Retrospective metabolomics via dual-dimensional deconvolution using ZT Scan DIA 2.0

Yuki Matsuzawa[1,$], Kanako Tokiyoshi[1,$], Bujinlkham Buyantogtokh[1,$], Takaki Oka[1], Jason Causon[2,*], Rana Yamamoto[1], Manami Takeuchi[3], Ushio Takeda[3], Mikiko Takahashi[4], Mayu Hasegawa[5], Gordana Ivosev[2], David Cox[2], Paul RS Baker[2], Anjali Chelur[2], Nic Bloomfield[2], Junki Miyamoto[5], Takeshi Harayama[6,7], Lu Deng[2], Hiroshi Tsugawa[1,4,7,8,*]

1. Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588, Japan
2. SCIEX, 71 Four Valley Drive, Concord, Ontario L4K 4V8, Canada
3. K.K. ABSciex Japan, Shinagawa Tokyo 140-0001, Japan
4. RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
5. Department of Applied Biological Science, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai-cho, Fuchu, Tokyo, 183-8509, Japan
6. Institut de Pharmacologie Moléculaire et Cellulaire, Université Côte d'Azur - CNRS UMR7275 - Inserm U1323, 660 Route des Lucioles, 06560 Valbonne, France
7. Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, 2-24-16 Nakamachi, Koganei-shi, Tokyo 184-8588, Japan
8. RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

[$]These authors contributed equally to this work

***Corresponding Authors**

Jason Causon (J.C. Jason.Causon@sciex.com), Hiroshi Tsugawa (H. T.: htsugawa@go.tuat.ac.jp)

Abstract

Herein, we present a scanning data-independent acquisition (DIA) approach (ZT Scan DIA 2.0) combined with dual-dimensional tandem mass spectrometry spectral filtering and deconvolution along both the quadrupole and retention time axes to reconstruct compound-specific MS2 spectra from complex mixtures. This approach outperformed conventional data-dependent acquisition (DDA) and window-based DIA methods in terms of annotation rates for hydrophilic metabolomics (114–160%) and lipidomics (105–136%). Moreover, this approach achieved dot product score distributions comparable with those obtained with a 1-Da precursor isolation window. Furthermore, the platform enables lipid isomer separation through the retrospective analysis of complete DIA datasets covering 1,017 and 2,353 molecules for human plasma and mouse liver tissues, respectively. In addition, the platform yields compound-specific ground truth MS2 spectra that surpass DDA in terms of spectral purity. This establishes a transformative foundation for repository-scale metabolomics in line with the findable, accessible, interoperable, and reusable data principles.

Main text

Mass spectrometry (MS) is the gold standard for comprehensive profiling of small molecules. These molecules are collectively referred to as the metabolome, lipidome, or exposome, depending on the biological context[1-3]. The challenge of small-molecule profiling lies in their vast diversity of chemical structures and concentrations within living organisms. As of July 6, 2025, the Human Metabolome Database (HMDB) listed 248,138 compounds in human blood, with concentrations ranging from 1 pM for aldosterone to 32 mM for hydrogen carbonate—a dynamic range spanning over 10^7-fold[4]. Capturing all of these molecules simultaneously remains unrealistic owing to the limited dynamic range of MS (10^4–10^5). Consequently, optimized extraction protocols and analytical conditions are employed and generate >200 terabytes of MS data, which is housed in public repositories[5-7]. Although large-scale machine learning applications increasingly benefit from rich and structured knowledge bases, as exemplified by recent advances in language models, the demand for highly reusable and scalable omics datasets to drive transformative progress in biology has increased[8, 9].

Liquid chromatography (LC) coupled with electrospray ionization-based tandem MS (MS/MS) is the most widely used platform for untargeted metabolomics and lipidomics. Because molecular annotation is a prerequisite for biological interpretation, properties, such as retention time (RT), precursor *m/z*, isotope pattern, and product ion (MS2) spectra, were computationally leveraged to assign molecular formulae, classes, and structures[1]. Among these, MS2 spectra are the most informative descriptors for characterizing molecular substructures and are typically acquired using either data-dependent acquisition (DDA) or data-independent acquisition (DIA)[10]. DDA provides high-purity MS2 spectra using narrow precursor isolation windows (<1 Da), which facilitate accurate metabolite annotation based on spectral similarity metrics, such as dot product or spectral entropy, against reference libraries. In contrast, DIA acquires comprehensive MS2 data for all detectable precursor ions across the RT and *m/z* dimensions, thereby enabling the detection of low-abundance molecules and semi-quantification of structural isomers based on unique fragment ion chromatographic profiles. However, a major challenge of DIA-based untargeted metabolomics is the mathematical deconvolution of highly multiplexed MS2 spectra, in which contaminant ions decrease spectral similarity scores and increase false-negative rates[11-13]. Moreover, the accurate characterization of co-eluting isobaric structural isomers remains a pressing issue in lipidomics, requiring the optimal utilization of diagnostic fragment ions that define lipid classes and acyl chain compositions[14].

Suitably, in this study, we present a novel untargeted analysis framework based on ZT Scan DIA 2.0. The platform was designed to acquire MS2 spectral data over a broader *m/z* range than that of the first-generation ZT Scan DIA[15]. This is essential for untargeted metabolomics and lipidomics. ZT Scan DIA is a DIA method that employs a sliding scanning technique to continuously isolate precursor ions for MS2 data acquisition, in contrast to window-based scanning approaches, such as SWATH-DIA (**Supplementary Fig. 1**). The unique data structure generated by ZT Scan DIA enables a new concept in spectral deconvolution by leveraging the MS2 signal traces across the quadrupole (Q1) data bins (**Fig. 1a**). MS2 spectra acquired with isolation windows ranging from 5–20 Da are distributed into Q1 bins centered on the precursor isolation range, where the Q1 bin resolution is automatically determined by the width of the isolation window ranging from 1.0 Da for narrow windows (e.g., 5 Da) to 4.0 Da for wider windows

(e.g., 20 Da). During each scan cycle, the MS2 signals from a given analyte were recorded across multiple overlapping Q1 bins corresponding to the isolation window.

To reconstruct the compound-specific MS2 spectra, we developed a filtering algorithm termed Q1 axis-based deconvolution (Q1Dec). Q1Dec extracts MS2 ion chromatograms across Q1 bins and retains only those fragment ions whose intensity maxima are located within the Q1 bin corresponding to the target precursor *m/z* or within its immediate neighboring bins (i.e., ±1 Q1 bin). All other signals were discarded as contaminants. An illustrative example of compound detected thanks to Q1Dec is trazodone, which is a serotonin reuptake inhibitor used as an antidepressant. Notably, this compound is not listed in the HMDB-listed compounds in human blood. This indicates that accurate spectral deconvolution in untargeted workflows allows for the detection of previously undetectable molecules (**Fig. 1b**). In this case, the ion at *m/z* 101.0585 was excluded because its signal peaked in the Q1 bin centered at *m/z* 366.71, which does not correspond to the precursor *m/z* of protonated trazodone (theoretical *m/z* 372.1586; Q1 bin centered at *m/z* 372.50). In contrast, the fragment ion at *m/z* 176.0808 was retained as its intensity maximum aligned with the correct Q1 bin. The Q1Dec-processed MS2 spectrum achieved dot product and reverse dot product similarity scores of 0.877 and 0.914, respectively. These values are substantially higher than those of the raw spectrum (0.391 and 0.906) and exceeded the commonly used 0.6–0.8 thresholds for confident annotation in untargeted metabolomics[16, 17].

We further demonstrated the utility of Q1Dec by resolving the co-eluted exposome compounds γ-undecalactone and dicyclohexylamine, which share the same RT in hydrophilic metabolomics data from NIST SRM 1950 human plasma. These compounds have precursor *m/z* values of 185.154 and 182.190, respectively, differing by only ~3 Da, whereas the precursor isolation window in our ZT Scan DIA 2.0 method was set to 7.2 Da. A diagnostic ion at *m/z* 83.084 was observed in both compounds and was retained in the Q1Dec MS2 spectra as the local maxima for this fragment in the Q1 bins, corresponding to *m/z* 185.1541 and 182.1896, respectively. The annotation of dicyclohexylamine was successful because of the substantial improvement in spectral similarity, indicated by an increased dot product score from 0.510 (raw MS2) to 0.825 (Q1Dec MS2) when matched against the reference MS2 spectrum. In contrast, the product ion at *m/z* 100.107 derived from dicyclohexylamine was not excluded from the Q1Dec MS2 spectrum of γ-undecalactone owing to a low signal-to-noise ratio, which caused its local maximum to appear within the Q1 bin at *m/z* 185. Nevertheless, the Q1Dec spectrum was sufficiently cleaned for confident annotation of γ-undecalactone, as reflected in an improved dot product score of 0.755 compared to 0.632 using the raw MS2 spectrum.

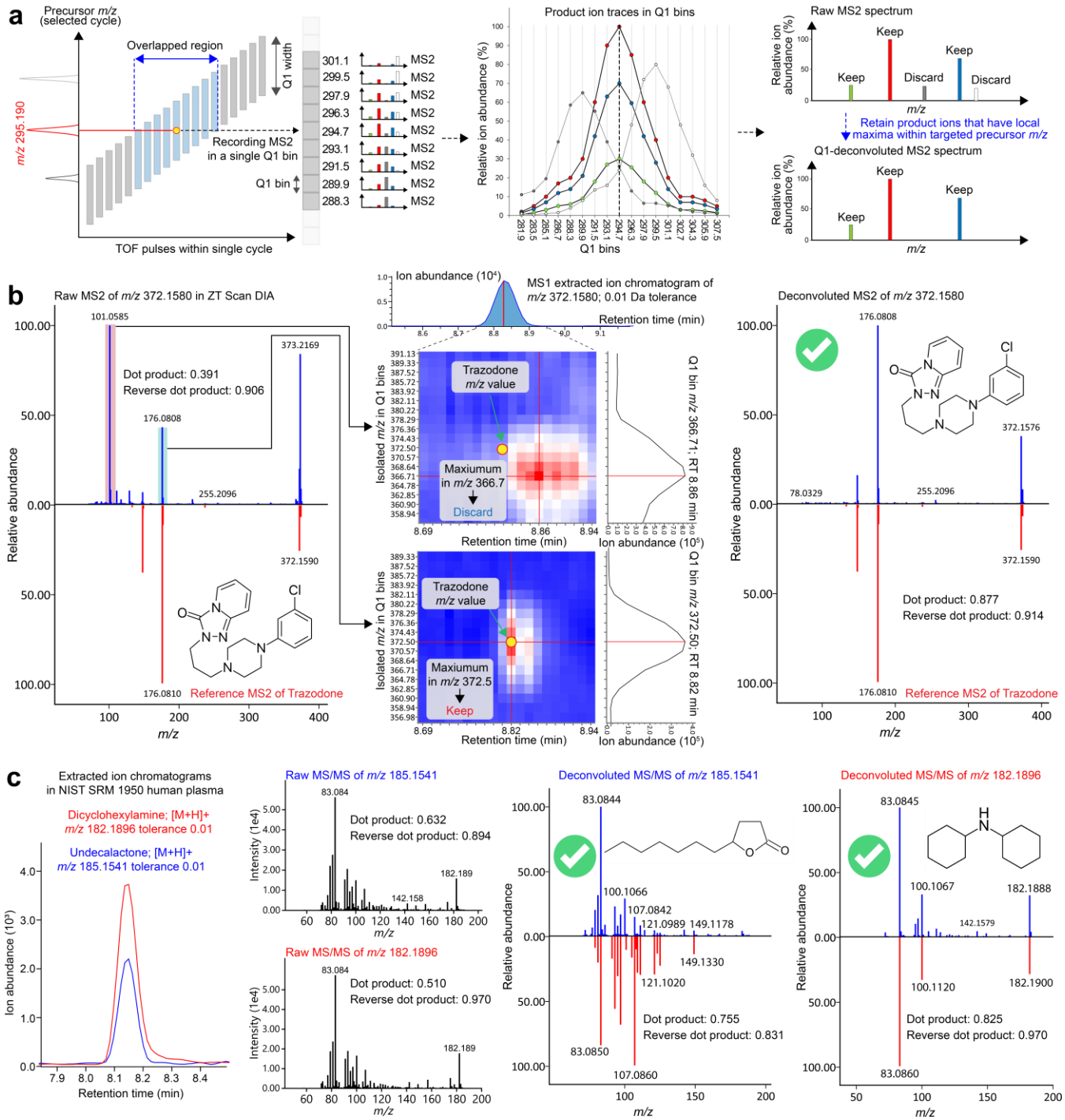


Fig. 1. Concept and application of Q1-axis-based MS2 spectral filtering. (a) Schematic overview of Q1 axis-based data processing (Q1Dec), a spectrum filtering strategy that purifies product ion (MS2) spectra using ion intensity traces across Q1 bins. When a precursor isolation window of 7.5 Da is scanned across a precursor range, fragment ions of *m/z* 295.190 precursor ion are recorded across multiple Q1 bins spanning from *m/z* 287.69–302.69. Each Q1 bin has a width of 1.5 Da. Q1Dec retains product ions only if their intensity maxima fall within the Q1 bin of the target precursor or its immediate neighbors (±1 Q1

bin); all others are excluded. (b) Example of Q1Dec applied to trazodone. The raw MS2 spectrum (blue) includes contaminant signals, whereas the Q1Dec-processed spectrum aligns with the reference spectrum (red). A true fragment ion at *m/z* 176.0808 shows its maximum intensity in the Q1 bin at *m/z* 372.50, consistent with the precursor *m/z* of protonated trazodone (theoretical *m/z* 372.1586). In contrast, the ion at *m/z* 101.0585 is excluded as a contaminant, as its signal peaks in a mismatched Q1 bin (*m/z* 366.71). Heatmap axes represent scan time (x-axis) and Q1 bins (y-axis); red indicates higher ion intensity. (c) Application of Q1Dec to two co-eluted exposome compounds (γ-undecalactone and dicyclohexylamine)that share retention time but differ in terms of precursor *m/z* values. Raw MS2 spectra contain overlapping signals and background ions. Q1Dec resolves both compounds based on diagnostic fragments, enabling confident annotation with dot product and reverse dot product scores exceeding 0.755 and 0.831 for γ-undecalactone and 0.825 and 0.970 for dicyclohexylamine, respectively.

We further evaluated the use of RT-based deconvolution with Q1 filtering (RTDec, originally termed MS2Dec) to refine the Q1Dec-processed spectra (**Fig. 2**)[11]. For example, we applied Q1Dec followed by RTDec to the MS2 spectrum of 6-gingerol (**Fig. 2a**). Q1Dec effectively removed the major contaminant ions at *m/z* 288.289. However, certain nonspecific fragment ions, such as *m/z* 133.099, which exhibited a local maximum in the Q1 bin of *m/z* 295.77, remained in the spectrum after Q1Dec alone. RTDec addresses this limitation by evaluating the similarity of the elution profiles. This approach compares the chromatographic trace of each ion against a model peak generated from a unique mass fragment trace specific to the target precursor. This layer of LC deconvolution successfully excludes ions arising from the solvent background and co-eluting species with distinct RTs, resulting in a more purified MS2 spectrum. The dot product score of the final annotated spectrum substantially improved from 0.56, which was obtained using Q1Dec, alone to 0.89.

We further assessed the performance of dual-dimensional processing using polar metabolomics MS2 spectra from NIST SRM 1950 human plasma acquired under identical LC conditions with DDA, SWATH-DIA, and ZT Scan DIA 2.0 (three technical replicates each) (**Fig. 2b**). The MS2 spectra from six configurations, namely the original SWATH-DIA, SWATH-DIA + RTDec, original ZT Scan DIA 2.0, ZT Scan DIA 2.0 + Q1Dec, ZT Scan DIA 2.0 + RTDec, and ZT Scan DIA 2.0 + Q1Dec + RTDec, were compared against the DDA spectra. To define a realistic benchmark for DDA vs. DDA performance, we aligned peaks across the three DDA replicates, consolidated their MS/MS spectra into an MSP library, where the representative MS/MS spectrum was selected from a file among the three files, and searched for each file against this library. We then excluded spectra that matched identically (dot product = 1) (see **Methods**). As a result, ZTScan DIA 2.0, with both Q1Dec and RTDec, produced a dot product score distribution most similar to that of DDA versus the DDA benchmark, outperforming the other acquisition and deconvolution strategies tested. This trend persisted when the spectra were searched against MS/MS reference libraries, including MassBank, MoNA, and NIST20 (**Fig. 2c**). Spectra with reverse dot product similarity above 80% typically display bimodal dot product distributions, and the right-hand peak represents high-confidence matches[18-20]. Notably, only DDA and ZT Scan DIA 2.0 with Q1Dec + RTDec yielded clear bimodal distributions of dot product scores. This allowed us to define the onset of the high-scoring peak (~60%) as an empirical threshold for confident annotation. This strategy is consistent with established metabolomics practices in which bimodal similarity score distributions are used to justify non-

arbitrary thresholds. By applying dot product and reverse dot product thresholds of 60 and 80%, respectively, ZT Scan DIA 2.0 with Q1Dec and RTDec assigned MS2 spectra for 99.8% of the precursor ions and achieved annotation rates that were 120 and 145% higher than those of DDA and SWATH-DIA on average (**Fig. 2d and 2e**).

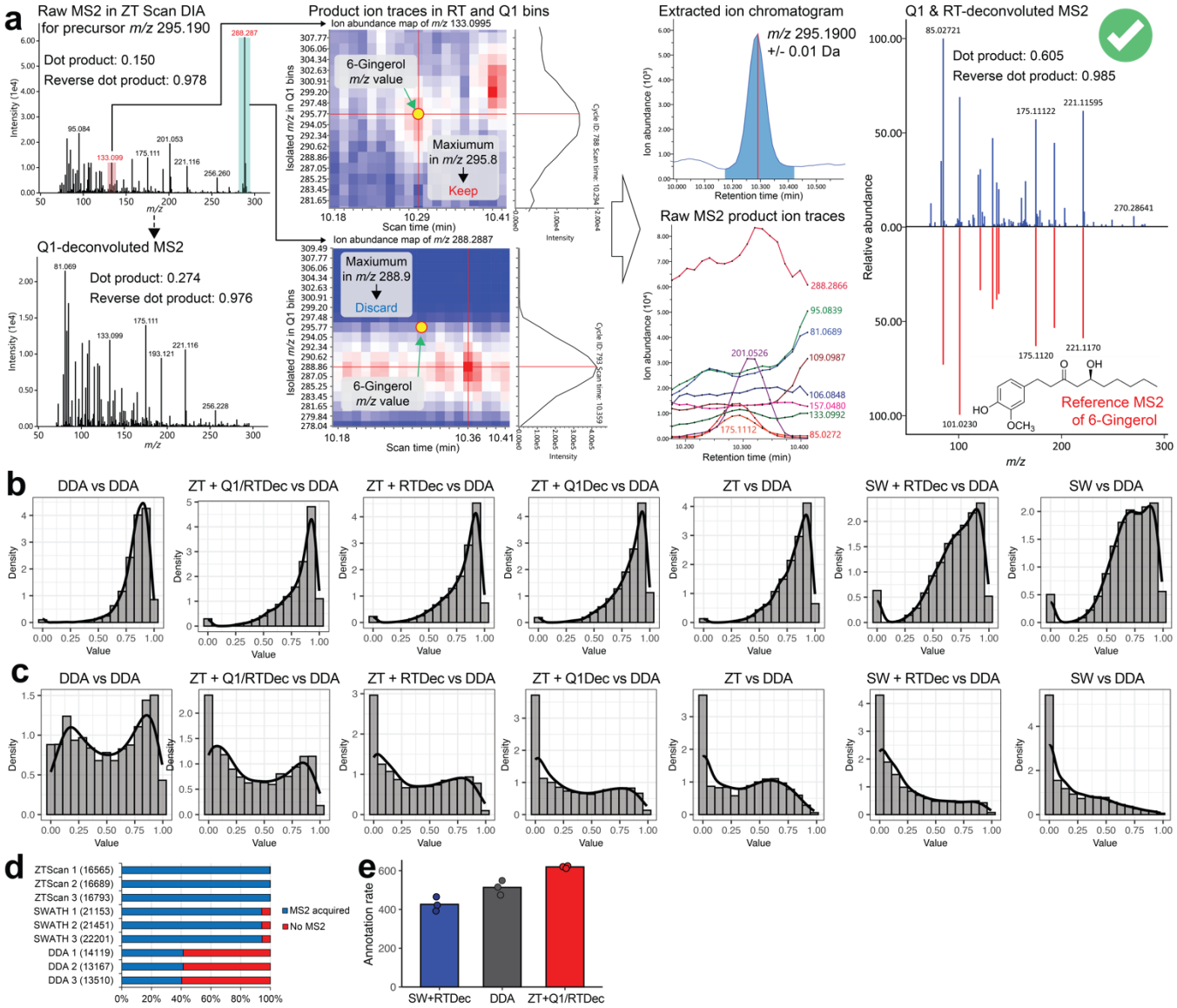


Fig. 2. Validation of dual-dimensional MS2 processing using hydrophilic metabolomics data. (a) Example of Q1 axis-based spectral filtering (Q1Dec) and retention time-based deconvolution (RTDec) applied to the MS2 spectrum of 6-gingerol. A contaminant ion at *m/z* 288.287 was excluded owing to a mismatch between its Q1 bin maximum and the precursor *m/z* of 6-gingerol, whereas the fragment ion at *m/z* 133.099 was retained based on alignment with the correct Q1 bin at *m/z* 295.77. Subsequent RTDec applied to Q1Dec-processed MS2 chromatograms further purified the spectrum, thus increasing the dot product score from 0.56 (raw) and 0.75 (Q1Dec) to 0.89. (b) Weighted dot product score distributions obtained by searching MS2 spectra against data-dependent acquisition (DDA) derived reference spectra

from the same biological sample. The x-axis denotes dot product score, and the y-axis represents density. Results are shown for seven methods: DDA, ZT Scan DIA 2.0 + Q1Dec + RTDec (ZT + Q1/RTDec), ZT Scan DIA 2.0 + Q1Dec (ZT + Q1Dec), ZT Scan DIA 2.0 + RTDec (ZT + RTDec), original ZT Scan DIA 2.0 (ZT), SWATH-DIA + RTDec (SW + RTDec), and original SWATH-DIA (SW). (c) Weighted dot product score distributions from spectral matching against experimental MS2 libraries of authentic standard compounds. Axis definitions are consistent with panel (b). (d) Tandem mass spectrometry (MS/MS) assignment rates to precursor ions detected. Owing to spectral deconvolution processes, no MS2 data were acquired for several precursor ions, even for DIA methods. (e) Comparison of annotation rates in untargeted metabolomics using DDA, SWATH-DIA + RTDec, and ZT Scan DIA 2.0 + Q1Dec + RTDec. The number of annotated features exceeding dot product and reverse dot product thresholds of 0.60 and 0.80, respectively, was counted for each method.

Finally, we highlight the advantages of ZT Scan DIA 2.0 in untargeted lipidomics. The product ion at *m/z* 184.07 is a well-known diagnostic fragment of the phosphocholine head group found in phosphatidylcholine (PC) and sphingomyelin species. However, contamination of this diagnostic ion in the MS/MS spectra can lead to misannotation. We investigated this issue by generating an ion heatmap of *m/z* 184.0726 across the RT range, in which PC 16:0_18:1, the most abundant phospholipid in blood and many mammalian cells and tissues[21-23], was eluted in the ZT Scan DIA data from NIST SRM 1950 human plasma (**Fig. 3a**). The heatmap revealed that this diagnostic ion is broadly distributed across Q1 bins from *m/z* 750.57–770.78, with a local maximum centered at *m/z* 760.54, corresponding to the precursor ion *m/z* 760.5841 of PC 16:0_18:1 in the MS1 dimension. Importantly, the same *m/z* 184.07 fragment was also observed in the DDA spectra for the neighboring precursors *m/z* 768.5594 and 770.6047. In such cases, the presence of the PC marker ion could mislead their annotation toward PC 35:4 and PC O-36:3, and their correct annotations as phosphatidylethanolamine (PE) 18:0_20:4 and an unknown compound, respectively, required careful examination of the chromatographic elution behavior and lipid class-specific retention patterns. In contrast, Q1Dec applied to ZT Scan DIA 2.0 data successfully removed the *m/z* 184.0726 signal from the MS2 spectrum of *m/z* 770.6047. Similarly, this diagnostic ion was eliminated from the MS2 spectrum at *m/z* 768.5594, thereby facilitating the accurate annotation of PE 18:0_20:4. Annotation in the presence of such contaminant marker ions becomes particularly challenging when authentic standards are unavailable. Such is the case in untargeted lipidomics, especially when RT behavior is considered under reversed-phase chromatography conditions. Importantly, our results demonstrated that scanning DIA combined with appropriate processing along with two dimensions yields an MS2 spectrum that more closely reflects the ground truth, with increased MS2 acquisition and annotation rates compared with those obtained via DDA (**Fig. 3b and 3c**).

We also demonstrated how MS2 fragment ions can be leveraged to resolve co-eluting lipid structural isomers and expand lipid feature detection and quantification in ZT Scan DIA 2.0 data. As an initial evaluation, we examined the correlation between the precursor ion peak heights and total ion counts of the corresponding MS2 spectra (**Fig. 3d**). Although certain abundant product ions, particularly PCs and triacylglycerols (TG), exhibited saturation relative to their MS1 precursor intensities, the overall correlation coefficient exceeded 0.89. This indicated that MS2 fragment ions can also serve as quantitative measures. To facilitate the comprehensive use of MS2 data in metabolic profiling, we implemented new

221 functions in MS-DIAL that allow quantification based on isomer-specific MS2 fragment ions when MS1
222 features are assigned multiple co-eluting isomers. For this, MS1 features that are defined by RT and
223 precursor *m/z* were extended to multiple features that are defined by RT, precursor *m/z*, and isoform-
224 specific MS2 fragments, whenever the peak was detected in all biological samples and if the lipid was
225 annotated at the molecular species level (see **Methods** and **Supplementary Fig. 2**). This allowed the
226 discrimination of co-eluting lipid isomers and their separate quantification based on MS2 fragment ion
227 intensity. Applying this approach to aligned peak features from NIST SRM 1950 plasma and mouse liver
228 tissue, we observed annotation rate increases of 131 and 117%, respectively, compared with that of
229 conventional two-dimensional feature detection (**Fig. 3e and 3f**). The total number of non-redundant
230 annotated features reached 1,017 in the plasma and 2,353 in the liver. This surpassed the annotation
231 coverage reported in our previous DDA-based study using a TripleTOF 6600 system, which characterized
232 541 and 746 lipid features in the same or similar biological samples, respectively[22]. Although direct
233 comparisons of annotation rates are limited by differences in acquisition mode and instrumentation, our
234 results suggest that ZT Scan DIA 2.0, with three-dimensional feature detection, yields broader lipidome
235 coverage than that of conventional untargeted lipidomic methods using standard extraction protocols[24].
236 The advantage of MS2-based peak quantification was further exemplified by the separation of PC
237 18:1_20:4 and PC 16:0_22:5 isomers in the lipidomes of young (two-month-old) and aged (24-month-
238 old) mice fed either a high-fat diet (HFD) or a control diet (CE2) for one month (**Fig. 3g and Fig. 3h**).
239 Both compounds were annotated at the molecular species level using MS-DIAL, with PC 18:1_20:4
240 selected as the representative annotation in the aligned peak set across the liver samples. By extracting
241 unique fragment ion traces specific to each isomer, we uncovered distinct abundance patterns: PC
242 18:1_20:4 levels were elevated in aged mice across both diets, whereas PC 16:0_22:5 levels were reduced
243 under HFD conditions and further decreased in aged mice fed the HFD. The accumulation of
244 polyunsaturated fatty acid (PUFA)-containing phospholipids, such as PC 18:1_20:4, is a characteristic
245 lipidomic signature of aging[21]. However, diet- and age-associated reductions in PC 16:0_22:5 levels have
246 rarely been documented. Under HFD conditions, n-3 PUFAs decline, and 22:5 species in PC have been
247 characterized as n-3 PUFAs in the mouse liver[25-27]. Although the underlying mechanisms remain unclear,
248 such isomer-specific trends may reflect changes in the activities of enzymes, such as fatty acid desaturases
249 and acyltransferases. Notably, the use of MS2 chromatograms enabled the detection of distinct changes in
250 PC 16:0_22:5 that were not captured by MS1-based quantification, primarily reflecting the behavior of
251 PC 18:1_20:4. This highlights the risk of overlooking critical metabolic differences when relying solely
252 on the MS1 data.
253 Using a comprehensive lipidomics dataset of mouse liver tissues, we performed principal
254 component analysis (PCA) on the auto-scaled data of log-transformed lipid abundances. This analysis
255 revealed distinct score distributions among the four groups (CE2, young; CE2, aged; HFD, young; and
256 HFD, aged) along the first (PC1) and second (PC2) principal component axes (**Fig. 3i**). We further
257 compared the young and aged groups under the control diet condition by orthogonal partial least squares
258 discriminant analysis (OPLS-DA) and examined lipids with variable importance in projection (VIP),
259 which captured age-related differences (**Supplementary Fig. 3**). This analysis revealed an age-associated
260 increase in bis(monoacylglycero)phosphate (BMP) and several glycerolipids containing PUFAs,

consistent with a previous report[21]. Notably, age-related differences in bacterial lipids such as *N*-acyl glycine and acyl diacylglyceryl glucuronide, and TGs esterified with fatty acid esters of hydroxy fatty acids, as well as in monogalactosyldiacylglycerols with long-chain saturated fatty acids, represent novel findings that have not been reported in previous aging lipidomics studies. Differentially expressed lipids (DEL) were also characterized via volcano plot analysis using Student's t-tests with Benjamini–Hochberg correction (adjusted $P < 0.05$) and a fold change ≥ 1.5 or ≤ 0.5 relative to control. This revealed that only a small number of DELs could be attributed to aging, owing to the large inter-individual variability in lipid expression levels at 24-month-old mice. The most notable lipid signature captured in our lipidomics experiment was that HFD feeding in young mice induced substantial changes in the hepatic lipidome (297 lipid species), whereas HFD feeding in aged mice had almost no statistically significant effect on the liver lipid profiles (**Fig. 3j**). Using the lipid names present in the liver lipidomics dataset and those identified as DELs in the CE2/M2 vs. HFD/M2 comparison, enrichment analysis of lipid classes and fatty acid properties revealed that, under HFD feeding, acyl chains such as monounsaturated fatty acids (18:1) and 20:4 (likely arachidonic acid), as well as lipid classes including sphingomyelin (SM), ether-linked TG (Ether TG), and BMP were extracted as the significantly enriched terms (**Fig. 3k**). The profile of the most abundant molecules within these enriched terms further showed that HFD feeding led to an increase in lipids containing saturated fatty acids (16:0, 18:0), monounsaturated fatty acids, and 20:4 (**Fig. 3l**). These trends closely reflected the fatty acid composition of the HFD used in this study (D12492; Research Diets, Inc.). In contrast, CE2 feeding uniquely enriched acyl chains such as 20:5 (likely eicosapentaenoic acid) and odd-chain fatty acids (15:0, 19:0), along with lipid classes including PC, phosphatidylinositol (PI), and lysophosphatidylethanolamine (LPE). The higher abundance of lipids containing 20:5 under CE2 feeding compared with HFD is likely attributable to the absence of eicosapentaenoic acid in the HFD chow. By contrast, the relative increase in lipids containing odd-chain 15:0 and 19:0 is intriguing. Although more experiments are required to investigate the mechanisms, the link between the gut microbiome and odd-chain fatty acids raises the possibility that differences in gut microbiota composition between CE2- and HFD-fed mice contribute to these changes. Although the molecular mechanisms underlying the absence of lipidomic changes in aged mice fed a HFD remain unclear, aged mice are known to exhibit elevated inflammatory markers across various tissues[28], which may contribute to reduced hepatic lipid uptake during high-fat feeding. Detailed mechanistic studies are beyond the scope of this work and will be reported elsewhere.

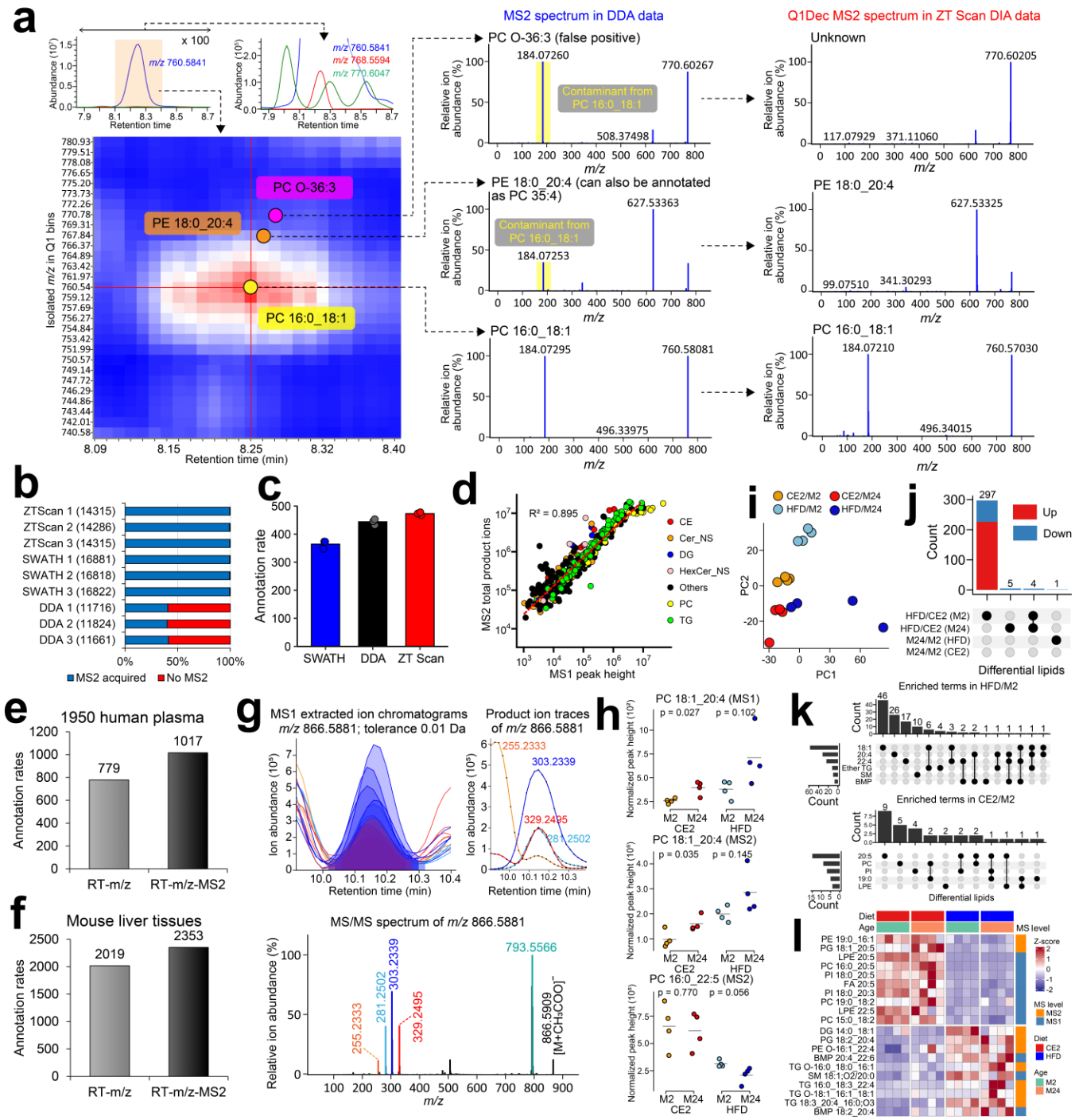


Fig. 3. Evaluation of untargeted lipidomics workflow using ZT Scan DIA 2.0 and MS2 deconvolution. (a) Extraction of ground-truth MS2 spectra from untargeted lipidomics data. Extracted ion chromatograms (EIC) for *m/z* 760.5841, 768.5594, and 770.6047 are shown. An ion heatmap of *m/z* 184.072 in the MS2 spectrum of protonated phosphatidylcholine (PC) 16:0_18:1 (*m/z* 760.5841) is presented along retention time (RT) and quadrupole (Q1) axes. MS2 spectra corresponding to the three peaks were extracted from both data-dependent acquisition (DDA) and ZT Scan DIA 2.0 data, with Q1 axis-based spectral processing (Q1Dec) applied to the latter. (b) Comparison of MS2 acquisition rates among ZT Scan DIA 2.0, SWATH-

DIA, and DDA techniques. (c) Comparison of annotation rates in untargeted lipidomics using SWATH-DIA, DDA, and ZT Scan DIA 2.0. Retention time-based deconvolution (RTDec) was applied for SWATH-DIA deconvolution, and Q1Dec was used for ZT Scan DIA 2.0. (d) Correlation between MS1 and MS2 quantification. Pearson correlation coefficients were calculated between MS1 precursor peak heights and summed MS2 product ion intensities. (e) Comparison of lipid feature counts using RT–*m/z* versus RT–*m/z*–MS2 dimensional criteria. Features were extracted from aligned data across NIST SRM 1950 plasma (N = 3). Positive and negative ion mode results were merged, and duplicate annotations were removed to count unique features. (f) Same analysis as (e) performed on mouse liver tissue data (four experimental conditions; N = 4). (g) Showcase of the MS2 chromatogram separations to distinguish lipid structural isomers. EICs of *m/z* 866.5881, initially annotated as PC 18:1_20:4 in MS-DIAL, are shown for four conditions of mice (N = 4). The MS2 chromatograms for unique fragment ions (*m/z* 281.2502 and 303.2339 for PC 18:1_20:4 and *m/z* 255.2333 and 329.2495 for PC 16:0_22:5) were extracted along RT to distinguish co-eluting isomers. The MS2 snapshot of chromatogram peak top in MS1 dimension was also described. (h) Comparison of normalized MS2 peak intensities across biological groups. Product ion intensities specific to each structural isomer were summed for MS2-based quantification. Statistical significance was assessed using two-tailed Student's t-test. (i) Principal component analysis (PCA) score plots generated from auto-scaled data of log-transformed lipid abundances. (j) Upset plot illustrating differentially expressed lipids (DEL) identified based on adjusted P values <0.05 (two-sided Student's t-tests with Benjamini–Hochberg correction) and fold changes ≥1.5 or ≤ 0.5 relative to control. (k) Upset plot highlighting the results of lipid molecule enrichment analysis. (l) Heatmap visualization of DELs, with highly abundant lipids highlighted according to enrichment in lipid ontology analyses.

In conclusion, our mathematical deconvolution algorithms applied to the ZT Scan DIA 2.0 data enable the reconstruction of high-purity, compound-specific MS2 spectra that often surpass those obtained from conventional DDA in terms of spectral accuracy. These algorithms were fully integrated into the graphical user interface of MS-DIAL, which now features a dynamic MS2 loader that substantially reduces memory consumption. This allows users to handle ZT Scan DIA 2.0 files exceeding 20 gigabytes in size (**Supplementary Fig. 4**). This platform improves annotation rates in both untargeted metabolomics and lipidomics, enables the deconvolution of lipid structural isomers through MS2-based chromatographic features, and supports retrospective analysis because all molecular information is embedded in the raw data. By providing access to ground-truth-like MS2 spectra at scale, our approach lays the foundation for Findable, Accessible, Interoperable, and Reusable data-driven research in the life sciences and promotes more transparent, reproducible, and reusable metabolomic workflows.

Methods

Dynamic MS2 loader environment

The ability to directly import vendor-specific MS raw data files without file conversion, such as mzML, is one of the notable features of MS-DIAL. WIFF (.wiff), WIFF2 (.wiff2), and WiffScan (.wiff.scan) files are typically generated from recent SCIEX machines, where MS spectral data are stored in a WiffScan file and spectral retrieval can be performed from WIFF or WIFF2 files via software development kits. Both the WIFF and WIFF2 data on direct imports are supported by MS-DIAL. The dynamic MS2 loader functions were developed for the WIFF2 format file, where the ZT Scan DIA 2.0 metadata information was fully supported.

For data processing that included the processes of peak picking, MS2 peak purification, peak annotation, and peak alignment, the previous MS-DIAL program stored all MS1 and MS2 data in random-access memory because the data size of MS2 was similar to or typically lower than that of MS1. However, the situation changed in ZT Scan DIA 2.0, where the MS2 data size (10–20 GB) became 100–200 times larger than that of the MS1 (10–200 MB). This is also true for other high-end machines, such as the Orbitrap Astral of Thermo Fisher Scientific. Furthermore, the current MS-DIAL program only requires MS1 data for the peak picking and peak alignment processes, and not all MS2 data are used for spectral deconvolution and annotation. These facts motivated us to develop a dynamic MS2 loader system that calls MS2 data on demand. Thus, the source codes for MS2 data retrieval were completely rewritten in this study to execute on-demand MS2 extraction, whereas the spectra were stored in the cache once the data were retrieved. This allowed one to save memory space and reduce the required time for ZT Scan DIA 2.0 data analysis. Although this environment is currently available in SCIEX WIFF2 format, the dynamic MS2 loader will become available for all MS vendor data in the future. Importantly, the system is available not only for ZT Scan DIA 2.0, but also for other DIA methods, such as SWATH-DIA, in addition to the DDA method.

Q1Dec: Q1-axis information-based filtering

After the peak picking process was performed to recognize peak features based on the RT and MS1 *m/z* dimension(RT-mz), the Q1Dec function was performed for each RT-mz feature. The MS2 spectral data having the closest precursor *m/z* value (Q1-bin, typically 1.0–1.5 Da step-size) to the actual *m/z* property of the RT-mz feature was retrieved. In addition, the MS2 spectra of the two neighboring Q1 bins from the central Q1 bin were retrieved, and each product ion trace was evaluated. If a product ion trace had a local maximum in the center Q1 bin or one neighboring bin from the center, the MS2 peak was maintained or otherwise excluded.

RTDec: RT-axis information-based MS2 deconvolution

The MS2 spectrum constructed using the Q1Dec function was used as the starting spectrum for the RTDec process. Furthermore, users could set two thresholds before starting the deconvolution process: the MS2 absolute peak height cutoff (default zero) and the relative MS2 peak height cutoff (default zero). These were set to 100 and 0.0%, respectively, for hydrophilic metabolomics data, whereas the thresholds of the MS2 absolute peak height were determined using the MS2 peak heights of barcode (noise) ions. This

resulted in threshold values of 50, 100, and 160 for DDA, SWATH-DIA, and ZT Scan DIA 2.0, respectively, for untargeted lipidomics data. The MS2 relative abundance cut-off was set at 0.0%. These thresholds are useful for reducing the processing time for deconvolution via ZT Scan DIA 2.0 and clean-up of noise ions whose abundance typically increases when a short accumulation time (higher scanning speed for example) is used, as performed in ZT Scan DIA 2.0. Filtering was applied to the Q1Dec deconvoluted MS2 spectrum.

By using product ion values in the starter MS2 spectrum with a user-defined binning value (0.05 Da in this study), the MS2 chromatograms are extracted from the Q1 bin matched with the precursor *m/z* of interest. The RT range was set as the peak maximum plus or minus twice the peak width. The computational workflow, including MS2 peak quality estimation, MS2 model chromatogram construction, and the least squares method to provide the deconvoluted MS2 chromatograms, were the same as those of the MS2Dec algorithm described in a previous study.

MS2 feature extractions based on MS-DIAL lipid in silico MS/MS templates or MS2 correlation matrix

DIA techniques provide an environment for the use of MS2 chromatogram data for peak annotation and quantification of co-eluted structural isomers. In this study, we developed several functions to utilize MS2 spectra for MS-DIAL (**Supplementary Fig. 2 and Supplementary Fig. 5**). In lipidomics, the rt-*m/z* peak features satisfying the following three prerequisites are expanded into the MS2 dimension for further feature extraction. First, the peak must be detected for all samples. The gap-filled peaks in the peak alignment process are beyond the scope of this study. Second, the lipids of the second or later candidates must be annotated at the molecular species level, where up to three co-eluted isomers (candidates) are considered in this study. Finally, the MS2 data should be available for all samples, although this requisite is always true for DIA. The function to provide the peak features in the rt-*m/z*-MS2 dimensions was applied for the peak alignment results. Furthermore, a function to determine the unique MS2 *m/z* value (unique mass) for isomer separation and quantification was developed. For this function, the *m/z* value, which is not used for annotation in other isomer candidates, is automatically determined, while users define their own unique mass in the MS-DIAL graphical user interface. If the annotation is performed in experimental MS2 spectral libraries such as MassBank, MoNA, and NIST, the product ion that is contained as a minor peak with less than 1% relative intensity in the other candidates is also considered a candidate for a unique mass. Only one unique mass value is currently used in the MS-DIAL program.

In contrast to the above environments, where annotations are essential and unique mass determination is based on the reference MS2 spectra, MS-DIAL offers a different method (MS2 correlation matrix) to extract unique metabolite features that contain both known and unknown peaks (**Supplementary Fig.5**). Dot-product similarity scores were calculated for all sample pairs in an aligned peak, where users could set their own thresholds to define the number of unique components. In this function, the unique mass is defined from the experimental MS2 spectra, where the intensity values are summed within a 0.5-Da bin, and the product ion peak with more than 70% relative peak intensity is recognized as a unique mass candidate as the default setting. If unique mass values were shared among some components, the components were treated as a component group. This was similar to the protein

414 grouping algorithm performed in the conventional shotgun proteomics data analysis pipeline. The feature
415 and its application will be detailed elsewhere.
416
417 *Benchmark data acquisitions to compare DDA, SWATH-DIA, and ZT Scan DIA 2.0 technologies*
418 Human plasma samples (NIST SRM 1950) were purchased from the National Institute of Standards and
419 Technology (NIST). The Bligh and Dyer technique was used for lipid extraction. Briefly, 10 μL of blood
420 sample was mixed with 1,000 μL of an ice-cold MeOH/$CHCl_3$/H_2O (10:4:4, v/v/v) solvent. Solvents were
421 vortexed for 10 s, centrifuged at 15,000 × g for 10 min at 4°C, and the lipid layer of the supernatant was
422 transferred to a clean tube. Supernatant was dried using a speedvac and resuspended into a final
423 concentration of 1 μL plasma per 10 μL of injection solvent. Hydrophilic metabolite extraction was
424 performed via methanol protein precipitation with four volumes of ice-cold methanol. The solvent was
425 vortexed for 10 s, and the solution was centrifuged at 15,000 × g for 10 min. Supernatant was dried using
426 a speedvac and resuspended in water to a final concentration of 1 μL extract equivalent to 0.1 μL plasma.
427 Three technical replicates were used for each experiment.
428 LC-MS/MS analysis was performed using an Agilent Integrated System for LC comprising a
429 G7167B autosampler, G7120A binary pump, G7116B column management system, and ZenoTOF 8600
430 system for MS (SCIEX, Toronto, Canada). For hydrophilic metabolomics, the metabolite separation was
431 performed using the Kinetex F5 column (150 × 2.1 mm, 2.6 μm, Phenomenex Inc., CA, US) and the
432 following mobile phases: (A) 0.1% formic acid in H_2O and (B) 0.1% formic acid in acetonitrile (ACN).
433 The column was maintained at 30°C and a flow rate of 0.2 mL/min. A sample volume of 3 μL was used
434 for the injection. The separation was performed using the following gradient: 0 min 0.0% (B), 2.0 min
435 0.0% (B), 14.0 min 95% (B), 16.0 min 95% (B), 16.1 min 0% (B), and 23.0 min 0% (B). The common
436 MS conditions were as follows: ion source gas 1, 40 psi; ion source gas 2, 60 psi; curtain gas, 40 psi;
437 charged aerosol detector (CAD) gas, 7; source temperature, 500°C; spray voltage, 4000 V; *m/z* range of
438 time-of-flight (TOF)-MS, 70–1000 Da; *m/z* range of TOF MS/MS, 50–1000 Da; QJet declustering
439 potential, 20 V; collision energy, 30 eV; Q1 resolution, unit; MS1 accumulation time, 100 ms. The
440 following conditions were used for ZT Scan DIA 2.0: Q1 width, 7.2 Da; Q1 speed, 1390 Da/s; and Q1
441 precursor range, 60–1000. For SWATH DIA, the precursor isolation window was set to 50 with a 1-Da
442 overlap, and the Q1 precursor range was set to 80–950 to match the cycle time, as used in the ZT Scan
443 DIA 2.0. For DDA, the conditions of the top 50 precursors per survey were used.
444 For lipidomics, the lipid separation was carried out with the column of Luna C18 (150 × 2.0 mm,
445 2.1 μm, Phenomenex Inc., CA, US) and the mobile phases of (A) ACN/isopropyl alcohol (IPA)/H_2O (3:2:5,
446 v/v/v) and (B) ACN/IPA (1:9, v/v) where both contained 10 mM ammonium acetate. The column was
447 maintained at 55°C and a flow rate of 0.4 mL/min. A sample volume of 5 μL was used for the injection.
448 The separation was performed using the following gradient: 0 min 10.0% (B), 2.7 min 45.0% (B), 2.8 min
449 53% (B), 9 min 65% (B), 9.1 min 89% (B), 11.0 min 92% (B), 11.1 min 100% (B), 11.9 min 100% (B),
450 12.0 min 10% (B), and 15.0 min 10% (B). The common MS conditions were as follows: ion source gas 1,
451 40 psi; ion source gas 2, 60 psi; curtain gas, 40 psi; CAD gas, 7; source temperature, 350°C; spray voltage,
452 5500 V; *m/z* range of TOF-MS, 70–1000 Da; *m/z* range of TOF MS/MS, 50–1000 Da; QJet declustering

453 potential, 20 V; collision energy, 30 eV; Q1 resolution, unit; MS1 accumulation time, 100 ms; MS2
454 accumulation time, 5 ms.
455 The MS conditions for ZT Scan DIA were as follows: ion source gas 1, 50 psi; ion source gas 2,
456 70 psi; curtain gas, 40 psi; CAD gas, 7; source temperature, 250°C; spray voltage, 5500 V; QJet
457 declustering potential, 20 V; collision energy, 42 eV; collision energy spread 15 eV; time bins to sum
458 (MS1), 6; time bins to sum (MS2), 6; MS1 accumulation time, 100 ms; *m/z* range of TOF-MS, 200–
459 1250; *m/z* range of TOF-MS/MS, 100–1250; Q1 resolution, unit; zeno pulsing, enabled; zeno threshold,
460 2000000 cps; collision energy, 37 eV; collision energy spread, 0 eV; Q1 width, 6.8 Da; Q1 scan speed,
461 1015 Da/s; total MS/MS scan time, 1040 ms.
462
463 *MS-DIAL parameters*
464 MS-DIAL was built as version "5.6" where dynamic MS2 loading is available when a WIFF2 file format
465 is imported. Importantly, we developed the "rawdatahander.dll" package independently to handle raw MS
466 data. The package uses a vendor-specific application programming interface to convert the profile
467 spectrum into a centroid spectrum. This means that the current rawdatahandler.dll always parses the
468 spectrum as a centroid. Given that the MS-DIAL program uses the rawdatahander package, MS-DIAL
469 always retrieves centroid mass spectra as long as users import the vendor's raw data. Therefore, "centroid"
470 was selected as the ms1 and ms2 data type in the MS-DIAL parameter settings. In the analysis file settings
471 window, acquisition type was set to "DDA," "SWATH," or "ZTScan" according to the data sets. The
472 following parameters were used: max number of isotope recognitions, 5; number of threads, 6; minimum
473 peak amplitude, 500; MS/MS absolute abundance cut-off, 50, 100, and 160 for DDA, SWATH-DIA, and
474 ZT Scan DIA 2.0 for untargeted lipidomics, respectively; 100 for hydrophilic metabolomics data; MS/MS
475 relative abundance cut-off, 0; and exclusion after precursor *m/z*, 0.5. For hydrophilic metabolomics, the
476 following annotation parameters were used: simple dot product, 600; weighted dot product, 600; reverse
477 dot product, 800; minimum matched peak number, three; and minimum matched peak percentage, 0%. In
478 lipidomics, the annotation parameters were set as follows: dot product, 150; weighted dot product, 150;
479 reverse dot product, 300 (default in MS-DIAL); minimum matched peak number, 1; and minimum
480 matched peak percentage, 0%. Default parameter values were used for the other parameters. All data were
481 analyzed using a Windows OS personal computer with the following system: processor, AMD EPYC 7452
482 32-Core processor with 2.35 GHz; installed RAM, 256 GB, graphics card, NVIDIA Quadro P4000 (8 GB),
483 and storage, 13.2 TB.
484
485 *Evaluation of dot product scores when compared with those of DDA*
486 Hydrophilic metabolomic data from NIST 1950 human plasma (technical replicates = 3) were used, and
487 three data acquisition methods (DDA, SWATH-DIA, and ZT Scan DIA) were utilized. In the alignment
488 results of the DDA data, the RT, *m/z*, and MS/MS spectral information of the detected peaks were saved
489 in NIST MSP format. The MSP file was imported into MS-DIAL as a spectral library. The RT tolerance
490 for annotation was set to 0.1 min, and the options of "use retention time for scoring" and "use retention
491 time for filtering" were ticked. The same parameter settings as those described above for the hydrophilic
492 metabolome analysis were used for the others. For the ZT Scan DIA data, four different mass spectral

deconvolution conditions were used: both Q1Dec and RTDec (labeled as ZT + Q1/RTDec), Q1Dec (labeled as ZT + Q1Dec), RTDec (labeled as ZT + RTDec), or no deconvolution (ZT). The SWATH-DIA data were analyzed in two ways: RTDec (labeled SW + RTDec) and no deconvolution (labeled SW). The DDA data were processed using the same MSP library. Peak picking and annotation results for each sample file were exported. A weighted dot-product score greater than zero was used as a representative dot-product score. Considering that MS-DIAL extracts a representative MS/MS spectrum from the samples in the alignment result, where the representative spectrum with the highest total score for annotation or the highest total ion (when the peak has no annotation information) is selected, one of the three samples should have an identical MS2 spectrum stored in the MSP library. Thus, the weighted dot product score having identical, i.e., "one" in DDA data analyses, was excluded to estimate the reproducibility of product ion intensities among analyses.

Evaluation of annotation rates and dot product scores when compared with those of the experimental MS2 spectral libraries

An in-house MS/MS spectral library containing 1,492,414 spectra of 259,277 compounds in positive ion mode was used. The spectra were obtained from the RIKEN Plasma, KI_GIAR zic-HILIC, BMDMS-NP, CASMI 2012, CASMI 2016, EMBL-MCF, EMBL-MCF 2.0 HRMS, Fiehn HILIC, GNPS, HCD natural product library, MassBank, MetaboBase, Metlin, NIST14, NIST20, Pathogen Box, PEP-NP, QiaoLab PGN, ReSpect, and Vaniya-Fiehn Natural Products libraries. The options "use retention time for scoring" and "use retention time for filtering" were unknown. The other parameters were the same as those used in "*Evaluation of dot product scores when compared with those of DDA*."

Evaluation of annotation rates in lipidomics

Lipid annotation was performed using a rule-based decision-tree algorithm optimized for each lipid subclass. The MS-DIAL version 5.6 series was used, in which a total of 160 lipid subclasses were managed. False hits were manually excluded based on the RT behavior of lipids in reversed-phase chromatography, according to a previous study. The lipids annotated as the adduct forms of $[M-H_2O+H]^+$, $[M+H]^+$, or $[M+NH_4]^+$ for the positive ion mode and $[M-H]^-$ and $[M+CH_3COO]^-$ for the negative ion mode were used to compare the annotation coverage among the analytical techniques.

Experimental design to showcase three-dimensional feature extractions in ZT Scan DIA

The animal experiments were conducted in accordance with the ethical protocol approved by the Tokyo University of Agriculture and Technology (R5-50). C57BL/6J male mice were purchased from SLC (Shizuoka, Japan). Mice at 11 weeks or two years of age were used for the experiments and sacrificed for tissue collection. All mice were initially maintained on CE-2 chow (CLEA Japan, Tokyo, Japan), and a subset of animals was switched to an HFD (D12492; Research Diets, New Jersey, USA) four weeks before sacrifice. The livers were harvested, and tissues were immediately frozen after dissection and stored at −80°C until lipid extraction. NIST SRM 1950 plasma was also analyzed as a benchmark.

Lipids were extracted on ice using a biphasic solvent system comprising ice-cold methanol (MeOH; FUJIFILM Wako Pure Chemical, Osaka, Japan), MTBE (Sigma-Aldrich, Tokyo, Japan), and

533 ultrapure water (UP-0090a-0U1; Organo, Tokyo, Japan). Unless otherwise stated, all solvents were kept
534 on ice throughout the procedure. An internal standard (IS) mixture containing EquiSPLASH LIPIDOMIX
535 (Avanti Polar Lipids, USA) and FA 16:0-d_3/FA 18:0-d_3 (SRL, Canada) was used for normalization. The
536 detailed composition of the IS mixture is provided in **Supplementary Data 2**. For plasma samples from
537 NIST SRM 1950 (20 μL, n = 3), 225 μL of MeOH containing internal standards was added. After
538 vortexing for 10 s, 750 μL of MTBE was added. The mixture was vortexed again, sonicated for 10 min,
539 shaken at 4°C for 30 min, and centrifuged at 16,000 × g for 3 min at 4°C. The upper organic phase (800 μL)
540 was collected, mixed with 154 μL of water, vortexed, and centrifuged again. A total of 560 μL of the upper
541 layer was collected, divided equally into four tubes, dried under vacuum, and stored at −80 °C. Frozen
542 mouse liver tissues were homogenized at 2500 rpm for 15 s using a multibead shocker equipped with a
543 metal cone (YASUI KIKAI, Japan). One milliliter of MeOH was added to the homogenate. An aliquot
544 corresponding to 5.0 mg of tissue was transferred and diluted to a final volume of 225 μL with MeOH
545 containing internal standards. Lipid extraction was performed as previously described. All dried lipid
546 extracts were reconstituted in 40 μL of MeOH, vortexed for 30 s, centrifuged at 16,000 × g for 3 min at
547 4°C, and transferred to LC-MS vials fitted with micro-inserts (Agilent Technologies, USA).
548 The LC-MS/MS analysis was performed using a Nexera XR UHPLC system (SIL-40C XR
549 autosampler, LC-40D XR binary pump, CTO-40C column oven, and SCL-40 system controller; Shimadzu,
550 Kyoto, Japan) coupled to a ZenoTOF 8600 mass spectrometer (SCIEX, Toronto, Canada). Lipids were
551 separated on an Acquity UPLC Peptide BEH C18 column (50 × 2.1 mm, 1.7 μm; Waters, Milford, MA,
552 USA) maintained at 45°C with a flow rate of 0.3 mL/min. The mobile phases consisted of (A)
553 ACN/methanol/water (1:1:3, v/v/v) with 5 mM ammonium acetate and 10 nM ethylenediaminetetraacetic
554 acid and (B) ACN/isopropanol (1:9, v/v) with the same additives (FUJIFILM Wako Pure Chemical, Osaka,
555 Japan). A sample volume of 2 μL was used for the injection. The separation was conducted under the
556 following gradient: 0 min 0.1% (B), 1 min 0.1% (B), 5 min 40% (B), 7.5 min 64% (B), 12 min 71% (B),
557 12.5 min 82.5% (B), 19 min 85% (B), 19.1 min 99.9% (B), 20.5 min 99.9% (B), 20.6 min 0.1% (B), and
558 25 min 0.1% (B). The temperature of each sample was maintained at 4°C.
559 The MS conditions were as follows: ion source gas 1, 50 psi; ion source gas 2, 70 psi; curtain gas,
560 40 psi; CAD gas, 7; source temperature, 250 °C; spray voltage, 5500 V/−4500 V; QJet declustering
561 potential, ±20 V; collision energy, 42 eV; collision energy spread 15 eV; time bins to sum (MS1), 4; time
562 bins to sum (MS2), 6; MS1 accumulation time, 100 ms; *m/z* range of TOF-MS, 200–1250; *m/z* range of
563 TOF-MS/MS, 100–1250; Q1 resolution, unit; zeno pulsing, enabled; zeno threshold, 2000000 cps. For
564 DDA, the following conditions were used: MS2 accumulation time, 25 ms; maximum candidate ion, 35;
565 Intensity threshold exceeds, 10 counts/s; Exclude former candidate ions, disabled; Inclusion/Exclusion
566 list, disabled. For SWATH-DIA, MS2 accumulation time, 17 ms; isolation window width, 20 Da. For ZT-
567 scan DIA, the following conditions were used: collision energy, ±37 eV; collision energy spread, 0 eV;
568 time bins to sum (MS1), 6; Q1 width, 6.8 Da; Q1 scan speed, 1015 Da/Sec; total MSMS scan time, 1040
569 ms.
570 Data analysis was performed using MS-DIAL 5.6 with the same parameter set as described in the
571 previous section. Data curation, reduction, and normalization were performed as previously described.
572 Briefly, the annotation results were manually curated, and lipids that did not satisfy the RT behaviors in

reverse-phase LC were excluded. The positive and negative ion mode data were integrated, where only one adduct form was retained; for example, the acetate adduct form was used for PC, whereas the protonated and sodium adduct forms of PC were excluded. The quantitative values of MS1 and MS2 were normalized to the MS1 peak heights of the internal standards. Although the normalized MS1 quantitative value reflects the concentrations of lipid molecules in pmol per mg of tissue, the MS2 quantitative value was simply divided by the MS1 peak height of the ISs and multiplied by the average MS1 peak height of the IS among samples to reflect the magnitude of the original intensities. These details are described in **Supplementary Data 2**.

Data visualization and statistical analyses

All data visualization, significance testing, and multivariate analyses were conducted using the R programming language. For principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA), variables were log-transformed and standardized by auto-scaling. In OPLS-DA, only the CE2 diet conditions in liver lipidomics dataset was used to compare young (2-month-old) and aged (24-month-old) mice. Variable importance in projection (VIP) values were calculated, and the top 40 variables were visualized as a heatmap.

For liver lipidomics data, statistical comparisons were performed for the following groups: CE2/M2 vs. CE2/M24, HFD/M2 vs. HFD/M24, CE2/M2 vs. HFD/M2, and CE2/M24 vs. HFD/M24. Significance testing was conducted using Student's t-test (two-sided, equal variance assumed; $N = 4$), with P values adjusted by the Benjamini–Hochberg method. Lipids with an adjusted P value ≤ 0.05 and a fold change ≥ 1.5 or ≤ 0.5 relative to control were defined as differentially expressed lipids (DELs).

Lipid molecule enrichment analysis was performed using the clusterProfiler package. First, a lipid metadata library (**Supplementary Data 3**) was generated from lipid molecule names in the liver lipidomics dataset. Lipid class names and acyl chain labels were stored in a long-format data object, which was provided as input to clusterProfiler with the DELs information to evaluate enrichment of lipid classes and acyl chains using over-representation analysis (ORA). Lipids belonging to enriched terms were selected for heatmap visualization according to the following criteria: (1) select one lipid molecule from each enriched term, prioritizing those with the highest mean abundance across all samples; (2) repeat selection from terms in order of P value from the enrichment analysis until a total of 10 lipids were chosen, prioritizing those with higher abundance in subsequent selections.

605 ## Supplementary Figures

606 **Supplementary Fig. 1. Overview of tandem mass spectrometry (MS/MS) data acquisition strategies**
607 **in liquid chromatography (LC)-MS/MS.** In conventional LC-MS/MS, full precursor ion scanning
608 (MS1) and tandem MS acquisition (MS2) are performed sequentially. In data-dependent acquisition
609 (DDA), precursor ions are isolated with narrow windows (typically <1 Da) to generate high-purity MS2
610 spectra. In contrast, data-independent acquisition (DIA), such as SWATH-DIA implemented on SCIEX
611 instruments, utilizes wider isolation windows ranging from 5–50 Da. SWATH-DIA applies a block-wise
612 acquisition strategy, pausing briefly (typically <1 ms) between isolation windows. ZT Scan DIA 2.0 differs
613 fundamentally by employing a continuous sliding precursor isolation strategy, scanning across the *m/z*
614 range with 5–20-Da windows and eliminating the need for pause times between window transitions.
615

616 **Supplementary Fig. 2. Summary of lipid name-centric deconvolution for distinguishing lipid**
617 **isomers using MS2 product ions.** Lipid candidates annotated at the molecular species level were
618 extracted across biological samples. Unique fragment ions specific to each lipid candidate were identified,
619 and their MS2 fragment signals from each sample were used as quantitative values for lipid feature
620 analysis.
621

622 **Supplementary Fig. 3. Orthogonal partial least squares discriminant analysis (OPLS-DA) and**
623 **variable importance in projection (VIP) used to capture age-related differences.** The OPLS-DA score
624 plots are shown in the left panel, where the x-axis represents the latent variable related to the response
625 variable (young vs. aged), and the y-axis represents the latent variable orthogonal to the response variable.
626 Based on VIP values, the top 40 lipids were visualized using heatmap analysis.
627

628 **Supplementary Fig. 4. Summary of dynamic MS2 data loading in new MS-DIAL program (version**
629 **5.6).** The approximate file sizes of DDA, SWATH-DIA, and ZT Scan DIA 2.0 under a 20–25 min LC
630 condition are shown, along with an example of the data storage structure. In SCIEX wiff.scan format files,
631 the spectra of the MS1 dimension are always stored in experiment (exp.) 1. From experiment 2 onward,
632 MS2 spectral data are stored. For DDA, if the "Top 10" setting is used, a total of 10 experiments are
633 generated, with each experiment storing the MS2 data. In SWATH-DIA and ZT Scan DIA 2.0, the number
634 of experiments generated and the storage of MS2 data are determined by the precursor window size and
635 the precursor scanning range. For example, in ZT Scan DIA 2.0, if the precursor isolation window is set
636 to 5 Da and the *m/z* scanning range is set to 100–1000, a total of 1001 experiments are generated (with Q1
637 bins spaced at 1 Da intervals). In earlier versions of MS-DIAL, even for datasets like ZT Scan DIA 2.0
638 containing massive amounts of MS2 spectral data, all spectra were loaded into memory at once. As a result,
639 data loading required more than one hour even when using 100% of CPU resources. To address this, we
640 developed an algorithm to dynamically retrieve only the necessary MS2 data on demand. With this
641 approach, MS1 data can be loaded within 5–10 seconds, enabling peak detection algorithms to start
642 immediately. The right panel illustrates the concept of dynamic MS2 reading. After peak detection is

643 performed along the RT and *m/z* axes, MS2 feature searching is conducted only in the vicinity of the
644 detected peaks.
645
646 **Supplementary Fig. 5. Summary of spectrum-centric deconvolution strategy.**
647
648

Data availability

All raw MS data are available in the MB-POST repository (https://repository.massbank.jp/) under index number MPST000090.2. Although the processed table data used to create the figures were submitted as source data, the original peak lists and alignment tables were deposited in MB-POST together with the raw data. Lipidomic data from SRM 1950 human plasma and mouse liver tissues are available in Supplementary Data 1 and Supplementary Data 2, respectively. The table data for lipid enrichment analysis is available in Supplementary Data 3. The lipidomics minimal reporting checklist for liver lipidomics is available in Supplementary Note 1.

Code availability

MS-DIAL source code is available at https://github.com/systemsomicslab/MsdialWorkbench.

Acknowledgments

This research was supported by the Japan Agency for Medical Research and Development (AMED) under Infectious Diseases Research and Infrastructure (JP25wm0325071, H.T.), the Japan Science and Technology Agency (JST) Exploratory Research for Advanced Technology (ERATO) (JPMJER2101 to H. T.), JST FOREST program (JPMJFR230H to H.T.), JST NBDC (JPMJND2305 to H.T.), the JSPS KAKENHI (24K02011, 24H00043, 24H00392, 24K21269, 25H01425, and 25H01426 to H.T.), the Inserm project (international research project AtypicoLipid to T.H. and H.T.), and Technologically Advanced research through Marriage of Agriculture and engineering as Groundbreaking Organization (TAMAGO to H.T. and J.M.).

Author Contributions

L.D. and H.T. designed the study. T.H. and H.T. organized the lipidomics experiments and the relevant data analyses. Y.M., Mikiko T., and H.T. developed the MS-DIAL environment. K.T., R.Y., M.H., and J.M. performed biological experiments. K.T., B.B., T.O., T.H., and H.T. analyzed the data. Manami T., U.T., G.I., D.C., P.B., A.C., N.B., and L.D. organized the ZT Scan DIA instruments and performed the LC-MS/MS analyses. H.T. wrote the manuscript, and G.I. and T.H. improved the draft. All authors have thoroughly discussed this project and helped improve the manuscript.

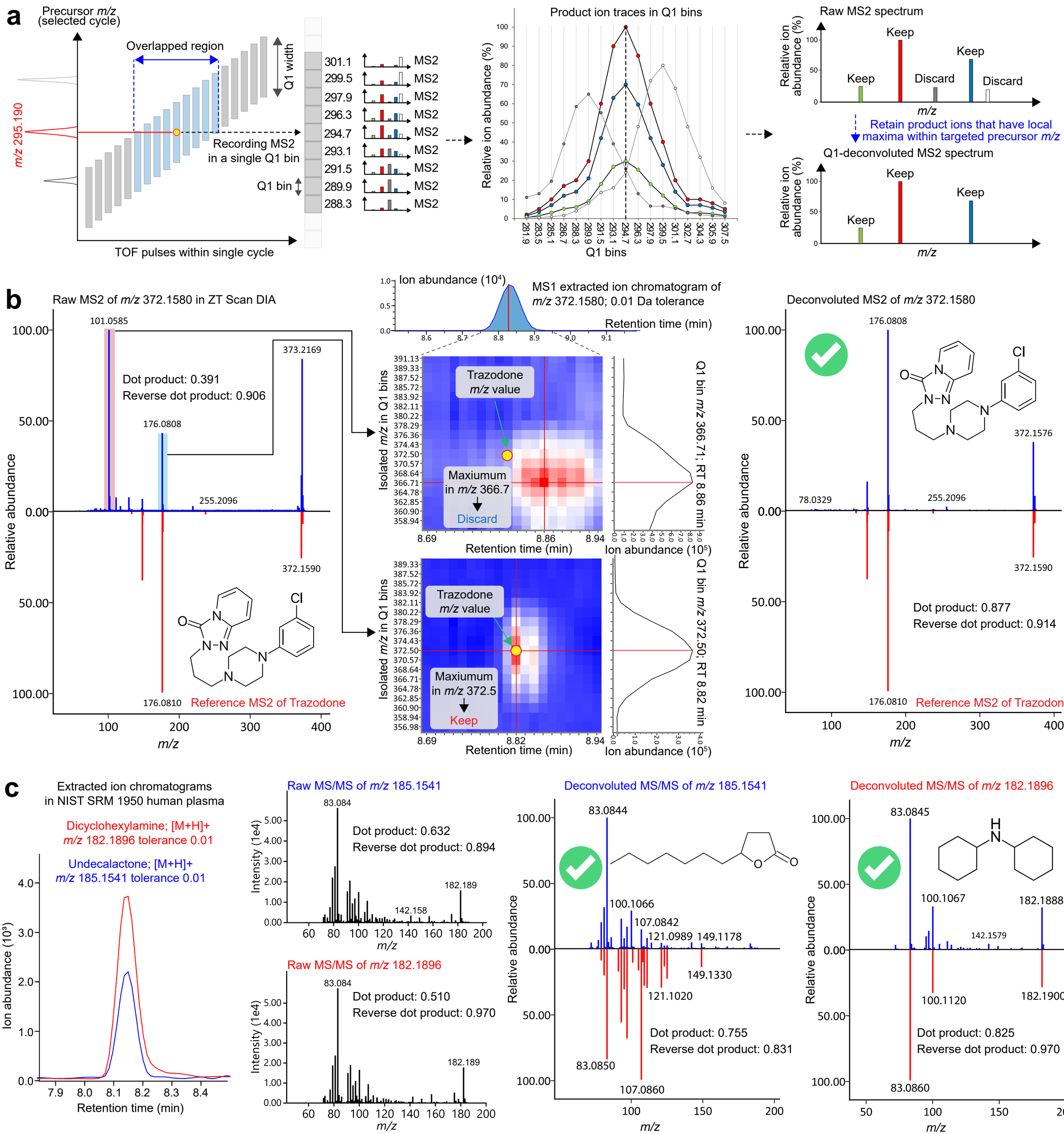
Competing interests

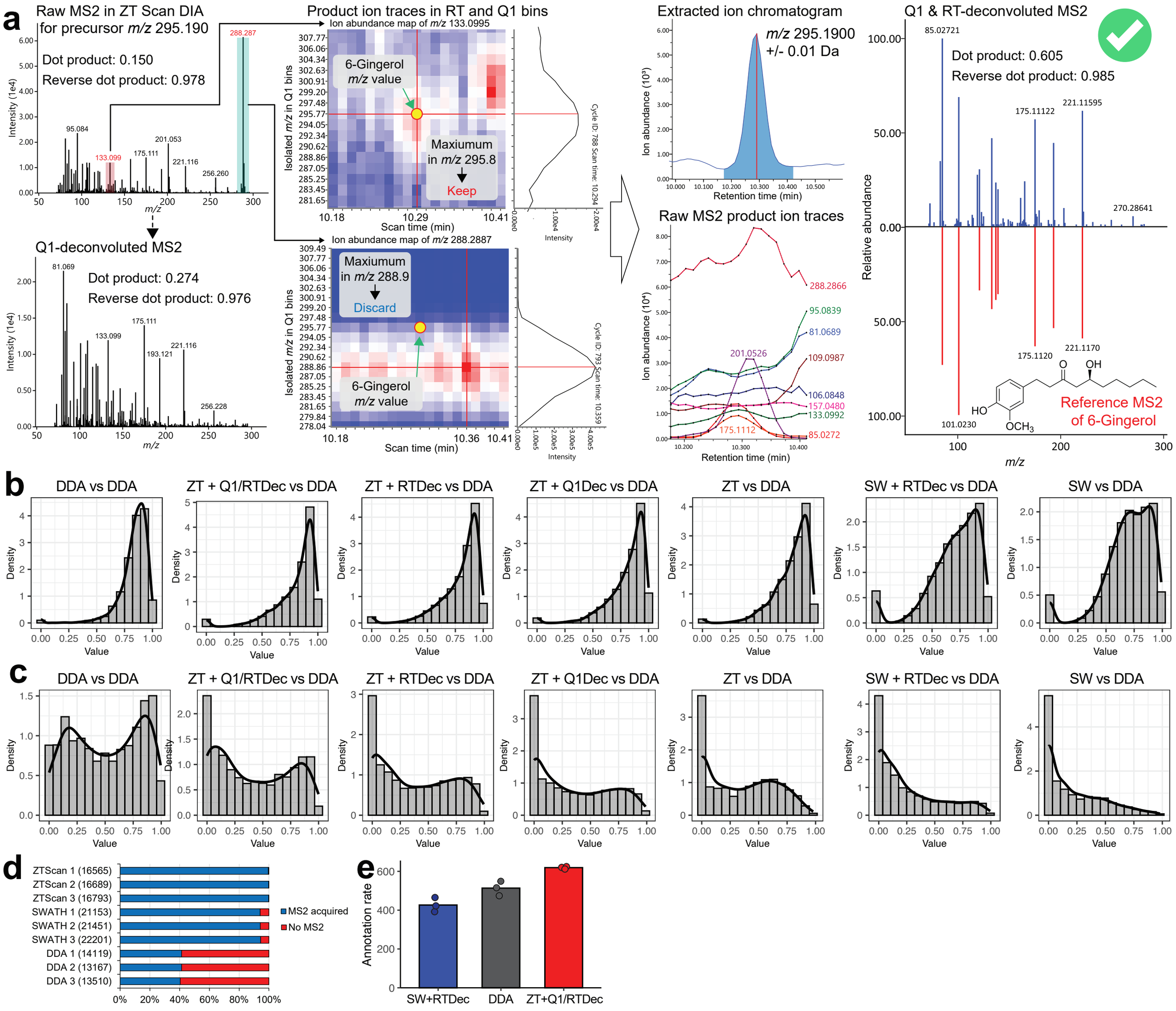
Manami.T. and U.T. are research scientists at AB Sciex, Japan. G.I., D.C., P.B., A.C., N.B. and L.D. are research scientists at SCIEX, Canada. The other authors declare no competing interests.

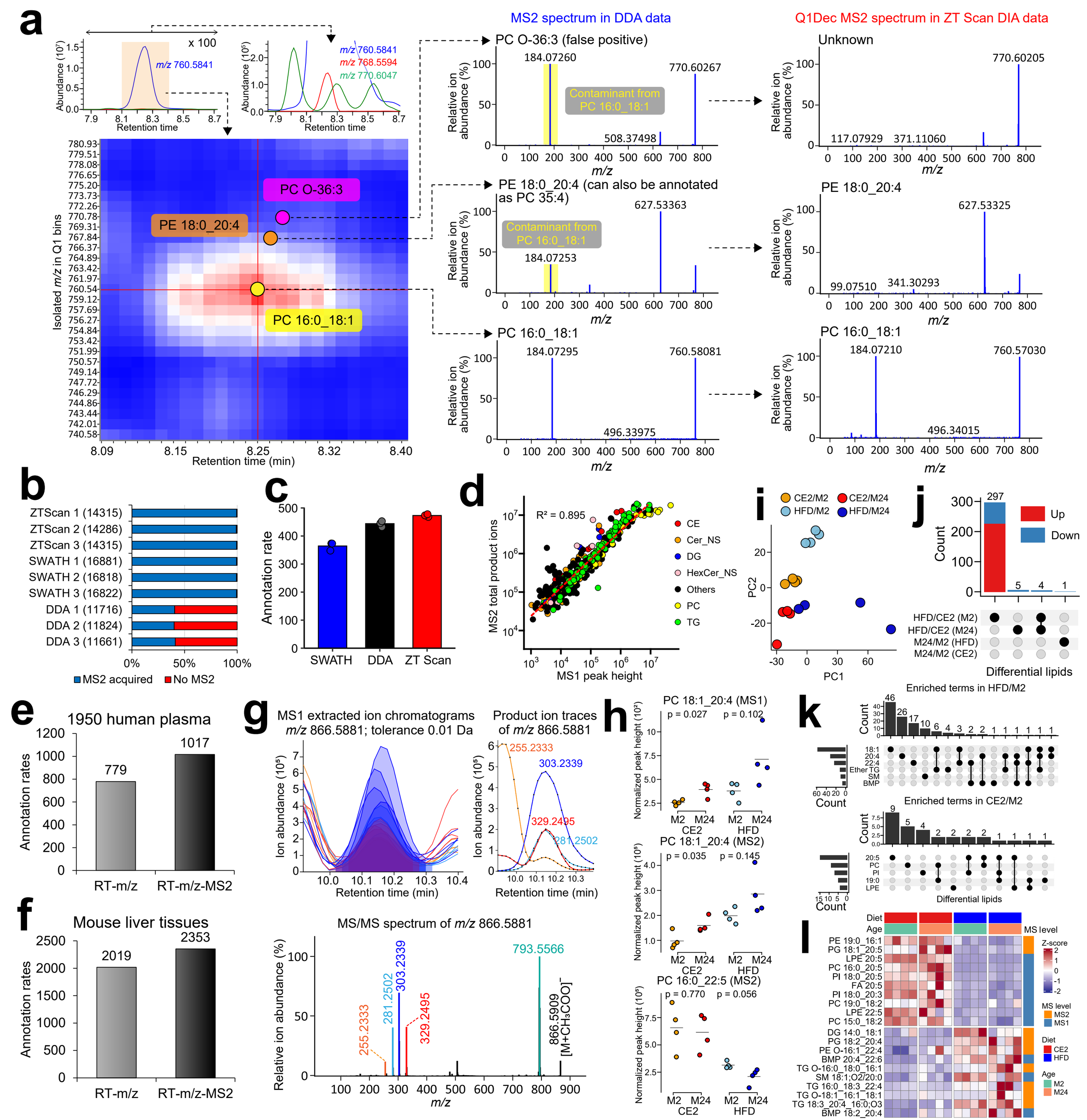
References

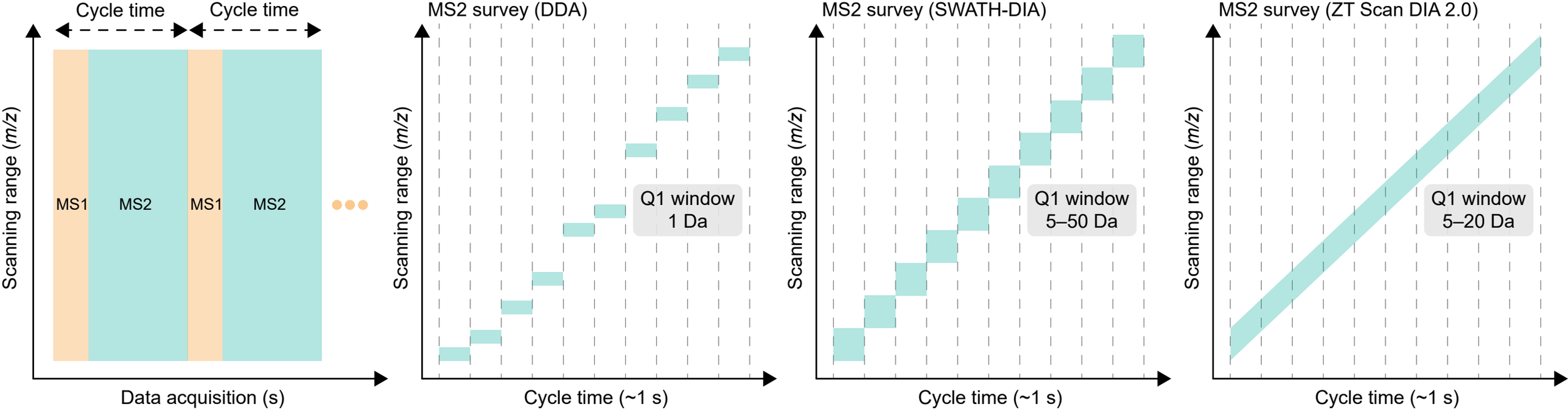
1. Alseekh, S. et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods* **18**, 747-756 (2021).
2. Hu, X. et al. A scalable workflow to characterize the human exposome. *Nat Commun* **12** (2021).
3. Ni, Z.X. et al. Guiding the choice of informatics software and tools for lipidomics research applications. *Nat Methods* **20**, 193-204 (2023).
4. Wishart, D.S. et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res* **50**, D622-D631 (2022).
5. Ara, T. et al. DDBJ update in 2023: the MetaboBank for metabolomics data and associated metadata. *Nucleic Acids Res* **52**, D67-D71 (2023).
6. Sud, M. et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res* **44**, D463-470 (2016).
7. Yurekten, O. et al. MetaboLights: open data repository for metabolomics. *Nucleic Acids Res* **52**, D640-D646 (2023).
8. Bushuiev, R. et al. Self-supervised learning of molecular representations from millions of tandem mass spectra using DreaMS. *Nat Biotechnol* (2025).
9. El Abiead, Y. et al. Enabling pan-repository reanalysis for big data science of public metabolomics data. *Nat Commun* **16** (2025).
10. Guo, J. & Huan, T. Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography-Mass Spectrometry Based Untargeted Metabolomics. *Anal Chem* **92**, 8072-8080 (2020).
11. Tsugawa, H. et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **12**, 523-+ (2015).
12. Stancliffe, E., Schwaiger-Haber, M., Sindelar, M. & Patti, G.J. DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution. *Nat Methods* **18**, 779-787 (2021).
13. Yin, Y., Wang, R., Cai, Y., Wang, Z. & Zhu, Z.J. DecoMetDIA: Deconvolution of Multiplexed MS/MS Spectra for Metabolite Identification in SWATH-MS-Based Untargeted Metabolomics. *Anal Chem* **91**, 11897-11904 (2019).
14. Liu, Y.K., Xia, Y. & Zhang, W.P. Structural Lipidomics Enabled by Isomer-Resolved Tandem Mass Spectrometry. *Anal Chem* **97**, 4275-4286 (2025).
15. Messner, C.B. et al. Ultra-fast proteomics with Scanning SWATH. *Nat Biotechnol* **39**, 846-854 (2021).
16. Stein, S.E. & Scott, D.R. Optimization and Testing of Mass-Spectral Library Search Algorithms for Compound Identification. *J Am Soc Mass Spectr* **5**, 859-866 (1994).
17. Li, Y.Y. et al. Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. *Nat Methods* **18**, 1524-+ (2021).
18. Scheubert, K. et al. Significance estimation for large scale metabolomics annotations by spectral matching. *Nat Commun* **8** (2017).
19. Wallace, W.E., Ji, W., Tchekhovskoi, D.V., Phinney, K.W. & Stein, S.E. Mass Spectral Library Quality Assurance by Inter-Library Comparison. *J Am Soc Mass Spectrom* **28**, 733-738 (2017).
20. Broeckling, C.D. et al. Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal Chem* **88**, 9226-9234 (2016).
21. Tsugawa, H. et al. A lipidome landscape of aging in mice. *Nat Aging* **4**, 709-726 (2024).

22. Tsugawa, H. et al. A lipidome atlas in MS-DIAL 4. *Nat Biotechnol* **38**, 1159-+ (2020).
23. Burla, B. et al. MS-based lipidomics of human blood plasma: a community-initiated position paper to develop accepted guidelines. *J Lipid Res* **59**, 2001-2017 (2018).
24. Aldana, J., Romero-Otero, A. & Cala, M.P. Exploring the Lipidome: Current Lipid Extraction Techniques for Mass Spectrometry Analysis. *Metabolites* **10** (2020).
25. Alvarez, D. et al. Effects of a High-Fat Diet and Docosahexaenoic Acid during Pregnancy on Fatty Acid Composition in the Fetal Livers of Mice. *Nutrients* **15** (2023).
26. Wang, X.H., Li, C.Y., Muhammad, I. & Zhang, X.Y. Fatty acid composition in serum correlates with that in the liver and non-alcoholic fatty liver disease activity scores in mice fed a high-fat diet. *Environ Toxicol Phar* **44**, 140-150 (2016).
27. Uchino, H., Tsugawa, H., Takahashi, H. & Arita, M. Computational mass spectrometry accelerates C = C position-resolved untargeted lipidomics using oxygen attachment dissociation. *Commun Chem* **5** (2022).
28. Ferrucci, L. & Fabbri, E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat Rev Cardiol* **15**, 505-522 (2018).


a
Precursor m/z (selected cycle)
Overlapped region
Q1 width
m/z 295.190
Recording MS2 in a single Q1 bin
Q1 bin
TOF pulses within single cycle
301.1
299.5
297.9
296.3
294.7
293.1
291.5
289.9
288.3
MS2
Product ion traces in Q1 bins
Relative ion abundance (%)
Q1 bins
Raw MS2 spectrum
Keep
Discard
m/z
Retain product ions that have local maxima within targeted precursor m/z
Q1-deconvoluted MS2 spectrum
b
Raw MS2 of m/z 372.1580 in ZT Scan DIA
101.0585
373.2169
Dot product: 0.391
Reverse dot product: 0.906
176.0808
255.2096
372.1590
176.0810
Reference MS2 of Trazodone
Relative abundance
Ion abundance (10⁴)
MS1 extracted ion chromatogram of m/z 372.1580; 0.01 Da tolerance
Retention time (min)
Isolated m/z in Q1 bins
Trazodone m/z value
Maxiumum in m/z 366.7
Discard
Q1 bin m/z 366.71; RT 8.86 min
Ion abundance (10⁵)
Maxiumum in m/z 372.5
Keep
Q1 bin m/z 372.50; RT 8.82 min
Deconvoluted MS2 of m/z 372.1580
176.0808
372.1576
78.0329
255.2096
372.1590
Dot product: 0.877
Reverse dot product: 0.914
176.0810
Reference MS2 of Trazodone
c
Extracted ion chromatograms in NIST SRM 1950 human plasma
Dicyclohexylamine; [M+H]+ m/z 182.1896 tolerance 0.01
Undecalactone; [M+H]+ m/z 185.1541 tolerance 0.01
Ion abundance (10³)
Retention time (min)
Raw MS/MS of m/z 185.1541
83.084
Dot product: 0.632
Reverse dot product: 0.894
182.189
142.158
Intensity (1e4)
Raw MS/MS of m/z 182.1896
83.084
Dot product: 0.510
Reverse dot product: 0.970
182.189
Deconvoluted MS/MS of m/z 185.1541
83.0844
100.1066
107.0842
121.0989
149.1178
149.1330
121.1020
Dot product: 0.755
Reverse dot product: 0.831
83.0850
107.0860
Deconvoluted MS/MS of m/z 182.1896
83.0845
100.1067
182.1888
142.1579
100.1120
182.1900
Dot product: 0.825
Reverse dot product: 0.970
83.0860


a
Raw MS2 in ZT Scan DIA for precursor m/z 295.190
Dot product: 0.150
Reverse dot product: 0.978
Q1-deconvoluted MS2
Dot product: 0.274
Reverse dot product: 0.976
Product ion traces in RT and Q1 bins
Ion abundance map of m/z 133.0995
6-Gingerol m/z value
Maxiumum in m/z 295.8
Keep
Ion abundance map of m/z 288.2887
Maxiumum in m/z 288.9
Discard
Extracted ion chromatogram
m/z 295.1900 +/- 0.01 Da
Raw MS2 product ion traces
Q1 & RT-deconvoluted MS2
Dot product: 0.605
Reverse dot product: 0.985
Reference MS2 of 6-Gingerol
b
c
DDA vs DDA
ZT + Q1/RTDec vs DDA
ZT + RTDec vs DDA
ZT + Q1Dec vs DDA
ZT vs DDA
SW + RTDec vs DDA
SW vs DDA
d
ZTScan 1 (16565)
ZTScan 2 (16689)
ZTScan 3 (16793)
SWATH 1 (21153)
SWATH 2 (21451)
SWATH 3 (22201)
DDA 1 (14119)
DDA 2 (13167)
DDA 3 (13510)
MS2 acquired
No MS2
e
Annotation rate
SW+RTDec
DDA
ZT+Q1/RTDec


a
x 100
m/z 760.5841
m/z 768.5594
m/z 770.6047
Abundance
Retention time
Isolated m/z in Q1 bins
PC O-36:3
PE 18:0_20:4
PC 16:0_18:1
Retention time (min)
MS2 spectrum in DDA data
Q1Dec MS2 spectrum in ZT Scan DIA data
PC O-36:3 (false positive)
184.07260
770.60267
508.37498
Contaminant from PC 16:0_18:1
Unknown
770.60205
117.07929
371.11060
PE 18:0_20:4 (can also be annotated as PC 35:4)
627.53363
184.07253
PE 18:0_20:4
627.53325
99.07510
341.30293
PC 16:0_18:1
184.07295
760.58081
496.33975
PC 16:0_18:1
184.07210
760.57030
496.34015
Relative ion abundance (%)
m/z
b
ZTScan 1 (14315)
ZTScan 2 (14286)
ZTScan 3 (14315)
SWATH 1 (16881)
SWATH 2 (16818)
SWATH 3 (16822)
DDA 1 (11716)
DDA 2 (11824)
DDA 3 (11661)
MS2 acquired
No MS2
c
Annotation rate
SWATH
DDA
ZT Scan
d
R² = 0.895
MS2 total product ions
MS1 peak height
CE
Cer_NS
DG
HexCer_NS
Others
PC
TG
i
CE2/M2
CE2/M24
HFD/M2
HFD/M24
PC1
PC2
j
Count
Up
Down
297
5
4
1
HFD/CE2 (M2)
HFD/CE2 (M24)
M24/M2 (HFD)
M24/M2 (CE2)
Differential lipids
e
1950 human plasma
Annotation rates
779
1017
RT-m/z
RT-m/z-MS2
f
Mouse liver tissues
2019
2353
g
MS1 extracted ion chromatograms
m/z 866.5881; tolerance 0.01 Da
Ion abundance
Retention time (min)
Product ion traces of m/z 866.5881
255.2333
303.2339
329.2495
281.2502
MS/MS spectrum of m/z 866.5881
793.5566
866.5909 [M+CH3COO]-
h
PC 18:1_20:4 (MS1)
p = 0.027
p = 0.102
PC 18:1_20:4 (MS2)
p = 0.035
p = 0.145
PC 16:0_22:5 (MS2)
p = 0.770
p = 0.056
Normalized peak height
M2
M24
CE2
HFD
k
Enriched terms in HFD/M2
Enriched terms in CE2/M2
Count
Differential lipids
l
Diet
Age
MS level
Z-score
PE 19:0_16:1
PG 18:1_20:5
LPE 20:5
PC 16:0_20:5
PI 18:0_20:5
FA 20:5
PI 18:0_20:3
PC 19:0_18:2
LPE 22:5
PC 15:0_18:2
DG 14:0_18:1
PG 18:2_20:4
PE O-16:1_22:4
BMP 20:4_22:6
TG O-16:0_18:0_16:1
SM 18:1;O2/20:0
TG 16:0_18:3_22:4
TG O-18:1_16:1_18:1
TG 18:3_20:4_16:0;O3
BMP 18:2_20:4
MS2
MS1
CE2
HFD
M2
M24


Cycle time
Cycle time
MS1
MS2
MS1
MS2
Scanning range (m/z)
Data acquisition (s)
MS2 survey (DDA)
Q1 window
1 Da
Scanning range (m/z)
Cycle time (~1 s)
MS2 survey (SWATH-DIA)
Q1 window
5–50 Da
Scanning range (m/z)
Cycle time (~1 s)
MS2 survey (ZT Scan DIA 2.0)
Q1 window
5–20 Da
Scanning range (m/z)
Cycle time (~1 s)

MS-DIAL function for lipid name-centric deconvolution

Sample #	Annotation
1	PC 16:1_22:5
2	PC 18:2_20:4 or PC 16:1_22:5
3	PC 18:2_20:4 or PC 16:1_22:5
4	PC 18:2_20:4 or PC 16:1_22:5
5	PC 18:1_20:5
6	PC 18:1_20:5

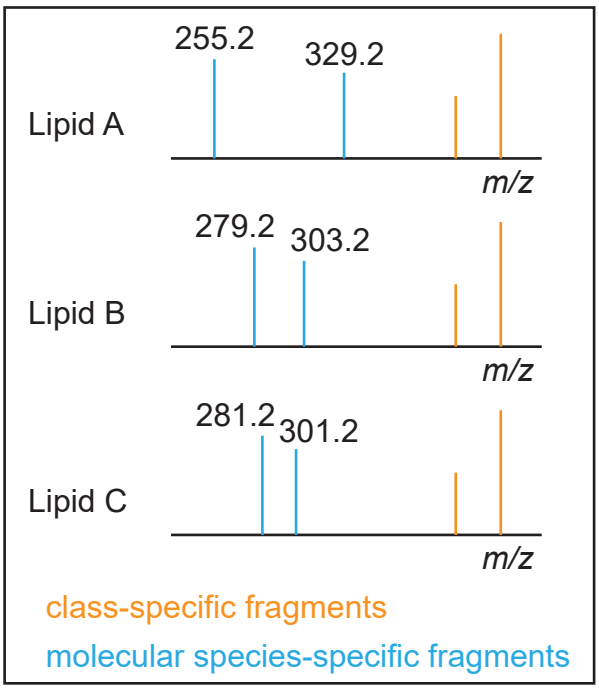
Extract unique annotations

A. PC 16:1_22:5

B. PC 18:2_20:4

C. PC 18:1_20:5

Reference spectra



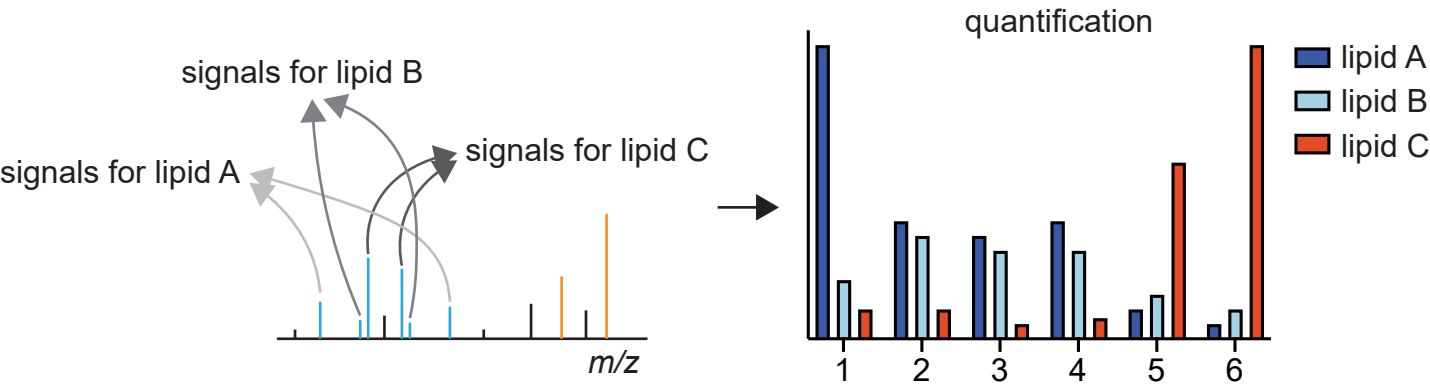
Extract molecular species-specific fragments

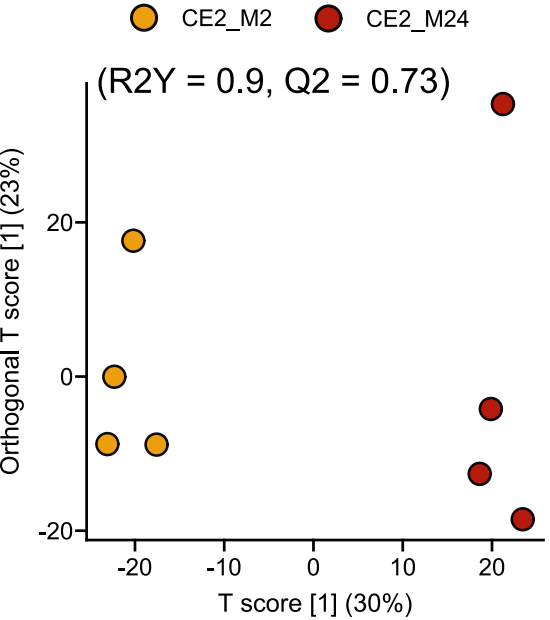
A. 255.2, 329.2

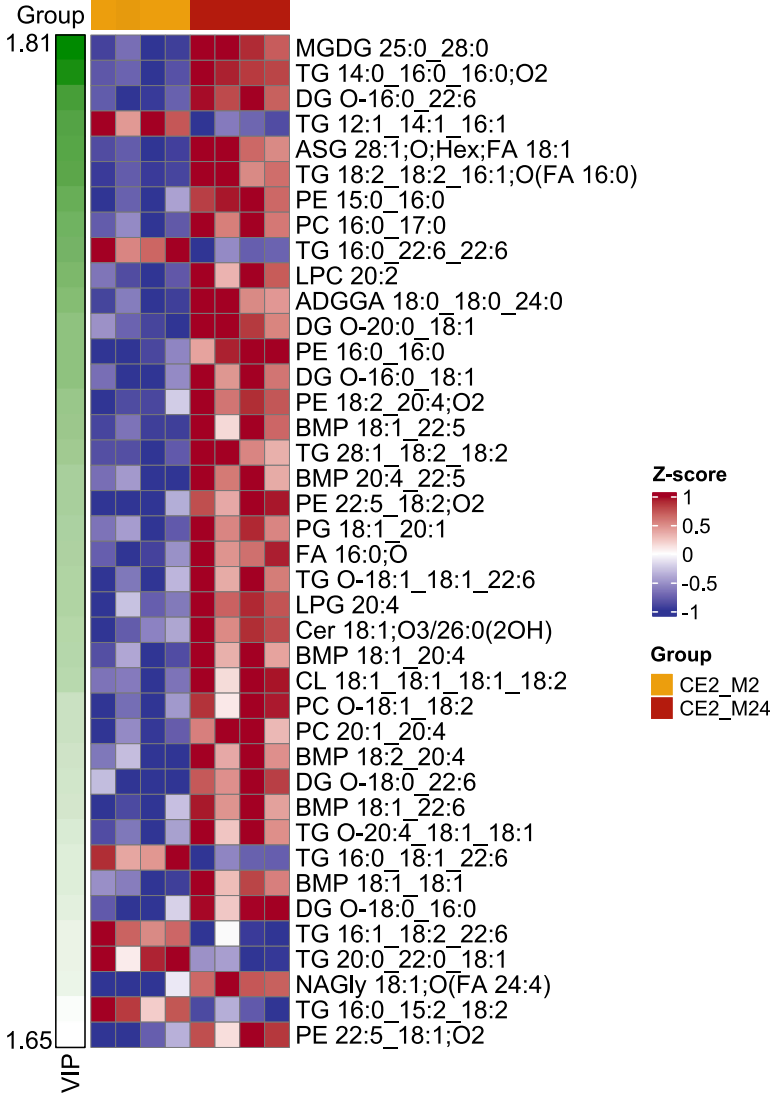
B. 279.3, 303.2

C. 281.2, 301.2

Extract MS2 fragment signals from sample-derived spectra



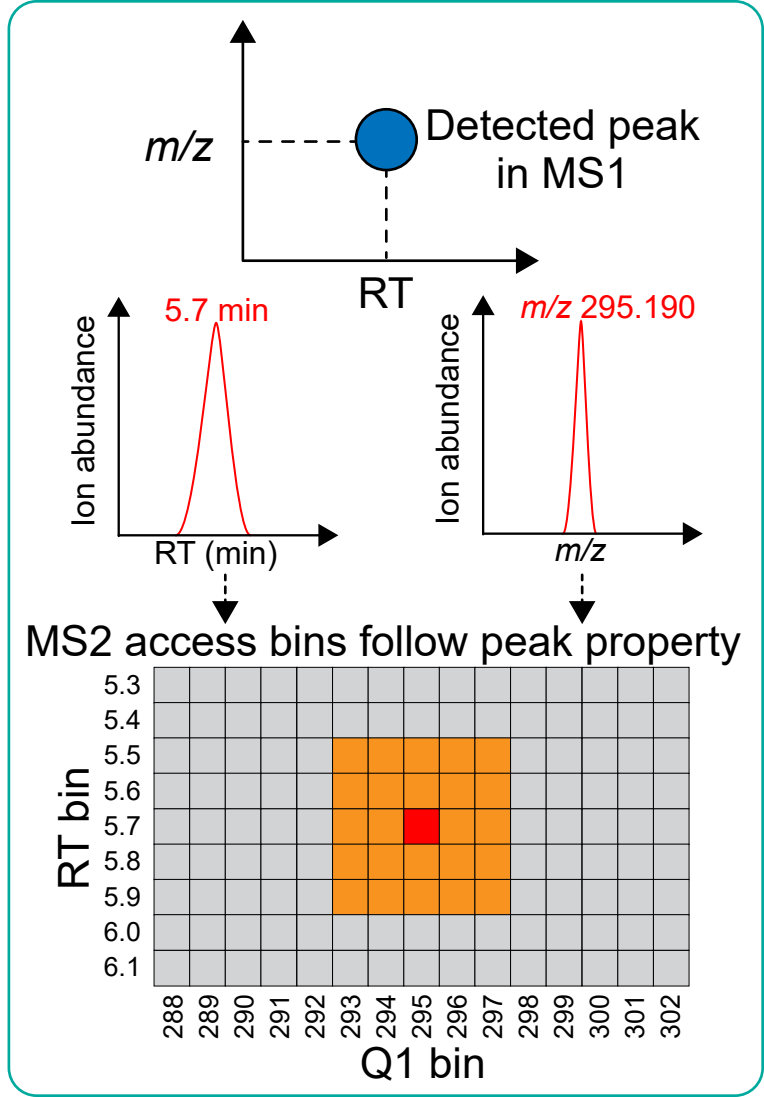

CE2_M2
CE2_M24
(R2Y = 0.9, Q2 = 0.73)
Orthogonal T score [1] (23%)
T score [1] (30%)
20
0
-20
-20
-10
0
10
20


Group
1.81
1.65
VIP
MGDG 25:0_28:0
TG 14:0_16:0_16:0;O2
DG O-16:0_22:6
TG 12:1_14:1_16:1
ASG 28:1;O;Hex;FA 18:1
TG 18:2_18:2_16:1;O(FA 16:0)
PE 15:0_16:0
PC 16:0_17:0
TG 16:0_22:6_22:6
LPC 20:2
ADGGA 18:0_18:0_24:0
DG O-20:0_18:1
PE 16:0_16:0
DG O-16:0_18:1
PE 18:2_20:4;O2
BMP 18:1_22:5
TG 28:1_18:2_18:2
BMP 20:4_22:5
PE 22:5_18:2;O2
PG 18:1_20:1
FA 16:0;O
TG O-18:1_18:1_22:6
LPG 20:4
Cer 18:1;O3/26:0(2OH)
BMP 18:1_20:4
CL 18:1_18:1_18:1_18:2
PC O-18:1_18:2
PC 20:1_20:4
BMP 18:2_20:4
DG O-18:0_22:6
BMP 18:1_22:6
TG O-20:4_18:1_18:1
TG 16:0_18:1_22:6
BMP 18:1_18:1
DG O-18:0_16:0
TG 16:1_18:2_22:6
TG 20:0_22:0_18:1
NAGly 18:1;O(FA 24:4)
TG 16:0_15:2_18:2
PE 22:5_18:1;O2
Z-score
1
0.5
0
-0.5
-1
Group
CE2_M2
CE2_M24

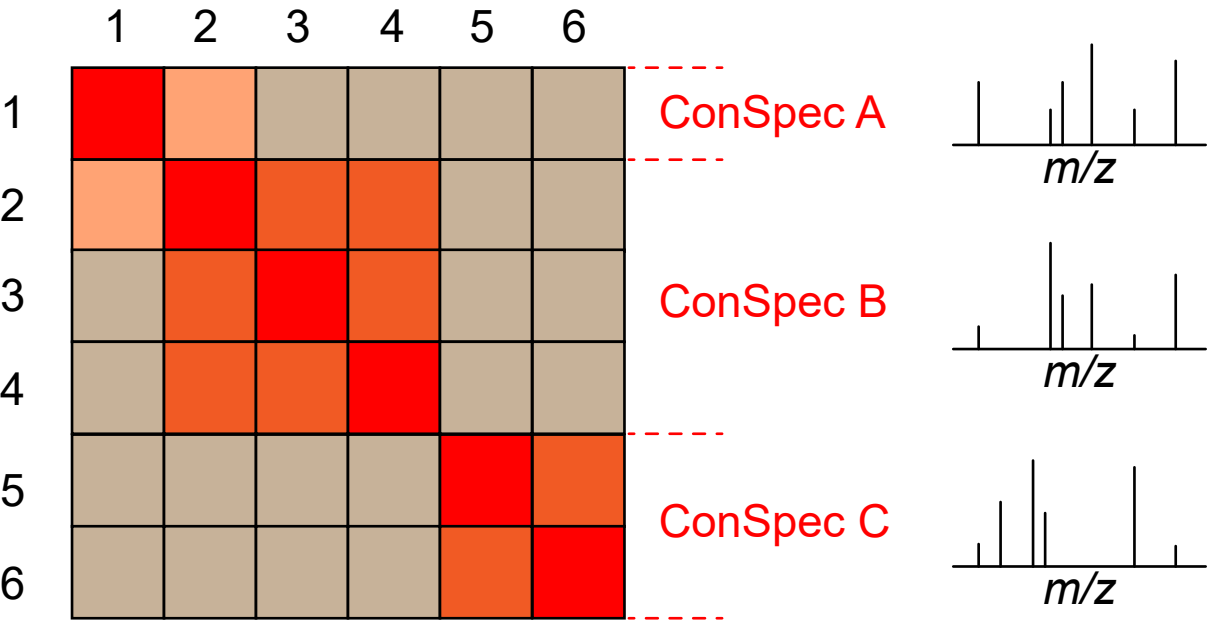
20–25 min LC-MS/MS

	Wiff.Scan DDA	Wiff.Scan SWATH	Wiff.Scan ZT scan
File size	50–100 MB	0.5–1 GB	10–30 GB
Data storage	1 MS1 exp. 3–20 MS2 exp.	1 MS1 exp. 50–100 MS2 exp.	1 MS1 exp. 500–1000 MS2 exp.
Parse time	<10 s with a single thread	<30 s with a single thread	~1 h with 100% CPU

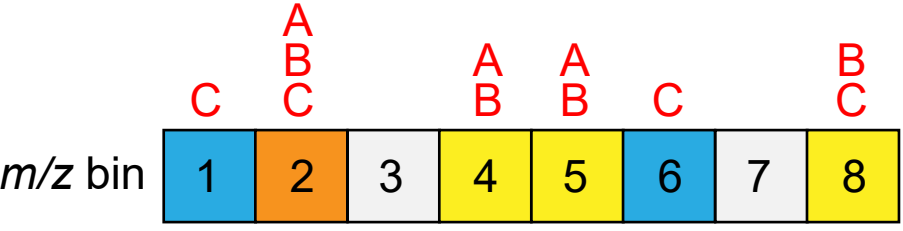
1. Only parsing MS1 exp for peak picking (<5 s)
2. MS2 called on-demand
   for deconvolution/annotation (save memory space)



Spectrum-centric deconvolution



ConSpec: Consensus spectrum
-Peak summed by 0.5 Da bin (default)
-Dot product > 0.8 (default)
-Shared by 100% samples of a group (default)
or *m/z* in Ref. spec (>0.5% rel.abs) (option)



Unique mass definition and MS2 quantification
-mapping *m/z* of 70% rel.abs in 0.5 Da bin (default)
-metabolites grouping (AB and C in the illustration)
-calculate MS2 peak height/area by unique masses