PAPER

Check for updates

Unsupervised machine learning for mass spectrometry imaging data analysis with *in vivo* isotope labeling

Raven L. Buckman Johnson, [ID] Vy T. Tat and Young Jin Lee [ID] *

Mass spectrometry imaging (MSI) has emerged as a powerful tool for spatial metabolomics, but untargeted data analysis has proven to be challenging. When combined with *in vivo* isotope labeling (MSI*i*), MSI provides insights into metabolic dynamics with high spatial resolution; however, the data analysis becomes even more complex. Although various tools exist for advanced MSI analyses, machine learning (ML) applications to MSI*i* have not been explored. In this study, we leverage Cardinal to process MSI*i* datasets of duckweeds labeled with either $^{13}CO_2$ or D_2O. We apply spatial shrunken centroid (SSC) segmentation, an unsupervised ML algorithm, to differentiate metabolite localizations and investigate isotope labeling of untargeted metabolites. In the SSC segmentation of three-day ^{13}C-labeled duckweed dataset, five spatial segments were identified based on distinct lipid isotopologue distributions, in contrast to classification of only three tissue regions in previous manual analysis based on galactolipid isotopologues. Similarly, SSC segmentation of five-day D-labeled dataset revealed five spatial segments based on distinct metabolite and isotopologue profiles. Further, this untargeted segmentation analysis of MSI*i* dataset provided insights on tissue-specific relative flux of each metabolite by calculating the fraction of *de novo* biosynthesis in each segment. Overall, the application of unsupervised machine learning to MSI*i* datasets has proven to significantly reduce analysis time, increase throughput, and improve the clarity of spatial isotopologue distributions.

Introduction

Ranging from atomic[1,2] to molecular[3,4] studies, mass spectrometry imaging (MSI) is most well-known for its ability to provide crucial insights into the spatial distributions of metabolites *via in situ* measurements.[4,5] Despite its broad applicability, MSI analyses are restricted to collecting information at fixed times which does not provide information about metabolic activity.[6] By introducing isotopically labeled tracers, such as heavy water (D_2O) or [U-^{13}C]glucose, to living organisms, metabolic dynamics can be monitored through their incorporation into downstream metabolites.[6–8] Typical modes of *in vivo* isotope tracing analysis require chromatographic separations; however, these methods are limited by their lack of spatial information.[6,9–11] Matrix-assisted laser desorption/ionization (MALDI)-MSI offers both high spatial and mass resolution, making it an ideal technique for *in situ* analysis of *in vivo* isotope labeled analytes.[12–16] By combining MSI with *in vivo* isotope labeling (referred to as MSI*i*), metabolic dynamics and

flux analysis can be probed in high spatial resolution with reliable accuracy.[17–19]

Recently, MSI*i* was adopted to reveal spatiotemporal dynamics in lipid biosynthesis during plant development.[14–16] Among those, MSI*i* of *Lemna minor* (duckweed) with deuterium (D)- and 13-carbon (^{13}C)-labeling showed three distinct binomial isotopologue distributions corresponding to the partial labeling of each galactolipid building block.[15] Interestingly, each isotopologue group demonstrated distinct localizations to the parent, intermediate, or daughter frond tissues, thus, revealing a spatial dependence in galactolipid biosynthesis.[15] While this study successfully demonstrated MSI*i*, data analysis was a major bottleneck. Tedious data processing to manually identify isotopologues and subsequently investigate analyte spatial distributions ultimately limited throughput to targeted metabolites.[15,17,20] MSI is considered high-dimensional data due to its large number of mass spectral and spatial features,[21–23] thus, it becomes more cumbersome when isotope labeling is added. While several software tools exist for MSI data analysis with a wide range of capabilities, each provides benefits and drawbacks that make MSI*i* analysis challenging. For example, MSiReader[24] incorporates percent isotope enrichment (PIE) that requires a manual region-of-interest

Department of Chemistry, Iowa State University, Ames, IA, USA.
E-mail: yjlee@iastate.edu

selection to analyze targeted analytes.[25–27] METASPACE,[28] a widely used MSI platform, is not capable of analyzing isotopologues. Meanwhile, IsoScope,[17] designed to work with MSI*i* data, is limited to targeted compounds. Therefore, additional tools are required to analyze isotope patterns imposed by *in vivo* isotope labeling in MSI*i* data and maintain high throughput.

Machine learning (ML) is a well-established and growing field of research that employs computer systems to adapt statistical models or algorithms without following explicit, user-defined instructions.[21,29] Several recent review articles highlighted numerous examples of ML algorithms and platforms for streamlined MSI data processing.[21,23,30,31] Additionally, a wide variety of ML applications and packages have been developed that include supervised,[32,33] unsupervised,[34,35] semi-supervised,[36] or deep learning algorithms.[37] Unsupervised learning techniques create models from uncategorized data and are allowed to explore patterns, associations, or structures in datasets without user intervention.[21,29,31,38] These algorithms are particularly useful for large, complicated datasets that contain patterns which the user may, or may not, be aware of. In contrast, supervised and semi-supervised ML algorithms rely on a set of data, categorized by the user, to generate models, and are used to predict the outcome of another dataset.[39,40] MSI*i* analysis could benefit from unsupervised machine learning to explore unexpected isotope labeling patterns in a rapid fashion; however, there have been no such applications so far.

Cardinal, an R-based statistical package developed to process and analyze MSI data, was first introduced by K. Bemis, *et al.* in 2015 and has quickly evolved into a robust platform for MSI analysis.[41–43] The built-in spatial shrunken centroid (SSC) algorithm in Cardinal can be used for supervised classification or unsupervised segmentation.[41–43] The unsupervised segmentation creates a subset of spectral features, uses probabilistic modeling and data reduction methods to compare subsets, then produces spatially segmented images.[42,43] SSC segmentation provides a unique opportunity to enhance detection of spatially distributed isotopologue groups in MSI*i* datasets without the requirement of *a priori* identification of metabolites.[42,43] In our previous study on MSI*i* of *L. minor*, we focused only on galactolipid biosynthesis and the corresponding spatiotemporal changes to isotopologue distributions due to the time-consuming manual analysis.[15] By developing an analysis workflow to apply SSC segmentation to ^{13}C- and D-labeled duckweed data in the current study, we improve throughput and reliability for a broader range of analytes in MSI*i* analysis.

Materials and methods

Sample preparation and data collection details were described in our previous publication.[15] Briefly, *L. minor* was propagated in 0.5x Schenk and Hildebrandt (SH) media with a 16/8 hr light cycle. To label duckweed plants with deuterium, healthy fronds were transferred to Petri dishes containing 50% D_2O, $H_2O : D_2O$ (50 : 50 mol : mol), in the single parent-frond stage and grown for five days. For ^{13}C labeling, fronds were grown in an Erlenmeyer flask system that allowed for the purging of $^{12}CO_2$ with CO_2-free air and replenishment of $^{13}CO_2$ by reacting $Ba^{13}CO_3$ with lactic acid. ^{13}C-labeled fronds were harvested after two and three days. Harvested fronds were separated into two halves along the longitudinal direction *via* a fracturing method.[44] Fracturing allows the internal mesophyll cell layers to be exposed for MALDI-MSI. The fractured duckweed samples were sprayed with 2,5-dihydroxybenzoic acid (DHB) as a matrix and sputtered with gold to produce a conductive surface. MSI was performed using an Orbitrap mass spectrometer (Q Exactive HF, Thermo Scientific, San Jose, CA, USA) equipped with a medium pressure MALDI source (Spectroglyph, Kennewick, WA, USA) at ~7 torr. Raw mass spectrometry files were aligned and converted to .imzML format with ImageInsight (ver. 0.1.0.1516, Spectroglyph).

Analysis of the unlabeled and labeled duckweed data was performed in R (ver. 4.4.1) with Cardinal (ver. 3.8.3).[43] With Cardinal, .imzML files were read into memory and subsequently sum normalized (referred to as total ion count, TIC) and peak picked with a relative tolerance of 2 or 3 mDa. Detailed explanation of the fundamentals behind the SSC segmentation were provided in the original paper by K. Bemis *et al.*,[42] and briefly summarized in SI 1. For all datasets, SSC segmentation was employed in a two-staged manner. In the first stage, images were partitioned into two segments (denoted as "tape" or "tissue"). The second stage, unsupervised segmentation on the tissue-only pixels was performed and the spatial distribution of isotopologues was evaluated. An example of the workflow followed in this study is available on GitHub (**https://github.com/buckm065/IsotopeLabelingCardinalAnalysis**). Computations were performed using 128 GB of RAM on Iowa State University's Nova cluster. Labeled MSI data from previously published work and used in this study is available through Figshare (**https://doi.org/10.25380/iastate.28540523.v1**), and unlabeled MSI data used in this study is available through METASPACE (**https://metaspace2020.org/project/duckweed-machinelearning**).

Results and discussion

Unlabeled duckweed

For effective MSI*i* data analysis, the removal of the background, prior to the analysis of tissue regions, should be performed. In Fig. 1A, an optical image of unlabeled duckweed sample is shown where tape, from the sample fracturing, remained and was clearly distinguishable from the green duckweed tissue. It is common to utilize a manual region-of-interest (ROI) selection prior to MSI data analysis to remove any off-tissue areas. Manual ROI selections, however, rely on an "eyeballed" border selection around the tissue and are prone to error. There have been some efforts to automatically separate off-tissue background from on-tissue ROI in MSI analysis;[33,37]
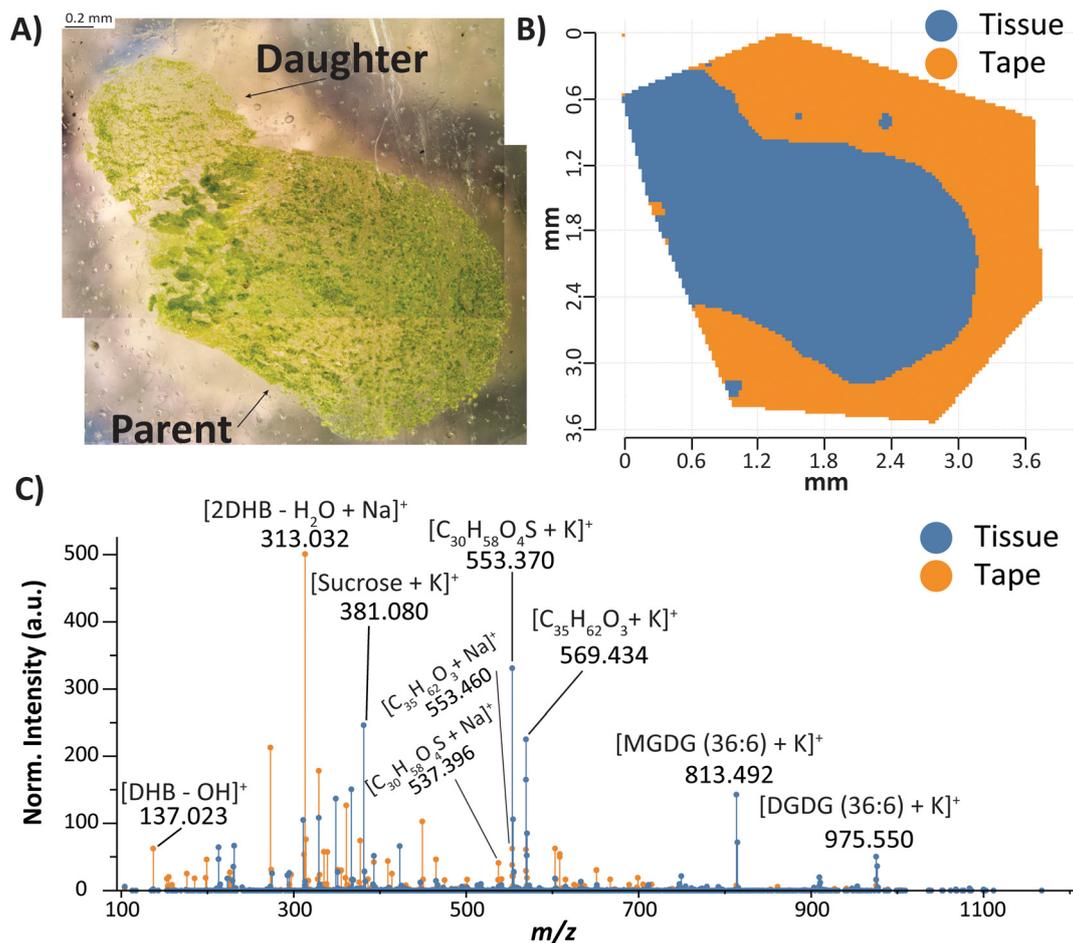
Fig. 1 (A) Optical image of unlabeled duckweed after fracturing. (B) Spatial segments generated by Cardinal using SSC ($r = 3$, $k = 2$, and $s = 16$); parameters were selected to ensure the best partitioning of tape and tissue in accordance with the optical image. (C) Mean mass spectra for segments shown in B.

however, it can be readily performed with SSC segmentation. By setting the maximum number of clusters (k) to 2 in the SSCs, the tape background and the duckweed tissue were easily separated (Fig. 1B); the segmentation was in good agreement with the optical image. Mean mass spectra of the tape (orange) and tissue (blue) segments showed that the most abundant peaks in the tape were matrix related (Fig. 1C). The same matrix peaks were also present in the tissue region but in lower abundance. This preliminary background removal process was applied to all other datasets used in this study.

A subset of pixels was created from the "tissue" pixels observed in Fig. 1B, then SSC segmentation was performed a second time. Five models were generated with varying sparsity shrinkage parameters ($s = 2, 4, 8, 16,$ and 32; see Fig. S1 and Table S1); in all, four segments were observed. It was determined that s = 8 provided reasonable performance metrics (see SI 2) and maintained enough features for later comparison to isotope-labeled metabolites. The four segments (Fig. 2A) were categorized based on differences in t-statistics and m/z co-localizations. Positive t-statistics indicated features with higher abundance than the global mean spectrum

(Fig. 2B). In contrast, negative t-statistics indicated that feature intensities were lower, or possibly absent, compared to the global mean. For example, sucrose (m/z 381.080, $[C_{12}H_{22}O_{11} + K]^+$; labeled with an * in Fig. 2B) showed higher abundance in the parent and intermediate/daughter segments ($t = 14$ and 12, respectively), and lower abundance in the tissue edge and budding pouch segments ($t = -14$ and -33, respectively); the MS image of sucrose (Fig. S2) supports this observed trend. In Fig. S3, additional examples of selected m/z values from the top five most positive t-statistics of each segment to demonstrate the most extreme differences in localizations and serve as indicators for the driving forces behind the segmentation results.

In the outer portion of the parent frond, sucrose (m/z 381.080, Fig. S2) and potassium sulfate (m/z 212.843, Fig. S3) exhibited more positive t-statistics than in the daughter/intermediate segment (Fig. 2). High abundance of sucrose in the parent frond was likely due to the short-term storage of excess glucose.[45] Potassium sulfate originated from plant media, although it is not clear why it was more abundant in the outer parent region. In the budding pouch region of Fig. 2, the
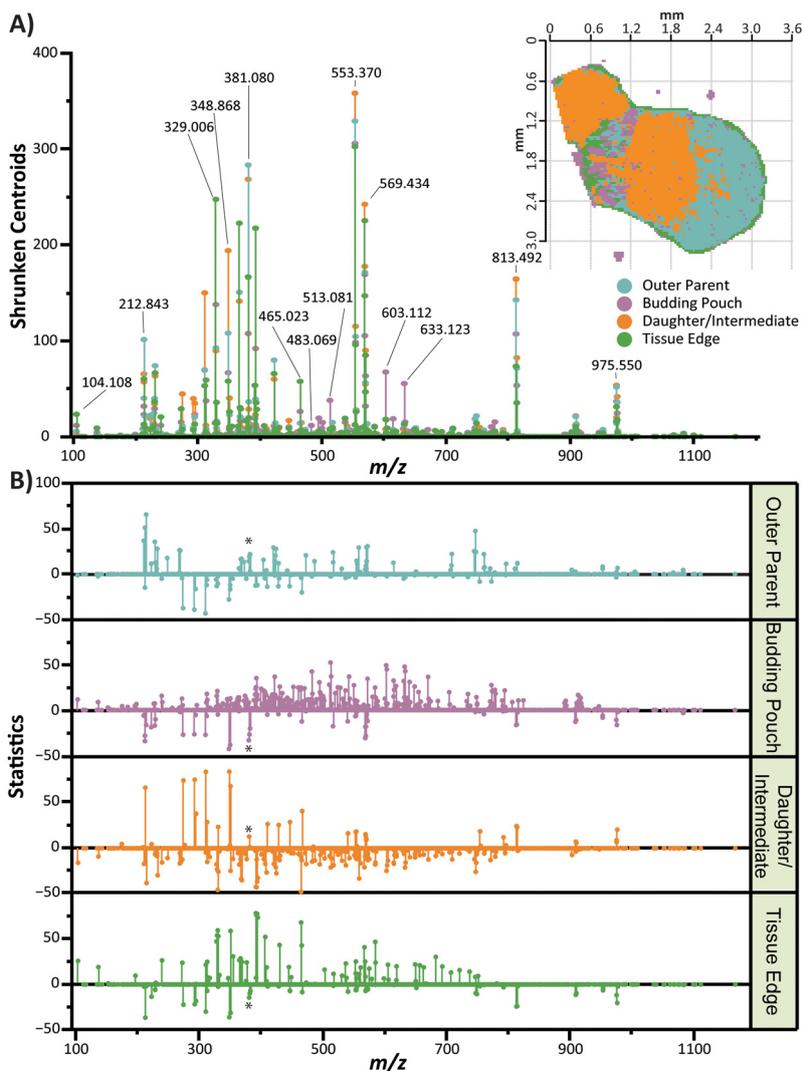
Fig. 2 (A) Spatial segments (inset figure; $r = 1$, $k = 4$, and $s = 8$) and shrunken centroids of unlabeled duckweed showing four regions: the outer edge of the parent frond (cyan), budding pouch (purple), daughter/intermediate (orange) region, and tissue edge (green). (B) The t-statistics of four SSC segments. Marked with an asterisk is m/z 381.080 annotated as sucrose.

pocket from which new fronds develop, characteristic ions at m/z 483.069, 513.081, 603.112, and 633.123 were consistent with flavonoids, or other closely-related compounds, according to annotations provided by METASPACE. Galactolipids, major chloroplast lipids present in all photosynthetic cells of plants, such as MGDG 36 : 6 (m/z 813.492) and DGDG 36 : 6 (m/z 975.550), were homogenously distributed throughout the duckweed tissue. The edge of the tissue was attributed to a sample preparation artifact, a ring of residue that remained after vacuum drying; an example microscope image of this residue is shown in Fig. S4. Therefore, SSC segmentation could effectively differentiate one or two pixel layers along the tissue border, which might not have been possible with manual ROI selection. A few characteristic ions in the tissue edge included choline ([$C_5H_{14}NO$]$^+$ at m/z 104.108) and matrix ions (i.e. [2DHB–H + Na]$^+$ and [3DHB–2H$_2$O + K]$^+$ at m/z 329.006 and 465.023, respectively). Two high abundance ions

at m/z 553.370 and 569.434 were present in all parts of the tissue with minor intensity differences between segments; their chemical formulae were tentatively identified as [$C_{30}H_{58}O_4S$ + K]$^+$ and [$C_{35}H_{62}O_3$ + K]$^+$, matching with well-known plastic additives.[46,47] $C_{30}H_{58}O_4S$ and $C_{35}H_{62}O_3$ likely originated from the tape, but preferentially formed potassium adducts in the tissue region while detected as sodium adducts in the off-tissue region at m/z 537.396 and 553.460 (Fig. 1C).

SSC segmentation was also performed without background removal to compare with the two-step segmentation results. To achieve similar segmentation results to that shown in Fig. 2A, $k = 7$ was required (Fig. S5). Four segments in the tissue region were similar to Fig. 2A; however, additional segments were observed in the off-tissue region. Due to SSC being sensitive to high abundance m/z features, such as matrix peaks or chemical background, the additional segments revealed were due to fluctuations in matrix intensity or matrix cluster peaks that

were not relevant to the analysis of duckweed metabolites. Therefore, removal of the off-tissue regions could improve computational speed without impacting the overall segmentation results.

^{13}C-labeled duckweed

After successfully applying SSC segmentation to the MSI of unlabeled duckweed, a similar process was applied to previously published MSI*i* datasets of ^{13}C-labeled duckweed.[15] Previously, manual data analysis was limited to major galactolipids and their isotopologues, which revealed that the spatial distributions of isotopologues corresponded to various stages of plant growth. After three days of ^{13}C-labeling, galactolipid isotopologue patterns were classified into four groups, depending on the labeling of each building block: unlabeled, galactose-only (Group 1), galactose and one fatty acyl chain (Group 2), and the entire molecule (Group 3). The isotopologues approximately followed binomial distributions based on the concentration of ^{13}CO$_2$ (~90%) and the number of carbons in the labeled molecular moiety. The unlabeled and Group 1 isotopologues were localized in the parent frond, while Groups 2 and 3 were mostly present in the 'intermediate' region and daughter fronds, respectively. However, this 'intermediate' region was ill-defined and subjective.[15]

SSC segmentation was applied to MSI*i* data of ^{13}C-labeled duckweed to test its effectiveness for the detection of various galactolipid isotopologue groups. First, the tape background (Fig. S6) was removed, then six SSC segmentation models (Table S2) were generated for the tissue-only region. Fig. 3 shows the segmentation of three-day ^{13}C-labeled duckweed with $r = 2$, $k = 5$, and $s = 8$. Five distinct segments were observed: inner and outer parent frond, inner and outer daughter fronds, and budding pouch (Fig. 3A); no tissue edge was observed in this sample. This is in contrast to only three regions classified *via* manual analysis: parent, daughter, and intermediate.[15] In Fig. 3B, the shrunken centroids of the MGDG 36:6 isotopologue distribution are shown as an example. As described in our previous work, the isotopologue distribution of MGDG 36:6 can be broken into four groups: unlabeled, and Groups 1, 2, and 3. The parent frond was transferred to the ^{13}CO$_2$-chamber at the beginning of the experiment; therefore, it was composed of unlabeled 'old' lipids. During the three-day incubation period, newly photosynthesized ^{13}C$_6$-galactose was transferred to unlabeled diacylglycerol (DAG) from UDP-^{13}C$_6$-galactose, producing Group 1 MGDG 36:6 in both the outer and inner parent segments (Fig. 3B). The inner parent and budding pouch exhibited greater abundance of Group 2 isotopologues; this region was previously referred to as 'intermediate tissue'.[15] Group 2 labeling was attributed to unlabeled lysophosphatidic acid (LPA), acylated by newly synthesized ^{13}C-labeled fatty acid, then galactosylated by ^{13}C$_6$-galactose. It was hypothesized that, at the time of transfer to the ^{13}CO$_2$-chamber, the inner parent and budding pouch regions were still in the developing stage with highly abundant unlabeled LPA.[15] Interestingly, there was a subtle difference in the Group 2 MGDG 36:6 isotopologue distri-

butions between the inner parent and budding pouch. The Group 2 labeling in the inner parent and budding pouch segments peaked at ^{13}C$_{22}$ and ^{13}C$_{24}$, respectively (Fig. 3B). Unusually wide Group 2 distribution was previously noted compared to the expected binomial distribution fitting and ascribed to additional glycerol backbone labeling in Group 2.[15] Here, SSC segmentation could reveal the minute differences in the spatial distributions of these intermediately labeled lipids.

In the daughter fronds, which were mainly composed of newly synthesized cells, Group 3 MGDG 36:6 isotopologues were most abundant and a notable lack of unlabeled or Group1 isotopologues was observed. While the difference between the outer and inner daughter segments was not distinguishable based on the isotopologues of MGDG 36:6, it was clearer in other lipids (Fig. 3C). For example, a lower abundance of Group 3 DGDG 36:6 was noted in the outer daughter compared to the inner daughter. Likewise, higher abundances of Group 3 pheophytin *a* and an unidentified lipid were observed in the outer daughter frond. These differences were also evident in the shrunken centroids, shown in Fig. S7. SSC segmentation of two-day ^{13}C-labeled duckweed displayed similar results to those observed in the three-day sample. Details of the SSC segmentation model parameters are provided in Table S3. As shown in Fig. S8A, only four distinct segments were noted for the two-day labeled, corresponding to the outer parent, inner parent, daughter frond, and tissue edge. A simpler segmentation was expected due to the younger development stage of the daughter fronds than those shown in the three-day sample. Still, the four isotopologue groups of MGDG 36:6 (Fig. S8B) and DGDG 36:6 (Fig. S8C) were the main driving forces of the segmentation.

When interpreting segmented mass spectra (or shrunken centroids), one should carefully consider possible matrix effects, the change of ionization efficiencies in different parts of the tissues due differences in major species. In this particular application, matrix effects were not considered to be a major hindrance considering that all spatial segments are components of a leaf prepared by fracturing. These segments were all dominated by mesophyll cell membrane lipids, such as MGDGs, DGDGs, and chlorophylls. One may consider that the budding pouch might have a slight difference in matrix effect compared to parent or daughter fronds. However, the summed signal of the isotopologues of MGDG 36:6 in the budding pouch was not markedly different from other segments (Fig. 3B), suggesting there was minimal matrix effects, if any.

D-labeled duckweed

In autotrophic plants, all hydrogen atoms needed for biosynthesis originate from water.[48] Therefore, substitution of H$_2$O with D$_2$O is an excellent method to probe into lipid biosynthesis and other metabolic pathways that cannot be investigated with ^{13}CO$_2$.[48,49] For *L. minor*, relatively healthy reproduction was maintained with weekly media changes in up to D$_2$O concentrations of 50%; however, notable decreases in growth rate and frond sizes were observed under these conditions.[15] To directly compare the SSC segmentation of the ^{13}C-labeled

Analyst

Analyst Paper
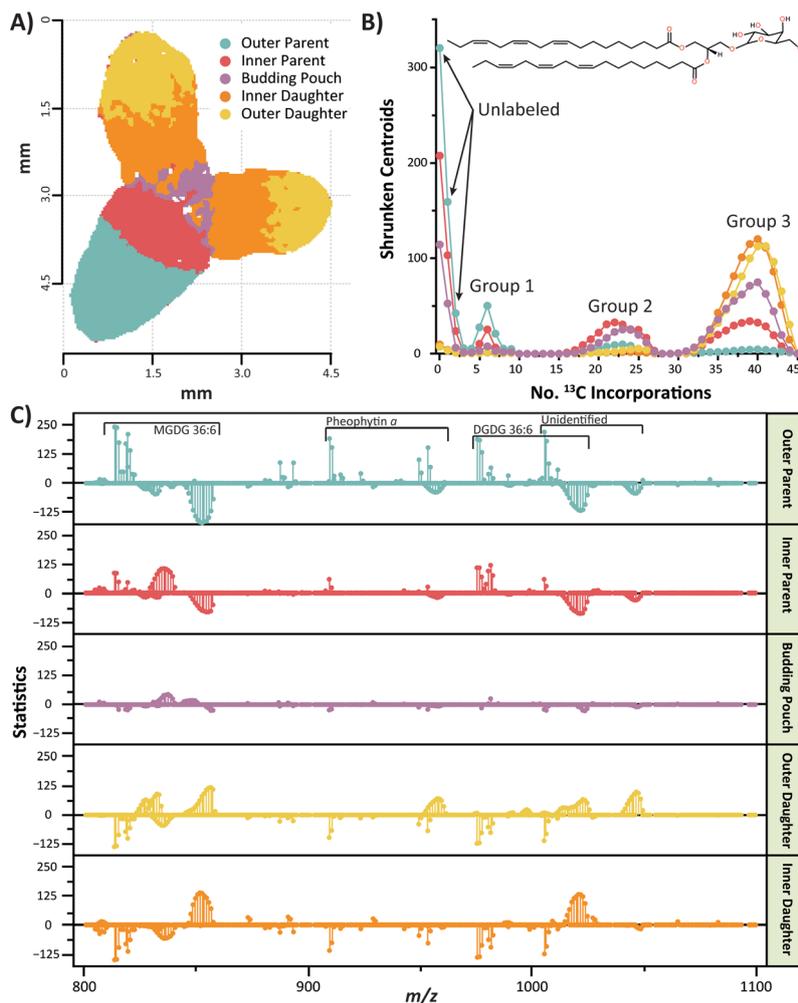
Fig. 3 (A) Spatial segmentation of three-day ^{13}C-labeled duckweed sample. (B) Isolated isotopologue distribution of MGDG 36:6. (C) The t-statistics of the three-day labeled sample.

MSIi datasets, a subset of five-day D-labeled data was defined for the lipid mass range of m/z 800–1200. Segmentation of the tissue-only region showed only two segments corresponding to duckweed anatomy (r = 3, k = 2, and s = 8); k was initially set to 5 but the final number of usable segments was 2. The two segments were annotated as parent and daughter (see Fig. S9). As was observed in the ^{13}C-labeled datasets, unlabeled and Group 1 isotopologues were localized to the parent frond while Group 3 was almost exclusively observed in the daughter frond. Inefficient segmentation of D-labeled, compared to ^{13}C-labeled, duckweed was mostly attributed to the lower signal intensities for D-labeled peaks (~6 times lower). The difference in signal was, in part, due to broader isotopologue distributions with D-labeling (50% D$_2$O) than ^{13}C-labeling (90% ^{13}CO$_2$) but also to smaller tissue size in 50% D$_2$O culture.

While the lipid mass range segmentation yielded few meaningful conclusions, the entire mass range (m/z 100–1200) provided more promising avenues to explore. The results of six SSC segmentation models are summarized in Table S4. Five segments were observed and annotated as the outer and inner

parent, budding pouch, daughter fronds, and tissue edge (Fig. 4); the shrunken centroids and t-statistics for this model are provided in Fig. S10. The top five t-statistics from each segment (Table S5) were mostly from non-biological origins such as inorganic salts, matrix clusters, plastic additives, and polymer contaminations. Sucrose (m/z 381.080) in outer parent and asparagine (m/z 208.973) in the daughter fronds are two distinct metabolites with multiple isotopologues. Furthermore, after comparing the t-statistics of labeled in Fig. S10B to the unlabeled in Fig. 2B, another metabolite was noted with high abundance isotopologues and distinct localization: choline (m/z 104.108) localized to the budding pouch.

Fig. 5 shows the MS images of choline, asparagine, and sucrose and their respective isotopologue distributions. Choline and its isotopologues were most abundant in the budding pouch segment; unsurprisingly, the localization was similar in the unlabeled duckweed (not shown). Since the budding pouch is the pocket from which new fronds will emerge, higher abundance of choline could be due the synthesis of new membrane lipids (such as phosphatidylcholine)

This journal is © The Royal Society of Chemistry 2025

*Analyst*, 2025, **150**, 4404–4413 | **4409**
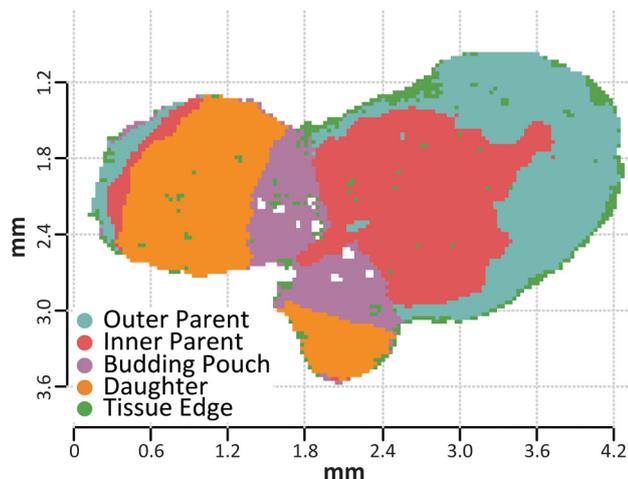
Fig. 4 Segmentation of five-day D-labeled duckweed with five segments corresponding to the outer parent (cyan), inner parent (red), budding pouch (purple), daughter (orange), and tissue edge (green). Parameters used to achieve this segmentation were $r = 3$, $k = 5$, and $s = 8$.

in this region to facilitate daughter frond maturation. Interestingly, asparagine demonstrated an unusually high abundance in deuterated tissues compared to unlabeled

tissues (Fig. 5A *vs.* Fig. S11). When the total asparagine signal, including the monoisotope and isotopologues, were normalized to the signal at m/z 212.843, potassium sulfate from the media, it was only ~1.9% in the unlabeled duckweed, whereas in the D-labeled tissue it was ~194.1%. This higher abundance could indicate D_2O-induced stress, which can inhibit the production of proteins and lead to a build up of amino acids.[50] Sucrose was completely D-labeled and unlabeled sucrose was not observed after five days in 50% D_2O (Fig. 5B), suggesting its high turnover rate. Its localization to the parent frond was attributed to the conversion of excess glucose to sucrose in matured cells. Several D-labeled metabolites, such as flavonoids, were not clearly visualized due to low abundances; however, their isotopologues were higher in abundance in the tissue edge (Fig. S12). A previous report showed that flavonoids were more abundant on the top exterior surfaces of duckweed than the interior middle layer.[51] Hence, flavonoids were detected in higher abundance in the tissue edge segment where the residue from the top side of the duckweed remained on the tape after vacuum drying. MGDG 36:6 showed similar localization as previously discussed in ^{13}C-labeling or D-labeling but Group 2 labeling was noted in very low abundance with no apparent localization in either the daughter or parent frond. MALDI-MS/MS was performed to confirm the
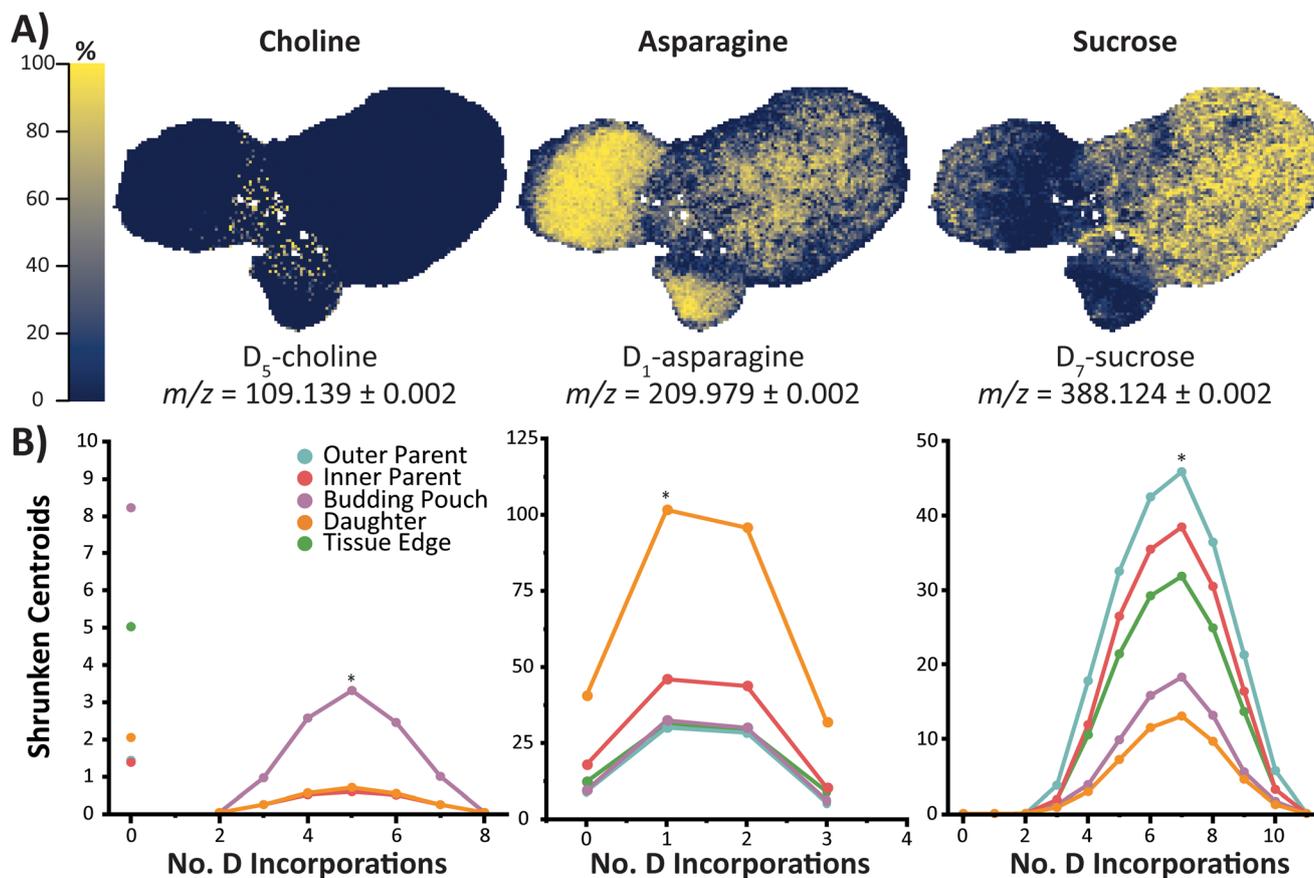
Fig. 5 (A) Representative MS images of D-labeled choline ($D_0 = 104.108$), asparagine ($D_0 = 208.973$), and sucrose ($D_0 = 381.080$). (B) D-labeling isotopologue distributions for choline, asparagine, and sucrose. Isotopologues shown in A are marked with an asterisk in B.

proposed chemical formulae for select ions as shown in Fig. S13.

Insights into metabolite turnover

In the study of *de novo* fatty acid biosynthesis with *in vivo* isotope labeling, isotopomer spectral analysis has been widely used to obtain two model parameters: the fraction of newly synthesized fatty acids (the $g(t)$ value) and the fractional contribution of ^{13}C-tracer to lipogenic acetyl-CoA (the D-value).[52] The $g(t)$ values are particularly informative as they directly relate to absolute *de novo* biosynthesis rates of fatty acids. Here, we generalize to any metabolite and define the fraction of *de novo* biosynthesis (f) in eqn (S4) which provides insight into metabolite turnover rates (SI 3). By calculating f for a metabolite in each SSC segment, the relative metabolite turnover can be compared for different regions of the duckweed plant. Fraction of ^{13}C- and D-labeling, or fraction of *de novo* biosynthesis, were calculated for selected lipids and metabolites in each segment of MSI*i* data and are summarized in Tables 1 and 2. For ^{13}C-labeling of lipids, f values in the daughter frond was ~96% or less when labeled for two days and then increased to almost 100% after three days (Table 1). In the outer parent, f values were below 20% even after three days of labeling, indicating no or minimal turnover but showed higher turnover in the inner parent and budding pouch segments (mostly 50–70%), generally agreeing with tissue development. A similar pattern was observed for D-labeling, with higher turnover in the daughter frond and much lower turnover in the parent frond. Some lipids, especially DGDG 34:3, showed much lower f values than others, partially due to their low abundances, which exacerbated ion suppression in Orbitrap,[53–57] in addition to their slow turnover.

While lipids exhibited a gradual increase of f values with tissue development, small metabolites showed a wider range of f values (Table 2). Flavonoids, generally, followed a similar trend as the lipids, gradually increasing toward the younger parts of the tissue. Asparagine and sucrose have consistent and high f values in all segments, lending to the conclusion that these metabolites are rapidly consumed and produced in all developmental stages. Unlike asparagine and sucrose, choline had the low f values less than 60%, especially in parent fronds. A lower f values for choline even in younger tissues compared to asparagine or sucrose was likely due to its complex synthetic pathways, whereas sucrose and asparagine are rapidly synthesized from glucose and its energy cycle in every cell types.

Conclusions

In this study, we demonstrated that SSC segmentation is a useful tool for untargeted MSI*i* data analysis to elucidate insights into metabolomic dynamics with high spatial resolution. In this proof-of-principle study, the usefulness of unsupervised machine learning was presented for MSI*i* data analysis of *in vivo* isotope labeled duckweeds, which showed improved performance from previous manual data analyses. We illustrated the usefulness of this analysis workflow on four examples: unlabeled, two- and three-day ^{13}C-labeled, and five-day D-labeled duckweeds. Using Cardinal and SSC allowed for a highly efficient and streamlined analyses of chemical profiles between the off-and on-tissue regions as well as spatial distributions of isotopologues. In addition, the unsupervised segmentation was able to segment one- or two-pixel layers around the tissue resulting from the vacuum drying process. Further, for the first time tissue specific metabolic flux information, relative metabolite turnover, could be inferred from the fraction of *de novo* biosynthesis in each SSC segment through untargeted MSI*i* analysis.

Table 1 Fraction of *de novo* biosynthesis (%) of lipids calculated from the mean mass spectrum of each SSC segment in three- and two-day ^{13}C-labeled and five-day D-labeled datasets

	MGDG 36:6	Pheophytin *a*	DGDG 34:3	DGDG 36:6
Three-day ^{13}C				
Outer parent	7.5	1.3	ND	17.9
Inner parent	70.2	58.1	13.1	70.7
Budding pouch	73.9	71.1	6.0	67.5
Outer daughter	99.0	99.9	100.0	99.0
Inner daughter	99.3	99.9	100.0	99.3
Two-day ^{13}C				
Outer parent	6.0	<1	ND	15.1
Inner parent	36.8	11.3	ND	28.2
Daughter	96.3	91.8	17.7	93.9
Tissue edge	72.1	29.3	ND	56.3
Five-day D *m/z* 800–1200				
Parent	2.0	7.4	ND	9.9
Daughter	81.1	85.2	ND	27.7

Calculated from SSC segmentation results shown in Fig. 3 and Fig. S8, S9, respectively. ND: not detected.

Table 2 Fraction of *de novo* biosynthesis of various metabolites calculated from the mean mass spectrum of each SSC segment in five-day D-labeled duckweed for the mass range *m/z* 100–1200

	Choline	Asparagine	Sucrose	$C_{22}H_{20}O_{10}$	$C_{23}H_{22}O_{11}$	$C_{26}H_{28}O_{14}$	$C_{27}H_{30}O_{15}$	MGDG 36:6	Pheophytin *a*
Outer parent	ND	88.6	100.0	ND	17.4	4.2	12.7	2.1	1.0
Inner parent	14.8	84.8	100.0	ND	17.9	7.3	33.8	2.7	17.6
Budding pouch	59.3	89.9	100.0	100	100.0	92.9	95.3	88.8	93.9
Daughter	50.4	84.9	100.0	ND	100.0	100.0	100.0	97.4	98.2
Tissue edge	48.4	86.9	100.0	48.5	58.1	33.5	48.4	29.3	26.1

Calculated from SSC segmentation result shown in Fig. 4. ND: not detected.

One caveat in this analysis is that the segments might be subject to different matrix effects, making quantitative interpretation unreliable. It is not expected to be a major limitation in this particular application to duckweed, as the fractured leaves are dominated by mesophyll cell membranes. Additionally, signal normalization was performed to account for pixel-to-pixel signal variations. However, some other tissue samples with quite distinct chemical compositions might have significant matrix effect among spatial segments. Even in such cases, relative isotopologue abundances of the same compound can be compared because the isotopologues would have the identical ionization efficiencies in the same segment. It is to be noted, as with all machine learning applications, the investigator should be critical of the results as not all reported t-statistics are analytically relevant or biologically pertinent. Improvements to this workflow could be made with improved signal intensities of isotopologues and reduction of chemical interferences, such as tape and matrix background contributions. Future work should focus on applying SSC segmentation to a variety of *in vivo* isotope labeling systems and improving the current method to automatically extract isotopologue distributions of each metabolite, cross-reference to existing metabolite databases, and generate a summary of results.

Author contributions

Y. L. conceived the idea. R. B. J. performed all data analysis and wrote the manuscript with the help of Y. L. V. T. obtained data for isotope labeled duckweeds and helped with the experiment on unlabeled duckweeds. All authors have given approval to the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for the unlabeled duckweed sample is publicly available on METASPACE (**https://metaspace2020.org/project/duckweed-machinelearning**). Data for isotope labeled duckweed are publicly available on Figshare (**https://doi.org/10.25380/iastate.28540523.v1**).

Code availability: An example of the code used in this analysis is publicly available on GitHub (**https://github.com/buckm065/IsotopeLabelingCardinalAnalysis**).

Supplementary information is available. See DOI: **https://doi.org/10.1039/d5an00649j**.

Acknowledgements

References

1 T. Fu, N. Elie and A. Brunelle, *Phytochemistry*, 2018, **150**, 31–39.

2 P. Becker, T. Nauser, M. Wiggenhauser, B. Aeschlimann, E. Frossard and D. Günther, *Anal. Chem.*, 2024, **96**, 4952–4959.

3 L. Wu, K. Qi, C. Liu, Y. Hu, M. Xu and Y. Pan, *Anal. Chem.*, 2022, **94**, 15108–15116.

4 Y. J. Lee, D. C. Perdian, Z. Song, E. S. Yeung and B. J. Nikolau, *Plant J.*, 2012, **70**, 81–95.

5 M. Vats, B. Cillero-Pastor, E. Cuypers and R. M. A. Heeren, *Analyst*, 2024, **149**, 4553–4582.

6 C. Jang, L. Chen and J. D. Rabinowitz, *Cell*, 2018, **173**, 822–837.

7 G. A. Gowda, N. Shanaiah and D. Raftery, *Adv. Exp. Med. Biol.*, 2012, **992**, 147–164.

8 C. B. Clish, *Cold Spring Harbor Mol. Case Stud.*, 2015, **1**, a000588.

9 F. Ma, L. J. Jazmin, J. D. Young and D. K. Allen, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 16967–16972.

10 B. Faubert, A. Tasdogan, S. J. Morrison, T. P. Mathews and R. J. DeBerardinis, *Nat. Protoc.*, 2021, **16**, 5123–5145.

11 F. J. Naser, M. M. Jackstadt, R. Fowle-Grider, J. L. Spalding, K. Cho, E. Stancliffe, S. R. Doonan, E. T. Kramer, L. Yao, B. Krasnick, L. Ding, R. C. Fields, C. K. Kaufman, L. P. Shriver, S. L. Johnson and G. J. Patti, *Cell Metab.*, 2021, **33**, 1493–1504.

12 K. B. Louie, B. P. Bowen, S. McAlhany, Y. Huang, J. C. Price, J.-H. Mao, M. Hellerstein and T. R. Northen, *Sci. Rep.*, 2013, **3**, 1656.

13 A. C. Grey, M. Tang, A. Zahraei, G. Guo and N. J. Demarais, *Anal. Bioanal. Chem.*, 2021, **413**, 2637–2653.

14 Y. J. Lee, P. N. P. Hapuarachchige, E. A. Larson, N. Le and T. T. Forsman, *J. Am. Soc. Mass Spectrom.*, 2024, **35**, 1434–1440.

15 V. T. Tat and Y. J. Lee, *Plant Cell Physiol.*, 2024, **65**, 986–998.

16 S. Na and Y. J. Lee, *Front. Plant Sci.*, 2024, **15**, 1379299.

17 L. Wang, X. Xing, X. Zeng, S. R. Jackson, T. TeSlaa, O. Al-Dalahmah, L. Z. Samarah, K. Goodwin, L. Yang, M. R. McReynolds, X. Li, J. J. Wolff, J. D. Rabinowitz and S. M. Davidson, *Nat. Methods*, 2022, **19**, 223–230.

18 M. Schwaiger-Haber, E. Stancliffe, D. S. Anbukumar, B. Sells, J. Yi, K. Cho, K. Adkins-Travis, M. G. Chheda, L. P. Shriver and G. J. Patti, *Nat. Commun.*, 2023, **14**, 2876.

19 E. Buglakova, M. Ekelöf, M. Schwaiger-Haber, L. Schlicker, M. R. Molenaar, M. Shahraz, L. Stuart, A. Eisenbarth, V. Hilsenstein, G. J. Patti, A. Schulze, M. T. Snaebjornsson and T. Alexandrov, *Nat. Metab.*, 2024, **6**, 1695–1711.

20 K. C. O'Neill and Y. J. Lee, *Front. Plant Sci.*, 2020, **11**, 639.

21 N. Verbeeck, R. M. Caprioli and R. Van de Plas, *Mass Spectrom. Rev.*, 2020, **39**, 245–291.

22 W. Gardner, S. M. Cutts, D. R. Phillips and P. J. Pigram, *Biopolymers*, 2021, **112**, e23400.

23 H. Hu and J. Laskin, *Adv. Sci.*, 2022, **9**, 2203339.

24 M. T. Bokhart, M. Nazari, K. P. Garrard and D. C. Muddiman, *JASMS*, 2018, **29**, 8–16.

25 A. L. Mellinger, K. P. Garrard, S. Khodjaniyazova, Z. N. Rabbani, M. P. Gamcsik and D. C. Muddiman, *J. Proteome Res.*, 2022, **21**, 747–757.

26 A. L. Mellinger, R. R. Kibbe, Z. N. Rabbani, D. Meritet, D. C. Muddiman and M. P. Gamcsik, *Free Radicals Biol. Med.*, 2021, **193**, 677–684.

27 A. L. Mellinger, D. C. Muddiman and M. P. Gamcsik, *J. Proteome Res.*, 2022, **21**, 1800–1807.

28 A. Palmer, P. Phapale, I. Chernyavsky, R. Lavigne, D. Fay, A. Tarasov, V. Kovalev, J. Fuchser, S. Nikolenko, C. Pineau, M. Becker and T. Alexandrov, *Nat. Methods*, 2017, **14**, 57–60.

29 N. Grira, M. Crucianu and N. Boujemaa, A Review of Machine Learning Techniques for Processing Multimedia Content, Report of the MUSCLE European Network of Excellence, 2004, vol. 1, pp. 9–16.

30 T. Alexandrov, *Ann. Rev. Biomed. Data Sci.*, 2020, **3**, 61–87.

31 A. Jetybayeva, N. Borodinov, A. V. Ievlev, M. I. U. Haque, J. Hinkle, W. A. Lamberti, J. C. Meredith, D. Abmayr and O. S. Ovchinnikova, *J. Appl. Phys.*, 2023, **133**, 020702.

32 P. Mittal, M. R. Condina, M. Klingler-Hoffmann, G. Kaur, M. K. Oehler, O. M. Sieber, M. Palmieri, S. Kommoss, S. Brucker, M. D. McDonnell and P. Hoffmann, *Cancers*, 2021, **13**, 5388.

33 T. Zhou, L. Chen, J. Guo, M. Zhang, Y. Zhang, S. Cao, F. Lou and H. Wang, *BMC Bioinf.*, 2021, **22**, 185.

34 W. M. Abdelmoula, B. G.-C. Lopez, E. C. Randall, T. Kapur, J. N. Sarkaria, F. M. White, J. N. Agar, W. M. Wells and N. Y. R. Agar, *Nat. Commun.*, 2021, **12**, 5544.

35 D. Guo, M. C. Föll, K. A. Bemis and O. Vitek, *Bioinformatics*, 2023, **39**, btad067.

36 K. Ovchinnikova, L. Stuart, A. Rakhlin, S. Nikolenko and T. Alexandrov, *Bioinformatics*, 2020, **36**, 3215–3224.

37 K. Ovchinnikova, V. Kovalev, L. Stuart and T. Alexandrov, *BMC Bioinf.*, 2020, **21**, 129.

38 E. N. Castanho, H. Aidos and S. C. Madeira, *Brief. Bioinf.*, 2024, **25**, bbae342.

39 M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain and A. J. Aljaaf, in *Supervised and Unsupervised Learning for Data Science*, ed. M. W. Berry, A. Mohamed and B. W. Yap, Springer International Publishing, Cham, Switzerland AG, 2020, ch. 1, pp. 3–21, DOI: 10.1007/978-3-030-22475-2_1.

40 S. H. Shetty, S. Shetty, C. Singh and A. Rao, in *Fundamentals and Methods of Machine and Deep Learning*, 2022, pp. 1–16, DOI: 10.1002/9781119821908.ch1.

41 K. D. Bemis, A. Harry, L. S. Eberlin, C. Ferreira, S. M. van de Ven, P. Mallick, M. Stolowitz and O. Vitek, *Bioinformatics*, 2015, **31**, 2418–2420.

42 K. D. Bemis, A. Harry, L. S. Eberlin, C. R. Ferreira, S. M. van de Ven, P. Mallick, M. Stolowitz and O. Vitek, *Mol. Cell. Proteomics*, 2016, **15**, 1761–1772.

43 K. A. Bemis, M. C. Föll, D. Guo, S. S. Lakkimsetty and O. Vitek, *Nat. Methods*, 2023, **20**, 1883–1886.

44 A. T. Klein, G. B. Yagnik, J. D. Hohenstein, Z. Ji, J. Zi, M. D. Reichert, G. C. MacIntosh, B. Yang, R. J. Peters, J. Vela and Y. J. Lee, *Anal. Chem.*, 2015, **87**, 5294–5301.

45 O. Stein and D. Granot, *Front. Plant Sci.*, 2019, **10**, 95.

46 C. Diamante, M. Z. Fiume, W. F. Bergfeld, D. V. Belsito, R. A. Hill, C. D. Klaassen, D. C. Liebler, J. G. Marks, R. C. Shank, T. J. Slaga, P. W. Snyder and F. A. Andersen, *Int. J. Toxicol.*, 2010, **29**, 137S–150S.

47 R. Liu and S. A. Mabury, *Environ. Sci. Technol. Lett.*, 2020, **7**, 14–19.

48 R. S. Nett, X. Guan, K. Smith, A. M. Faust, E. S. Sattely and C. R. Fischer, *AIChE J.*, 2018, **64**, 4319–4330.

49 Y. Xu, A. A. Koroma, S. E. Weise, X. Fu, T. D. Sharkey and Y. Shachar-Hill, *Plant Physiol.*, 2024, **194**, 475–490.

50 R. J. Cooke, S. Grego, J. Oliver and D. D. Davies, *Planta*, 1979, **146**, 229–236.

51 J. J. Rensner and Y. J. Lee, *Anal. Chem.*, 2022, **94**, 11129–11133.

52 M. R. Antoniewicz, *Exp. Mol. Med.*, 2018, **50**, 1–13.

53 J. C. L. Erve, M. Gu, Y. Wang, W. DeMaio and R. E. Talaat, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 2058–2069.

54 J. Eiler, J. Cesar, L. Chimiak, B. Dallas, K. Grice, J. Griep-Raming, D. Juchelka, N. Kitchen, M. Lloyd, A. Makarov, R. Robins and J. Schwieters, *Int. J. Mass Spectrom.*, 2017, **422**, 126–142.

55 X. Su, W. Lu and J. D. Rabinowitz, *Anal. Chem.*, 2017, **89**, 5940–5948.

56 S. Khodjaniyazova, M. Nazari, K. P. Garrard, M. P. V. Matos, G. P. Jackson and D. C. Muddiman, *Anal. Chem.*, 2018, **90**, 1897–1906.

57 C. Neubauer, A. Crémière, X. T. Wang, N. Thiagarajan, A. L. Sessions, J. F. Adkins, N. F. Dalleska, A. V. Turchyn, J. A. Clegg, A. Moradian, M. J. Sweredoski, S. D. Garbis and J. M. Eiler, *Anal. Chem.*, 2020, **92**, 3077–3085.