<div align="center">Original Article</div>

Data Processing of Product Ion Spectra: Methods to Control False Discovery Rate in Compound Search Results for Untargeted Metabolomics

Fumio Matsuda [*,1,2]

Graduate School of Information Science and Technology, Osaka University, Osaka 565–0871, Japan
Osaka University Shimadzu Omics Innovation Research Laboratories, Osaka University, Osaka 565–0871, Japan

Several database search methods have been employed in untargeted metabolomics utilizing high-resolution mass spectrometry to comprehensively annotate acquired product ion spectra. Recent technical advancements in *in silico* analyses have facilitated the sorting of the degree of coincidence between a query product ion spectrum, and the molecular structures in the database. However, certain search results may be false positives, necessitating a method for controlling the false discovery rate (FDR). This study proposes 4 simple methods for controlling the FDR in compound search results. Instead of preparing a decoy compound database, a decoy spectral dataset was created from the measured product-ion spectral dataset (target). Target and decoy product ion spectra were searched against an identical compound database to obtain target and decoy hits. FDR was estimated based on the number of target and decoy hits. In this study, 3 decoy generation methods, polarity switching, mirroring, and spectral sampling, were compared. Additionally, the second-rank method was examined using second-ranked hits in the target search results as decoy hits. The performances of these 4 methods were evaluated by annotating product ion spectra from the MassBank database using the SIRIUS 5 CSI:FingerID scoring method. The results indicate that the FDRs estimated using the second-rank method were the closest to the true FDR of 0.05. Using this method, a compound search was performed on 4 human metabolomic data-dependent acquisition datasets with an FDR of 0.05. The FDR-controlled compound search successfully identified several compounds not present in the Human Metabolome Database.

1. INTRODUCTION

In untargeted metabolomics utilizing high-resolution mass spectrometry, researchers employ several database search methods to comprehensively annotate acquired product ion spectra.[1] Molecular formula searches are performed using accurate m/z values, isotopic abundance patterns of the precursor ions, and several empirical rules of molecular formula.[2,3] A spectral similarity search approach was employed to estimate the compound structure using spectral databases such as MassBank[4] (Table 1). Recent technical advancements in *in silico* analyses have facilitated the sorting of the degree of coincidence between a query product ion spectrum, and the molecular structures in the database. Several software packages that implement these methods, including MS-FINDER,[5] SIRIUS,[6,7] MetFrag,[8,9] and CFM-ID,[10,11] have been developed

in recent years (see Table 1). These software packages generated a ranked list of compound scores for each queried product ion spectrum. In this study, the top-ranked compound in the list is referred to as a "hit." A compound search was performed for each product ion spectrum in the metabolomic dataset. Consequently, because some hits may be false positives, a score threshold must be set to control the false discovery rate (FDR) within an acceptable range.[12]

In proteomics, the FDR of the peptide identification results is estimated using the target–decoy method (Fig. 1A).[13,14] For peptide identification, crude protein extract was digested with trypsin. The sample was used for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis in the data-dependent acquisition (DDA) mode to acquire product ion spectra from trypsin-digested peptides comprehensively. Peptide identification was accomplished by searching

---

*Correspondence to: Fumio Matsuda, Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 1–5 Yamadaoka, Suita, Osaka 565–0871, Japan, e-mail: fmatsuda@ist.osaka-u.ac.jp

Table 1. Database searching methods for structural annotation of product ion spectra.

|	Molecular formula search	Similar spectra search	Compound search
Query	m/z value and isotope abundance of precursor ion	Product ion spectrum	m/z value of precursor ion and product ion spectrum
Database or methodology	Compound database or 7 golden rules[2]	Measured mass spectra database	Compound database
Output	A list of molecular formula	Ranking of compounds based on spectra similarity score	Ranking of compounds based on coincidence score
Webtool or Software	ChemCalc[3]	MassBank[4]	MS-FINDER[5], SIRIUS[6,7], MetFrag[8,9], CFM-ID[10,11]
Method to estimate FDR	Model-based[17]	Target–decoy approach[18–20]	This study

FDR, false discovery rate.
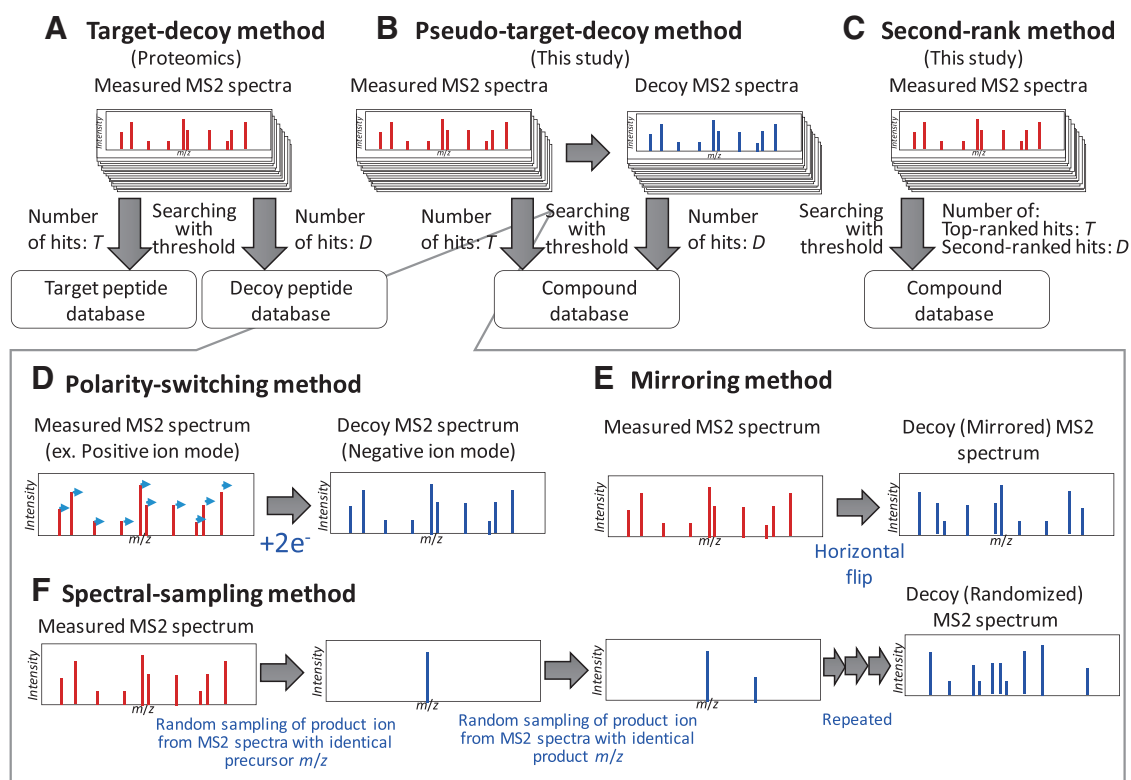
Fig. 1. Methods to estimate FDR in product ion spectra-based peptide and compound identification results. (A) Target–decoy method used for peptide identification. (B) Pseudo-target–decoy method for compound identification was developed in this study. (C) Second-rank method for compound identification was developed in this study. For generating decoy MS2 spectra, 3 methods to generate decoy MS2 spectra were considered, including (D) polarity-switching, (E) mirroring method, and (F) spectral-sampling methods. FDR, false discovery rate.

each product ion spectrum against the target peptide database[15] (Fig. 1A). The target database comprised all possible trypsin-digested peptides derived from all genome-encoded protein sequences. The peptides were identified from the top-ranked peptides in the search results. To control FDR, peptide identification is also performed using the decoy peptide database (Fig. 1A). The decoy database included the reverse sequences of all peptides in the target database. Because all hits against the decoy database were false positives, the FDR at a particular score-threshold level was estimated as FDRs = D/T or FDRc = $2D/(D + T)$, where T and D represent the number of hits against the target and decoy databases, respectively.[16] A suitable score threshold was employed to control the FDR levels, such as FDR = 0.05.

The target–decoy method is based on the availability of a complete list of possible trypsin-digested peptides and the generation of a decoy database with characteristics similar to those of the target database. However, no complete list of human metabolites is available. Moreover, no established method exists to create a decoy compound database for metabolomics. For molecular formula searches, methods employing theoretical models have been reported as alternatives to the target–decoy method.[17] Several methods for creating decoy spectral databases have been reported for spectral similarity searches, including fragmentation tree-based,[18] ion entropy and accurate entropy-based,[19] and violation of the octet rule of chemistry-based[20] methods (Table 1). A Gaussian mixture model-based framework for estimating FDR was also proposed for gas chromatography-electron impact-MS data.[21] For compound searches, SIRIUS 5 recently implemented a new scoring method called the confidence score, which proposes that hits with 0.64 or higher correspond to

an FDR of 0.1.[22)] Because the versatility of these methods remains unknown, various other methodologies must be investigated to estimate the FDR in compound search results.

In this study, 4 approximation methods were proposed to estimate the FDR in compound search results. Three of these methods are based on the generation of decoy query MS2 spectra (Fig. 1B). A method using second-rank scores was also investigated (Fig. 1C). First, the performances of the 4 methods were validated using the measured product ion spectra from MassBank. These 4 methods were applied to human metabolome DDA datasets for metabolite annotation. Based on the FDR-controlled metabolite annotation results, metabolites not included in the human metabolome database (HMDB) were investigated.

2. EXPERIMENTAL PROCEDURES

2.1. Software for compound search and data processing

SIRIUS 5 (version 5.8.5)[6,7)] was downloaded from https://bioinformatik.uni-jena.de/software/sirius/. MSFINDER (version 3.61)[5)] was downloaded from https://github.com/systemsomicslab/MsdialWorkbench/releases. MetFrag CL (version 2.4.5)[8,9)] was downloaded from https://ipb-halle.github.io/MetFrag. Compound search tasks were performed using the settings shown in Data S1. Three compound databases, including the yeast metabolome database (YMDB, 16,042 compounds, https://www.ymdb.ca/),[23)] HMDB 220,945 compounds, https://www.hmdb.ca),[24)] and Biodatabase were used. The Biodatabase is an edited database of biological compounds available at SIRIUS 5.[6,7)] Although the detailed number is unclear, the Biodatabase includes biological compounds derived from HMBD, PubChem, and other compound databases. This study used the Biodatabase as a larger database because the number of compounds appears to be considerably larger than that in the HMDB. All data pre- and postprocessing tasks were performed using an in-house Python script.

2.2. Performance test using MassBank records

All MassBank records were downloaded from the MassBank Consortium GitHub page (https://github.com/MassBank/; downloaded on 15/4/2023). First, 10,711 spectra were selected by the following criteria: AC$MASS_SPECTROMETRY: MS_TYPE = MS2, AC$INSTRUMENT_TYPE includes TOF or FT, AC$MASS_SPECTROMETRY: IONIZATION = ESI, AC$MASS_SPECTROMETRY: FRAGMENTATION_MODE = CID or no description, MS$FOCUSED_ION: PRECURSOR_TYPE = [M+H]+ or [M−H]−, CH$COMPOUND_CLASS does not include "Environmental Standard," "Surfactant," "Non-natural," and "Non-Natural," and MS$FOCUSED_ION: PRECURSOR_M/Z <850, and the number of product ions ≥3. The criteria were employed to use the default settings for the software packages. Second, the selected records are divided into 4 datasets. From the 574 compounds, several spectra were commonly acquired in both the positive and negative ion modes and designated as the CommonPos (3388 spectra) and CommonNeg (3100 spectra) datasets, respectively. The OthersPos (2648 spectra) and OthersNeg (1575 spectra) datasets comprised the remaining positive and negative ion mode data, respectively.

The 3 software packages used in this study output a ranked list of compound scores for each query product ion spectrum. In this study, the top-ranked compound on the list was referred to as a hit. Even in the case of multiple top-ranked compounds with the same score, they were considered hits. When one of the top-ranked compounds was correct, the hit was considered a true-positive hit. To calculate the FDR for the pseudo-target–decoy approach, the number of hits above a given threshold level was determined for the search results from the target (T) and decoy (D) datasets. Two methods were used to calculate FDRs: FDRs = D/T and FDRc = $2D/(D + T)$.[16)] In the second-rank method, the second-highest score in the ranked list of compound scores was used as the second-ranked score. The FDR was estimated using FDR2nd = D/T, where T and D represent the number of hits above a given threshold level for the first and second-ranked scores.

2.3. Generation of decoy spectra

A decoy spectrum was generated from each measured (target) spectrum in the DDA dataset to ensure that the total number of decoy spectra matched the number of target spectra. The product ion spectrum, S, was defined as a set of m/z and intensity values of the product ions.

$$S = \{(m_1, int_1), (m_2, int_2),\ldots, (m_n, int_n)\},$$

Here, m_i, int_i, and n represent the m/z and intensity values of i-th product ion and the total number of product ions, respectively. In this study, all the product ions were considered singly charged ions.

For the polarity-switching method (Fig. 1D), a decoy product ion spectrum, S', was generated from S by the following method:

$$S' = \{(m_1 + d, int_1), (m_2 + d, int_2), \ldots, (m_n + d, int_n)\},$$

Here, values of d are 0.001097 and −0.001097 for the positive and negative ion modes, respectively. The value of d corresponds to the mass of 2 electrons ($2e^-$). The polarity of S' was opposite from that of S.

For the mirroring method (Fig. 1E), a decoy product ion spectrum, S' was generated as follows:

$$S' = \{(m_{prec}-m_1 + p, int_1), (m_{prec}-m_2+ p, int_2), \ldots, \\ (m_{prec}-m_n + p, int_n)\}$$

Here, m_{prec} represents the m/z value of the precursor ion. The values of p (mass of proton) are 1.007276 and −1.007276 for the positive and negative ion modes, respectively. If $m_i \geq m_{prec}$, the value m_i was used instead of $m_{prec}-m_i + p$.

The spectral sampling was based on a previously reported method[18)] (Fig. 1F). A blank decoy product ion spectrum, $S' = \{\}$, was generated. First, the m/z value of the product ion (m'_1) was randomly sampled from a subset of product ion spectra with an identical precursor m/z value in the DDA dataset and added to S' to generate

$$S' = \{(m'_1, int_1)\}$$

Next, an m/z value of product ion (m'_2) was randomly sampled from a subset of product ion spectra sharing an m'_1 in the DDA dataset. The 2 m/z values, m_j and m_k, were

considered identical when the rounded integer values of $m_j * 200$ were the same as those of $m_k \times 200$. This procedure is repeated until the total number of product ions reaches n.

$$S' = \{(m'_1, int_1), (m'_2, int_2), \ldots, (m'_n, int_n)\}$$

All decoy spectra data were produced using in-house Python scripts.

2.4. Compound search and determination of FDR

The target and decoy queries of the CommonPos datasets, a sample Python script for compound searching using SIRIUS 5, and an Excel worksheet for FDR calculation are available from GitHub (https://github.com/fumiomatsuda/FDR-estimation-in-compound-search-results). Four human metabolome DDA datasets were downloaded from the Metabolomics Workbench[25] and MetaboLights repositories.[26,27] Averaged spectra were generated from the product ion spectra of each dataset using the spectral averaging method described in our previous study.[28] For each averaged product ion spectrum, averaged spectra of the corresponding MS1 spectra were generated using the same method. In addition to protonated or deprotonated molecules, additional ion forms were considered in the compound search if the corresponding ions were observed in the averaged MS1 spectra. The datasets of the averaged spectra will be made available to the public in the near future from https://github.com/Shin-MassBank/MassBank-Human.

3. RESULTS

3.1. Pseudo target–decoy approach using polarity-switching method

To facilitate the application of the target–decoy approach without creating a decoy compound database, an approach for generating decoy product ion spectra was investigated (Fig. 1B). In the proposed approach, a set of decoy product ion spectra is generated from the original product ion spectra in a DDA dataset (targets). The number of decoy spectra was identical to that of target spectra. Target and decoy product ion spectra were searched against an identical compound database to obtain target and decoy hits. FDRs were calculated from the number of target (T) and decoy (D) hits at a score threshold. However, the proposed approach is not ideal because the decoy product ion spectra do not exhibit properties identical to the target spectra. Hence, we term this the pseudo-target–decoy approach (Fig. 1B).

This study investigated 3 methods for generating decoy spectra from measured product ion spectra: polarity-switching, mirroring, and spectral sampling. First, the polarity-switching method was investigated (Fig. 1D). This method is based on that metabolomics employs both positive and negative ion modes for data acquisition. This indicated that a single-compound database can be used to search for both polarities. For instance, the original product ion spectra acquired in the positive ion mode (targets) can be searched against a compound database to produce target hits. The original product ion spectra were then regarded as decoy product ion spectra acquired in the negative ion mode by adding the mass of 2 electrons ($2e^-$) to all m/z values. Decoy hits were obtained by searching the decoy product ion

spectra against an identical compound database in negative ion mode (Fig. 1B).

To verify the performance of this method, this study utilized a dataset of measured product ion spectra stored in the MassBank database. From MassBank records, this study used product ion spectra (MS2) obtained from the $[M+H]^+$ and $[M-H]^-$ of natural products by collision-induced dissociation (CID) and high-resolution mass analyzers (the detailed procedure is presented in Experimental Procedures). Selected MassBank records were further divided into 4 datasets. The CommonPos (3388 spectra) and CommonNeg (3100 spectra) datasets include spectra commonly acquired from identical 574 compounds in both positive and negative ion modes. The remaining positive and negative ion mode data were designated as OthersPos (2648 spectra) and OthersNeg (1575 spectra) datasets, respectively.

First, the properties of the CommonPos and CommonNeg datasets were compared because these spectra were obtained from the same compounds. Comparison of the frequency distributions of m/z values, intensities, and numbers of product ions revealed that the CommonPos and CommonNeg datasets had similar properties, except for a slightly smaller number of product ions in the negative ion mode (data not shown). Next, compound search tasks were performed for the 3388 spectra in the CommonPos dataset (Data S2). The compound search using SIRIUS 5 with the CSI:FingerID scoring method successfully provided a hit for 2439 query spectra (the compound database was HMDB, and other search conditions used default values; Data S1). Each hit was checked against the correct answer and classified as a true- or false-positive hit. The true- and false-positive hits exhibited distinct score distributions (Fig. 2A), indicating that the score threshold for control FDR of 0.05 was −10.27, with 907 hits falling within this range. In this study, FDRmes represents the true FDR measured by true- and false-positive hits. Similar distributions were observed for the CommonNeg dataset (Fig. S1). Moreover, similar distributions were observed even when using a smaller compound database (YMDB) or a larger database (Biodatabase) (Fig. S1). These results indicated that there was a positive–negative symmetry in the compound search results using SIRIUS 5 with the CSI:FingerID scoring method. However, other software packages could not confirm positive–negative symmetry and distinct distributions between true and false hits (Fig. S2). This is probably because of the distinct scoring methods employed in these software packages. Therefore, subsequent analyses were performed using SIRIUS 5 with the CSI:FingerID scoring method.

To test the pseudo-target–decoy approach using the polarity-switching method, the mass of the 2 electrons was added to all m/z values of the 3388 spectra in the CommonPos dataset to produce a decoy spectral dataset (Fig. 1D). The compound search in the negative ion mode using SIRIUS 5 with the CSI:FingerID scoring method produced 2105 hits (Data S2), with the frequency distribution shown in Fig. 2B. Using the score distributions, 2 FDR indices, FDRs = D/T and FDRc $2D/(T + D)$, were calculated and compared with the FDRmes (Fig. 2C and 2D), where T and D indicate the number of hits for the target and decoy datasets, respectively. Figure 2C shows the relationship between the threshold levels of CSI:FingerID scores and the FDRmes, FDRs, and FDRc levels. The comparison showed that the estimations by FDRc were closer to the true FDR (FDRmes) than those by FDRs (Fig. 2C).

Query: CommonPos, Decoy: Polarity switching: Scoring: SIRIUS 5 CSI, Database: HMDB
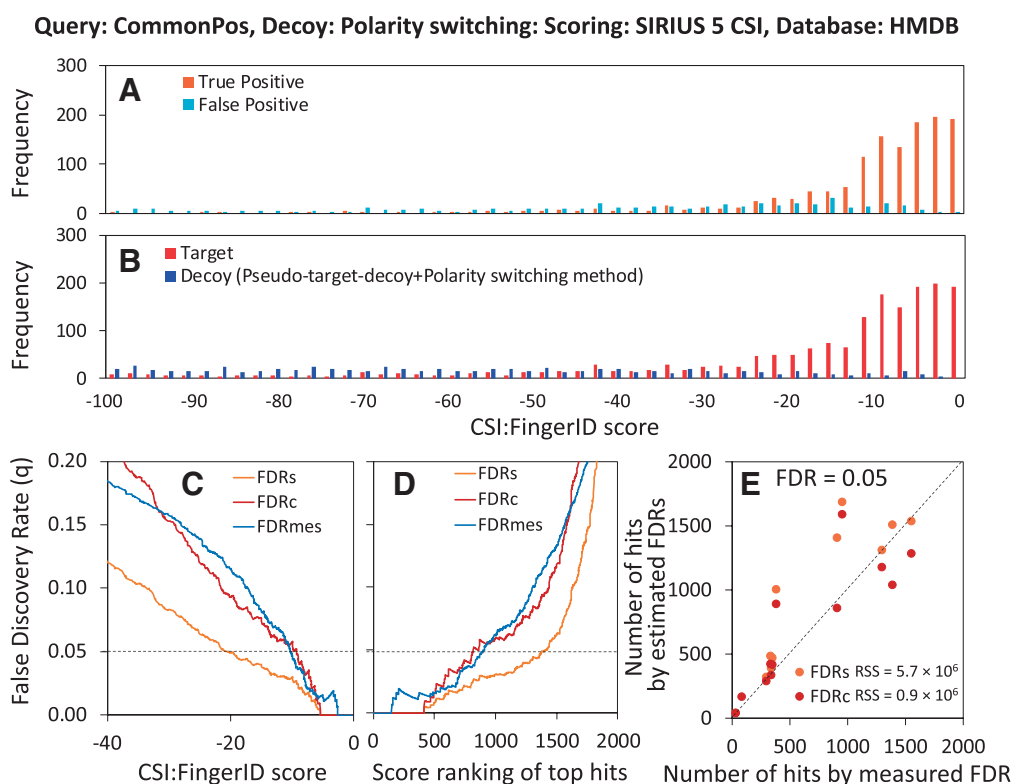
Fig. 2. Performance evaluation of the pseudo-target–decoy approach using the polarity-switching method. (A) Score distribution of true- and false-positive hits in the compound identification result. Total of 3388 high-resolution mass spectra data obtained at positive ion mode were collected from MassBank. The CommonPos dataset was served for the compound search by the SIRIUS 5 CSI:Finger ID scoring method using HMDB as the compound database. (B) Score distribution of searching results of target and decoy CommonPos datasets. (C) Relationship between the CSI:Finger ID score and FDR levels. FDRmes indicate true FDR levels measured from the number of true and false hits in panel A. FDRs and FDRs represent estimated FDR levels from the number of target (T) and decoy (D) hits in panel B. FDRs = D/T, FDRc = $2D/(T + D)$. (D) Relationship between the score ranking of top hits and FDR levels. (E) Comparison between the number of hits determined by the true FDR (FDRmes) and estimated FDR (FDRs and FDRc). The number of hit compounds at FDR = 0.05 was compared among all 12 combinations of the 4 datasets (CommonPos, CommonNeg, OhtersPos, and OthersNeg) and 3 compound databases (YMDB, HMDB, and BioDB). The RSS levels between the true FDR (FDRmes) and estimated FDR (FDRs and FDRc) are also shown. FDR, false discovery rate; RSS, residual sum of square.

Figure 2D also shows the relationship between the FDRs and the score ranking of the top hits. For example, the true number of hit compounds for FDRmes = 0.05 was 907. By contrast, the numbers estimated by FDRs and FDRc were 1408 and 860, respectively, indicating that the estimation by FDRc was better than that of FDRs (Fig. 2D). A similar trend was observed in the CommonNeg dataset (Fig. S3). Using the 4 datasets (CommonPos, CommonNeg, OhtersPos, and OthersNeg) and 3 compound databases of different sizes (YMDB, HMDB, and BioDB), the numbers of hit compounds at FDR = 0.05 were estimated for all 12 combinations (Fig. 2E, Table S1). The numbers of hit compounds for all 12 combinations were compared to that of the true FDR (FDRmes) by calculating the residual sum of squares (RSS). The RSS levels of FDRs and FDRc were 5.73×10^6 and 0.90×10^6, confirming better performance of FDRc. Moreover, the results showed that FDRc and FDRs estimated using the polarity-switching method often underestimated the FDR in certain cases (Fig. 2E, Table S1).

3.2. Pseudo-target–decoy approach using mirroring and spectral-sampling methods

Next, mirroring and spectral sampling methods were used to generate decoy spectra. In the mirroring method, a decoy spectrum is generated by horizontally flipping the target spectrum (Fig. 1E). Spectral sampling was used to generate decoy spectral databases for spectral similarity search.[18] The product ions were randomly sampled from a set of target spectra sharing identical precursors and product ions (Fig. 1F; See Experimental Procedures for details).

Using the decoy spectra generated by the mirroring and spectra-sampling methods, FDRs and FDRc levels were determined for the CommonPos dataset using the SIRIUS 5 CSI:FingerID scoring method (Fig. 3, Table S1). For the mirroring method, the results showed that the estimation by FDRc was a better approximation at FDR = 0.05 than that of FDRs (Fig. 3A and 3B). For instance, the number of hits estimated by FDRc at FDR = 0.05 was 860, similar to the number of hits (907) determined by the true FDR (FDRmes). However, the mirroring method significantly overestimated the FDR levels in the low-score region (Fig. 3A). The numbers of hit compounds at FDR = 0.05 were compared among all 12 combinations (Fig. 3C, Table S1). The results showed that FDRc always produced a more conservative FDR estimation than the FDRs. Moreover, the estimations by FDRc at FDR = 0.05 were similar to those of FDRmes among all 12 combinations, indicating that the mirroring method generated reasonable decoy spectra
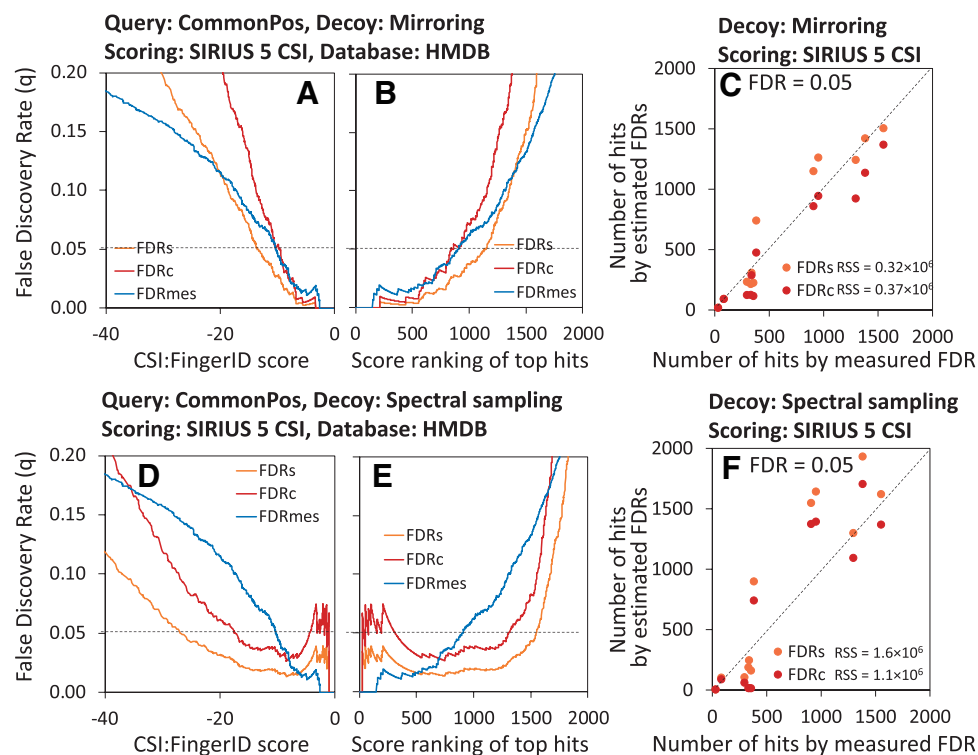
Fig. 3.   Performance evaluation of the pseudo-target–decoy approach using the mirroring (A–C) and the spectral sampling (D–F) methods. (A, D) Relationship between the CSI:Finger ID score and FDR levels. FDRmes indicate true FDR levels measured from the number of true and false hits. FDRs and FDRc represent estimated FDR levels from the number of target (T) and decoy (D) hits. FDRs = D/T, FDRc = $2D/(T + D)$ (B, E) Relationship between the score ranking of top hits and FDR levels. (C, F) Comparison between the number of hits determined by the true FDR (FDRmes) and estimated FDR (FDRs and FDRc). The number of hit compounds at FDR = 0.05 was compared among all 12 combinations of the 4 datasets (CommonPos, CommonNeg, OhtersPos, and OthersNeg) and 3 compound databases (YMDB, HMDB, and BioDB). The RSS levels between the true FDR (FDRmes) and estimated FDR (FDRs and FDRc) are also shown. FDR, false discovery rate; RSS, residual sum of square.

(Fig. 3C, Table S1). Indeed, the RSS levels of FDRs and FDRc were 0.32×10^6 and 0.37×10^6, which were smaller than that of the polarity-switching method (Fig. 3C, Table S1).

The performance of the spectral sampling method was evaluated using the same procedure. The RSS levels of FDRs and FDRc were 1.56×10^6 and 1.08×10^6 (Fig. 3D–3F, Table S1). The results show that the FDR estimation capability of the spectral sampling method is lower than that of the mirroring method. It has been reported that the spectral sampling method is useful in constructing a decoy database for spectral similarity searches.[18] However, this study demonstrated that the decoy spectra generated by the spectra-sampling method are unsuitable for FDR estimation in a compound search. Similar overestimations and underestimations were observed for the SIRIUS 5 with the confidence score method (Table S1).

3.3.   Second-rank method

Next, another approach was examined. As previously mentioned, the top-ranked hits included both true and false positives. Each compound search can also provide a second-ranked hit in addition to the top-ranked hit. Certain second-ranked hits should have high scores that are very close to the top-ranked hits and can be considered as failed attempts of false-positive hits. Thus, the distribution of high scores in the second-ranked hits should be similar to that of false positives in the top-ranked hits (Fig. 1C).

To test the second-rank method, the scores of second-ranked hits were obtained from the search results of the CommonPos dataset (Fig. 4A, Data S3). Treating these scores as decoys, score thresholds and numbers of hits were determined by FDRrank2 = D/T and compared with those from FDRmes (Fig. 4B and 4C). The results showed that the estimation using FDRrank2 deviated significantly from that using FDRmes, particularly in the low-score region. However, FDRrank2 provides a better approximation at FDR = 0.05. For instance, the number of hits estimated by FDRrank2 was 1067, which approximates the 907 hits obtained by FDRmes. By contrast, FDRrank3, calculated using third-ranked data, significantly underestimated the FDR. Similar trends were observed for the CommonNeg dataset (Fig. S4). A comparison across all 12 combinations of the 4 datasets and 3 compound databases at FDR = 0.05 showed that the RSS levels of FDRrank2 and FDRrank3 were 0.2×10^6 and 1.1×10^6 (Fig. 4D, Table S1). The results revealed that the estimation by FDRrank2 using the second-rank method was better than using the polarity-switching, mirroring, and spectral sampling methods (Table S1).

3.4.   Application to human metabolome DDA datasets

Previously, DDA metabolomic datasets have been shown to contain similar product ion spectra redundantly obtained from the same compound.[29,30] Integrating similar spectra into an averaged spectrum can improve the signal-to-noise ratio.[28]

Query: CommonPos, Decoy: Second-rank, Scoring: SIRIUS 5 CSI:FingerID, Database: HMDB
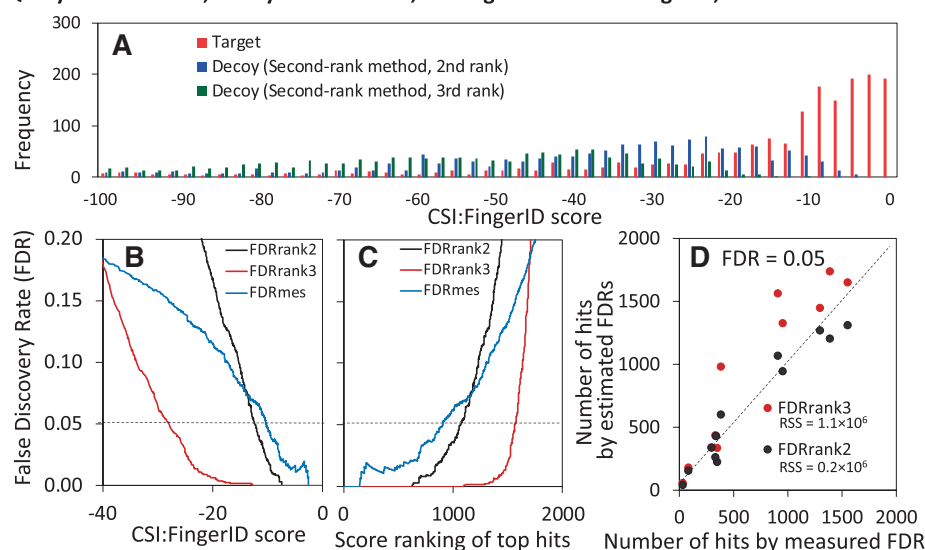

Fig. 4. Performance evaluation of the second-rank method. (A) Distribution of the top- (Target), second-, and third-ranked scores in the compound identification result. A total of 3388 high-resolution mass spectra data obtained at positive ion mode were collected from MassBank. The CommonPos dataset was served for the compound search by the SIRIUS 5 CSI:Finger ID scoring method using HMDB as the compound database. (B) Relationship between the CSI:Finger ID score and FDR levels. FDRmes indicate true FDR levels measured from the numbers of true and false hits. FDRrank2 and FDRrank3 represent the estimated FDR (D/T) levels from the numbers of target (T) and decoy (D) hits. (C) Relationship between the score ranking of top hits and FDR levels. (D) Comparison between the number of hits determined by the true FDR (FDRmes) and estimated FDR (FDRrank2 and FDRrank3). The RSS levels between the true FDR (FDRmes) and estimated FDR (FDRs and FDRc) are also shown. FDR, false discovery rate; HMDB, human metabolome database; RSS, residual sum of square.

Table 2. FDR-based compound searching of human metabolome DDA datasets.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Repository	Metabolomics Workbench	Metabolomics Workbench	Metabolomics Workbench	MetaboLights
ID	ST001171	ST002338	ST002338	MTBLS417
Title	Metabolomics of World Trade Center Exposed New York City Firefighters[31]	Interplay Between Cruciferous Vegetables and the Gut Microbiome: A Multi-Omic Approach[30]	Interplay Between Cruciferous Vegetables and the Gut Microbiome: A Multi-Omic Approach[30]	Customized Consensus Spectral Library Building for Untargeted Quantitative Metabolomics Analysis with Data Independent Acquisition Mass Spectrometry and MetaboDIA Workflow[32]
Author	Anna Nolan, Laboratory at NYU/Bellevue	Bouranis John, Oregon State University	Bouranis John, Oregon State University	Zhou Lei, Hyungwon Choi, National University of Singapore
Sample	Human, serum	Human, Feces	Human, Feces	Human, serum
Mass Spectrometer	Q Exactive HF-X Orbitrap, Thermo Scientific	TripleTOF 5600, AB Sciex	TripleTOF 5600, AB Sciex	TripleTOF 5600+, AB Sciex
HPLC	Reverse phase ODS	Reverse phase ODS	Reverse phase ODS	Reverse phase ODS
Polarity	Positive	Positive	Negative	Positive
Number of data files	248	40	40	60
The number of raw spectra in the dataset	3,242,674	256,580	220,008	548,663
The number of averaged spectra (with more than 3 fragment ion signal)	21,403 (8104)	3062 (1789)	2317 (1170)	3077 (1450)
The number of hits at FDR = 5%				
Polarity-switching (FDRc)	709	247	57	166
Mirroring (FDRc)	512	250	12	82
Spectral-sampling (FDRc)	355	186	11	42
Second-rank (FDRrank2)	642	182	41	100
The number of hits without HMDB ID in the second-rank results	56	24	1	12

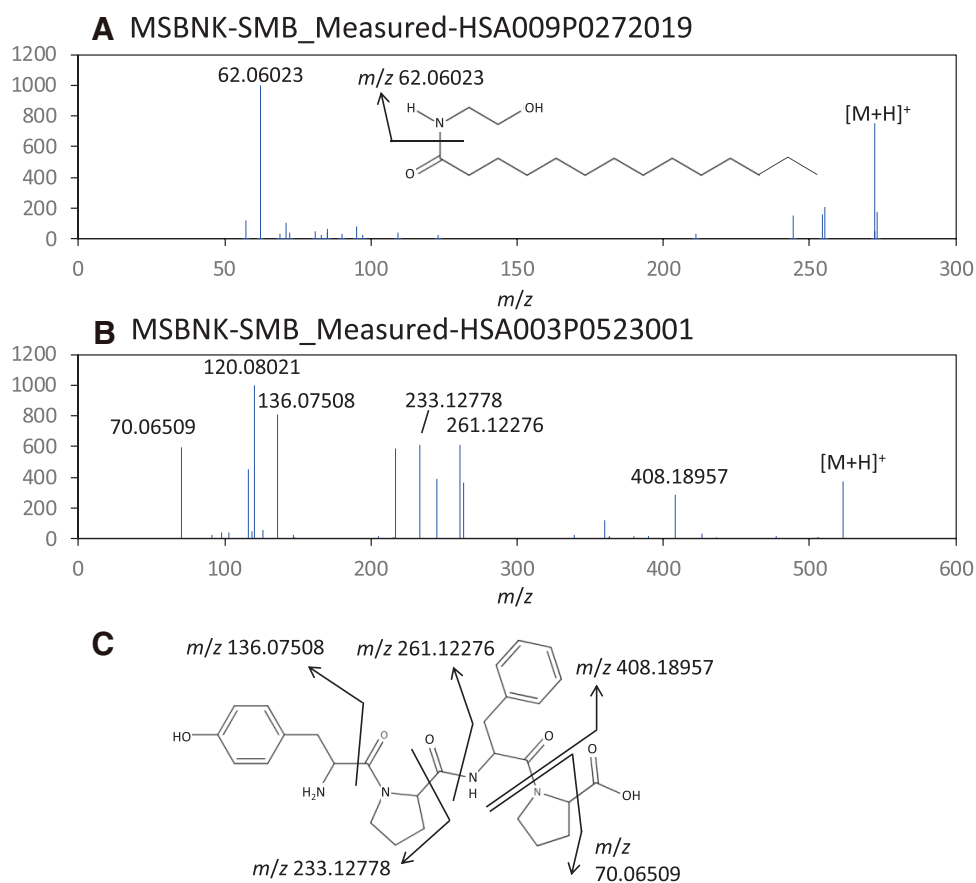FDR, false discovery rate; HMDB, human metabolome database.

Fig. 5. Two annotation examples with compounds not included in the HMDB. (A) Product ion spectrum derived from Dataset 1. The spectrum was annotated as *N*-myristoylethanolamide by the SIRIUS 5 CSI:Finger ID scoring method using Biodatabase as the compound database. (B) Product ion spectrum derived from Dataset 2, annotated as β-casomorphin 4. (C) Structure of β-casomorphin 4 and estimated assignment of fragment ions. HMDB, human metabolome database

Here, 4 human metabolomic DDA datasets were obtained from the public repositories.[30–32] Averaged spectral sets were produced from these datasets and used for compound searches using the SIRIUS 5 CSI:FingerID scoring method in the Biodatabase. The 4 methods developed in this study were used to control the FDR at 0.05. The results are summarized in Table 2. For Dataset 1, 3,242,674 product ion spectra across 248 data files were consolidated into 21,403 averaged spectra by spectral averaging. Among them, 8104 spectra included 3 or more product ions and were used for the compound search. The number of hits at FDR=0.05 was estimated to be 709, 512, 355, and 642 using the polarity-switching, mirroring, spectral sampling, and second-rank methods, respectively. Although the number of hits varied widely among the 4 methods, the results of the second-rank method were in the middle range. Similar patterns were observed for Datasets 2, 3, and 4 (Table 2).

A purpose of untargeted human metabolome analysis is to identify novel human metabolites. Thus, all hits were checked using the HMDB[24] to identify any novel human metabolites in the FDR-controlled compound search results. The HMDB contains 220,945 known small-molecule metabolites found in the human body. Among the 642 hits in Dataset 1 by the second-rank method, 56 hits did not have HMDB identifiers (Data S4). For example, the top hit for the spectrum shown in Fig. 5A

was *N*-myristoylethanolamide, whose CSI:FingerID score (0.18) was the highest among the 56 non-HMDB hits (Data S5). The compound search results were confirmed using the measured spectrum of *N*-myristoylethanolamide from MassBank (MSBNK-BGC_Munich-RP003002; similarity score = 0.7437; data not shown). *N*-myristoylethanolamide is a lipid mediator in mammals, indicating that its detection in human serum samples is plausible.[33,34] Furthermore, in Dataset 2, the spectrum shown in Fig. 5B hit to β-casomorphin 4 with the highest CSI:FingerID score (−1.01) among the 24 non-HMDB hits (Data S6). Although MassBank does not contain any measured spectra of this compound (Data S5), the product ion spectrum was consistent with the possible fragmentation of β-casomorphin 4 by manual curation. β-Casomorphin 4 is a tetrapeptide (Tyrosyl-prolyl-phenylalanyl-proline, Fig. 5C) and a degradation product of the milk protein β-casein, likely to be detected in human fecal samples.[35]

4. DISCUSSION

This study examined 4 methods for controlling the FDR in the compound search results of untargeted metabolomics: polarity-switching, mirroring, spectral sampling, and second-rank methods (Fig. 1). It is important to note that

the calculated FDRs are approximate estimates because these methods rely on invalid heuristics or assumptions. Estimations at FDR = 0.01 or 0.10 were particularly prone to large errors (Figs. 2–4). Among the 4 methods developed, the second-rank method provided the best FDR estimation in the performance test using the 4 spectral datasets from MassBank (Fig. 4D and Table S1). Furthermore, compound search results of 4 human metabolome DDA datasets with FDR = 0.05 showed that the number of hits determined by the second-rank method was neither extremely large nor small, suggesting that it avoided significant over- or under-estimation (Table 2). The FDR-controlled compound search results identified compounds not present in the HMDB, such as N-myristoylethanolamide and β-casomorphin 4 (Fig. 5).

The 4 methods examined in this study offer the advantage of simplicity and do not require any modifications to the structural elucidation software. Moreover, the polarity switching, mirroring, and second-rank methods do not require random sampling techniques. However, the FDR estimates provided by these methods are rough approximations of the true FDR. To improve the estimation, the application of FDR estimation methods for spectral similarity searches, such as the fragmentation tree-based construction of decoy spectra,[18] is promising for the construction of decoy query spectra datasets.

Moreover, the versatility of these methods requires further investigation. Software packages and scoring methods are being developed rapidly. Recently, SIRIUS 6, with an updated scoring method, was made available to the public from the developer's webpage (https://bioinformatik.uni-jena.de/software/sirius/). Furthermore, this study used the product ion spectral data stored in MassBank for method verification owing to their structural variety. Additional verification is required to handle product-ion spectral datasets with less structural variation, such as those in lipidomics. In the future, more accurate FDR estimation is expected to be achieved by developing methods based on more valid assumptions or methods for constructing valid decoy compound databases.

ACKNOWLEDGMENTS

SUPPLEMENTARY AND SPECTRUM DATA

Table S1. Number of hits in the compound searching results by SIRIUS 5 CSI:Finger ID scoring method when false discovery rate (FDR) is 0.05

Figure S1. Score distribution of true positive and false positive hits in the compound identification result.

Figure S2. Score distribution of true positive and false positive hits in the compound identification result using other compound search methods.

Figure S3. Performance evaluation of the pseudo-target-decoy approach using the polarity-switching method.

Figure S4. Performance evaluation of the second-rank method.

Data S1. Parameters used for MetFrag, MSFINDER, and SIRIUS5.

Data S2. Compound search results of the CommonPos dataset using SIRIUS 5 CSI:FingerID scoring method with FDR estimation by a pseudo-target-decoy approach using the polarity-switching method.

Data S3. Compound search results of the CommonPos dataset using SIRIUS 5 CSI:FingerID scoring method with FDR estimation by the second-rank method.

Data S4. Compound search result of Dataset1 using SIRIUS 5 CSI:FingerID scoring method with FDR estimation using the second-rank method.

Data S5. Query spectra data for SIRIUS5.

Data S6. Compound search result of Dataset2 using SIRIUS 5 CSI:FingerID scoring method with FDR estimation using the second-rank method.

REFERENCES

1) F. Matsuda. Rethinking mass spectrometry-based small molecule identification strategies in metabolomics. *Mass Spectrom. (Tokyo)* 3: S0038, 2014.

2) T. Kind, O. Fiehn. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 8: 105, 2007.

3) L. Patiny, A. Borel. ChemCalc: A building block for tomorrow's chemical infrastructure. *J. Chem. Inf. Model.* 53: 1223–1228, 2013.

4) H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, Y. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka. MassBank: A public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45: 703–714, 2010.

5) H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, M. Arita. Hydrogen rearrangement rules: Computational MS/MS fragmentation and structure elucidation using MS-FINDER software. *Anal. Chem.* 88: 7946–7958, 2016.

6) K. Dührkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu, S. Böcker. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16: 299–302, 2019.

7) M. Ludwig, M. Fleischauer, K. Duhrkop, M. A. Hoffmann, S. Bocker. De novo molecular formula annotation and structure elucidation using SIRIUS 4. *Methods Mol. Biol.* 2104: 185–207, 2020.

8) C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, S. Neumann. MetFrag relaunched: Incorporating strategies beyond *in silico* fragmentation. *J. Cheminform.* 8: 3, 2016.

9) C. Ruttkies, S. Neumann, S. Posch. Improving MetFrag with statistical learning of fragment annotations. *BMC Bioinformatics* 20: 376, 2019.

10) A. Chao, H. Al-Ghoul, A. D. McEachran, I. Balabin, T. Transue, T. Cathey, J. N. Grossman, R. R. Singh, E. M. Ulrich, A. J. Williams, J. R. Sobus. In silico MS/MS spectra for identifying unknowns: A critical examination using CFM-ID algorithms and ENTACT mixture samples. *Anal. Bioanal. Chem.* 412: 1303–1315, 2020.

11) F. Wang, J. Liigand, S. Tian, D. Arndt, R. Greiner, D. S. Wishart. CFM-ID 4.0: More accurate ESI-MS/MS spectral prediction and compound identification. *Anal. Chem.* 93: 11692–11700, 2021.

12) Y. Benjamini, Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57: 289–300, 1995.

13) H. Choi, A. I. Nesvizhskii. False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J. Proteome Res.* 7: 47–50, 2008.

14) D. L. Tabb. What's driving false discovery rates? *J. Proteome Res.* 7: 45–46, 2008.

15) J. E. Elias, S. P. Gygi. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4: 207–214, 2007.

16) S. Aggarwal, A. K. Yadav. False discovery rate estimation in proteomics. *Methods Mol. Biol.* 1362: 119–128, 2016.

17) F. Matsuda, Y. Shinbo, A. Oikawa, M. Y. Hirai, O. Fiehn, S. Kanaya, K. Saito. Assessment of metabolome annotation quality: A method for evaluating the false discovery rate of elemental composition searches. *PLoS One* 4: e7490, 2009.

18) K. Scheubert, F. Hufsky, D. Petras, M. Wang, L. F. Nothias, K. Duhrkop, N. Bandeira, P. C. Dorrestein, S. Bocker. Significance estimation for large scale metabolomics annotations by spectral matching. *Nat. Commun.* 8: 1494, 2017.

19) S. An, M. Lu, R. Wang, J. Wang, H. Jiang, C. Xie, J. Tong, C. Yu. Ion entropy and accurate entropy-based FDR estimation in metabolomics. *Brief. Bioinform.* 25: bbae056, 2024.

20) X. Wang, D. R. Jones, T. I. Shaw, J. H. Cho, Y. Wang, H. Tan, B. Xie, S. Zhou, Y. Li, J. Peng. Target-decoy-based false discovery rate estimation for large-scale metabolite identification. *J. Proteome Res.* 17: 2328–2334, 2018.

21) J. E. Flores, L. M. Bramer, D. J. Degnan, V. L. Paurus, Y. E. Corilo, C. S. Clendinen. Gaussian mixture modeling extensions for improved false discovery rate estimation in GC-MS metabolomics. *J. Am. Soc. Mass Spectrom.* 34: 1096–1104, 2023.

22) M. A. Hoffmann, L. F. Nothias, M. Ludwig, M. Fleischauer, E. C. Gentry, M. Witting, P. C. Dorrestein, K. Duhrkop, S. Bocker. High-confidence structural annotation of metabolites absent from spectral libraries. *Nat. Biotechnol.* 40: 411–421, 2022.

23) M. Ramirez-Gaona, A. Marcu, A. Pon, A. C. Guo, T. Sajed, N. A. Wishart, N. Karu, Y. Djoumbou Feunang, D. Arndt, D. S. Wishart. YMDB 2.0: A significantly expanded version of the yeast metabolome database. *Nucleic Acids Res.* 45(D1): D440–D445, 2017.

24) D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M. A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. Macinnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel, L. Querengesser. HMDB: The human metabolome database. *Nucleic Acids Res.* 35(Database): D521–D526, 2007.

25) M. Sud, E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, K. S. Nair, S. Sumner, S. Subramaniam. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 44(D1): D463–D470, 2016.

26) N. S. Kale, K. Haug, P. Conesa, K. Jayseelan, P. Moreno, P. Rocca-Serra, V. C. Nainala, R. A. Spicer, M. Williams, X. Li, R. M. Salek, J. L. Griffin, C. Steinbeck. MetaboLights: An open-access database repository for metabolomics data. *Curr. Protoc. Bioinformatics* 53: 14.13.1–14.13.18, 2016.

27) K. Haug, K. Cochrane, V. C. Nainala, M. Williams, J. Chang, K. V. Jayaseelan, C. O'Donovan. MetaboLights: A resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* 48(D1): D440–D444, 2020.

28) F. Matsuda, S. Komori, Y. Yamada, D. Hara, N. Okahashi. Data processing of product ion spectra: Quality improvement by averaging multiple similar spectra of small molecules. *Mass Spectrom. (Tokyo)* 11: A0106, 2022.

29) F. Matsuda. Data processing of product ion spectra: Redundancy of product ion spectra of small molecules in data-dependent acquisition dataset. *Mass Spectrom. (Tokyo)* 12: A0138, 2023.

30) J. A. Bouranis, L. M. Beaver, D. Jiang, J. Choi, C. P. Wong, E. W. Davis, D. E. Williams, T. J. Sharpton, J. F. Stevens, E. Ho. Interplay between cruciferous vegetables and the gut microbiome: A multi-omic approach. *Nutrients* 15: 42, 2022.

31) G. Crowley, J. Kim, S. Kwon, R. Lam, D. J. Prezant, M. Liu, A. Nolan. PEDF, a pleiotropic WTC-LI biomarker: Machine learning biomarker identification and validation. *PLOS Comput. Biol.* 17: e1009144, 2021.

32) G. Chen, S. Walmsley, G. C. M. Cheung, L. Chen, C. Y. Cheng, R. W. Beuerman, T. Y. Wong, L. Zhou, H. Choi. Customized consensus spectral library building for untargeted quantitative metabolomics analysis with data independent acquisition mass spectrometry and MetaboDIA workflow. *Anal. Chem.* 89: 4897–4906, 2017.

33) J. Keereetaweep, A. Kilaru, I. Feussner, B. J. Venables, K. D. Chapman. Lauroylethanolamide is a potent competitive inhibitor of lipoxygenase activity. *FEBS Lett.* 584: 3215–3222, 2010.

34) P. Garg, R. S. Duncan, S. Kaja, A. Zabaneh, K. D. Chapman, P. Koulen. Lauroylethanolamide and linoleoylethanolamide improve functional outcome in a rodent model for stroke. *Neurosci. Lett.* 492: 134–138, 2011.

35) D. D. Nguyen, S. K. Johnson, F. Busetti, V. A. Solah. Formation and degradation of beta-casomorphins in dairy processing. *Crit. Rev. Food Sci. Nutr.* 55: 1955–1967, 2015.