OpenDVP: An experimental and computational framework for community-empowered deep visual proteomics

Jose Nimo[1,2,11], Sonja Fritzsche[1,3], Daniela S. Valdes[1], Minh Trinh[1], Tancredi Pentimalli[1,2,4], Melissa Klingeberg[1,3], Simon Schallenberg[5,11], Frederick Klauschen[6,7,8], Florian Herse[1,2,9], Stefan Florian[5,11], Nikolaus Rajewsky[1,2,4,10,11,12,13,14,15] and Fabian Coscia[1,11*]

[1]Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany
[2]Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
[3]Humboldt-Universität zu Berlin, Institute of Biology, Berlin, Germany
[4]Laboratory for Systems Biology of Regulatory Elements, Berlin Institute for Medical Systems Biology (BIMSB), Max-Delbrück-Centrum for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany
[5]Institute of Pathology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
[6]Institute of Pathology, Ludwig Maximilians University Hospital Munich, Munich, Germany
[7]German Cancer Consortium (DKTK), Partner Site Munich, and German Cancer Research Center (DKFZ), Heidelberg, Germany
[8]BIFOLD - Berlin Institute for the Foundations of Learning and Data, Berlin, Germany.
[9]Experimental and Clinical Research Center, a cooperation between the Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association and the Charité - Universitätsmedizin Berlin, Germany
[10]Cluster of Excellence, NeuroCure, Charitéplatz 1, 10117 Berlin, Germany
[11]German Cancer Consortium (DKTK), Partner Site Berlin, Heidelberg, Germany
[12]German Center for Cardiovascular Research (DZHK), Site Berlin, Berlin, Germany
[13]National Center for Tumor Diseases (NCT), Site Berlin, Berlin, Germany
[14]German Center for Child and Adolescent Health (DZKJ), Site Berlin, Germany
[15]Max Delbrück Center for Molecular Medicine (MDC) and Cluster of Excellence ImmunoPreCept and Einstein Center for Early Disease Interception (EC-EDI)

*To whom correspondence should be addressed: fabian.coscia@mdc-berlin.de

Abstract

Deep visual proteomics (DVP) is an emerging approach for cell type-specific and spatially resolved proteomics. However, its broad adoption has been constrained by the lack of an open-source end-to-end workflow in a community-driven ecosystem. Here, we introduce openDVP, an experimental and computational framework for simplifying and democratizing DVP. OpenDVP integrates open-source software for image analysis, including MCMICRO, QuPath, and Napari, and uses the scverse data formats AnnData and SpatialData for multi-omics integration. It offers two workflows: a fast-track pipeline requiring no image analysis expertise and an artificial intelligence (AI)-powered pipeline with recent algorithms for image pre-processing, segmentation, and spatial analysis. We demonstrate openDVP's versatility in three archival tissue studies, profiling human placenta, early-stage lung cancer, and locally relapsed breast cancer. In each study, our framework provided insights into health and disease states by integrating spatial single-cell phenotypes with exploratory proteomic data. Finally, we introduce deep proteomic profiling of cellular neighborhoods as a scalable approach to accelerate spatial discovery proteomics across biological systems.

54 # Introduction

55 Recent advances in spatial omics have revolutionized our understanding of tissue biology and
56 disease processes [1]. In particular, spatial proteomics is experiencing remarkable progress in
57 sensitivity, throughput, and spatial resolution [2]. The ability to map proteins with high spatial
58 resolution has provided novel insights into cellular organization, tissue heterogeneity, and
59 disease mechanisms with unprecedented detail. The significance of these advancements lies in
60 the premise that proteomic profiles provide a direct perspective on the functional and
61 phenotype-centric cellular states that govern health and disease. Historically, spatial proteomics
62 has depended on targeted, antibody-based methodologies [3], which offer excellent spatial
63 resolution, yet require prior knowledge for antibody panel design. Antibody-based methods,
64 which generally profile up to ~60 proteins [4], capture a comparatively small fraction of the
65 proteome, which is estimated to encompass more than 10,000 different proteins per single cell
66 type [5]. A highly synergistic approach involves integrating targeted methods with ultrasensitive
67 liquid chromatography-mass spectrometry (LC-MS)-based proteomics. This multiscale approach
68 has paved the way for exploratory, comprehensive analyses of cell-type and spatially resolved
69 tissue proteomes [6,7]. Recently, we co-developed deep visual proteomics (DVP)[8], which combines
70 tissue imaging (immunofluorescence [IF] or immunohistochemistry [IHC]), machine learning-
71 based image analysis, automated laser microdissection (LMD), and ultrasensitive MS for the
72 exploratory proteomic profiling of cell types or regions of interest (ROI). DVP enables the
73 systematic mapping of thousands of proteins, elucidating their networks and signaling pathways
74 within complex tissue architectures, and providing unprecedented insights into health and
75 disease states. For example, DVP recently enabled the discovery of a curative treatment strategy
76 for a fatal human skin disease [9] and mapped single-cell proteotoxicity in human liver [10].

77 For image analysis, the Biological Image Analysis Software (BIAS) [8] has enabled full support of
78 DVP experiments. However, broader adoption and methodological development have been
79 limited by the absence of non-proprietary, open-source workflows that support scalable image
80 analysis, laser microdissection experimental design, multimodal data integration, and
81 interoperability across imaging and omics platforms. These limitations have constrained
82 systematic benchmarking, extensibility, and community-driven DVP innovations.

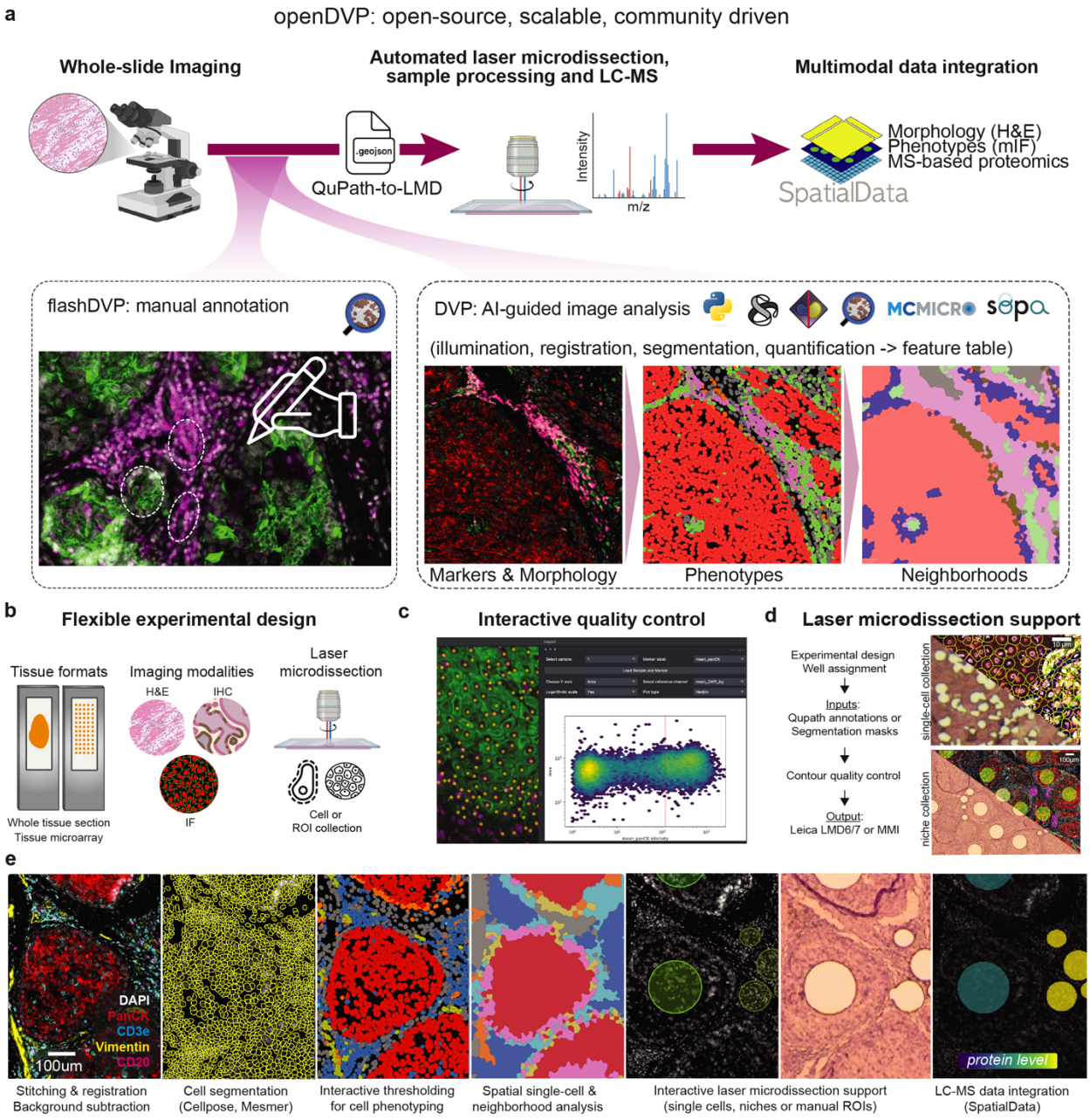
83 To address these gaps, we developed openDVP, an open-source modular framework designed to
84 complement and extend existing DVP solutions, including BIAS. OpenDVP provides flexible,
85 interoperable workflows that support simplified experimental strategies for users with limited
86 computational expertise and advanced pipelines optimized for scalability and complex analyses.
87 Built to integrate state-of-the-art bioimage analysis, spatial statistics, and ultrasensitive mass
88 spectrometry–based proteomics, openDVP enables scalable acquisition and analysis of hundreds
89 of spatially resolved proteomes. Across three archival tissue studies, we demonstrate the
90 versatility of the framework and show that cellular neighborhood–guided proteome profiling
91 offers a balanced alternative to technically demanding single-cell isolation, enabling efficient
92 spatial discovery proteomics at scale.

93 Results

94 Overview of the openDVP framework

95 We developed openDVP, an open-source modular framework for user-empowered deep visual
96 proteomics (**Fig. 1a**) (https://github.com/CosciaLab/openDVP). Our end-to-end framework
97 encompasses optimized functions for all the key steps of the DVP workflow, supporting image
98 preprocessing and analysis, automated laser microdissection, interactive image data
99 visualization, and multimodal data integration. OpenDVP supports the most common imaging
100 and proteomics data formats by current and extending functionalities from existing open-source
101 software, such as MCMICRO[11], SOPA[12], QuPath [13], Napari [14], and SpatialData [15]. Importantly, the
102 modules' inputs and outputs are standard file formats for seamless integration of community-

103 published or in-house tools. We introduce two complementary workflows: a simplified fast-track
104 pipeline, termed flashDVP, which does not require image analysis expertise, and an extended AI-
105 powered pipeline capable of in-depth spatial tissue analyses. FlashDVP serves as an intuitive
106 entry point for users seeking to rapidly convert raw images (i.e., H&E, IHC, or IF) into exploratory
107 spatial proteomics results. The extended AI-powered DVP workflow enables more experienced
108 users to customize and optimize the pipeline using advanced single-cell spatial analysis tools for
109 more complex phenotype-to-proteotype associations.



110

Fig. 1: Overview of the openDVP framework

(a) Main workflow of openDVP begins with tissue imaging to produce TIFF files. FlashDVP comprises manual image annotation, coordinate transfer with the QuPath-to-LMD webapp, and laser microdissection. OpenDVP provides a framework to process images into feature tables and analyze matrices to phenotype cells, using open-source tools. These data layers provide the basis for contour export and laser microdissection via QuPath-to-LMD. **(b)** OpenDVP supports tissue microarrays or whole-slide images, common imaging modalities, and different LMD collection strategies. **(c)** The Napari-cell-gater plugin enables phenotyping through visual feedback thresholding, plotting cell features in FACS-like density plots. **(d)** QuPath-to-LMD webapp allows users to design, transfer, and validate annotations from QuPath and openDVP to an LMD-ready file format. **(e)** Exemplary openDVP workflow showing the results of multiplexed immunofluorescence (mIF) imaging, cell segmentation, cell phenotyping, cellular neighborhood analysis, contour annotations, and multimodal data integration using SpatialData.

Providing a versatile solution to support diverse spatial proteomics experiments

We built openDVP for compatibility with a wide range of imaging applications that support whole-slide images (WSI) and tissue microarrays (TMA) based on conventional widefield or confocal microscopy (H&E, IHC, IF, and mIF) (**Fig. 1b**). While WSI can provide a comprehensive view of large tissue areas, which are particularly important for heterogenous tissues, TMA applications support cohort-size spatial proteome analyses. Importantly, the imaging pipelines use BioFormats compatible file formats (e.g., OME-TIFF) and do not rely on any proprietary software or specific microscopy hardware for broad accessibility.

Efficient and extensible image processing

Following image acquisition, the first step in DVP workflow is the processing of images, including illumination correction and stitching image tiles. OpenDVP builds on top of the established open-source image processing pipelines MCMICRO and SOPA (**Fig. 1a**). MCMICRO was built using Nextflow [16], following nf-core guidelines, to ensure high standards of modularity, consistency, and interoperability. SOPA is built on python-based Snakemake [17], enabling python-fluent users to add new modules or modify existing steps. We developed custom modules to expand segmentation masks by a defined number of pixels, quantify marker intensity quantiles per cell, and pyramidize images for smooth quality control. These open-source pipelines offer key advantages: (1) Broad compatibility, supporting most imaging formats and modalities. (2) Modular architecture, integrating latest benchmarked algorithms and machine learning models, such as Cellpose [18] and Mesmer [19]. (3) Scalability, enabling efficient distribution of processes for stitching, registration and segmentation. For example, segmenting eleven WSI with a total of 6.34

145 million cells took less than two hours (**Extended Data Fig. 1a-c**). (4) Active maintenance and
146 development by a growing community of bioimage analysts.
147
148 **Interactive quality control and marker thresholding for single-cell phenotyping**
149 Visual quality control is critical in image processing pipelines to assess illumination correction,
150 stitching and registration results, and segmentation accuracy. OpenDVP also provides demo
151 datasets and Jupyter notebook-based tutorials to help users integrate and inspect results using
152 two complementary interfaces, Napari and QuPath. We chose them for their ability to visualize
153 larger-than-memory images and community activity in developing new features. To facilitate
154 image-based cell phenotyping based on marker expression, we developed an interactive Napari
155 plugin (**Fig. 1c**) (github.com/CosciaLab/napari-cell-gater). Our plugin enables users to visualize
156 multichannel images, overlay segmentation masks, plot mutually exclusive markers (FACS
157 analysis-like), and provide real-time feedback of cells being labelled positive or negative based
158 on selected thresholds. Although more laborious, this approach gives users full autonomy in their
159 filter strategy for the complex task of single-cell phenotyping. For transparency, our plugin
160 generates summary tables with cutoff values for each marker-sample pair. Finally, our image-
161 processing pipeline creates a filtered single-cell matrix used for supervised or unsupervised
162 analysis to identify cell types, states, and cellular neighborhoods. These guide laser
163 microdissection and exploratory LC-MS-based proteome analysis.
164
165 **A versatile interface between image analysis and laser microdissection**
166 Laser microdissection is essential in the DVP workflow for isolating single cells or ROIs precisely
167 and automatically. Segmentation-based cell/nucleus isolation provides the highest biological
168 granularity, particularly relevant for single-cell applications or rare phenotypes. The isolation of
169 larger ROIs (e.g., multi-cellular niches) represents a robust alternative to single-cell LMD cutting.
170 To support both scenarios, we developed QuPath-to-LMD, an open-source web app for
171 experimental design, contour export, and validation of LMD-ready masks, supporting the Leica
172 LMD7 and MMI CellCut systems (**Fig. 1d**) (github.com/CosciaLab/Qupath_to_LMD). QuPath-to-
173 LMD provides an interface between histopathology and omics-based profiling, as slide
174 annotations can be directly transferred into LMD-ready contours for downstream (prote)omic

175 analysis. We validated single-cell and ROI-based contour alignment on the LMD7 system using
176 different magnifications (20x and 40x) and provided example images for selecting optimal tissue
177 reference points (**Extended Data Fig. 1d-e**).

178 **Integrating imaging and MS-based proteomics data through the scverse ecosystem**

179 DVP generates two types of proteomic data: imaging and LC-MS data. For data integration,
180 visualization, and multi-modal analysis, openDVP's python code uses scverse standards and the
181 AnnData and SpatialData formats, ensuring compatibility with popular analysis packages such as
182 scimap [20], scanpy [21], and spatialproteomics [22]. To our knowledge, this framework represents the
183 first instance of integrating mass spectrometry-based proteomic data into the scverse ecosystem,
184 streamlining data analysis for large-scale spatial tissue proteomics. For example, through
185 SpatialData adoption, users can overlay their images with cellular phenotype information (e.g.,
186 cell types and cellular neighborhoods) and matching quantitative proteome information. We
187 developed a data analysis package to facilitate storing, integrating, visualizing, and analyzing
188 spatial proteomics data, available for download from the PyPi registry (**Extended Data Fig. 1f**). A
189 vignette of the openDVP workflow is shown in **Fig. 1e.**

190 OpenDVP quantifies cell-type resolved proteomes of human placental tissue

191 Understanding biological foundations of human health and disease requires comprehensive
192 characterization of individual cells, the fundamental units of life. Recent single-cell and spatial
193 omics advancements have catalyzed large-scale initiatives, such as the Human Cell Atlas [23] and
194 the LifeTime Initiative [24], which aim to systematically profile all human cell types across tissues
195 to unravel relationships between cells, tissue organization, and function. However, certain cell
196 types (e.g., cardiomyocytes, adipocytes, and neurons) pose analytical challenges due to their
197 resistance to dissociation into single-cell suspensions. Their interconnectivity and/or fragility
198 make them inaccessible for sorting-based single-cell profiling, and they are often morphologically
199 too complex for accurate cell segmentation. One example is the syncytiotrophoblast (STB) of the
200 human placenta, a syncytial layer derived from fused cytotrophoblasts (CTB) that lacks distinct
201 cell boundaries (**Fig 2a**), making cell segmentation and single-cell-based laser microdissection
202 challenging. Despite STB's critical roles in nutrient exchange, hormone secretion, and immune

203 modulation at the maternal-fetal interface [25], its syncytial nature has prohibited its
204 comprehensive molecular profiling, particularly at the global proteomic level. With this in mind,
205 we applied flashDVP as a simplified and image segmentation-free version of DVP. We performed
206 four-color immunofluorescence (IF) staining of 5µm thick first trimester human placental FFPE
207 sections mounted on PPS frame slides and performed centimeter-scale whole-slide imaging (**Fig**
208 **2b**).
209 The placental barrier, appearing in tree-like villi structures, controls the nutrient and gas
210 exchange between maternal and fetal blood. It consists of E-cadherin+ bipotential CTBs that fuse
211 to form a multinucleated E-cadherin−, HLAG− STB layer that is in direct contact to maternal blood.
212 HLAG+ trophoblasts constitute the invasive CTB derived lineage that anchor the placenta to the
213 maternal uterine tissue (**Fig 2a**). CD163+ extraembryonic macrophages (Hofbauer cells, HBC) in
214 the stroma play roles in tissue remodeling, immune surveillance, and support the formation and
215 maintenance of fetal blood vessels in the villi [26]. With QuPath, we manually annotated 150 cells
216 per replicate for each cell type (STB, CTB, and HBC, **Fig 2c-d**) to reach an equal tissue amount per
217 sample. LMD contours were created using the QuPath-to-LMD web app, transferred to a Leica
218 LMD7 microscope for laser microdissection and processed using our ultralow-input tissue
219 proteomics protocol [27]. Samples were measured in label-free diaPASEF mode and analyzed with
220 DIA-NN [28] in library-free mode. We identified 3,994 unique proteins across groups (**Extended**
221 **Data Fig. 2a**), including canonical markers for each cell type (**Fig. 2d, Extended Data Fig. 2b-c**).
222 Principal component analysis separated cell phenotypes regardless of tissue origin (**Fig. 2e**).
223 Differential abundance analysis revealed cell type-specific markers that showed good
224 concordance with public transcriptomics data [29,30] (**Fig. 2f-g, Extended Data Table 1**). Pathway
225 analysis of protein clusters confirmed STB-specific functions in steroid hormone metabolism,
226 vesicular transport, and energy demands (**Fig. 2h**). The regulation of SUMO proteins, linked to
227 placental development and hypertension in pregnancy in bulk studies [31,32], was specific to STBs
228 despite their lower abundance, underscoring the value of cell-type resolved proteomic analyses.
229 CTB and HBC showed an enrichment in RhoJ GTPase and VEGF signaling pathways, while HBCs
230 displayed immune regulation pathways, supporting their immunotolerant role. We found HBC-
231 specific enrichment for proteins involved in mRNA splicing, aligning with the role of alternative

splicing as an essential feature of placental dysfunction [33]. Given the physical proximity between STBs and CTBs (**Fig. 2a, c**), cell communication analysis between these connected trophoblasts cells revealed high paracrine activity of both cell types, with autocrine communication strongest in the CTB monolayer (**Fig. 2i, Extended Data Table 1**). Strong signaling via cadherins (CDH), junctional adhesion molecules (JAM), fibronectins (FN1), and desmosomes was inferred, supporting the importance of cell adhesion integrity in fusion-competent CTBs. We identified robust communication involving collagens, laminins, and extracellular matrix proteins critical for implantation and placentation [34]. Our data also revealed that endothelial cell-selective adhesion molecule (ESAM) as a unique and previously unreported autocrine signaling molecule of STBs. ESAM is an important component of tight junctions and a negative regulator of platelet activation; the latter is known to underpin pregnancy health and coexist with complications [35]. These findings establish flashDVP as a powerful 'fast-track' approach for cell type and spatially resolved tissue proteomics.

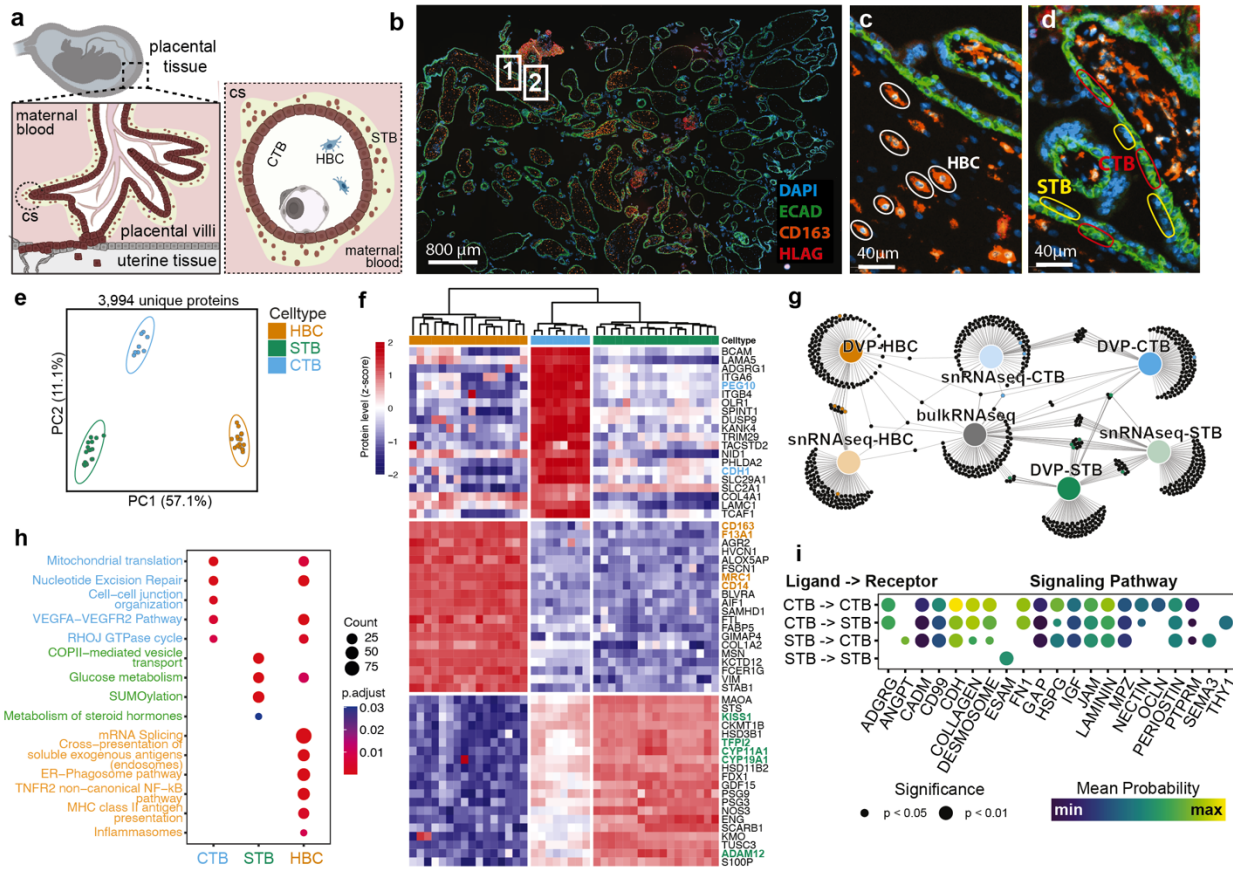


Fig. 2: High-resolution cell-type resolved proteomics of human placental tissue

(a) Left: Schematic of placental villi samples and fetal structures at the maternal-fetal interface. Right: Cross-section showing cytotrophoblast (CTB) monolayer replenishing syncytiotrophoblast (STB) layer, with Hofbauer cells (HBC) in the stroma. cs, cross-section. **(b)** Representative immunofluorescence image of placental tissue stained for E-cadherin, CD163, HLA-G, and DNA (DAPI), used for cell type annotation, laser microdissection, and proteomic profiling. Box inserts correspond to panels c (1) and d (2). **(c,d)** Representative regions of interest manually selected in QuPath for isolation of immune (c) and trophoblast (d) compartments. CTB monolayer is defined by E-cadherin+ and HLAG-, whereas the multinucleated STB layer is E-cadherin- and HLAG- in the outer layer of the villi. CD163 expression marks macrophages (HBC). **(e)** Principal component analysis of cell type resolved proteomes from three placental tissues. Colors indicate cell types. Confidence ellipses (95% CI) computed from PC scores covariance. **(f)** Cell type-specific protein marker heatmap as inferred by ANOVA (FDR < 0.05) showing relative abundance profiles per replicate. Canonical markers for each phenotype highlighted in orange (HBC), green (STB), and light blue (CTB). Z-scored relative protein abundances are shown. Distance metric: "euclidean", method: "complete". **(g)** Comparison of flashDVP-generated phenotype-specific proteins and published genes marking CTB, STB, and HBC phenotypes. Shared and unique markers highlight the importance of proteomic data in complementing other omics. **(h)** Dotplot of enriched functional pathways of phenotype-separating clusters from (g). Enrichment based on ranked proteins using the Reactome database. Color shows adjusted p-values of overrepresented pathways and dot size indicates category sizes ("Count"). **(i)** Significant cell-cell communication in trophoblast compartment inferred using CellChat[36]. Dot colors show mean communication probability of ligand-receptor pairs in labelled pathway, size represents p-value from one-sided permutation test.

OpenDVP coupled to spatial transcriptomics uncovers niche-specific drug targets

We next tested whether flashDVP could facilitate cell-type resolved tissue proteomics by integrating additional spatial omics layers. We analyzed a serial FFPE tissue section from an early-stage lung adenocarcinoma that we previously characterized by spatial transcriptomics (ST, Nanostring CosMx) [37] (**Fig. 3a**). Using 960 cancer-related genes, we profiled over 340,000 cells, identifying 18 cell types and their spatial distribution. These rich data revealed diverse immune and tumor cell niches of therapeutic value, including a small tumor niche enriched for pro-metastatic mesenchymal tumor cells with high *NDRG1* and *LGALS1* expression, located near myofibroblasts and SPP1+ tumor-associated macrophages (TAMs) (**Fig. 3b**). Pseudotime analysis identified this 'epithelial to mesenchymal (EMT)-niche' as the potential start site of tumor invasion. We used the CosMx data to guide the collection of the EMT-niche tumor cells, and revealed their proteomes to understand their functional state and identify potential drug targets. IF staining of a serial tissue section against epithelial and mesenchymal markers, overlaid with CosMx phenotypes, confirmed the mesenchymal nature of EMT niche tumor cells on protein level (panCK+ and VIM+) (**Fig. 3c-d**). We next annotated multiple ROIs to proteomically profile tumor and stromal niches (**Extended Data Fig. 3a**). ROIs contained 50-100 cells (40,000 – 50,000 μm^2, 5 µm thick). We quantified 4,138 and 4,976 proteins in stromal and tumor compartments respectively, totaling 5,956 proteins (**Fig. 3e, Extended Data Fig. 3b, Extended Data Table 2**).

284 Group replicates clustered together and showed high proteome correlations (Pearson's r > 0.9,
285 **Extended Data Fig. 3c**). PCA analysis showed EMT-niche tumor cells and stroma-infiltrating tumor
286 cells were closer to stromal samples in component 1, reflecting their mesenchymal phenotype
287 (**Fig. 3f**). Integration with CosMx data confirmed that tumor cell-intrinsic signatures dominated
288 the EMT-niche tumor proteomes (**Fig. 3g-h**). Proteomics further validated the CosMx and IF data
289 showing panCK+, VIM+, and E-cadherin- tumor cells in the EMT-niche and upregulated NDRG1
290 and LGALS1 protein levels. Many genes identified by ST were also significantly upregulated at the
291 protein level in EMT niche tumor cells, emphasizing the value of spatial niche information in 3D
292 for aligning and integrating tissue sections for multi-omics analyses (**Fig. 3a**). However, MS-based
293 proteomics captured higher dynamic ranges of analyte abundances, resulting in different
294 gene/protein rankings and drug target prioritizations (**Fig. 3g-h, Extended Data Fig. 3d**). We
295 identified 180 FDA-approved drug targets of which 16 cancer-related genes were significantly
296 upregulated in EMT niche tumor cells (**Extended Data Fig. 3e, Extended Data Table 2**). For
297 example, *PDGFRB*, an actionable gene for imatinib, a receptor tyrosine kinase inhibitor approved
298 for various malignancies including gastrointestinal stromal tumors and myeloproliferative
299 neoplasms [38]. SLC2A1/GLUT1 was the most up-regulated protein in EMT niche tumor cells.
300 Consistently, pathway analysis revealed high glycolysis expression in EMT niche tumor cells, along
301 with hypoxia, angiogenesis, and EMT pathways (**Extended Data Fig. 3f**). As targeting glucose
302 metabolism is currently under clinical investigation for cancer therapy [39,40], we assessed the
303 tumor-specific expression of GLUT1 and its association with patient outcomes. We validated high
304 tumoral GLUT1 expression in a large-scale transcriptome study of 483 lung adenocarcinomas and
305 347 control samples [41,42] (**Fig. 3i)** as well as on IHC data obtained from the human protein atlas [43]
306 (**Fig. 3j)**. High *GLUT1* expression was a strong prognostic factor associated with poor patient
307 outcome, as revealed by the significantly different overall survival (p = 0.0003, n=239 per group)
308 (**Fig. 3k**). Notably, re-evaluation of the spatial transcriptomic data not only confirmed the tumor-
309 specific expression of *GLUT1*, but also emphasized its strong EMT-niche specificity (**Extended**
310 **Data Fig. 3g**). Together, these results underline the benefit of cellular niche-guided spatial
311 multiomics as a promising approach for identifying and validating niche-specific, personalized
312 drug targets.

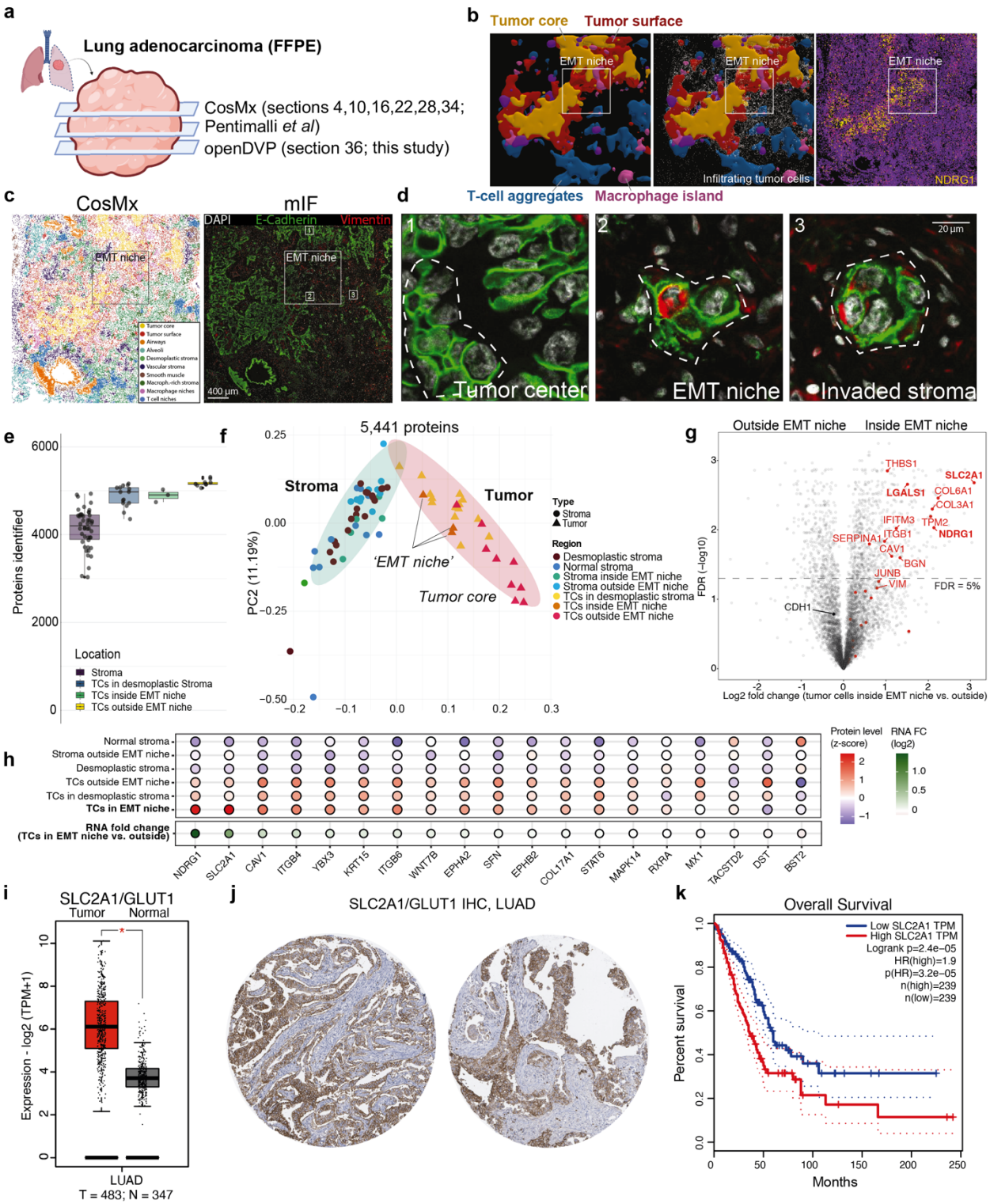


Fig. 3: Cell-type resolved deep proteomic profiling of transcriptionally defined cellular niches

(a) Overview of 5 µm FFPE lung adenocarcinoma sections showing regions for CosMx spatial transcriptomics and DVP analysis. **(b)** 3D reconstruction of transcriptomically defined niches, showing tumor-infiltrating cells. NDRG1 RNA level was upregulated in tumor cells from the surface and EMT niche, as previously published by us [37]. **(c)** Left: 2D multicellular niche map from CosMx data (Section 34[37]). Right: immunofluorescence of epithelial (E-cadherin,

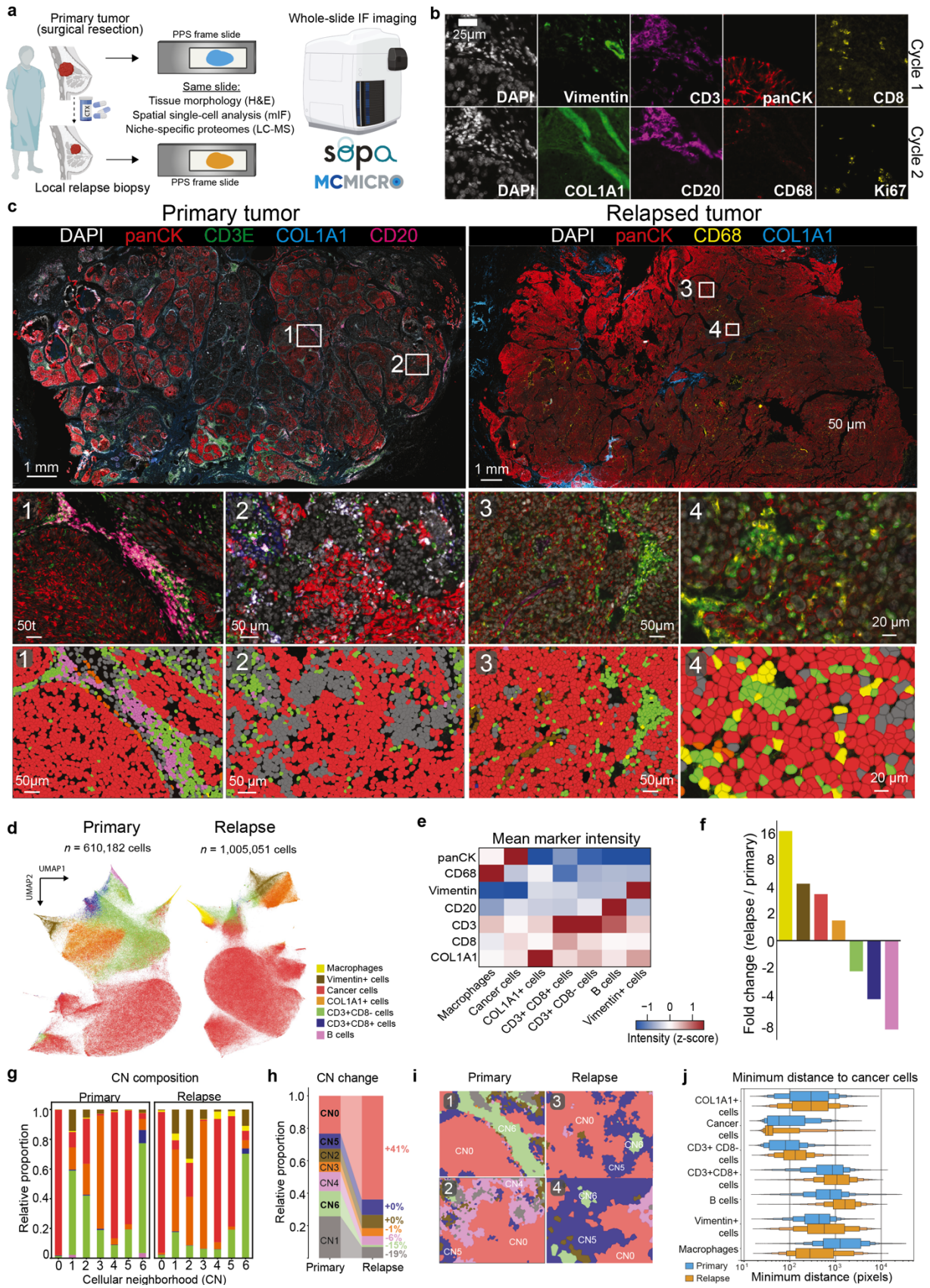
green) and mesenchymal (vimentin, red) markers in section 36. Insets 1–3 correspond to panel (d). Nuclei with DAPI (white). Scale bar: 400 µm. **(d)** Magnified views: [1] tumor center (E-cadherin+/Vimentin-), [2] EMT niche (E-cadherin+/Vimentin+, arrow), and [3] invading tumor cells (E-Cadherin+/Vimentin+, arrow). Scale bar: 20 µm. **(e)** Proteomics identified a mean of 4,138 proteins in stroma and 4,976 in tumor regions (50–100 cells/sample). **(f)** Principal component analysis of 5,441 proteins shows a separation of tumor and stroma samples along PC1. EMT niche tumor samples cluster near invasive and stromal samples. **(g)** Volcano plot showing differential protein abundance in tumor cells inside vs. outside the EMT niche. CosMx-upregulated genes [37] are shown in red. Proteins above the dashed line have a permutation-based false-discovery rate (FDR) below 0.05. **(h)** Relative protein levels of top upregulated genes in EMT niche tumor cells. **(i)** Mean RNA expression of *SLC2A1*/*GLUT1*, the most EMT-niche upregulated tumor protein in the proteomics data, in lung cancer cohort (n = 483 tumor, 347 healthy; TCGA LUAD) shows higher expression in cancer tissues [41]. **(j)** GLUT1 IHC staining of lung adenocarcinoma tissues confirmed tumor-specific expression. Images (left: ID 426; right: ID 1303) from human protein atlas [43]. **(k)** Kaplan–Meier analysis showed lower overall survival in patients with high *SLC2A1*/*GLUT1* RNA expression, grouped by median. For boxplots in **(e)** and **(i)** the central lines indicate the median, the box the interquartile range, whiskers extend to (1.5*IQR) of the box.

OpenDVP reveals tumor microenvironment remodeling in primary and relapsed TNBC

Current DVP implementations enable the linkage of visually and machine learning defined tissue phenotypes to quantitative proteomic profiles but involve trade-offs between spatial resolution, proteome depth, and experimental throughput. Our original DVP concept [8] achieves workflow robustness and high proteome coverage by pooling phenotypically coherent cell populations, whereas single-cell DVP (scDVP) [44] maximizes spatial resolution at the expense of analysis depth and ease of use. These limitations motivate alternative strategies that preserve spatial interpretability while substantially increasing throughput and discovery potential.

Recent advances in spatial biology have established cellular neighborhoods (CNs) as fundamental organizational units that shape tissue function, disease progression, and therapeutic response [45–47]. Triple-negative breast cancer (TNBC), the most aggressive breast cancer subtype [48], exemplifies how CN organization within the tumor microenvironment (TME) correlates with clinical outcomes [49], including critical roles for macrophage- and fibroblast-enriched niches [50–52]. Resolving CN-specific proteomes therefore offers an opportunity to identify molecular mediators of cell–cell interactions and disease progression. To enable CN-resolved proteome profiling, we first established an optimized protocol for centimeter-scale whole-slide multiplex immunofluorescence (mIF) imaging on PPS membrane slides, using a nine-marker panel for this proof-of-concept study (**Fig. 4a-b**). We analyzed paired TNBC tissue sections from the primary

354 surgical resection and a matched local relapse obtained two years after adjuvant chemotherapy.
355 The primary specimen contained extensive ductal carcinoma in situ (DCIS; 38 mm) with focal
356 invasion into the surrounding stroma, whereas the relapse consisted of a large (45 mm) invasive
357 carcinoma without residual DCIS (**Fig. 4c**). Each whole-slide image stack (>150 GB) was processed
358 using MCMICRO and SOPA for cell segmentation, phenotyping, and neighborhood analysis, with
359 matched H&E sections providing additional morphological context (**Extended Data Fig. 1a-c**). We
360 segmented 610,182 and 1,005,051 cells from the primary and relapse samples, respectively,
361 identifying five major cell types —tumor, T cells, B cells, macrophages, and stromal cells (**Fig. 4c–**
362 **e; Extended Data Fig. 4a-d**). While panCK$^+$ tumor cells dominated both tissues, marked shifts in
363 immune and stromal composition were observed (**Fig. 4f; Extended Data Table 3**). Compared
364 with the primary tumor, the relapse exhibited pronounced depletion of B and CD8$^+$ T cells
365 alongside a 16-fold increase in CD68$^+$ macrophages and a fourfold expansion of VIM$^+$ stromal
366 cells, consistent with immune evasion and tumor progression [53,54]. Cellular neighborhood analysis
367 using scimap's spatialLDA identified seven recurrent CNs (CN0–6) with distinct cellular
368 compositions (**Fig. 4g; Extended Data Table 3**). Several CNs showed strong abundance
369 differences between the two samples (**Fig. 4g–i**), accompanied by spatial reorganization of
370 immune cells. Proximity analysis revealed that macrophages were located closer to tumor cells
371 in the relapse, whereas lymphocytes preferentially surrounded DCIS regions in the primary tumor
372 (**Fig. 4j**), indicating a transition from a lymphocyte-rich peritumoral niche to a macrophage-
373 dominated intratumoral microenvironment.


a
Primary tumor (surgical resection)
PPS frame slide
Same slide:
Tissue morphology (H&E)
Spatial single-cell analysis (mIF)
Niche-specific proteomes (LC-MS)
Local relapse biopsy
PPS frame slide
Whole-slide IF imaging
sepa
MCMICRO
b
25µm
DAPI
Vimentin
CD3
panCK
CD8
Cycle 1
DAPI
COL1A1
CD20
CD68
Ki67
Cycle 2
c
Primary tumor
DAPI panCK CD3E COL1A1 CD20
1 mm
Relapsed tumor
DAPI panCK CD68 COL1A1
50 µm
1 mm
50µm
20 µm
d
Primary
n = 610,182 cells
Relapse
n = 1,005,051 cells
UMAP1
UMAP2
Macrophages
Vimentin+ cells
Cancer cells
COL1A1+ cells
CD3+CD8- cells
CD3+CD8+ cells
B cells
e
Mean marker intensity
panCK
CD68
Vimentin
CD20
CD3
CD8
COL1A1
Macrophages
Cancer cells
COL1A1+ cells
CD3+ CD8+ cells
CD3+ CD8- cells
B cells
Vimentin+ cells
Intensity (z-score)
f
Fold change (relapse / primary)
g
CN composition
Primary
Relapse
Relative proportion
Cellular neighborhood (CN)
h
CN change
Relative proportion
CN0
CN5
CN2
CN3
CN4
CN6
CN1
+41%
+0%
+0%
-1%
-6%
-15%
-19%
Primary
Relapse
i
Primary
Relapse
CN0
CN6
CN4
CN5
j
Minimum distance to cancer cells
COL1A1+ cells
Cancer cells
CD3+ CD8- cells
CD3+CD8+ cells
B cells
Vimentin+ cells
Macrophages
Primary
Relapse
Minimum distance (pixels)

374

375 **Fig. 4. Image acquisition and analysis of primary and recurrent triple-negative breast cancer (TNBC).**
376 **(a)** FFPE 5 µm sections from primary tumor and same-patient post-chemotherapy relapse sample were mounted on
377 PPS frame slides. **(b)** Slides underwent multiplex immunofluorescence (mIF) staining and processing using MCMICRO
378 pipeline with SOPA. **(c)** Whole-slide mIF images of samples with zoomed regions show overlaid cell phenotype
379 information. **(d)** UMAP embeddings of segmented cells from each slide, based on morphology and marker intensity
380 (excluding nuclear stains and background); unidentified cells excluded; n_neighbors = 25. **(e)** Heatmap of z-scored
381 mean marker intensities across phenotyped cells. Phenotyping performed by thresholding markers per slide,
382 rescaling intensities, and scimap phenotype function. **(f)** Fold change in cell type counts between recurrent and
383 primary samples. **(g)** Relative composition of cellular neighborhoods per sample excluding unidentified cells. h)
384 Selected regions with cellular neighborhoods overlaid. **(i)** Stacked bar plots showing cellular neighborhood frequency
385 in each sample, calculated using scimap's spatialLDA function. **(j)** Boxen plots of minimum distance from cells to
386 nearest cancer cell. The central line indicates the median, while the first box represents the interquartile range (IQR,
387 25th to 75th percentiles), each subsequent level of boxes corresponds to half the previous number of datapoints.
388

389 Cellular neighborhood–guided DVP enhances proteome coverage and scalability

390 Motivated by these spatial differences, we hypothesized that cellular neighborhood–guided DVP
391 (cnDVP) could enable scalable spatial proteomics while preserving the core DVP principle of
392 linking visual phenotypes to quantitative proteomic profiles. To systematically evaluate this
393 approach, we compared cnDVP with conventional DVP [8] and scDVP [44] using matched regions from
394 the primary and relapse samples. To minimize spatial confounders, we profiled directly adjacent
395 cells of identical phenotype and CN assignment (**Fig. 4e,g; Fig. 5a**). For DVP, we excised 100
396 individual cell shapes per well, whereas cnDVP required only a single shape encompassing 100
397 phenotyped cells, resulting in an approximately 100-fold faster tissue collection. For scDVP, we
398 collected individual tumor or immune cells. As cell size largely determines proteome coverage [55],
399 scDVP has remained challenging for small cell types such as immune cells (**Extended Data Fig.
400 5a**). From 50 single immune cells (CD3+/CD8- lymphocytes or CD68+ macrophages) only three
401 samples remained after filtering for more than 400 proteins. For cancer cells, we identified up to
402 1,500 proteins per shape (**Fig. 5b**). CnDVP achieved the highest proteome coverage in both tumor
403 and immune niches, with a 26% increase in protein identifications relative to DVP and the largest
404 number of uniquely detected proteins (**Fig. 5b-c**). The majority of proteins were consistently
405 detected across all approaches (**Extended Data Fig. 5b**), and tissue- and niche-specific
406 quantitative signatures were highly correlated (**Fig. 5d-e, Extended Data Fig. 5c-e**). These results
407 establish cnDVP as an efficient and generalizable strategy that substantially improves throughput
408 while largely preserving cell-type specificity and spatial resolution.

Conceptually, cnDVP expands the DVP method family implemented in openDVP, enabling users to select among complementary profiling strategies based on tissue architecture, biological question, and experimental constraints (**Fig. 5f**).

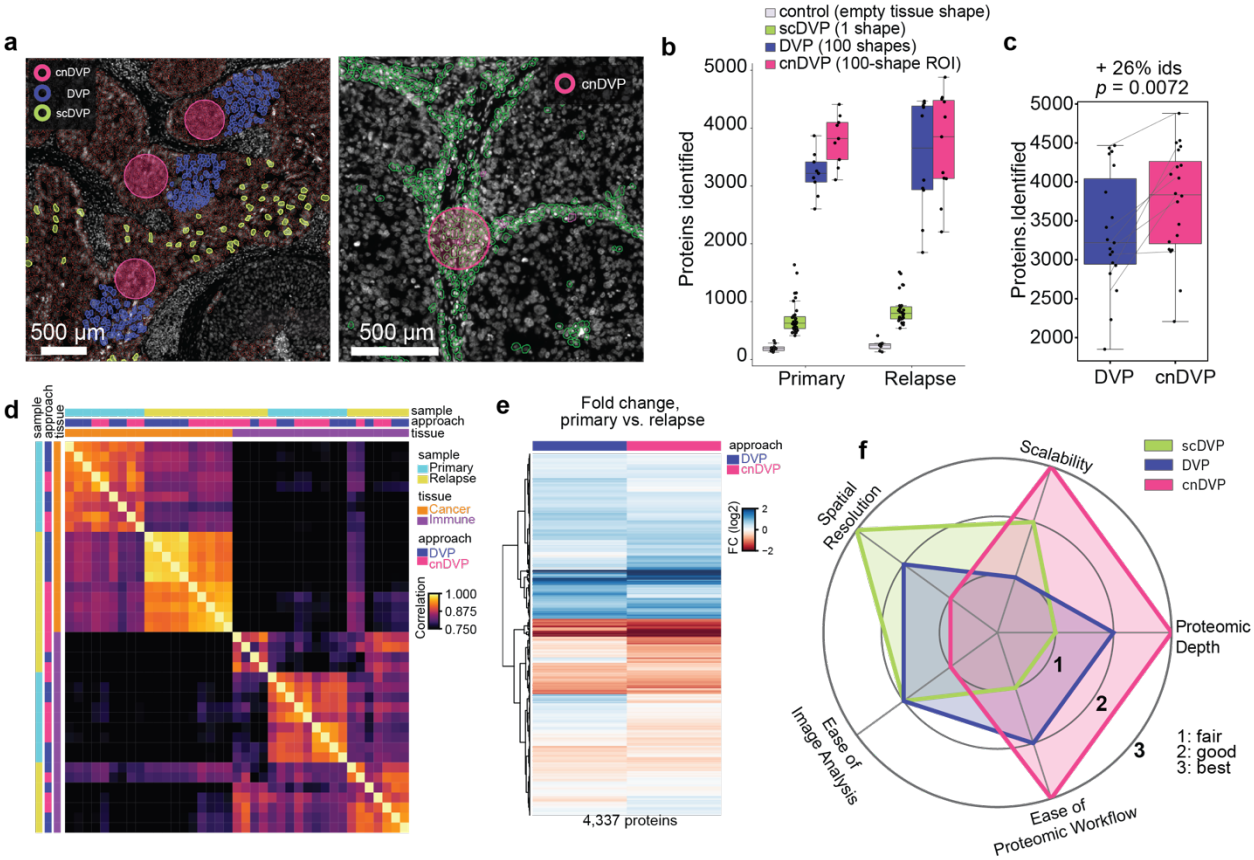


Figure 5: Cellular neighborhood guided DVP improves proteomic depth and throughput.

(a) Images of primary TNBC tissue, showing the nuclear stain with segmentation masks for phenotyped cancer cells with red outlines on the left and CD3+ cells with green outlines on the right. Overlayed are collection annotations for the three DVP approaches. In magenta cnDVP, in blue DVP, and green for scDVP. **(b)** Boxplots with number proteins identified with DIANN search across DVP approaches. **(c)** Boxplot with number of proteins identified in matching cancer-specific samples analyzed by DVP and cnDVP. Lines denote spatial co-localization, showing an increase of proteome coverage using the cnDVP approach. P-value is based on a two-sided paired T-test between paired samples. For both b and c, the central lines indicate the median, while the box represents the interquartile range (IQR, 25th to 75th percentiles), whiskers extend to 1.5* IQR of the box **(d)** Proteome correlation heatmap using Spearman correlations for all immune and cancer samples analyzed with DVP and cnDVP. Metric: "euclidean" and method: "ward" **(e)** Heatmap of log2FC values for cancer samples between the primary and relapse breast cancer tissues across DVP and cnDVP approaches. Note, similar fold changes were retrieved for both approaches. **(f)** The radar plot summarizes the key differences between three DVP approaches; scDVP (light green), DVP (blue), and cnDVP (pink); across five key properties: Scalability, Proteomic Depth, Ease of Proteomic Workflow, Ease of Image Analysis, and Spatial Resolution. The radial axes represent a qualitative ranking scale where 1 = "fair", 2 = "good", and 3 = "best".

429 **CN-guided DVP delineates spatial tumor heterogeneity and molecular drivers of**
430 **relapse**

431 To investigate how CN-resolved proteomes reflect tumor progression and treatment resistance,
432 we performed CN-guided LMD sampling for exploratory proteomics (**Fig. 6a**), focusing on tumor-
433 specific (CN0), tumor–immune interface (CN5), and immune-enriched (CN6) regions. Regions of
434 interest were distributed across tissue sections to capture spatial heterogeneity (**Fig. 6b**), yielding
435 203 low-input samples after QC filtering, and processed within one week using an advanced
436 throughput (80 samples per day) DIA-based workflow. We quantified ~4,800 proteins in tumor-
437 specific CNs and 3,500–4,000 proteins in immune-enriched niches (**Extended Data Fig. 6a-b**).
438 Proteomes exhibited strong CN specificity, with shared core proteins and niche-restricted
439 signatures (**Fig. 6c**). Integration with SpatialData and public scRNA-seq datasets [56] confirmed
440 epithelial tumor signatures in CN0 and immune programs in CN6 (**Fig. 6d**). Principal component
441 analysis separated proteomes by tissue origin and spatial niche (**Fig. 6e-f**), and unsupervised
442 clustering identified distinct proteomic programs associated with primary versus relapsed tumor
443 samples (**Fig. 6g-h**). Tumor-specific CNs showed significantly higher proteomic heterogeneity in
444 the primary sample than in the relapse, as reflected by lower global correlation, increased protein
445 variability, and stronger spatial autocorrelation (**Fig. 6i,j; Extended Data Fig. 6c–e**). This
446 observation also transcended to single tumor shapes (**Extended Data Fig. 6e**), possibly reflecting
447 reduced intratumoral diversity through clonal selection following chemotherapy [57,58].
448 Differential abundance analysis revealed pronounced niche-specific alterations in the relapse
449 sample. Immune niches exhibited upregulation of interferon-γ signaling, complement, and heme
450 metabolism pathways (**Fig. 6k-l**), alongside enrichment of myeloid/macrophage programs
451 (**Extended Data Fig. 6f**). Tumor niches displayed a metabolic shift toward glycolysis and activation
452 of EMT, MYC, and interferon signaling pathways (**Fig. 6n**), hallmark features of chemoresistance
453 and progressive disease[59,60]. Spatial mapping revealed widespread interferon-high tumor
454 phenotypes in the relapse but not in the primary tumor (**Fig. 6o**), consistent with molecular
455 features of inflammatory TNBC and tumor-immune evasion [61,62].

456 Together, these results demonstrate how openDVP integrates whole-slide imaging, cellular
457 neighborhood analysis, and exploratory proteomics to resolve spatial tumor heterogeneity and
458 uncover potential drivers of disease progression.

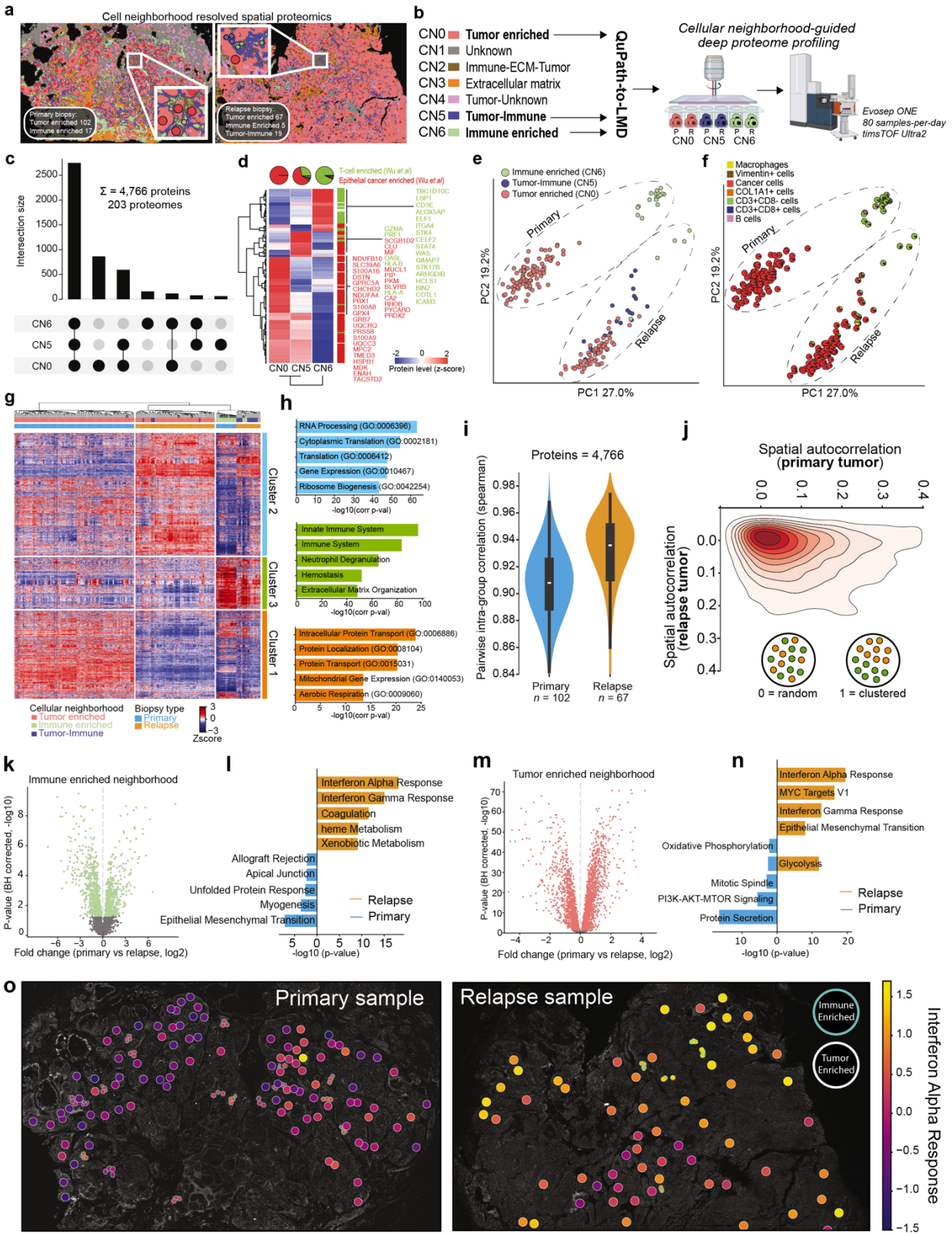


Figure 6. Spatial proteomic analysis of triple-negative breast cancer (TNBC) samples comparing primary and recurrent disease. (a) Whole-slide overlay of cellular neighborhood (CN) annotations with contours for laser

463 microdissection-guided proteomics. Sample collection summaries are shown in boxes. **(b)** Three selected CNs for
464 proteomics: Tumor Enriched, Immune Enriched, and Tumor-Immune. Regions were collected using laser
465 microdissection and analyzed using a rapid (80 samples-per-day) LC-MS strategy. **(c)** UpSet plot showing protein
466 identification overlap across CNs. **(d)** Heatmap showing signature markers from a recent RNAseq study [56]. Protein
467 levels were z-scored, averaged per CN, and subjected to unsupervised hierarchical clustering (method: "ward",
468 distance metric: "euclidean"). **(e)** Principal component analysis (PCA) of all samples, colored by CN. **(f)** PCA
469 embedding with samples as pie charts reflecting cell phenotype proportions derived from image analysis. Note, both
470 PCA plots show clustering by sample type (primary vs. relapse), followed by CN. **(g)** Heatmap displaying 4,570
471 proteins across 203 samples, clustered hierarchically (method: "ward", distance metric: "euclidean"). Proteins
472 filtered for 70% valid values in any CN group. **(h)** Protein clusters analyzed using EnrichR pathway enrichment using
473 all identified proteins as background. **(i)** Violin plots showing pairwise intragroup correlations, highlighting greater
474 proteomic variability in the primary tumor. P-value <0.001, Mann-Whitney two-sided test. Central lines indicate the
475 median, box represents the interquartile range, whiskers extend to 1.5* IQR of the box **(j)** Density plot of Moran's I
476 spatial autocorrelation for proteins across samples, showing greater spatial variability in the primary tissue. P-value
477 <0.001, Mann-Whitney two-sided test. **(k)** Volcano plot showing regulated proteins between immune-enriched CNs
478 of primary and relapse tumors, FDR cut-off of 5%. **(l)** Barplot showing enriched MsigDB hallmarks between primary
479 and relapse immune-enriched samples. **(m)** Volcano plot showing regulated proteins between tumor-enriched CNs
480 of primary and relapse samples, FDR cut-off = 5%. **(n)** Barplot showing enriched MsigDB hallmarks between primary
481 and relapse tumor enriched samples. **(o)** Overlay of interferon alpha response pathway (Hallmarks) showing z-scored
482 pathway levels of tumor and immune-enriched samples. Note, the relapse sample featured consistently higher IFN
483 levels across tissue. Large circles represent tumor specific CNs (white outline) and small circles immune specific CNs
484 (cyan outline).

485 Discussion

486 OpenDVP is the first Python-based foundational platform designed to make deep visual
487 proteomics broadly accessible and community driven. Built on the scverse ecosystem, it adheres
488 to FAIR principles by ensuring that processed data are findable, accessible, interoperable, and
489 reusable. Its modular architecture enables seamless integration of established tools for image
490 segmentation, cell phenotyping, and spatial analysis, including MCMICRO, QuPath, and Napari,
491 thereby supporting flexible and reproducible end-to-end DVP workflows.
492 For broad accessibility, openDVP provides two complementary workflow options. The first,
493 flashDVP, implements a simplified 'annotate–collect–measure' strategy that enables rapid spatial
494 discovery within days. This approach is particularly well suited for exploratory studies of acellular
495 regions such as the extracellular matrix, or tissue contexts where single-cell phenotyping is less
496 applicable. Applied to human placental tissue, flashDVP enabled comprehensive cell-type–
497 resolved proteomic profiling, identifying pathways critical for placental biology. In an aggressive
498 lung cancer previously analyzed by spatial transcriptomics, flashDVP confirmed transcriptional

499 findings at the protein level and uncovered additional niche-specific drug target candidates,
500 illustrating the complementarity of multimodal spatial omics.
501 The second workflow represents an AI-powered DVP pipeline optimized for high-performance
502 computing and large-scale imaging data. This pipeline integrates state-of-the-art tools for
503 memory-efficient single-cell and spatial analyses and was applied to study tumor relapse in triple-
504 negative breast cancer. While our original DVP workflow and single-cell DVP offer maximal
505 biological granularity, indispensable for certain biomedical applications [44,63,64] , they remain
506 technically challenging due to gravity-based collection constraints, contour alignment errors, and
507 potential cell-type admixture during single-cell LMD. These limitations motivated the
508 development of an alternative strategy that balances biological resolution with scalability,
509 robustness, and experimental efficiency. By operating at the level of cellular neighborhoods,
510 cnDVP overcomes throughput limitations of existing DVP workflows while preserving spatial and
511 cell-type specificity. We anticipate cnDVP as particularly useful for homotypic niches. For
512 heterotypic niches, such as the case for tumor and immune cell co-localization, cnDVP could offer
513 a direct and quantitative approach for protein-centric ligand-receptor analyses. Applied to a
514 TNBC relapse case, cnDVP revealed extensive tumor microenvironment remodeling and
515 pronounced proteomic differences between primary and relapse tumors. Our data revealed
516 elevated interferon signaling, EMT, and glycolytic programs in relapse-associated niches -
517 features linked to tumor progression and chemoresistance and potentially targetable through
518 JAK/STAT inhibition or metabolic interventions [40,65,66].
519 Looking ahead, we anticipate that cellular neighborhood–guided proteome profiling will be
520 particularly powerful when combined with higher-plex DVP assays [67] (e.g., 20–30 markers),
521 substantially reducing undetected phenotypes and enabling finer-resolution dissection of tissue
522 architecture and disease-associated states.
523
524 In summary, openDVP represents a significant methodological advance in spatial proteomics by
525 providing an open, extensible, and scalable platform for deep visual proteomics. As imaging, AI,
526 and proteomics technologies continue to mature, widespread adoption of openDVP across
527 tissues and disease contexts has the potential to accelerate biological discovery, enable

systematic biomarker identification, and uncover novel therapeutic targets. Community-driven efforts built on openDVP will further facilitate the development of centralized spatial proteomics resources that integrate antibody- and mass spectrometry–based data, ultimately advancing our understanding of tissue biology in health and disease.

Acknowledgements

We thank our colleagues at the Max Delbrück Center (MDC) and Charité for their support and fruitful discussions. Furthermore, we acknowledge the MDC technology platform 'Proteomics' and 'Advanced light microscopy' for their great support. Gina Dörpholz, Pia Larsen and Jeannine Engel for administrative support and Janett König for laboratory support. Wouter-Michiel Vierdag for help with Napari plugin development and SpatialData. The authors also gratefully appreciate Martin Gauster (Medical University of Graz) and Andreas Glasner (Femina-Med) for patient recruitment collection of first-trimester human placental tissue (Medical University of Graz). J.N. and F.C. and S.S. acknowledge funding support by the Federal Ministry of Education and Research (BMBF), as part of the National Research Initiatives for Mass Spectrometry in Systems Medicine, under grant agreement No. 161L0222 and 03LW0239K. F.H. and F.C. were supported by Deutsche Forschungsgemeinschaft (HE 6249/5-3). This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101115681) and support by the ERC (ERC starting grant). S.Fl. was supported by grant 01ZX1917A from the Federal Ministry for Education and Research (BMBF) through the e:Med Initiative for Systems Medicine.

Author contributions

Conceptualization: J.N. and F.C.

Methodology: J.N., S.F., M.K., and M.T.

Resources: T.P., S.Fl., S.S., F.H., and F.K.

Experiments, J.N., S.F., M.K., and D.S.V.

Data curation: J.N., S.F., M.T., and D.S.V.

Data analysis: J.N., S.F., D.S.V., T.P., and F.C.

Figures: J.N., S.F., D.S.V., T.P., and F.C.

Project administration: J.N. and F.C.

Supervision: F.H, S.Fl., N.R., and F.C.

Funding acquisition: N.R., F.H., F.K, and F.C.

Writing of the original draft: F.C. All authors have reviewed and edited the manuscript.

Declaration of interests

The authors declare that they have no competing interests.

566 # Methods

567 We structured the Methods section by analytical process to clarify workflow differences among
568 the three exemplary applications. Detailed information on data sources, image preprocessing,
569 proteomic analysis, and all other relevant settings is provided under dedicated subtitles for each
570 vignette within these process-based descriptions.
571
572 **Sample collection and patient cohort**
573 Placenta samples
574 Placental tissue was collected from electively terminated pregnancies (gestational age, 7 – 11
575 weeks) with informed consent. Exclusion criteria were maternal age under 18 years, BMI >25
576 kg/m2, and maternal pathologies. Ethical approval was obtained from Medical University Graz
577 Ethics Committee (31-019 ex18/19). After surgical extraction, tissue was stored at 4°C in culture
578 medium (DMEM/F12 1:1, 1 g/dL glucose) and processed within 4 h. Villous tissue was rinsed in
579 cold (4°C) 0.9% NaCl solution to remove blood, fixed with 10% formalin, paraffin embedded
580 (FFPE), and dehydrated per standard protocols.
581
582 Lung cancer sample
583 We analyzed the primary tumor of a non-small cell lung cancer (NSCLC) patient. The patient, a 63-
584 year-old woman with a 40 pack-year smoking history and no physical limitations (ECOG
585 performance status 0), was diagnosed in March 2020 with a metabolically active tumor at the
586 apex of the right upper lobe on positron emission tomography (PET) imaging. A transbronchial
587 lung biopsy confirmed a TTF1-positive lung adenocarcinoma (LUAD) with acinar growth pattern.
588 Two months after diagnosis, the patient underwent right upper lobectomy. Gross examination
589 was performed according to standardized protocols. The specimen was fixed in 10% buffered
590 formalin and paraffin-embedded (FFPE). Histological examination, including diagnosis, tumor
591 grading, pTNM classification, angioinvasion, lymphatic invasion, and tumor stage, was performed
592 according to the 8th edition of the TNM classification (AJCC). Following diagnosis, the FFPE tissue
593 block was stored at room temperature in the archive of the Institute of Pathology at the Charité

594 University Hospital, Campus Mitte. The study was performed according to the ethical principles
595 for medical research of the Declaration of Helsinki, and approval was approved by the Ethics
596 Committee of the Charité University Medical Department in Berlin (EA4/243/21).
597 A comprehensive 3D spatial transcriptomic analysis of this tumor has been described recently
598 based on 34 5-µm-thick consecutive sections (sections 4, 10, 16, 22, 28, and 34) [37] and a section-
599 to-section distance of 30 µm. From the same tumor block, we profiled section 36 using
600 proteomics after image registration in QuPath.
601
602 Triple negative breast cancer samples
603 Archival leftover diagnostic tissue from the tumor resection specimen and a specimen obtained
604 two years later after relapse was selected from the archives of the Institute of Pathology, Charité
605 Universitätsmedizin Berlin. At the time of diagnosis and primary resection, the patient had
606 carcinoma in situ (DCIS), 38 mm in diameter, with multifocal stromal invasion. The multifocal
607 invasive tumor had a total diameter of 9 mm, lacked expression of estrogen and progesterone
608 receptors, and had an Her2-score of 1+, considered equivalent to a lack of HER2-gene
609 amplification. The diagnosis was multifocal invasive triple-negative breast cancer (stage pT1b
610 (m)). In 2022, the patient experienced a relapse with a diameter of 34 mm and ulceration of the
611 skin (stage yrpT4b). The patient provided written consent for the use of leftover diagnostic
612 material for research purposes. The project was approved by the IRB of Charité ("Ethics
613 Committee"), project number EA1/253/19.
614
615 Membrane slide preparation
616 Prior to tissue mounting, slides were washed with 70% EtOH for 5min, rinsed in ddH2O and
617 incubated with poly-L-lysine at 0.01% (w/v) (Sigma-Aldrich, cat.no. P1524) for 10min to optimize
618 tissue adherence. Slides were then dried overnight at room temperature. All samples were
619 sectioned at a thickness of 5 µm and mounted on metal frame PPS membrane slides (Leica,
620 cat.no. 11600294), followed by overnight drying at 37 °C. Immediately prior to deparaffinization,
621 slides were incubated at 60 °C for 30 min.
622
623

624 **Cyclic and non-cyclic immunofluorescence staining and imaging**

625 Placenta tissue

626 Placenta tissue sections were deparaffinized, and subjected to antigen retrieval in Tris/EDTA (pH
627 9) at 93 °C for 20 min. After cooling (20 min), slides were placed in warm distilled water (5 min)
628 and cooled again (5 min). Sections were washed with PBST (PBS + 0.1% Tween 20) and blocked
629 with Ultra V Block (20 min, RT). For triple staining (CD163, CD31, E-cadherin), primary antibodies
630 (Table 1) were diluted in PBST + 1% normal goat serum (NGS) and incubated overnight at 4°C.
631 fter washing, secondary antibodies in PBST + 1% NGS (mouse IgG-AF647 and rabbit IgG-AF488,
632 Table 1) were incubated for 1 h at RT in the dark. Slides were washed and mounted with SlowFade
633 Diamond containing DAPI. Rabbit immunoglobulin and mouse IgG controls showed no stains.
634 Imaging was performed using Zeiss Axioscan 7.

635

636 Lung and breast cancer tissue

637 Breast cancer and lung cancer tissue sections were deparaffinized and rehydrated (2x 5 min in
638 Neo-Clear, 2x 2 min in 99% EtOH, 1x 2 min in 80% EtOH, 1x 2 min in 70% EtOH, 3x 1 min in 1x
639 PBS). For decrosslinking, heat-mediated antigen retrieval was performed in pH 9 Tris buffer
640 (Dako, #K800421-2, 1:50 in ddH_2O) in a steamer for 30min. Sections were washed in 1x PBS and
641 pre-quenched in 4.5% H_2O_2 (in 25mM NaOH in PBS) for 2x 30 min. Cover glasses were mounted
642 with 10% glycerol (in PBS) and the first image acquisition for background subtraction was
643 performed using the Zeiss Axioscan 7 Slidescanner with a 20x/0.5NA objective at 2 × 2 binning.
644 Slides were soaked in PBS for cover glass detachment for 5 min and then washed 3x 5 min in PBS
645 to remove residual glycerol. Tissues were then blocked for 30 min in 3% BSA (in PBS, Serva
646 #11948.01) at room temperature, followed by antibody staining in a humid chamber at 4 °C
647 overnight. Antibodies are listed in the table below, all diluted 1:50 in the blocking buffer. The
648 next day, sections were washed in PBS, mounted, and imaged with similar settings, while
649 exposure times were adjusted to the fluorescent signal. For cyclic (multiplex)
650 immunofluorescence of breast cancer samples, cover glasses were removed as before, sections
651 were washed, and fluorescent signal was bleached as previously described for 30 min. Sections
652 were washed 3x 5 min in PBS, and the second cycle of antibody staining was performed.

653 Afterwards, sections were incubated for 3 min in hematoxylin, 10 min in tap water, dipped twice
654 in ddH_2O, 30sec in eosin, dipped twice in ddH_2O, and dehydrated by dipping in increasing
655 concentrations of EtOH (70-99%) for visible staining.
656 Immunofluorescent staining of lung cancer samples was performed in a single cycle followed by
657 visible staining explained as described above.
658

Antibody target protein	Vendor	Catalog no.	Clone	Fluorophore	Research resource identifier	Dilution
α-SMA	Abcam	ab184675	1A4	Alexa Fluor® 488	AB_2832195	1:50
CD3E	NovusBio	NBP2-54392AF647	C3e/1308	Alexa Fluor® 647	AB_3095415	1:50
CD8A	Thermo Fisher Scientific	53-0008-80	AMC908	Alexa Fluor® 488	AB_2574412	1:50
CD20	eBioscience	53-0202-80	L26	Alexa Fluor® 488	AB_10734357	1:50
CD68	CST	79594	D4B9C	PE	AB_2799935	1:50
E-cadherin	CST	3199	24E10	Alexa Fluor® 488	AB_10691457	1:50
Pan Cytokeratin	Thermo Fisher Scientific	41-9003-80	AE1/AE3	eFluor™ 570	AB_11218704	1:50
Vimentin	NovusBio	NBP1-97670AF750	V9	Alexa Fluor® 750	AB_10013743	1:50
COL1A1	NovusBio	NB600-408AF750	Polyclonal	Alexa Fluor® 750	AB_10000511	1:50
Ki67	CST	11882	D3B5	Alexa Fluor® 488	AB_2687824	1:50
CD163	Miltenyi	130-127-908	REA1309	PE	AB_2928274	1:50
E-cadherin	Cell Signalling	3195	24E10	-	AB_2291471	1:200
HLA-G	EXBIO	11-291-C100	MEM-G/1	-	AB_10734353	1:200
Anti-goat Ms IgG	Thermo Fisher Scientific	A21235	-	Alexa Fluor® 647	AB_2535804	1:1000
Anti-goat Rb IgG	Thermo Fisher Scientific	A11008	-	Alexa Fluor® 488	AB_143165	1:1000

659
660
661
662

663 **Image analysis and contour export for laser microdissection**

664 Lung cancer

665 Immunofluorescent and H&E staining were overlayed using the Warpy extension (v0.2.6) in
666 Qupath (v0.4). ROI selection was based on panCK, Vimentin and E-Cadherin staining. Annotations
667 were exported as geoJSON files and converted to the LMD compatible format (.xml) using the
668 QuPath-to-LMD webapp.

669

670 Triple negative breast cancer

671 Triple-negative breast cancer whole slide images were processed with MCMICRO: Tiles were
672 illumination corrected with BASIC v1.1.1, then stitched and registered with ASHLAR v1.17.1.
673 Channels were background-subtracted from autofluorescence images using Backsub v0.4.1. The
674 image stack was processed using a modified SOPA snakemake pipeline. The nuclear signal from
675 the first cycle was segmented using cellpose v2's "nuclei" model, with parameters: diameter 25,
676 flow_threshold 0.8, cellprob_threshold –8, min_area 250, clip_limit 0.2, gaussian_sigma 1. For
677 whole-slide images, we used SOPA's patchify function to tile images into 5000 pixels tall/wide
678 squares with 100 pixel overlap to distribute segmentation tasks across HPC. Segmentation masks
679 were expanded by five pixels and used to quantify mean, standard deviation, and three quantiles
680 (25% 50% 75%) of marker signal per cell. For quality control, image stacks were scaled to 8-bit
681 and pyramidized with tile-size 4096. The following additional filtering steps were applied with
682 visual quality control: (1) removal of manually labelled artifact regions, (2) removal of cells with
683 too low/high nuclear signal, (3) removal of cells with too small/large areas, (4) removal of dropout
684 cells where DAPI ratio between first and last cycle was not between 0.15-1.05. Cell phenotyping
685 was done using the napari-cell-gater (github.com/CosciaLab/napari-cell-gater) to determine
686 marker thresholds and scale marker signals for binomial distribution. The plugin allows the user
687 to load an image, segmentation mask, and quantification matrix of any sample(s); then to
688 recursively determine a threshold for each marker-sample combination. This is done with
689 dynamic feedback, and by choosing any two features (intensity or morphology based) to populate
690 a scatterplot. Lastly, we used scimap's phenotype function using the generated signature-matrix
691 after marker thresholding to label the cells. Artifacts present in the CD8 channel were manually

692 removed prior analysis. For cellular neighborhood analysis, AnnData objects were merged and
693 analyzed by scimap's spatial_lda function with knn=30 and k=7.
694
695 **Laser microdissection and proteomic sample preparation**
696 General procedure
697 We used the Leica LMD7 system with software version 8.3.0.08259. Tissue was collected using
698 the 20x objective (HC PL FL L 20x/0.40 CORR) for lung cancer and breast cancer samples, and 63x
699 objective (HC PL FLUOTAR L 63x/0,70 CORR PH2) for placenta and scDVP samples. Laser settings
700 were optimized per tissue type and listed below. Contours were collected into low-binding 384-
701 well plates (Eppendorf 0030129547) or into proteoChip EVO96 (C-PEVO-96-10). To ensure tissue
702 settling at the bottom of the wells, 20µL of acetonitrile was added, followed by centrifugation for
703 2min at 2000g and vacuum drying for 10 min at 60 °C. Wells were finally inspected in the LMD7
704 system to confirm collection.
705 Sample purification was performed using Evotips Pure (Evosep) following manufacturer's
706 instructions, and samples were kept with Buffer A until LC-MS measurement.
707
708 <u>First trimester placenta</u>
709 Contours were collected by laser microdissection using the 63x magnification objective. About
710 150 cells were collected per sample, in total about 50,000 μm^2. Cell lysis was performed with 4 µL
711 of 60 mM TEAB, followed by brief centrifugation (2,000 RCF, 1 min) and heating at 95 °C for
712 60 min in a Bio-Rad thermal cycler (384-well module, lid at 110 °C). ACN (1 µL; 20% final) was
713 added, and the samples were incubated at 75 °C for 60 min. After cooling, 2 µL LysC (2 ng/µL in
714 LC-MS grade H_2O) was added and digestion was performed at 37 °C for 4 h. Trypsin (2 µL, 2 ng/µL;
715 Promega Trypsin Gold) was added for overnight digestion at 37 °C. Digestion was stopped with
716 1% TFA (v/v), and samples were vacuum-dried (~60 min, 60 °C) and stored at –20 °C. Before LC-
717 MS, peptides were purified using Evotips Pure, eluted into a 96-well plate, vacuum-dried, and
718 resuspended in 4.2 µL MS loading buffer (3% ACN, 0.2% TFA); the plate was vortexed (10 s),
719 centrifuged (5 min, 2,000 RCF) and 4µL injected into the EASY-nLC-1200 system.
720 <u>Lung cancer</u>

721 Contours were collected by laser microdissection using the 20x magnification objective. All
722 samples' total area ranged between 50,000-100,000 μm^2. Cell lysis and protein digestion was
723 performed as described above for the placenta samples. After trypsin digestion, samples were
724 vacuum-dried (~60 min, 60 °C) and stored at –20 °C. Immediately before measurement, C18
725 based peptide purification (Evotip Pure) was performed, and samples were maintained in Buffer
726 A prior LC-MS analysis.

727 Triple negative breast cancer

728 The total area for each sample was 25,000 μm^2. Protein extraction and digestion used a DDM-
729 based protocol as described above. A lysis buffer containing 0.025% DDM, 5 mM TCEP, 20 mM
730 CAA, and 0.1 M TEAB in water was dispensed into wells (4 µL per sample) using a MANTIS Liquid
731 Dispenser. Plates were sealed with PCR ComfortLid seals and incubated at 95 °C for 60 min. After
732 cooling, 2µL of LysC (4 ng/µL in 0.1 M TEAB [pH 8.5] with 30% acetonitrile) was added, and
733 samples were digested for 4 h at 37 °C in a thermal cycler. Subsequently, 3µL of trypsin (2ng/µL
734 in 0.1 M TEAB [pH 8.5] with 10% acetonitrile) was added, and digestion proceeded overnight at
735 37 °C. Samples were vacuum dried for peptide purification.

736

737 **Liquid chromatography and mass spectrometry (LC-MS)**

738 First trimester placenta

739 Peptides were analyzed using an EASY-nLC 1200 (Thermo Fisher) coupled to a trapped ion
740 mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF SCP, Bruker
741 Daltonik). Samples were loaded onto a 20 cm in-house packed C18 column (75 µm ID, 1.9 µm
742 ReproSilPur C18-AQ silica beads, Dr. Maisch, Germany) and separated at 250 nL/min with a
743 gradient of increasing concentrations of buffer B (0.1% formic acid, 90% ACN in LC-MS grade H2O)
744 to 60% buffer A (3% ACN, 0.1% formic acid in LC-MS grade H2O). The gradient duration was
745 21 min. MS acquisition was performed in diaPASEF mode using the vendor's default '3x8'
746 method. Eight dia-PASEF scans were conducted, each divided into three ion mobility windows,
747 spanning a mass-to-charge ratio (m/z) range of 400-1000 with 25 Th windows, and covering an
748 ion mobility range from 0.64 to 1.37 Vs cm-2. The mass spectrometer was set to high sensitivity
749 mode, featuring an accumulation and ramp time of 100 ms, a capillary voltage of 1750V, and a

750 collision energy that increased linearly from 20 eV at 1/K0 = 0.6 Vs cm-2 to 59 eV at 1/K0 = 1.6 Vs
751 cm-2. The collision energy was adjusted linearly based on ion mobility, starting at 59 eV at 1/K0
752 = 1.6 V s cm-2 and decreasing to 20 eV at 1/K0 = 0.6 V s cm-2.
753 Lung cancer
754 Lung cancer samples were directly eluted from Evotips using the Evosep One coupled to a
755 trapped ion mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF Ultra,
756 Bruker Daltonik). Separation was performed using an Evosep 30SPD gradient with an IonOpticks
757 Aurora Elite column. Samples were measured in dia-PASEF acquisition mode. The MS acquisition
758 settings were the same as described above.
759 Triple negative breast cancer
760 Liquid chromatography was performed using the Evosep One LC system connected to a trapped
761 ion mobility spectrometer with a quadrupole time-of-flight mass spectrometer (timsTOF Ultra 2,
762 Bruker Daltonik). Separation was performed using an Evosep 80SPD gradient with an IonOpticks
763 Aurora Rapid column. Samples were measured in dia-PASEF acquisition mode in high sensitivity
764 mode with the MS method details as described above.
765
766 **Proteomic raw file and data analysis**
767 First trimester placenta
768 Raw data were processed in DIA-NN (v1.8.1) using a predicted human spectral library (UniProt,
769 2021). Default settings were used with adjustments: mass range 100–1,700 m/z, precursor
770 charge states 2–4, max 2 miscleavages, MS1/MS2 mass accuracy 15 ppm, match-between-runs
771 enabled, and quantification set to 'Robust LC'. Protein-level outputs (pg_matrix.tsv and
772 unique_genes_matrix.tsv) were analysed in Perseus [68](v.1.16.0.5). Missing values were imputed
773 from a normal distribution (width = 0.3, downshift = 1.8) after filtering for ≥70% quantified values
774 per cell type group. Matrices were further analyzed in R (v4.1.2). To identify phenotype-specific
775 protein markers, highly variable proteins were selected based on an ANOVA F-values greater than
776 four. Differential abundance was assessed using Welch's two-sided t-test followed by Benjamini–
777 Hochberg FDR correction. Identified marker proteins (FDR < 0.05) overlapping with public
778 transcriptomic datasets were visualized using DiVenn [69]. Receptor ligand communication analysis

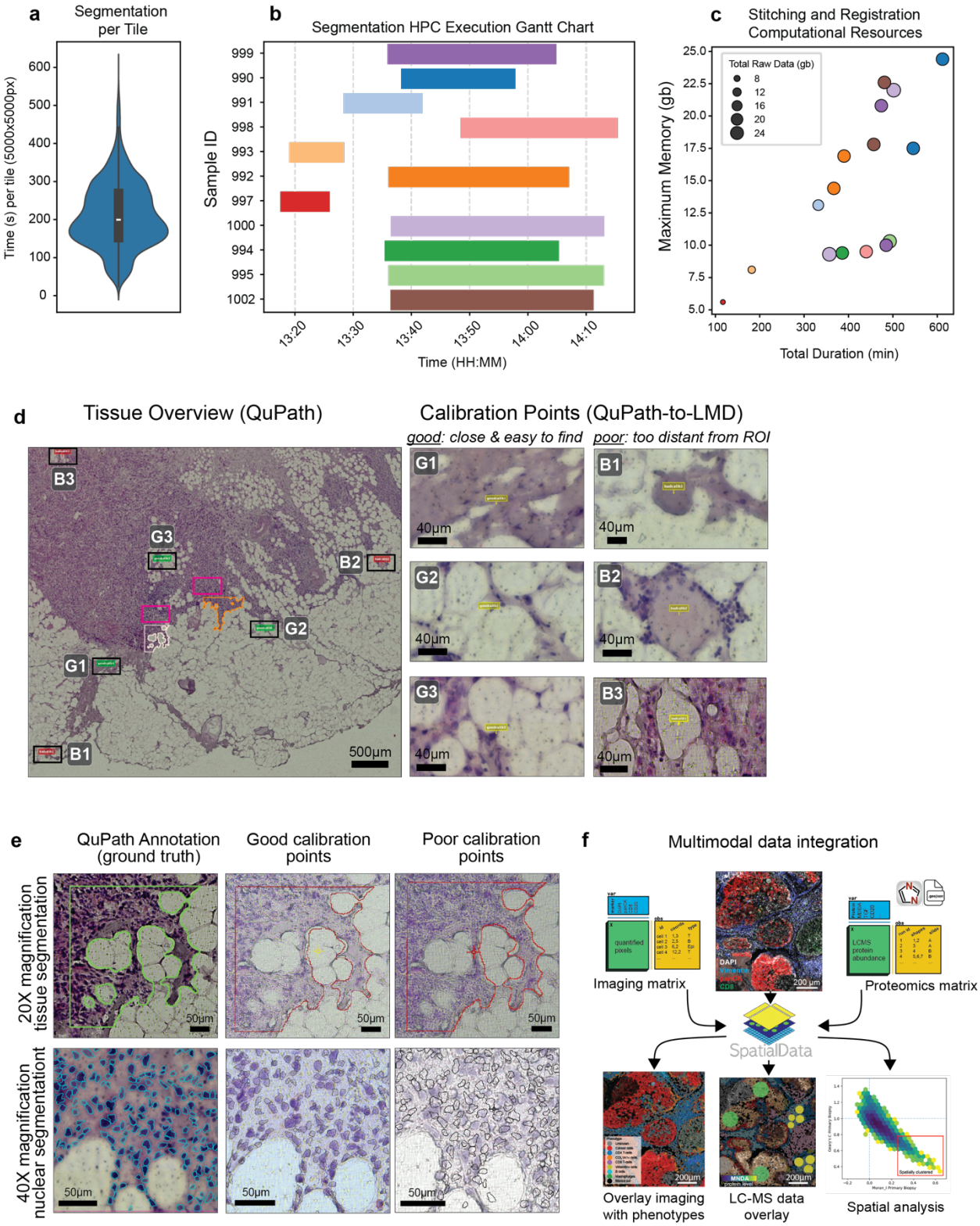
779 was done using CellChat and included receptors and ligands indexed as "secreted," "cell-cell
780 contact" or "ECM-receptor" in their database. We used standard parameters (excluding per-
781 protein smoothing) of 1000 permutations for the probabilistic inference of each cell pair and
782 transmitter-receiver interaction, considering heterometric structures and interactor mediator
783 proteins.
784
785 Lung cancer
786 Raw files were analyzed in DIA-NN (version 1.8.1), using a predicted human spectral library
787 (UniProt, 2021) with small changes to default parameters. Briefly, mass range 400-1000, missing
788 cleavages 2, max variable modifications 1, Ox(M) and Ac(N-term) modifications, precursor charge
789 range 2-4, precursor m/z range 400-1000, fragment ion m/z range 200-1000, mass accuracy 15,
790 MS1 accuracy 15, and MBR active. Raw files were pre-processed in Perseus as mentioned above.
791 Missing values were imputed from a normal distribution (width = 0.3, downshift = 1.8) after
792 filtering for ≥70% quantified values per cell type group. Downstream analysis was performed in
793 R (v4.3.2). Differential abundance was assessed using Welch's two-sided t-test followed by
794 Benjamini–Hochberg FDR correction of 5%. For comparisons with RNA levels, the spatial
795 transcriptomics dataset previously published in collaboration with Pentimalli et al. [37] was used.
796 For analyses in figures 3i-k, publicly available data from the LUAD TCGA dataset was used.
797
798 Triple negative breast cancer
799 Raw files were analyzed in DIA-NN (version 1.9) using a predicted human spectral library
800 (UniProt, release 2022) with small changes to default parameters. Briefly, maximum number of
801 variable modifications 1, mass accuracy 15, MS1 accuracy 15, and MBR active.
802 Proteomic data analysis was performed in Python v3.12.8, using AnnData v0.11.2 and SpatialData
803 v0.4.0 as the main data formats. Analysis was performed using opendvp functions with
804 geopandas v1.0.1 and pandas v2.2.3 in the background. Plotting was performed using matplotlib
805 v3.10.0, seaborn v0.13.2, and pycomplexheatmap. The DIA-NN output was processed and
806 adapted to the AnnData object using the openDVP DIANN_to_adata function, and quality control
807 samples were filtered out. Proteins that did not have at least 70% valid values in any cellular

808 neighborhood were filtered out. The remaining NaN values were imputed by low-abundance
809 normal distribution per protein (width=0.3, downshift=1.8). Scanpy v1.11.0 was used to perform
810 PCA dimensionality reduction and plotting. Hierarchical clustering and plotting of heatmap took
811 place with pycomplexheatmap v1.8.1, using method "average," and metric "cityblock." T-tests
812 and ANOVA were performed using pingouin v0.5.5. Spearman's correlation was calculated using
813 scipy v1.16.
814
815 **Data availability**
816 Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium
817 via the PRIDE partner repository [70] with the identifier PXD072882. RNA-seq data were obtained
818 from GEPIA [41], an interactive web server for analyzing RNA sequencing expression data from the
819 TCGA and GTEx projects, using a standard processing pipeline. The LUAD RNA-Seq dataset GEPIA
820 used was based on the UCSC Xena project (http://xena.ucsc.edu).
821 Image data have been deposited to the BioImage Archive repository with the accession number
822 S-BIAD-2194.
823
824 **Code availability**
825 The source code of openDVP is fully opened and accessible on GitHub, including example
826 tutorials: https://github.com/CosciaLab/openDVP. Extensive API documentation of functions can
827 be found at: https://coscialab.github.io/openDVP/index.html. The QuPath-to-LMD webapp code
828 can be found at https://github.com/CosciaLab/Qupath_to_LMD. Lastly, napari-cell-gater can be
829 found at https://github.com/CosciaLab/napari-cell-gater.
830
831

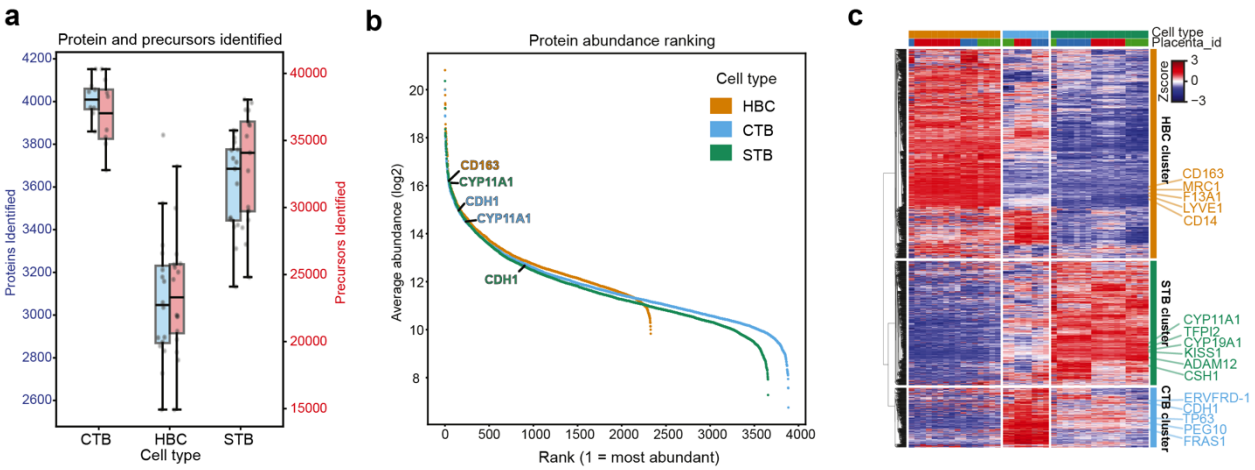
832 **Declaration of generative AI and AI-assisted technologies in the writing process**

833 During the preparation of this manuscript, the authors used Paperpal to improve the readability
834 and language of the text. After using this tool, the authors reviewed and edited the content as
835 needed and take full responsibility for the content of the publication

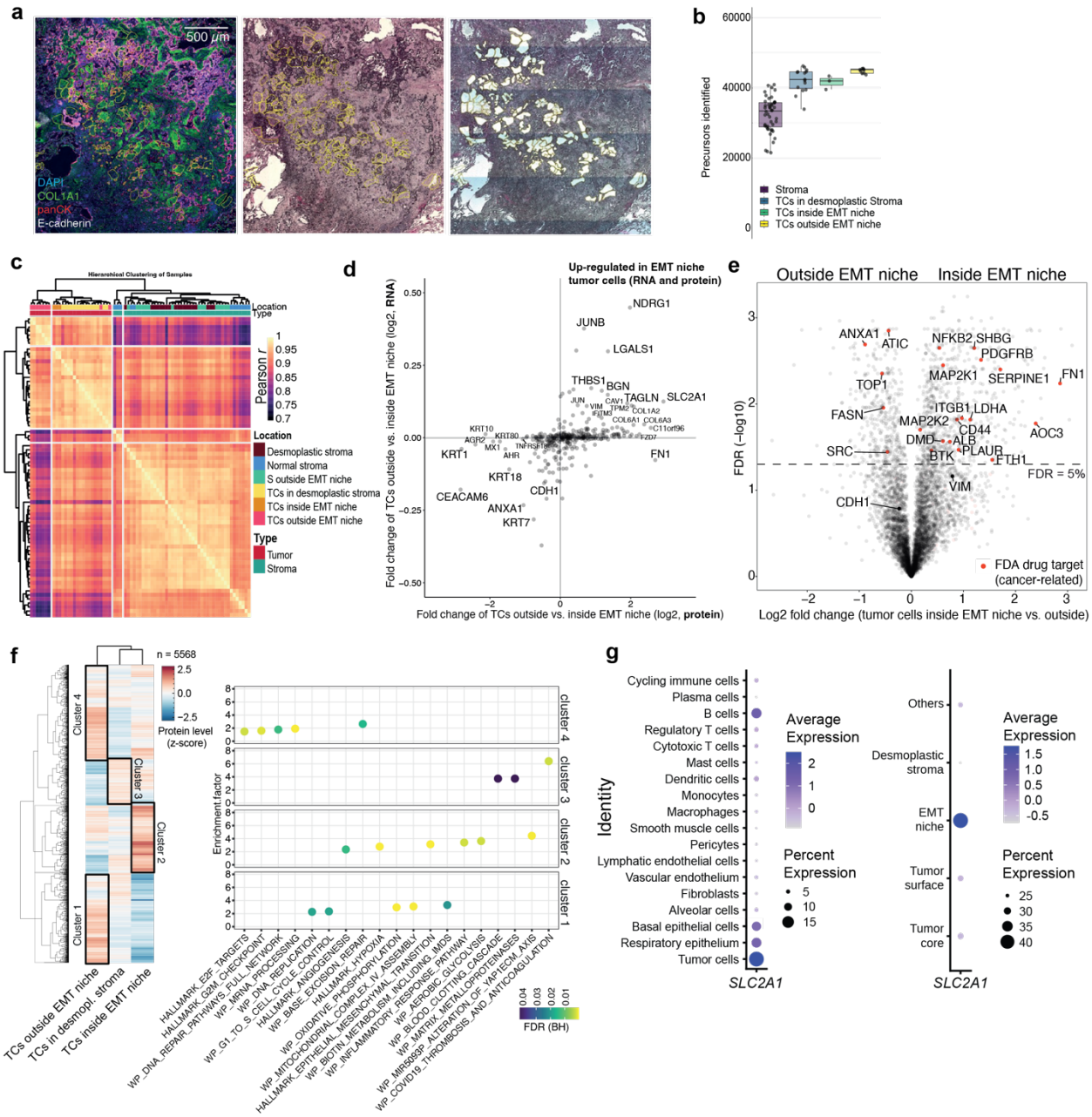
Extended Data Figures



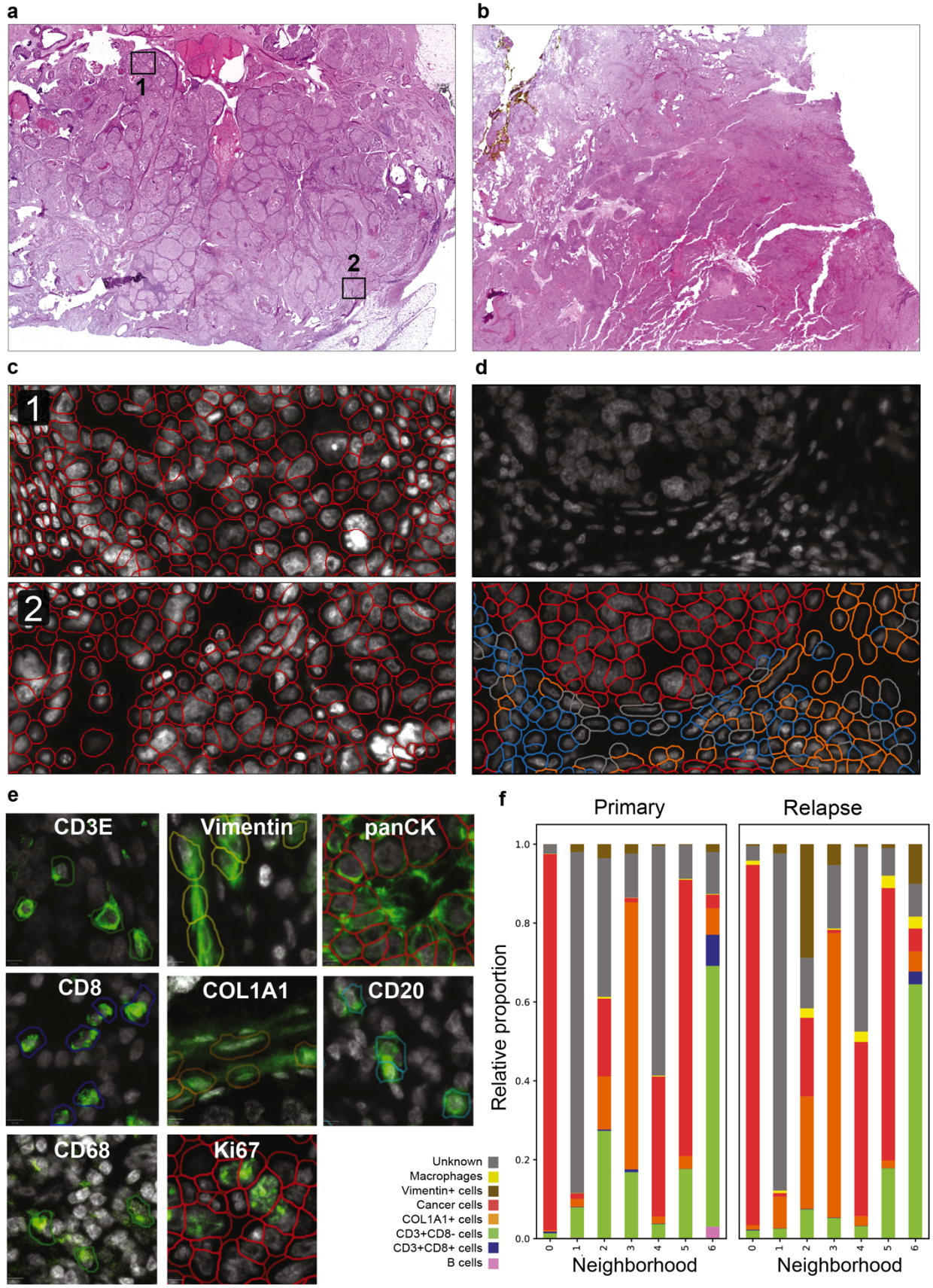
Extended Data Fig. 1. Benchmarking of computational resources and accurate reference point selection for reliable contour transfer for laser microdissection.
(a) Violinplot showing duration in seconds of tile (5000x5000px) segmentation with Cellpose2 "nuclei" model. The central line indicates the median, while the box represents the interquartile range (IQR, 25th to 75th percentiles), whiskers extend to 1.5* IQR of the box **(b)** Gantt chart showing parallelization capabilities of MCMICRO and SOPA, 11 whole-slides with a total of 6+ million cells were segmented in less than two hours. **(c)** Bubble scatterplot with proportional computational resources to input dataset size. Stitching and registration via ASHLAR, is the slowest process in image processing pipeline. **(d)** Tissue overview and reference point definition in QuPath. G1–G3 indicate three good calibration points, located close to the objects of interest and easily identifiable. B1–B3 indicate three poor calibration points, which are located farther away and are difficult to recognize. The center shows four annotated regions: two magenta boxes corresponding to regions subjected to nuclear (H&E-based) segmentation in QuPath, and two irregular regions corresponding to tissue segmentation annotations. **(e)** Accurate annotation transfer depends on reliable calibration points. Top row: large tissue segmentation annotation and its contour transfer to the LMD7. Poor calibration points result in an upward shift of the entire contour. Bottom row: nuclear segmentation and contour transfer. In this case, poor calibration points lead to relatively large positional errors, which would result in improper single-cell collection. **(f)** Schematic illustrating how openDVP leverages the scverse SpatialData format to integrate antibody-based imaging, image-analysis–derived matrices, and LC–MS-based spatial proteomics data. This integration enables overlay of multiple information layers to guide laser microdissection and downstream spatial proteomics analyses.



Extended Data Fig. 2 Proteome coverage and cell-type specific placental tissue proteomes. (a) Box plots showing the number of proteins (blue) and precursors (pink) identified per cell type; each dot represents a sample. The central lines indicate the median, the box the interquartile range, whiskers extend to (1.5*IQR) of the box. **(b)** Rankplot showing the dynamic range of proteins measured with labelled canonical markers and their rank for each cell type. **(c)** Unsupervised hierarchical clustering of 3,994 proteins, showing three distinct cell-type specific protein clusters used in **Fig. 2h.** Canonical markers of each cell type are labelled. The metric for clustering is "cityblock," and the method is "average." Clustering was performed by splitting the dendrogram into three largest groups. CTB, cytotrophoblast; STB, syncytiotrophoblast; HBC, Hofbauer cell.

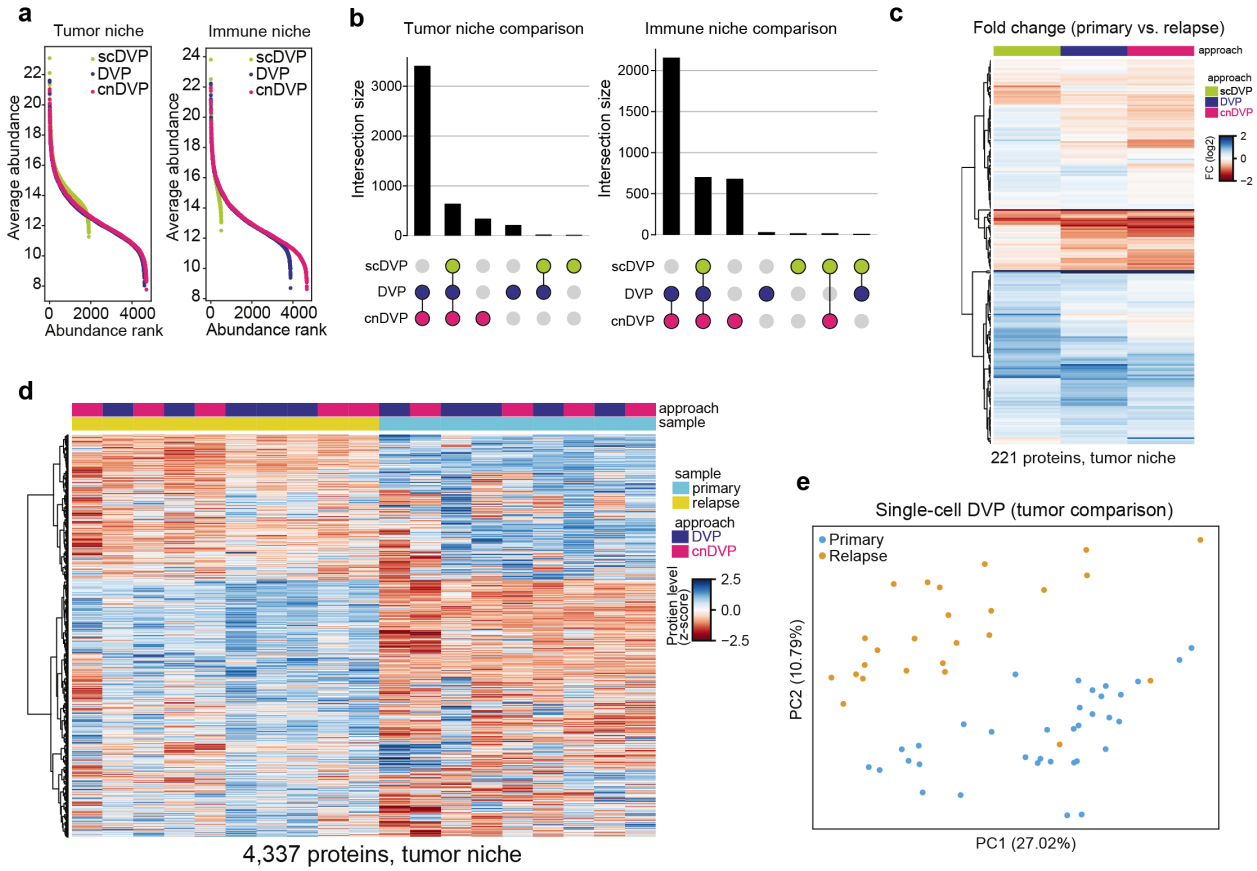


Extended Data Fig. 3 Integration of spatial transcriptomics and proteomics in lung cancer. **(a)** Imaging modalities (H&E and IF) used to guide laser microdissection. Yellow annotations show ROIs used for proteomic profiling. Left: Multiplex immunofluorescence used to locate tumor and stromal niches identified by spatial transcriptomics. H&E staining showing ROIs before (middle panel) and after LMD collection (right panel). **(b)** Boxplot showing the number of identified precursors per sample type. The central lines indicate the median, the box the interquartile range, whiskers extend to (1.5*IQR) of the box **(c)** Pearson's correlation heatmap for all analyzed samples, method: "complete", distance metric: "euclidean". **(d)** Scatterplot comparing fold changes between tumor cells inside and outside the EMT niche, for both spatial proteomics and spatial transcriptomics. **(e)** Volcano plot showing differentially abundant proteins between tumor cells inside and outside the EMT niche as defined by CosMx spatial transcriptomics. FDA-approved drug targets are highlighted in red. E-Cadherin (*CDH1*) and Vimentin (*VIM*) are shown for reference. **(f)** Heatmap showing clusters of proteins between tumor cells inside the EMT niche, outside the EMT niche, and in the desmoplastic stroma, method: "complete", distance metric: "euclidean". The dot plot shows enriched Hallmarks and WikiPathways pathways for the different clusters using ClusterProfiler R package [71]. **(g)** Dot plot of SLC2A1 expression across all cell types (left) and in tumor cells distributed in different niches (right). Dot size indicates the percentage of gene-expressing cells; color represents average expression (log-normalized and scaled).

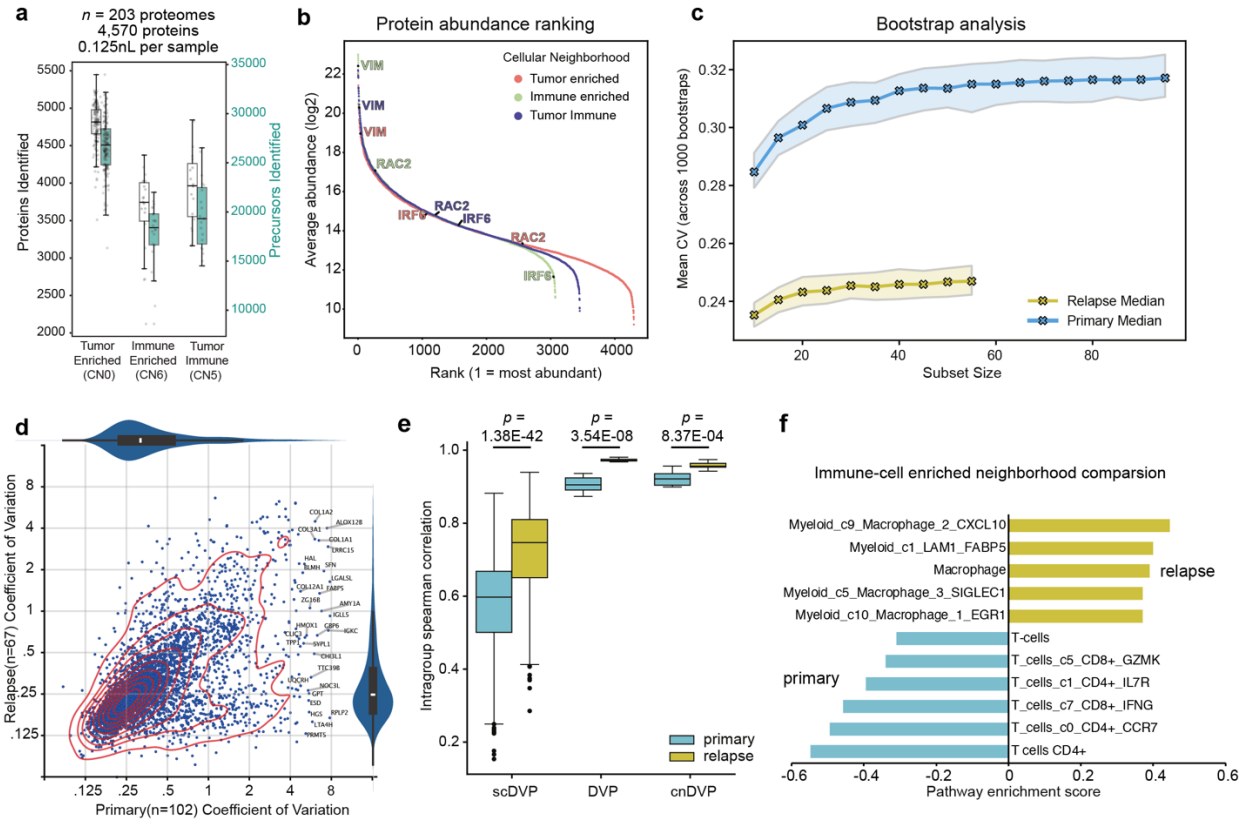

a
1
2
b
c
1
2
d
e
CD3E
Vimentin
panCK
CD8
COL1A1
CD20
CD68
Ki67
f
Primary
Relapse
Relative proportion
Neighborhood
Unknown
Macrophages
Vimentin+ cells
Cancer cells
COL1A1+ cells
CD3+CD8- cells
CD3+CD8+ cells
B cells

881

Extended Data Fig. 4 Consistent segmentation across large distances and staining variations. **(a)** H&E image of the primary TNBC tissue. **(b)** H&E image of the relapse TNBC tissue. Both tissues showed minor tissue losses due to cyclic imaging on PPS frame slides. **(c)** Nuclear stain overlayed with segmentation mask. These zoom-ins belong to distant regions from the primary TNBC tissue, areas shown and numbered in plot (a), showing segmentation consistency across entire whole-slide. **(d)** Zoom-ins of the primary tissue. The white color shows DAPI as the nuclear counterstain. Bottom image shows segmentation masks coloured by assigned phenotype labels. **(e)** Exemplary images for each IF marker used, overlaid with segmentation mask. Colors represent phenotyped cells for each marker (not for Ki67). **(f)** Barplots with cellular neighborhood compositions for each sample, showing the proportion of unclassified cells in grey. Related to Fig. 4g.



Extended Data Fig. 5, Comparative proteomic performance of complementary DVP strategies.
(a) Rank plots illustrating the dynamic range of identified protein groups across three DVP approaches. Protein groups were retained if detected in at least five and three valid values for tumor and immune niches, respectively. **(b)** UpSet plots showing overlap of protein groups identified by the three DVP approaches. Only protein groups with at least 60% valid values per group are shown. **(c)** Heatmap of $\log_2$ fold-change values between primary and relapse cancer samples for each DVP approach. Protein groups were stringently filtered to retain features with at least 70% valid values in each DVP approach. Missing values were imputed from a Gaussian distribution of lower-intensity values (mean shift −1.8, standard deviation 0.3, per feature). Hierarchical clustering was performed using Euclidean distance and the Ward linkage method. **(d)** Heatmap of $\log_2$-transformed protein abundances in cancer samples profiled by the DVP and cnDVP approaches. Protein groups were retained if they contained at least 70% valid values in either the primary or relapse sample group. Missing values were imputed as in (c), and values were z-scored separately for the DVP and cnDVP approaches. Unsupervised hierarchical clustering was based on Euclidean distance and the Ward method. **(e)** Principal component analysis (PCA) of single-cell DVP (scDVP) cancer samples, colored by primary or relapse status. Protein groups were filtered to retain those with at least 50% valid values in either group (923 protein groups total). Missing values were imputed as described in (c).



Extended Data Fig. 6, Neighborhood-guided DVP reveals proteomic differences between primary and recurrent TNBC. **(a)** Boxplot showing proteins and precursors identified per group of cellular neighborhoods. **(b)** Rankplot showing the dynamic range of protein abundance for each neighborhood. Vimentin (VIM) and RAC2 show the highest abundance in the immune-enriched samples, and IRF6 in tumor-enriched samples. **(c)** Line plot showing the spread of the mean CV of 1000 bootstrap simulations, repeated across different subset sizes for both samples. Note, plateauing was observed around 30-40 samples. **(d)** Scatterplot of intra-tissue protein level variability (CV) comparing the primary and relapse sample. Only samples of the tumor-specific CN0 were used analysis. Note, the primary tumor showed overall higher proteome variability. **(e)** Boxplots showing intragroup spearman correlation between cancer only samples for the three DVP approaches scDVP, DVP, and cnDVP. Protein groups are shown with at least 70% valid values for each sample group (primary or relapse) and approach (DVP, scDVP and cnDVP). P-values were calculated with a two-sided independent t-test. **(f)** Barplot of enriched cell type signatures between the primary and relapse sample. Only samples of the immune enriched CN6 were used for analysis. Breast cancer-specific cell type signatures were obtained from publicly available scRNAseq data described in Wu *et al*. [56]. Note, the primary sample showed higher T cell signatures (e.g., CD4+, CD8+), whereas the relapsed immune niche featured higher myeloid/macrophage scores. For boxplots in (**a**),(**d**), and (**e**) the central lines indicate the median, the box the interquartile range, whiskers extend to (1.5*IQR) of the box

References

1. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science (1979)* 381, 1–35 (2023).
2. Method of the Year 2024: spatial proteomics. *Nat Methods* 21, 2195–2196 (2024).
3. de Souza, N., Zhao, S. & Bodenmiller, B. Multiplex protein imaging in tumour biology. *Nat Rev Cancer* 24, 171–191 (2024).
4. Lewis, S. M. *et al.* Spatial omics and multiplexed imaging to explore cancer biology. *Nat Methods* 18, 997–1012 (2021).
5. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* vol. 537 347–355 Preprint at https://doi.org/10.1038/nature19949 (2016).
6. Horvath, P. & Coscia, F. Spatial proteomics in translational and clinical research. *Molecular Systems Biology* 1–5 (2025) doi:10.1038/s44320-025-00101-9.
7. Mund, A., Brunner, A.-D. & Mann, M. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol Cell* 82, 2335–2349 (2022).
8. Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nature Biotechnology 2022 40:8* 40, 1231–1240 (2022).
9. Nordmann, T. M. *et al. Spatial Proteomics Identifies JAKi as Treatment for a Lethal Skin Disease*. *Nature* vol. 635 (Springer US, 2024).
10. Rosenberger, F. A. *et al.* Deep Visual Proteomics maps proteotoxicity in a genetic liver disease. *Nature 2025 642:8067* 642, 484–491 (2025).
11. Schapiro, D. *et al.* MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. *Nat Methods* 19, 311–315 (2022).
12. Blampey, Q. *et al.* Sopa: a technology-invariant pipeline for analyses of image-based spatial omics. *Nat Commun* 15, (2024).
13. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Scientific Reports 2017 7:1* 7, 1–7 (2017).
14. Chiu, C.-L., Clack, N. & community, the napari. napari: a Python Multi-Dimensional Image Viewer Platform for the Research Community. *Microscopy and Microanalysis* 28, 1576–1577 (2022).
15. Marconato, L. *et al.* SpatialData: an open and universal data framework for spatial omics. *Nature Methods 2024* 1–5 (2024) doi:10.1038/s41592-024-02212-x.
16. DI Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* 35, 316–319 (2017).
17. Köster, J. *et al.* Sustainable data analysis with Snakemake. *F1000Res* 10, (2021).
18. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat Methods* 19, 1634–1641 (2022).
19. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology 2021* 1–11 (2021) doi:10.1038/s41587-021-01094-0.

966 20. Nirmal, A. J. & Sorger, P. K. SCIMAP: A Python Toolkit for Integrated Spatial Analysis
967 of Multiplexed Imaging Data. *J Open Source Softw* 9, 6604 (2024).
968 21. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression
969 data analysis. *Genome Biol* 19, 1–5 (2018).
970 22. Meyer-Bender, M. *et al.* Spatialproteomics - an interoperable toolbox for analyzing highly
971 multiplexed fluorescence image data. 1–26 (2025).
972 23. Regev, A. The Human Cell Atlas -Slides. *Elife* 6, e27041 (2017).
973 24. Rajewsky, N. *et al.* LifeTime and improving European healthcare through cell-based
974 interceptive medicine. *Nature* 587, 377–386 (2020).
975 25. Burton, G. J. & Fowden, A. L. The placenta: A multifaceted, transient organ.
976 *Philosophical Transactions of the Royal Society B: Biological Sciences* 370, (2015).
977 26. Bezemer, R. E., Faas, M. M., van Goor, H., Gordijn, S. J. & Prins, J. R. Decidual
978 macrophages and Hofbauer cells in fetal growth restriction. *Front Immunol* 15, 1379537
979 (2024).
980 27. Makhmut, A. *et al.* A framework for ultra-low-input spatial tissue proteomics. *Cell Syst*
981 14, 1002-1014.e5 (2023).
982 28. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN:
983 neural networks and interference correction enable deep proteome coverage in high
984 throughput. *Nat Methods* 17, 41–44 (2020).
985 29. Gonzalez, T. L. *et al.* Sex differences in the late first trimester human placenta
986 transcriptome. *Biol Sex Differ* 9, (2018).
987 30. Nonn, O. *et al.* Senescent Syncytiotrophoblast Secretion During Early Onset
988 Preeclampsia. *Hypertension* 82, (2024).
989 31. Baczyk, D., Drewlo, S. & Kingdom, J. C. P. Emerging role of SUMOylation in placental
990 pathology. *Placenta* 34, 606–612 (2013).
991 32. Baczyk, D., Audette, M. C., Coyaud, E., Raught, B. & Kingdom, J. C. Spatiotemporal
992 distribution of small ubiquitin-like modifiers during human placental development and in
993 response to oxidative and inflammatory stress. *Journal of Physiology* 596, 1587–1600
994 (2018).
995 33. Ruano, C. S. M. *et al.* Alternative splicing in normal and pathological human placentas is
996 correlated to genetic variants. *Hum Genet* 140, 827–848 (2021).
997 34. Turco, M. Y. & Moffett, A. Development of the human placenta. *Development*
998 *(Cambridge)* 146, (2019).
999 35. Reese, J. A. *et al.* Platelet Counts during Pregnancy. *New England Journal of Medicine*
1000 379, 32–43 (2018).
1001 36. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat*
1002 *Commun* 12, 1–20 (2021).
1003 37. Tancredi Massimo Pentimalli, A. *et al.* Combining spatial transcriptomics and ECM
1004 imaging in 3D for mapping cellular interactions in the tumor microenvironment. *Cell Syst*
1005 0, 101261 (2025).

1006 38. David, M. *et al.* Durable responses to imatinib in patients with PDGFRB fusion gene–
1007 positive and BCR-ABL–negative chronic myeloproliferative disorders. *Blood* 109, 61–64
1008 (2007).
1009 39. Stine, Z. E., Schug, Z. T., Salvino, J. M. & Dang, C. V. Targeting cancer metabolism in
1010 the era of precision oncology. *Nat Rev Drug Discov* 21, 141–162 (2022).
1011 40. Ligorio, F. *et al.* Early downmodulation of tumor glycolysis predicts response to fasting-
1012 mimicking diet in triple-negative breast cancer patients. *Cell Metab* 37, 330-344.e7
1013 (2025).
1014 41. Tang, Z. *et al.* GEPIA: A web server for cancer and normal gene expression profiling and
1015 interactive analyses. *Nucleic Acids Res* 45, W98–W102 (2017).
1016 42. Goldman, M. J. *et al.* Visualizing and interpreting cancer genomics data via the Xena
1017 platform. *Nature Biotechnology 2020 38:6* 38, 675–678 (2020).
1018 43. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science (1979)* 357,
1019 (2017).
1020 44. Rosenberger, F. A. *et al.* Spatial single-cell mass spectrometry defines zonation of the
1021 hepatocyte proteome. *Nat Methods* 20, 1530–1536 (2023).
1022 45. Liu, Y. *et al.* Conserved spatial subtypes and cellular neighborhoods of cancer-associated
1023 fibroblasts revealed by single-cell spatial multi-omics. *Cancer Cell* 43, 905-924.e6 (2025).
1024 46. Ma, C. *et al.* Pan-cancer spatially resolved single-cell analysis reveals the crosstalk
1025 between cancer-associated fibroblasts and tumor microenvironment. *Mol Cancer* 22, 170
1026 (2023).
1027 47. Barkley, D. *et al.* Cancer cell states recur across tumor types and form specific
1028 interactions with the tumor microenvironment. *Nature Genetics 2022 54:8* 54, 1192–1201
1029 (2022).
1030 48. Bianchini, G., De Angelis, C., Licata, L. & Gianni, L. Treatment landscape of triple-
1031 negative breast cancer — expanded options, evolving needs. *Nat Rev Clin Oncol* 19, 91–
1032 113 (2022).
1033 49. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* 578,
1034 615–620 (2020).
1035 50. Ranek, J. S. *et al.* QUICHE reveals structural definitions of anti-tumor responses in triple
1036 negative breast cancer. (2025).
1037 51. Miroshnychenko, D. *et al.* Stroma-Mediated Breast Cancer Cell Proliferation Indirectly
1038 Drives Chemoresistance by Accelerating Tumor Recovery between Chemotherapy
1039 Cycles. *Cancer Res* 83, 3681–3692 (2023).
1040 52. Wang, X. Q. *et al.* Spatial predictors of immunotherapy response in triple-negative breast
1041 cancer. *Nature* 621, 868–876 (2023).
1042 53. de Visser, K. E. & Joyce, J. A. The evolving tumor microenvironment: From cancer
1043 initiation to metastatic outgrowth. *Cancer Cell* 41, 374–403 (2023).
1044 54. Friedman-DeLuca, M., Karagiannis, G. S., Condeelis, J. S., Oktay, M. H. & Entenberg, D.
1045 Macrophages in tumor cell migration and metastasis. *Front Immunol* 15, 1–16 (2024).

1046 55. Rosenberger, F. A., Thielert, M. & Mann, M. Making single-cell proteomics biologically
1047 relevant. 20, 320–323 (2023).
1048 56. Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat*
1049 *Genet* 53, 1334–1347 (2021).
1050 57. Lomakin, A. *et al.* Spatial genomics maps the structure, nature and evolution of cancer
1051 clones. *Nature* 611, 594–602 (2022).
1052 58. Casasent, A. K. *et al.* Multiclonal Invasion in Breast Tumors Identified by Topographic
1053 Single Cell Sequencing Article Multiclonal Invasion in Breast Tumors Identified by
1054 Topographic Single Cell Sequencing. *Cell* 172, 205-210.e12.
1055 59. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell*
1056 *Metabolism* vol. 23 27–47 Preprint at https://doi.org/10.1016/j.cmet.2015.12.006 (2016).
1057 60. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* vol. 144
1058 646–674 Preprint at https://doi.org/10.1016/j.cell.2011.02.013 (2011).
1059 61. Wang, X. *et al.* Genomic and transcriptomic analyses identify distinctive features of
1060 triple-negative inflammatory breast cancer. *NPJ Precis Oncol* 8, 1–14 (2024).
1061 62. Benci, J. L. *et al.* Tumor Interferon Signaling Regulates a Multigenic Resistance Program
1062 to Immune Checkpoint Blockade. *Cell* 167, 1540 (2016).
1063 63. Mardamshina, M. Multiplexed Deep Visual Proteomics Unveils Spatial Heterogeneity and
1064 Rare Endocrine States in Human Adult Pancreatic Islets. 1–48 (2025).
1065 64. Post, F. *et al.* Deep visual proteomics reveals an in vivo-like phenotype of orthotopically
1066 transplanted human colon organoids. *Cell Syst* 0, 101396 (2025).
1067 65. Stevens, L. E. *et al.* JAK–STAT Signaling in Inflammatory Breast Cancer Enables
1068 Chemotherapy-Resistant Cell States. *Cancer Res* 83, 264–284 (2023).
1069 66. Long, L., Fei, X., Chen, L., Yao, L. & Lei, X. Potential therapeutic targets of the
1070 JAK2/STAT3 signaling pathway in triple-negative breast cancer. *Front Oncol* 14,
1071 1381251 (2024).
1072 67. Zheng, X., Mund, A. & Mann, M. Deciphering functional tumor-immune crosstalk
1073 through highly multiplexed imaging and deep visual proteomics. *Mol Cell* 85, 1008-
1074 1023.e7 (2025).
1075 68. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of
1076 (prote)omics data. *Nature Methods* vol. 13 731–740 Preprint at
1077 https://doi.org/10.1038/nmeth.3901 (2016).
1078 69. Sun, L. *et al.* Divenn: An interactive and integrated web-based visualization tool for
1079 comparing gene lists. *Front Genet* 10, (2019).
1080 70. Vizcaíno, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic*
1081 *Acids Res* 44, D447–D456 (2016).
1082 71. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing
1083 biological themes among gene clusters. *OMICS* 16, 284–287 (2012).
1084