L.; Gao, X.; Zhang, H.; Yao, P.; Hu, Y.; Ma, Y.; Wang, F.; Jin, Q.; Li, H.; et al. Early Prediction of Developing Type 2 Diabetes by Plasma Acyl-carnitines: A Population-Based Study. *Diabetes Care* **2016**, *39*, 1563–1570. [CrossRef] [PubMed]
40. Guasch-Ferré, M.; Ruiz-Canela, M.; Li, J.; Zheng, Y.; Bullo, M.; Wang, D.D.; Toledo, E.; Clish, C.; Corella, D.; Estruch, R.; et al. Plasma Acylcarnitines and Risk of Type 2 Diabetes in a Mediterranean Population at High Cardiovascular Risk. *J. Clin. Endocrinol. Metab.* **2019**, *104*, 1508–1519. [CrossRef]
41. Guasch-Ferré, M.; Hruby, A.; Toledo, E.; Clish, C.B.; Martínez-González, M.A.; Salas-Salvadó, J.; Hu, F.B. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care* **2016**, *39*, 833–846. [CrossRef] [PubMed]
42. Vanweert, F.; Schrauwen, P.; Phielix, E. Role of branched-chain amino acid metabolism in the pathogenesis of obesity and type 2 diabetes-related metabolic disturbances BCAA metabolism in type 2 diabetes. *Nutr. Diabetes* **2022**, *12*, 35. [CrossRef] [PubMed]
43. Ouyang, Y.; Qiu, G.; Zhao, X.; Su, B.; Feng, D.; Lv, W.; Xuan, Q.; Wang, L.; Yu, D.; Wang, Q.; et al. Metabolome-Genome-Wide Association Study (mGWAS) Reveals Novel Metabolites Associated with Future Type 2 Diabetes Risk and Susceptibility Loci in a Case-Control Study in a Chinese Prospective Cohort. *Glob. Chall.* **2021**, *5*, 2000088. [CrossRef]
44. Fuhrer, T.; Zamboni, N. High-throughput discovery metabolomics. *Curr. Opin. Biotechnol.* **2015**, *31*, 73–78. [CrossRef] [PubMed]
45. Kempa, E.E.; Hollywood, K.A.; Smith, C.A.; Barran, P.E. High throughput screening of complex biological samples with mass spectrometry–from bulk measurements to single cell analysis. *Analyst* **2019**, *144*, 872–891. [CrossRef] [PubMed]
46. Kozlova, A.; Shkrigunov, T.; Gusev, S.; Guseva, M.; Ponomarenko, E.; Lisitsa, A. An Open-Source Pipeline for Processing Direct Infusion Mass Spectrometry Data of the Human Plasma Metabolome. *Metabolites* **2022**, *12*, 768. [CrossRef] [PubMed]
47. Forsberg, E.M.; Huan, T.; Rinehart, D.; Benton, H.P.; Warth, B.; Hilmers, B.; Siuzdak, G. Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. *Nat. Protoc.* **2018**, *13*, 633–651. [CrossRef]
48. Thompson, C.J.; Witt, M.; Forcisi, S.; Moritz, F.; Kessler, N.; Laukien, F.H.; Schmitt-Kopplin, P. An Enhanced Isotopic Fine Structure Method for Exact Mass Analysis in Discovery Metabolomics: FIA-CASI-FTMS. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 2025–2034. [CrossRef] [PubMed]
49. Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C.E. Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. *Curr. Opin. Biotechnol.* **2019**, *55*, 44–50. [CrossRef]
50. Baygi, S.F.; Banerjee, S.K.; Chakraborty, P.; Kumar, Y.; Barupal, D.K. IDSL.UFA Assigns High-Confidence Molecular Formula Annotations for Untargeted LC/HRMS Data Sets in Metabolomics and Exposomics. *Anal. Chem.* **2022**, *94*, 13315–13322. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.