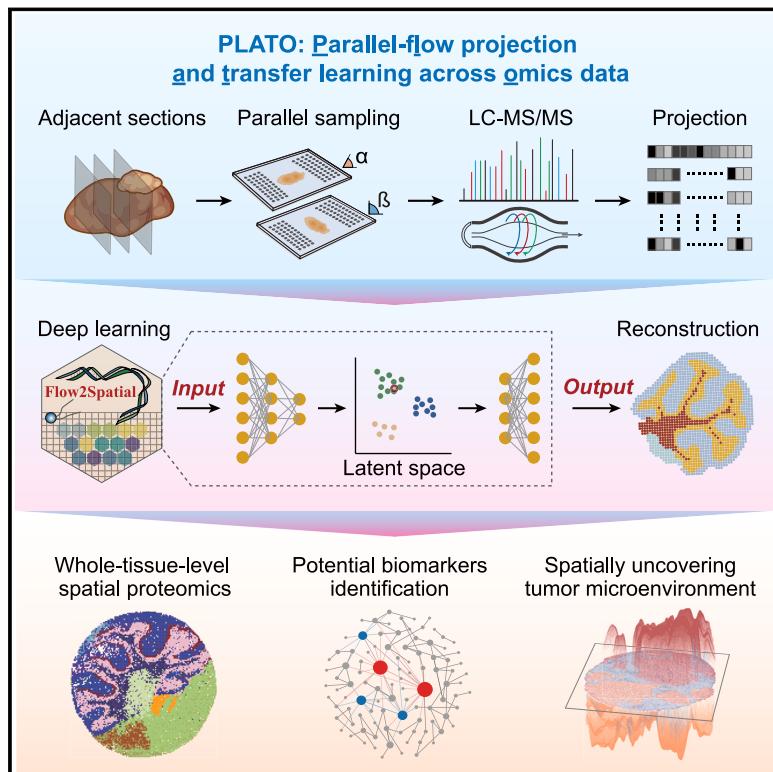


High-resolution spatially resolved proteomics of complex tissues based on microfluidics and transfer learning

Graphical abstract



Authors

Beiyu Hu, Ruiqiao He, Kun Pang, ..., Weihu Wang, Peifeng Ji, Fangqing Zhao

Correspondence

jipeifeng@biols.ac.cn (P.J.), zhfq@ioz.ac.cn (F.Z.)

In brief

PLATO, a high-resolution and high-throughput spatial mass spectrometry proteomics platform, identifies distinct tumor subtypes and key dysregulated proteins in human breast cancer.

Highlights

- PLATO enables high-resolution spatial proteomics across entire tissues
- Spatially profiles thousands of proteins using mass spectrometry and microfluidics
- Is compatible with diverse tissues, from model organisms to human samples
- Identifies distinct tumor subtypes and dysregulated proteins in human breast cancer

Article

High-resolution spatially resolved proteomics of complex tissues based on microfluidics and transfer learning

Beiyu Hu,^{1,2,6} Ruiqiao He,^{1,3,6} Kun Pang,^{1,3,6} Guibin Wang,^{4,6} Ning Wang,^{1,3} Wenzhuo Zhu,¹ Xin Sui,⁵ Huajing Teng,⁵ Tianxin Liu,¹ Junjie Zhu,^{1,2} Zewen Jiang,¹ Jinyang Zhang,¹ Zhenqiang Zuo,¹ Weihu Wang,⁵ Peifeng Ji,^{1,*} and Fangqing Zhao^{1,2,3,7,*}

¹Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

²Key Laboratory of Systems Biology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, China

³University of Chinese Academy of Sciences, Beijing 100049, China

⁴State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China

⁵Key Laboratory of Carcinogenesis and Translational Research, Department of Radiation Oncology, Peking University Cancer Hospital and Institute, Beijing 100142, China

⁶These authors contributed equally

⁷Lead contact

*Correspondence: jipeifeng@biols.ac.cn (P.J.), zhfq@ioz.ac.cn (F.Z.)

<https://doi.org/10.1016/j.cell.2024.12.023>

SUMMARY

Despite recent advances in imaging- and antibody-based methods, achieving in-depth, high-resolution protein mapping across entire tissues remains a significant challenge in spatial proteomics. Here, we present parallel-flow projection and transfer learning across omics data (PLATO), an integrated framework combining microfluidics with deep learning to enable high-resolution mapping of thousands of proteins in whole tissue sections. We validated the PLATO framework by profiling the spatial proteome of the mouse cerebellum, identifying 2,564 protein groups in a single run. We then applied PLATO to rat villus and human breast cancer samples, achieving a spatial resolution of 25 μm and uncovering proteomic dynamics associated with disease states. This approach revealed spatially distinct tumor subtypes, identified key dysregulated proteins, and provided novel insights into the complexity of the tumor microenvironment. We believe that PLATO represents a transformative platform for exploring spatial proteomic regulation and its interplay with genetic and environmental factors.

INTRODUCTION

Dissociation-based single-cell technologies have enabled the deep characterization of cellular heterogeneity and complexity.¹ However, how cells constitute tissue organization, which is vital for understanding the design principles of complex tissues and tumors, remains an open issue. Spatial genomics and transcriptomics^{2–4} have recently emerged to bridge this gap by aiding in investigations on tissues *in situ* at cellular and subcellular resolution. Nevertheless, these sequencing-based technologies have been primarily confined to indirectly measuring cellular states, as most biological processes (BPs) are controlled by proteins. The abundance relationship between protein and transcript is complicated and dependent on the experimental context, which challenges biological interpretation. Furthermore, protein posttranslational modifications (PTMs), involved in a large number of BPs, extend far beyond the coverage of nucleic acid measurements. High throughput and spatial protein

profiling, therefore, promises to revolutionize our understanding of BPs not only from a fundamental cell biology perspective but also from a clinical perspective.

Recent advances in multiplexing technologies have promised to simultaneously map up to hundreds of proteins in the same tissue section by using antibodies tagged with DNA,⁵ fluorophore,⁶ or metal.^{7,8} Nevertheless, these approaches, along with other protein profiling schemes based on immunoassay, ultimately have suffered from several inherent limitations. Generating antibodies, in practice, is prohibitively time consuming and expensive. Moreover, technologies based on these approaches share a major caveat in the limited number of antigens they probe, which is far short of the actual complexity of the proteome.⁹ The situation is even more fundamentally skewed, considering alternatively spliced transcripts and PTMs. In this regard, unbiased and proteome-scale spatial protein mapping approaches are urgently needed to overcome such challenges.

Mass spectrometry (MS)-based proteomic approaches offer label-free analyses with high specificity and deep proteomic coverage^{10,11} and have recently been applied to reveal key regulators in response to drug perturbation at single-cell resolution.¹² As a complement to MS, laser capture microdissection (LCM) permits areas of interest within a tissue to be isolated and subsequently provides comprehensive molecular coverage of proteomes.^{13,14} For instance, a recent development of deep visual proteomics (DVP),¹⁵ which combines ultrahigh-sensitivity MS and LCM, provides the ability to probe small to cellular objects in depth. These methods hold great promise for proteome-wide spatial protein profiling of focused tissue areas, but they first necessitate highly specialized laboratory facilities. Moreover, these approaches are typically low throughput and inevitably involve bottlenecks in whole-tissue level investigations in which extremely high-throughput applications are necessary. An alternative approach involves the use of matrix-assisted laser desorption ionization MS imaging (MALDI-MSI). The strength of MALDI-MSI is the ability to directly survey spatial proteomic complexity down to the cellular or subcellular scale, but this method has proven to be difficult because of several challenges, including simultaneous peptide quantification and identification,¹⁶ as well as low-abundance proteins. Most recently, a microscaffold-assisted spatial proteomics (MASP) strategy was developed,¹⁷ which utilizes micro-compartmentalization of tissues based on a 3D-printed microscaffold to map thousands of proteins at whole-tissue level. However, a large number of liquid chromatography-tandem mass spectrometry (LC-MS/MS) measurements (~900 samples) are still necessary, and the resolution is low (400 μm).

Thus far, it is highly desirable to develop new methods for high spatial resolution, proteome-scale mapping, covering all regions of a tissue slice, which do not require sophisticated imaging but leverage the power of MS to achieve high throughput and cost efficiency. Here, we present a high-resolution spatially resolved proteomics framework based on parallel-flow projection and transfer learning across omics data (PLATO) to localize thousands of proteins covering all regions of a tissue slice. We first demonstrated that our approach allows for spatial expression patterns to be reconstructed at the whole-tissue level using the mouse cerebellum. We further validated the ability of PLATO to measure the protein expression of intestinal villi at 25 μm spatial resolution. Finally, we applied PLATO to heterogeneous human breast cancer tissues, identifying two distinct tumor subtypes characterized by dysregulated protein profiles and a complex tumor microenvironment.

RESULTS

The PLATO framework

A key limitation of current LC-MS/MS-based spatial proteomics is its low detection throughput. To overcome this, we developed PLATO, a microfluidics-based and transfer learning-based spatial proteomics framework, which is aimed at determining high-resolution spatial protein patterns across entire tissues by using minimal LC-MS/MS measurements (Figure 1A). The PLATO workflow begins with cryosectioning three consecutive tissue slices: the middle slice is used for generating reference

omics data through histological staining or spatial omics (e.g., spatial transcriptomics), while the first and last slices are subjected to microfluidics-based proteomic profiling at different angles. Each slice is covered by parallel microchannels, and after on-chip digestion, peptides are collected for LC-MS/MS analysis. These measurements, referred to as parallel-flow projections, resemble ray-based tomography.^{18–20}

To reconstruct protein spatial distributions, we developed Flow2Spatial, a transfer learning algorithm that utilizes reference omics data (e.g., H&E staining, spatial transcriptomics, or spatial metabolomics) to train a deep learning model for predicting protein distributions from the parallel-flow projections. Instead of relying solely on mRNA-protein correlations, Flow2Spatial employs clustering features and orthogonal projections to infer spatial protein patterns (see **STAR Methods**).

We validated PLATO by addressing several key aspects: (1) optimizing on-chip MS sample preparation, including chip design and digestion methods; (2) verifying cross-contamination, reliability, and reproducibility in protein quantification; and (3) assessing Flow2Spatial's accuracy in reconstructing spatial patterns through simulated datasets and experimental approaches such as immunofluorescence (IF) and LCM-based proteomics.

On-chip proteomic preparation and cross-contamination evaluation

To achieve parallel and high-throughput tissue sampling, we used a microfluidic chip featuring 48 or 70 parallel microchannels, with each channel measuring 25–100 μm in width. The width of the channels matched the thickness of the channel walls across all chips (Figures 1B and S1A). The channels were designed to be equal in length to minimize fluid resistance differences, improving sample collection efficiency (Figure S1B). Since no ready-to-use methods were available for chip-based nanogram-scale proteomics, we systematically optimized the on-chip proteomic preparation workflow. We compared a surfactant-aided pipeline using 4-hexylphenylazosulfonate (AZO)²¹ and n-dodecyl-β-D-maltoside (DDM)²² with a direct-lysis pipeline. The direct-lysis method resulted in a 2-fold increase in the number of identified protein groups, compared with the surfactant-aided approach (Figures 1C and S1C). This improvement likely stems from trypsin adsorption to the microchannel before digestion, which reduces nonspecific protein adsorption (Figure S1D). Additionally, the direct-lysis method provided effective, unbiased on-chip tissue digestion (Figures 1D and S1E).

Next, we evaluated cross-contamination between microchannels, as proteins may diffuse across adjacent channels during tissue lysis. Although no leakage was observed by imaging fluorescent molecule flow (Figures 1E and S1F), a more stringent MS-based evaluation was conducted to assess protein diffusion (Figure 1F). Specifically, after placing the 25-μm chip on tissue, we added digestion buffer to three adjacent microchannels, including an *E. coli* lysis mix in the middle channel. MS analysis detected no *E. coli* protein signals in the adjacent channels. Similar tests were performed on 100-μm chips with consistent results (Figure S1G). These results demonstrate that the on-chip proteomic preparation workflow can effectively extract proteins with minimal cross-contamination.

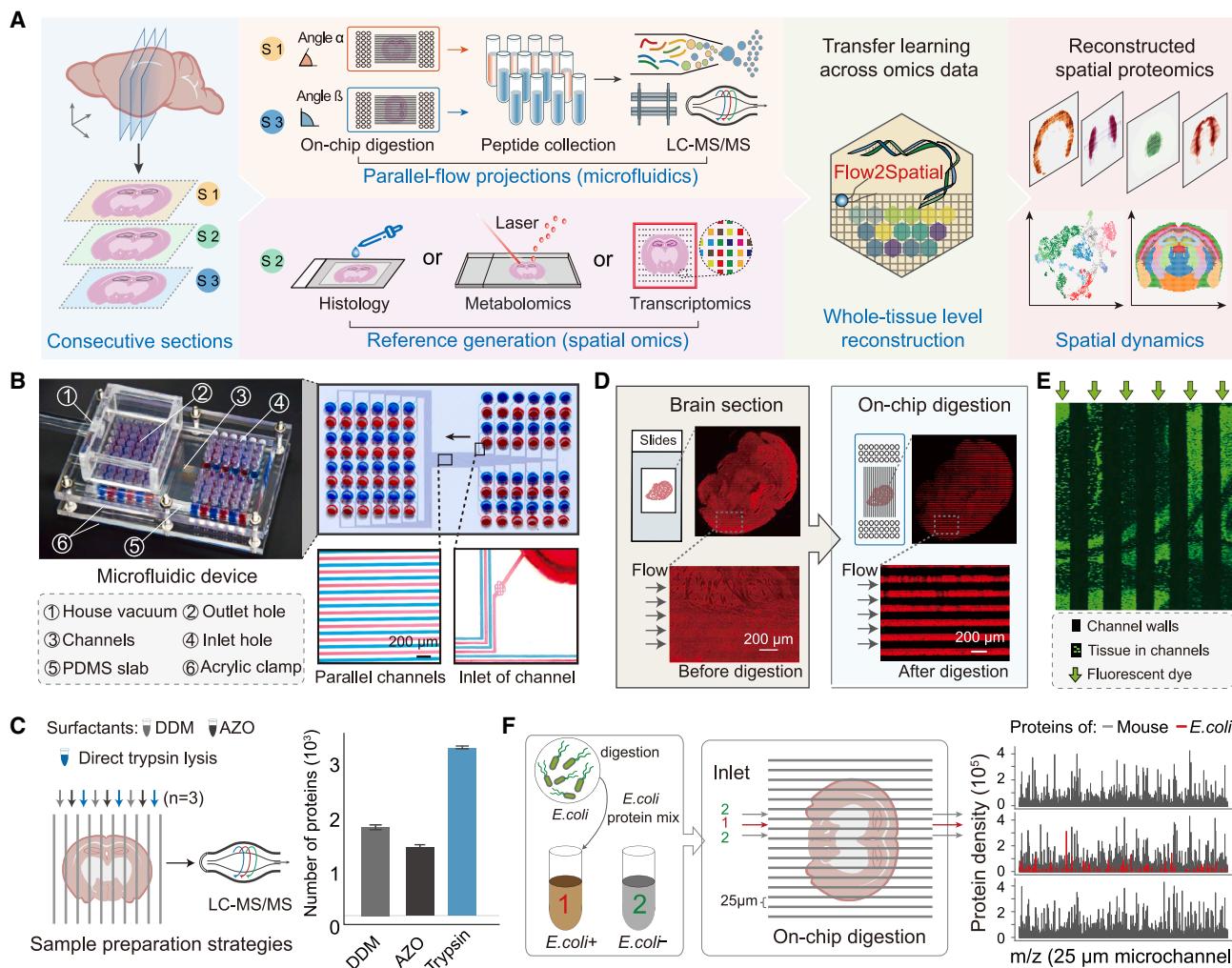


Figure 1. Design and experimental validation of the PLATO framework

(A) Schematic representation of the PLATO framework. The tissue was first cut along the axis to obtain three consecutive sections (with thicknesses of 8–10 μm), and the middle sections were subjected to H&E staining or spatial omics. Subsequently, the first and third slices were subjected to microfluidic chip-based parallel-flow proteomic profiling at two different angles. Slices were digested on the chip. Peptides in each microchannel were drawn out, collected, and subjected to LC-MS/MS quantification. The measurements for each angle are referred to as the parallel-flow projection. Finally, reconstruction of protein expression patterns was performed by using the Flow2Spatial algorithm, based on the images of the middle slices and two sets of parallel-flow projections.

(B) A mouse expressing red fluorescent protein was used to validate tissue digestion. Brain slices were covered by a microfluidic chip and subjected to on-chip digestion. Most tissue regions underneath the microchannels were digested, whereas those underneath the walls remained intact.

(C) Optimization of on-chip tissue digestion. Left, tissues underneath the microchannel were lysed by MS-friendly surfactants such as AZO and DDM. Then, the lysed tissue was processed by reduction, alkylation, and trypsin digestion. Meanwhile, tissues in the microchannels were directly lysed by trypsin without extra reduction and alkylation processes. Right, the number of detected protein groups in each condition.

(D) Leakage evaluation using fluorescent dye of propidium iodide (PI).

(E) Leakage evaluation of on-chip digestion using spike-in bacterial proteins. Upon covering chip on the tissue, three adjacent microchannels were chosen to add digestion buffer, with additional protein lysis buffer of *E. coli* added to the middle microchannel (middle). Upon digestion, the flows were collected and analyzed using LC-MS/MS (right), followed by identifying *E. coli* proteins within the two side microchannels.

See also Figure S1.

Quantitative evaluation of on-chip proteomic samples

To evaluate the performance of PLATO in protein quantification, which is crucial for accurate spatial reconstruction, we systematically assessed its sensitivity, proteome coverage, and reproduc-

ability. We conducted measurements on a dilution series of mouse cerebellum tissue lysates, ranging from 1 to 8 μL , with a protein concentration of approximately 30 ng/ μL . As expected, metrics improved with larger lysate volumes. Notably, even at the lowest

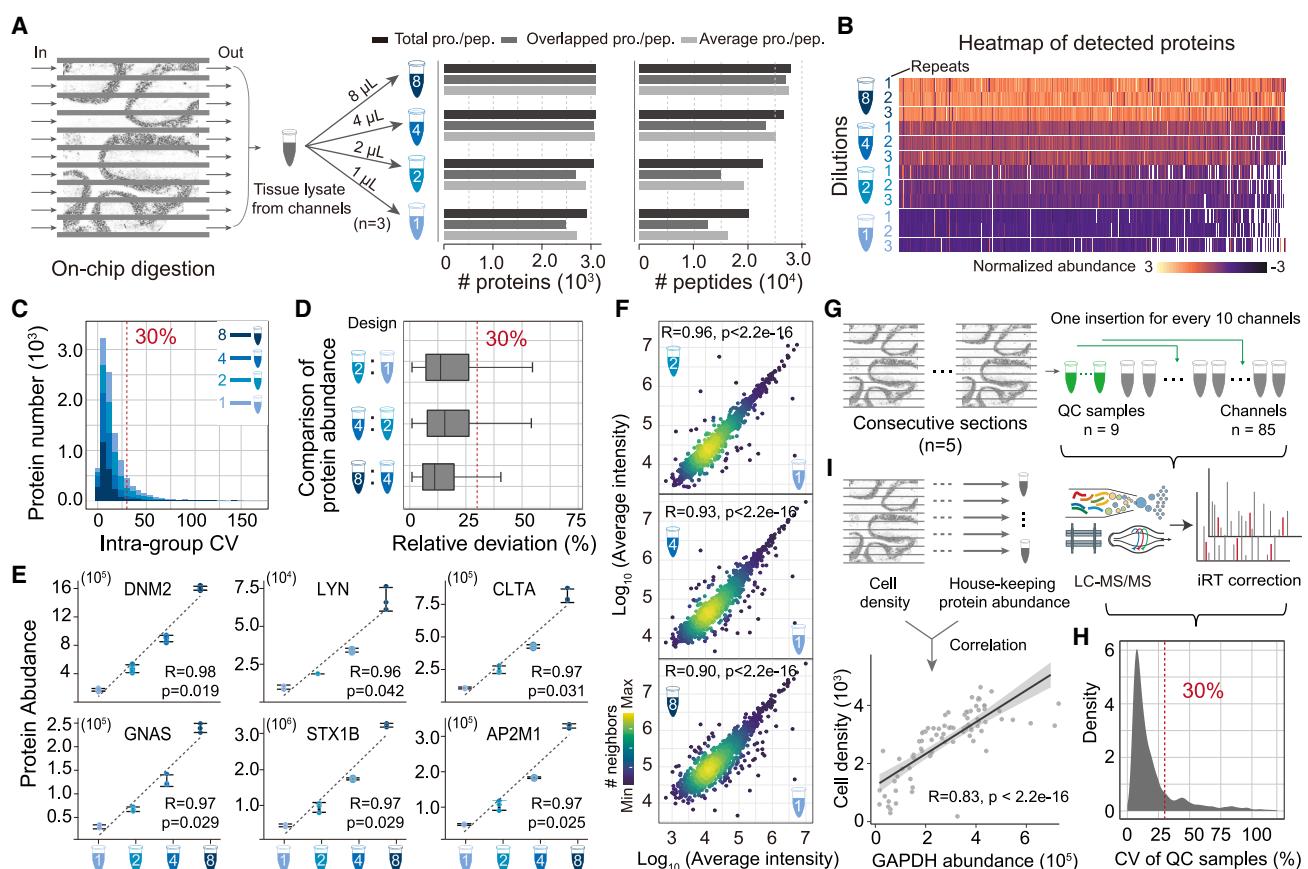


Figure 2. Protein quantification and reproducibility evaluation

(A) Schematic representation of the protein quantification performance validation workflow, using mouse cerebellum tissue. A dilution series (1, 2, 4, and 8 μ L) of mouse cerebellum tissue lysates was obtained by mixing the products of 10 microchannels from 1 to 8 μ L, with each diluted sample analyzed in triplicate (left and middle). These samples were subsequently analyzed by LC-MS/MS. Next, the detected protein groups and peptides, as well as overlapping among replicates, were identified (right). The colored tubes denote different lysate concentrations.

(B) Abundance of detected protein groups in each sample. Each line denotes a detected protein group and is colored according to its abundance.

(C) Histogram showing the coefficient of variation (CV) of each protein group's abundance in four groups with different dilutions. The red line indicates 30% CV for visual reference.

(D) Boxplots showing the relative deviation between the actual quantitative abundance ratio and the theoretical ratio of each protein between adjacent dilutions groups (1 and 2 μ L; 2 and 4 μ L; 4 and 8 μ L). The red line indicates a 30% relative deviation for visual reference.

(E) Representative examples of proteins showing quantitation of protein abundance calculated from peak area in four groups with different dilutions. Data are shown as the mean \pm SD from $n = 3$ independent measurements.

(F) Quantitative reproducibility in a rank order plot for the dilution with the lowest level (1 μ L) of lysate concentration with other sets of dilutions (2, 4, and 8 μ L from top to bottom).

(G) Schematic representation of the protein quantification performance validation workflow, using a larger number of samples. On-chip digestion of five consecutive mouse cerebellum sections was performed (upper left). This generates 9 QC samples and 85 microchannel-derived samples (upright). These samples were analyzed by LC-MS/MS with the data-independent acquisition (DIA) model (bottom). Finally, the raw intensity of the identified protein groups was adjusted by using the intensity of iRT.

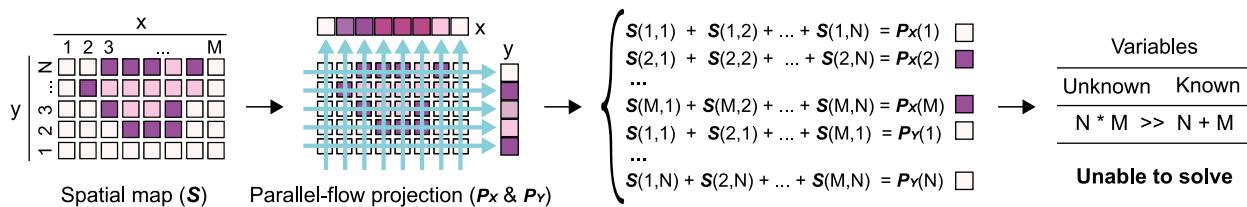
(H) Histogram showing CV of each protein group's abundance in QC samples. The red line indicates 30% CV for visual reference.

(I) Correlation between the abundance of GAPDH protein and cell density, calculated from the H&E image. Each circle represents the cell density and GAPDH abundance in the channel sample.

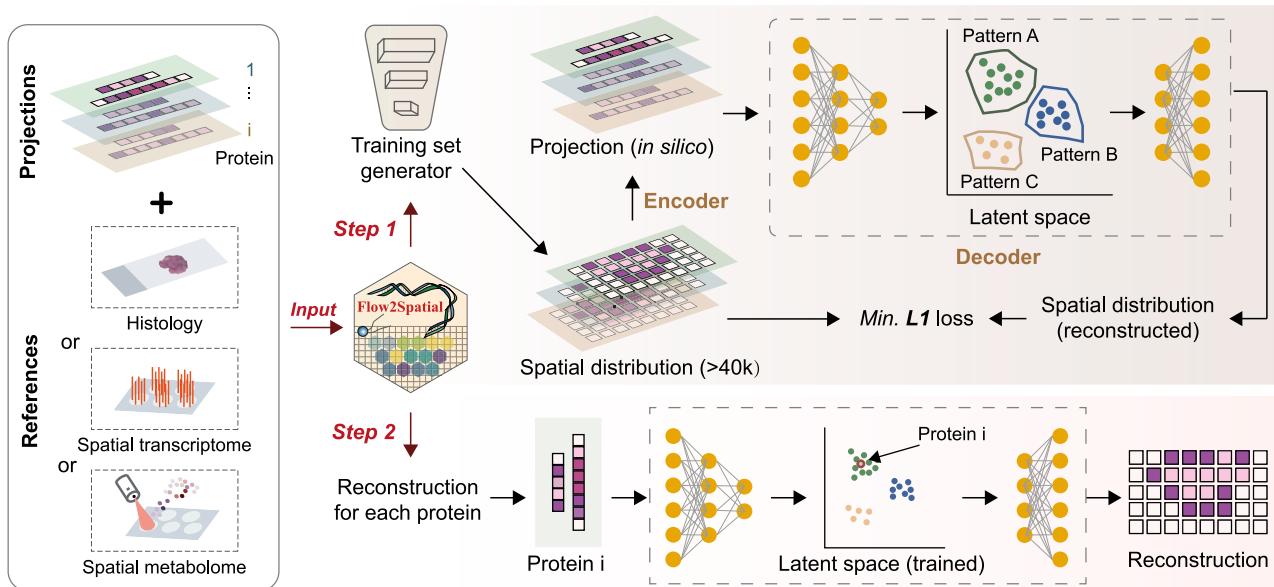
amount of 1 μ L (~30 ng), 85.32% of 2,931 proteins were reproducibly detected (Figures 2A and 2B). Most proteins showed low quantification variation, as reflected by their coefficients of variation (CVs) (Figures 2C and 2D). Quantification accuracy was validated by comparing experimental values with theoretical abundances, showing a strong linear response, such as for dynamin-2 (DNM2), tyrosine-protein kinase Lyn (LYN), and clathrin light

chain A (CLTA) (Figure 2E). The abundance ranking of proteins across different dilutions also exhibited a high correlation (Figure 2F). Robustness was further confirmed by analyzing 85 microchannel-derived samples, yielding low CVs in quality control (QC) samples (Figures 2G and 2H). Additionally, we observed a linear correlation between glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression and cell density (Figure 2I). These

A The problem: spatial reconstruction from orthogonal projection



B The solution: Flow2Spatial reconstruction based on transfer learning



C

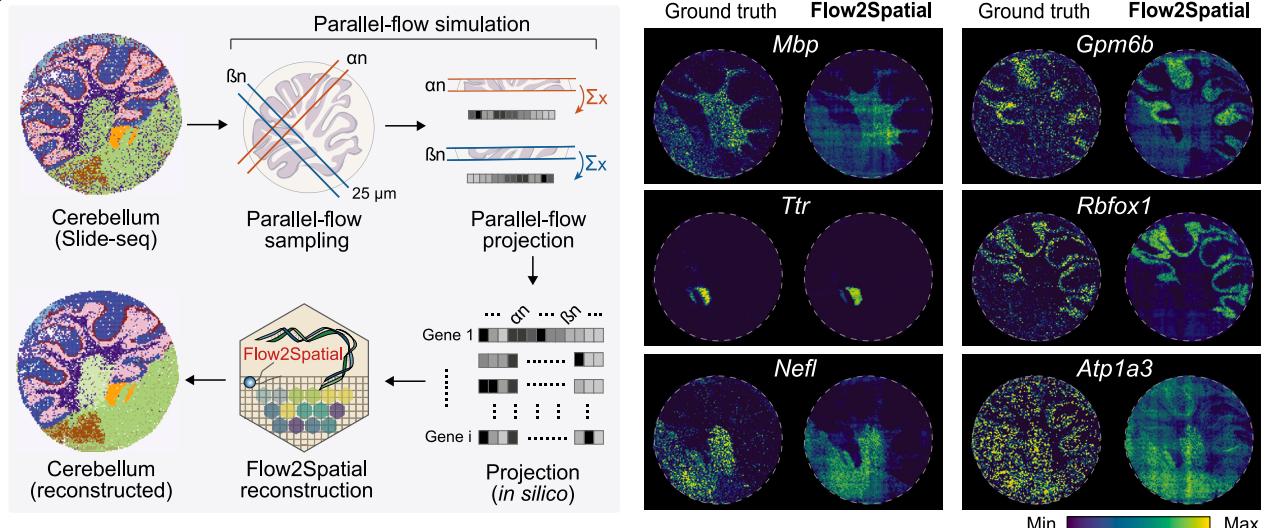


Figure 3. Workflow of the Flow2Spatial algorithm and performance evaluations on simulated datasets

(A) Reconstructing spatial patterns solely on orthogonal projections proves to be insufficient.

(B) Workflow of the Flow2Spatial algorithm. The transfer learning process is a two-step procedure, encompassing transfer learning (upper) and reconstruction (lower). In the first step, Flow2Spatial leverages readily available omics data, such as histological staining, spatial transcriptome, or spatial metabolome, from the middle section. This data serve as the foundation for constructing a generator tasked with producing a novel spatial dataset by randomly integrating fully measured spatial omics data, which aims to enhance the diversity of spatial distributions. Subsequently, the entire set of generated spatial data is employed as the training dataset for the deep learning model. In detail, Flow2Spatial employs an autoencoder-like deep learning model to establish connections between

(legend continued on next page)

results demonstrate the reliability and reproducibility of PLATO, supporting spatial reconstruction efforts.

Reconstruction workflow of Flow2Spatial

Previous studies have shown that classic computer tomography cannot resolve the reconstruction problem when the number of parallel-flow projections is less than three^{18,23} (Figure 3A). We thus developed a deep learning-based method, Flow2Spatial, to reconstruct the original spatial distribution from the parallel-flow projected values of a few angles. In theory, if we were to use all potential spatial distributions in a 2D space as a training dataset, we could build a deep learning model that connects these spatial distributions to the values of their corresponding projection strips. However, the size of such a training dataset grows exponentially as the number of spots in the 2D space increases. For example, in a basic scenario where a gene is expressed or not in a 2D space with dimensions of $N \times M$, the training dataset would become as large as $2^{N \times M}$. In practice, creating and handling such a massive deep learning model is nearly impossible.

To address this challenge, we simplify the problem by incorporating information of other readily available spatial omics data from adjacent tissue sections, which exhibit high correlation with spatial proteomic patterns. This approach is based on the well-established notion of the strong correlation of spatial molecular patterns across different omics data, as supported by previous studies.^{5,24–35} Our findings also demonstrate that consistent spatial clusters can be generated by different spatial omics (Figure S2), making it feasible for the deep learning model to reconstruct spatial distributions within a closely related latent space. We refer to this approach as “transfer learning across omics data,” with the accessible omics data used for training termed “the reference.”

Specifically, Flow2Spatial operates through two key steps: transfer learning and reconstruction (Figure 3B). In the first step, Flow2Spatial integrates easily accessible omics data from the middle tissue section, such as histological staining, spatial transcriptomics (ST), or spatial metabolomics. It builds a generator to create a new spatial dataset by randomly combining these fully measured spatial omics, thereby enhancing the diversity of spatial distributions. Subsequently, all generated spatial data serve as the training set for the deep learning model (Figure S3). Specifically, Flow2Spatial employs an autoencoder-like deep learning model to establish connections between parallel-flow projections and spatial information. During the encoding phase, *in silico* projections of the newly generated spatial dataset are performed, simulating the microfluidic chip’s parallel-flow projection to create pseudostrips that span the entire section. In the decoding phase, a residual

network (ResNet)-based deep learning model³⁶ is trained using the pseudostrip values and corresponding ground truth data to minimize L1 loss (Figures 3B and S3). Finally, the model embeds all spatial patterns from the training dataset into a latent space, based on the relationships among the projected strip values. In the reconstruction step, the trained decoder reconstructs the spatial distribution of the experimentally identified proteome.

Spatial distribution of proteins can be accurately reconstructed by Flow2Spatial

To evaluate the potential of Flow2Spatial for resolving spatial heterogeneity, we first carried out *in silico* simulation experiments using a Slide-seq dataset of mouse cerebellum,³⁷ of which the ST profile was used as ground truth (Figure 3C). To mimic the microfluidic chip-based sampling strategy, the slice was analogously sliced along two orthogonal axes. For each angle, the tissue was sliced into consecutive 25-μm pseudostrips spanning the entire tissue, and in each pseudostrip, the expression of genes was obtained by summing the expression values of the spots underneath this pseudostrip. We compared ground truth profiles with the spatial reconstructions obtained from Flow2Spatial and Tomographer.¹⁸ After spatial reconstruction, we computed Spearman’s correlation coefficient, the relative error, and the total absolute difference among spots between reconstructions and the ground truth for all genes. Based on these metrics, we concluded that Flow2Spatial outperformed Tomographer in spatial pattern reconstruction of given genes (Figures S4A and S4B), particularly in detecting fine structures (Figures 3C and S4B) and clustering different region types (Figures 3C and S4C). Collectively, these results demonstrate the robustness of our approach and show its ability to dissect tissues at high resolution.

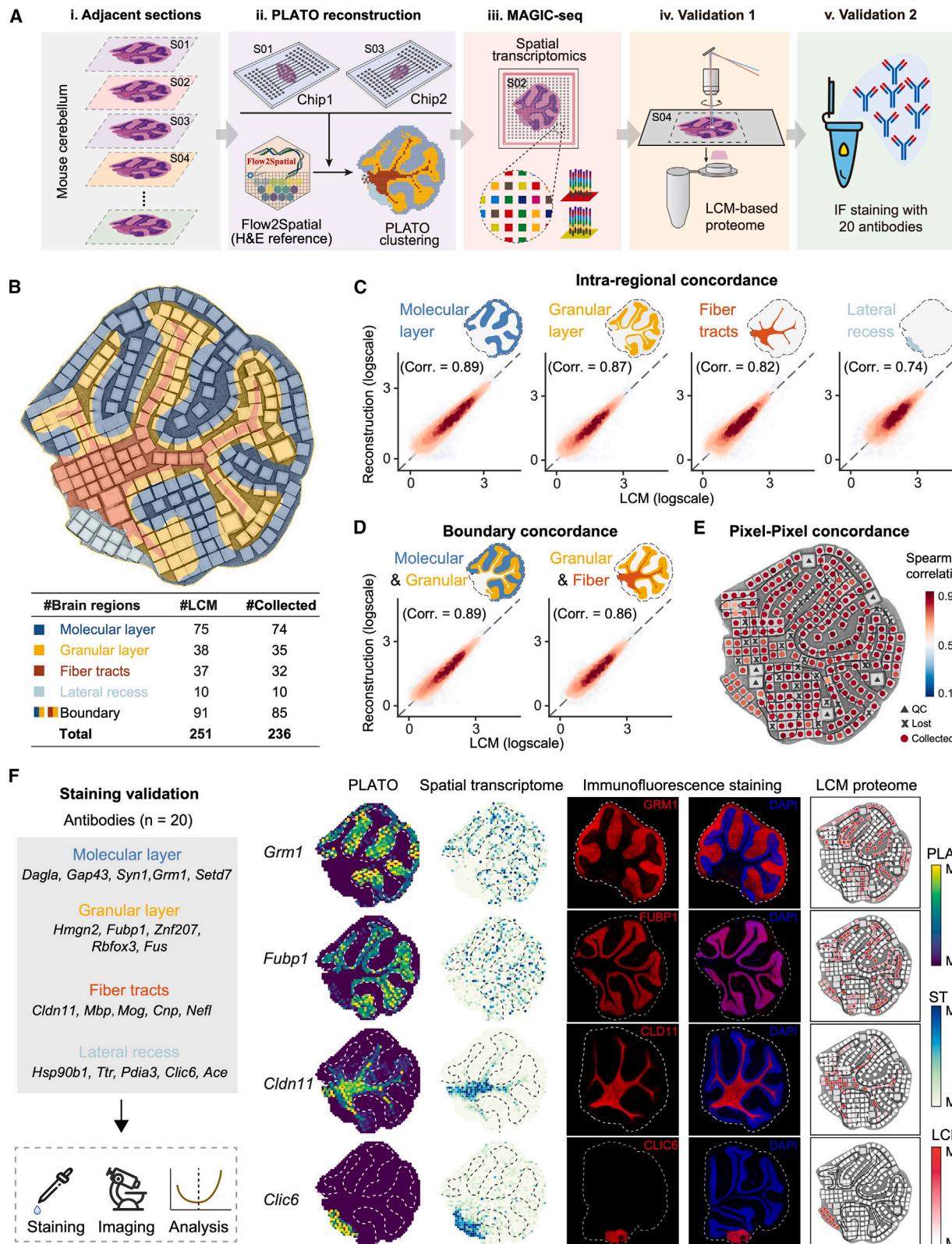
Spatial proteomics mapping of mouse cerebellum

To validate the performance of Flow2Spatial, we employed PLATO to generate a spatial protein map of the cerebellum (H&E staining as reference), followed by extensive validation using spatial transcriptomics, LCM-based spatial proteomics, and IF staining (Figure 4A). A microfluidic chip featuring 48 parallel channels, each 50 μm wide with equal channel and wall widths, was used for LC-MS/MS detection. This resulted in the identification of 6,086 protein groups, of which 5,722 were retained after QC (Figure S4D). Finally, a median of 2,564 proteins per spot was reconstructed. This represents a significant improvement, by at least an order of magnitude, over traditional antibody-based or MALDI-based spatial proteomics methods.^{6,7} Clustering based on these proteins revealed four distinct structural proteomic landscapes corresponding to histologically defined regions: the molecular layer, granular layer, fiber tracts, and lateral recess (Figure S4E).

parallel-flow projections and spatial information. This model embeds all spatial patterns in the training set into a latent space, based on the relationships across the values of strips after projection. In the reconstruction step, the trained decoder is used to reconstruct the spatial distribution of the experimentally detected proteome.

(C) Left, validation of the Flow2Spatial algorithm, using simulated data from Slide-seq spatial transcriptomics. Briefly, to mimic the microfluidic chip-based sampling strategy, the tissue was sliced along two orthogonal axes *in silico*. For each angle, the tissue was sliced into consecutive 25 μm pseudostrips spanning the entire tissue, where in each pseudostrip, the expression of genes was obtained by summing the expression values of the spots underneath this pseudostrip. Next, the Flow2Spatial algorithm is employed to reconstruct the spatial patterns of detected genes. Right, visualization of gene expression generated by Slide-seq and reconstructions.

See also Figures S2 and S3.



(legend on next page)

To assess the accuracy of Flow2Spatial in identifying spatial protein patterns, we collected 236 tissue voxels ($\sim 100 \times 100 \times 10 \mu\text{m}$) for LCM-based spatial proteomics (Figure 4B). LC-MS/MS analysis yielded an average of 1,849 protein groups after outlier removal (Figure S4F). When comparing the spatial proteomes between PLATO and LCM, we observed strong Spearman correlations across brain regions: molecular layer (0.89), granular layer (0.87), fiber tracts (0.82), and lateral recess (0.74) (Figure 4C). These correlations closely matched the baseline Spearman's coefficient of 0.80 observed among LCM samples from the same region (Figure S4G). Notably, high correlations were also maintained at regional boundaries (molecular layer and granular layer: 0.89, granular layer and fiber tracts: 0.86) (Figure 4D). To further evaluate pixel-level concordance, we calculated the Spearman's coefficients between each LCM sample and the corresponding spots in the PLATO map (Figure 4E) and the spatial transcriptomics data (Figure S4H). Only PLATO exhibited high expression correlations with the corresponding LCM samples, underscoring the accuracy of Flow2Spatial reconstructions.

Finally, IF staining of 20 region-enriched proteins (Figure 4F) confirmed that their spatial distributions closely aligned with the Flow2Spatial reconstructions (Figures 4F and S5). Collectively, these findings demonstrate that PLATO reliably maps the anatomical locations of proteins with high accuracy.

Ablation experiment on PLATO performance

To confirm that various spatial omics datasets can serve as references for PLATO, we employed Flow2Spatial to reconstruct these proteins using three different spatial omics datasets as references, including histology (H&E staining), ST, and spatial metabolome (MALDI-MSI). Notably, we observed a high degree of concordance in the clustering results of these reconstructions (Figure S6A). After calculating Spearman coefficient for the results reconstructed from references of H&E staining and ST, we found a median correlation of 0.85 (interquartile range: 0.82–0.87) (Figure S6B). Collectively, these findings demonstrate that Flow2Spatial consistently reveals the molecular organization of tissue heterogeneity, irrespective of the reference dataset used.

It should be noted that PLATO's ability to reconstruct the spatial distribution of a specific protein does not rely on the expression of its original mRNA but on the learned distribution patterns of various molecules. To verify this, we sought to investigate whether mRNAs with low correlation to proteins could still

effectively contribute to protein reconstruction within the PLATO framework. We divided reference genes of ST into two groups based on their expression similarity between mRNA and protein: highly correlated and lowly correlated. Each group served as a reference for subsequent transfer learning and spatial reconstruction. As expected, the results of Flow2Spatial based on two types of references demonstrated a strong correlation, with a median Spearman's coefficient of 0.89 (interquartile range: 0.86–0.91) (Figure S6C). Moreover, regionally enriched genes and clustering results also emphasized that the correlation level between mRNA and its corresponding protein does not influence the performance of Flow2Spatial reconstruction (Figure S6C).

To test the specificity of Flow2Spatial, we conducted experiments to assess whether Flow2Spatial could accurately reconstruct protein distributions when provided with incorrectly input reference images. We first swapped the values in the strips obtained from two different angles in PLATO. Specifically, we replaced the values obtained from one angle with those from another angle. After Flow2Spatial reconstruction, we observed a disordered clustering result (Figure S6D). Notably, regional markers such as claudin-11 (CLD11) and chloride intracellular channel protein 6 (CLIC6) exhibited changes in their orientation in the reconstructions. We further randomly shuffled the values in the strips before performing Flow2Spatial reconstruction. As expected, we found that both the clustering patterns and regional markers were disrupted (Figure S6E). These results underscore the high level of specificity exhibited by the Flow2Spatial algorithm.

Spatial proteomics mapping of the intestinal villus

To validate whether our method allows for high spatial resolution mapping, we further applied PLATO to large intestinal villus tissue from an adult rat. A microfluidic chip featuring 70 parallel channels, each $25 \mu\text{m}$ wide with equal channel and wall widths, was utilized (Figure 5A). Detection using LC-MS/MS successfully yielded a total of 1,986 protein groups (Figure 5B; Table S1), subsequently followed by Flow2Spatial reconstruction. Finally, a median of 1,183 proteins per spot was reconstructed, and spatial expression of representative proteins is shown in Figure 5C. Clustering based on these proteins identified four basic structural proteomic landscapes that were histologically discernible as epithelial cells, lamina propria, and muscularis (Figure 5C), in which representative proteins in these structures were then validated by IF (Figure 5D). For example, MYH11, a

Figure 4. Experimental validation of PLATO

(A) Workflow illustrating the experimental validation of PLATO on the mouse cerebellum tissue. Consecutive sections of cerebellar tissue were processed for spatial proteomics using PLATO, followed by spatial transcriptomics (MAGIC-seq), laser capture microdissection (LCM)-based proteomics, and immunofluorescence (IF) antibody staining.

(B) Top, image of the cerebellar section after LCM with regional annotations. Bottom, the number of LCM samples collected.

(C and D) Spearman's correlation of protein expression between LCM-based proteomics and Flow2Spatial reconstructions, analyzed both within specific regions and at regional boundaries. Each subgraph includes a schematic representation of the cerebellar region, with dot color indicating their density. Protein abundance data were log transformed and normalized using the normalizeBetweenArrays function from the limma package.

(E) Pixel-by-pixel Spearman's correlation of protein expression between LCM-based proteomics and Flow2Spatial reconstructions. QC, quality control samples represented by triangles. Lost: failed collections during microdissection and filtered samples represented by x-shaped markers. Collected: finally used samples represented by circles.

(F) Validation of Flow2Spatial by IF staining and LCM-based proteome. Left, the schema of IF staining by 20 antibodies. Right, four region-enriched proteins were validated by IF staining and LCM proteome.

See also Figures S4, S5, and S6.

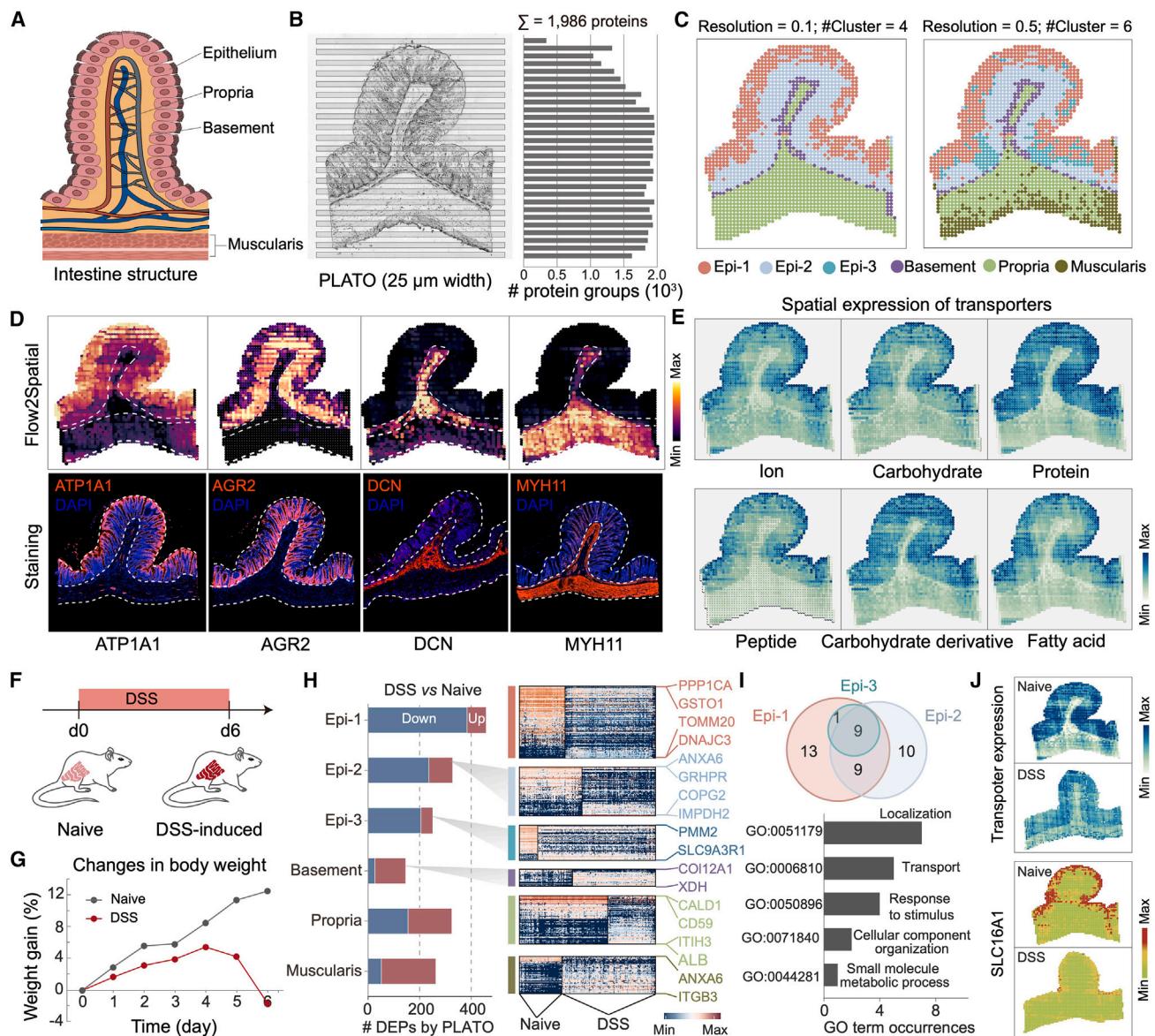
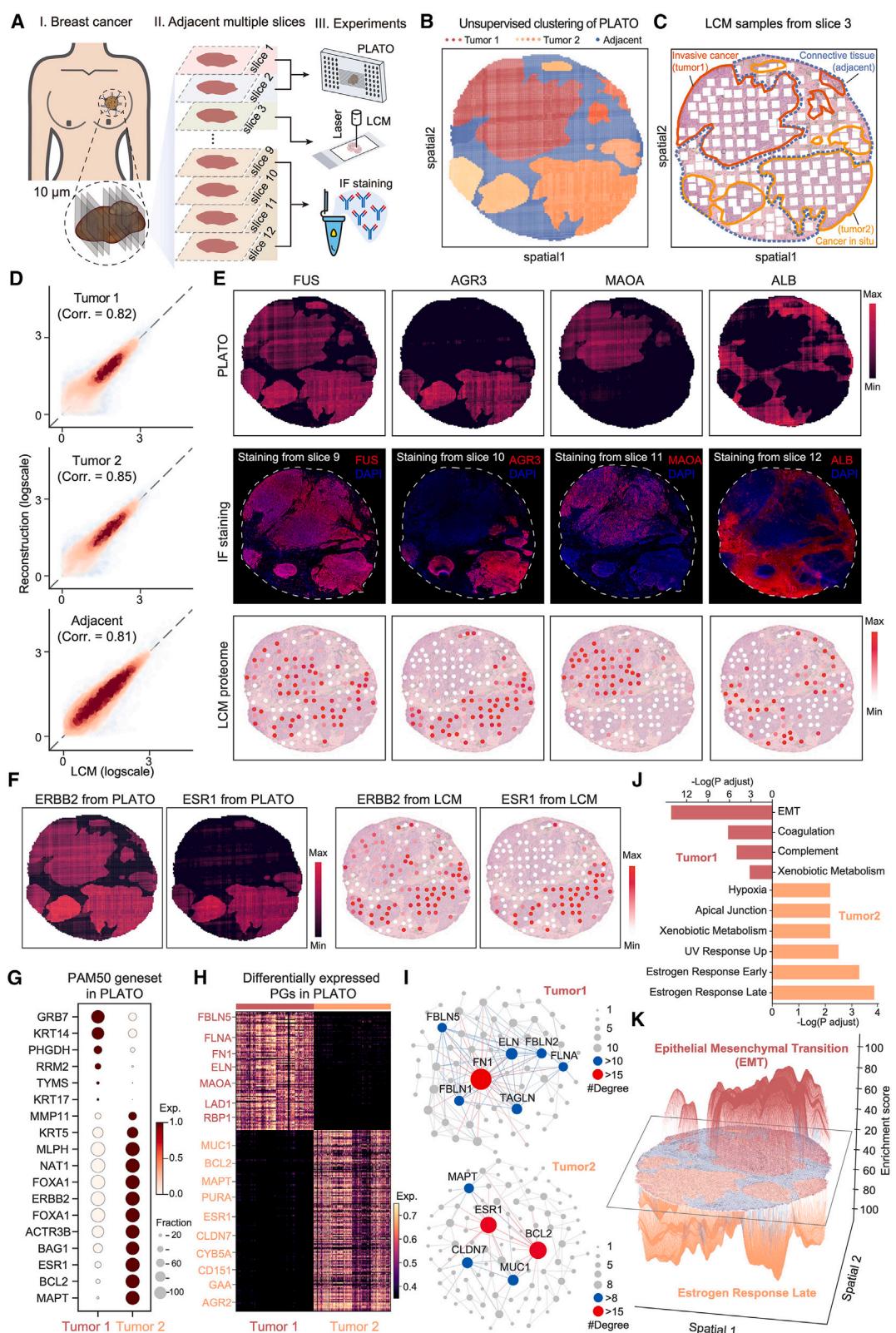


Figure 5. Spatial proteomics mapping of villi

- (A) Structure of intestinal villus.
- (B) Implementation of PLATO in rat intestinal villi. Left, a bright-field image of villus labeled with the microfluidic channels. Right, the number of obtained protein groups for each microchannel.
- (C) Clustering based on spatial reconstructions. Two clustering resolutions are displayed, with higher resolution identifying more clusters. At resolution 0.1, four distinct clusters were detected, while resolution 0.5 identified six clusters. Each cluster is visualized in a different color for clarity. Epi-1, -2, and -3 represent epithelial layers 1, 2, and 3 in the clustering results, respectively.
- (D) IF staining validation of four marker proteins.
- (E) Spatial expression of different transporters.
- (F) Schematic representation of the experiment: colitis was induced by DSS administration in drinking water for 6 days.
- (G) Changes in body weight in each condition.
- (H) Number of upregulated and downregulated genes in each cluster between DSS and naive (left). Both upregulated and downregulated changes represent variations in DSS relative to naive. Heatmap of protein expression in each cluster for each condition. Protein groups with dramatic changes are highlighted.
- (I) Overlap of enriched GO biological processes (BPs) of the three epithelial layers (top). Count of GO BP terms enriched in at least two layers of the epithelium (bottom).
- (J) Expression of transporters (up) and representative protein of SLC16A1 in villi.

See also [Figure S7](#).



(legend on next page)

marker of smooth muscle cells, was exclusively expressed in the muscularis. We also identified transporters closely related to nutrient absorption of villi in different epithelial layers (Figure 5E). Interestingly, some transporters exhibited a region-specific expression pattern. For example, transporters correlated with ions, carbohydrates, and peptides showed greater enrichment in the outer epithelial layer, whereas those related to proteins and carbohydrate derivatives were prevalently expressed in all three epithelial layers.

We next sought to unveil the temporal proteomic dynamics of the large intestine during inflammation by taking advantage of the widely used dextran sodium sulfate (DSS)-induced model of colitis. In brief, we treated wild-type rats with DSS in drinking water for 6 days (Figure 5F), during which DSS exposure resulted in continuous body weight loss (Figure 5G) as well as diarrhea, rectal bleeding (Figure S7B), and shortened colon and intestine length (Figure S7C). Histological analysis confirmed epithelial damage, such as desquamation of the epithelial layer on day 6 (Figure S7D). We then performed PLATO on both naive and DSS-treated villi (Figures S7E and S7F; Table S1) and computed differentially expressed protein groups (DEPs) for each cluster. As shown in Figure 5H, epithelial layers contained the largest number of DEPs, most of which were significantly downregulated during inflammation; one example is CD59, a well-known epithelial marker of ulcerative colitis and Crohn's disease.³⁸ Further, Gene Ontology (GO) enrichment analysis of these DEPs showed that the outer epithelial layer experienced the most dysregulation of BPs after DSS treatment, compared with that of the inner layer (Figures 5I and S7G). We then counted the GO BP terms enriched in at least two layers of the epithelium and found that certain terms related to localization and transport were the most common. For example, transporters of organic solutes (choline transporter-like protein 1 [SLC44A1]) and inorganic ions (chloride anion exchanger [SLC26A3]) in the large intestine appeared dysfunctional after DSS treatment (Figure S7H). The spatial distribution of transporter proteins was altered after DSS treatment (Figure 5J). Collectively, these results demonstrated that PLATO can faithfully capture DSS-induced spatial proteomic changes at a pixel size of 25 μm .

PLATO-based spatial proteomics of human breast cancer

To demonstrate the applicability of PLATO in clinical research, we applied it to the spatial analysis of human breast cancer tissues. Breast cancer, one of the most prevalent cancers in women, presents significant public health challenges³⁹ because of its marked heterogeneity, which leads to varied clinical outcomes. Subtype-specific treatment strategies are therefore crucial.⁴⁰ PLATO shows great potential in deciphering this heterogeneity in both fresh frozen (Figure 6) and formalin-fixed paraffin-embedded (Figure S8) samples.

We profiled the tumor's spatial proteome using a fresh frozen breast cancer sample from a 77-year-old female patient diagnosed with HER2+, ER 70%, PR-. Using a microfluidic chip with 70 parallel channels, each 25 μm wide, PLATO identified approximately 4,000 protein groups per channel from two angles (Figures 6A and S9A). Clustering analysis revealed three distinct spatial clusters: two tumor regions and one adjacent region (Figure 6B), which aligned well with pathologist annotations (Figure S9B). To validate PLATO, we performed LCM-proteome analysis on an adjacent tissue slice, collecting 145 tissue voxels ($\sim 100 \times 100 \times 10 \mu\text{m}$) from the identified clusters (Figures 6C, S9C, and S9D). On average, 3,500 proteins per voxel were found in tumor regions, significantly more than in adjacent areas (Figure S9E). A high Spearman correlation was observed between PLATO and LCM results across all regions (tumor 1: 0.82, tumor 2: 0.85, adjacent: 0.81) (Figure 6D), and similarly strong correlations were found at region boundaries (adjacent and tumor 1: 0.80, adjacent and tumor 2: 0.81) (Figure S9F). Pixel-level comparisons further confirmed PLATO's accuracy (Figure S9G). IF staining also validated the spatial distribution of four representative proteins in PLATO (Figure 6E).

Next, we examined the spatial distribution of breast cancer classification markers, including HER2/ERBB2, ER/ESR1, and PR,³⁹ in the PLATO results. Two tumor subtypes were identified: tumor 1 (HER2+, ER-, PR-) and tumor 2 (HER2+, ER+, PR-) (Figure 6F). These findings were validated through LCM-proteome analysis (Figure 6F), showing greater precision than the clinical diagnosis (HER2+, ER 70%, PR-), as PLATO revealed that the 70% ER expression was a composite of the two subtypes.

Figure 6. Spatial proteomics of frozen breast cancer tissue

- (A) Schematic representation of the experiment: breast cancer tissue slices were subjected to PLATO, laser capture microdissection (LCM)-based proteome, and immunofluorescence (IF) staining, respectively.
- (B) Unsupervised clustering of PLATO spots based on protein abundance.
- (C) Schematic diagram of microdissection points. White square boxes represent the LCM sampling area. Colored lines represent the pathological annotations.
- (D) Regional concordance between LCM-based proteome and PLATO reconstructions. Each anatomical region is represented by all samples from the region. Spearman's coefficient was calculated. The protein abundance was log transformed and normalized using the normalizeBetweenArrays function from the limma package.
- (E) Spatial distribution of representative proteins by PLATO, IF staining, and LCM proteome.
- (F) Left, the spatial distribution of epithelial growth factor receptor 2 (ERBB2) and estrogen receptor 1 (ESR1) by PLATO. Right, the spatial distribution of ERBB2 and ESR1 by LCM proteome.
- (G) Dot plot showing the protein expression levels of PAM50 gene set in two tumor subtypes.
- (H) Differentially expressed protein groups between the two tumor subtypes.
- (I) Network analysis of differential protein interactions in the two tumor subtypes. Each node represents a protein, and the node size represents the degree of proteins in the network.
- (J) MSigDB hallmark gene set enrichment based on the differentially expressed proteins across the two tumor subtypes. Pathway significance was determined using Fisher's exact test, with p values adjusted for false discovery rate (FDR).
- (K) Spatial distribution of epithelial mesenchymal transition (EMT) and estrogen response late gene set enrichment score, which was visualized by RidgeSpace package. Scores reflect the average expression levels of enriched DEPs within each gene set.
- See also Figures S8 and S9.

Additionally, proteins associated with the PAM50 subtypes were differentially expressed between tumor 1 and tumor 2, confirming their distinct characteristics^{41,42} (Figure 6G). The PLATO reconstruction of ERBB2 provided a more comprehensive view than the LCM results, suggesting a potential heterogeneity within tumor 1. We then explored the DEPs between these subtypes to understand the tumor microenvironment's heterogeneity (Figure 6H). Tumor 1 was enriched in extracellular matrix (ECM) proteins, such as fibulin-5 (FBLN5), filamin-A (FLNA), fibronectin (FN1), elastin (ELN), and ladinin-1 (LAD1), suggesting a more aggressive tumor phenotype.⁴³ In contrast, tumor 2 was enriched in hormone signaling-related proteins like estrogen receptor 1 (ESR1), a clinical biomarker for breast cancer, and mucin 1 (MUC1), which stabilizes and activates estrogen receptor,⁴⁴ highlighting its distinct biology and relationship with estrogen signaling.

To further investigate the functional roles of these DEPs, we constructed protein-protein interaction networks (Figure 6I). FN1, which was reported to play a crucial role in the mammary mesenchymal compartment during breast cancer development,⁴⁵ was identified as a key protein in tumor 1, while ESR1 and BCL2 were central in tumor 2, the latter being linked to estrogen receptor-positive tumors.⁴⁶ Pathway enrichment analysis revealed epithelial-to-mesenchymal transition (EMT) activity in tumor 1, suggesting a more aggressive phenotype, while estrogen response pathways were predominant in tumor 2 (Figures 6J and 6K). Both tumors exhibited enrichment in xenobiotic metabolism pathways, potentially contributing to drug resistance in breast cancer.⁴⁷ Collectively, PLATO effectively identifies spatially distinct tumor subtypes, highlights key dysregulated proteins, and provides insights into the complexity of the tumor microenvironment.

DISCUSSION

With the development of LCM-based methods and the improvement of sensitivity of MS, several remarkable methodologies have been developed that allow for spatially resolved proteomic profiling of tissue sections.^{13,15} Although these techniques exhibit specific advantages in resolution and applicability, they share a major caveat in the limited tissue areas and sampling throughput. For example, a recent study developed the nano-POTS approach,¹³ which successfully obtained >2,000 proteins with 100 μm spatial resolution from mouse uterine tissue, but it was only applicable to an extremely small tissue region containing 24 data points.¹³ Likewise, the DVP¹⁵ strategy employs an artificial intelligence-driven LMD technique to perform spatially resolved proteomics profiling at single-cell or subcellular levels. However, it is difficult to construct protein maps on the whole-tissue level. The MASP¹⁷ method successfully improves the detection scale to whole-tissue level, based on spatially resolved micro-compartmentalization of tissue using a 3D-printed micro-scaffold, but it requires a large number of measurements and is of low resolution. For example, 900 LC-MS/MS measurements for mouse brain tissue at a spatial resolution of 400 μm are needed. Overall, without mention of sample degradation and batch effects during large-scale experiments, it is evident that these methods are prohibitively time consuming, labor intensive,

and expensive. In contrast to previous methods, our approach is a fundamentally different method compatible with the easily obtainable LC-MS/MS measurements and can be set up with high throughput and a large field of view. For example, in this study, we obtained approximately 3,000 protein groups from adult mouse cerebellum, with only ~60 MS measurements. In addition to its independence from specialized laboratory facilities, the PLATO framework shows several significant advantages, including the following: (1) it is versatile and easy to operate with a simple PDMS slab clamped on the tissue slide, without sophisticated fluidic handling, and thus can be readily adopted by researchers who have no training in microfluidics; (2) high spatial resolution (e.g., 25 μm) can be achieved by requiring only a few parallel sampling projections; and (3) whole-tissue mapping capacity (e.g., several thousand spots) enables a comprehensive survey of region-to-region variations. Therefore, we believe that this method will remarkably facilitate efforts toward understanding the spatially organized translational regulation responsible for biological mechanisms.

An important computational challenge of the PLATO framework is how to recover spatial patterns from parallel-flow projections. State-of-the-art methods to address this challenge are based on computer tomography,^{19,20} which takes the measurements obtained from each microchannel as an analog of a parallel-beam sum projection. Although advances in this approach significantly reduce the number of measurements by one order of magnitude, compared with the spot-by-spot sampling approach,¹⁸ at least five slices are needed to implement the protocol, which ignores the heterogeneity of different slices and requires a large number of measurements (e.g., ~700 strips for ST reconstruction of mouse brain). More critically, these methods showed low resolution and an inability to discover discontinuous, checkerboard-like patterns.^{18,48} In contrast, the PLATO framework utilizes a transfer learning-based algorithm to address this challenge. By training the model using images that provide spatial patterns of the middle slice, we build connections between microchannels and their spatial expression. The trained model restores the spatial patterns of detected proteins, based on the measurements from microfluidic chip experiments. By employing this strategy, Flow2Spatial remarkably reduces the number of slices to three and measurements to less than 100, and this number is at least one order of magnitude less than that of previous approaches. Advantages resulting from the reduction of MS measurements include (1) reducing heterogeneity introduced by consecutive slices, especially for complex tissues such as the brain, and (2) saving time and costs for sample preparation and measurement; in particular, the latter is essential for spatial proteomics in situations in which no efficient protein barcoding technology is available. We believe this method will pave the way for achieving spatial resolution in a wide range of molecular detection techniques.

Limitations of the study

While PLATO represents a pioneering advancement in whole-tissue proteomic imaging that offers remarkable proteome depth, several areas for improvement remain. First, although the current version utilizes microfluidic channels as narrow as 25 μm, it does not yet achieve single-cell resolution. Enhancing PLATO's

resolution could be accomplished by using higher-precision microfluidic channels, potentially as narrow as 10 μm , which would bring it closer to single-cell imaging. Alternatively, integrating deconvolution algorithms that align cell-type information from spatial transcriptomics could help resolve cellular composition at each spot. Second, further applications of PLATO would benefit from incorporating faster and more reliable MS. A more rapid instrument could enable PLATO to serve as a routine clinical diagnostic tool, capable of delivering results in less than 24 h. Improving the instrument's reliability would reduce random fluctuations in the MS signal, enhancing the overall accuracy of the measurements. Third, while the current PLATO framework focuses on protein profiling, it does not yet capture protein PTMs. Given LC-MS/MS's potential for PTM detection, PLATO could be expanded to include PTM analysis by incorporating additional enrichment steps, such as affinity capture or covalent coupling,^{49,50} into the microfluidic workflow.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Fangqing Zhao (zhfq@ioz.ac.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the iProX partner repository^{51,52} with the dataset identifier PXD045687. The MAGIC-seq sequencing data of mouse cerebellum are available from the Genome Sequence Archive (GSA) in the National Genomics Data Center, Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences under the accession code PRJCA032225. In detail, the raw sequencing data have been deposited under the accession code CRA020331. The Slide-seq dataset of mouse cerebellum was obtained from https://singlecell.broadinstitute.org/single_cell/study/SCP815/.³⁷ The source code for Flow2Spatial is available at <https://github.com/bioinfo-biols/Flow2Spatial>.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (32025009, 32130020, 32400533, and 32300538) and National Key R&D Project (2022YFA1303900, 2021YFA1300500, 2022YFC2703200, and 2024YFF1206600), Natural Science Foundation of Beijing (Z230007), Beijing Hospitals Authority Clinical Medicine Development of Special Funding Support (ZLRK202327), and the Postdoctoral Innovation Talent Support Program (BX20240367). We would like to thank Dr. Wenbin Du (Institute of Microbiology, Chinese Academy of Sciences) for valuable advice on designing microfluidic device and for sharing the soft lithography facilities.

AUTHOR CONTRIBUTIONS

F.Z. and P.J. conceived the study. F.Z., B.H., and P.J. designed the microfluidic chip-based experimental workflow. R.H. and F.Z. designed the Flow2Spatial algorithm. B.H., G.W., N.W., T.L., J.Z., Z.J., Y.H., and Z.Z. performed the experiments and generated the data. K.P., J.Z., P.J., and F.Z. analyzed the data. P.J., B.H., R.H., and F.Z. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Animals
 - Human samples
- **METHOD DETAILS**
 - Tissue handling
 - Microfluidic device fabrication
 - Proteomics sample preparation of PLATO
 - Spatial transcriptomics with MAGIC-seq
 - Proteomics of laser capture microdissection
 - Imaging mass spectrometry of metabolites
 - Immunofluorescence staining
 - LC-MS/MS
 - Optimization of chip-based proteomics
 - Cross-contamination between microchannels
 - PLATO for mouse cerebellum
 - PLATO for intestinal villus
 - PLATO for human breast cancer
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Overview of the Flow2Spatial algorithm
 - Training dataset generator in Flow2Spatial
 - Deep learning model of Flow2Spatial
 - Proteomic data processing
 - Filtering and normalization of the proteomic data
 - MALDI-imaging MS data analysis
 - Processing of MAGIC-seq raw data
 - Filtering and normalization of ST data
 - Dimension reduction, clustering, and marker identification
 - Enrichment analysis
- **ADDITIONAL RESOURCES**
 - Abbreviations

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2024.12.023>.

Received: June 27, 2024

Revised: October 12, 2024

Accepted: December 17, 2024

Published: January 23, 2025

REFERENCES

1. Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* 173, 1307. <https://doi.org/10.1016/j.cell.2018.05.012>.
2. Liu, Y., Yang, M., Deng, Y., Su, G., Enninfel, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020). High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* 183, 1665–1681.e18. <https://doi.org/10.1016/j.cell.2020.10.026>.
3. Rodrigues, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vandenburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363, 1463–1467. <https://doi.org/10.1126/science.aaw1219>.
4. Eng, C.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.C., and Cai, L. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* 568, 235–239. <https://doi.org/10.1038/s41586-019-1049-y>.
5. Liu, Y., DiStasio, M., Su, G., Asashima, H., Enninfel, A., Qin, X., Deng, Y., Nam, J., Gao, F., Bordignon, P., et al. (2023). High-plex protein and whole

- transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat. Biotechnol.* 41, 1405–1409. <https://doi.org/10.1038/s41587-023-01676-0>.
6. Hickey, J.W., Neumann, E.K., Radtke, A.J., Camarillo, J.M., Beuschel, R.T., Albanese, A., McDonough, E., Hatler, J., Wiblin, A.E., Fisher, J., et al. (2022). Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. *Nat. Methods* 19, 284–295. <https://doi.org/10.1038/s41592-021-01316-y>.
 7. Giesen, C., Wang, H.A.O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P.J., Grolimund, D., Buhmann, J.M., Brandt, S., et al. (2014). Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* 11, 417–422. <https://doi.org/10.1038/nmeth.2869>.
 8. Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* 178, 473–490.e26. <https://doi.org/10.1016/j.cell.2019.05.027>.
 9. Tan, W.C.C., Nerurkar, S.N., Cai, H.Y., Ng, H.H.M., Wu, D., Wee, Y.T.F., Lim, J.C.T., Yeong, J., and Lim, T.K.H. (2020). Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Commun. (Lond.)* 40, 135–153. <https://doi.org/10.1002/cac2.12023>.
 10. Slavov, N. (2021). Increasing proteomics throughput. *Nat. Biotechnol.* 39, 809–810. <https://doi.org/10.1038/s41587-021-00881-z>.
 11. Petelski, A.A., Emmott, E., Leduc, A., Huffman, R.G., Specht, H., Perlman, D.H., and Slavov, N. (2021). Multiplexed single-cell proteomics using SCope2. *Nat. Protoc.* 16, 5398–5425. <https://doi.org/10.1038/s41596-021-00616-z>.
 12. Brunner, A.D., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., Hoerning, O.B., Bache, N., Apalategui, A., Lubeck, M., et al. (2022). Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* 18, e10798. <https://doi.org/10.1525/msb.202110798>.
 13. Piehowski, P.D., Zhu, Y., Bramer, L.M., Stratton, K.G., Zhao, R., Orton, D.J., Moore, R.J., Yuan, J., Mitchell, H.D., Gao, Y., et al. (2020). Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100-μm spatial resolution. *Nat. Commun.* 11, 8. <https://doi.org/10.1038/s41467-019-13858-z>.
 14. Xu, R., Tang, J., Deng, Q., He, W., Sun, X., Xia, L., Cheng, Z., He, L., You, S., Hu, J., et al. (2018). Spatial-Resolution Cell Type Proteome Profiling of Cancer Tissue by Fully Integrated Proteomics Technology. *Anal. Chem.* 90, 5879–5886. <https://doi.org/10.1021/acs.analchem.8b00596>.
 15. Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovács, F., Brunner, A.D., Migh, E., Schweizer, L., Santos, A., Bzorek, M., et al. (2022). Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* 40, 1231–1240. <https://doi.org/10.1038/s41587-022-01302-5>.
 16. Guo, G., Papanicolaou, M., Demarais, N.J., Wang, Z., Schey, K.L., Timpton, P., Cox, T.R., and Grey, A.C. (2021). Automated annotation and visualisation of high-resolution spatial proteomic mass spectrometry imaging data using HIT-MAP. *Nat. Commun.* 12, 3241. <https://doi.org/10.1038/s41467-021-23461-w>.
 17. Ma, M., Huo, S., Zhang, M., Qian, S., Zhu, X., Pu, J., Rasam, S., Xue, C., Shen, S., An, B., et al. (2022). In-depth mapping of protein localizations in whole tissue by micro-scaffold assisted spatial proteomics (MASP). *Nat. Commun.* 13, 7736. <https://doi.org/10.1038/s41467-022-35367-2>.
 18. Schede, H.H., Schneider, C.G., Stergiadou, J., Borm, L.E., Ranjak, A., Yamawaki, T.M., David, F.P.A., Lönnberg, P., Tosches, M.A., Codeluppi, S., and La Manno, G. (2021). Spatial tissue profiling by imaging-free molecular tomography. *Nat. Biotechnol.* 39, 968–977. <https://doi.org/10.1038/s41587-021-00879-7>.
 19. Kruse, F., Junker, J.P., van Oudenaarden, A., and Bakkers, J. (2016). Tomo-seq: A method to obtain genome-wide expression data with spatial resolution. *Methods Cell Biol.* 135, 299–307. <https://doi.org/10.1016/bs.mcb.2016.01.006>.
 20. Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., and van Oudenaarden, A. (2014). Genome-wide RNA Tomography in the zebrafish embryo. *Cell* 159, 662–675. <https://doi.org/10.1016/j.cell.2014.09.038>.
 21. Brown, K.A., Chen, B., Guardado-Alvarez, T.M., Lin, Z., Hwang, L., Ayaz-Guner, S., Jin, S., and Ge, Y. (2019). A photocleavable surfactant for top-down proteomics. *Nat. Methods* 16, 417–420. <https://doi.org/10.1038/s41592-019-0391-1>.
 22. Laganowsky, A., Reading, E., Hopper, J.T.S., and Robinson, C.V. (2013). Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* 8, 639–651. <https://doi.org/10.1038/nprot.2013.024>.
 23. Candes, E.J., and Tao, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory* 52, 5406–5425. <https://doi.org/10.1109/TIT.2006.885507>.
 24. Vicari, M., Mirzazadeh, R., Nilsson, A., Shariatgorji, R., Bjärterot, P., Larsson, L., Lee, H., Nilsson, M., Foyer, J., Ekwall, M., et al. (2024). Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nat. Biotechnol.* 42, 1046–1050. <https://doi.org/10.1038/s41587-023-01937-y>.
 25. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
 26. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chatopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14, 865–868. <https://doi.org/10.1038/nmeth.4380>.
 27. Zeng, H., Huang, J., Ren, J., Wang, C.K., Tang, Z., Zhou, H., Zhou, Y., Shi, H., Aditham, A., Sui, X., et al. (2023). Spatially resolved single-cell translationalomics at molecular resolution. *Science* 380, eadd3067. <https://doi.org/10.1126/science.add3067>.
 28. Zhao, T., Chiang, Z.D., Morris, J.W., LaFave, L.M., Murray, E.M., Del Priore, I., Meli, K., Lareau, C.A., Nadaf, N.M., Li, J., et al. (2022). Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature* 601, 85–91. <https://doi.org/10.1038/s41586-021-04217-4>.
 29. Ben-Chetrit, N., Niu, X., Swett, A.D., Sotelo, J., Jiao, M.S., Stewart, C.M., Potenski, C., Mielinis, P., Roelli, P., Stoeckius, M., and Landau, D.A. (2023). Integration of whole transcriptome spatial profiling with protein markers. *Nat. Biotechnol.* 41, 788–793. <https://doi.org/10.1038/s41587-022-01536-3>.
 30. De Sio, G., Smith, A.J., Galli, M., Garancini, M., Chinello, C., Bono, F., Pagni, F., and Magni, F. (2015). A MALDI-Mass Spectrometry Imaging method applicable to different formalin-fixed paraffin-embedded human tissues. *Mol. Biosyst.* 11, 1507–1514. <https://doi.org/10.1039/C4MB00716F>.
 31. Fremond, S., Andani, S., Barkey Wolf, J., Dijkstra, J., Melsbach, S., Jobson, J.J., Brinkhuis, M., Roothaan, S., Jurgenliemk-Schulz, I., Lutgens, L.C.H.W., et al. (2023). Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts. *Lancet Digit. Health* 5, e71–e82. [https://doi.org/10.1016/S2589-7500\(22\)00210-2](https://doi.org/10.1016/S2589-7500(22)00210-2).
 32. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F.K., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al. (2024). Towards a general-purpose foundation model for computational pathology. *Nat. Med.* 30, 850–862. <https://doi.org/10.1038/s41591-024-02857-3>.
 33. Ahn, B., Moon, D., Kim, H.-S., Lee, C., Cho, N.H., Choi, H.-K., Kim, D., Lee, J.-Y., Nam, E.J., Won, D., et al. (2024). Histopathologic image-based deep learning classifier for predicting platinum-based treatment responses in high-grade serous ovarian cancer. *Nat. Commun.* 15, 4253. <https://doi.org/10.1038/s41467-024-48667-6>.
 34. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., and Rajpoot, N. (2019). Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* 58, 101563. <https://doi.org/10.1016/j.media.2019.101563>.

35. He, B., Bergenstråhlé, L., Stenbeck, L., Abid, A., Andersson, A., Borg, Å., Maaskola, J., Lundberg, J., and Zou, J. (2020). Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat. Biomed. Eng.* 4, 827–834. <https://doi.org/10.1038/s41551-020-0578-x>.
36. He, K.M., Zhang, X.Y., Ren, S.Q., and Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
37. Cable, D.M., Murray, E., Zou, L.S., Goeva, A., Macosko, E.Z., Chen, F., and Izarri, R.A. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* 40, 517–526. <https://doi.org/10.1038/s41587-021-00830-w>.
38. Scheinin, T., Böhling, T., Halme, L., Kontiainen, S., Bjørge, L., and Meri, S. (1999). Decreased expression of protectin (CD59) in gut epithelium in ulcerative colitis and Crohn's disease. *Hum. Pathol.* 30, 1427–1430. [https://doi.org/10.1016/s0046-8177\(99\)90163-6](https://doi.org/10.1016/s0046-8177(99)90163-6).
39. Nolan, E., Lindeman, G.J., and Visvader, J.E. (2023). Deciphering breast cancer: from biology to the clinic. *Cell* 186, 1708–1728. <https://doi.org/10.1016/j.cell.2023.01.040>.
40. Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J., and Cardoso, F. (2019). Breast cancer. *Nat. Rev. Dis. Primers* 5, 66. <https://doi.org/10.1038/s41572-019-0111-2>.
41. Asleh, K., Negri, G.L., Spencer Miko, S.E., Colborne, S., Hughes, C.S., Wang, X.Q., Gao, D., Gilks, C.B., Chia, S.K.L., Nielsen, T.O., et al. (2022). Proteomic analysis of archival breast cancer clinical specimens identifies biological subtypes with distinct survival outcomes. *Nat. Commun.* 13, 896. <https://doi.org/10.1038/s41467-022-28524-0>.
42. Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J. Clin. Oncol.* 27, 1160–1167. <https://doi.org/10.1200/JCO.2008.18.1370>.
43. Zhang, J., Hu, Z., Horta, C.A., and Yang, J. (2023). Regulation of epithelial-mesenchymal transition by tumor microenvironmental signals and its implication in cancer therapeutics. *Semin. Cancer Biol.* 88, 46–66. <https://doi.org/10.1016/j.semcan.2022.12.002>.
44. Wei, X., Xu, H., and Kufe, D. (2006). MUC1 Oncoprotein Stabilizes and Activates Estrogen Receptor alpha. *Mol. Cell* 21, 295–305. <https://doi.org/10.1016/j.molcel.2005.11.030>.
45. Li, C.L., Yang, D., Cao, X., Wang, F., Hong, D.Y., Wang, J., Shen, X.C., and Chen, Y. (2017). Fibronectin induces epithelial-mesenchymal transition in human breast cancer MCF-7 cells via activation of calpain. *Oncol. Lett.* 13, 3889–3895. <https://doi.org/10.3892/ol.2017.5896>.
46. Dawson, S.J., Makretsov, N., Blows, F.M., Driver, K.E., Provenzano, E., Le Quesne, J., Baglietto, L., Severi, G., Giles, G.G., McLean, C.A., et al. (2010). BCL2 in breast cancer: a favourable prognostic marker across molecular subtypes and independent of adjuvant therapy received. *Br. J. Cancer* 103, 668–675. <https://doi.org/10.1038/sj.bjc.6605736>.
47. Murray, G.I., Weaver, R.J., Paterson, P.J., Ewen, S.W.B., Melvin, W.T., and Burke, M.D. (1993). Expression of xenobiotic metabolizing enzymes in breast cancer. *J. Pathol.* 169, 347–353. <https://doi.org/10.1002/path.1711690312>.
48. Codeluppi, S., Borm, L.E., Zeisel, A., La Manno, G., van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* 15, 932–935. <https://doi.org/10.1038/s41592-018-0175-z>.
49. Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P., and Jørgensen, T.J.D. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* 4, 873–886. <https://doi.org/10.1074/mcp.T500007-MCP200>.
50. Witze, E.S., Old, W.M., Resing, K.A., and Ahn, N.G. (2007). Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* 4, 798–806. <https://doi.org/10.1038/nmeth1100>.
51. Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., Li, K., Zhang, G., Jin, Z., He, F., et al. (2019). iProX: an integrated proteome resource. *Nucleic Acids Res.* 47, D1211–D1217. <https://doi.org/10.1093/nar/gky869>.
52. Chen, T., Ma, J., Liu, Y., Chen, Z., Xiao, N., Lu, Y., Fu, Y., Yang, C., Li, M., Wu, S., et al. (2022). iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.* 50, D1522–D1527. <https://doi.org/10.1093/nar/gkab1081>.
53. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* 17, 41–44. <https://doi.org/10.1038/s41592-019-0638-x>.
54. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: open source software for digital pathology image analysis. *Sci. Rep.* 7, 16878. <https://doi.org/10.1038/s41598-017-17204-5>.
55. Fang, Z., Liu, X., and Peltz, G. (2023). GSEAp: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 39, btac757. <https://doi.org/10.1093/bioinformatics/btac757>.
56. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15. <https://doi.org/10.1186/s13059-017-1382-0>.
57. Kaminow, B., Yunusov, D., and Dobin, A. (2021). STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. Preprint at bioRxiv. <https://doi.org/10.1101/2021.05.05.442755>.
58. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* 11, e0163962. <https://doi.org/10.1371/journal.pone.0163962>.
59. Zhu, J., Pang, K., Hu, B., He, R., Wang, N., Jiang, Z., Ji, P., and Zhao, F. (2024). Custom microfluidic chip design enables cost-effective three-dimensional spatiotemporal transcriptomics with a wide field of view. *Nat. Genet.* 56, 2259–2270. <https://doi.org/10.1038/s41588-024-01906-4>.
60. Makhmut, A., Qin, D., Fritzsche, S., Nimo, J., König, J., and Coscia, F. (2023). A framework for ultra-low-input spatial tissue proteomics. *Cell Syst.* 14, 1002–1014.e5. <https://doi.org/10.1016/j.cels.2023.10.003>.
61. Martin, J.C., Bériou, G., and Josien, R. (2016). Dextran Sulfate Sodium (DSS)-Induced Acute Colitis in the Rat. *Methods Mol. Biol.* 1371, 197–203. https://doi.org/10.1007/978-1-4939-3139-2_12.
62. Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L.H., Meng, C., et al. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* 15, e8503. <https://doi.org/10.1525/msb.20188503>.
63. Mund, A., Brunner, A.-D., and Mann, M. (2022). Unbiased spatial proteomics with single-cell resolution in tissues. *Mol. Cell* 82, 2335–2349. <https://doi.org/10.1016/j.molcel.2022.05.022>.
64. Felzenswalb, P.F., and Huttenlocher, D.P. (2004). Efficient graph-based image segmentation. *Int. J. Comput. Vis.* 59, 167–181. <https://doi.org/10.1023/B:VISI.0000022288.19776.77>.
65. Diamond, S., and Boyd, S. (2016). CVXPY: A Python-Embedded Modeling Language for Convex Optimization. *J. Mach. Learn. Res.* 17, 83.
66. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* 9, 5233. <https://doi.org/10.1038/s41598-019-41695-z>.
67. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008, 10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--|--|-----------------------------------|
| Antibodies | | |
| Anti-MBP rabbit monoclonal antibody | ABclonal Technology | Cat#A11162; RRID: AB_2861512 |
| Anti-MOG rabbit monoclonal antibody | ABclonal Technology | Cat#A3992; RRID: AB_2863171 |
| Anti-NEFL rabbit monoclonal antibody | ABclonal Technology | Cat#A20269; RRID: AB_3661634 |
| Anti-CNP rabbit monoclonal antibody | ABclonal Technology | Cat#A19033; RRID: AB_2862525 |
| Anti-SYN1 rabbit monoclonal antibody | ABclonal Technology | Cat#A24122; RRID: AB_3661635 |
| Anti-NEUN rabbit monoclonal antibody | ABclonal Technology | Cat#A0951; RRID: AB_2757475 |
| Anti-MAOA rabbit polyclonal antibody | ABclonal Technology | Cat#A14024; RRID: AB_2760879 |
| Anti-ALB rabbit monoclonal antibody | ABclonal Technology | Cat#A24161; RRID: AB_3661636 |
| Anti-CLDN11 rabbit polyclonal antibody | Sangon Biotech | Cat#D261566; RRID: AB_3661751 |
| Anti-DAGLA rabbit polyclonal antibody | Sangon Biotech | Cat#D163668; RRID: AB_3661757 |
| Anti-GAP43 rabbit polyclonal antibody | Sangon Biotech | Cat#D163002; RRID: AB_3661758 |
| Anti-GRM1 rabbit polyclonal antibody | Sangon Biotech | Cat#D260458; RRID: AB_3661759 |
| Anti-SETD7 rabbit polyclonal antibody | Sangon Biotech | Cat#D222485; RRID: AB_3661806 |
| Anti-HMGN2 rabbit polyclonal antibody | Sangon Biotech | Cat#D225187; RRID: AB_3661760 |
| Anti-FUBP1 rabbit polyclonal antibody | Sangon Biotech | Cat#D163947; RRID: AB_3661807 |
| Anti-ZNF207 rabbit polyclonal antibody | Sangon Biotech | Cat#D124006; RRID: AB_3661808 |
| Anti-FUS rabbit polyclonal antibody | Sangon Biotech | Cat#D123360; RRID: AB_3661809 |
| Anti-HSP90B1 rabbit polyclonal antibody | Sangon Biotech | Cat#D220724; RRID: AB_3661810 |
| Anti-TTR rabbit polyclonal antibody | Sangon Biotech | Cat#D120267; RRID: AB_3661811 |
| Anti-CLIC6 rabbit polyclonal antibody | Sangon Biotech | Cat#D263475; RRID: AB_3661812 |
| Anti-ACE rabbit polyclonal antibody | Sangon Biotech | Cat#D260020; RRID: AB_3661813 |
| Anti-PDIA3 rabbit polyclonal antibody | Beijing Solarbio Science & Technology | Cat# K002098P; RRID: AB_3661619 |
| Anti-DCN rabbit polyclonal antibody | Beijing Solarbio Science & Technology | Cat#K002913P; RRID: AB_3661620 |
| Anti-MYH11 rabbit polyclonal antibody | Beijing Solarbio Science & Technology | Cat#K002095P; RRID: AB_3661621 |
| Anti-AGR3 rabbit polyclonal antibody | ProteinTech Group | Cat#11967-1-AP; RRID: AB_2877809 |
| Anti-AGR2 rabbit polyclonal antibody | ProteinTech Group | Cat# 12275-1-AP; RRID: AB_2225096 |
| Anti-ATP1A1 rabbit polyclonal antibody | ProteinTech Group | Cat#14418-1-AP; RRID: AB_2227873 |
| Alexa Fluor 488 anti-rabbit secondary antibody | Invitrogen | Cat#A-21206; RRID: AB_2535792 |
| Alexa Fluor 555 anti-rabbit secondary antibody | Invitrogen | Cat# A-31572; RRID: AB_162543 |
| Biological samples | | |
| C57BL/6 Adult mouse | Beijing Vitalstar Biotechnology | N/A |
| Sprague–Dawley male rats | Beijing Vital River Laboratory Animal Technology | N/A |
| Human breast cancer | Peking University Cancer Hospital | N/A |
| Chemicals, peptides, and recombinant proteins | | |
| 4-hexylphenylazosulfonate | Sigma-Aldrich | Cat#919233-100ML |
| n-dodecl-β-D-maltoside | Sigma-Aldrich | Cat#D4641-500MG |
| Iodoacetamide | Sigma-Aldrich | Cat#I1149-5G |
| Dithiothreitol | Sigma-Aldrich | Cat#D0632-1G |
| iRT Standard | Biognosys | Cat#Ki-3002-1 |
| Trypsin Gold | Promega | Cat#V5280 |
| dextran sodium sulfate | MP Biomedicals | Cat#160110 |
| Cryo-Gel | Leica | Cat#39475237 |

(Continued on next page)

Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---------------------------------------|---|
| Maxima H Minus Reverse Transcriptase (200 U/L) | Thermo Fisher Scientific | Cat#EP0751 |
| dNTP mix | New England Biolabs | Cat#N0447L |
| RNase Inhibitor | Enzymatics | Cat#Y9240L |
| Kapa Hotstart HiFi ReadyMix | Kapa Biosystems | Cat#KK2602 |
| Bst 2.0 WarmStart DNA polymerase | New England Biolabs | Cat#M0538L |
| Critical commercial assays | | |
| Qubit dsDNA HS Assay Kit | Invitrogen | Cat#Q33230 |
| NEBNext Ultra II FS DNA Library Prep Kit for Illumina | New England Biolabs | Cat#E7805S |
| H&E staining kit | Beijing Solarbio Science & Technology | Cat#G1120 |
| Micro BCA™ Protein Assay Kit | Thermo Fisher Scientific | Cat#23235 |
| Deposited data | | |
| Mass spectrometry raw data | This paper | http://proteomecentral.proteomexchange.org; PXD045687; |
| Mouse cerebellum spatial transcriptomics data | This paper | GSA: CRA020331 |
| Slide-seq mouse cerebellum data | Cable et al. ³⁷ | https://singlecell.broadinstitute.org/single_cell/study/SCP815/ |
| Software and algorithms | | |
| Spectronaut v.15.2.210819 | Biognosys | https://biognosys.com/software/spectronaut/ |
| Leica Laser Microdissection software v 8.2.0.6739 | Leica Microsystems | https://www.leica-microsystems.com/products/microscope-software/p/leica-lmd-software/ |
| SCiLS Lab v2021b | Bruker Daltonics | https://www.bruker.com/en/products-and-solutions/mass-spectrometry/ms-software/scils-lab.html |
| DIA-NN v1.8.1 | Demichev et al. ⁵³ | https://github.com/vdemichev/DiaNN |
| Qupath v 0.5.1 | Bankhead et al. ⁵⁴ | https://qupath.github.io/ |
| RidgeSpace | N/A | https://github.com/bioinfo-biols/RidgeSpace |
| Gseapy v 0.10.1 | Fang et al. ⁵⁵ | https://gseapy.readthedocs.io/en/latest/index.html |
| Scanpy v1.9.3 | Wolf et al. ⁵⁶ | https://scanpy.readthedocs.io/en/stable/ |
| STAR v2.7.10b | Kaminow et al. ⁵⁷ | https://github.com/alexdobin/STAR/tree/master |
| seqkit v2.0.0 | Shen et al. ⁵⁸ | https://bioinf.shenwei.me/seqkit/ |
| pyimzML | N/A | https://pyimzml.readthedocs.io/en/latest/ |

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animals

All mice used in this work were C57BL/6. Specifically, the wild-type mouse and the expressed red fluorescent protein of tdTOMATO was purchased from Beijing Vitalstar Biotechnology Corporation (Beijing, China). Sprague–Dawley male rats aged 8 weeks were purchased from Beijing Vital River Laboratory Animal Technology Corporation (Beijing, China). All animal experiments were conducted according to the National Institute of Health Guide for the Care and Use of Laboratory Animals. All procedures were approved by the Animal Ethics Committee at the Institute of Zoology, Chinese Academy of Sciences.

Human samples

The breast cancer sample was collected from a 77-year-old female patient of Han Chinese ethnicity, with clinical information obtained through informed consent under a protocol approved by the Ethics Committee of Beijing Cancer Hospital. Approximately 3 mm tissue blocks were excised from the sample using a biopsy punch. Some of these tissue punches were snap frozen in Cryo-Gel (Leica, the Netherlands) as embedding medium and stored at -80 °C, labeled as fresh frozen samples. The remaining punches were processed into formalin-fixed paraffin-embedded (FFPE) blocks. Specifically, the punched tissue was fixed in 10%

formalin for 48 hours, followed by sequential dehydration in methanol and isopropanol. The dehydrated tissues were then transferred into embedding cassettes and infiltrated with paraffin wax to create the FFPE blocks.

METHOD DETAILS

Tissue handling

For the frozen tissue, it was equilibrated to -18 °C in a cryostat (Leica CM3050S) for 30 minutes. Then, the tissue was cut into 10-μm-thick consecutive sections, and three consecutive sections were collected. The first tissue section was placed on the center of the chip glass slide at a predefined angle. The chip glass slide was a custom-designed glass slide with dimensions of 86 × 48 × 1.5 mm. The second tissue section was placed on a poly-L-lysine-coated glass slide (Citotest, China) for histological staining, an indium tin oxide (ITO)-coated glass slide (Bruker Daltonics, USA) for imaging mass spectrometry (IMS), or ST glass slide (MAGIC seq) for spatial transcriptomics.⁵⁹ The third tissue was placed on the center of the chip glass slide at an angle perpendicular to the first tissue section. Finally, bright field full images were obtained for tissues using a spinning disk confocal microscope (Olympus, SpinSR), and tissue sections were stored at -80 °C prior to use.

Microfluidic device fabrication

The microfluidic chip mold was fabricated by CapitalBio Corporation (Beijing, China) using standard soft lithography with SU-8 photoresist. The following types of microfluidic chips were produced: (i) channels 100 μm wide and approximately 100 μm high, (ii) channels 50 μm wide and approximately 60 μm high, and (iii) channels 25 μm wide and approximately 30 μm high. Polydimethylsiloxane (PDMS) (Dow Corning, USA) in a 9:1 (w/w) ratio was poured onto the mold and cured at 65 °C for 5 hours. After curing, the PDMS slab was peeled off, and holes were punched for all inlets and outlets. The inlet and outlet diameters ranged from 2.8 to 3.5 mm, each capable of holding 20 to 40 μL of liquid.

Proteomics sample preparation of PLATO

In this experiment, two adjacent tissue sections were used, oriented perpendicularly to each other. Both sections were processed simultaneously for protein sample preparation to avoid batch effects. Two fabricated PDMS slabs were attached to the tissue section slides, and a homemade acrylic clamp was used to reversibly seal the microfluidic chip. After device assembly, a bright field full image was scanned to record the relative positions of the channels and tissue using a spinning disk confocal microscope (Olympus SpinSR, Japan). Following imaging, 12 μL of lysis buffer was added to each inlet. The lysis buffer contained 25 ng/μL trypsin (Promega, USA), 0.56% iRT standard (Biognosys, Switzerland), and 0.1% n-dodecyl-β-D-maltoside (DDM, Sigma-Aldrich) in 50 mM ammonium bicarbonate (ABC) solution. A global cover connected to a house vacuum was placed over the outlets to draw the buffer from the inlets through the tissue surface to the outlets, maintaining a vacuum pressure of 0.03 MPa for approximately 2 minutes. The chip was then placed in a wet box and incubated at 37 °C for 30 minutes. The lysis buffer was again pulled through the channels using the vacuum pump at 0.03 MPa for another 2 minutes and incubated at 37 °C for an additional 30 minutes. After incubation, all the lysis buffer was collected from the chip outlets and transferred to corresponding Eppendorf tubes. The tubes were then incubated at 37 °C for 4 hours to complete protein digestion. Following digestion, each sample was individually quantified by LC-MS/MS.

Spatial transcriptomics with MAGIC-seq

Spatial transcriptomics was performed using MAGIC-seq as previously described.⁵⁹ Two microfluidic chips (A and B), each with 32 μm-wide channels, were used. The channels of chip A were oriented perpendicularly to those of chip B. Initially, chip A was attached to a carboxyl-modified glass slide and sealed with an acrylic clamp. Barcode A solutions (A1–A70, 40 μM) were prepared in 0.1 M MES buffer (Coolaber, SL33002X) containing N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC; TCI, D1601) and N-hydroxysuccinimide (NHS; Thermo Fisher Scientific, 24500). Each inlet of the chip received 5 μL of Barcode A solution, which was pulled into the channels to react with the carboxyl groups. After the reaction, the glass slide was washed with blocking buffer, 2X SSC buffer, and distilled water. Next, 5 μL of ligation mix was added to the inlets of chip B and pulled into the channels for a 2-hour reaction. The ligation mix consisted of 25 μM barcode B (B1–B70, 20 μM), 5 μM Cy3-labeled linker, 0.2 mg/mL BSA, 20 U/μL T4 DNA ligase (NEB, M0202L). After the reaction, the glass slide was washed and scanned using a spinning disk confocal microscope (Olympus SpinSR, Japan).

After the fabrication of the DNA array, a tissue section was placed onto the array. The section was stained with H&E (Solarbio, G1120) and scanned using the spinning disk confocal microscope (Olympus SpinSR, Japan). Tissue permeabilization was carried out for 12 minutes using 70 μL of permeabilization solution (0.1% pepsin in 0.1 N HCl). Next, a 90-minute reverse transcription (RT) was performed with 70 μL of RT mix, containing 1 mM dNTP (NEB, N0447L), 1× RT buffer, 2.5 μM template switch oligos, 2 U/μL RNase inhibitor (Enzymatics, Y9240L), 0.2 mg/mL BSA, 10 U/μL Maxima H minus reverse transcriptase (Thermo Fisher Scientific, EP0751). Following reverse transcription, the tissue was removed, and second-strand synthesis was conducted. The resulting DNA was amplified by PCR (KAPA Biosystems, KK2602), and approximately 100 ng of DNA was used for sequencing library construction using the NEBNEXT Ultra II FS DNA Library Prep Kit for Illumina (NEB, E7805S). The libraries were sequenced using the Element AVITI platform (Element Biosciences) in PE150 mode.

Proteomics of laser capture microdissection

Tissue section was placed on a membrane-coated glass slide (4.0 μm , PEN-membrane, Leica) for LCM-proteome. Before microdissection, the entire tissue section image was captured using a spinning disk confocal microscope (Olympus SpinSR, Japan) and imported into QuPath version 0.5.1.⁵⁴ Regions of interest (ROIs) were manually created as $\sim 100 \times 100 \mu\text{m}$ square areas, and reference points were set for alignment with laser capture microdissection. The ROIs were exported in GeoJSON format and converted to XML format.⁶⁰ The Leica LMD7 system was then used to excise the $\sim 100 \times 100 \mu\text{m}$ tissue voxels, with the following cutting settings: Power 39, aperture 4, speed 3, head current 75%, and pulse frequency 660. Each dissected voxel was collected into the cap of a 0.5 mL Eppendorf tube. The tubes were centrifuged at 8,000 rcf for 5 minutes to collect the tissue at the bottom. Afterward, 7 μL of 50 mM ABC buffer was added to each tube, followed by sonication for 10 minutes. Trypsin (30 ng/tube) was then added, and digestion was performed at 37 °C for 5 hours. Following digestion, the tissue lysate was analyzed using LC-MS/MS. Samples with significantly fewer detected proteins compared to the average were excluded from further analysis.

Imaging mass spectrometry of metabolites

Tissue sections were placed on an ITO-coated glass slide (Bruker, Germany) for metabolomics MALDI-MSI analysis. The procedure was as follows: (i) The matrix solution was freshly prepared by dissolving 1,5-Diaminonaphthalene (1,5-DAN, 5 mg/mL) in a 70:30 ethanol/water (v/v) mixture. (ii) An automated pneumatic sprayer (HTX Technologies, USA) was used to evenly spray 5 mL of the heated matrix solution onto the tissue sections. (iii) Metabolite imaging was performed using a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker Daltonics, Germany), with sodium trifluoroacetate used for mass spectrometry calibration. Data collection was conducted in negative ion mode with the following MALDI imaging parameters: laser focus set to small, laser power at 85%, raster size of $50 \times 50 \mu\text{m}$, and a collection window ranging from m/z 100 to 1200. The imaging process was managed using FlexImaging 6.0 software (Bruker Daltonics).

Immunofluorescence staining

Immunofluorescence staining was performed on an adjacent slide to generate validation data. The primary antibodies were obtained from several sources, including ABclonal Technology Co., Ltd., with anti-rabbit antibodies against MBP, MOG, NEFL, CNP, SYN1, NEUN, MAOA, and ALB. Additional primary antibodies were sourced from Sangon Biotech (Shanghai) Co., Ltd., targeting CLDN11, DAGLA, GAP43, GRM1, SETD7, HMGN2, FUBP1, ZNF207, FUS, HSP90B1, TTR, and CLIC6. Primary antibodies from Beijing Solarbio Science & Technology Co., Ltd. included PDIA3, DCN, and MYH11. Other antibodies such as AGR3, AGR2 and ATP1A1 were obtained from Proteintech (USA). Visualization was achieved using Alexa Fluor 488 and Alexa Fluor 555 anti-rabbit secondary antibodies (Invitrogen, USA).

For fresh frozen tissue sections, they were fixed in 4% formaldehyde for 20 minutes, followed by three washes with PBS containing 0.3% Triton X-100. The tissue was then blocked with 1% bovine serum albumin (BSA) in PBS for 60 minutes at room temperature (RT). After blocking, the tissue was washed three times with PBS. The primary antibody, diluted in PBS, was applied to the tissue and incubated overnight at 4 °C, followed by three additional washes with PBS. After this, the secondary antibody was added and incubated for 2 hours at RT. The tissue was washed again three times with PBS, with each wash lasting 5 minutes. Imaging was performed using a spinning disk confocal microscope (Olympus SpinSR, Japan).

For formalin-fixed, paraffin-embedded (FFPE) sections, they were deparaffinized with xylene for 2 × 10 minutes. The sections were then rehydrated through a graded ethanol series: 100%, 95%, 90%, 80%, 70%, followed by rinsing in distilled water. Antigen retrieval was performed by boiling the sections in sodium citrate antigen retrieval buffer (pH 6.0) for 10 minutes, followed by natural cooling to room temperature. The sections were blocked with 1% BSA in PBS for 60 minutes at room temperature (RT). After blocking, the sections were washed three times with PBS. The primary antibody, diluted in PBS, was applied to the tissue and incubated overnight at 4 °C, followed by three additional washes with PBS. After this, the secondary antibody was added and incubated for 2 hours at RT. The tissue was washed again three times with PBS, with each wash lasting 5 minutes. Imaging was performed using a spinning disk confocal microscope (Olympus SpinSR, Japan).

LC-MS/MS

The digested samples were directly injected into a C18 analysis column (75 μm I.D. × 20 cm, 1.9 μm , 120 Å, Dr. Maisch GmbH) using an EASY-nLC 1200 system (Thermo Scientific) with 12 μL A solution (0.1% formic acid, FA) at a maximum pressure of 400 bar. The separation gradient was set as follows: 0–2 minutes, 6–12% mobile phase B (80% acetonitrile, 0.1% FA); 2–18 minutes, 12–30% B; 18–22 minutes, 30–42% B solution; and 95% B solution within 4 minutes. A Q Exactive HF mass spectrometer (Thermo Scientific, USA) was used for data-independent acquisition (DIA) analysis. The full scan was set at a resolution of 120,000 at a range of 398 to 1202 m/z; the DIA scan parameter was set with a resolution of 30,000; NCE: 28%; AGC target: 3e6 and maximal injection time: 100 ms. There were three MS1 scans and 27 DIA windows per cycle.

Optimization of chip-based proteomics

In this experiment, a microfluidic chip containing 48 parallel channels with a 100 μm channel width was used. The mouse brain tissue of C57BL/6J was used in this experiment. Three extremely small amounts of protein preparation pipeline were tested, including 4-hexylphenylazosulfonate (AZO, Sigma, USA)-based pipeline, n-dodecyl- β -D-maltoside (DDM, Sigma, USA)-based pipeline, and

the pipeline of direct lysis with trypsin. For the surfactant-based pipeline, 20 μ L lysis solution containing 50 mM ammonium bicarbonate, 10 mM dithiothreitol (DTT, Sigma, USA), and surfactant (0.1% AZO or 0.2% DDM) was added to each inlet of the chip. The lysis solution was further pulled into microchannels, and the chip was placed in a humid chamber at 60 °C for 1 hour to lyse tissue. Then, all lysis solution was pulled to the chip outlet and transferred to a 0.5 mL Eppendorf low-binding tube. Next, 1 μ L iodoacetamide (IAA, Sigma, USA) solution (300 mM in 50 mM ABC) was added to each tube and incubated in the dark for 30 min at room temperature. Then, 1 μ L trypsin (200 ng in 50 mM ABC) was added to each tube and incubated at 37 °C for 4 hours. Next, the AZO surfactant was rapidly degraded by exposing AZO-based samples to UV irradiation. For direct trypsin lysis, 20 μ L trypsin solution (10 ng/ μ L in 50 mM ABC) was added to each chip inlet and further pulled into the microchannels. Then, the chip was kept in a humid chamber at 37 °C for 1 hour to lyse tissue. Next, each lysate in the microchannel was transferred to a low-binding tube for another 4 hours of incubation at 37 °C. Finally, each sample in the Eppendorf tube was quantified by LC–MS/MS separately. Mice expressing red fluorescence protein (tdTOMATO) were used to assess the ability of tissue digestion in microchannels of direct trypsin lysis. Specifically, mouse cerebellum sections were directly digested by trypsin in microchannels, and red fluorescence images of tissue sections were obtained by a confocal microscope before and after tissue lysis.

The following experiment was performed to verify that trypsin adsorbs to the channel when digesting tissue in a microchannel, thereby reducing nonspecific adsorption of tissue proteins. First, a mouse brain section and 200 μ L of 50 mM ABC were collected into a PCR tube. Next, the tissue was lysed by freezing the tube to -80 °C for 5 min and heating it to -90 °C for 10 min. Next, three pretreatments to the microchannels were performed separately, including flowing 20 μ L BSA (10 ng/ μ L in 50 mM ABC), 20 μ L trypsin (10 ng/ μ L in 50 mM ABC), or 20 μ L ABC (50 mM) into the microchannels. After pretreatment, the solution in the microchannels was removed. Next, 20 μ L of tissue lysate was added into each chip inlet and pulled into the pretreated channels for 1 hour. Next, all solutions in the microchannels were collected into low-binding tubes separately and lysed with trypsin. Finally, four groups were analyzed with LC–MS/MS, including BSA, trypsin, ABC-treated microchannels and the original lysate mix.

Cross-contamination between microchannels

The mouse brain tissue of C57BL/6J was used in this experiment. The fluorescent dye propidium iodide (PI, Invitrogen, USA) was pulled into channels and stained tissue at room temperature for 10 minutes. After staining, the tissue was imaged with confocal microscopy. Furthermore, cross-contamination was evaluated by LC–MS/MS. Briefly, we first prepared the *E. coli* lysate mix by repeating freezing, thawing, and sonication (-80 °C 5 min, 90 °C 10 min, sonication 5 min). The protein concentration of the *E. coli* lysate was quantified by a Micro BCA™ Protein Assay Kit (Thermo Scientific, USA). Then, a microfluidics chip (25 μ m or 100 μ m) was placed on a brain tissue slide, and three adjacent microchannels were used. Next, 20 μ L trypsin solution (10 ng/ μ L in 50 mM ABC) was added with 12 ng *E. coli* protein mix flowing into the middle channel to digest tissue, and 20 μ L trypsin solution (10 ng/ μ L in 50 mM ABC) without *E. coli* flowed into the two side channels to digest tissue. After digestion, the tissue lysate was collected from the channels and detected by LC–MS/MS.

PLATO for mouse cerebellum

The cerebellum from C57BL/6J mice was snap-frozen using Cryo-Gel as the embedding medium. Five consecutive 10 μ m sections were cryosectioned for PLATO and related experimental validations: Section 1 and Section 3 were placed on the PLATO chip glass slides with angles of 0° and 90° for PLATO proteomics, respectively. A microfluidic chip with a 50 μ m channel width was used in this experiment and the protein samples were prepared following the PLATO workflow. Section 2 was processed using spatial transcriptomics with a spot size of 32 μ m, as described in prior reports.⁵⁹ Sections 4 was placed on a membrane-coated glass slide (4.0 μ m, PEN-membrane, Leica) for LCM-proteome. Leica LMD7 system was used to excise 100×100 μ m tissue voxels. Section 5 was placed on an ITO-coated glass slide (Bruker, Germany) for metabolomics MALDI-MSI.

PLATO for intestinal villus

The rat colitis model was induced by dextran sodium sulfate (DSS) (MP Biomedicals, USA) using a previously described protocol.⁶¹ Briefly, 5.5% DSS (MW = 36~50 kDa) in water was provided to rats by free drinking for 6~10 days. Meanwhile, a daily record of the body weight and feces was maintained. The rats were sacrificed when blood in feces was observed for the first time. The middle part of the colon tissue was snap frozen using Cryo-Gel as an embedding medium. A microfluidic chip with a 25 μ m channel width was used in this experiment. The colon tissue was sectioned laterally to a thickness of 8 μ m, and three consecutive sections were collected. Sections 1 and 3 were placed separately on the PLATO chip glass slides and further processed according to the PLATO workflow. Section 2 was placed on an ITO-coated glass slide (Bruker, Germany) for MALDI-MSI (matrix-assisted laser desorption ionization-mass spectrometry imaging).

PLATO for human breast cancer

A microfluidic chip with a 25 μ m channel width was used in this experiment. Fresh frozen breast cancer tissue was sectioned to a thickness of 8 μ m, and seven consecutive sections were collected. Sections 1 and 2 were placed separately on chip glass slides and processed according to the PLATO workflow. Sections 3 was placed on a membrane-coated glass slide (4.0 μ m, PEN-membrane, Leica) for LCM-proteome, following the procedure of laser capture microdissection. Sections 9, 10, 11 and 12 were subjected to immunofluorescence staining according to the fresh frozen tissue staining procedure.

For PLATO-FFPE, serial FFPE tissue sections were cut with a microtome at 8 μm and placed on chip glass slides, membrane-coated glass slide and poly-L-lysine-coated glass slide. After overnight air dry at room temperature, these slides were heated at 60°C for 20 minutes to ensure better tissue adhesion. The sections were then deparaffinized with xylol, followed by rehydration in a series of ethanol concentrations: 100%, 95%, 90%, 80%, 70%, and then rinsed in pure water. And section 1 and 2, the cross-links were reversed by heating at 85 °C in antigen retrieval buffer (sodium citrate, pH 6.0) for 20 minutes, followed by natural cooling to room temperature and then rinsing in pure water. Microfluidic chips with a 25 μm channel width were then used, and sections were processed according to the PLATO workflow. After dewax, voxels with volume of 320,000 μm^3 were manually excised from section 3 and collected into the cap of a 0.5 mL Eppendorf tube. Then, the tissue voxel was centrifuged to the bottom of the tube at a speed of 8,000 rcf for 5 min. Next, each tube was added with 12 μL 50 mM ABC buffer and heated at 95 °C for 20 minutes in a thermocycler. Then, trypsin (40 ng/tube) was added, and the sample was digested at 37 °C for 5 hours. After digestion, the tissue lysate was detected by LC-MS/MS. Immunofluorescence staining was performed on the remaining sections according to the FFPE tissue staining procedure.

QUANTIFICATION AND STATISTICAL ANALYSIS

Overview of the Flow2Spatial algorithm

Our approach is grounded in the well-established notion of the strong correlation of spatial molecular patterns across different omics data, as supported by previous studies.^{24–29,33,35} Although mRNA expression is generally weakly associated with its corresponding protein expression,^{12,62,63} Flow2Spatial does not rely on one-to-one expression correlations between mRNA and protein. Instead, it utilizes the distribution features of all mRNAs (clustering features) and orthogonal protein projections to predict protein distributions. Under this premise, the spatial distribution from one type of omics data can be transferred to infer the spatial patterns of other omics data. Briefly, Flow2Spatial trains a model based on the images (signals) of the middle slice, such as hematoxylin-eosin (H&E) staining, fluorescence in situ hybridization (ISH), spatial transcriptomics, or spatial metabolomics. It learns the spatial characteristics of each signal. Based on the trained model that establishes the link between parallel-flow projections and transferred spatial patterns, we can then reconstruct the spatial distribution of the detected proteins.

The Flow2Spatial procedure for reconstructing a 2D spatial matrix from 1D slitting values goes through the following steps: transferring and learning. The former step assumes that the spatial patterns of molecules from different omics are highly correlated and resemble each other (*premise*). We can leverage one easily accessible spatial omics as a reference and transfer its spatial information to help solve a new task, herein spatial proteome reconstruction. Grounded in this assumption, Flow2Spatial reconstructs a 2D spatial matrix with fine structure.

Let us assume the 2D spatial matrix mentioned above as \mathbf{X} . We obtain \mathbf{s} as the parallel-flow projection of \mathbf{X} as follows:

$$\mathbf{s} = \text{Projection}(\mathbf{X}, \alpha)$$

where α is the parameter of slicing channels (lines) in the microfluidics chip. Here, we transfer spatial information from easily accessible omics (\mathbf{X}^t) and mimic the parallel-flow projection with α as follows:

$$\mathbf{s}^t = \text{Projection}(\mathbf{X}^t, \alpha)$$

We thus train a deep learning model (DNN) by using $(\mathbf{s}^t, \mathbf{X}^t)$ as training data and transfer the learned rules to solve the reconstruction problem (from \mathbf{s}^t to \mathbf{X}^t). Based on *Assumption*, the trained DNN model is the solver from \mathbf{s} to reconstruct \mathbf{X} .

Training dataset generator in Flow2Spatial

We built three types of training data generators in this study.

The first is termed the spatial-omics generator. This generator transfers easily accessible omics data from the middle slice in the PLATO framework as training data, including H&E staining, spatial transcriptomics, or spatial metabolomics.

The second one is named the histological information generator, which makes use of histological architecture information. This generator first splits tissue slices into multiple segments based on histological staining or spatial omics clustering results. For staining images, we obtained histological segments based on staining intensity by the Felzenszwalb segmentation algorithm⁶⁴ or manual delineation. Then, the generator assumes that the composition and protein expression levels in spots within a certain segment are highly similar. We thus generate a new spatial dataset \mathbf{X}^n to allow for expression levels that are similar for spots within each segment and uncorrelated between segments as follows:

$$X_p^i = R_j, \text{if } p \in \text{segment}_j$$

$$R_j = \text{Rand}(\cdot), \text{s.t. } R_j \geq 0$$

where X^i is the i-th generated dataset in \mathbf{X}^n , and p is a spot in X^i . R_j is the randomly assigned value for each segment, j represents the j-th segment, and $\text{Rand}(\cdot)$ is a function that generates random values. Euclidean distance is then used to select the spatial

distribution in \mathbf{X}^n that is most similar to the parallel-flow projections of each detected protein. For protein m , we sought to find X^q in \mathbf{X}^n that have the most minimized distance to s_m in \mathbf{s} by calculating $\|Projection(X^q, \alpha) - s_m\|_2^2$.

To further approximate the distribution of spatial proteome, we optimize the sum of the squared differences between the \mathbf{X}^t projection and \mathbf{s} with the package cvxpy.⁶⁵ For protein m , we perform convex optimization to iteratively optimize X^m in \mathbf{X}^t as follows:

$$\min \|X^m - X^q\|_2^2 + \|Projection(X^m, \alpha) - s_m\|_2^2$$

$$\text{s.t. } X^m \geq 0$$

The resulting converged optimal X^m is the spatial distribution provided by this generator, which is then incorporated into the training dataset.

The last generator is a random generator, in which we randomly select two generated spatial distributions from either the first or second generators and randomly add or subtract these chosen distributions as follows:

$$X^{tn} = \frac{X^{ti} + X^{tj}}{2}$$

$$\text{or } X^{tn} = X^{ti} - X^{tj} \text{ where } X^{tn} \geq 0$$

where X^{tn} means the new generated dataset by this random generator. This generator will iterate until the size of the generated data is larger than 40,000. \mathbf{X}^t yielded from the three generators will be used as the training dataset in a later section.

Deep learning model of Flow2Spatial

The architecture of the Flow2Spatial model mimics an autoencoder. In the encoder part, we simulate the parallel-flow projection process to obtain the summation of each pseudostrip from \mathbf{X}^t .

$$\mathbf{s}^t = Encoder(\mathbf{X}^t, \alpha)$$

In the decoder part, we train a deep learning model to reconstruct \mathbf{X}^t based on the values of pseudostrips as follows:

$$\mathbf{X}^t = Decoder(\mathbf{s}^t, \theta)$$

where θ is the trainable weight in the DNN decoder model. Our goal is to minimize the L1-loss between \mathbf{X}^t and \mathbf{X}^t based on gradient descent:

$$Loss = \|\mathbf{X}^t, \mathbf{X}^t\|_1$$

$$\min Loss$$

To reduce model complexity, we normalize these 1D pseudostrip values from different molecules with the same average value:

$$\mathbf{s}^t = \mathbf{s}^{t*} / mean(\mathbf{s}^t)$$

where \mathbf{s}^{t*} represents the raw pseudostrip values and \mathbf{s}^t is the normalized values used for the decoder model.

The architecture of the Flow2Spatial decoder consists of a pretrained Residual Network (ResNet) together with multiple deconvolutional layers. We use ResNet34 to sample spatial structure information from the values of these pseudostrips. After one max pooling layer, we used three layers of transposed convolution to finally yield the reconstructed 2D matrix.

To train the Flow2Spatial model, the whole generated dataset was randomly split into training and testing data at a proportion of 9:1. Flow2Spatial was implemented with PyTorch and Python. In this study, Adam optimization with an initial learning rate of 1e-5 was used to update the parameters. By setting the batch size to 32, we trained the model for 100 epochs on one NVIDIA GPU. The trained Flow2Spatial decoder model was used for real data reconstruction.

Proteomic data processing

Spectronaut v.15.2.210819 software (Biognosys, Schlieren, Switzerland) was used for identification and quantification with the DirectDIA model. The raw files were searched against the UniProt mouse database (17,090 entries), Rattus database (57,342 entries), *E. coli* database (4,595 entries) and iRT peptide sequence using the BGS factory setting. Peptides FDR\ PSMs FDR\ Proteins FDR were all set as 1%, and the best 3-6 fragments were chosen per peptide. QUANT 2.0 (SN Standard) was set for the protein LFQ method, and MS1 was set for the quantity MS level. Based on iRT proteins with consistent theoretical abundance, the abundance

of ontology proteins was corrected and used for reconstruction. For DIA analysis of human breast and mouse cerebellum samples, the raw files were search in DIA-NN v1.8.1 against the generated library of UniProt human database (20,398 entries), UniProt mouse database (17,090 entries) and iRT peptide sequence.⁵³ Additional settings were –original-mods, –ms1-isotope-quant, –report-lib-info, –mass-acc-quant 10.0.

Filtering and normalization of the proteomic data

For Flow2Spatial channel samples, proteins detected in only one angle sample are filtered. Proteins supported by at least 2 peptides were retained. For Flow2Spatial channel samples, proteins detected in only one angle sample are filtered. For LCM samples, proteins present in 30% of the samples are retained and samples were excluded if the number of detected proteins was below 600. Proteins in QC samples with a coefficient of variation below 30% were filtered. The Flow2Spatial data were converted into AnnData objects using Scanpy (v1.9.3).⁵⁶ Abnormal channel data were corrected using neighboring channel information. The protein groups expression data was subjected to logarithmic transformation for dimension reduction, clustering, and marker identification.

MALDI-imaging MS data analysis

Raw data files were loaded into SCiLS Lab software (v2021b, Bruker Daltonics). Feature selection was performed in the dataset by using the feature finding algorithm in SCiLS Lab with default parameters. Features detected in only one spot were filtered. Reduced feature list and spectra quantified by peak area were exported. Feature list and spectra data were then imported in python with pyimzML package and converted into AnnData objects using Scanpy (v1.9.3)⁵⁶ for downstream analysis.

Processing of MAGIC-seq raw data

Spatial barcodes and UMI sequences were extract from read1 of the original sequencing data using the seqkit (v2.0.0) tool.⁵⁸ The STARsolo (v2.7.10b)⁵⁷ was then utilized to process the spatial transcriptomic data by aligning to the mouse reference genome (mm10), obtaining the raw gene expression matrix for spatial transcriptomics. The spatial coordinate sequences were then aligned to the H&E images based on the chip marker coordinates, finally yielding the gene expression matrix with spatial coordinate information for downstream analysis.

Filtering and normalization of ST data

The processed gene expression matrices were converted into AnnData objects using Scanpy (v1.9.3).⁵⁶ Spots with fewer than 1000 genes and genes expressed in fewer than 10 spots were filtered. The filtered gene expression data were then normalized by total counts over all genes so that every spot had 10,000 counts after normalization. The normalized expression matrix was then subjected to logarithmic transformation for the identification of spatially variable genes.

Dimension reduction, clustering, and marker identification

Dimensionality reduction was performed using principal component analysis. A two-dimensional UMAP embedding was constructed from the established top principal components for each spot. Clustering was performed by constructing a shared nearest neighbor (SNN) graph based on the spatially resolved data using established components and clusters identified through the Louvain⁶⁶ or Leiden⁶⁷ algorithm. The marker protein groups for each cluster and differential expression for different groups were determined by the Wilcoxon rank-sum test.

Enrichment analysis

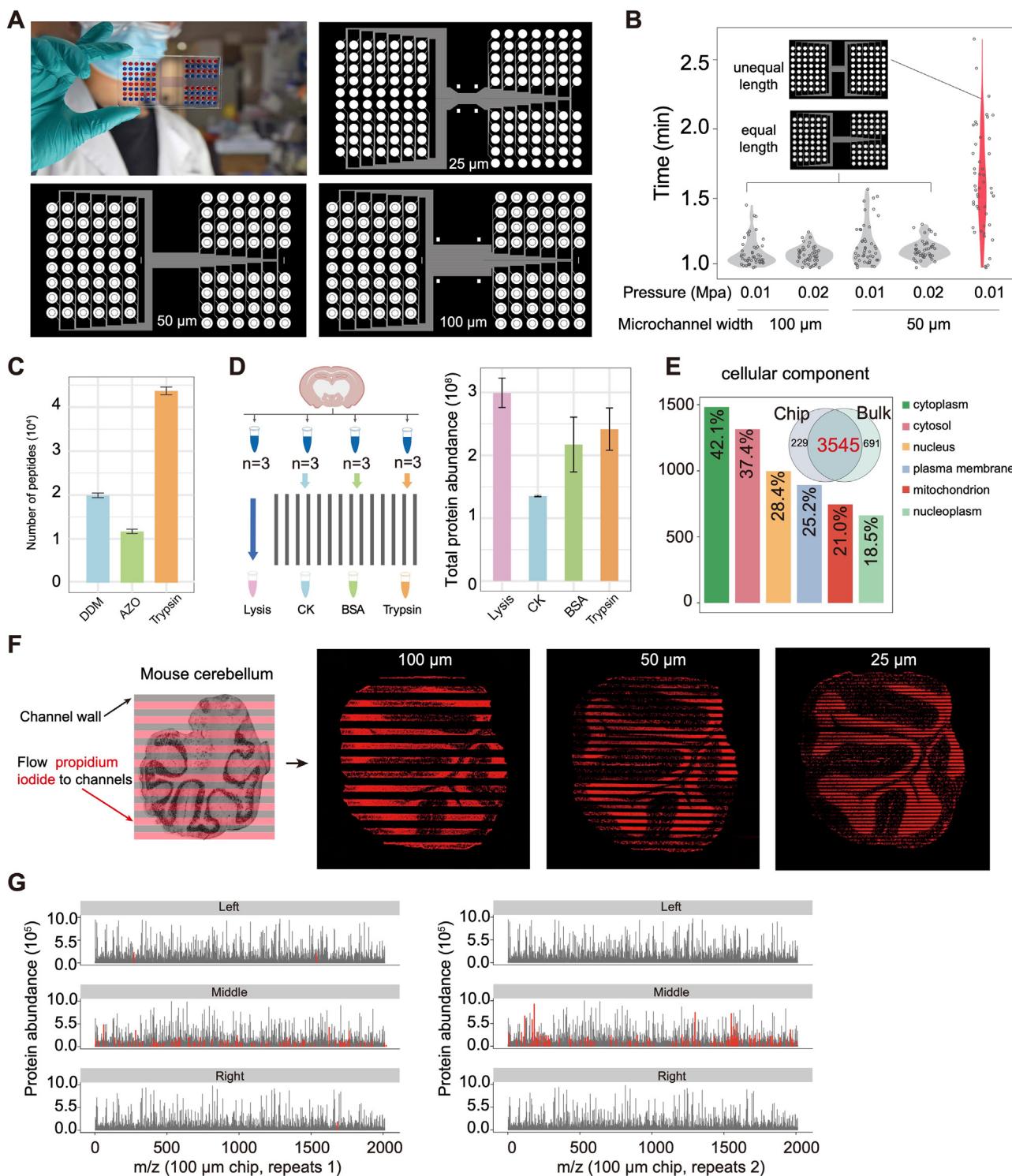
To explore broad signatures of cell-type-specific responses, Gene Ontology (GO) and gene set enrichment analysis were executed using python implementation of gseapy (v 0.10.1).⁵⁵ The reference gene sets were obtained from the Gene Ontology Resource, Kyoto Encyclopedia of Genes and Genomes (KEGG), and hallmark gene sets in Human Molecular Signatures Database (MSigDB).

ADDITIONAL RESOURCES

Abbreviations

In the manuscript, HER2/ERBB2 means human epithelial growth factor receptor 2, ER/ESR1 means estrogen receptor, and PR means progesterone receptor.

Supplemental figures



(legend on next page)

Figure S1. Microfluidics-based parallel sampling to resolve the spatial distribution of proteomic data, related to Figure 1

- (A) Picture of microfluidic chip illustrated with food dye, and the AutoCAD designs of microfluidic chips with different channel width (25, 50, 100 μm).
(B) Assessing fluid resistance distribution of channels in different types of microfluidic chips, including equal channel length and unequal channel length chip. Fluid resistance distribution of channels was inferred by recording the time required for 20 μL of fluid to flow from the inlet to the outlet.
(C) Optimization of the PLATO workflow for nanogram-scale samples. Specifically, AZO, DDM, and direct-lysis with trypsin were used for on-chip digestion. Then, the lysates were collected and analyzed by LC-MS/MS. Figure shows the number of detected peptides in each condition.
(D) Evaluation of nonspecific adsorption of the PDMS chip channel surface for protein samples. Lysis: tissue lysate directly detected by LC-MS/MS without flowing into channels. CK, BSA, and trypsin: before detected by LC-MS/MS, tissue lysate flowing into the channels pretreated by ammonium bicarbonate solution (CK), BSA, or trypsin.
(E) Cellular component (CC) enrichment of LC-MS/MS data. The data of histogram obtained from on-chip lysate and bulk proteomic data obtained from whole-tissue section lysate.
(F) Chips with different channel width including 100, 50, and 25 μm were used to evaluate the molecular leakage across channels. Specifically, fluorescent dye propidium iodide (PI) flows into channels to stain the section of cerebellum.
(G) Evaluation of peptides diffusing across channels by detecting the LC-MS/MS signals of *E. coli* in adjacent microchannels.

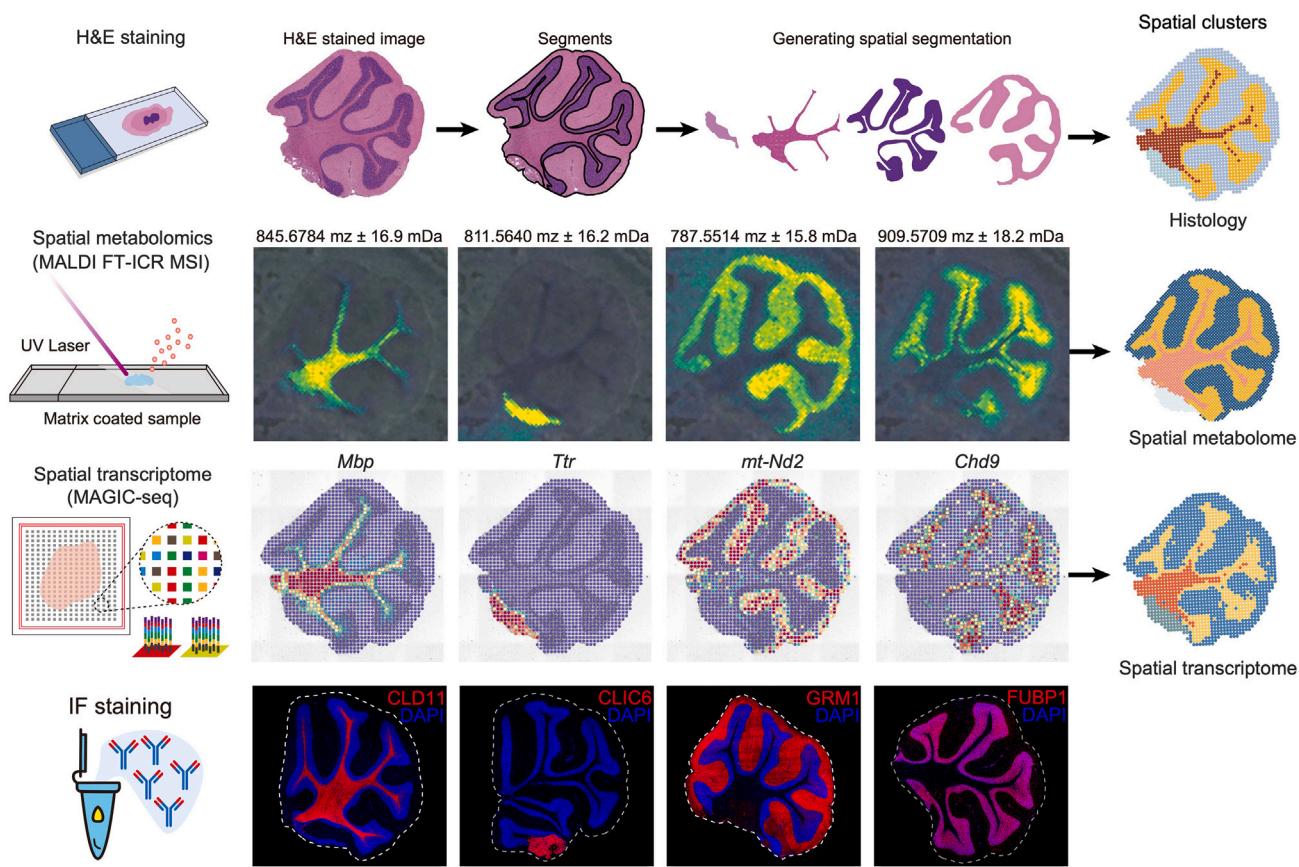
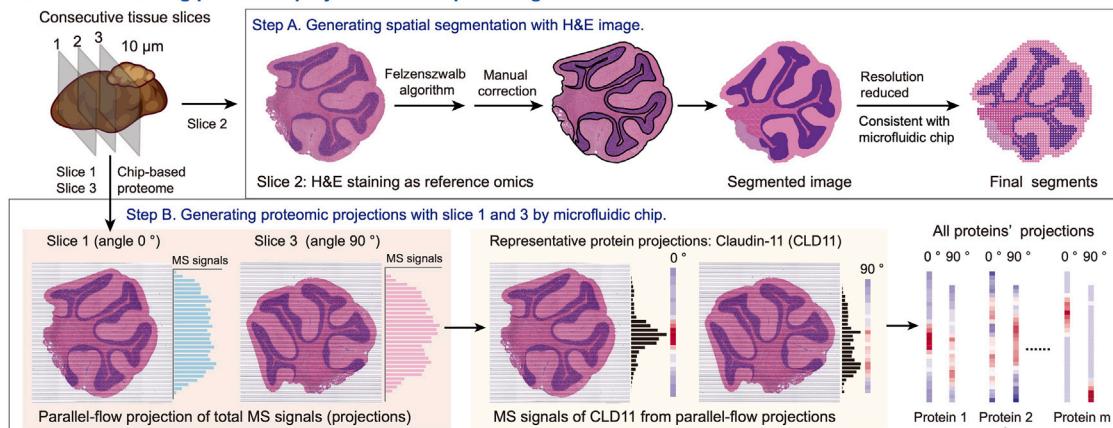


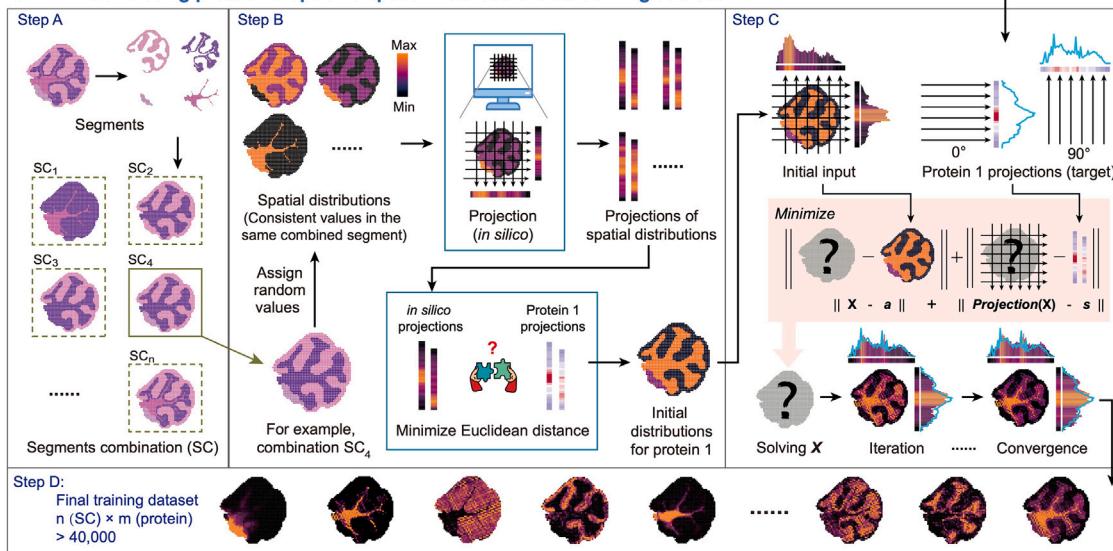
Figure S2. Correlation evaluation among different spatial omics, related to Figure 3

To assess the concordance across various spatial omics, we obtained a series of consecutive cerebellum sections, each 10 μm thick. One section underwent histological staining (H&E staining), another was used for spatial metabolome imaging via matrix-assisted laser desorption ionization Fourier-transform ion cyclotron resonance mass spectrometry imaging (MALDI FT-ICR MSI), and the third section was processed for spatial transcriptomics using MAGIC-seq. Immunofluorescence (IF) staining was performed to visualize the histological structure of the cerebellum at the protein level. The representative spatial patterns are shown together with the clustering results from the different spatial omics, including histology, spatial metabolomics, and spatial transcriptomics.

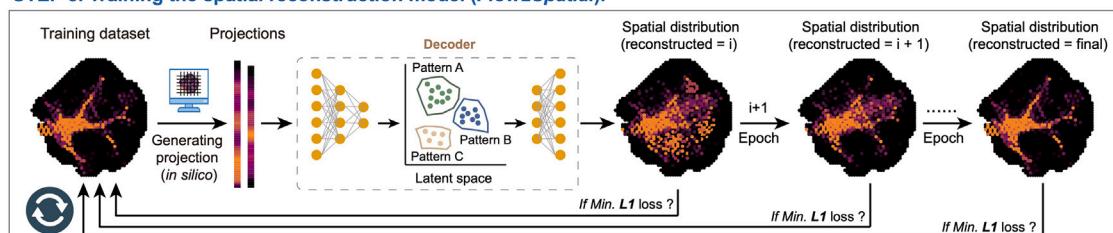
STEP 1. Performing proteomic projections and spatial segmentation.



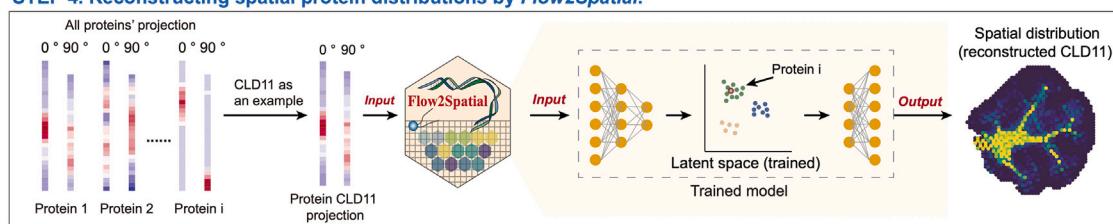
STEP 2. Generating proteome-specific spatial distributions as training dataset.



STEP 3. Training the spatial reconstruction model (*Flow2Spatial*).



STEP 4. Reconstructing spatial protein distributions by *Flow2Spatial*.



(legend on next page)

Figure S3. Step-by-step reconstruction process of PLATO using H&E as a reference, related to Figure 3

Step 1: performing proteomic projections and spatial segmentation. To minimize differences in signal distribution across slices, three consecutive cerebellar slices were collected for spatial proteome analysis of PLATO. The first and third slices underwent parallel-flow projection using microfluidic chips at 0° and 90° angles, followed by sensitive LC-MS/MS proteome analysis. The second slice served as a reference for H&E staining. After H&E staining, the H&E image was segmented using the Felzenszwalb algorithm with manual curation. These segments served as distribution constraints, ensuring uniform expression levels among pixels within the same segment in subsequent step. Step 2: generating proteome-specific spatial distributions as training dataset. In this step, we aimed to create a training dataset closely resembling the spatial proteome. To achieve this, we built a data generator capable of producing spatial distributions by learning the associated proteomic features. Initially, we speculated potential proteome clusters by randomly combining reference segments. For each combination of segments, we generated multiple spatial distributions by randomly assigning values to individual segments. Then, we selected the generated spatial distribution that best matched the parallel-flow projections of specific proteins, determined by Euclidean distance. This selected distribution underwent optimization to align with the proteome parallel-flow projections. After convex iteration, the resulting converged optimal spatial distribution, informed by proteomic data, was then incorporated into the training dataset. Step 3: training spatial reconstruction model with the training dataset. All spatial data generated served as the training dataset for our deep learning reconstruction model. We employed an autoencoder-like deep learning architecture to bridge parallel-flow projections with spatial information. Within the encoder component, we simulated *in silico* projection for each spatial distribution in the training dataset, replicating the microfluidic chip-based parallel-flow projection process. Concurrently, within the decoder component, we trained a ResNet-based model to reconstruct spatial distributions from the proteome parallel-flow projection values. Finally, the trained model embedded all spatial patterns present in the training dataset into a latent space, reflecting the relationships among the projected values. Step 4: reconstructing spatial protein distributions by the trained model. To faithfully reconstruct the spatial proteome, the trained decoder was used as the reconstruction model to regenerate the original spatial distributions from the proteome parallel-flow projections. Here, CLD11 was taken as an example to demonstrate the performance of PLATO.

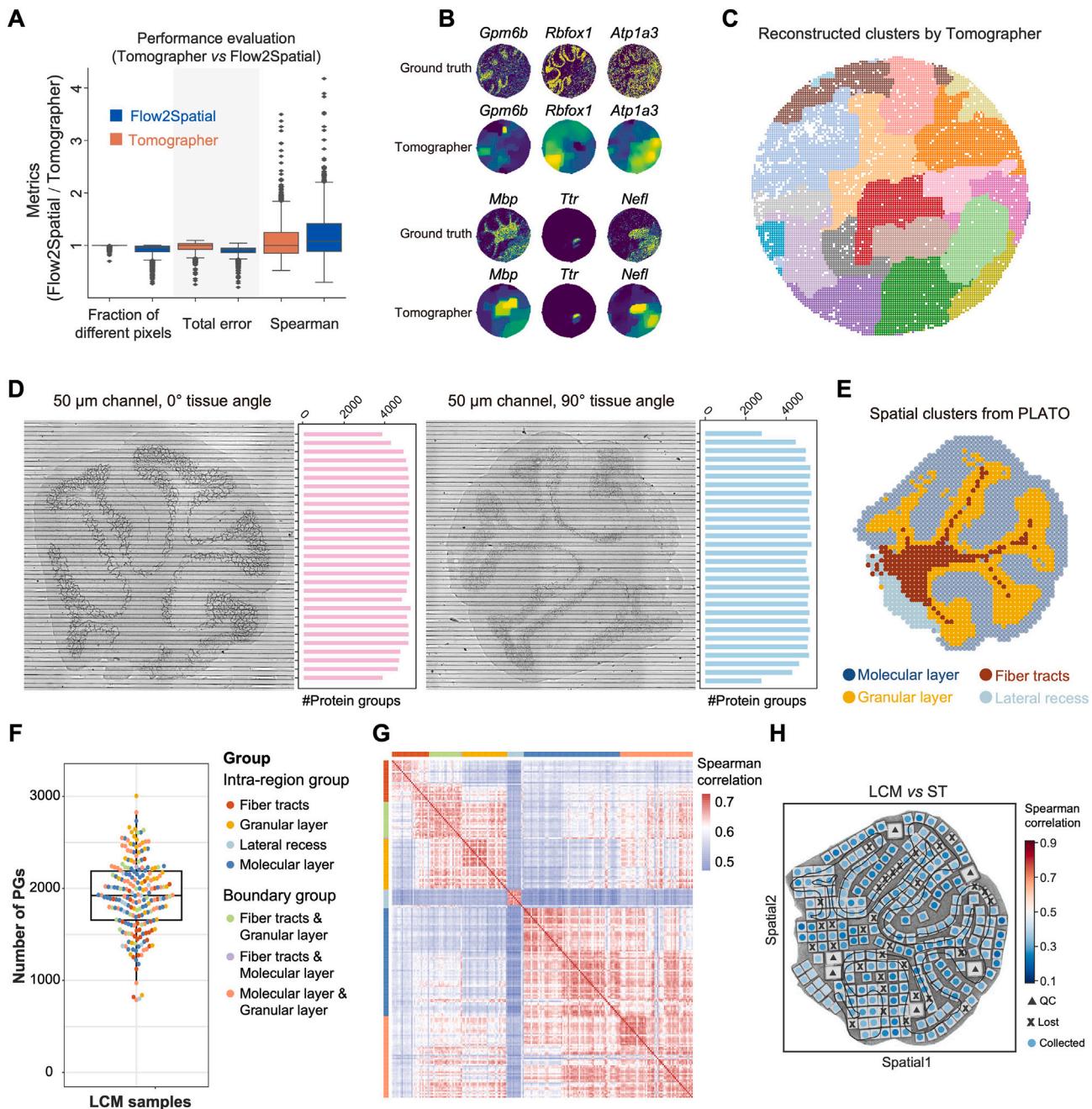


Figure S4. Spatial distribution of proteins can be accurately reconstructed by Flow2Spatial, related to Figure 4

- (A) Reconstruction accuracy compared with Tomographer over all genes evaluated, using different metrics of similarity to the ground truth: Spearman's R, relative total error, and fraction of different pixels.
- (B) Visualization of gene expression generated by ground truth and reconstructions.
- (C) Unbiased determination of molecular anatomy obtained by pixel clustering of Tomographer.
- (D) Number of detected protein groups of each microchannel from 0° and 90° cerebellum slices.
- (E) Spatial clustering results based on reconstructions of Flow2Spatial in cerebellum.
- (F) Boxplot shows the number of detected protein groups in LCM samples.
- (G) Heatmap exhibits expression correlation between LCM samples.
- (H) Pixel-by-pixel Spearman's correlation of expression between LCM proteome and spatial transcriptomics (ST). Triangle shapes represent quality control samples, x-shaped markers indicate failed collections during microdissection, and circles mark successfully collected and filtered samples.

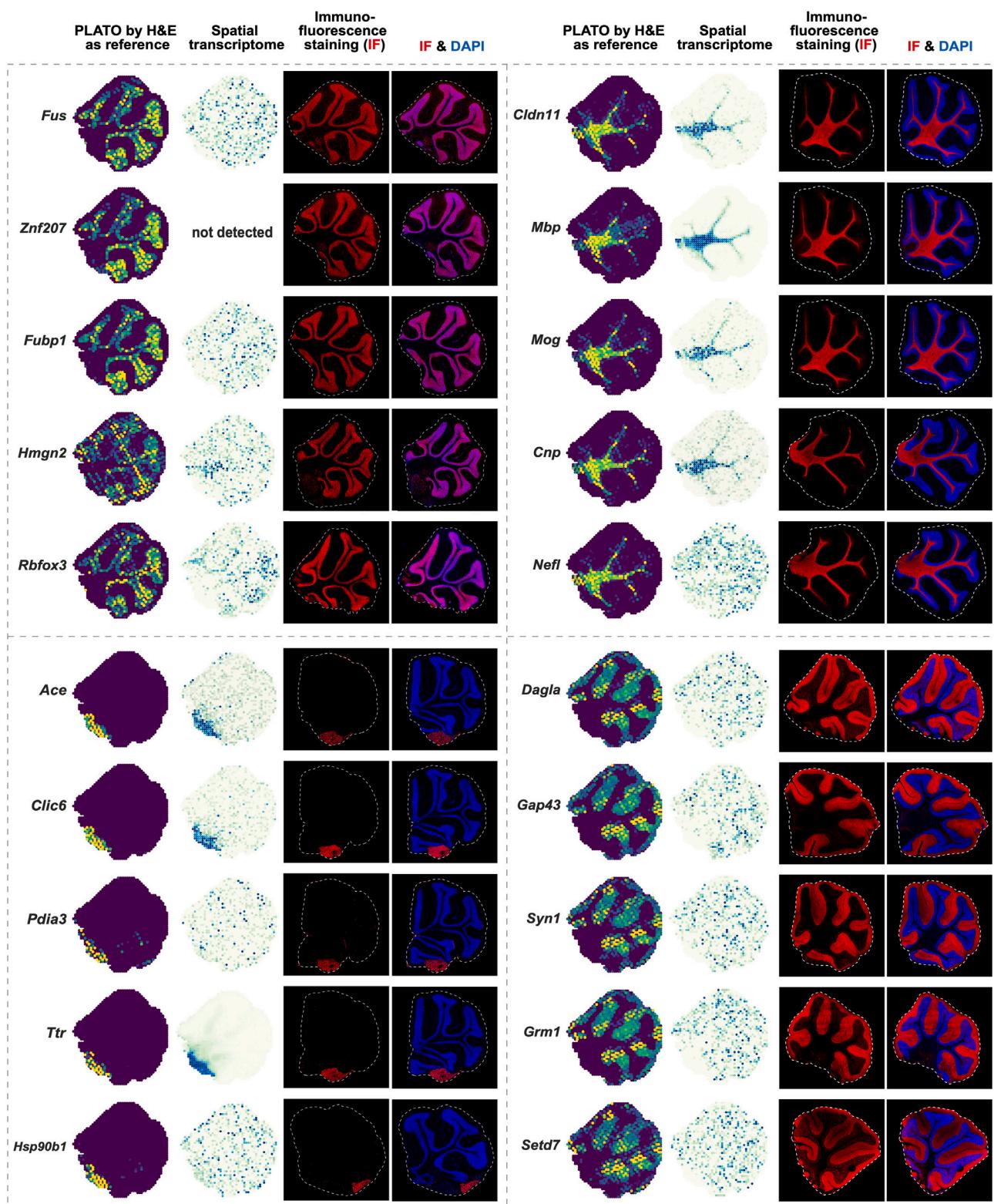


Figure S5. Validation of individual protein reconstructions by IF staining, related to Figure 4

Individual genes were captured by different methods, including PLATO (protein level), spatial transcriptomics (ST) (RNA level), and IF staining (protein level).

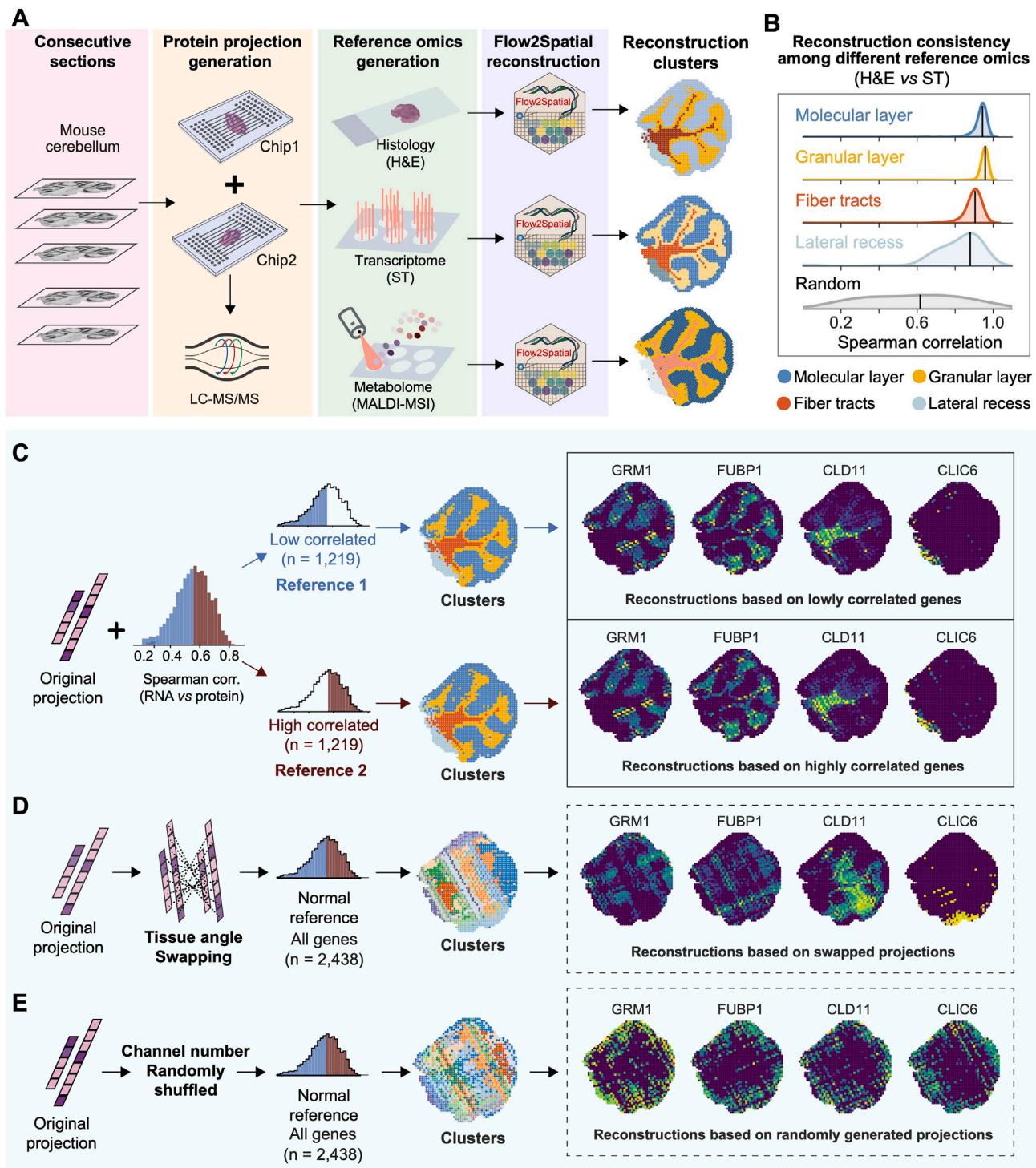


Figure S6. Evaluation of different references on PLATO performance, related to Figure 4

(A) Flow2Spatial reconstructions using three different spatial omics datasets as reference: H&E staining image, spatial transcriptome, and spatial metabolome (MALDI-MSI). The spatial clusters are based on these reference omics.

(B) Comparison of PLATO reconstructions using different reference omics. Spearman correlations were calculated for all spots in each region, using H&E and spatial transcriptomics (ST) as references.

(C-E) Robustness validation of Flow2Spatial through three simulations.

(legend continued on next page)

-
- (C) Reference genes were divided into two groups based on the expression similarity between mRNAs and their corresponding proteins. Flow2Spatial reconstructions were then executed separately by using these two gene groups as reconstruction reference.
- (D) Using normal spatial transcriptomics data as a reference, the tissue angle was swapped, inputting an incorrect protein projection into Flow2Spatial by converting angle labels from 0° to 90° and vice versa.
- (E) Again using normal spatial transcriptomics data as a reference, the channel numbers were randomly shuffled before inputting the protein projection into Flow2Spatial, resulting in an incorrect channel order.

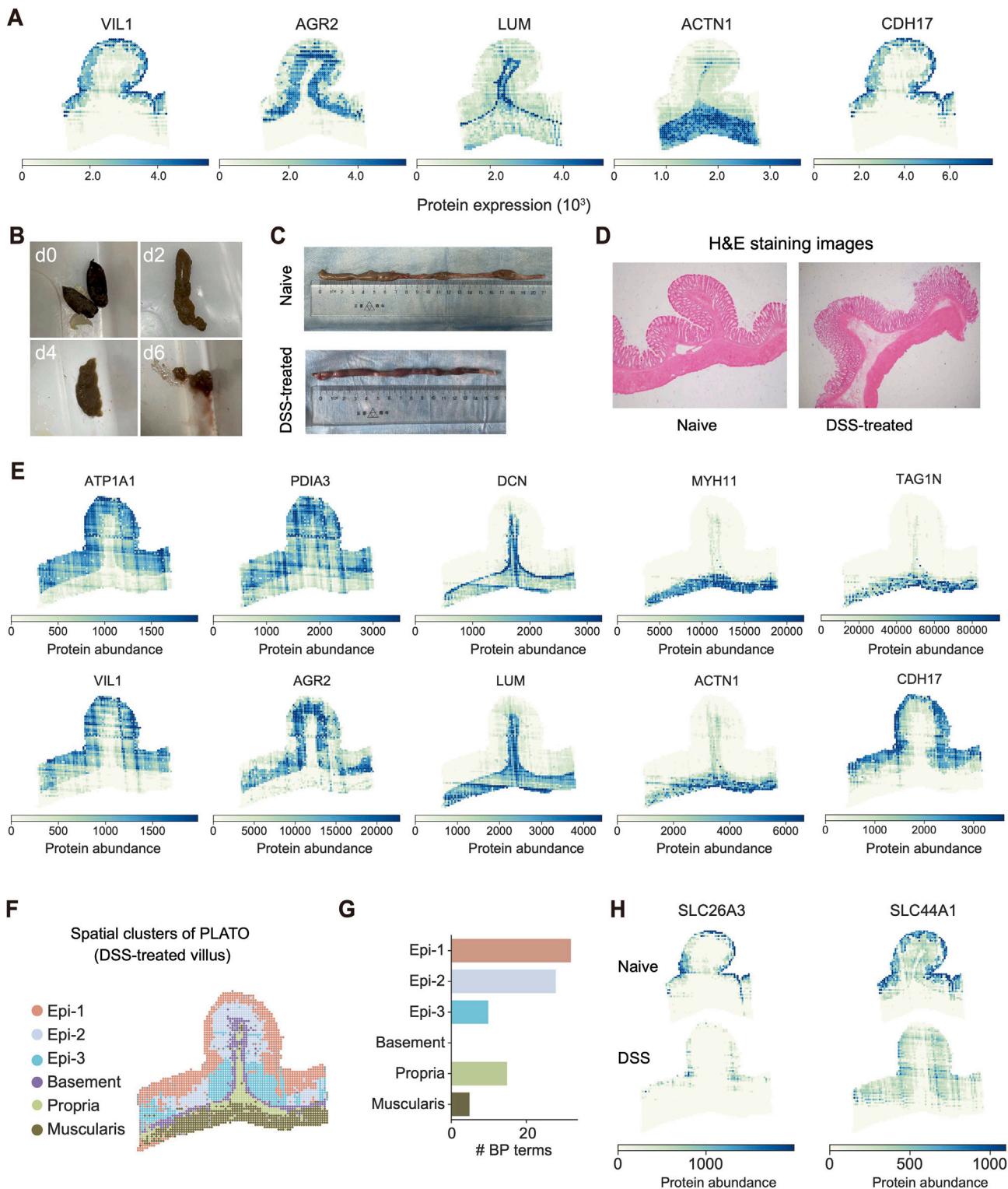
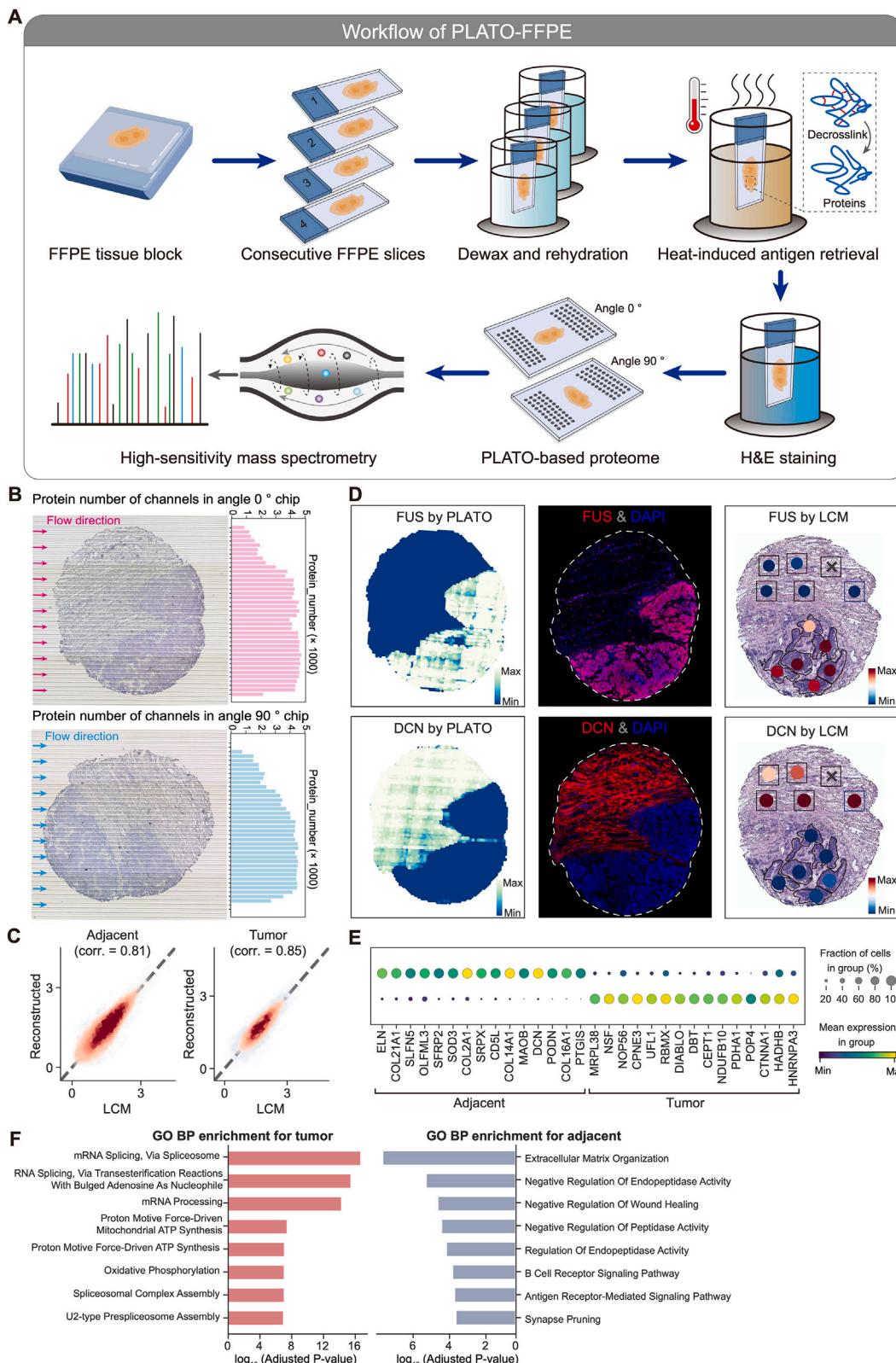


Figure S7. Spatial proteomics mapping of villus, related to Figure 5

- (A) Spatial expression of representative proteins from naive villus.
 (B) Images of rat stools after DSS treatment. Four pictures represent 0, 2, 4, and 6 days after DSS treatment.
 (C) Intestinal tract length between naive and DSS-treated mice.
 (D) H&E staining of villus between naive and DSS-treated mice.

(legend continued on next page)

-
- (E) Spatial expression of representative proteins from DSS-treated villus.
 - (F) Clustering based on reconstructions of Flow2Spatial in villus after DSS treatment. The clusters were visualized in different colors. Epi-1, -2, and -3 represent epithelial layer 1, 2, and 3 in the clustering results, respectively.
 - (G) The number of GO biological processes enriched in different clusters of the villus.
 - (H) Spatial expression of transporters in each condition.



(legend on next page)

Figure S8. Spatial proteomic mapping of FFPE breast cancer tissue, related to Figure 6

- (A) Workflow of PLATO for FFPE breast cancer tissue. Initially, consecutive FFPE slices of breast cancer tissue were obtained from an FFPE tissue block and attached to customized adhesive glass slides. Following drying and heating, the slices underwent dewaxing and rehydration through successive immersions in xylene and ethanol. Subsequently, we adopted a heat-induced antigen retrieval method to reverse the crosslinking of fixed proteins. The tissue slices were then stained with hematoxylin and eosin (H&E). Next, two PDMS chips were directly placed on two tissue slices, respectively. Finally, the tissues within the micro-channels were lysed directly by trypsin and analyzed using LC-MS/MS.
- (B) Number of obtained protein groups of each microchannel from 0° and 90° slices.
- (C) Regional concordance between LCM-based proteome and PLATO reconstructions. Each anatomical region is represented by all samples from the region. Spearman's coefficient was calculated.
- (D) Spatial distribution of two proteins by PLATO, immunofluorescence staining, and LCM-based proteome.
- (E) Dot plot showing the expression levels of top eight marker proteins in clustering regions.
- (F) Gene Ontology functional enrichment by the top 100 marker proteins of two clustering regions.

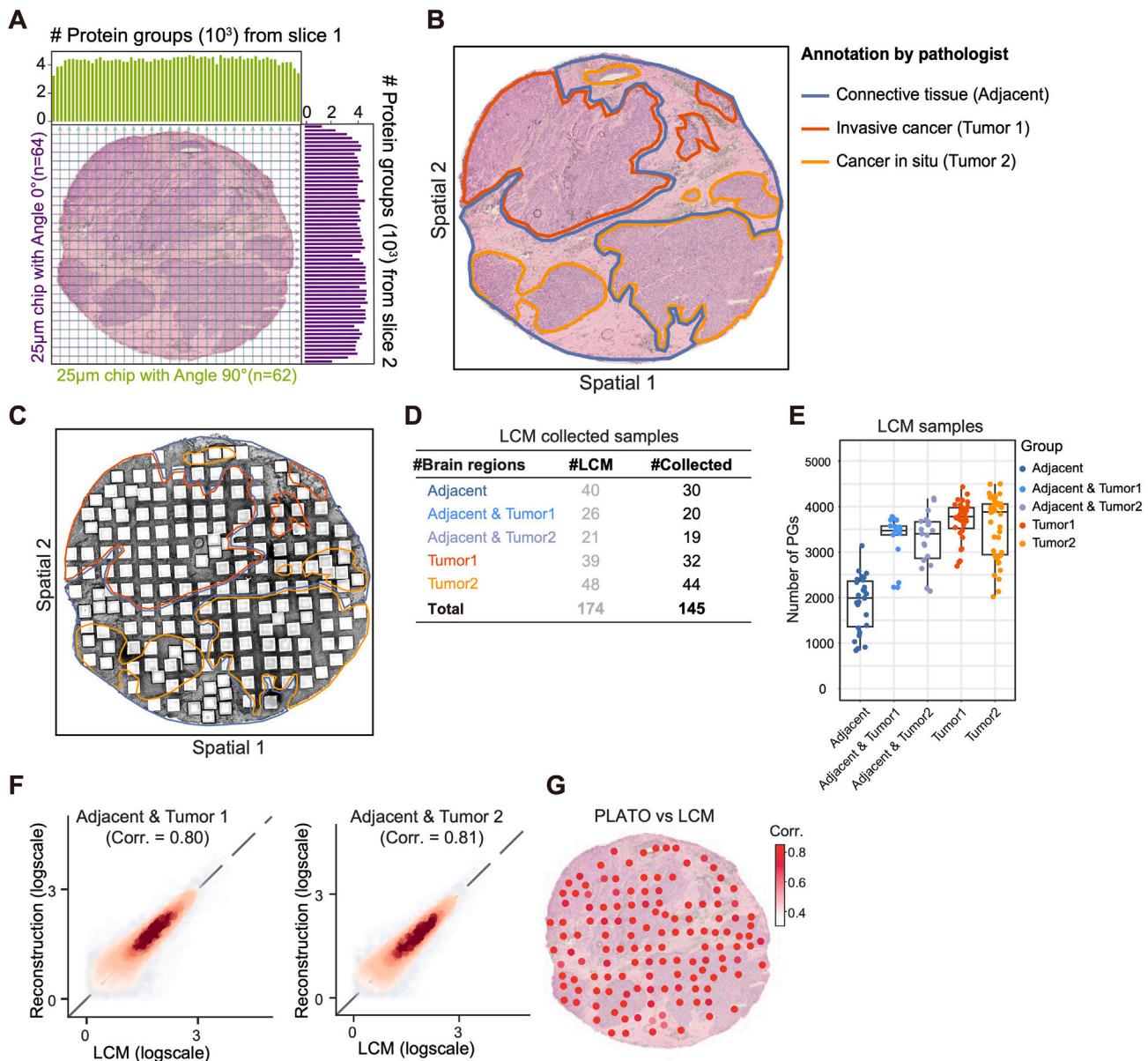


Figure S9. Spatial proteomic mapping of fresh frozen breast cancer tissue, related to Figure 6

- Number of protein groups obtained from each microchannel for 0° and 90° slices.
- Pathological annotations of breast cancer based on H&E staining images.
- Schematic diagram of microdissection points, with shaped dots indicating sampling areas and colored lines representing pathological annotations.
- Statistics on the number of LCM samples collected.
- Number of obtained protein groups from LCM samples.
- Concordance of regional boundaries between LCM-based proteome and PLATO reconstructions, with Spearman's coefficients calculated.
- Spearman's correlation of protein expression between LCM-based proteome and PLATO reconstruction, analyzed spot by spot.