analytical chemistry

pubs.acs.org/ac

Article

Correlation-Based Deconvolution (CorrDec) To Generate High-Quality MS2 Spectra from Data-Independent Acquisition in Multisample Studies

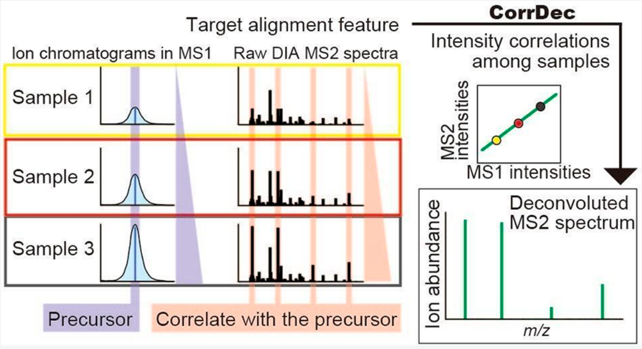
Ipputa Tada,[§] Romanas Chaleckis,[§] Hiroshi Tsugawa, Isabel Meister, Pei Zhang, Nikolaos Lazarinis, Barbro Dahlén, Craig E. Wheelock,* and Masanori Arita*

Cite This: *Anal. Chem.* 2020, 92, 11310−11317

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Data-independent acquisition mass spectrometry (DIA-MS) is essential for information-rich spectral annotations in untargeted metabolomics. However, the acquired MS2 spectra are highly complex, posing significant annotation challenges. We have developed a correlation-based deconvolution (CorrDec) method that uses ion abundance correlations in multisample studies using DIA-MS as an update of our MS-DIAL software. CorrDec is based on the assumption that peak intensities of precursor and fragment ions correlate across samples and exploits this quantitative information to deconvolute complex DIA spectra. CorrDec clearly improved deconvolution of the original MS-DIAL deconvolution method (MS2Dec) in a dilution series of chemical standards and a 224-sample urinary metabolomics study. The primary advantage of CorrDec over MS2Dec is the ability to discriminate coeluting low-abundance compounds. CorrDec requires the measurement of multiple samples to successfully deconvolute DIA spectra; however, our randomized assessment demonstrated that CorrDec can contribute to studies with as few as 10 unique samples. The presented methodology improves compound annotation and identification in multisample studies and will be useful for applications in large cohort studies.

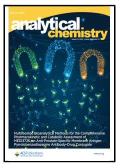


For compound identification, high-resolution tandem mass spectra (MS2) with public spectral library and associated computational tools are indispensable. A number of resources are available including MassBank,[1] GNPS,[2] CSI:FingerID,[3] and MS-FINDER.[4] In the classical data-dependent acquisition mass spectrometry (DDA-MS), ions are isolated in a narrow *m*/*z* window to obtain clean spectra (typically 1 Da, sometimes up to 9 Da).[5,6] In contrast, for data-independent acquisition mass spectrometry (DIA-MS), wider *m*/*z* windows of 10−1000 Da are used to obtain complex spectra from coeluting precursors, thereby requiring computational approaches to interpret.[7]

To overcome the trade-off between cleanness and comprehensiveness of DIA spectra, various deconvolution tools have been proposed, such as OpenSWATH,[8] Specter,[9] MetDIA,[10] MS-DIAL,[11] and decoMetDIA.[12] The first three tools were designed for targeted analyses that utilize predefined spectral libraries to deconvolute spectra. The latter two can deconvolute MS2 spectra de novo by fitting MS2 chromatograms to their precursor chromatogram in a single sample (i.e., using retention time). These powerful methods are suitable for the SWATH (Sequential Window Acquisition of all THeoretical fragment ion spectra) type of DIA data.[7] However, MS2 spectra become highly complex when precursor ions of all *m*/*z* are fragmented together: e.g., all ion fragmentation (AIF), MS^{ALL}, or MS^{E}.[7] In particular, busy chromatographic regions with multiple coeluting compounds pose a significant challenge. In the case of the original MS-DIAL, at least two data-point differences between the chromatographic peak tops is required for deconvolution, which is a challenging condition for AIF data. Previous tools are therefore not suitable for untangling complex MS2 spectra from the AIF acquisition and its equivalent.

We present a new MS2 deconvolution method based on the correlation of ion abundances between precursor and product ions among biological samples, named CorrDec (Correlation-based Deconvolution). This method, implemented in MS-DIAL version 3.22 and later, is designed to deconvolute MS2 spectra from untargeted, multisample AIF metabolomics without requiring a predefined spectral library. The method

Received: May 8, 2020
Accepted: July 10, 2020
Published: July 10, 2020



is based on three assumptions: (1) metabolite concentrations differ across study samples in multisample studies; (2) the MS2 fragmentation pattern is identical under identical experimental conditions; (3) intensities of fragment ions correlate with those of their precursors.

Correlation has been widely used in mass spectrometry-based metabolomics.[13,14] For example, the Pearson correlation is used in CAMERA to estimate the similarity of different mass chromatograms to extract compound spectra and to annotate adduct ions and isotopic peaks.[15] For DIA, data correlation-based approaches such as RAMClust assigns precursor-product relationships based on detected features in MS1 and MS2.[16,17] In contrast to the previous approaches, CorrDec is not designed to retrieve as many characteristic product ions as possible from the DIA-MS2 spectra. Rather, it excludes noise peaks effectively by integrating multisample profiles. We demonstrate the concept and utility of CorrDec in a dilution series of chemical standards in urine and a case study from a urinary metabolomics cohort.

■ EXPERIMENTAL SECTION

Correlation-Based Deconvolution. CorrDec starts with the aligned peak list from multiple samples. The peak list consists of "aligned features", which include the averages of retention time, m/z, peak height, and width obtained from the detected peaks in the samples, their ion abundances, and corresponding MS2 spectra. The peak height is used for the quantification of MS1 and MS2 peaks. The MS2 deconvolution is performed as follows (Figure 1).

Step 1: For each aligned feature Ft1, Pearson correlations are calculated between all product ions and their precursors. The MS2 spectra of Ft1 for all samples are retrieved to create a "MS2Mat" data matrix, consisting of the ion abundances of each product ion (P) binned by an m/z threshold in all samples (0.01 in this study). The precursor ion abundances of all samples are retrieved to create a "MS1Vec" data vector, and Pearson correlations are calculated for all pairs of the features in MS1Vec and product ions in MS2Mat (Figure 1A). For each product ion, its existence ratio within the samples (the number of samples having the product ion above the threshold (1000 in this study) divided by the number of all samples) is also recorded.

Step 2: All correlation values in all features are integrated into a matrix based on the m/z of the product ion using the same m/z threshold (0.01 in this study) as MS2Mat (Figure 1B).

Step 3: Each product ion is assessed using the correlation value $\mathrm{Corr}_{MS1vsMS2}$ for its inclusion to the deconvoluted spectrum of Ft1. Three criteria are applied (Figure 1C):

(Criterion 1) $\mathrm{Corr}_{\mathrm{MS1vsMS2}}$ > minimum threshold,
(Criterion 2) $\mathrm{Corr}_{\mathrm{MS1vsMS2}}$> $\mathrm{MaxCorr}_{\mathrm{Ft}}$-margin1, and
(Criterion 3) $\mathrm{Corr}_{\mathrm{MS1vsMS2}}$> $\mathrm{MaxCorr}_{\mathrm{P}}$-margin2.

Criterion 1 is an overall cutoff to suppress noise signals. Correlations between the ion abundances of a MS1 precursor ion and the ion abundances of a MS2 product ion must be higher than a predefined minimum correlation threshold for all peaks. The recommended threshold is between 0.3 and 0.7, and we used 0.7 in this study. A lower threshold indicates a higher possibility to introduce noise peaks into spectra.

Criterion 2 is a threshold to filter product ions for fragments from each MS1 feature Ft1. $\mathrm{MaxCorr}_{\mathrm{Ft}}$ is the maximum of all correlations for Ft1, and relatively low-correlating peaks from ionization enhancement and/or biochemical proximation are removed. The recommended margin1 is between 0.1 and 0.3, and we used 0.2 in this study. A larger margin indicates a higher possibility to introduce noise peaks into spectra. For example, in Figure 1B, the MS2 peak P1 (0.73) is removed because $\mathrm{MaxCorr}_{\mathrm{Ft}}$ for the feature is 0.98.

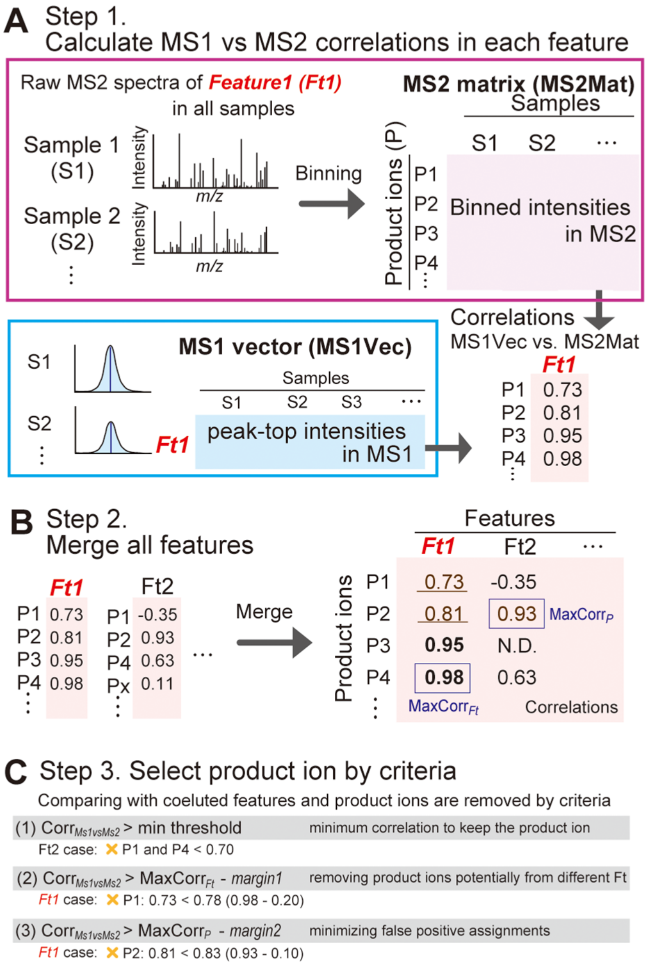

A Step 1.
Calculate MS1 vs MS2 correlations in each feature
Raw MS2 spectra of Feature1 (Ft1) in all samples
Sample 1 (S1)
Sample 2 (S2)
Intensity
m/z
Binning
MS2 matrix (MS2Mat)
Samples
S1 S2 ···
Product ions (P)
P1
P2
P3
P4
Binned intensities in MS2
Correlations
MS1Vec vs. MS2Mat
MS1 vector (MS1Vec)
S1
S2
Ft1
Samples
S1 S2 S3 ···
peak-top intensities in MS1
Ft1
P1 0.73
P2 0.81
P3 0.95
P4 0.98
B Step 2.
Merge all features
Ft1
P1 0.73
P2 0.81
P3 0.95
P4 0.98
Ft2
P1 -0.35
P2 0.93
P4 0.63
Px 0.11
Merge
Features
Ft1 Ft2 ···
Product ions
P1 0.73 -0.35
P2 0.81 0.93 MaxCorrP
P3 0.95 N.D.
P4 0.98 0.63
MaxCorrFt
Correlations
C Step 3. Select product ion by criteria
Comparing with coeluted features and product ions are removed by criteria
(1) CorrMs1vsMs2 > min threshold minimum correlation to keep the product ion
Ft2 case: ✕ P1 and P4 < 0.70
(2) CorrMs1vsMs2 > MaxCorrFt - margin1 removing product ions potentially from different Ft
Ft1 case: ✕ P1: 0.73 < 0.78 (0.98 - 0.20)
(3) CorrMs1vsMs2 > MaxCorrP - margin2 minimizing false positive assignments
Ft1 case: ✕ P2: 0.81 < 0.83 (0.93 - 0.10)

Figure 1. Flowchart of the CorrDec method for a target feature Ft1. A. For each feature, the Pearson correlations are calculated for all pairs of precursor (MS1 vector) and product ions (MS2 matrix). B. All correlation values of all features are merged into a single matrix. C. Product ions satisfying the three criteria (see the main text for details) are selected to produce the deconvoluted MS2 spectrum of Ft1.

Criterion 3 is used to avoid false-positive assignments by Criterion 2 when the same product ion shows high correlation values for multiple precursor ions. For each product ion Px, a maximum correlation $\mathrm{MaxCorr}_{\mathrm{P}}$ with its neighboring features (eluting within ±0.5 × peak width of Ft1) is determined. When the correlation value between the Ft1 and Px is less than $\mathrm{MaxCorr}_{\mathrm{P}}$-margin2, Px is excluded from the deconvoluted spectrum of Ft1. The recommended range is between 0.1 and 0.3, and we used 0.1 in this study. A larger margin indicates a higher possibility to introduce noise peaks into spectra. For example, the product ion P2 is excluded from the Ft1 deconvoluted spectrum because the value of 0.81 is less than $\mathrm{MaxCorr}_{\mathrm{P}}$ (0.93) − 0.1 (Figure 1B).

These threshold values require tuning when applied to different data sets. The m/z value and the intensity in a deconvoluted spectrum are represented by their respective median value of m/z and intensities in biological samples,

https://dx.doi.org/10.1021/acs.analchem.0c01980

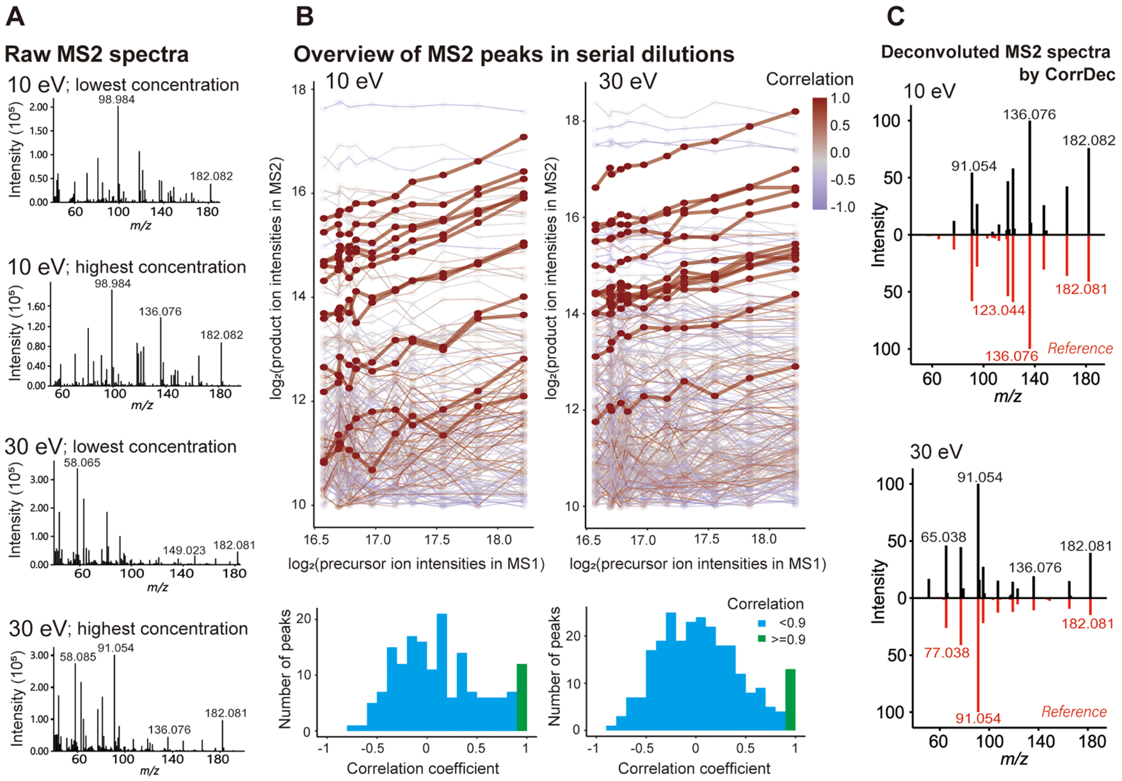


Figure 2. Demonstration of the CorrDec method using tyrosine dilution series spiked into diluted urine as background matrix. A. Raw MS2 spectra of tyrosine $[M + H]^+$ (m/z: 182.082) at the lowest (69 nM) and the highest (4 μM) spiked concentrations in dilution series. Raw MS2 spectra contain over one hundred peaks masking the ions derived from tyrosine, especially at low spiked-in concentrations. B. Linked scatter plots visualizing the intensity correlations between the MS1 m/z 182.082 and MS2 peaks in 11 dilution series samples. Only 12 out of 193 (10 eV) and 13 out of 280 peaks (30 eV) correlated >0.9 (highlighted lines). C. Deconvoluted MS2 spectra (above, in black) matched well with the library reference spectra (below, in red). The MS2 similarities of deconvoluted spectra were 90.5% (10 eV) and 86.5% (30 eV), while the MS2 similarities of raw spectra at 0, 10, and 30 eV were less than 30% in the all samples.

where the intensities are normalized by the abundance of the precursor ion in each sample.

Sample Information and Data Acquisition. Information on samples and experiments are detailed in Supporting Information. Liquid chromatography (LC)-MS measurements in AIF mode were performed as described previously.[18,19] Metabolites were separated on a 15 min gradient using HILIC chromatography with acidified water and acetonitrile. Data were acquired in positive ionization mode on an Agilent 6550 Q-TOF-MS system with a mass range of 40−1200 m/z in AIF mode with three alternating collision energies (0, 10, and 30 eV). The data acquisition rate was 2 scans/s for each segment.

A dilution series of eight chemical standards (proline betaine, trigonelline, dimethylglycine, trimethylamine *N*-oxide, tyrosine, glycine betaine, proline, 3-hydroxy-kynurenine; Table S1) was prepared using 10-fold diluted urine as a matrix. The starting spiked-in concentration of 4 μM in urine was diluted 1.5-fold with an equal amount of urine 10 times, resulting in an 11-point series to the final concentration of 69 nM (Figure S1). In addition, we also acquired data with a smaller dilution step (1.07-fold, 3.27−4.00 μM) for tyrosine. Two dilution series of trimethoprim (1.07-fold starting at 0.06 μM and 1.5-fold starting at 0.3 μM and) were acquired in urine samples with differing matrix composition.

Urine samples (n = 224) were used as the proof of concept for assessing the CorrDec performance. A detailed description of the full study is given in the original publication.[20] Samples were measured in four analytical batches, with pooled quality control (QC) sample injections every five samples and a water blank at the end of the batch sequence. The data sets have been deposited in the EMBL-EBI MetaboLights repository[21] with the identifiers MTBLS787 (chemical standards) and MTBLS816 (urine metabolomics).

Chemical Standard Library. An in-house MS2 spectral library combined from various open and closed sources containing 128,039 experimental MS2 spectra (high-resolution, mostly DDA) for 13 597 compounds was used for identification. The retention times (RT) for 280 compounds were obtained from purchased chemical standards.[18,22]

Data Processing and Analysis. The CorrDec method was implemented into MS-DIAL.[11] Data were processed in MS-DIAL version 4.12 (peak detection, alignment, and deconvolution). Important parameters were as follows: minimum peak height MS1:3000, noise level of MS2:1000, total identified score cutoff: 80%, detected in at least 20% of all samples, not in blank (maximum sample intensity/average blank intensity >5). As our library contained records of both DDA and DIA spectra, we used the deconvoluted spectra with and without the ions heavier than the precursor during the identification process; the higher matching score was kept. Detailed data processing settings of MS-DIAL can be found in the Tables S2 and S3. The MS2 spectra were deconvoluted independently using the MS2Dec[11] and the CorrDec methods (after the alignment of features).

For the urine data, we manually confirmed and curated the alignment results to correct missed or doubtful peak picking, feature alignment, and compound identification. We also annotated all features using three criteria: (i) accurate mass (AM) match (tolerance: 0.01 Da), (ii) RT match (tolerance: 1 min), and (iii) MS2 spectrum match (similarity >80%). The MS2 similarity was scored by the simple dot product without any weighting,[23] for clearer understanding of our method:

$$\text{MS2 similarity}(\%) = 100 \times \frac{\sum (\text{Am Ar})^2}{\sum \text{Am}^2 \sum \text{Ar}^2}$$

where Am and Ar are the arrays of m/z intensities in a measured and reference mass spectrum, respectively. To avoid erroneous high similarity matches resulting from only a few peaks, we adopted the following additional criteria for MS2 spectrum match: (1) if AMRT, a match of at least two MS2 peaks with the reference spectra, and (2) if AM only, a match of at least three MS2 peaks with the reference spectra. The MS2 similarities with reference spectra were compared between the CorrDec and the MS2Dec using three different collision energies (0, 10, and 30 eV).

Random Sampling Analysis. We evaluated the performance of CorrDec for different sample sizes by randomized resampling analysis of the urine metabolomics data set. After chromatographic alignment was performed using all samples, we reselected the study and QC samples for deconvolution by the CorrDec. The number of samples varied from four to the number of detected samples (depending upon the chosen compound) with 100 iterations. For each iteration, we calculated the MS2 similarity between the deconvoluted spectrum from the resampling and the reference spectrum. The MS2 similarity of resampling was the average of 100 iterations.

RESULTS AND DISCUSSION

CorrDec Demonstration Using Compound Dilution Series in Urine. Using a dilution series of chemical standards, we verified high correlation of the intensities of MS2 fragments with those of their precursors. We measured the 11-point dilution series (0.069−4 μM) of eight chemical standards in AIF mode with diluted urine as the matrix. In such a setup, only the concentrations of the spiked compound vary (partially masked by the endogenous compounds present in the matrix) while concentrations of other compounds in the urine matrix remain stable. In the case of the tyrosine dilution series (Figure 2), the MS2 spectra of tyrosine contained 193 and 280 peaks for 10 and 30 eV, respectively. The similarity scores (simple dot product) of all raw MS2 spectra with the reference spectra were less than 30%. When processed by CorrDec, 12 peaks in 10 eV and 13 peaks in 30 eV showed >0.9 correlations with their precursors, clearly deviating from the normal distribution formed by the correlation values of the other peaks (Figure 2B bottom). These highly correlated peaks exhibited intensities proportional to the dilution (Figure 2B top in the log scale), and the MS2 similarity scores with the reference spectra were 90.5% and 86.5% for 10 and 30 eV, respectively. Similar results were obtained for the other 7 compounds: MS2 spectra were successfully generated with high MS2 matches (1 compound >80%, 6 compounds >90% for at least one collision energy) by CorrDec (Table S4 and Figure S2).

In addition to the MS2 spectra at 10 and 30 eV, deconvoluted spectra were also obtained for 0 eV, justifying the use of in-source fragmentation for metabolite identification.[16,17,24,25] The degree of in-source fragmentation depends on the ionization source setting. In this study, in-source fragmentation is facilitated by a high fragmentor voltage (380 V), and the MS2 spectra at 0 eV for all eight chemical standards provided >80% matches in the library (Table S4), corroborating the usability of the in-source fragmentation.

To further confirm the usability of CorrDec for the cases where the concentration of the compounds varies only little across samples, we measured a 1.07-fold dilution series of tyrosine and trimethoprim. For tyrosine, the same QC background was used because the variation of endogenous tyrosine overwhelmed the 1.07-fold variation; we could obtain clean CorrDec spectra in such cases. With the constant QC background, CorrDec could generate MS2 spectra, showing >80% MS2 match at 10 eV using four samples (spiked-in concentration range 3.27−4.00 μM; Figure S3). To ensure the performance of CorrDec with the minimum concentration variations on different urine backgrounds, we measured an exogenous compound, trimethoprim, under different background matrixes. CorrDec could again deconvolute spectra using four samples (Figure S3). We could confirm that small concentration changes between the samples (<25%) suffice for the correlation-based method,[16] when heavy coelution is avoided.

Urine Metabolomics Data Set. To verify the practical performance of CorrDec, we analyzed a LC-MS (HILIC chromatography) metabolomics data set consisting of 224 unique urine samples, 58 pooled QCs, and 4 blanks acquired in positive ionization AIF mode. Data were processed by MS-DIAL version 4.12. In the CorrDec deconvolution process, we discarded product ions that appeared in <50% of all samples for computational efficiency. This threshold of 50% is arbitrary and should be set for each study considering the sample number and the desired level of reliability. The remaining 4159 features were aligned, among which the alignment of 64 features was manually corrected to separate fortuitously merged, coeluting compounds. By matching AM, RT, and MS2 spectra to the reference library, 105 compounds were confidently identified at the MSI level 1.[26]

For all of the 105 compounds, both MS2Dec and CorrDec could generate MS2 spectra. The number of spectra achieving >80% match, however, were 34 and 85 for MS2Dec and CorrDec, respectively (Figure 3A). Furthermore, the distribution of MS2 similarity scores reveals that MS2Dec spectra showed <60% match for 50 compounds. Median similarity values were 59.1% and 91.3% for MS2Dec and CorrDec, respectively (Figure 3B). The reason for the disparity is that CorrDec is especially effective in obtaining cleaner spectra for compounds of low abundance or smaller peak intensity (Figure 3C).[27]

In addition to the 105 compounds identified at the AMRT and MS2 match level, we could also identify six metabolites as high match (>80%) to the applied MS2 library using CorrDec spectra but not with MS2Dec spectra. These compounds have been previously reported in human urine: imidazole acetic acid,[28] homocitrulline,[29] aminohippuric acid,[30] isobutyryl (C4) carnitine,[30] liquiritigenin,[31] and AICA-riboside.[32] Among the 111 identified compounds, over half (61) were amino acids and their metabolites: standard amino acids (13), methylated (9), acetylated (6), other amino acid metabolites (22), and conjugates (11). The other major compound groups include products of nucleic acid metabolism (13), and food/drug metabolites (8) (Table S5 and Figure S4).

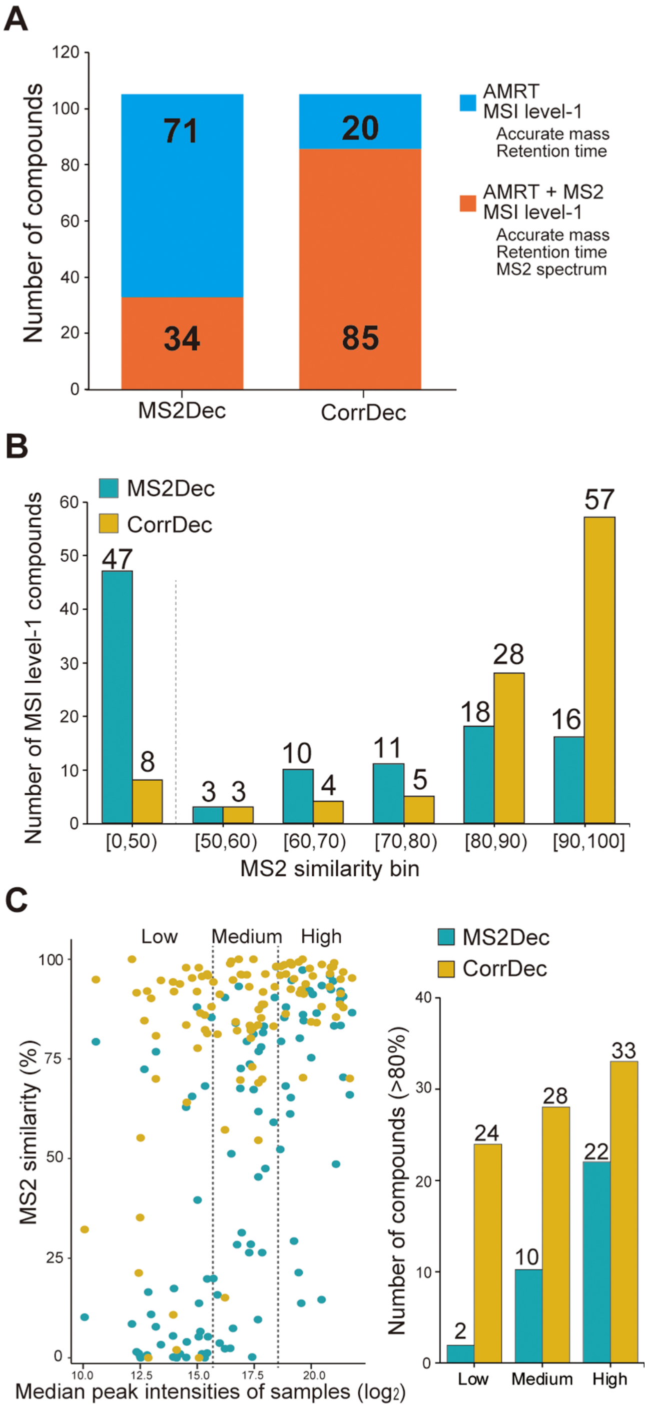


Figure 3. CorrDec MS2 spectra provide increased confidence in compound identification than those obtained by MS2Dec in the urinary metabolomics DIA data set. A. Number of compounds in each identification category identified using MS2Dec and CorrDec. B. Distribution of the MS2 similarity scores for the MSI level-1 compounds spectra deconvoluted by the CorrDec and MS2Dec. C. MS2 similarity scores from CorrDec were higher than MS2Dec, especially for low-intensity peaks.

In addition to the MS2 library matching, CorrDec can provide a more reliable MS2 spectra for structure-prediction tools such as MS-FINDER.[4] For example, we could annotate two features based on their CorrDec spectra as acetaminophen sulfate and valerylcarnitine (Figure S5), two compounds not present in our MS2 spectral library but likely present in urine.[30,33] The method is particularly suitable for compounds with variable levels in the samples, such as drugs and dietary components. Indeed, among the confirmed 85 AMRT+MS2 compounds (Figure 3A), 25 were first annotated by MS2 match only and were later purchased for confirmation.

MS2 Spectra Deconvolution of Coeluting Compounds. Nontargeted LC methods often contain regions with multiple coeluting compounds. In our analytical method, the distribution of the 4159 features ranged from a few to over 250 peaks per 20 s (approximate average peak width at base) across the 0.8−15 min of gradient elution (Figure S6A,B). Such coeluting peaks pose a challenge to deconvolution methods relying on mass chromatograms. With CorrDec, even completely coeluting compounds could be deconvoluted, such as abundant glutamine and sparse *N*-acetylcarnosine (Figure 4A,B). The relative peak intensities of the two compounds fit well with the reported average concentrations in the literature: 18−72 and 1−2 μM/mmol creatinine for glutamine[30] and *N*-acetylcarnosine (see Supporting Information), respectively. Using MS2Dec, the deconvoluted spectrum of *N*-acetylcarnosine contained all fragment peaks of glutamine, reducing the MS2 match to only 29.3%. The MS2Dec deconvoluted spectrum of glutamine showed the MS2 match of 86.5% (Figure 4C). With the same data set, CorrDec could deconvolute the MS2 spectrum of *N*-acetylcarnosine with 95.1% match and provided an equivalent high match for glutamine (84.1%) as well (Figure 4D). Low abundance metabolites such as *N*-acetylcarnosine arguably constitute the larger part of most metabolomics data sets.[27] The high-quality MS2 spectra deconvoluted by the CorrDec enabled us to untangle the complex AIF data set, by improving the identifications and annotations of smaller peaks in chromatographically dense sections.

The benefits of CorrDec are summarized as (1) cleaner MS2 spectra, and (2) statistical annotations (frequency and correlation) for MS2 peaks. CorrDec can generate clean spectra without noise signals from the matrix, mobile phase, or mass spectrometer artifacts, enabling better match to spectral databases or libraries. In the deconvolution process, each MS2 peak is assigned with a correlation value and frequency among samples. Using these statistical annotations, advanced users can manually interpret deconvoluted MS2 spectra of unknown or marginally matching metabolites.

On the other hand, CorrDec has two disadvantages: (1) the requirement of multiple samples with varying compound concentrations, and (2) the possibility of removing shared fragments among coeluting compounds. First, in principle, CorrDec cannot be performed on a single sample and at least three samples are required to calculate the correlation. While we observed that four spiked samples were sufficient to obtain >80% similarity match (Figure S3), we investigated further to estimate the required sample size using random resampling of the urine metabolomics data set in the next section. Second, if coeluting compounds produce the same m/z product ions, their intensity correlations become small enough to be removed from deconvoluted spectra depending on the CorrDec parameters. MS2Dec spectra are useful to complement such missing peaks, and advanced users can recover them through careful interpretation of statistical annotations and MS2 chromatograms. One such example for a group of betaines is provided in Figure S7.

Verification by Random Resampling. Estimating the number of samples required for CorrDec is difficult; it depends

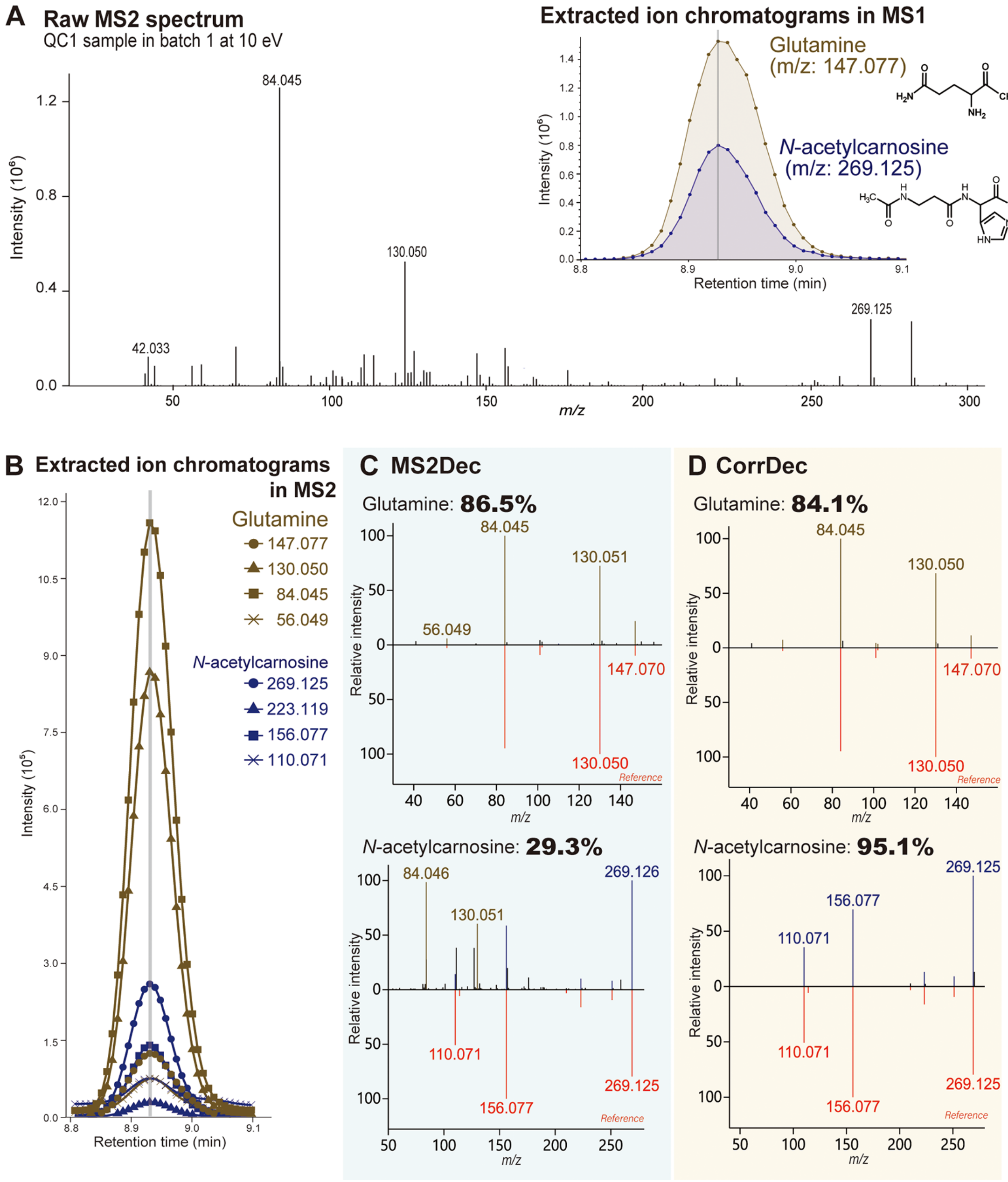


Figure 4. CorrDec can successfully deconvolute the MS2 spectra of completely coeluting compounds, glutamine and *N*-acetylcarnosine. A. The raw MS2 spectrum and extracted ion chromatograms in MS1 (0 eV) of completely coeluting glutamine and *N*-acetylcarnosine as well as B. their fragments in MS2 (10 eV) from the urine data (QC1 sample in batch 1). C. MS2 spectra of glutamine and *N*-acetylcarnosine deconvoluted by the MS2Dec. D. MS2 spectra of glutamine and *N*-acetylcarnosine deconvoluted by the CorrDec.

on multiple factors (study design, sample matrix, metabolite, etc.). Here for a rough estimation, we used 85 compounds confidently annotated (AMRT and MS2 match) in the urine study to perform random resampling analysis. For each of the 85 compounds, we created a scatter plot between the number of samples and MS2 similarity with the preselected library reference spectrum (Figure S8). On the basis of the median MS2 similarity from 100 iterations for each resampling, we plotted the number of compounds (total 85) of high MS2 similarity scores for each sampling size (Figure 5). Already with 10 samples, 47% (40 of 85) of the compounds showed >80% MS2 similarity; when using 30 samples, the number rose to 85% (72 of 85). Therefore, small studies with tens of samples can benefit from the CorrDec method. Note that urine is more variable compared to homeostatic fluids such as blood. A larger number of samples might be required for successful application of CorrDec in studies with less metabolite variations. On the other hand, the quality of MS2 spectra are largely dependent on compound classes and study designs. Defining the best parameters or the minimum sample number required for all studies is therefore difficult.

In MS-DIAL, CorrDec is not intended to replace MS2Dec. Both deconvolution methods are based on different concepts and have different usage scenarios. The CorrDec method provides a reasonably clean deconvoluted MS2 spectrum per feature and sample set, and it is suitable for annotating and identifying a feature at the level of the *whole sample set*. MS2Dec can deconvolute MS2 spectra for each feature in a

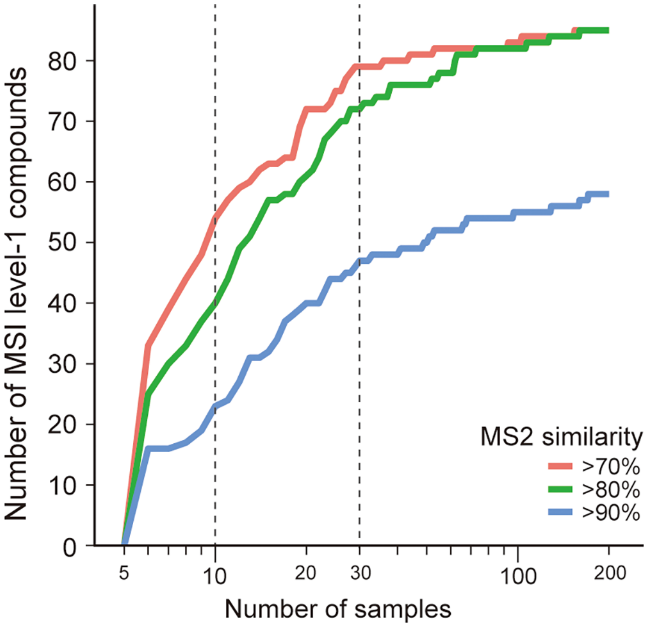


Figure 5. Summary of the randomized resampling analysis for the 85 CorrDec AMRT+MS2 compounds (Figure 3) to assess the relationship between the number of samples (urinary metabolomics data set) used for the CorrDec and quality of the deconvoluted MS2 spectra compared library MS2 spectrum.

single sample; therefore, while noisier, the MS2Dec can be utilized to evaluate the feature identification *for each sample* in the data set. In DIA metabolomics, MS2 spectra are obtained from only a small number of MS scans. For such complex and noisy data, traditional deconvolution methods such as multivariate curve resolution (MCR) are difficult to apply because the multivariate method requires proper constraints to deconvolute spectra. When the number of coeluting compounds and peak shapes are much interfered by noise, error-minimization is not a good algorithmic choice. For such a data set, MS2Dec and CorrDec methods can function in complement to clean MS2 spectra. Lastly, regardless of how clean the MS2 spectra or how good the MS2 library similarity matches are, it is still necessary to further confirm compound annotations with chemical standards.

■ CONCLUSIONS

We have developed CorrDec, a new MS2 spectra deconvolution method for DIA data based on the correlations of the peak intensities across samples. CorrDec has been implemented in MS-DIAL and is available in version 3.22 or later (version 4 also covers ion-mobility data processing[34]). The improved quality of the MS2 spectra and the ability to deconvolute completely coeluting compounds are the main advantages over retention-time based deconvolution methods. Therefore, CorrDec enables more reliable compound annotations and identifications in multisample studies.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c01980.

Figures S1−S8, legends of Tables S1−S8 (PDF)
Tables S1−S8 (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Masanori Arita − *RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 240-0045, Japan; National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan;* orcid.org/0000-0001-6706-0487; Phone: +81-55-981-9449; Email: arita@nig.ac.jp

Craig E. Wheelock − *Gunma University Initiative for Advanced Research (GIAR), Gunma University, Maebashi, Gunma 371-8511, Japan; Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm 171-77, Sweden;* orcid.org/0000-0002-8113-0653; Email: craig.wheelock@ki.se

Authors

Ipputa Tada − *Department of Genetics, The Graduate University for Advanced Studies, SOKENDAI, Mishima, Shizuoka 411-8540, Japan;* orcid.org/0000-0003-4149-7191

Romanas Chaleckis − *Gunma University Initiative for Advanced Research (GIAR), Gunma University, Maebashi, Gunma 371-8511, Japan; Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm 171-77, Sweden;* orcid.org/0000-0001-8042-1005

Hiroshi Tsugawa − *RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 240-0045, Japan; RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 240-0045, Japan;* orcid.org/0000-0002-2015-3958

Isabel Meister − *Gunma University Initiative for Advanced Research (GIAR), Gunma University, Maebashi, Gunma 371-8511, Japan; Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm 171-77, Sweden;* orcid.org/0000-0001-9063-0492

Pei Zhang − *Gunma University Initiative for Advanced Research (GIAR), Gunma University, Maebashi, Gunma 371-8511, Japan; Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm 171-77, Sweden;* orcid.org/0000-0003-2054-928X

Nikolaos Lazarinis − *Division of Respiratory Medicine and Allergy, Department of Medicine, Karolinska University Hospital Huddinge, Stockholm 141-86, Sweden*

Barbro Dahlén − *Division of Respiratory Medicine and Allergy, Department of Medicine, Karolinska University Hospital Huddinge, Stockholm 141-86, Sweden*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c01980

Author Contributions

§I.T. and R.C. are first authors. I.T., R.C., C.E.W., and M.A. designed the study. The development and implementation of CorrDec into MS-DIAL were performed by I.T and H.T. R.C., I.M., P.Z., and IT prepared samples and acquired LC-MS data. Data processing and compound identification were performed by I.T. and R.C. Data analysis, figures, and tables were produced by I.T. and R.C. N.L. and B.D. performed the asthma intervention study. The manuscript was written through the support of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Hideru Obinata, Stacey Reinke, and David Broadhurst for discussions and suggestions. This work was supported by JSPS KAKENHI grant numbers JP18J23133, 17H03621, 15H05898, 15K21738, 18H02432, 18K19155, and 19K17662. We acknowledge support from the Gunma University Initiative for Advanced Research (GIAR), the STINT Foundation, the Swedish Heart Lung Foundation (HLF 20170734, HLF 20180290), the Swedish Research Council (2016-02798), the Stockholm County Council Research Funds (ALF), the Asthma and Allergy Research Foundation, the Centre for Allergy Research and Karolinska Institutet, NBDC, and AMED (JP17gm1010006, JP18gm0910005). This work was supported in part by The Environment Research and Technology Development Fund (ERTDF) (grant no. 5-1752). I.M. was supported by Japan Society for the Promotion of Science (JSPS) postdoctoral fellowship (P17774).

■ REFERENCES

(1) Horai, H.; et al. *J. Mass Spectrom.* **2010**, *45*, 703−714.

(2) Wang, M.; et al. *Nat. Biotechnol.* **2016**, *34*, 828−837.

(3) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12580−12585.

(4) Tsugawa, H.; Nakabayashi, R.; Mori, T.; Yamada, Y.; Takahashi, M.; Rai, A.; Sugiyama, R.; Yamamoto, H.; Nakaya, T.; Yamazaki, M.; Kooke, R.; Bac-Molenaar, J. A.; Oztolan-Erol, N.; Keurentjes, J. J. B.; Arita, M.; Saito, K. *Nat. Methods* **2019**, *16*, 295−298.

(5) Nikolskiy, I.; Mahieu, N. G.; Chen, Y. J.; Tautenhahn, R.; Patti, G. J. *Anal. Chem.* **2013**, *85*, 7713−7719.

(6) Lawson, T. N.; Weber, R. J. M.; Jones, M. R.; Chetwynd, A. J.; Rodríguez-Blanco, G.; Di Guida, R.; Viant, M. R.; Dunn, W. B. *Anal. Chem.* **2017**, *89*, 2432−2439.

(7) Zhu, X.; Chen, Y.; Subramanian, R. *Anal. Chem.* **2014**, *86*, 1202−1209.

(8) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. *Nat. Biotechnol.* **2014**, *32*, 219−223.

(9) Peckner, R.; Myers, S. A.; Jacome, A. S. V.; Egertson, J. D.; Abelin, J. G.; MacCoss, M. J.; Carr, S. A.; Jaffe, J. D. *Nat. Methods* **2018**, *15*, 371−378.

(10) Li, H.; Cai, Y.; Guo, Y.; Chen, F.; Zhu, Z. J. *Anal. Chem.* **2016**, *88*, 8757−8764.

(11) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. *Nat. Methods* **2015**, *12*, 523−526.

(12) Yin, Y.; Wang, R.; Cai, Y.; Wang, Z.; Zhu, Z.-J. *Anal. Chem.* **2019**, *91*, 11897−11904.

(13) Brown, M.; Wedge, D. C.; Goodacre, R.; Kell, D. B.; Baker, P. N.; Kenny, L. C.; Mamas, M. A.; Neyses, L.; Dunn, W. B. *Bioinformatics* **2011**, *27*, 1108−1112.

(14) Alonso, A.; Marsal, S.; Julià, A. *Front. Bioeng. Biotechnol.* **2015**, *3*, 23.

(15) Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T. R.; Neumann, S. *Anal. Chem.* **2012**, *84*, 283−289.

(16) Broeckling, C. D.; Heuberger, A. L.; Prince, J. A.; Ingelsson, E.; Prenni, J. E. *Metabolomics* **2013**, *9*, 33−43.

(17) Broeckling, C. D.; Afsar, F. A.; Neumann, S.; Ben-Hur, A.; Prenni, J. E. *Anal. Chem.* **2014**, *86*, 6812−6817.

(18) Naz, S.; Gallart-Ayala, H.; Reinke, S. N.; Mathon, C.; Blankley, R.; Chaleckis, R.; Wheelock, C. E. *Anal. Chem.* **2017**, *89*, 7933−7942.

(19) Chaleckis, R.; Naz, S.; Meister, I.; Wheelock, C. E. *Methods Mol. Biol.* **2018**, *1730*, 45−58.

(20) Lazarinis, N.; Bood, J.; Gomez, C.; Kolmert, J.; Lantz, A. S.; Gyllfors, P.; Davis, A.; Wheelock, C. E.; Dahlén, S. E.; Dahlén, B. *J. Allergy Clin. Immunol.* **2018**, *142*, 1080−1089.

(21) Haug, K.; Salek, R. M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendraker, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; Maguire, E.; González-Beltrán, A.; Sansone, S. A.; Griffin, J. L.; Steinbeck, C. *Nucleic Acids Res.* **2013**, *41*, D781−786.

(22) Tada, I.; Tsugawa, H.; Meister, I.; Zhang, P.; Shu, R.; Katsumi, R.; Wheelock, C. E.; Arita, M.; Chaleckis, R. *Metabolites* **2019**, *9*, 251.

(23) Moorthy, A. S.; Wallace, W. E.; Kearsley, A. J.; Tchekhovskoi, D. V.; Stein, S. E. *Anal. Chem.* **2017**, *89*, 13261−13268.

(24) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Benton, H. P.; Siuzdak, G. *Anal. Chem.* **2018**, *90*, 480−489.

(25) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Guijas, C.; Majumder, E. L.; Benton, H. P.; Siuzdak, G. *Anal. Chem.* **2019**, *91*, 3246−3253.

(26) Sumner, L. W.; et al. *Metabolomics* **2007**, *3*, 211−221.

(27) Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C. E. *Curr. Opin. Biotechnol.* **2019**, *55*, 44−50.

(28) Tsuruta, Y.; Tomida, H.; Kohashi, K.; Ohkura, Y. *J. Chromatogr., Biomed. Appl.* **1987**, *416*, 63−69.

(29) Pohjanpelto, P.; Niemi, K.; Sarmela, T. *Acta Ophthalmol.* **1979**, *57*, 443−446.

(30) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorndahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; Dame, Z. T.; Poelzer, J.; Huynh, J.; Yallou, F. S.; Psychogios, N.; Dong, E.; Bogumil, R.; Roehring, C.; Wishart, D. S. *PLoS One* **2013**, *8*, No. e73076.

(31) Li, C.; Homma, M.; Oka, K. *Biol. Pharm. Bull.* **1998**, *21*, 1251−1257.

(32) Hornik, P.; Vyskocilová, P.; Friedecký, D.; Adam, T. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2006**, *843*, 15−19.

(33) Bales, J. R.; Sadler, P. J.; Nicholson, J. K.; Timbrell, J. A. *Clin. Chem.* **1984**, *30*, 1631−1636.

(34) Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H., Okahashi, N., Yamada, Y., Tada, I., Bonini, P., Higashi, Y., Okazaki, Y., Zhou, Z., Zhu, Z. J., Koelmel, J., Cajka, T., Fiehn, O., Saito, K., Arita, M., Arita, M. A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol.* **2020**, DOI: 10.1038/s41587-020-0531-2.