analytical chemistry

pubs.acs.org/ac Article

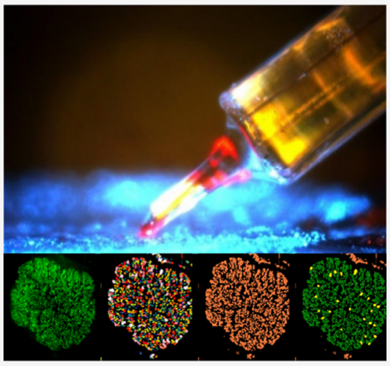
High-Throughput Analysis of Tissue-Embedded Single Cells by Mass Spectrometry with Bimodal Imaging and Object Recognition

Sylwia A. Stopka,* Ellen A. Wood, Rikkita Khattar, Beverly J. Agtuca, Walid M. Abdelmoula, Nathalie Y. R. Agar, Gary Stacey, and Akos Vertes*

Cite This: *Anal. Chem.* 2021, 93, 9677−9687 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In biological tissues, cell-to-cell variations stem from the stochastic and modulated expression of genes and the varying abundances of corresponding proteins. These variations are then propagated to downstream metabolite products and result in cellular heterogeneity. Mass spectrometry imaging (MSI) is a promising tool to simultaneously provide spatial distributions for hundreds of biomolecules without the need for labels or stains. Technological advances in MSI instrumentation for the direct analysis of tissue-embedded single cells are dominated by improvements in sensitivity, sample pretreatment, and increased spatial resolution but are limited by low throughput. Herein, we introduce a bimodal microscopy imaging system combined with fiber-based laser ablation electrospray ionization (f-LAESI) MSI with improved throughput ambient analysis of tissue-embedded single cells (n > 1000) to provide insight into cellular heterogeneity. Based on automated image analysis, accurate single-cell sampling is achieved by f-LAESI leading to the discovery of cellular phenotypes characterized by differing metabolite levels.



1. INTRODUCTION

Biological tissues are composed of complex cellular networks and unique cell types with specialized functions. These cells exhibit varied expression levels of transcripts, proteins, and metabolites resulting in cellular heterogeneity. To capture cell-to-cell variations, ultimately a multiomics approach is required to provide insight into biochemical composition at the cellular level.[1] Similarly, crucial is the spatial mapping of the single-cell profile of these biomolecules in tissues to provide insight into the spatial aspects of cellular heterogeneity.[2,3] A grand challenge in single-cell analysis is the limited cell volume especially for analytes with low concentrations.[4] This has been resolved in single-cell genomics and transcriptomics by utilizing nucleic acid amplification to improve analyte detection, and these methods are becoming mainstream.[5] Established high-throughput techniques are dominated by single-cell RNA sequencing (scRNA-Seq) aided by flow cytometry and microfluidic devices.[6−8]

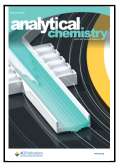
However, single-cell proteomics and metabolomics are not as well established and heavily rely on high sensitivity and selectivity approaches to detect, identify, and quantitate these species. Mass spectrometry (MS) has been used extensively for small-molecule profiling from single cells using a wide array of techniques, including capillary microsampling MS with ion mobility separation (IMS), single-probe MS, capillary electrophoresis MS, and optical fiber-based laser ablation electrospray ionization (LAESI).[9−12] However, a significant drawback to these methods is the lack of high throughput that makes it difficult to reach representative sampling for statistical analysis.

Mass spectrometry imaging (MSI) has been used for single-cell analysis to spatially map biomolecules within tissues and cell cultures at the single-cell level.[13,14] The most widely used MSI techniques for single-cell imaging are matrix-assisted laser desorption/ionization (MALDI) and secondary ion mass spectrometry (SIMS) with lateral resolutions capable of <10 and <1 μm, respectively.[15−18] High-throughput single-cell methods based on MALDI-MSI are emerging to collect enough data that represent a population for statistically meaningful conclusions.[19−21] However, both MALDI and SIMS require extensive sample preparation and a vacuum environment for the analysis.

Ambient ionization sources can sample the native state of biological tissues and single cells with minimal sample preparation.[22,23] A wide array of ambient ionization MS sources has been adapted for the direct analysis of tissue-embedded single cells, including optical fiber-based LAESI (f-LAESI), single-probe MS, and nanodesorption electrospray ionization (nano-DESI), improving the spatial resolution for the latter to <10 μm.[24−27]

Despite the significant advances in MSI to achieve spatial resolutions commensurate with cell sizes, the rectangular grid

Received: February 6, 2021
Accepted: June 28, 2021
Published: July 8, 2021



https://doi.org/10.1021/acs.analchem.1c00569
Anal. Chem. 2021, 93, 9677−9687

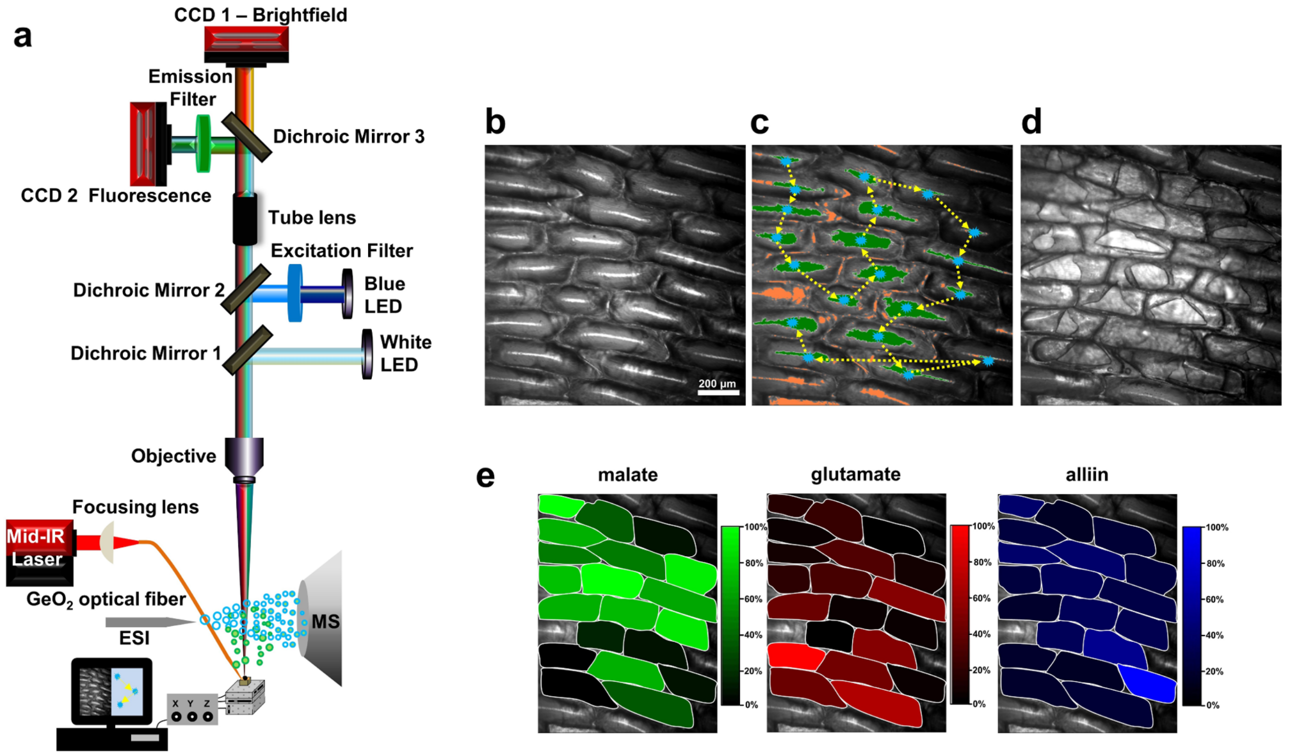


Figure 1. (a) Schematic of home-built bimodal imaging and single-cell sampling f-LAESI ion source. Simultaneous brightfield and fluorescence imaging of the tissue provides information on cell morphology and the presence of a fluorescent tag, respectively. Recognizing cells as objects in the image leads to the determination of centroid coordinates (x_i, y_i). Sampling is achieved by focusing the mid-IR laser beam through an etched optical fiber tip into a cell for ablation. The *XY* translation stage presents all *n* identified centroids consecutively to the fiber tip. (b) Brightfield image of epidermal cells in *Allium cepa* bulb. (c) Centroid positions (x_i, y_i) for $n = 20$ cells are marked by blue dots and the ablation sequence is shown by yellow arrows. (d) Postablation image confirms sampling of all selected cells. (e) Molecular distributions based on ion abundances for malate, glutamate, and alliin represented on false-color scales.

sampling pattern applied in most MALDI imaging experiments remains an impediment to obtaining true single-cell information from a tissue. As in natural tissues, the cells are not following an equidistant grid pattern, the resulting mismatch between the cell and sampling positions results in occasional cosampling of two or more adjacent cells. To overcome this limitation, we demonstrated cell-by-cell molecular imaging of tissues by f-LAESI-MSI.[28] In this approach, cell centroids were determined from a brightfield image of the tissue and their coordinates were used in the definition of sampling positions. As a result, each voxel in the chemical image corresponded to a single cell. As the optical microscope and the mass spectrometer were not integrated, this system was not capable of high-throughput operation. It also lacked fluorescence imaging, thereby further limiting its application.

Fluorescent labels are often used to identify a subpopulation of cells distinguished by a cellular state, the presence of a receptor or a biomarker. Combining cell selection by fluorescence microscopy and single-cell analysis of the selected set give insight into the chemical composition of predefined subpopulations.[9,12] Cell morphology is commonly observed by brightfield microscopy. Automatic recognition of cells in brightfield images is a common segmentation task.[29] Additional information can be extracted by morphometric analysis of the recognized cells in combination with fluorescence, e.g., fluorescence in situ hybridization (FISH), and chemical composition, e.g., nano-SIMS.[30−32] Several ambient ionization methods have been utilized in microscopy-guided analysis of targeted regions directly in tissue sections.[33−35] Morphometric data, collected from brightfield and fluorescence images, e.g., cell centroid positions, enables the automation of sequential f-LAESI-MS analysis of tissue-embedded single cells.

Here, we introduce a bimodal imaging system based on simultaneous brightfield and fluorescence imaging combined with f-LAESI-MS for microscopy-guided high-throughput single-cell analysis. Automated image segmentation, morphometry, and metabolic analysis of >1000 tissue-embedded cells reveal the presence of latent subpopulations distinguished by suppressed and upregulated levels of certain metabolites.

2. EXPERIMENTAL SECTION

2.1. Optical Microscopy. A dual-channel microscope capable of simultaneous brightfield and fluorescence imaging was built to study biological tissues in situ (see Figure 1). A microscope body (CEA1600, Thorlabs, Newton, NJ) held all of the imaging components and the laser optical train. To reduce environmental vibrations, the system was mounted on a passive isolation table. To accommodate all 1 in. optical components and for easy modification, a 30 mm cage system was used to steer the laser beam, fiber coupling, and mounting optics. Attached to the microscope body was an epi-illumination arm in a dovetail configuration in which three illumination kinematic cage cubes were installed in parallel. In the current configuration, the incorporated optical components were for green fluorescent protein (GFP) and brightfield microscopy. The GFP-fluorescence excitation was produced by a 470 nm blue LED light source (M470L3, Thorlabs, Newton, NJ), which passed through a 469 nm excitation filter (MF469-35, Thorlabs, Newton, NJ) and a dichroic mirror (MD498, Thorlabs, Newton, NJ) that reflected light at 415−470 nm and transmitted light at 490−720 nm. The brightfield light source was a cold white LED source (MCWHL5, Thorlabs, Newton, NJ). The produced light was collimated and passed through a beam splitter (BSS10R, Thorlabs, Newton, NJ) coated for 400−700 nm that reflected 30% and transmitted 70% of the intensity. All of the wavelengths were focused onto the tissue sample using either a 5×, 10×, or 20× magnification long working distance objective (MY20X-804, Thorlabs, Newton, NJ). A 1× magnification tube lens (WFA4100, Thorlabs, Newton, NJ) projected the images on CCD sensors. Two monochromatic 4 megapixel scientific-grade CCD cameras (4070M-GE, Thorlabs, Newton, NJ) were mounted orthogonally to each other onto a two-camera mount (2SCM1-DC, Thorlabs, Newton, NJ). The kinematic cage cube (2SCM1-DC, Thorlabs, Newton, NJ) that split the two channels contained a dichroic mirror (MD498, Thorlabs, Newton, NJ) that transmitted 505−800 nm and reflected 452−490 nm light. In the optical path of the fluorescence channel, an emission filter (MF525-39, Thorlabs, Newton, NJ) with a wavelength maximum of 525 nm was installed before the CCD detector. The light transmitted through the dichroic mirror was used for the brightfield channel and required no further filters. To reduce stray light and artifacts in the images, the whole optical path was enclosed from the CCDs to a few inches above the sample.

2.2. Fiber LAESI-MS. A tunable mid-IR laser source (IR Opolette HE 2731, Opotek, Carlsbad, CA) emitting light at 2.94 μm wavelength with a 7 ns laser pulse length and 20 Hz repetition rate was used for fiber ablation experiments. The laser beam was steered using gold-coated mirrors and coupled into the home-built microscope. To produce laser pulses with ~1.0 ± 0.17 mJ energy, the beam passed through an external attenuator. It was focused through a 50 mm focal length plano-convex CaF_2 lens into the blunt distal end of the optical fiber. For precise coupling, the focusing lens and fiber chuck were attached to a fiber mount tilt stage (F-91TS, Newport, Irvine, CA).

A 1 m long germanium oxide optical fiber with a 250 μm core diameter (HP Fiber, Infrared Fiber Systems, Inc., Silver Spring, MD) was used to deliver laser pulses from the laser coupling mount to the sample. Both ends of the fiber were stripped of the Hytrel and polyimide coatings using a heated 1-methyl-2-pyrrolidinone solution. The coupling end was cleaved, while the ablation end was etched in 4% nitric acid solution to produce a sharp tip with an $R = \sim 5$ μm radius of curvature, resulting in a 10 μm tip diameter. This end was secured in a bare fiber chuck and mounted onto a micromanipulator (MN-151, Narishige, Tokyo, Japan) with an automated translation stage along the Z-axis. More details regarding fiber etching and coupling can be found in our earlier work.[12,24]

The etched fiber tip was brought sufficiently close to the sample surface to induce the ablation of single cells by the laser pulses. The ablation plume was intercepted by a stable electrospray from a spay solution of 2:1 (v/v) $MeOH/CHCl_3$ at a flow rate of 500 nL/min. To produce negative ions through the LAESI mechanism, a spray voltage of −2.7 kV was applied.[12] The generated ions were analyzed by a quadrupole time-of-flight mass spectrometer (Synapt G2S, Waters, Milford, MA).

2.3. Integrated Imaging and Sampling Software. The acquisition and analysis of images, as well as cell recognition and automated sampling, were controlled by MetaMorph software (Meta Series Software 7.10.2) working with a compatible sample stage, illumination controllers, and cameras. Initially, the voxel to micrometer calibration was defined for all magnifications and stored. Using the journal feature within the software, separate journals were created for both acquisition channels (brightfield and fluorescence). Each journal triggered the corresponding illumination shutter and camera acquisition channel for the manually selected magnification in use. Triggering the channels individually reduced the overexposure in fluorescence mode. The two images were projected on the CCDs within 50 ms of each other using the twin camera/split-view simultaneous acquisition option. To control the illumination, LED control drivers were placed in triggering mode and connected to a TTL/Analog device (DDA02, Cairn Research, Faversham, U.K.). The journal also controlled the XY-axis sample translation stage (MLS203, Thorlabs, Newton, NJ) and the Z-axis stage (MMP Series one, Mad City Labs, Madison, WI) for lowering and lifting the sample. A stage journal was created that centered the desired sampling point, i.e., the selected single cell, to the middle of the field of view (FOV) under the sharpened optical fiber tip. Then, the sample was lifted for ablation by the laser pulse delivered through the fiber tip.

The ablation event was initiated using the TTL/Analog trigger input of the Q-switch for the mid-IR laser using a digital delay generator (DG535, Stanford Research Systems, Sunnyvale, CA), whereas the flashlamp was triggered internally. A loop was created in the journal to define an ablation event for each cell that also set the delays between laser shots. Setting the delays to 50 ms resulted in an event that consisted of 10 laser shots over an ~1 s period for the ablation of each cell. The ablation events were monitored using an oscilloscope. The schematic of the instrument is shown in Figure 1a.

The fiber tip was manually adjusted in the XYZ, yaw, and pitch directions using a five-axis translator (F-91TS, Newport, Irvine, CA) to the center of the FOV using a journal that created a region of interest (ROI) in the live image. Once in position, the fiber tip was defined as a fixed point for cell-by-cell sampling.

2.4. Sample Preparation. Organic purple onion (*A. cepa*) bulbs were obtained from a local store. The top two layers were removed, and all tissues were obtained from the third layer. Using a scalpel, a small section of this layer was cut and separated. Tweezers were used to remove an intact layer of the inner epidermal tissue from the concave surface that consisted of a monolayer of cells. This tissue was placed on a precut 25 × 25 mm^2 microscope slide and mounted onto a Peltier cooling stage. Maintaining a low sample temperature minimized metabolic changes during the analysis.

Soybean seeds were inoculated with *Bradyrhizobium japonicum* (USDA110) and were grown under greenhouse conditions for 21 days. Intact root nodules were harvested and flash frozen. For sample preparation, the nodules were embedded in 2.5% carboxymethyl cellulose (CMC) medium and cryosectioned to 30 μm thickness to produce a cellular quasi-monolayer of the tissue. The section was thaw mounted onto a precut microscope slide for analysis.

2.5. High-Resolution Imaging of Whole Tissue. As high-resolution images clearly show that the cells produced only a small FOV, to image the whole tissue multiple FOV segments had to be captured and stitched together. This was achieved by the Scan Slide MetaMorph application journal that produced a high-resolution image of the entire tissue in both the brightfield and fluorescence channels. Before acquiring each FOV image, a 500 ms delay was set to ensure the stage was stabilized. The whole tissue area had to be defined by fiduciary markers at the upper left and lower right corners. Once all high-resolution FOV images of the tissue were obtained, image overlaps of 10% were used for alignment and stitching. This raw image then underwent image processing for cell recognition and morphometry.

2.6. Image Processing for Cell-by-Cell Imaging. Once a raw fluorescence image was acquired, it underwent background subtraction to enhance contrast, highlight the cells, and remove artifacts. To isolate and define individual cell boundaries, image segmentation was applied. The approximate minimum and maximum cell sizes (e.g., 20 and 70 μm, respectively, for infected root nodule cells) were defined, sorted based on cell area in the image and color-coded. The segmented image was binarized to capture individual cells and morphometric analysis was performed. The objects were filtered using area, length, and shape factor and recognized as individual cells with (x_i, y_i) centroid coordinates determined by the software. Cell volumes were calculated by multiplying the areas of the cells in the image measured by morphometry with an estimated thickness of 100 μm for the cellular monolayer.

2.7. Automated Sampling for Cell-by-Cell Imaging. The list of (x_i, y_i) cell centroid coordinates in the image had to be linked with the corresponding stage positions. After the conversion factors were established, a journal drove the sample stage to consecutively position each cell at the center of the FOV, where the optical fiber tip was positioned slightly above the tissue surface. At each position, the sample stage was lifted in the Z direction until the selected cell was in close proximity to the fiber tip. Ten laser shots were delivered that caused the cell to ablate. Ablating a relatively constant volume of cell content, independent of cell volume, is a prerequisite for the relative quantitation of metabolite abundances between cells. In other words, analyzing an, e.g., ∼2 nL sample of cell content, irrespective of cell volume, assured that cellular metabolite concentrations, not their absolute amounts, determined relative signal intensities between cells. In addition, ion intensities for each cell are normalized to the sum of sample-related ion intensities to obtain relative intensities less dependent on the sampling process. As the ablation plume intercepted the electrospray, a single-cell spectrum was captured by the mass spectrometer. After the ablation event, the stage was lowered to the original Z position and moved to the next cell. The mean ion abundance values are not meant to be correlated with the cellular concentrations of the related metabolite because different chemical species can have very different ionization efficiencies, resulting in different ion intensities. Rather, ion intensities for a particular metabolite are compared for different cells. In addition, the definition of metabolic noise reveals that it is a relative metric that eliminates all linear distortions of ion intensities because it only depends on intensity ratios not on the intensities themselves.

2.8. Statistical Analysis, Dimensionality Reduction, and Data Visualization. Several of the detected ions were of low abundance and the signal occasionally fell below the limit of detection. Thus, a left-censored missing not at random (MNAR) imputation method was applied for missing values.[36]

To explore cellular heterogeneity, abundance distributions after MNAR imputation were plotted in the form of histograms. Unimodal distributions were modeled by normal and lognormal probability density functions, whereas bimodal distributions were fitted by their linear combinations.

High-dimensional analysis of single-cell MS data was performed using uniform manifold approximation and projection (UMAP) for dimension reduction. UMAP is a nonlinear dimensionality-reduction method that has been found efficient in preserving local structures in the original high-dimensional single-cell data.[37] Preserving such local structures in a lower-dimensional representation (e.g., in two-dimensional (2D)) can reveal distinct cell populations. The high-dimensional spectra were projected into a two-dimensional space where similar spectra were projected close to each other, whereas dissimilar ones were projected further apart.

3. RESULTS

3.1. Image Analysis. Simultaneous brightfield and fluorescence microscope imaging combined with f-LAESI-MS (see Figure 1a) yields a multidimensional data set composed of morphometric parameters, e.g., cell centroid coordinates (x_i, y_i) and cell-specific mass spectra $((m/z)_{i,j}, I_{i,j})$, where i enumerates the cells and index j corresponds to spectral features. The high-resolution images with cellular resolution reveal phenotypic differences between the cells based on location and morphology and identify their infection status based on the fluorescence signal. For example, brightfield images of a soybean (*Glycine max*) root nodule section revealed several cell phenotypes localized at the outer cortex and inner infection zone (see Figure 5e). Simultaneous fluorescence imaging of the same tissue revealed the cells that were infected by GFP-tagged rhizobia (Figure 5a). Determining the metabolic composition of these cell types can provide insight into the pathways active in biological nitrogen fixation without spectral interference from adjacent uninfected cells or cells of different phenotypes.

To achieve single-cell analysis, the diameter of the sampling spot had to be smaller than a cell. Coupling the laser beam into an optical fiber with the sampling tip etched to a diameter of ∼10 μm provided ∼10 μm ablation spot size that conferred sufficient resolution for single-cell sampling. For automated cell-by-cell analysis, the (x_i, y_i) centroid coordinates of targeted phenotypes were obtained through image processing (Figure 1b,c).[28] This started with identifying the cells in the tissue section as objects, selecting the targeted subset, and determining their centroid coordinates (x_i, y_i).

3.2. High-Throughput Sampling. Cells of the *A. cepa* (onion) bulb epidermis are used as a model system because they are largely uniform in shape and size. Once cell centroid positions (x_i, y_i) for cell i are determined in the microscope image, they are converted into a list of coordinates for the automated XY translation stage. The tip of the optical fiber is in a fixed position at the center of the FOV, and the sample stage moves through the coordinate list to position the cells one by one under it for ablation. For each cell, an $((m/z)_{i,j}, I_{i,j})$ mass spectrum is obtained (see, e.g., Figure 2a) and registered with the (x_i, y_i) coordinates. To link a cell from the image to the corresponding mass spectrum, an ion chromatogram is recorded linking the cell ablation events through time to the

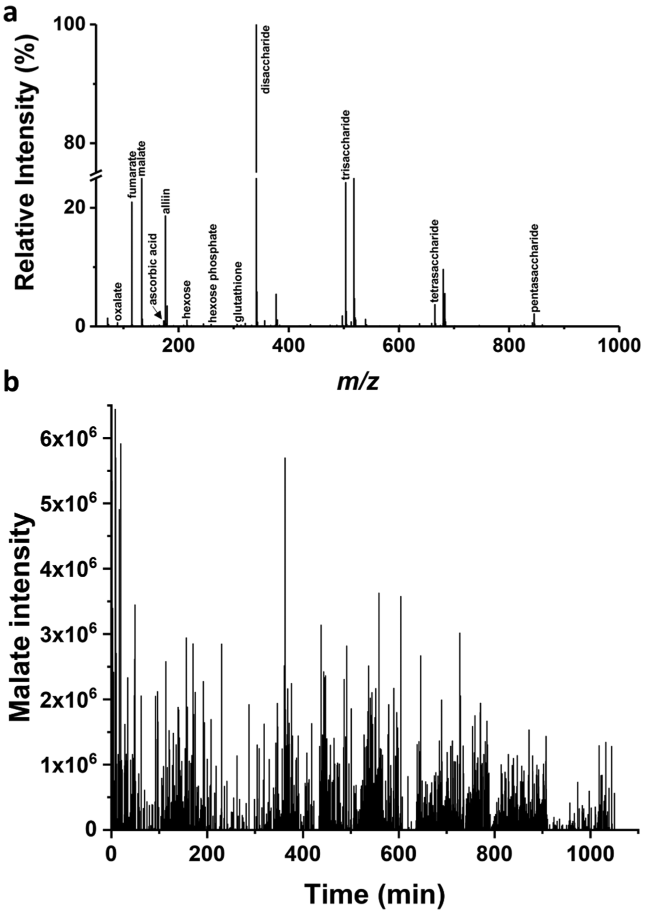


Figure 2. (a) Representative mass spectrum of a single *A. cepa* epidermal cell by f-LAESI. A total of 108 spectral features were detected. (b) Ion chromatogram of malate in 1084 single *A. cepa* epidermal cells captured by high-throughput f-LAESI-MS. Each peak corresponds to the malate abundance in an individual cell. Cellular heterogeneity is reflected by a wide range of intensities.

cell coordinates. For example, tracking a cell-related ion, e.g., malate, reveals the exact time of each ablation event (Figure 2b). Following the ablation, an image is captured to document that the correct cell was selected and no neighboring cells were affected (Figure 1d). Based on these data, each cell can be viewed as a voxel with (x_i, y_i) in a molecular image corresponding to the $(m/z)_{i,j}$ ion, where the $I_{i,j}$ abundance is represented on a false-color scale (Figure 1e). This automated process results in a sampling rate of ∼2 cells/min. For example, from an *A. cepa* epidermal monolayer, ∼120 single cells are sampled in 1 h. To improve the statistical power of our data set, 1084 single cells were analyzed in ∼17.5 h.

In addition to cell-by-cell imaging to explore spatial distributions, the collected data can also be used to explore cellular heterogeneity. For example, looking at the malate abundance distribution over this cell population, cell-to-cell variations with a lognormal distribution are observed, as it is explained in the next section.

3.3. Single-Cell Spectra. Using the f-LAESI-MS platform, primary and secondary metabolites are tentatively identified in single cells. In a typical negative ion mode spectrum, over 100 spectral features are detected, including the strong signals from disaccharides, trisaccharides, tetrasaccharides, and pentasaccharides in the higher mass region (Figure 2a). Various small metabolites are tentatively identified including oxalate, ascorbate, malate, and alliin, a defense compound known to be present in onion. Identifications relied on accurate mass measurements and an extensive tandem MS data set from similar samples.[24]

3.4. Cellular Heterogeneity and Metabolic Noise. To characterize cellular heterogeneity, abundance distributions and metabolic noise are investigated for 35 metabolites from a population of $n = 1084$ cells. The stochastic nature of enzyme copy numbers in single cells and random environmental factors modulate reaction fluxes and lead to metabolite concentration variations. This randomness over a cell population is captured by metabolic noise, η_m^2. Histograms of metabolite abundances can be modeled by lognormal or bimodal distributions. For example, glutamate and alliin both exhibit a lognormal distribution (Figure 3a,b), whereas a bimodal distribution is observed for disaccharides. The latter is a convolution of a lognormal and a normal component, representing two hidden subpopulations (Figure 3c). Numerous other metabolites follow a lognormal or bimodal pattern (see Supplementary Figures 1−5).

Determination of the measured noise does not require finding an underlying statistical model, as it is expressed as $\eta^2 = \sigma^2/\mu^2$, where μ is the mean and σ is the standard deviation of the abundance distribution. The technical noise for f-LAESI was found to be $\eta_t^2 = 0.02$ and $\eta_t^2 = 0.03$ for glutamate and glucarate standards, respectively.[12,38] This is significantly lower than the measured noise, η^2, for all of the studied metabolites. This technical noise, however, does not account for the variations in the sampled volume. A qualitative argument could be made to support the assumption of relatively low variations in the ablated sample volume. As the laser pulse energy exhibited <5% RSD variations, and the water concentration in the cells was relatively uniform, delivering similar amounts of energy is expected to result in similar sampled volumes. Thus, the metabolic noise is obtained by a minor correction: $\eta_m^2 = \eta^2 - \eta_t^2$ (Supplementary Table 1). For the 35 featured metabolites, the η_m^2 falls between $0.20 \leq \eta_m^2 \leq 2.76$.

Primary metabolites exhibited both low ($\eta_m^2 \leq 0.50$) and high ($\eta_m^2 > 0.50$) metabolic noise. For example, ascorbic acid ($\eta_m^2 = 0.36$), oxalate ($\eta_m^2 = 0.40$), and pyruvate ($\eta_m^2 = 0.54$) are among the metabolites with lower noise, whereas glutathione ($\eta_m^2 = 0.89$), gluconic acid ($\eta_m^2 = 2.09$), and gallic acid ($\eta_m^2 = 2.44$) exhibit much higher noise.

To explore if the mean abundance of a metabolite ion affected the metabolic noise, the latter was plotted against the mean values and the data showed a weak negative correlation (see Figure 3d). It is important to emphasize that we do not compare ion intensities between different metabolites or to metabolite concentrations in the tissue. The abundance histograms in Figure 3 reflect the distributions of ion intensities for a particular metabolite originating from different cells. As mentioned above, in metabolic noise values, all linear distortions of ion intensities are eliminated because they only depended on intensity ratios (intensity in a cell over the intensity average for all cells) not on the intensities themselves.

Morphometric analysis of the images provided cell volume values for each analyzed cell. This can be used to explore a correlation between the cell volume and metabolite abundance distributions. Cell volumes for the entire population (n = 1084) exhibited a lognormal distribution with a median of 4.5 nL and a range of 1.4−10.0 nL (Figure 4c).

Pearson, ρ_P, and Spearman, ρ_S, correlation coefficients were calculated between cell volumes and 28 metabolite ion

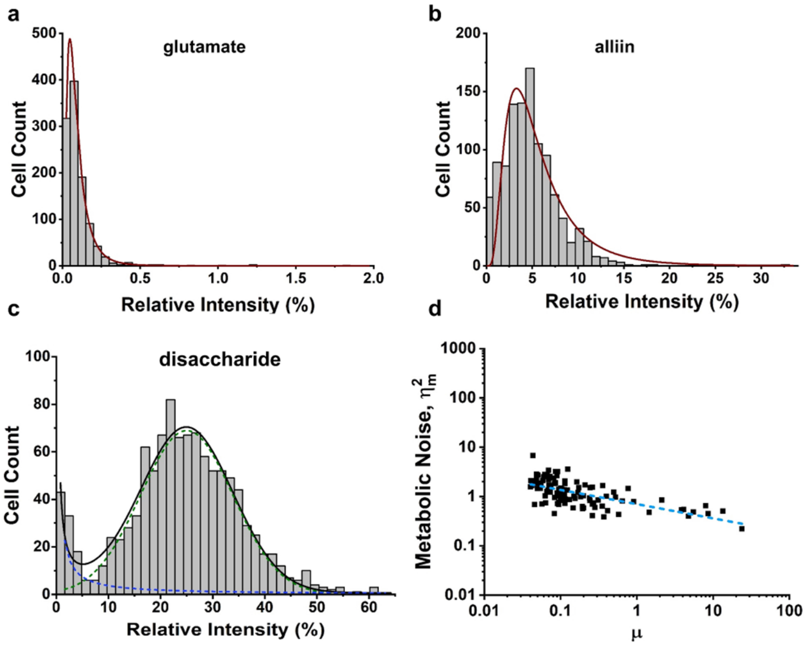


Figure 3. Metabolite abundance distributions for *A. cepa* epidermal cells (n = 1084) of (a) glutamate, (b) alliin, and (c) disaccharide. Distributions followed either a lognormal or bimodal normal curve. (d) Metabolic noise as a function of mean relative intensity from 1084 single *A. cepa* cells shows a slightly negative Pearson's correlation with $\rho_P = -0.64$.

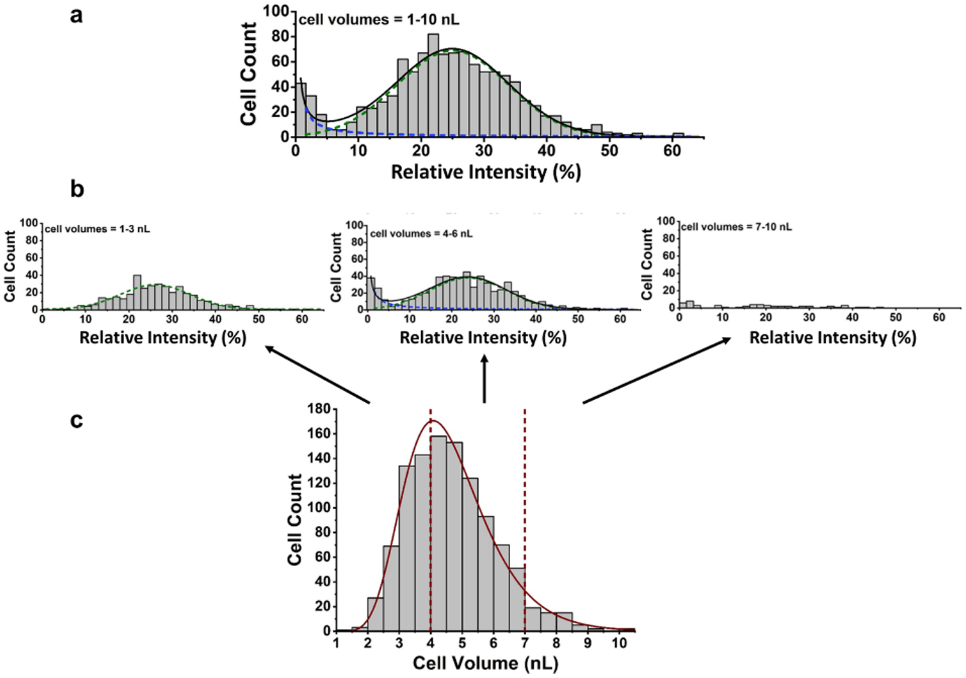


Figure 4. (a) Abundance distribution of disaccharides for a cell population of n = 1084 exhibits a bimodal distribution. (b) Abundance distributions for selected cell volume ranges: 1−4 nL (n = 377), 4−7 nL (n = 649), and 7−10 nL (n = 58) were investigated within the disaccharide distribution. (c) Lognormal cell volume distribution was observed for the analyzed 1084 single *A. cepa* cells.

intensities to characterize the strength of potential linear and nonlinear relationships, respectively. None or very weak correlations were observed for the studied ions with the cell volume. Specifically, for di-, tri-, and pentasaccharide ion intensities very weak negative correlations of ρ_P = −0.27, −0.32, and −0.24 were observed, respectively. All other correlations between volumes and ion abundances were even weaker, i.e., practically nonexistent, exhibiting $-0.14 < \rho_P < 0.10$ (Supplementary Table 2). This can be rationalized by recognizing that the ablation process does not consume the entire cell content, especially in the case of midsize (4−7 nL) and large (7−10 nL) cells. Postablation observation of the cells by microscopy confirmed this point, i.e., part of the cell content in these cases remained in the confines of the remaining cell walls. As the volume sampled by 10 laser shots is smaller than most of the cell volumes determined by morphometry, there is no significant correlation expected between metabolite abundances and the cell volume.

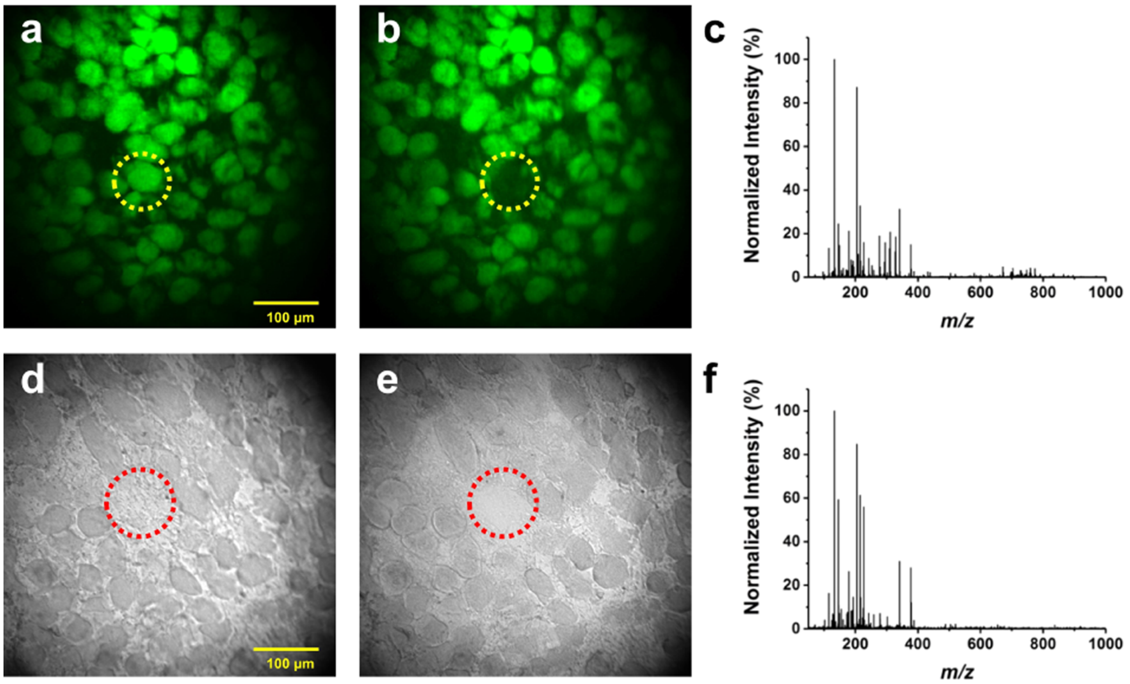


Figure 5. Selective sampling of specific cell types by f-LAESI. Fluorescence imaging of a 30 μm thick cross-section of a root nodule reveals only cells infected with nitrogen-fixing bacteria. Images show (a) before and (b) after the ablation of a single infected cell and the corresponding (c) mass spectrum. The corresponding brightfield images display both infected and uninfected cell types and reveal the uninfected cell clusters. Images show (d) before and (e) after the ablation of a small cell cluster of uninfected cells (n = 10 cells) with the (f) corresponding mass spectrum.

Conversely, ablating a relatively constant volume of the cell content, independent of cell volume, is a prerequisite for relative quantitation.

Dissecting the bimodal disaccharide abundance distribution reveals that cells with low (1−4 nL) volumes only contribute to the normal distribution (Figure 4a,b). In contrast, the midsize and large cells in the 4−7 and 7−10 nL volume range, respectively, contribute to both the lognormal and normal components of the bimodal distribution (Figure 4b). Thus, single-cell analysis combined with morphometry reveals two subpopulations of cells distinguished by their disaccharide abundance and volume. Small cells constitute a subpopulation with elevated disaccharide content, whereas among larger cells, two phenotypes exist, one with low and one with elevated disaccharide abundance.

There is indirect evidence in the literature for the lack of cell volume dependence of disaccharide abundances. Following the average cell sizes in onion epidermis as a function of scale number, it was found that for a mature spherical bulb, between scales 6th (outer) and 11th (inner), the average length and diameter of cells dropped by factors of ~2.5 and ~2.6, respectively.[39] This corresponded to a cell volume reduction by a factor of ~17 in a span of five scales. In an earlier f-LAESI-MS study of epidermal cells from five consecutive scales, we found no significant difference between disaccharide signal intensities as a function of scale number.[24]

Calculating correlation coefficients between 378 possible mixed pairs of the 28 ions revealed only a few significant correlations with $\rho_P > 0.70$ or $\rho_S > 0.70$. Pearson correlation calculations indicated for fumarate and malate, $\rho_P = 0.79$, citrate and gallic acid, $\rho_P = 0.71$, pentose and oxalate, $\rho_P = 0.73$, pentose and arabinonate, $\rho_P = 0.76$, pentose and hexose, $\rho_P = 0.83$, glutathione and hexose phosphate, $\rho_P = 0.83$, UDP-hexose and hexose phosphate, $\rho_P = 0.79$, and UDP-hexose and glutathione, $\rho_P = 0.81$ (Supplementary Table 2). In these analyses, hexose stands for glucose and fructose, and pentose includes ribose, arabinose, and xylulose.

Low correlation for many of the metabolite pairs is not surprising because of the complexity of metabolic pathways. Most metabolites are precursors for multiple reaction products, and at the same time, products in multiple reactions. Changes in their concentrations are determined by multiple upstream precursors and multiple downstream reaction products, i.e., they are branching nodes of the network. Thus, there is no strong correlation between the precursor and product concentrations of any single reaction.

3.5. Dimensionality Reduction and Data Visualization.

The 1084 single-cell spectra with ~100 spectral features each constitutes a high-dimensional data set. To assess the major features in the data, dimensionality reduction is performed based on UMAP analysis. The lower-dimensional representation of this data set is shown in Supplementary Figure 6a. It reveals the presence of a distinct spectral feature that corresponds to a batch-effect (highlighted by yellow) characterized by the presence of the m/z 202.970 peak originating from perfluorobutanoic acid, a known electrospray contaminant. The spectra are further analyzed by UMAP for three metabolites, glutamate, alliin, and disaccharides, to visualize their heterogeneity (Supplementary Figure 6b−d). The bimodal nature of the disaccharide distribution is represented by color differences, whereas glutamate and alliin show no separation among the cells.

3.6. Selective Targeting of Specific Cell Types.

Targeting cell types can be important for the analysis of rare cells or when the cell type of interest is intimately interspersed with other cell types. A well-known example of this is the infection zone of root nodules in legumes. Here, cells infected by nitrogen-fixing rhizobia are distributed among uninfected

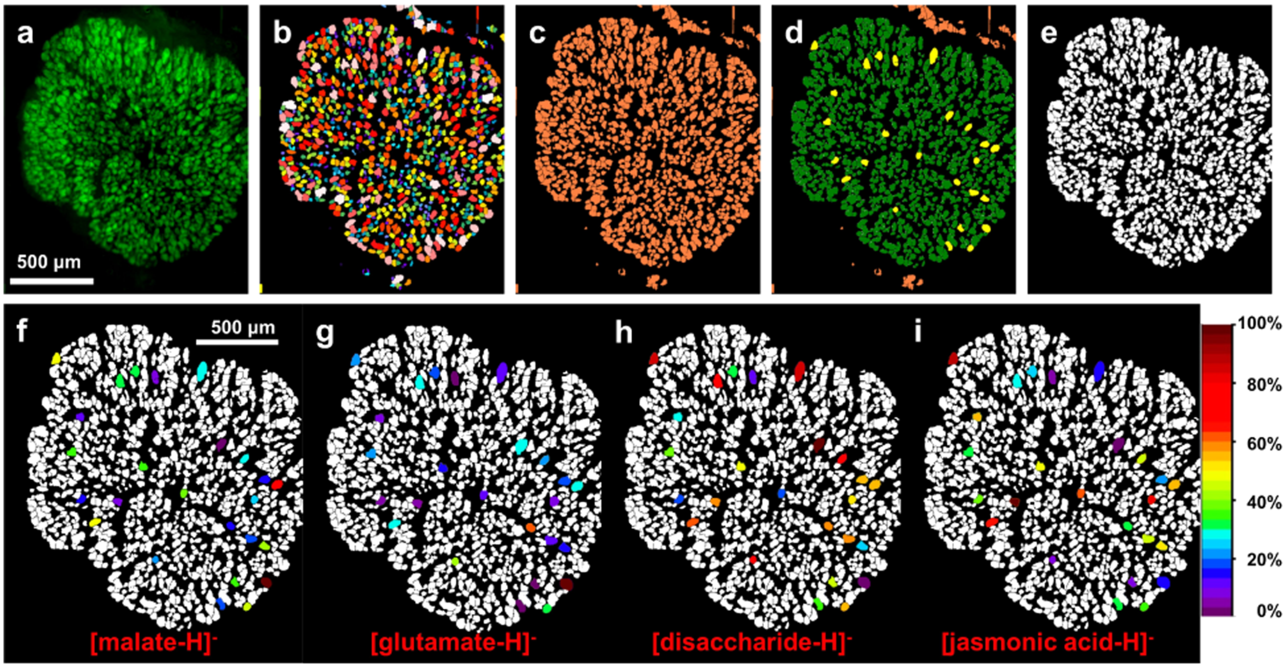


Figure 6. Workflow for selective targeting of infected soybean root nodule cells for analysis. (a) High-resolution fluorescence image was obtained by image stitching. (b) Image segmentation was applied to the raw image to provide cell border separation and grouping based on color. (c) Cell centroid locations were determined, and (d) 25 cells were selected to be sampled using the automated system. (e) Image mask was created for false-color spatial mapping of (f–i) several metabolites.

cells on a very fine scale. To demonstrate selective targeting of a specific cell type, frozen root nodules from 21 day old soybean plants, infected by GFP-labeled rhizobia, are sectioned. Brightfield and fluorescence images are captured, to identify the infected cells based on their fluorescence and the uninfected cells based on their smaller size and the lack of fluorescence. Figure 5d shows the brightfield image of a root nodule cross section depicting all of the cell types present. The corresponding fluorescence image (Figure 5a) highlights only the infected cells. This enables the selective targeting of infected cells and uninfected cell clusters. Analysis of a single infected cell is demonstrated by comparing before (Figure 5a) and after ablation (Figure 5b) images and capturing the corresponding mass spectrum (Figure 5e). Due to the small size of the uninfected cells, a cluster of them ($n = 10$) is ablated (Figure 5d,e) and analyzed (Figure 5f).

The mass spectra of the two cell types are compared (Figure 5c,f) and multiple metabolites, including malate, disaccharides, glutamate, and citrate, are identified in both. The most prominent difference is in the lipid region where membrane lipids are significantly more abundant in the infected cells. Phosphatidic acid (PA) species, including PA (16:0/18:2), and PA (18:2/18:2), phosphatidylglycerol (PG) species, including PG (16:1/18:2), and PG (16:0/18:1), and phosphatidylinositol (PI) species, including PI (16:0/18:2), are observed. The presence of these lipids in root nodules has been established by f-LAESI-MS of single infected cells on a 21T-FTICR instrument with ultrahigh mass accuracy.[38]

Randomly selected cells of the same type from different parts of the tissue can be analyzed automatically and with high throughput. To demonstrate this, a *G. max* root nodule section is scanned using the fluorescence channel to reveal only infected cells (Figure 6a). Image processing is performed to define cell borders, remove image artifacts, and define the cells of interest, and an image mask is created (Figure 6b–e). Spurious objects were rejected by intensity thresholding and/or by excluding objects that were too large or too small to be cells. Once the cells are recognized as objects, the (x_i, y_i) centroid coordinates for each cell are stored for f-LAESI sampling. Processing the image in Figure 6a results in the identification of 936 infected root nodule cells. From this pool, 25 cells are randomly selected for analysis. Abundances of several metabolites, including malate, glutamate, disaccharide, and jasmonic acid, are spatially mapped onto the 25 selected cells (Figure 6f–i).

4. DISCUSSION

High-throughput single-cell analysis methods to capture metabolic heterogeneity are available; however, they are applicable mainly for free-floating cells and report on the abundances of a few metabolites. Fluorescence microscopy enables the selection of specific subpopulations for single-cell analysis without interference from other cell types. For example, fluorescence microscopy was used to guide capillary microsampling for electrospray ionization IMS-MS of individual HepG2/C3A cells in specific mitotic stages.[9] Due to manual control of capillary microsampling, this type of analysis is extremely labor-intensive and low throughput. Typically, only a few cells are analyzed in 1 h. Much higher throughput can be achieved by studying isolated cell populations using object recognition in fluorescence microscopy followed by MALDI-MS of a selected subpopulation.[19] A promising technique, laser microdissection/liquid vortex capture/MS, was automated to recognize cell location in whole tissues and free-floating cells by optical microscopy and image processing. A sampling rate of 100 algae cells/h was achieved, and algal cell types were differentiated based on lipid signatures.[21] In our automated f-LAESI workflow, the fluorescence channel provided the ability to sample particular phenotypes, and to perform analysis with comparable 120 cells/h throughput.

Due to the large number of analyzed cells ($n > 1000$) and the >100 mass spectral features for each of them, an advanced multivariate statistical approach is needed. Dimensionality reduction has been fundamental for the visualization of high-dimensional data with nonlinearities.[40] As UMAP is a nonlinear dimensionality-reduction method, it preserves both local and global data structures needed to identify distinct cell populations.[37] Additionally, this platform can accommodate millions of data points; thus, it is applicable for high-throughput large-scale experiments. Displaying spectral intensities of metabolites on a false-color scale in UMAP can provide insight into the presence of subpopulations. For example, when looking at the histogram plots of glutamate and alliin, both exhibit a lognormal distribution (Figure 3a,b). However, the disaccharide distribution consists of a lognormal and normal component, indicating the presence of two subpopulations (Figure 3c). This can also be observed through the wide color gradient in the UMAP representation (Supplementary Figure 6d).

Elevated Pearson correlation coefficients indicated tracking in the abundances of eight metabolite ion pairs. Some of these pairs were found close to each other on specific metabolic pathways. Fumarate and malate ($\rho_P = 0.79$) are directly interconverted in the tricarboxylic acid (TCA) cycle by fumarate hydratase [EC:4.2.1.2]; thus, their strong correlation can be rationalized. However, in-source fragmentation can also contribute to this correlation, as the loss of water from the malate ion produces fumarate.[41] The correlation between pentose and arabinonate ($\rho_P = 0.76$) can be attributed to the two-step interconversion between arabinose and arabinonate through arabonolactone, catalyzed by arabinonolactonase [EC:3.1.1.15] and arabinose dehydrogenases [EC:1.1.1.46] and [EC:1.1.1.376]. There is evidence in the literature for a multistep link between a pentose (arabinose) and oxalate ($\rho_P = 0.73$). In *B. japonicum*, formation of oxalate was linked to the degradation of arabinose.[42] Hexose phosphate and glutathione ($\rho_P = 0.83$) are linked through NADPH produced by the pentose phosphate pathway and used in the reduction of oxidized glutathione by glutathione reductase [EC:1.8.1.7]. The pentose phosphate pathway also connects pentoses with hexoses potentially explaining the observed strong correlation in our single-cell data ($\rho_P = 0.83$).

The correlation between UDP-hexose and hexose phosphate ($\rho_P = 0.79$) probably corresponds to the UDP-glucose—glucose phosphate pair involved in starch and sucrose metabolism, where glucose phosphate is a direct hydrolysis product of UDP-glucose. Starch and sucrose metabolism is a primary pathway in *A. cepa* bulb swelling.[43] Correlations between UDP-hexose and hexose phosphate in combination with the correlations of hexose phosphate and glutathione also mean that UDP-hexose and glutathione are corelated ($\rho_P = 0.81$). Finally, the origin of the correlation between the single-cell abundances of citrate and gallic acid ($\rho_P = 0.71$) is unclear.

Most laser-based MSI methods analyze tissue voxels arranged on a rectangular grid. Even at the high spatial resolution, analysis of tissue-embedded single cells is hard to achieve due to the occasional cosampling of two or more cells within an imaging voxel. This results in scrambled spectra. With the automated f-LAESI system, as long as the cells are larger than the fiber tip diameter, the mass spectrum for each voxel is from a single cell, as laser sampling occurs at cell centroids as opposed to locations defined by fixed increments on a rectangular grid.

The feasibility of f-LAESI for MSI is tested on rhizobia infected soybean root nodule cells randomly selected from an entire nodule cross section (Figure 6). Three-dimensional molecular tomography of such nodules has been performed by MALDI-MSI at a step size of 50 μm. According to that study, several metabolites exhibit asymmetrical molecular distributions throughout the infection zone, where infected and uninfected cells are interspersed.[44] Although useful data were obtained, the lack of single-cell resolution resulted in potentially co-measuring the metabolite compositions of infected and uninfected cells. In contrast, using automated f-LAESI-MSI, the metabolite composition of single infected cells is captured.

Looking at cell-to-cell variations in our experiment, glutamate ($\eta_m^2 = 0.44$) showed greater noise levels than malate ($\eta_m^2 = 0.12$), jasmonic acid ($\eta_m^2 = 0.16$), and disaccharide ($\eta_m^2 = 0.10$). In previous studies using f-LAESI-IMS-MS and f-LAESI 21T-FTICR on infected cells in *G. max* root nodules, very similar metabolic noise, $\eta_m^2 = 0.08$, was observed for disaccharides at $n = 60$ and $n = 124$ cell numbers, respectively.[12,38] Metabolic noise measured by f-LAESI-IMS-MS at $n = 60$ for glutamate, $\eta_m^2 = 0.27$, and malate, $\eta_m^2 = 0.22$ are similar to the values found in this study.[12] Glutamate, malate, and jasmonic acid were not observed using the f-LAESI 21T-FTICR system due to the low mass cutoff in data acquisition.

In all three studies, the cell numbers for *G. max* have been small to marginal to accurately establish the shape of abundance distribution functions. To assess the number of cells needed to capture these distributions, the $n = 1084$ data set is randomly downsampled to $n = 1000$, 600, 400, 200, 100, and 50 cells. Fumarate abundances at $n = 1084$ exhibit a bimodal distribution with a lognormal and a normal component (see Supplementary Figure 2). Using reduced sample sizes, histograms are created for fumarate abundances and fitted with the same bimodal model (see Supplementary Figure 7). As the number of cells is reduced from $n = 1000$ to $n = 600$, 400, 200, 100, and 50 cells, the fit deteriorates as it is reflected in reduced coefficients of determination, $r^2 = 0.96$, 0.95, 0.87, 0.78, 0.79, and 0.44, respectively. It appears that the minimum required cell number to capture the shape of the bimodal fumarate distribution is $n > 200$. In the case of simpler unimodal distributions, somewhat lower cell numbers, e.g., $n > 100$, might be sufficient. In general, the number of cells needed depend on the number and relative size of the subpopulations present.

5. CONCLUSIONS

Automated analysis of $n > 1000$ tissue-embedded single cells is demonstrated under ambient conditions using f-LAESI-MS to explore cell-to-cell variations of metabolite abundances. Detecting bimodal abundance distributions leads to the identification of cellular subpopulations with low and high metabolite levels. Improving the throughput of single-cell metabolomics puts it on more even footing with single-cell transcriptomics and proteomics and prepares the ground for the systemic analysis of individual cells through multiomics. Combining this data with morphometry can uncover correlations between cellular composition and morphological phenotypes. Selectively analyzing cells tagged by affinity-based or genetically encoded fluorescent probes using f-LAESI-MS can reveal the relationship between functional phenotypes and cellular metabolism. Object recognition in optical microscopy can be used to establish single-cell voxels for cell-by-cell MSI experiments. Resolving difficulties presented by significantly smaller mammalian cell sizes, primarily by improving detection sensitivity, will open the door to the analysis of single mammalian cells in a tissue environment by f-LAESI-MS.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.1c00569.

> Metabolite abundance and UMAP distributions from *A. cepa* epidermal cells (n = 1084); metabolic noise as a function of mean relative intensity; visualization of 1084 *A. cepa* single-cell data using UMAP; assessment of cell numbers needed to determine abundance distributions; and metabolites annotated (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Sylwia A. Stopka − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States; Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, United States; Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, United States;* orcid.org/0000-0003-3761-6899; Phone: +1 (617) 525-9746; Email: sstopka@bwh.harvard.edu

Akos Vertes − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States;* orcid.org/0000-0001-5186-5352; Phone: +1 (202) 994-2717; Email: vertes@gwu.edu

Authors

Ellen A. Wood − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States*

Rikkita Khattar − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States*

Beverly J. Agtuca − *Divisions of Plant Sciences and Biochemistry, Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, Missouri 65211, United States*

Walid M. Abdelmoula − *Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, United States;* orcid.org/0000-0003-3117-7389

Nathalie Y. R. Agar − *Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, United States; Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, United States; Department of Cancer Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, United States;* orcid.org/0000-0003-3149-3146

Gary Stacey − *Divisions of Plant Sciences and Biochemistry, Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, Missouri 65211, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.1c00569

Author Contributions

A.V. and S.A.S. conceived the study, and S.A.S., E.A.W., and R.K. conducted the experiments. B.J.A. grew the *G. max* plants. W.M.A. and N.Y.R.A. performed the statistical analysis. S.A.S. and A.V. performed the data analysis, and S.A.S. and A.V. wrote the manuscript with input from G.S.

Notes

The authors declare the following competing financial interest(s): In compliance with Harvard Medical School and Partners Healthcare guidelines on potential conflict of interest, we disclose that N.Y.R. Agar is a consultant/advisory board member for BayesianDx and InviCRO. All other authors have no conflict of interest.

■ ACKNOWLEDGMENTS

This material is based on the work supported by the U.S. National Science Foundation Plant Genome Research Program under Award Number IoS-1734145 and by the U.S. Department of Energy, Office of Biological and Environmental Research under Award Number DE-SC0013978.

■ REFERENCES

(1) Zhang, L. W.; Vertes, A. *Angew. Chem., Int. Ed.* **2018**, *57*, 4466−4477.

(2) Altelaar, A. F. M.; Luxembourg, S. L.; McDonnell, L. A.; Piersma, S. R.; Heeren, R. M. A. *Nat. Protoc.* **2007**, *2*, 1185−1196.

(3) Neumann, E. K.; Do, T. D.; Comi, T. J.; Sweedler, J. V. *Angew. Chem., Int. Ed.* **2019**, *58*, 9348−9364.

(4) Zenobi, R. *Science* **2013**, *342*, No. 1243259.

(5) Ziegenhain, C.; Vieth, B.; Parekh, S.; Reinius, B.; Guillaumet-Adkins, A.; Smets, M.; Leonhardt, H.; Heyn, H.; Hellmann, I.; Enard, W. *Mol. Cell* **2017**, *65*, 631−643.

(6) Patel, A. P.; Tirosh, I.; Trombetta, J. J.; Shalek, A. K.; Gillespie, S. M.; Wakimoto, H.; Cahill, D. P.; Nahed, B. V.; Curry, W. T.; Martuza, R. L.; Louis, D. N.; Rozenblatt-Rosen, O.; Suva, M. L.; Regev, A.; Bernstein, B. E. *Science* **2014**, *344*, 1396−1401.

(7) Jaitin, D. A.; Kenigsberg, E.; Keren-Shaul, H.; Elefant, N.; Paul, F.; Zaretsky, I.; Mildner, A.; Cohen, N.; Jung, S.; Tanay, A.; Amit, I. *Science* **2014**, *343*, 776−779.

(8) Patabadige, D. E. W.; Mickleburgh, T.; Ferris, L.; Brummer, G.; Culbertson, A. H.; Culbertson, C. T. *Electrophoresis* **2016**, *37*, 1337−1344.

(9) Zhang, L. W.; Sevinsky, C. J.; Davis, B. M.; Vertes, A. *Anal. Chem.* **2018**, *90*, 4626−4634.

(10) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R. M.; Burgett, A. W. G.; Yang, Z. B. *Anal. Chem.* **2014**, *86*, 9376−9380.

(11) Onjiko, R. M.; Moody, S. A.; Nemes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 6545−6550.

(12) Stopka, S. A.; Khattar, R.; Agtuca, B. J.; Anderton, C. R.; Pasa-Tolic, L.; Stacey, G.; Vertes, A. *Front. Plant Sci.* **2018**, *9*, No. 1646.

(13) Gilmore, I. S.; Heiles, S.; Pieterse, C. L. *Annu. Rev. Anal. Chem.* **2019**, *12*, 201−224.

(14) McDonnell, L. A.; Heeren, R. M. A. *Mass Spectrom. Rev.* **2007**, *26*, 606−643.

(15) Li, L. J.; Garden, R. W.; Sweedler, J. V. *Trends Biotechnol.* **2000**, *18*, 151−160.

(16) Korte, A. R.; Yandeau-Nelson, M. D.; Nikolau, B. J.; Lee, Y. J. *Anal. Bioanal. Chem.* **2015**, *407*, 2301−2309.

(17) Fletcher, J. S. *Analyst* **2009**, *134*, 2204−2215.

(18) Lanni, E. J.; Dunham, S. J. B.; Nemes, P.; Rubakhin, S. S.; Sweedler, J. V. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1897−1907.

(19) Ong, T. H.; Kissick, D. J.; Jansson, E. T.; Comi, T. J.; Romanova, E. V.; Rubakhin, S. S.; Sweedler, J. V. *Anal. Chem.* **2015**, *87*, 7036−7042.

(20) Neumann, E. K.; Ellis, J. F.; Triplett, A. E.; Rubakhin, S. S.; Sweedler, J. V. *Anal. Chem.* **2019**, *91*, 7871−7878.

(21) Cahill, J. F.; Kertesz, V. *Front. Plant Sci.* **2018**, *9*, No. 1211.

(22) Feider, C. L.; Krieger, A.; DeHoog, R. J.; Eberlin, L. S. *Anal. Chem.* **2019**, *91*, 4266−4290.

(23) Yang, Y. Y.; Huang, Y. Y.; Wu, J. H.; Liu, N.; Deng, J. W.; Luan, T. G. *TrAC, Trends Anal. Chem.* **2017**, *90*, 14−26.

(24) Shrestha, B.; Vertes, A. *Anal. Chem.* **2009**, *81*, 8265−8271.

(25) Rao, W.; Pan, N.; Yang, Z. B. *J. Visualized Exp.* **2016**, *112*, No. e53911.

(26) Nguyen, S.; Sontag, R.; Carson, J.; Corley, R.; Ansong, C.; Laskin, J. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 316−322.

(27) Yin, R. C.; Burnum-Johnson, K. E.; Sun, X. F.; Dey, S. K.; Laskin, J. *Nat. Protoc.* **2019**, *14*, 3445−3470.

(28) Li, H.; Smith, B. K.; Shrestha, B.; Mark, L.; Vertes, A. *Mass Spectrometry Imaging of Small Molecules*; In He, L., Ed.; Humana Press: New York, 2015; Vol. *1203*, pp 117−127.

(29) Tscherepanow, M.; Zollner, F.; Hillebrand, A.; Kummert, F. *Advances in Mass Data Analysis of Images and Signals in Medicine, Biotechnology, Chemistry and Food Industry*; In Perner, P.; Salvetti, O., Eds.; Springer: New York, 2008; Vol. *5108*, pp 158−172.

(30) Puchkov, E. O. *Microbiology* **2019**, *88*, 1−14.

(31) Musat, N.; Halm, H.; Winterholler, B.; Hoppe, P.; Peduzzi, S.; Hillion, F.; Horreard, F.; Amann, R.; Jorgensen, B. B.; Kuypers, M. M. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17861−17866.

(32) Kopf, S. H.; Sessions, A. L.; Cowley, E. S.; Reyes, C.; Van Sambeek, L.; Hu, Y.; Orphan, V. J.; Kato, R.; Newman, D. K. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, E110−E116.

(33) Hsu, C. C.; White, N. M.; Hayashi, M.; Lin, E. C.; Poon, T.; Banerjee, I.; Chen, J.; Pfaff, S. L.; Macagno, E. R.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 14855−14860.

(34) Nguyen, S. N.; Liyu, A. V.; Chu, R. K.; Anderton, C. R.; Laskin, J. *Anal. Chem.* **2017**, *89*, 1131−1137.

(35) Mascuch, S. J.; Moree, W. J.; Hsu, C. C.; Turner, G. G.; Cheng, T. L.; Blehert, D. S.; Kilpatrick, A. M.; Frick, W. F.; Meehan, M. J.; Dorrestein, P. C.; Gerwick, L. *PLoS One* **2015**, *10*, No. e0119668.

(36) Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. *Sci. Rep.* **2018**, *8*, No. 663.

(37) Becht, E.; McInnes, L.; Healy, J.; Dutertre, C. A.; Kwok, I. W. H.; Ng, L. G.; Ginhoux, F.; Newell, E. W. *Nat. Biotechnol.* **2019**, *37*, 38−44.

(38) Samarah, L. Z.; Khattar, R.; Tran, T.; Stopka, S. A.; Brantner, C. A.; Parlanti, P.; Velickovic, D.; Shaw, J. B.; Agtuca, B. J.; Stacey, G.; Paša-Tolić, L.; Tolič, N.; Anderton, C. R.; Vertes, A. *Anal. Chem.* **2020**, *92*, 7289−7298.

(39) Suslov, D.; Verbelen, J. P.; Vissenberg, K. *J. Exp. Bot.* **2009**, *60*, 4175−4187.

(40) Zavalin, A.; Todd, E. M.; Rawhouser, P. D.; Yang, J. H.; Norris, J. L.; Caprioli, R. M. *J. Mass Spectrom.* **2012**, *47*, 1473−1481.

(41) Xu, Y. F.; Lu, W. Y.; Rabinowitz, J. D. *Anal. Chem.* **2015**, *87*, 2273−2281.

(42) Koch, M.; Delmotte, N.; Ahrens, C. H.; Omasits, U.; Schneider, K.; Danza, F.; Padhi, B.; Murset, V.; Braissant, O.; Vorholt, J. A.; Hennecke, H.; Pessi, G. *Appl. Environ. Microbiol.* **2014**, *80*, 2094−2101.

(43) Zhang, C. S.; Zhang, H. W.; Zhan, Z. X.; Liu, B. J.; Chen, Z. T.; Liang, Y. *Front. Plant Sci.* **2016**, *7*, No. 1245.

(44) Velickovic, D.; Agtuca, B. J.; Stopka, S. A.; Vertes, A.; Koppenaal, D. W.; Paga-Tolic, L.; Stacey, G.; Anderton, C. R. *ISME J.* **2018**, *12*, 2335−2338.

https://doi.org/10.1021/acs.analchem.1c00569

Supporting Information

High-Throughput Analysis of Tissue-Embedded Single Cells by Mass Spectrometry with Bimodal Imaging and Object Recognition

Sylwia A. Stopka[a,b,c]*, Ellen A. Wood[a], Rikkita Khattar[a], Beverly J. Agtuca[d], Walid M. Abdelmoula[b], Nathalie Y. R. Agar[b,c,e], Gary Stacey[d], and Akos Vertes[a]*

[a]Department of Chemistry, The George Washington University, Washington, DC 20052

[b]Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

[c]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

[d]Divisions of Plant Sciences and Biochemistry, C. S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211

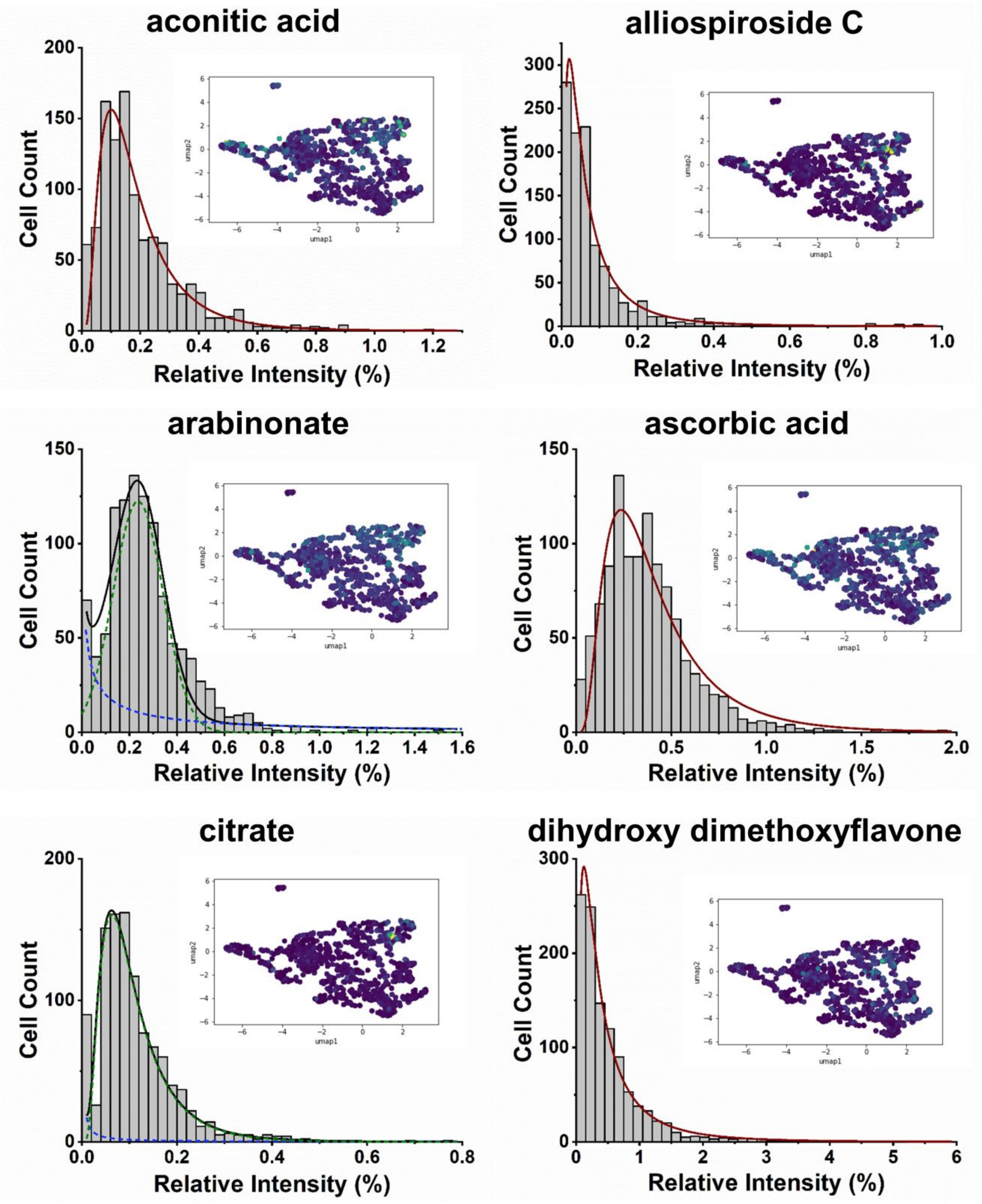
[e]Department of Cancer Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115

*Corresponding authors:
Prof. Akos Vertes
E-mail: vertes@gwu.edu, Phone: +1 (202) 994-2717, Fax: +1 (202) 994-5873
ORCID https://orcid.org/0000-0001-5186-5352

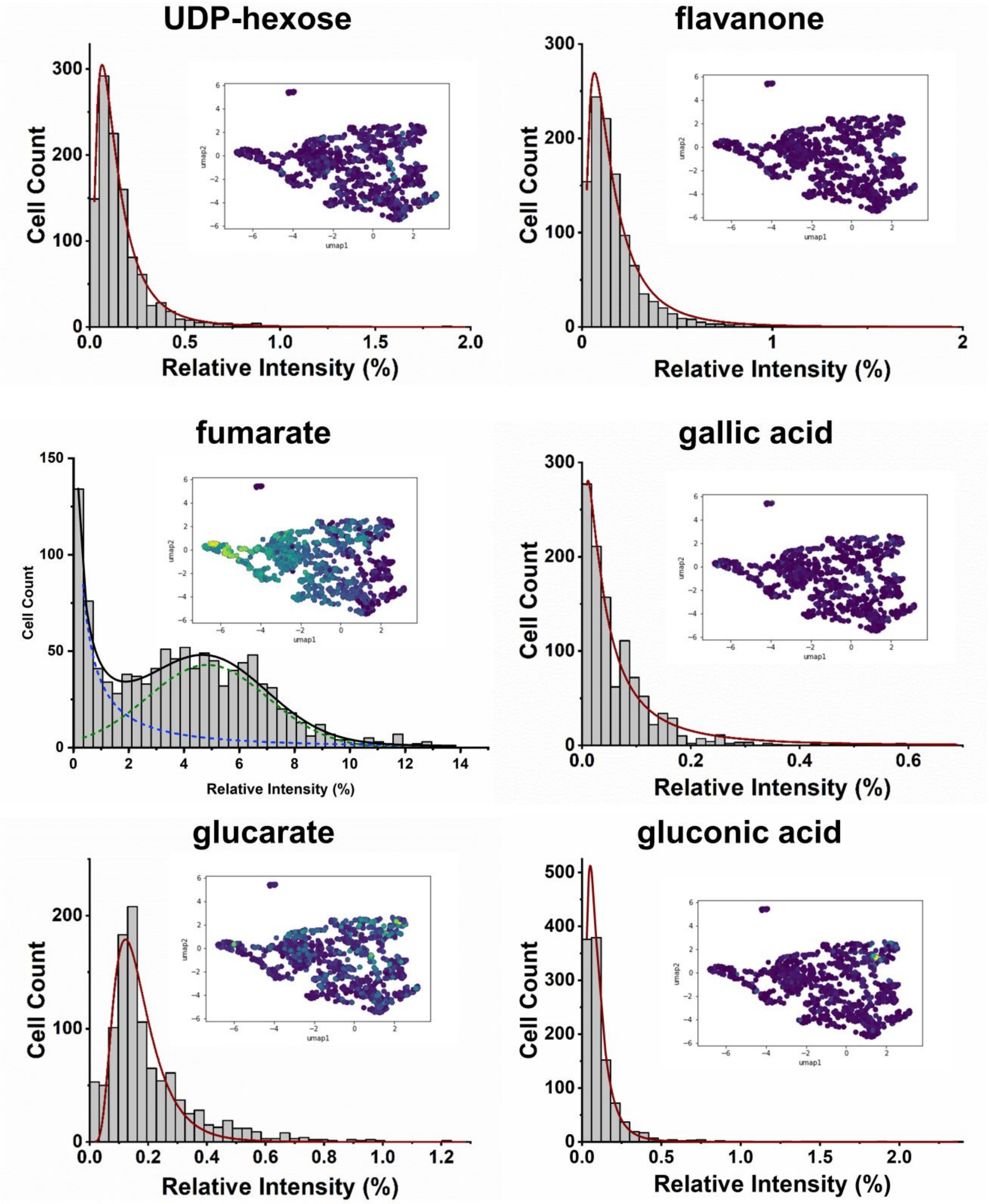
Dr. Sylwia A. Stopka
E-mail: sstopka@bwh.harvard.edu, Phone: +1 (617) 525-9746
ORCID https://orcid.org/0000-0003-3761-6899

Table of Content

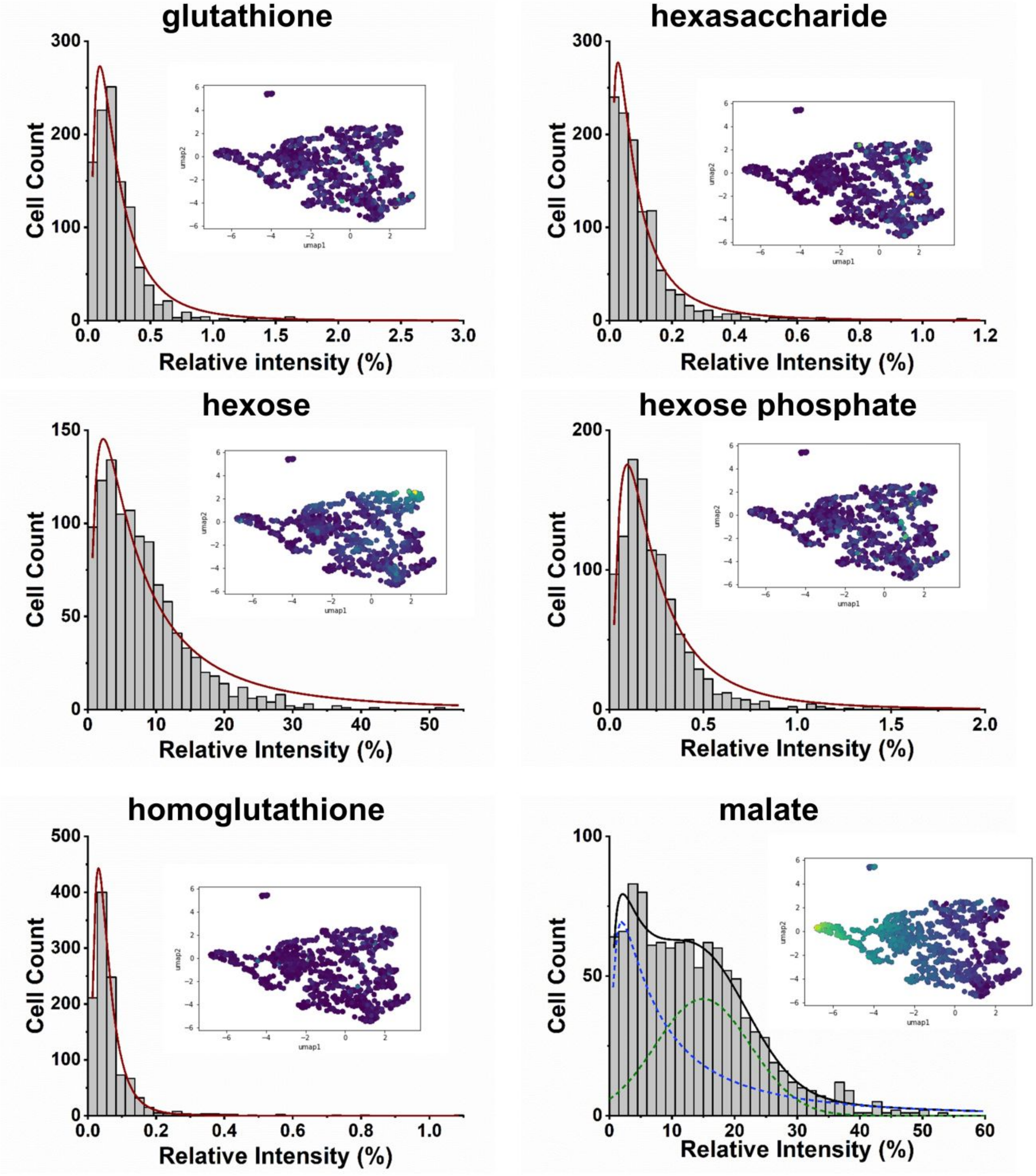
Figures:



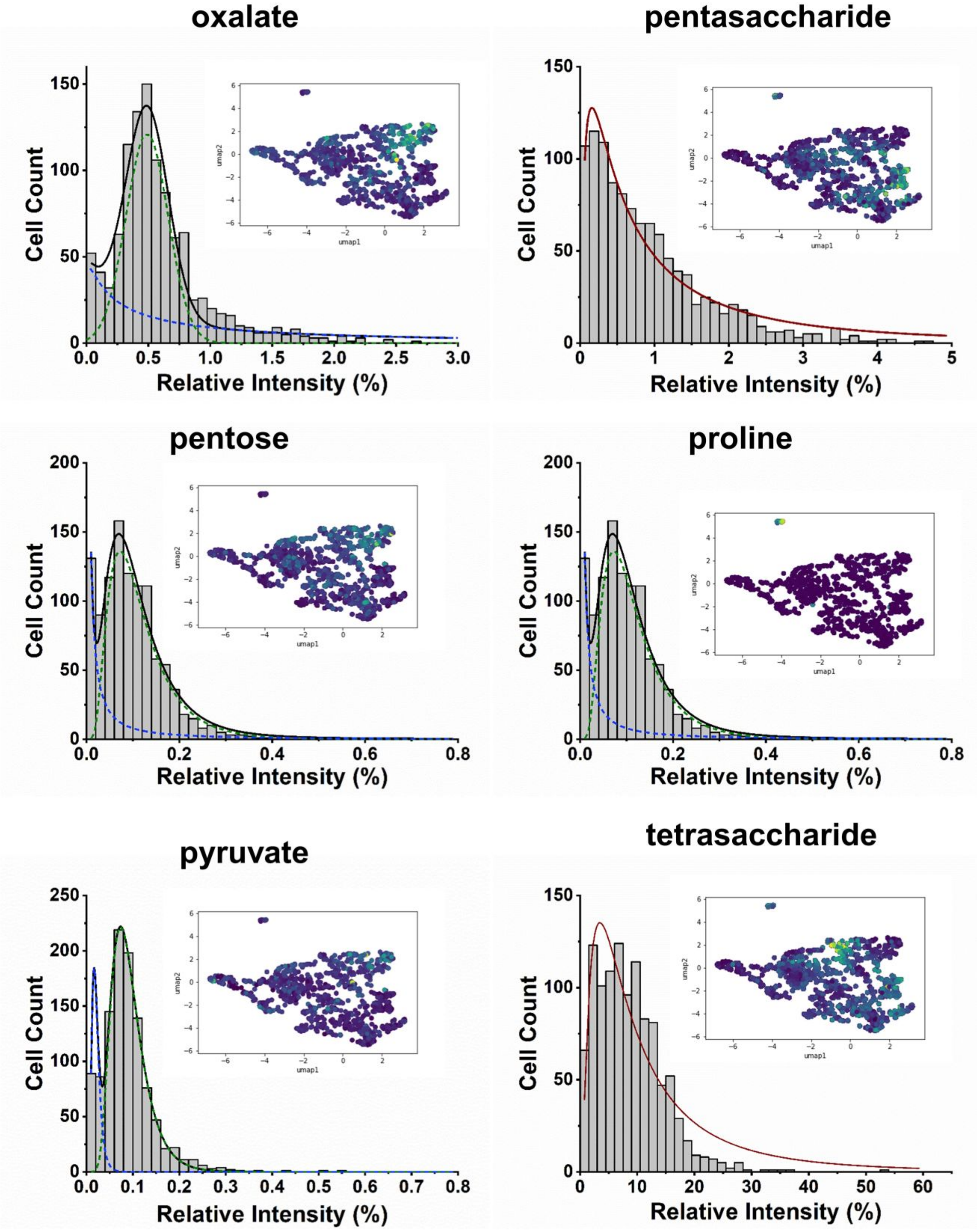
Supplementary Figure 1. Metabolite abundance and UMAP distributions from *A. cepa* epidermal cells (n = 1084) of aconitic acid, alliospiroside C, arabinonate, ascorbic acid, citrate, and dihydroxy dimethoxyflavone.



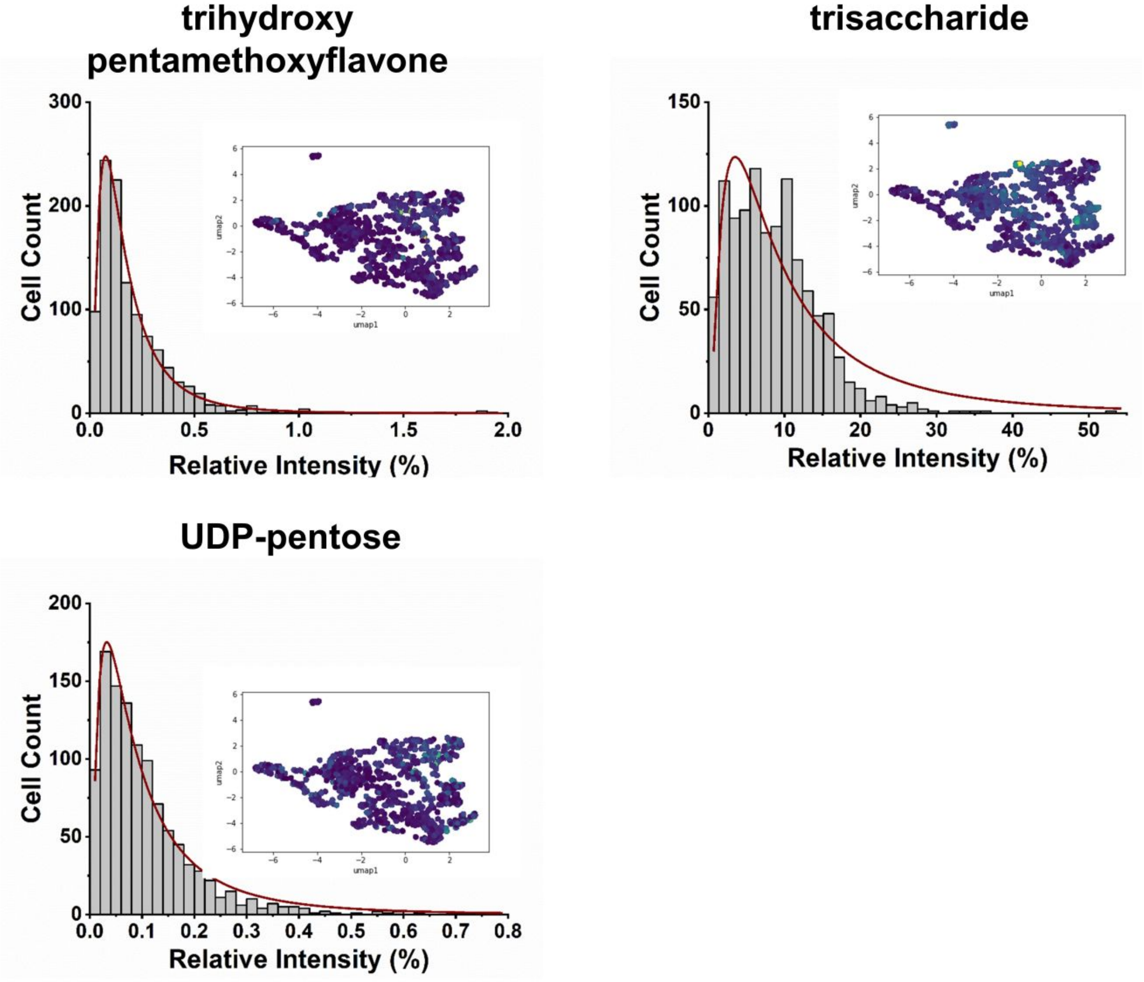
Supplementary Figure 2. Metabolite abundance and UMAP distributions from *A. cepa* epidermal cells (n = 1084) of UDP-hexose, flavanone, fumarate, gallic acid, glucarate, and gluconic acid.



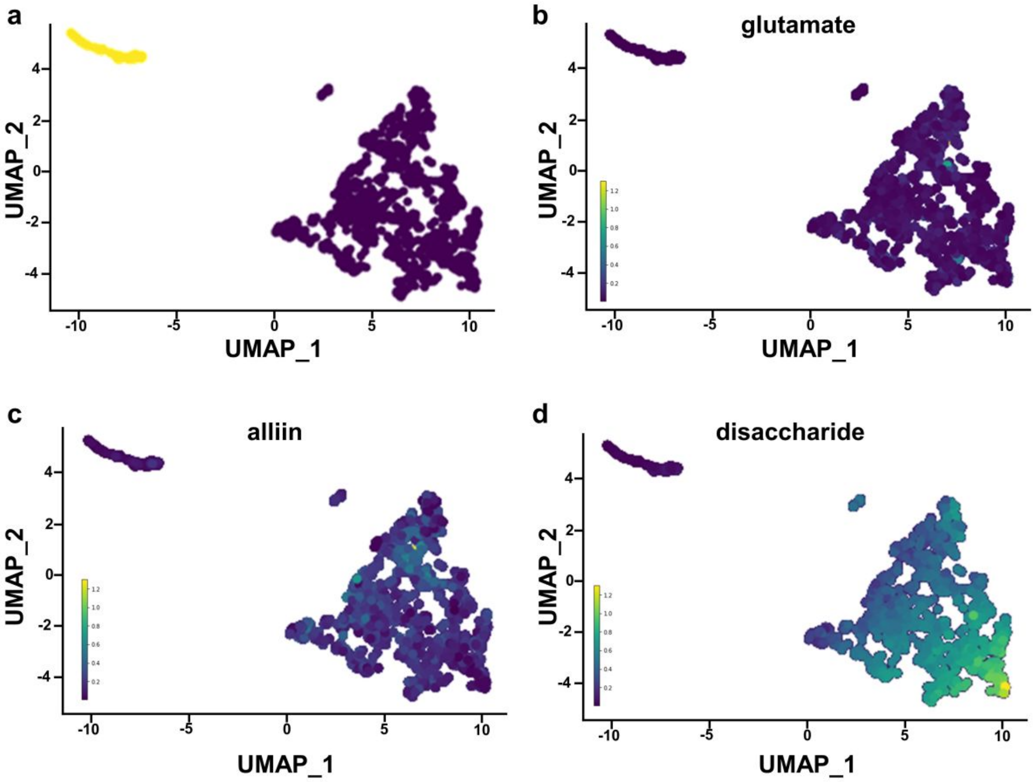
Supplementary Figure 3. Metabolite abundance and UMAP distributions from *A. cepa* epidermal cells (n = 1084) of glutathione, hexasaccharides, hexose, hexose phosphate, homoglutathione, and malate.



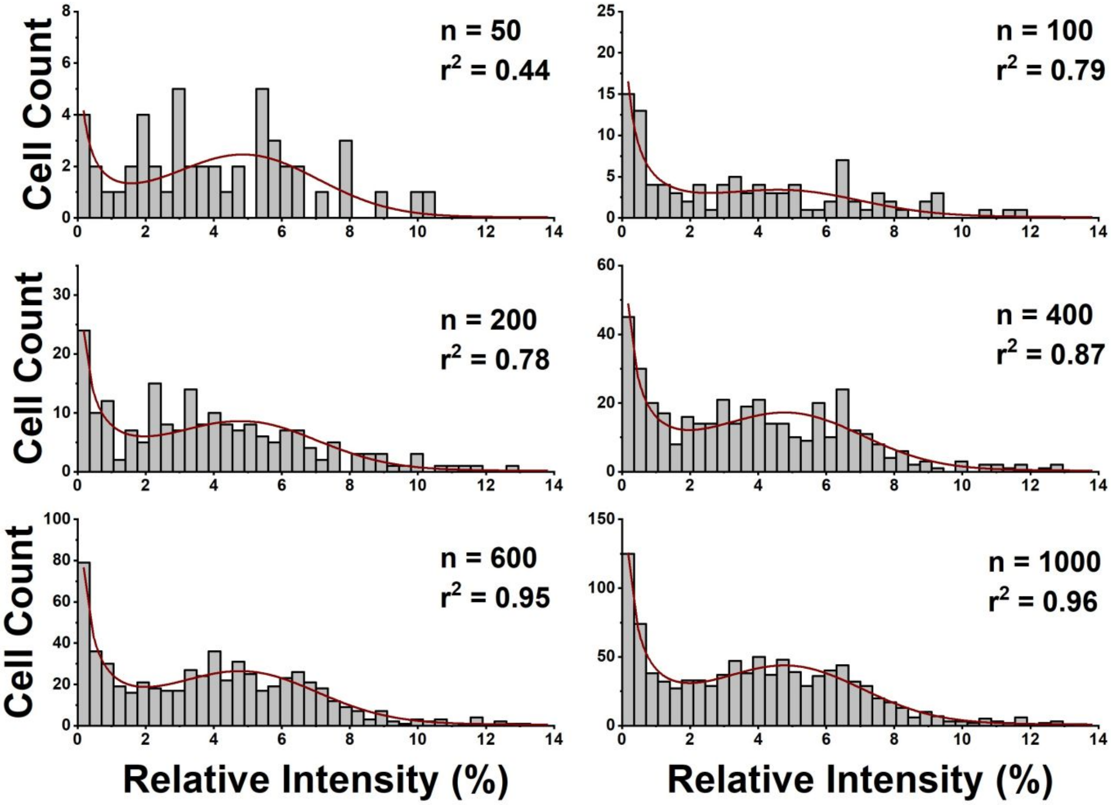
Supplementary Figure 4. Metabolite abundance and UMAP distributions from *A. cepa* epidermal cells (n = 1084) of oxalate, pentasaccharides, pentose, proline, pyruvate, and tetrasaccharide.



Supplementary Figure 5. Metabolite abundance and UMAP distributions from *A. cepa* epidermal cells (n = 1084) of trihydroxy pentamethoxyflavone, trisaccharide, and UDP-pentose.



Supplementary Figure 6. a Visualization of 1084 *A. cepa* single cell data using UMAP. The yellow-colored group of cells were clustered based on a batch effect. **b-d** Metabolite intensity of glutamate, alliin, and disaccharide are overlaid on the UMAP using false color scale, revealing distinct cell populations.



Supplementary Figure 7. Assessment of cell numbers needed to accurately capture fumarate abundance distribution in cell population. Dataset collected for n = 1084 cells is randomly downsampled to n = 1000, 600, 400, 200, 100, and 50 cells. Using the reduced sample sizes, six histograms are created for the fumarate abundances and fitted retaining the bimodal model established for the complete dataset (see Supplementary Figure 2). As the number of cells is reduced, the fit deteriorates reflected in reduced coefficients of determination, r^2 = 0.96, 0.95, 0.87, 0.78, 0.79, and 0.44, respectively.

Supplementary Table 1. Metabolites annotated from single epidermal *A. cepa* cells with corresponding metabolic noise, η_m^2, and w/w concentration, $c_{w/w}$.

Compound	**Formula**	**Adduct**	**m/z_{meas}**	**Δm (mDa)**	**μ**	**σ**	**$\eta_m^2 = \eta^2 - \eta_t^2$**	**$c_{w/w}$ (mg/100g)**	**c_{ave} (mM)**
oxalate	$C_2H_2O_4$	$[M-H]^-$	88.991	3.0	0.58	0.38	0.40	1.0[a], 4.03[b], 11.3[c]	0.60
fumarate	$C_4H_2O_4$	$[M-H]^-$	115.003	-1.2	3.85	2.82	0.51	0.63[b], 0.24[c]	0.04
malate	$C_4H_6O_5$	$[M-H]^-$	133.010	-4.2	13.46	9.58	0.48	43.6[c]	3.25
glutamate	$C_5H_9NO_4$	$[M-H]^-$	146.042	-3.9	0.09	0.11	1.47	325[c], 258[d]	19.82
pentose	$C_5H_{10}O_5$	$[M+Cl]^-$	149.040	-5.6	0.19	0.15	0.60		
arabinonate	$C_5H_{10}O_6$	$[M+Cl]^-$	165.042	2.1	0.26	0.17	0.40		
aconitic acid	$C_6H_6O_6$	$[M-H]^-$	173.001	-8.2	0.19	0.15	0.60		
ascorbic acid	$C_6H_8O_6$	$[M-H]^-$	175.025	0.3	0.37	0.23	0.36	6.69[b], 3.75[c], 7.4[d]	0.34
alliin	$C_6H_{11}NO_3S$	$[M-H]^-$	176.038	-0.4	4.76	3.00	0.37		
hexose	$C_6H_{12}O_6$	$[M-H]^-$	179.0611	5.0	8.02	6.48	0.63	1790[d]	99.41
		$[M+Cl]^-$	215.026	-6.8					
citrate	$C_6H_8O_7$	$[M-H]^-$	191.026	5.9	0.11	0.12	1.17	33.9[b], 48.5[c]	2.15
gluconic acid	$C_6H_{12}O_7$	$[M-H]^-$	195.044	-7.0	0.11	0.16	2.09		
glucarate	$C_6H_{10}O_8$	$[M-H]^-$	209.039	8.7	0.20	0.15	0.54		
flavanone	$C_{15}H_{12}O_2$	$[M-H]^-$	223.076	-0.5	0.17	0.21	1.50		
hexose phosphate	$C_6H_{13}O_9P$	$[M-H]^-$	259.021	-1.4	0.24	0.22	0.82		
glutathione	$C_{10}H_{17}N_3O_6S$	$[M-H]^-$	306.0758	-0.7	0.24	0.23	0.89		
homoglutathione	$C_{11}H_{19}N_3O_6S$	$[M-H]^-$	320.097	4.8	0.07	0.07	0.98		
disaccharide	$C_{12}H_{22}O_{11}$	$[M-H]^-$	341.109	0.5	23.91	11.22	0.20	2520[b], 990[d]	51.30
		$[M+H_2O-H]^-$	359.113	-5.6					
		$[M+Cl]^-$	377.085	-0.8					
trihydroxy pentamethoxyflavone	$C_{20}H_{20}O_{10}$	$[M-H]^-$	419.102	3.6	0.20	0.20	0.98		
trisaccharide	$C_{18}H_{32}O_{16}$	$[M-H]^-$	503.157	-4.8	8.62	5.78	0.42	1640[b]	32.53
		$[M-+H_2O-H]^-$	521.171	-0.7					
		$[M+Cl]^-$	539.136	-2.4					
UDP-pentose	$C_{14}H_{22}N_2O_{16}P_2$	$[M-H]^-$	535.035	-1.6	0.10	0.09	0.79		
UDP-hexose	$C_{15}H_{24}N_2O_{17}P_2$	$[M-H]^-$	565.044	-3.2	0.15	0.15	0.98		
tetrasaccharide	$C_{24}H_{42}O_{21}$	$[M-H]^-$	665.212	-2.6	8.62	5.78	0.42	1640[b]	19.66
pentasaccharide	$C_{30}H_{52}O_{26}$	$[M+Cl]^-$	845.268	-9.4	0.91	0.81	0.77	1640[b]	10.04
hexasaccharide	$C_{36}H_{62}O_{31}$	$[M+Cl]^-$	1025.288	-8.3	0.10	0.11	1.19		

Dihydroxy dimethoxyflavone	$C_{17}H_{14}O_6$	$[M-H]^-$	313.078	6.2	0.50	0.55	1.19		
gallic acid	$C_7H_6O_5$	$[M-H]^-$	169.012	-2.2	0.07	0.11	2.44	0.41[a]	0.02
proline	$C_5H_9NO_2$	$[M-H]^-$	114.048	-8.1	3.85	2.82	0.51	14.05[a], 12.0[d]	1.13
pyruvate	$C_3H_4O_3$	$[M-H]^-$	87.0025	-5.5	0.08	0.06	0.54	103.4[a]	11.75
alliospiroside C	$C_{38}H_{60}O_{13}$	$[M-H]^-$	723.403	7.2	0.08	0.11	1.87	49.1[a]	0.68

[a] Duke, James. 'Dr. Duke's Phytochemical and Ethnobotanical Databases. United States Department of Agriculture, Agricultural Research Service.

[b] Longvah, T., Ananthan, R., Bhaskar, K. & Venkaiah, K. *Indian food Composition Tables*. (2017).

[c] Galdon BR, Rodriguez CT, Rodriguez ER, Romero CD (2008) Organic acid contents in onion cultivars (Allium cepa L.). Journal of Agricultural and Food Chemistry 56 (15)

[d] U.S. Department of Agriculture, Agricultural Research Service. USDA National Nutrient Database for Standard Reference, Release 24. Nutrient Data Laboratory Home Page. 2011 http://www.ars.usda.gov/ba/bhnrc/ndl.

Supplementary Table 2. Pearson correlation coefficients, ρ_P, between cell volumes and 28 metabolite ion abundances.

	Volume	**pyruvate**	**oxalate**	**fumarate**	**malate**	**glutamate**	**pentose**	**arabinonate**	**gallic acid**	**aconitic acid**	**ascorbic acid**	**alliin**	**hexose**	**citrate**	**gluconic acid**	**glucarate**	**flavanone**	**hexose phosphate**	**glutathione**	**2OH-2CH3O-flavone**	**homoglutathione**	**disaccharide**	**3OH-5CH3O-flavone**	**trisaccharide**	**UDP-pentose**	**UDP-hexose**	**alliospiroside C**	**pentasaccharide**	**hexasaccharide**
Volume	1.0	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	0.0	0.0	-0.3	0.0	-0.3	0.1	0.0	0.0	-0.2	0.0
pyruvate		1.0	0.6	0.2	0.2	0.3	0.5	0.5	0.4	0.2	0.3	0.3	0.5	0.4	0.3	0.4	0.3	0.3	0.2	0.1	0.3	0.0	0.3	-0.1	0.4	0.2	0.2	-0.2	0.3
oxalate			1.0	0.2	0.1	0.4	0.7	0.5	0.1	0.4	0.3	0.4	0.6	0.2	0.5	0.4	0.2	0.3	0.1	0.3	0.3	0.0	0.3	-0.1	0.3	0.1	0.3	-0.1	0.3
fumarate				1.0	0.8	0.1	0.2	0.2	0.0	0.3	0.3	0.2	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	-0.2	0.1	0.0	0.0	-0.2	-0.2
malate					1.0	0.0	-0.1	0.1	0.1	0.2	0.2	0.2	-0.1	0.1	-0.1	0.0	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	0.0	0.0	-0.1	-0.1	-0.2	-0.3
glutamate						1.0	0.3	0.2	0.0	0.1	0.3	0.2	0.2	0.1	0.2	0.0	0.1	0.5	0.5	0.0	0.1	0.2	0.2	-0.1	0.1	0.4	0.1	0.0	0.1
pentose							1.0	0.7	0.2	0.3	0.3	0.2	0.7	0.3	0.4	0.6	0.3	0.3	0.1	0.3	0.3	0.0	0.5	0.0	0.4	0.3	0.3	-0.1	0.5
arabinonate								1.0	0.2	0.3	0.5	0.3	0.5	0.3	0.4	0.6	0.4	0.3	0.2	0.2	0.4	0.0	0.5	0.0	0.5	0.3	0.4	-0.1	0.5
gallic acid									1.0	0.2	0.1	-0.1	0.1	0.7	0.2	0.3	0.3	0.1	0.1	-0.1	0.2	0.0	0.1	0.0	0.2	0.1	0.2	-0.1	0.2
aconitic acid										1.0	0.5	0.3	0.1	0.2	0.2	0.3	0.2	0.1	0.1	0.2	0.2	-0.1	0.2	-0.1	0.4	0.0	0.3	-0.1	0.2
ascorbic acid											1.0	0.5	0.1	0.3	0.1	0.3	0.2	0.3	0.3	0.3	0.3	0.0	0.5	-0.1	0.4	0.3	0.3	-0.1	0.1
alliin												1.0	0.1	0.0	0.0	0.2	0.0	0.2	0.1	0.4	0.2	-0.1	0.4	0.0	0.2	0.1	0.0	0.0	-0.1
hexose													1.0	0.2	0.4	0.5	0.1	0.3	0.1	0.1	0.3	0.2	0.3	0.0	0.2	0.2	0.3	-0.2	0.4
citrate														1.0	0.2	0.4	0.4	0.2	0.2	0.1	0.3	0.0	0.3	0.1	0.3	0.3	0.3	0.0	0.3
gluconic acid															1.0	0.3	0.4	0.2	0.1	0.2	0.2	0.0	0.2	-0.1	0.4	0.1	0.5	0.0	0.4
glucarate																1.0	0.3	0.4	0.3	0.2	0.3	0.0	0.5	0.1	0.4	0.3	0.4	-0.1	0.4
flavanone																	1.0	0.2	0.1	0.0	0.1	0.1	0.2	0.0	0.4	0.2	0.3	0.0	0.3
hexose phosphate																		1.0	0.8	0.1	0.2	0.2	0.3	-0.1	0.2	0.7	0.2	0.0	0.2
glutathione																			1.0	0.1	0.1	0.3	0.1	-0.1	0.1	0.8	0.1	0.0	0.1
2OH-2CH3O-flavone																				1.0	0.1	0.0	0.4	0.0	0.3	0.1	0.2	0.1	0.1
homoglutathione																					1.0	0.0	0.3	0.0	0.4	0.2	0.3	0.0	0.3
disaccharide																						1.0	0.0	0.2	0.0	0.3	0.0	0.3	0.1
3OH-5CH3O-flavone																							1.0	0.0	0.4	0.2	0.2	-0.1	0.3
trisaccharide																								1.0	-0.2	-0.1	-0.1	0.5	0.1
UDP-pentose																									1.0	0.2	0.5	-0.1	0.3
UDP-hexose																										1.0	0.1	0.0	0.2
alliospiroside C																											1.0	-0.1	0.4
pentasaccharide																												1.0	0.1
hexasaccharide																													1.0

2OH-2CH3O-flavone = dihydroxy dimethoxyflavone
3OH-5CH3O-flavone = trihydroxy pentamethoxyflavone