Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ionScell enables microscopy-independent single-cell spectral extraction from MALDI mass spectrometry imaging.

Michel Salzet

michel.salzet@univ-lille.fr

Univerity Lille https://orcid.org/0000-0003-4318-0817

Laurine Lagache

Univerity Lille

Yanis Zirem

Université de Lille Protéomique Réponse Inflammatoire Spectrométrie de Masse - PRISM
https://orcid.org/0009-0003-1237-4448

Lea Ledoux

Univerity Lille

Julie Defrance

Univerity of Lille https://orcid.org/0009-0003-9866-973X

isabelle Fournier

Université de Lille https://orcid.org/0000-0003-1096-5044

Article

Keywords:

Posted Date: July 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9874513/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Reliable single-cell spectrum extraction from MALDI mass spectrometry imaging (MSI) is constrained by pixel-level signal mixing and by the lack of automated segmentation tools that operate without microscopy. We present ionScell, an open-source pipeline that derives cell boundaries directly from the Total Ion Current (TIC) image using adaptive thresholding, watershed segmentation and six-metric spectral quality control, without co-registered microscopy, fluorescent labels or manual ROI annotation. Across four cancer cell lines and two ionization modes (eight independent acquisition settings), ionScell generated thousands of QC-retained single-cell spectra per acquisition, discriminated binary NCH82/AU565 (AUC = 0.987; 96% accuracy) and ternary breast-cancer mixtures (macro-AUC = 0.926; 86% accuracy), and resolved NCH82 intra-line molecular heterogeneity supported by orthogonal single-cell proteomics. Cross-platform operation was demonstrated on a public MALDI-Orbitrap SpaceM HeLa/NIH3T3 co-culture dataset. On the same NCH82 dataset, ionScell recovered 2.3–3.1× more individually contoured cells than MSI Parser and provided outputs not available from the comparator, including per-cell ion maps, six-metric QC and probabilistic molecular-subpopulation assignments. ionScell is available as a Python package, Jupyter notebook and standalone GUI, with user documentation and example data.

Introduction

Spatially resolved omics technologies have fundamentally changed how we study tissue architecture and cellular heterogeneity. Among these, matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) has established itself as a uniquely versatile label-free approach: it detects lipids, peptides, and metabolites directly from biological surfaces while preserving their spatial context [1–3]. These properties make MALDI-MSI particularly well-suited for cancer research, where spatially structured molecular heterogeneity drives disease progression and therapy resistance [4, 5].

Despite continued improvements in laser optics and detector sensitivity, applying MALDI-MSI at true single-cell resolution remains technically and computationally demanding [6, 7]. Standard clustering-based approaches such as *k*-means, DBSCAN [8], PCA and hierarchical clustering, can separate cellular from background regions but fall short of reliably delineating individual cells. Adjacent or overlapping cells produce mixed spectra and poorly defined boundaries, ultimately limiting molecular interpretation at the single-cell level [9].

Several strategies have been proposed to address this challenge (**Supplementary Table S1**). Manual region selection yields accurate spectra but is impractical beyond a handful of cells per sample [10]. ROI-based segmentation in SCiLS Lab using curated lipid panels works well for homogeneous populations but cannot resolve individual cells [11]. Fluorescence microscopy co-registration improves segmentation accuracy at the cost of an extra experimental step and alignment errors [12]. Pre- and post-MALDI ablation mark alignment offers an alternative but introduces its own procedural complexity [13]. Ion-

based depth-first search (DFS) segmentation has also been proposed [14], but relying on a single selected ion limits robustness in molecularly heterogeneous samples.

Here we present ionScell, an open-source computational pipeline that bypasses these constraints (Fig. 1). Rather than requiring co-registered microscopy images or relying on a single informative ion, ionScell derives segmentation information directly from the MSI data, using the TIC image as a spatial proxy for cellular material. Adaptive TIC thresholding separates cell-rich regions from background and aggregates; watershed segmentation delineates individual cell boundaries; and size-based morphological filtering removes artefacts. A six-metric spectral quality-control module then excludes poor-quality spectra before downstream analysis.

Beyond segmentation, *ionScell* identifies molecular subpopulations (MSps) via UMAP dimensionality reduction followed by soft probabilistic clustering (GMM or Fuzzy C-Means [16]) with per-cell confidence scoring. The number of clusters is determined automatically by silhouette scoring [17], eliminating user-dependent bias. Downstream statistical analyses including differential ion testing, machine learning model training, and pathway enrichment are performed using standard bioinformatics tools [18].

We validated ionScell on breast cancer (MCF-7, AU565 and MDA-MB-231) and glioblastoma (NCH82) cell lines. Single-cell isolation was confirmed by microscopy overlay, cell-type discrimination in mixed preparations reached 96% accuracy, cross-platform operation was tested on a public SpaceM dataset, and within-line heterogeneity in NCH82 was independently supported by single-cell proteomics. This validation strategy was designed to address practical use by a broad life-science audience: automated cell recovery, method benchmarking, general applicability across biological systems and ionization modes, and biologically interpretable downstream outputs.

Results

Single-Cell Segmentation, Validation and Quality Control

To validate *ionScell's* ability to isolate individual cells purely from MSI-derived information, a single MCF-7 breast cancer cell was deposited onto an ITO-coated glass slide using the cellenONE® precision system, and bright-field images were acquired before and after MALDI acquisition (Fig. 2A). Following TIC normalization [19], the TIC map displayed a sharp intensity maximum at the cellular region, with signal decaying smoothly toward the periphery, consistent with higher molecular density within the intracellular compartment. The averaged mass spectrum showed the characteristic phospholipid fingerprint for a breast cancer cell, with prominent peaks in the *m/z* 650–950 region (Fig. 2B).

Automatic thresholding followed by watershed segmentation [20] and morphological filtering produced a contour closely matching the cell boundary visible in the TIC image (Fig. 2C). Ion distribution maps at *m/z* 782.50 (PC 36:4, [M + H]+) and *m/z* 609.10 (SM 34:1, [M + H]+) were spatially confined to the segmented region, independently confirming segmentation accuracy (Fig. 2D). Scaling to a multicellular

cytospin of MCF-7 cells, the TIC map revealed multiple discrete intensity maxima, each corresponding to a deposited cell (Fig. 2E). Automatic segmentation resolved adjacent and partially overlapping cells without manual intervention, confirmed by comparing contours with toluidine blue-stained optical images after matrix removal (Fig. 2F).

Segmentation throughput was consistent across all four cell lines and both ionization modes. Starting from cytospin preparations of approximately 5,000 deposited cells, watershed segmentation recovered a mean of 3,313 ± 1,475 cell objects per acquisition (yield approximately 66%); after six-metric QC, 2,896 ± 1,364 cells were retained per acquisition (87.4% retention). Rejections were dominated by low total intensity (approximately 71%) and low sparsity (approximately 29%). Full statistics per line and mode are provided in Supplementary Table S2. MCF-7 negative-mode retention was lower (68.0% versus 89.4–90.0% for other lines), attributable to the denser growth pattern of this luminal epithelial line, which produces a higher proportion of partially overlapping cells that fail morphological and spectral QC criteria. This effect was absent in positive mode (90.0%), suggesting a mode-dependent interaction between cell packing density and ion suppression in negative-mode lipid profiling that warrants further investigation.

Reliable downstream analysis requires exclusion of low-quality spectra. ionScell computes six orthogonal per-cell metrics (Fig. 2G) [21]: SNR, sparsity, dynamic range, total intensity, spectral entropy and peak count. A cell is rejected if it fails any single criterion. This multi-metric strategy captures artefact classes that are invisible to single-threshold filters, including matrix crystals (high SNR, low entropy), debris (low total intensity) and off-tissue pixels (near-zero sparsity), without discarding genuine biological variation (Fig. 2H–I).

Single-Cell Profiling Across Cancer Cell Lines in Dual Ionization Modes

To assess generalizability, ionScell was applied to cytospin preparations of NCH82 glioblastoma cells analyzed in both negative and positive ionization modes (Fig. 3); results for MCF-7, AU565 and MDA-MB-231 are provided as Supplementary Figures S1–S12. In both polarities, TIC maps of dense cell carpets showed multiple discrete intensity maxima cleanly resolved by watershed segmentation, including in regions of partial cell overlap (Fig. 3A–F). In positive mode, dominant spectral features in the m/z 700–900 range are consistent with glycerophosphocholines. In negative mode, peaks around m/z 480–520 and extending to higher m/z are consistent with phosphatidylinositols, sulfatides and anionic lipid species enriched in glioblastoma [22,23]. Dual-polarity acquisition thus provides complementary lipid coverage that would be inaccessible in a single-mode experiment [24].

QC filtering retained 3,509 (90.0%) and 3,310 (90.0%) cells in negative and positive mode, respectively, out of 3,889 and 3,678 segmented objects (Supplementary Figures S22–S23; Fig. 3D,I).

Benchmark: Comparison with SCiLS Lab and MSI Parser

To contextualize ionScell within the existing landscape, we compared its segmentation output on the NCH82 glioblastoma cytospin against two widely used approaches: SCiLS Lab ROI-based segmentation and MSI Parser v3 ion-based depth-first search (DFS) segmentation, evaluated in both negative and positive ionization modes (**Figure S25**).

SCiLS Lab comparison. In both ionization modes, bisecting *k*-means resolved the NCH82 cytospin into tissue and background clusters but did not delineate individual cells: the entire cellular layer formed a single ROI, requiring manual one-by-one annotation to isolate individual spectra, impractical at thousands of cells per acquisition. In contrast, ionScell resolved 3,509 cells (negative) and 3,310 cells (positive) after QC (**Figure S25A–B**).

MSI Parser comparison. DFS segmentation using a single seed ion (negative: m/z 673.4; positive: m/z 782.4) recovered 1,413 and 1,074 isolated cell positions, respectively, presented as scatter maps without per-cell contours or ion spatial maps. ionScell recovered 3,889/3,678 cells before QC and 3,509/3,310 cells after QC in negative and positive mode, respectively, corresponding to 2.3× more QC-retained cells in negative mode and 3.1× more in positive mode than MSI Parser (Figure S25C–G).

Cell recovery and QC statistics. Across all four cell lines and both ionization modes, ionScell segmented a mean of 3,313 cells per acquisition, with 87.4% retained after QC. Rejection rates were consistent across modes, with an overall mean rejection rate of 12.8% across all acquisitions (Supplementary Table S2). QC profiles and spectral characteristics were highly consistent across eight acquisition settings (four cell lines × two ionization modes), each processed from cytospin preparation through classification under identical analytical workflows. These acquisition settings are not replicate measurements of a single biological system; instead, they deliberately span four biologically distinct cancer cell lines (luminal breast, HER2-enriched breast, triple-negative breast and glioblastoma), two tissue origins and three molecular subtypes, each analyzed in dual ionization modes. This multi-system, cross-mode framework provides a structured validation of robustness across biologically and molecularly diverse contexts.

Cross-Platform Validation on a Public SpaceM Dataset

To assess cross-platform applicability, ionScell was applied to the publicly available SpaceM co-culture HeLa/NIH3T3 imzML dataset (METASPACE Rappez_2021_SpaceM; MALDI-Orbitrap, negative mode, 50 µm spatial resolution) [13]. This dataset differs fundamentally from full-field MSI acquisitions, as SpaceM relies on microscopy-guided, ROI-based sampling with optical co-registration, and the publicly available data represent selected regions of interest rather than continuous cellular fields. Given the coarser spatial resolution (50 µm per pixel) relative to our 10 µm acquisitions, segmentation parameters were adapted accordingly (min_cell_size 50 µm; max_cell_size 150 µm; TIC normalization). After segmentation and six-metric QC, ionScell retained 146 high-quality single-cell spectra from 177

segmented objects (82.5% retention; 31 rejected). Unsupervised UMAP followed by soft GMM clustering identified two molecular subpopulations (C0 and C1; mean confidence = 0.915), with no cells falling below the ambiguity threshold (0.5) (Supplementary Figure S29). The two clusters exhibited distinct spatial organization and spectral signatures consistent with the expected co-culture composition (HeLa and NIH3T3), demonstrating that ionScell can resolve mixed-species populations in an independent, instrument- and workflow-distinct MSI dataset. Importantly, this evaluation is not intended as a quantitative benchmark against SpaceM performance. Direct numerical comparison is confounded by fundamental differences in experimental design, including microscopy-guided ROI selection, spatial sampling density and the absence of full-field single-cell ground truth in the publicly available dataset. Instead, this analysis provides evidence of cross-platform compatibility, showing that ionScell can operate without re-engineering on data acquired using a distinct MALDI workflow and co-registration-based paradigm.

Discrimination of Distinct Cell Populations in Mixed Preparations

A central application of single-cell MSI is the identification of distinct cell types within heterogeneous samples. We benchmarked *ionScell* on two co-deposition experiments that increase molecular complexity.

Binary mixture: NCH82 and AU565

NCH82 (glioblastoma) and AU565 (HER2-enriched breast cancer) cells were co-deposited by cytospin and analyzed in positive ionization mode. These two lines were chosen as an initial benchmark because their distinct tissue origins translate into markedly divergent lipid compositions. Watershed segmentation successfully delineated individual cells across the mixed preparation (Fig. 4A). Unsupervised UMAP [25] with soft GMM clustering revealed two high-confidence clusters, C0 and C1, with clear separation in embedding space (Fig. 4B). Spatial back-projection showed stochastic intermingling consistent with random co-deposition (Fig. 4C).

Differential ion analysis identified 14 discriminant ions (adjusted $p < 0.05$; $|\log_2 FC| > 2$) spanning m/z 641.4–881.5; LPC/higher-mass PC species were enriched in C0 (AU565), while PC 32:1, 35:2 and 38:6 dominated C1 (NCH82) (Fig. 4D–E). Full list in Supplementary Figure S16. Random Forest achieved the highest classification performance (AUC = 0.987; 96% global accuracy; Fig. 4F–G), with m/z 748.6, 796.6 and 678.5 as top SHAP discriminants (**Supplementary Figure S15**). Cluster identities were confirmed by bulk reference spectra, assigning C0 to AU565 and C1 to NCH82 (Fig. 4H–J). Negative-mode results are in Supplementary Figure S17.

Ternary mixture: MDA-MB-231, MCF-7 and AU565

A three-way co-deposition was designed as a more demanding benchmark, as MDA-MB-231, MCF-7, and AU565 all originate from breast epithelium and share broad lipid compositional similarities. Watershed segmentation resolved individual cell boundaries across the mixed preparation (Fig. 5A). UMAP clustering identified three distinct subpopulations (C0, C1, C2) with clear inter-cluster separation (Fig. 5B**)** and spatial back-projection confirmed their stochastic distribution across the tissue section (Fig. 5C). Cohen's kappa spectral similarity scores were 0.76 (C0–C1), 0.79 (C0–C2) and 0.87 (C1–C2) (**Supplementary Figure S19**), reflecting the expected molecular proximity of three breast-derived lines, with C1 and C2 sharing the highest spectral overlap.

Differential ion analysis and classifier training across the three clusters identified cluster-specific signatures (pairwise volcano plots, Fig. 5D–E). The Passive Aggressive classifier achieved the best macro-averaged performance (AUC = 0.926; 86% global accuracy) (Fig. 5F–G), with asymmetric accuracy (MCF-7: 93.6%; AU565: 86.6%; MDA-MB-231: 69.7%) reflecting the molecular proximity of breast subtypes (Cohen's κ 0.76–0.87; Supplementary Figure S19). SHAP analysis ranked m/z 782.5 (PC 36:4), 760.5 (PC 34:1) and 732.5 (PC 30:0–32:1) as top discriminants; pairwise markers confirmed identities (**Supplementary Figures S20B–C**). Cluster identities were assigned as C0 = MDA-MB-231, C1 = MCF-7, C2 = AU565 (Fig. 5H–K). Negative-mode results are in Supplementary Figure S21.

Intra-Line Molecular Heterogeneity and Single-Cell Validation

Beyond discriminating known cell types, we asked whether ionScell could uncover biologically meaningful heterogeneity within a single cell line, a harder and more translationally relevant problem. Applied to the NCH82 negative-mode dataset, the automated QC log records a first-pass four-component solution (Supplementary Figure S22); the final biological analysis retained the three stable high-confidence molecular subpopulations shown in Fig. 6 (MSp 0–2; first-pass silhouette = 0.534; mean assignment confidence = 0.965). The three MSps showed distinct mean spectra (Fig. 6C) with non-random spatial distributions (Fig. 6B); ion maps at m/z 673.40 (PI 34:1) and 887.50 (PI 38:4) confirmed MSp-correlated enrichment (Fig. 6D–E). Positive-mode analysis of the same NCH82 cells independently identified two MSps (Supplementary Figure S24), establishing that the binary partition is robustly captured by cationic glycerophosphocholines and ether-PCs before orthogonal validation.

To determine whether these lipidomic subpopulations reflect genuine biological differences, single-cell proteomics was performed on NCH82 cells (n = 233 cells; dia-PASEF LC-MS/MS). Proteomics identified two functional classes (A and B) in direct concordance with the two MSps resolved by positive-mode MSI (Supplementary Figure S24). t-SNE projection separated the two classes with clear inter-cluster distance (Fig. 6F). Pathway enrichment analysis revealed divergent signatures: Class B was enriched in MYC targets, G2/M checkpoint, oxidative phosphorylation, protein secretion and unfolded protein response pathways, characteristic of highly proliferative and metabolically stressed cells, whereas Class A showed enrichment in DNA repair and basal inflammatory pathways, consistent with a more quiescent or differentiated state (Fig. 6G). Pharmacological profiling highlighted Class A associations with bortezomib, carfilzomib, cyclophosphamide and doxorubicin (Fig. 6H) [28]. Both classes expressed

immunoglobulin family members (IGHG1, IGHG2, IGHA1, IGKC and IGLC), but Class A displayed significantly higher and more stable IGHG1 levels, whereas Class B exhibited lower median expression with greater cell-to-cell variability (Fig. 6I–J), consistent with active proteomic remodeling in the proliferative subpopulation. The convergence across three independent readouts establishes a coherent molecular hierarchy: positive-mode lipidomics and proteomics both resolve a binary functional state (quiescent Class A versus proliferative Class B), while negative-mode MSI adds a third layer of resolution, a PI- and sulfatide-enriched substate (MSp 2) that is invisible to both cationic lipid profiling and the proteome. The molecular basis of this third state remains to be functionally validated; we hypothesize that enrichment in phosphatidylinositols and sulfatides, known reporters of PI3K/AKT activity and membrane remodeling [22,23], may reflect a transitional metabolic state between the quiescent and proliferative poles. This convergence across independent platforms validates the biological relevance of ionScell-derived subpopulations and illustrates the potential of anionic lipidomics to resolve cellular substates inaccessible to proteomics or cationic lipid profiling.

Discussion

ionScell addresses a persistent bottleneck in single-cell MSI: the extraction of reliable, high-quality spectra from individual cells without co-registered microscopy or manual region selection. By grounding segmentation entirely in TIC-derived spatial information, the pipeline removes experimental dependencies that have limited throughput and reproducibility in previous approaches [10–14]. The six-metric QC framework is a deliberate departure from simpler filtering strategies. Single-metric filters fail systematically on common MSI artefact types: matrix crystals with high SNR but trivially low spectral complexity, cellular debris with plausible entropy but negligible total signal, and off-tissue pixels with near-zero sparsity regardless of SNR. Requiring all six criteria simultaneously captures these failure modes without discarding genuine biological variation, producing cleaner clustering and fewer false-positive MSp assignments downstream.

Cytospin preparations targeting approximately 5,000 cells yielded a mean of 3,313 segmented cells, with 2,896 (87.4%) retained after QC (Supplementary Table S2), sufficient depth for GMM clustering, where even a 10% subpopulation exceeds the practical minimum of approximately 260 cells.

A recurring question in the field concerns direct performance comparison with existing single-cell MSI methods. SpaceM [13] and fluorescence co-registration approaches [12] represent elegant solutions that deliver high-quality single-cell data, but operate in a categorically different experimental regime requiring an additional optical imaging step, precise spatial co-registration and access to fluorescent labeling or fiducial marks. ionScell operates without any of these prerequisites, making it immediately applicable on standard MALDI-MSI data. Rather than attempting head-to-head numerical benchmarking that would conflate workflow complexity with algorithmic performance, we validated cross-platform applicability on the public SpaceM HeLa/NIH3T3 imzML dataset (METASPACE Rappez_2021_SpaceM), confirming that ionScell recovers coherent cell-type subpopulations on a MALDI-Orbitrap platform with minimal parameter adaptation (see Results). Comparison against SCiLS Lab and MSI Parser on the same NCH82

dataset provides additional context: SCiLS bisecting k-means cannot delineate individual cells at scale, while MSI Parser DFS recovered 2.3–3.1× fewer cells than ionScell and produces no per-cell contours or ion spatial maps (Figure S25). The appropriate frame of comparison is therefore fitness for experimental context: microscopy-guided approaches for samples requiring optical ground truth; ionScell for settings where throughput, simplicity and instrument independence are primary. Notably, the 91.3% cell-type classification accuracy reported by SpaceM [13] was obtained across six acquisitions (two co-culture replicates, two HeLa mono-cultures and two NIH3T3 mono-cultures; 846 total cells) with fluorescence reporters (mCherry/GFP) as ground truth and 88 METASPACE-annotated metabolites per cell. The ionScell cross-validation used a single negative-mode field of view from the public METASPACE dataset (14052018_coculture_HeLa_NIH3T3_N1_Neg; 50 µm pixel size; 50×50-pixel region of interest; 146 cells retained after QC), without fluorescence input and using only 52 lipid annotations at FDR 10%. Despite this far more constrained setting, ionScell resolved both subpopulations with a mean soft-clustering confidence of 0.915 and zero ambiguous cells (Supplementary Figure S29). Direct numerical comparison is not appropriate given the asymmetry in acquisition scale and annotation depth, but the result confirms that TIC-based, microscopy-free segmentation produces biologically coherent single-cell partitions on an independent MALDI-Orbitrap platform, without instrument-specific parameter re-optimization.

The soft clustering approach deserves emphasis. Replacing hard *k*-means assignment [29] with GMM or Fuzzy C-Means introduces a probabilistic dimension that reflects biological reality: cells rarely fall into perfectly discrete categories [30]. Per-cell confidence scores distinguish robust high-confidence assignments from ambiguous border cases. Automatic cluster number selection by silhouette scoring further removes a major source of user-dependent bias that affects many published MSI workflows. The interoperability of *ionScell* with downstream analysis platforms [18] extends *ionScell*'s reach into downstream statistical modeling. All differential ion analyses, classifier training, and pathway enrichment reported here were performed using open-source tools on *ionScell*-generated CSV files as input, enabling complex multi-step workflows without additional scripting.

The apparent discrepancy between the final three lipidomic MSps reported for negative-mode MSI and the two proteomic classes is resolved by a key observation: positive-mode MSI of the same NCH82 cells also yields exactly two MSps (Supplementary Figure S24), in direct concordance with the proteomic binary partition. This convergence between two independent platforms (proteomics and positive-mode lipidomics) strongly supports the biological reality of the two-state model. The third subpopulation detected exclusively in negative mode is therefore interpreted as a lipid-level substate that is molecularly invisible to cationic lipid profiling and to the proteome. Its functional identity remains to be established by dedicated experiments. We note, however, that phosphatidylinositols and sulfatides, the ions most enriched in MSp 2, are established reporters of PI3K/AKT signaling and membrane remodeling [22,23], raising the hypothesis that this substate may correspond to a transitional metabolic intermediate between the quiescent and proliferative poles captured by the other two readouts. The hierarchy is therefore: two functional states (proteome plus positive lipidome) and three lipid-metabolic states (negative lipidome), with the additional resolution provided by the anionic lipidome representing a

testable prediction for future functional validation. Independently, the cross-instrument validation on the SpaceM HeLa/NIH3T3 dataset confirms that ionScell's clustering output is not instrument-specific: two subpopulations concordant with the known cell-type composition were identified on a MALDI-Orbitrap platform, reinforcing the generalizability of both the segmentation and clustering modules.

Several limitations merit discussion. At 10 µm pixel size, cells smaller than the laser footprint produce mixed spectra; this instrument-level constraint is mitigated by size-aware morphological filtering and confidence-based cell exclusion (threshold 0.6), and will diminish as sub-5 µm instruments become routine.

All validation used cytospin monolayers; extension to tissue sections, organoids and patient-derived samples, where ECM and cell-density variation increase complexity, is a near-term priority. The watershed and QC logic are not inherently restricted to monolayers, but systematic characterization on tissue remains to be done.

Default parameters (3×3 footprint, SNR > 2.0, sparsity > 0.05, dynamic range > 3) were optimized on a MALDI timsTOF flex at 10 µm with Norharmane matrix and should be treated as evidence-based starting points. The adaptive threshold selector (method='auto') minimizes re-tuning needs; the GUI provides step-by-step visual feedback for iterative optimization.

Planned developments include LIPID MAPS/HMDB annotation, tissue-section validation, and ML-based boundary delineation. In conclusion, ionScell provides a fully automated, microscopy-free framework for single-cell spectral extraction and molecular subpopulation identification from MALDI-MSI data, with three user interfaces and full interoperability with downstream analysis platforms [18], positioning it as a practical tool for translational single-cell spatial omics.

Online Methods

Experimental Design and Statistical Rationale

Four human cancer cell lines were used: luminal breast cancer (MCF-7), HER2-enriched breast cancer (AU565), triple-negative breast cancer (MDA-MB-231) and glioblastoma (NCH82). Cells were deposited as cytospin monolayers on ITO-coated glass slides and analyzed by MALDI-MSI at 10 µm spatial resolution in both positive and negative ionization modes on a MALDI timsTOF flex (Bruker Daltonics). Raw MSI data were exported in imzML format [35] and processed with ionScell. To assess inter-type discrimination, binary (NCH82 + AU565) and ternary (MDA-MB-231 + MCF-7 + AU565) co-deposited preparations were analyzed. Unsupervised UMAP followed by soft probabilistic clustering identified molecular subpopulations (MSps); cells with confidence < 0.6 were flagged as ambiguous and excluded from MSp-level analyses. To validate MSI-derived heterogeneity through an orthogonal modality, single-cell proteomics was performed on NCH82 cells (LC–MS/MS, dia-PASEF, DIA-NN v2.1.0 [41]; precursor FDR 0.1%, protein FDR 1%). Proteomic data preprocessing applied 70% shifted-Gaussian imputation across the single-cell matrix. Differential protein expression was assessed by one-way ANOVA ($p < 0.01$)

with Benjamini–Hochberg false discovery rate correction. Pathway over-representation analysis (ORA) was performed against MSigDB Hallmark 2022 and Drug Target 2024 databases using the nEnrichment module of Profiler v1.2.4 [18]. All statistical analyses, differential ion testing, machine learning model training and pathway enrichment were performed in Profiler v1.2.4. As a cross-platform benchmark, the public co-culture dataset 14052018_coculture_HeLa_NIH3T3_N1_Neg (MALDI-Orbitrap, negative mode, 50 µm spatial resolution; METASPACE project Rappez_2021_SpaceM [13]) was downloaded in imzML/ibd format from https://metaspace2020.org. ionScell parameters were adapted for pixel size only: min_cell_size = 50 µm, max_cell_size = 150 µm.

Cell Culture and Sample Preparation

All cell lines were cultured in T25 flasks at 37°C / 5% CO_2 with media refreshed every 2–3 days and passaging at ~ 80% confluence. MDA-MB-231 and MCF-7: high-glucose DMEM + 10% FBS + 1% penicillin–streptomycin. AU565: RPMI-1640 + 10% FBS + 1% penicillin–streptomycin. NCH82: high-glucose DMEM + 10% FBS + 1% penicillin–streptomycin. After trypsinization, cells were resuspended in PBS at 10,000 cells/sample and deposited onto ITO-coated glass slides by cytocentrifugation. Selected slides were fixed with 1% PFA in PBS (10 min, RT). Norharmane matrix (7 mg/mL in MeOH:$CHCl_3$ 2:1 *v/v*) was applied by dry sublimation using an HTX M5 sprayer (flow rate 0.1 mL/min, 12 passes, 30°C). After matrix removal with methanol, cells were stained with toluidine blue for optical validation.

MALDI-MSI Data Acquisition

MSI data were acquired on a MALDI timsTOF flex (SmartBeam 3D laser, up to 10 kHz). Acquisition parameters: 10 µm spatial resolution; *m/z* 300–1,350; laser energy 70–80%; 200 shots per pixel; TIMS disabled (qTOF mode). Both positive and negative ionization modes were used. Data were exported in imzML format without normalization using SCiLS Lab v2026b (Bruker Daltonics).

ionScell Pipeline

The *ionScell* pipeline is implemented as the *SCMSIPipeline* Python class, available as: (i) a Python script (ionscell_*pipeline.py*); (ii) a Jupyter Notebook (*ionScell.ipynb*); and (iii) a standalone GUI (*ionScell GUI*). All three interfaces expose the same class and produce identical outputs. Screenshots of the GUI are provided in **Supplementary Figure S28**.

Data cube construction. MSI data in imzML format are loaded and mapped into a 3D array (x, y, *m/z*). Two binning strategies are available. Fixed-grid binning (make_mz_axis(); default bin width: 0.1 Da; recommended range for lipid profiling: 600–1,100 Da) is preferred for exploratory clustering. Adaptive binning (build_adaptive_mz_axis()) scales bin width proportionally to the local *m/z* value to match the instrument's resolving power (user-specified as target_resolving_power or ppm_bin; when both are absent, the algorithm estimates ppm resolution from peak spacings in a random subsample of 50

spectra, clamped to 2–500 ppm), and is recommended for annotation workflows. Large datasets exceeding available RAM can be streamed via memory-mapped arrays (memmap_path parameter).

Normalization. TIC, RMS, or median normalization is applied to ensure spectral comparability across pixels.

TIC image generation and thresholding. A summed TIC image is computed from the 3D data cube and optionally Gaussian-smoothed (sigma parameter; default σ = 0 for segmentation, σ = 0.5 for display). Adaptive thresholding (auto_TIC_threshold()) evaluates candidate strategies in parallel — GMM-based (2- and 3-component), Otsu's method, triangle, Li, Yen, isodata and percentile-based (default: 75th percentile) — and selects the threshold that maximizes a composite score combining foreground-background separability (between-class variance / total variance, weighted 0.6) and proximity to the biologically expected cell occupancy fraction (15%; Gaussian penalty, weighted 0.4). Only thresholds yielding a cell fraction between 5% and 60% of the image area are considered valid. This automated selection removes the most common source of user-dependent variability in label-free MSI segmentation [36].

Watershed segmentation. Local TIC maxima are detected using a sliding-window peak detector (peak_local_max, neighbourhood footprint default: 3×3 pixels) applied to the binary foreground mask; each maximum seeds one cell. Boundaries are delineated by flooding the inverted TIC image from these seeds (watershed_or_dfs_from_mask(), scikit-image watershed implementation). An alternative depth-first-search (DFS) segmentation mode is available for datasets where pixel intensity gradients are weak; it groups pixels sharing similar TIC values (intensity_tolerance parameter; default: 0.15 normalized units) around each seed. Both modes resolve adjacent and partially overlapping cells [20].

Morphological filtering. Segmented regions outside the biologically plausible size range are discarded. The size filter accepts absolute pixel counts or physical units (size_unit='um'; min_cell_size default: 2 pixels; max_cell_size: no upper limit by default, though values of 50–200 pixels are typical for 10 µm acquisition and mammalian cell lines). Single-pixel isolated regions are removed when remove_isolated=True (default). Manual exclusion of specific artefact IDs is available via remove_cells(); entire MSp clusters can be merged or removed via combine_clones() and remove_clone().

Spectral QC. Six per-cell metrics are computed by compute_spectral_quality_metrics(): (i) Signal-to-Noise Ratio (SNR: mean intensity / 10th-percentile intensity; threshold > 2.0); (ii) sparsity (fraction of non-zero bins; threshold > 0.05); (iii) dynamic range ($\log_{10}$ max / 5th-percentile intensity; threshold > 3); (iv) total intensity (cells below the 10th percentile of the population are rejected); (v) spectral entropy (Shannon entropy of the normalised spectrum, $-\sum p \log_2 p$; captures spectral complexity); and (vi) peak count above local median (number of bins exceeding the per-cell spectral median; threshold > 20). A cell is rejected by filter_low_quality_cells() if it fails any one of these criteria (logical AND conjunction). This multi-criterion strategy is deliberately conservative: each metric targets a distinct artefact class (matrix crystals: high SNR, low entropy; debris: low total intensity; off-tissue pixels: near-zero sparsity), and

requiring all six simultaneously removes failure modes that escape single-metric filters without discarding genuine biological variation.

Spectral extraction. Per-cell spectra are reconstructed by aggregating pixel intensities within each segmented region (mean, sum, or median; default: mean) using extract_cell_spectra().

Subpopulation identification. UMAP [25] is applied to the single-cell spectral matrix, followed by soft probabilistic clustering (GMM or Fuzzy C-Means [16]). Optimal cluster number is selected by silhouette scoring [17]. UMAP parameters are n_neighbors = 15 (controls local versus global structure preservation), min_dist = 0.1 (controls compactness of the embedding), metric = euclidean, n_components = 2 and random_state = 0 (reproducibility). Spectra are StandardScaler-normalized prior to UMAP. For GMM clustering (covariance_type = 'tied'; regularization covariance reg_covar = 1×10 − 6; n_init = 10; max_iter = 200), the optimal number of clusters k is selected by maximizing the silhouette score if the best score exceeds 0.25; otherwise BIC minimization is used as fallback (search range: k = 2 to min(10, n_cells/3)). For Fuzzy C-Means, the fuzziness exponent m = 2.0 (standard value) is used. Cells with maximum cluster membership probability below 0.6 are flagged as ambiguous and excluded from MSp-level analyses; this threshold is configurable.

Export and visualization. Results are written to annotated CSV or imzML [35] files containing ion intensities, cell identifiers, geometric features and MSp labels. Interactive visualizations are generated with Plotly [37]. CSV files are directly compatible with downstream analysis platforms [18].

Differential Ion Analysis and Machine Learning with Profiler

CSV files exported from ionScell were analyzed in Profiler v1.2.4 [18] without additional preprocessing because the data were already TIC-normalized within ionScell. Four modules were used: (i) Comparison module, for pairwise Cohen's kappa similarity scores; (ii) Biomarkers module, for Kruskal–Wallis testing with Benjamini–Hochberg correction [38] (FDR < 0.05), fold-change computation and volcano plot generation; (iii) Visualization module (Signal Profile), for mean spectral profiles per cluster as skyline overlays; and (iv) AI Modeling module, for PCA and t-SNE dimensionality reduction, supervised classification with five-fold stratified cross-validated model selection (stratify = y; no data leakage across folds; linear, tree-based and ensemble models), confusion matrices, ROC curves and SHAP/LIME feature-importance analysis [27].

Single-Cell Proteomics

NCH82 cells were sorted using the CellenONE system into Evo96 proteoCHIPs containing lysis buffer (0.2% DDM, 100 mM TEAB, 20 ng/µL trypsin). Samples were digested at 50°C for 1.5 h and quenched with 4 µL 0.1% formic acid. Peptides were loaded onto Evotips and analyzed using an Evosep One

system [39] coupled to a timsTOF Flex mass spectrometer operating in dia-PASEF mode [40] (22 MS/MS scans; precursor range m/z 100–1,700; ion mobility $1/K_0$ = 0.65–1.42 V·s·cm^{-2}). Raw data were processed with DIA-NN v2.1.0 [41] against the UniProt Homo sapiens database (release 2024_12) using a precursor FDR of 0.1% and protein FDR of 1%. Subsequent preprocessing, including missing-value imputation using a 70% shifted Gaussian distribution, differential expression analysis (one-way ANOVA with Benjamini–Hochberg correction; $p < 0.01$), dimensionality reduction (t-SNE), and pathway enrichment analyses (ORA; MSigDB Hallmark 2022 and Drug Target 2024), were performed using Profiler v1.2.4 [18].

Declarations

Software and Code Availability

ionScell (Python ≥ 3.9) source code, Jupyter Notebook, GUI executable, user manual and example datasets are available at https://github.com/yanisZirem/ionScell (v1.1.0) under the MIT License. A permanently archived version of the code is deposited at Zenodo (https://doi.org/10.5281/zenodo.20443701). Complete parameter documentation, worked examples, example input/output files and a requirements.txt file specifying all dependencies (numpy ≥ 1.24, pandas ≥ 2.0, scipy ≥ 1.11, scikit-image ≥ 0.21, scikit-learn ≥ 1.3, umap-learn ≥ 0.5, pyimzml ≥ 1.5, plotly ≥ 5.18, matplotlib ≥ 3.7, tqdm ≥ 4.66) are provided in the repository README so that reviewers can test the workflow directly.

Data Availability

All raw and processed proteomics data are deposited in the PRIDE repository (accession PXD078019). MALDI-MSI datasets are publicly available at https://doi.org/10.5281/zenodo.20070096 (29.8 GB; centroid imzML/ibd [35] format; AU565, MDA-MB-231, MCF-7 and NCH82 cell lines plus binary and ternary mixed populations, acquired in both ionization modes on a Bruker MALDI timsTOF flex at 10 µm spatial resolution, *m/z* 350–1100).

Competing Interests

Y.Z., Le.L., I.F. and M.S. are co-developers of the Profiler analysis platform [18] used for downstream analysis in this study. The statistical and machine learning modules integrated into Profiler are derived from an open-source Python pipeline previously published and independently peer-reviewed [44,45], mitigating concerns of analytical circularity. All other authors declare no competing interests.

Author Contributions

Y.Z. conceived and conceptualized the project, developed the ionScell framework, and wrote and revised the manuscript. J.D. performed MALDI mass spectrometry imaging (MSI) experiments and contributed to ionScell code development. La.L. performed MSI experiments, cultured cells and contributed to manuscript writing. Le.L. performed proteomic validation experiments and contributed to manuscript writing. I.F. and M.S. supervised and funded the project and revised the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgements

This research was supported by the French National Research Agency (ANR France) through the Click&Detect project (ANR 23-CE29-0018-01). We would like to thank the OrganOmics Core Facility of the University of Lille, which is accredited and funded by the GIS IBiSA and is a member of the French Proteomics Infrastructure (ProFI) UAR 2048.

References

1. Alexandrov T (2012) MALDI imaging mass spectrometry: statistical data analysis and current computational challenges. BMC Bioinformatics 13(Suppl 16):S11
2. Norris JL, Caprioli RM (2013) Analysis of tissue specimens by MALDI imaging mass spectrometry in biological and clinical research. Chem Rev 113(4):2309–2342
3. Sans M et al (2021) Direct profiling and imaging of tumor lipids in colorectal cancer biopsies by MALDI mass spectrometry. Biochimie 187:93–101
4. Schwamborn K, Caprioli RM (2010) Molecular imaging by mass spectrometry — looking beyond classical histology. Nat Rev Cancer 10:639–646
5. Lagache L, Zirem Y, Le Rhun É, Fournier I, Salzet M (2025) Predicting protein pathways associated to tumor heterogeneity by correlating spatial lipidomics and proteomics: the dry proteomic concept. Mol Cell Proteom 24:100891. https://doi.org/10.1016/j.mcpro.2024.100891
6. Abdikenov B et al (2019) Analytics of heterogeneous single-cell populations: MALDI mass spectrometry imaging and probabilistic graphical models for single-cell profiling of lipid distributions. Sci Data 6:103
7. Pelletier DA et al (2020) Mass spectrometry imaging of lipids with high spatial resolution using a Bruker rapifleX MALDI Tissuetyper. Anal Chem 92(18):12233–12241
8. Ester M et al (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), 226–231
9. Alexandrov T et al (2011) Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering. J Proteome Res 10(7):3249–3259

10. Cuypers E et al (2024) MALDI MSI protocol for spatial bottom-up proteomics at single-cell resolution. J Proteome Res 23:5052–5063
11. Sekera ER et al (2024) Single Cell MALDI-MSI Analysis of Lipids and Proteins within a Replicative Senescence Fibroblast Model. J Am Soc Mass Spectrom 35(12):2815–2823
12. Bien T et al (2022) Fluorescence-guided single-cell MALDI-MSI using microscopy co-registration. Anal Chem 94:6167–6175
13. Rappez L et al (2021) SpaceM reveals metabolic states of single cells. Nat Methods 18:799–805
14. Zhang H, Liu Y, Fields L, Shi X, Huang P, Lu H, Schneider AJ, Tang X, Puglielli L, Welham NV, Li L (2023) Single-cell lipidomics enabled by dual-polarity ionization and ion mobility-mass spectrometry imaging. Nat Commun 14:5185
15. Rappez L et al (2020) Spatial metabolomics pipeline for co-registering MALDI and confocal imaging. J Am Soc Mass Spectrom 31:1031–1038
16. Bezdek JC (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York
17. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65
18. Zirem Y, Ledoux L, Fournier I, Salzet M (2025) Profiler: an open web platform for multi-omics analysis. Bioinformatics 42(1):btaf644. https://doi.org/10.1093/bioinformatics/btaf644
19. Deininger SO et al (2011) Normalization in MALDI-TOF imaging datasets of proteins: practical considerations. Anal Bioanal Chem 401:167–181
20. Beucher S, Lantuejoul C (1979) Use of watersheds in contour detection. Proc. International Workshop on Image Processing, CCETT/IRISA
21. Bemis KD et al (2016) Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments. Bioinformatics 32:2418–2420
22. Hosseini M et al (2020) Comparative lipidomic analysis of human glioblastoma and normal brain identifies potential biomarkers. Sci Rep 10:15367
23. Fonteh AN et al (2008) Alterations in cerebrospinal fluid glycerophosphocholine and sphingomyelin are associated with Alzheimer's disease and other neurological diseases. Lipids Health Disease 7:1
24. Ellis SR et al (2018) In situ characterization of the rat brain lipidome by MALDI mass spectrometry imaging. Metabolomics 14:77
25. McInnes L, Healy J, Melville J (2018) UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw 3(29):861
26. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605
27. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions (SHAP). Adv Neural Inf Process Syst, 30
28. Corsello SM et al (2020) Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nat Cancer 1:235–248
29. McLachlan GJ, Peel D (2000) Finite Mixture Models. Wiley Series in Probability and Statistics

30. Stuart T et al (2019) Comprehensive integration of single-cell data. Cell 177(7):1888–1902e21
31. Perou CM et al (2000) Molecular portraits of human breast tumours. Nature 406:747–752
32. Sorlie T et al (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. PNAS 98:10869–10874
33. Zitnik M et al (2019) Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. Inform Fusion 50:71–91
34. Fahy E et al (2009) Update of the LIPID MAPS comprehensive classification system for lipids. J Lipid Res 50(Suppl):S9–S14
35. Schramm T et al (2012) imzML — a common data format for the flexible exchange and processing of MS imaging data. J Proteom 75(16):5106–5110
36. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybernetics 9(1):62–66
37. Plotly Technologies Inc (2015) Collaborative data science. Montreal, QC. https://plot.ly
38. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B 57(1):289–300
39. Bache N et al (2018) A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. Mol Cell Proteom 17(12):2284–2296
40. Meier F et al (2020) diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. Nat Methods 17(12):1229–1236
41. Demichev V et al (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods 17(1):41–44
42. Liberzon A et al (2015) The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst 1:417–425
43. Alexandrov T, Bartels A (2013) Testing for presence of known and unknown molecules in imaging mass spectrometry. Bioinformatics 29:2335–2342
44. Zirem Y, Ledoux L, Roussel L, Maurage CA, Tirilly P, Le Rhun É, Salzet M, Fournier I (2024) Real-time glioblastoma tumor microenvironment assessment by SpiderMass for improved patient management. Cell Rep Med 5(4):101491. https://doi.org/10.1016/j.xcrm.2024.101491
45. Zirem Y, Ledoux L, Salzet M, Fournier I (2024) Protocol to analyze 1D and 2D mass spectrometry data from glioblastoma tissues for cancer diagnosis and immune cell identification. STAR Protocols 5(3):103161. https://doi.org/10.1016/j.xpro.2024.103161
46. Rappez L, Phapale P, Mamedov S, Schulze T, Alexandrov T (2024) HT SpaceM: a high-throughput and reproducible method for small-molecule single-cell metabolomics. bioRxiv. https://doi.org/10.1101/2024.10.24.620114. 2024.10.24.620114
47. Cairns J et al (2025) Mass-guided single-cell MALDI imaging of low-mass metabolites reveals cellular activation markers. Adv Sci 12:e2410506. https://doi.org/10.1002/advs.202410506
48. Bien T et al (2023) Alignment-free single-cell MALDI-MSI of lipids. Anal Chem 95:4498–4507

49. Amezquita RA et al (2020) Orchestrating single-cell analysis with Bioconductor. Nat Methods 17:137–145. https://doi.org/10.1038/s41592-019-0654-x
50. Stoeckius M et al (2017) Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 14:865–868
51. Practical Considerations and Parameter Optimization
52. Sample preparation For cytospin preparations, ~ 5,000 cells per slide is recommended. Select the slide with the most homogeneous cell distribution under a light microscope before matrix application. Gentle pipette mixing (P200–P1000) immediately before loading helps break up clumps. Dry sublimation is preferred over spray coating to minimize analyte delocalization at 10 µm pixel size and organic matrix crystals size
53. Thresholding For low-intensity TIC images, Otsu's method (*method='otsu'*) is recommended. For high-intensity images, start with the 85th percentile (*method='percentile', percentile = 85*) and adjust iteratively. Over-segmentation (many small regions) typically means the threshold is too low; under-segmentation (merged cells) means it is too high
54. *m/z* binning. For instruments with resolving power ≥ 25,000, adaptive *m/z* binning (build_adaptive_mz_axis()) is recommended. For lower-resolution instruments, fixed-grid binning is equally effective
55. Iterative refinement The *ionScell* GUI provides visual feedback after each step, enabling iterative parameter adjustment without programming. This interactive mode is recommended for first-time analysis of a new cell type or instrument. Complete parameter documentation and worked examples are available in the repository README at https://github.com/yanisZirem/ionScell
56. Adapting to non-standard pixel sizes When processing datasets acquired at coarser spatial resolution (e.g., 50 µm), set min_cell_size and max_cell_size in physical units (size_unit='um'). For 50 µm acquisition, starting values of min_cell_size = 50 and max_cell_size = 150 are appropriate for mammalian cell lines (1×–3× cell diameter at that resolution). The six-metric QC thresholds are scale-independent and require no adjustment. TIC normalization should always be applied prior to segmentation regardless of acquisition resolution

Figures

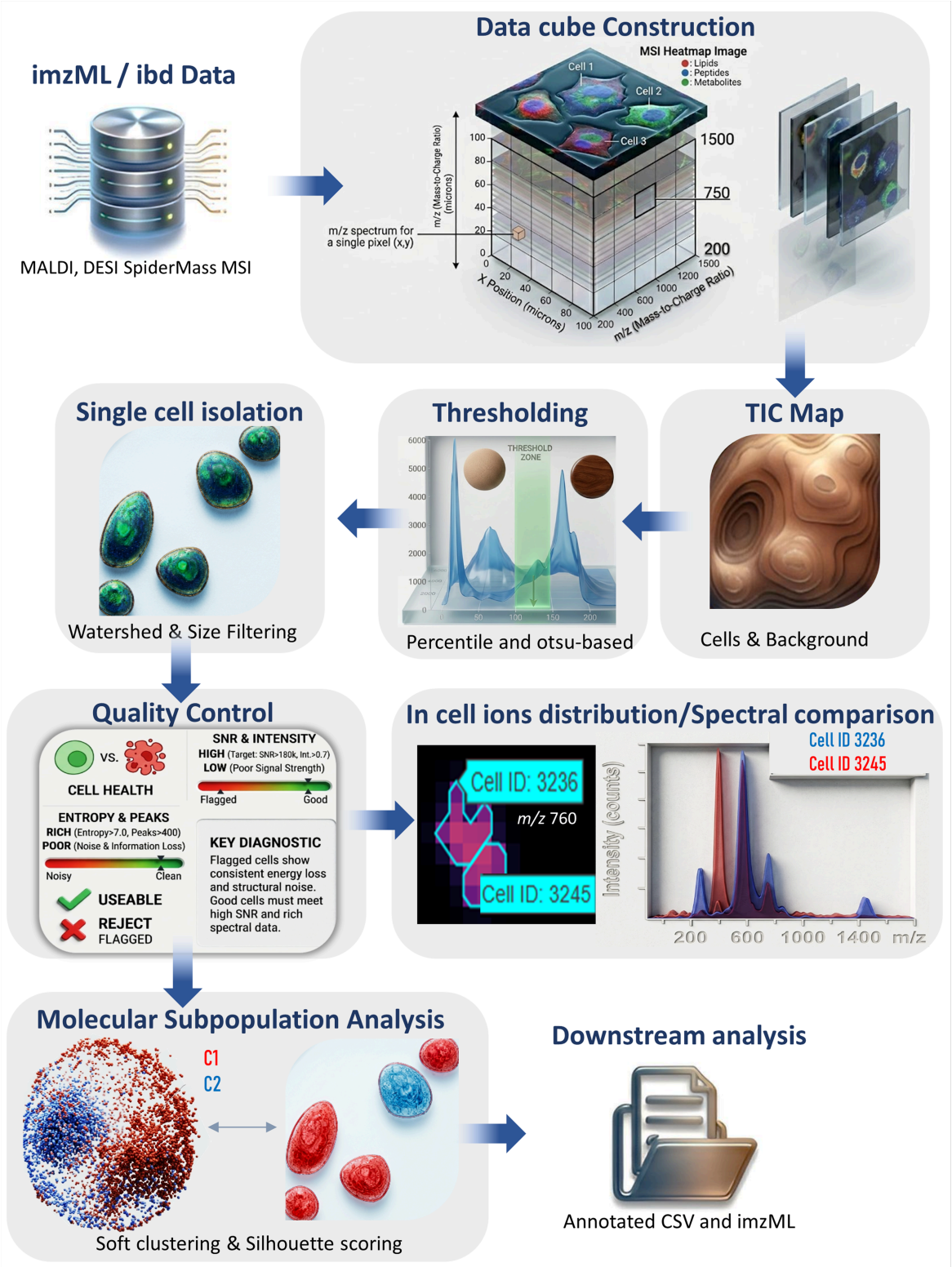


Figure 1

Overview of the *ionScell* pipeline. Starting from *imzML* input, the workflow proceeds through TIC normalization, adaptive thresholding (percentile or Otsu), watershed segmentation, morphological filtering and six-metric spectral quality control (QC). Per-cell spectra are subjected to UMAP dimensionality reduction and soft probabilistic clustering (GMM or Fuzzy C-Means) for molecular

subpopulation (MSp) identification. Results are exported as annotated CSV or *imzML*files and rendered interactively via Plotly.

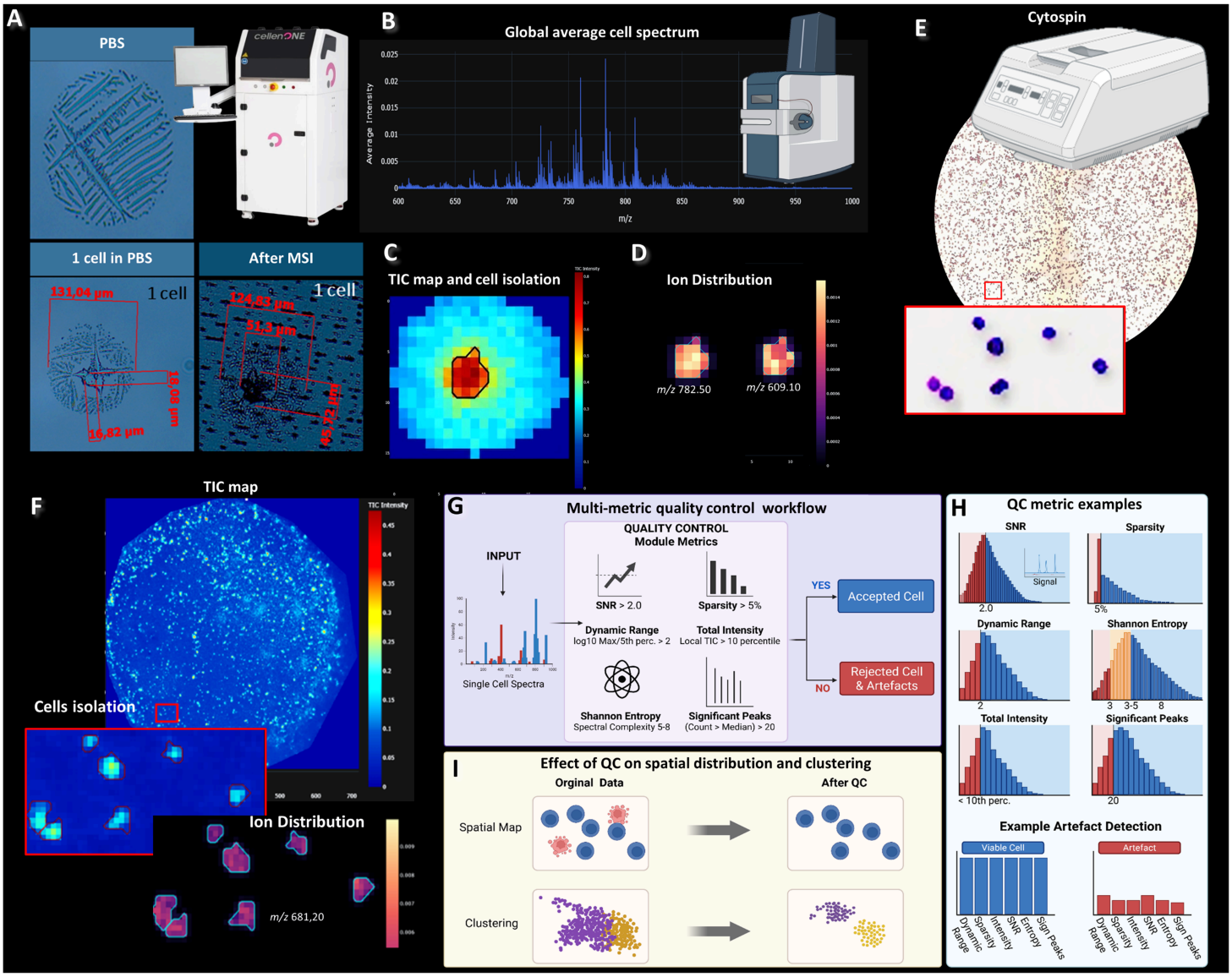


Figure 2

Single-cell validation and multi-metric quality control. (A–D) **Single-cell validation on a single MCF-7 cell**: (A) bright-field image before and after MALDI acquisition; (B) averaged mass spectrum with phospholipid region inset; (C) segmented cell contour overlaid on TIC map; (D) ion distribution maps at *m/z* 782.50 (PC 36:4, [M+H]+) and *m/z* 609.10 (SM 34:1, [M+H]+) spatially confined to the segmented region. (E–F) **Multicellular cytospin validation**: (E) TIC map with multiple discrete intensity maxima; (F) segmented contours confirmed by toluidine blue optical microscopy. (G–I) Six-metric QC module: (G) overview of the six complementary metrics with acceptance thresholds; (H) metric distributions for accepted (blue) and rejected (red) cells, illustrating artefact classes captured by each criterion; (I) effect of QC filtering on spatial cell maps and downstream UMAP clustering.

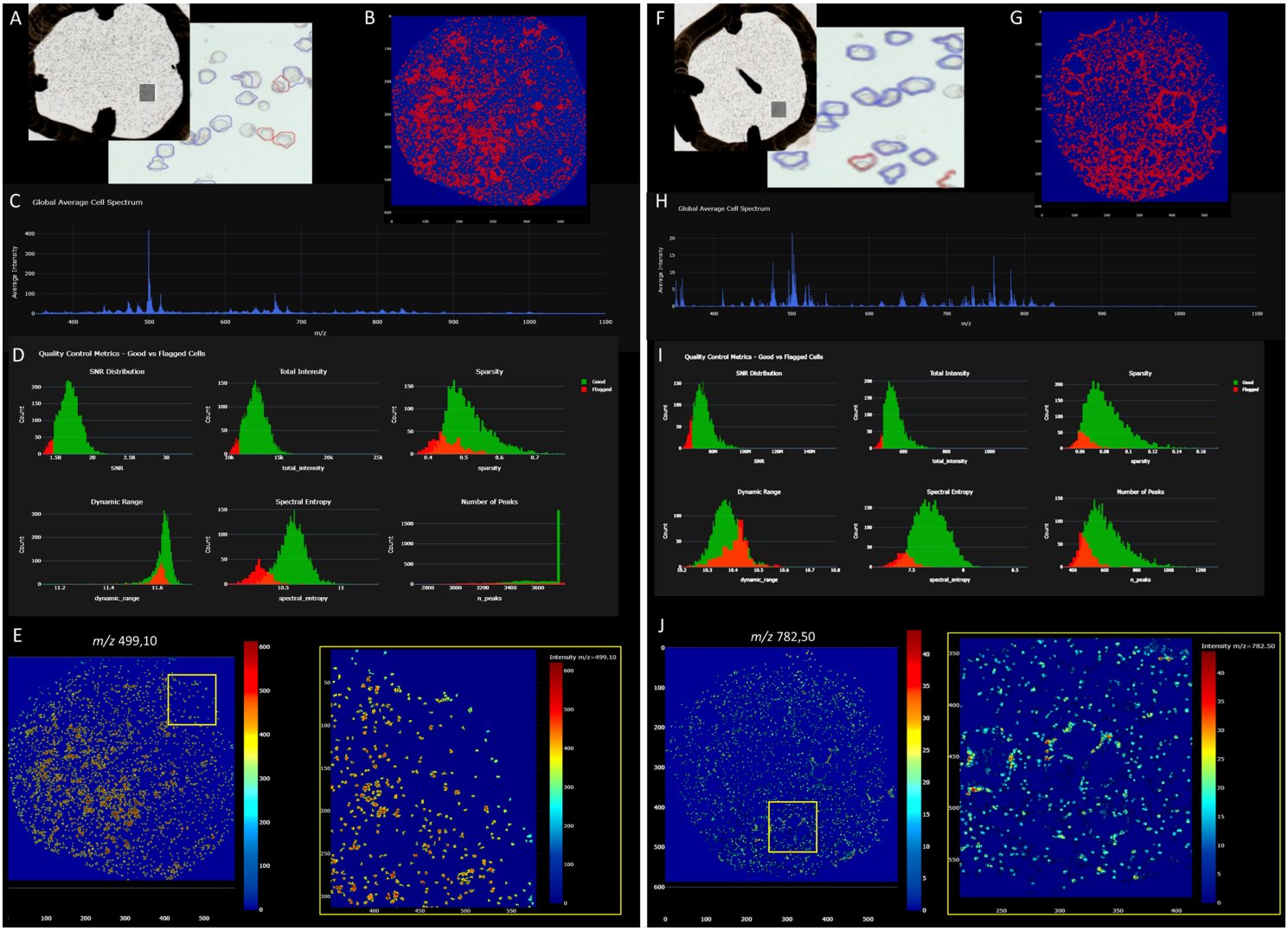


Figure 3

ionScell profiling of NCH82 glioblastoma cells in negative (left panels, A–E) and positive (right panels, F–J) ionization modes. (A, F) Optical images with segmented contours overlaid. (B, G) TIC intensity maps. (C, H) Global average spectra across all retained cells with spectral region insets. (D, I) Six-metric QC histograms (green: accepted; red: rejected). (E, J) Ion spatial distribution maps at *m/z* 499.10 (LPA 16:0, [M−H]−, negative) and *m/z* 782.50 (PC 36:4, [M+H]+, positive), co-localized with segmented cell boundaries. The two modes provide complementary anionic (phosphatidylinositols, sulfatides, lysophosphatidic acids) and cationic (glycerophosphocholines) lipid coverage at single-cell resolution.

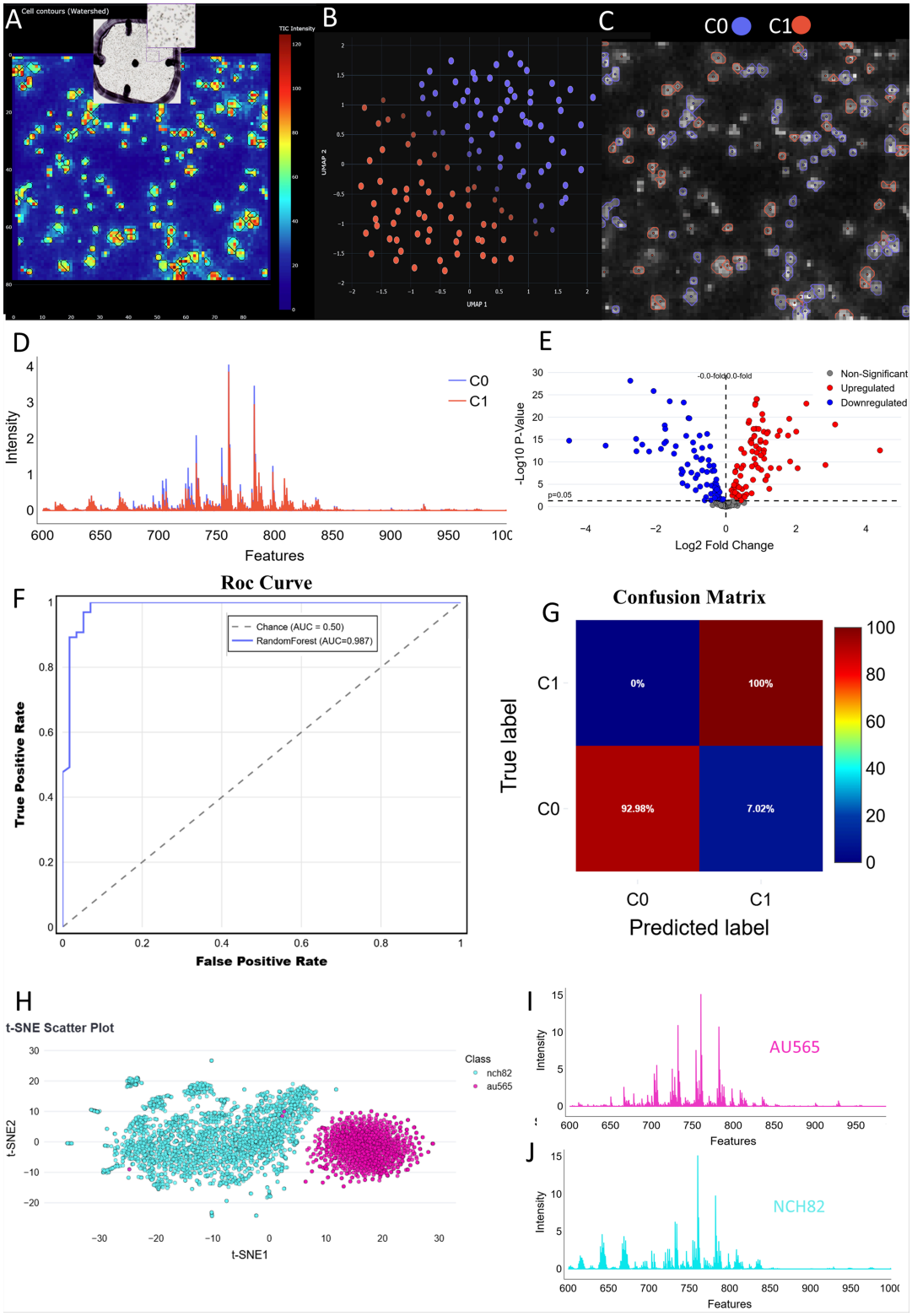


Figure 4

ionScell-based discrimination of a binary NCH82/AU565 cell mixture. (A) TIC map with segmented contours. (B) UMAP projection with soft GMM clustering (C0: blue; C1: red). (C) Spatial back-projection confirming stochastic co-deposition. (D) Skyline spectral overlays for C0 and C1. (E) Volcano plot of differential *m/z* features (fold-change vs. $-\log_{10}$ p-value). (F) ROC curves for five classifiers evaluated by cross-validated model selection; Random Forest achieves AUC = 0.987. (G) Confusion matrix (96%

global accuracy). (H) t-SNE reference projection confirming cluster identity assignments (C0 = AU565; C1 = NCH82). (I, J) Global average spectra for C0 (AU565) and C1 (NCH82), respectively, illustrating the spectral basis of cluster separation.

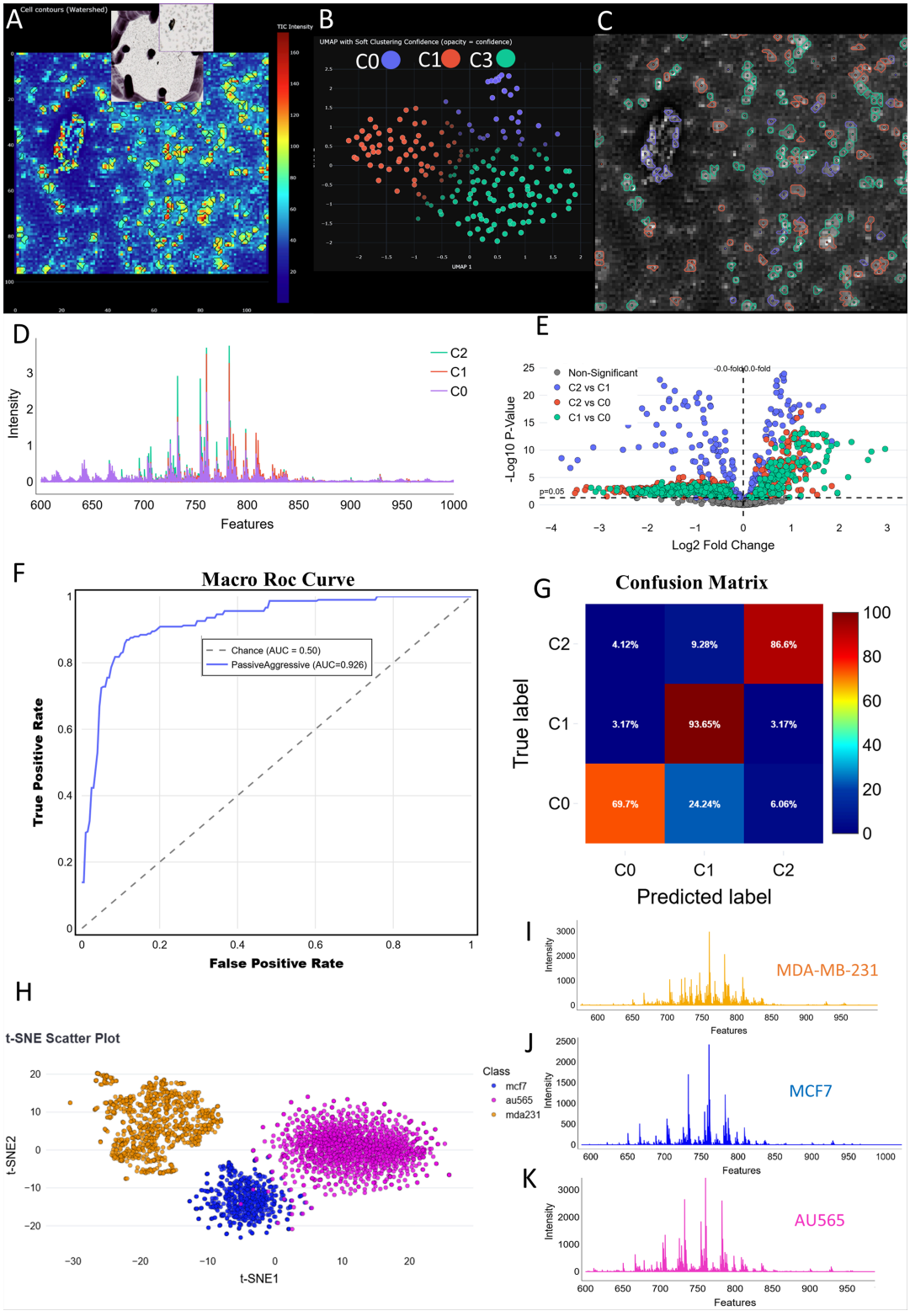


Figure 5

***ionScell*-based discrimination of a ternary MDA-MB-231, MCF-7 and AU565 cell mixture**. (A) TIC map with cell contours. (B) UMAP with three-way soft clustering (C0: blue; C1: red; C2: green). (C) Spatial back-projection confirming stochastic co-deposition. (D) Skyline spectral overlays for C0, C1 and C2. (E) Pairwise volcano plots. (F) ROC curves (Passive Aggressive classifier, AUC = 0.926; 86% global accuracy). (G) Confusion matrix showing asymmetric classification consistent with molecular proximity between breast cancer subtypes. (H) t-SNE reference projection. (I–K) Global average spectra for C0 (MDA-MB-231), C1 (MCF-7) and C2 (AU565), respectively, illustrating the spectral basis of three-way cluster separation.

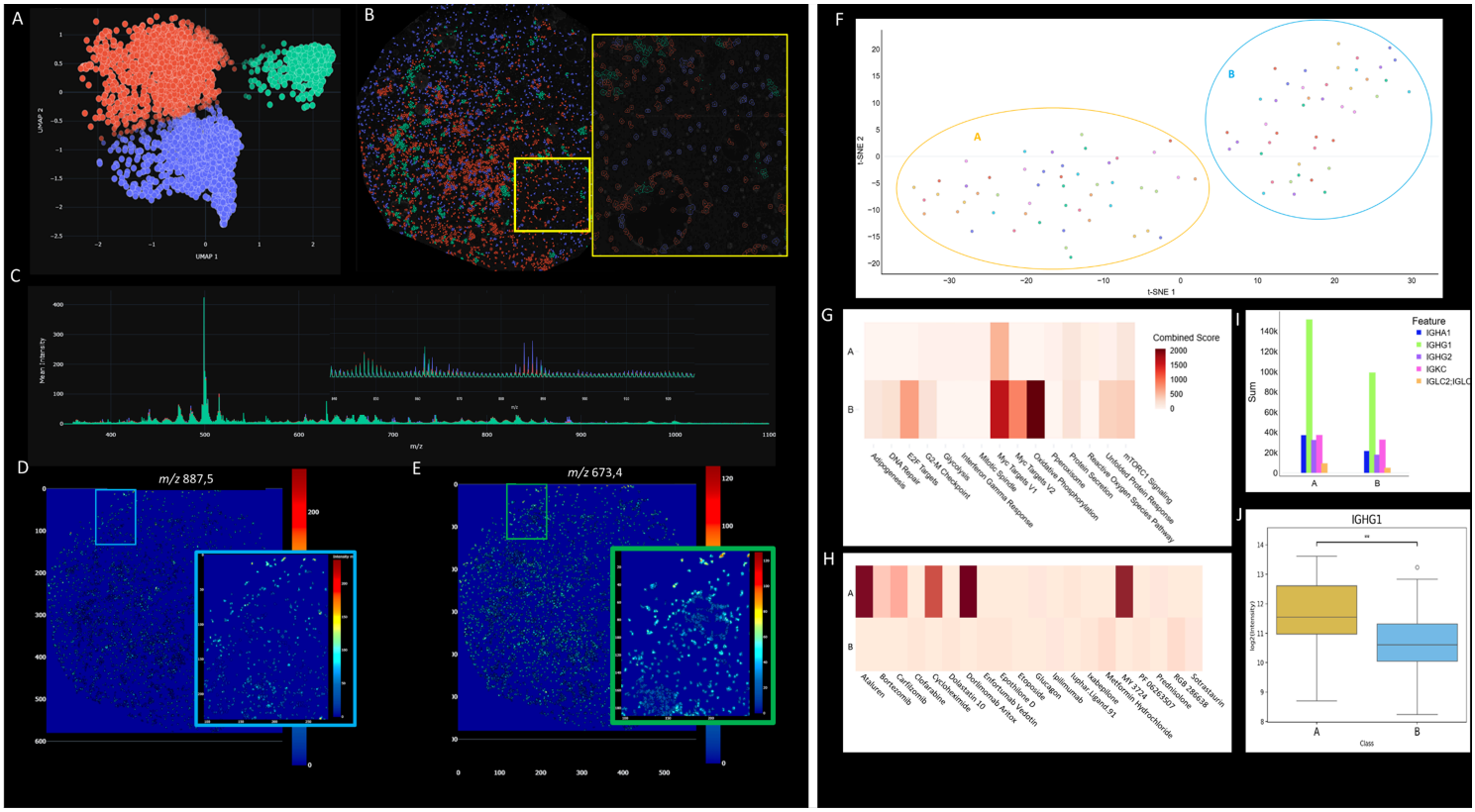


Figure 6

Intra-line molecular heterogeneity in NCH82 cells and multi-modal validation. (A–E) MSI-based subpopulation analysis (negative ion mode): (A) UMAP colored by MSp identity (MSp 0: blue; MSp 1: red; MSp 2: green); (B) spatial back-projection of MSp assignments on segmented NCH82 cells with a zoomed region of interest; (C) mean spectra per MSp; (D,E) ion maps at m/z 887.50 (PI 38:4, [M−H]−) and m/z 673.40 (PI 34:1, [M−H]−), respectively. (F–J) Single-cell proteomic validation (n = 233 cells, dia-PASEF): (F) t-SNE projection revealing Class A (quiescent) and Class B (proliferative); (G) MSigDB Hallmark pathway enrichment heatmap; (H) Drug Target 2024 enrichment in Class A; (I) immunoglobulin-family protein expression; (J) IGHG1 expression comparing Class A and Class B. Positive-mode MSI (Supplementary Figure S24) independently identifies two MSps concordant with the proteomic binary partition, while negative-mode MSI resolves a third anionic-lipid-defined substate invisible to both cationic profiling and proteomics.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ionScellsupp.docx