ShapeUpLMD: An Automated Pipeline for Spatially Optimized Laser Microdissection in Multi-omic Tissue Profiling

Joshua P. Schaaf[1,2*], Dave Mitchell[1,2*], Jonathan D. Ogata[1,2], Sakiyah A. TaQee[1,2], Katlin N. Wilson[1,2], Kelly A. Conrads[1,2], Jeremy Loffredo[1,2], Brian L. Hood[1,2], Tamara Abulez[1,2], Allison L. Hunt[3], Andrea Langeland[3], Paulette Mhawech-Fauceglia[1], Kathleen M. Darcy[1,2], Christopher M. Tarney[1], G. Larry Maxwell[1,3], Thomas P. Conrads[1,3*], Nicholas W. Bateman[1,2*]

1. Gynecologic Cancer Center of Excellence, Murtha Cancer Center Research Program, Department of Surgery, Uniformed Services University and Walter Reed National Military Medical Center, 8901 Wisconsin Avenue, Bethesda 20889, MD, USA.
2. The Henry M. Jackson Foundation for the Advancement of Military Medicine Inc., Bethesda, MD, USA 20817.
3. Department of Obstetrics and Gynecology, Inova Fairfax Medical Campus, 3300 Gallows Rd. Falls Church, VA 22042, USA.

*Authors contributed equally

Keywords: laser microdissection, spatial multi-omics, spatial proteomics, uterine serous carcinoma. Address correspondence to: Dr. Nicholas W. Bateman, email: batemann@whirc.org, or Dr. Thomas P. Conrads, email: conrads@whirc.org, 3289 Woodburn Rd, Suite 385, Annandale, VA 22003; Tel: 703-207-3362; Fax: 703-560-0919.

Abstract (250 word version)

Laser microdissection (LMD) enables enrichment of defined cellular populations from heterogeneous tissues, providing histologically-resolved molecular profiles with spatial resolution. However, the absence of standardized methods for generating optimized regions of interest (ROIs) limits reproducibility and scalability for spatial multi-omic workflows. To address this gap, we developed ShapeUpLMD, an open-source software tool and integrated workflow that automates the optimization of spatially defined ROIs for LMD, directly from digital pathology annotations.

Fresh-frozen uterine serous carcinoma (USC) tumors (n = 11) were sectioned onto polyethylene naphthalate slides, scanned and subsequently annotated by an expert pathologist. Tumor and non-tumor ROIs were used to train tumor-specific classifiers in HALO (Indica Labs). Classified tumor or unbiased whole tissue ROIs were refined and optimized by ShapeUpLMD prior to automated collection on a Leica LMD7 microscope. Across the cohort, predicted tumor ROIs increased effective tumor purity by 86 ± 6% relative to whole tissue while maintaining high spatial concordance between predicted and collected ROIs (accuracy = 0.9 ± 0.05). Data-independent acquisition mass spectrometry quantified upwards of ~6,500 proteins across spatially resolved ROIs, revealing regionally coherent clustering of adjacent tumor regions and abundance patterns consistent with tumor and non-tumor cell admixture in an unbiased spatial proteomic application.

ShapeUpLMD provides an automated, reproducible, and scalable framework that bridges digital pathology with LMD enabling high-fidelity spatial enrichment for multi-omic analyses. This workflow increases throughput, reduces inter-operator variability, and supports standardized, regionally resolved tissue collection for spatial systems biology applications. The software is available at https://github.com/GYNCOE/ShapeUpLMD.

Introduction

Laser microdissection (LMD) enables the precise isolation of specific cellular populations from histologically complex tissues, facilitating high-resolution spatial molecular analyses of tissue microenvironments. This technique mirrors workflows used in diagnostic pathology, where tissue sections are mounted onto slides, stained, and cellular architecture is visually assessed by expert pathology review to identify distinct cellular architectures or regions of interest (ROIs). By adding this spatial axis to molecular analyses, LMD serves as a powerful bridge between histopathological context and downstream molecular characterization.

Emerging evidence has shown that cells collected by LMD yield high quality molecular analytes including DNA, RNA and protein from solid tumor samples (1). This technique has proven especially valuable for microcompartment analysis of tumor tissues, which often exhibit marked heterogeneity arising from admixture of malignant, stromal, and immune cell populations (1,2). By enriching histologically homogeneous subpopulations, LMD enables disentanglement of molecular signals that would otherwise be obscured in bulk tissue analyses. More recently, LMD has become integral to spatial -omics workflows, such as deep visual proteomics (3), supporting high resolution characterization of molecular heterogeneity within tumor architectures. As LMD adoption increases in high-throughput multi-omic workflows, standardized collection protocols and quality control practices will be paramount to ensure reproducibility, scalability, and intra- and inter-study comparability.

Traditionally, ROI selection for LMD enrichment relies on manual annotation by expert pathologists - a time-intensive process subject to inter-individual variation. Manual annotation directly on the LMD microscope also exposes tissues to ambient conditions for extended periods, potentially compromising molecular integrity. Recent advances have enabled direct transfer of

digital tissue ROI annotations onto LMD instruments, supporting more efficient and automated collection of defined cell populations (4-6). Innovations in digital pathology and machine learning-based image analysis present new opportunities to automate ROI identification, thereby improving reproducibility and throughput. Current workflows lack strategies to algorithmically optimize ROI geometry, such as standardizing ROI shape and size, controlling inter-ROI spacing, and defining regionally resolved sampling patterns, to fully support spatial multi-omic analyses. Consequently, there remains a critical need for standardized and automated methods that generate optimized, spatially defined ROIs to minimize inter-operator variability and maximize reproducibility of collections at cohort-scale.

To address this need, we developed ShapeUpLMD, an open-source software tool that automates generation of spatially defined, LMD-optimized ROIs from digital pathology annotations. ShapeUpLMD extends our prior efforts integrating digital pathology with automated LMD (4) by introducing algorithms for ROI refinement, area and spacing optimization, and spatial partitioning to support spatially resolved tissue sampling. We demonstrate performance of this pipeline in a cohort of uterine serous carcinoma (USC) tumors and illustrate its application for spatial proteomics with a representative USC tissue specimen. Together, these efforts establish a scalable and reproducible pipeline for automated spatial enrichment in multi-omic workflows.

Methods

Patient Tissues and Classification of Tumor Regions of Interest

Fresh-frozen tumor tissues were collected from eleven USC patients enrolled in a WCG IRB approved #20110222 Tissue and Data Acquisition Study of Gynecologic Disease who underwent surgery at Inova Fairfax Medical Campus (Falls Church, VA, USA), University of Oklahoma Health Sciences (Oklahoma City, OK, USA) or at University of Virginia (Charlottesville, VA, USA); all experimental protocols involving human data in this study were in managed in accordance with the Declaration of Helsinki. Thin (8 µm) tissue sections were sectioned and placed on glass or polyethylene naphthalate membrane slides (PEN, Leica Microsystems, Wetzlar, Germany) into which fiducials were pre-cut for image co-registration as previously described (4). After staining with aqueous hematoxylin and eosin, tissue sections were scanned (Aperio ScanScope XT slide scanner, Leica Microsystems, Wetzlar, Germany) and digital images were reviewed by a board-certified pathologist (P.M.F.) to confirm disease diagnosis and define representative ROIs.

Representative tumor and adjacent non-tumor regions were annotated on adjacent PEN membrane tissue sections and HALO Image Analysis Platform v4.0.5107 (Indica Labs, Albuquerque, New Mexico, USA) was used to train tumor cell-specific classifiers using the DenseNet 2 module. A minimum of ten high-purity tumor regions were manually annotated on a PEN slide image, guided by annotations from expert pathology review of adjacent H&E tissue sections (please see Supplementary Figure 1 for more details). To ensure spatial and morphological diversity, five regions were placed near the periphery of the tissue section area, while the remainder were distributed more centrally when possible. Tumor shapes were then positioned adjacent to non-tumor elements or background glass to define edge boundaries for model training. A separate

132 annotation layer was created in which ten non-tumor annotations were added, each encompassing
133 a tumor annotation region noted above to maximize differentiation of tumor from non-tumor
134 regions. Non-tumor annotations were also chosen to include stromal elements and background
135 glass, taking care to ensure that each non-tumor regions contained only the embedded tumor
136 annotation as the sole source of tumor tissue. These annotations were used to create cell type-
137 specific classifiers using the DenseNet 2 image analysis module. Minimal object size was set to
138 10,000 μm^2 to exclude small, irrelevant shapes, and the resolution was set to 1 μm/pixel to allow
139 for tighter classification boundaries. Two classes were defined, tumor and non-tumor, with tumor
140 set as the priority class to ensure more conservative inclusion of ambiguous regions. Training was
141 run for approximately 3,000 iterations and the resulting high-resolution binary tumor cell-specific
142 classifier was used for downstream analysis.
143
144 **ShapeUpLMD – Development and Application**
145 ShapeUpLMD was developed in Python v3.10.16. Regions determined to be collected
146 were retrieved from the HALO Image Analysis Platform, generated using HALO AI v4.0.5107
147 (Indica Labs, Inc.) and the HALO GraphQL database. Annotations were converted from vertex
148 coordinates into shapely v2.0.6 polygon objects, making sure to account for all HALO shape types,
149 and whether that shape is to be collected or excluded (isExclusionRegion). Exclusion regions were
150 connected to their encompassing collected shape edge to avoid collecting negative regions due to
151 LMDs gravitational collection. The shortest path from the exclusion region to the encompassing
152 shape edge was used to diminish the loss of collection yield. Shapes were then reduced in size
153 through iteratively suggesting horizontal and vertical slices, choosing the proposed sliced sections
154 which provide the largest area of collectable material. For this application, we defined maximum

155 shape area to be 0.36 mm^2 and slice width as 60 µm (as applied for tumor-enriched collections),
156 both set to avoid tissue peeling during cutting. Regions between shapes were widened so that they
157 were at least 60 µm apart. This may introduce thin shapes, so shapes which had thin regions of less
158 than 60 µm wide were removed recursively. Shapes were finally simplified with the simplify
159 function in shapely using the Douglas-Peucker algorithm, as we identified that fewer vertices
160 enabled faster cutting times on Leica's LMD7. Shapes smaller than 0.01 mm^2 were removed
161 between multiple ShapeUpLMD steps, as they were deemed too small for collection. Shapes could
162 then be placed into spatially resolved groups by their proximity to algorithmically determined
163 points spread evenly within the tissue section. Finalized shapes were exported to an LMD7-
164 importable extensible markup language XML file using python code modified from the Malleator
165 tool (4). Slide image metadata such as micrometers per pixel and slide dimensions were read using
166 tiatoolbox (1.6.0). Environment files, HALO database connection files, and ShapeUpLMD code
167 are available on GitHub (https://github.com/GYNCOE/ShapeUpLMD).
168

169 **Collection of ROI shapes**

170 Tumor regions annotated in HALO were optimized for LMD using ShapeUpLMD. Briefly,
171 tumor layer coordinates were accessed through the HALO API and subdivided into uniformly
172 spaced polygons using the following parameters; 60 µm spacing between shapes, with minimum
173 and maximum dimensions of 100 × 100 µm and 600 × 600 µm, respectively. For unbiased
174 collections, we kept the same minimum size but updated the spacing to 50 µm and reduced the
175 maximum size to 250 × 250 µm. These constraints were implemented to minimize membrane and
176 tissue curling while ensuring adequate sample areas for recovery from collections using a 5×
177 objective. For spatial analyses of tumor-enriched regions, each subdivided region was assigned a

178 spatial cluster identifier using the spatial_segmentation_brute_force function in ShapeUpLMD,
179 and total cluster areas (mm^2) were balanced across the slide. For unbiased spatial analyses, spatial
180 regions were similarly generated, but by using rebalance_multipolygons_by_area function in
181 ShapeUPLMD to achieve more balanced area sizes across the tissue section of interest. Cluster-
182 specific XML annotation files were generated and imported into the Leica LMD7 system.
183 Alignment was verified using fiducial markers to maintain spatial correspondence with the tissue
184 section as previously described (4) and ROIs were cut using the following LMD microscope
185 software settings: power 60, aperture 2, speed 33, specimen balance 5, line spacing (draw + scan):
186 5, head current 100 %, pulse frequency 555 and offset 45. LMD tissue collections from each spatial
187 cluster were collected separately into pressure cycling technology (PCT) tubes (please see Sample
188 preparation for quantitative proteomic analysis below for more details).
189

190 **Evaluation of Manual Tumor, Automated Tumor, and Collected Regions**

191 Tumor cell-specific ROI annotations generated by manual inspection, by automated
192 classifier generation ROIs as noted above, by ShapeUpLMD-derived and optimized ROIs, and
193 finally collected ROIs were all compared using a series of in-house tools generated in Python.
194 Briefly, whole slide image annotation polygons corresponding to ground truth tumor, predicted
195 tumor, and tissue regions were converted into binary masks using custom functions utilizing the
196 packages NumPy, Shapely, and OpenCV. All performance metrics were computed within tissue
197 boundaries; pixel-wise comparisons between ground truth and predicted tumor masks were used
198 to calculate precision (the proportion of predicted tumor pixels that were correctly identified),
199 recall or sensitivity (the fraction of true tumor pixels correctly detected), F1 score (the harmonic
200 mean of precision and recall summarizing segmentation balance), and accuracy (the overall

201 proportion of correctly classified pixels). Collected regions were estimated by the HALO AI
202 DenseNet 2 model which determines areas collected from post-LMD collection scanned slides.
203 Slides were loaded using OpenSlide, and code was created to process and generate color-coded
204 overlay masks, thumbnail previews, and a timestamped CSV file summarizing all metric results
205 for subsequent visualization and statistical aggregation. Mann Whitney U testing was performed
206 using MedCalc v20.109.
207

208 **Sample preparation for quantitative proteomic analysis**

209 Tumor regions of interest (ROIs) were imported onto an laser microdissection microscope
210 (LMD7, Leica Microsystems, Wetzlar, Germany) from XML coordinate exports and co-registered
211 with mounted slide images as previously described (4). Tumor enriched ROIs were collected by
212 laser microdissection into 20 µL of 100 mM triethylammonium bicarbonate (TEAB)/10%
213 acetonitrile (ACN), pH 8.0 and unbiased ROIs were collected in 100mM ammonium
214 bicarbonate/10% ACN, in MicroTubes (Pressure BioSciences, Inc, South Easton, MA) followed
215 by digestion with a heat-stable form of trypsin (SMART Trypsin, ThermoFisher Scientific, Inc.)
216 using pressure cycling technology with a barocycler (2320EXT Pressure BioSciences, Inc).
217 Digested peptide concentrations were determined using the bicinchoninic acid assay (BCA assay,
218 ThermoFisher Scientific, Inc.) for tumor-enriched collections and equivalent volumes of peptide
219 digest were analyzed for tumor-enriched and unbiased ROI collections.
220

221 **Liquid-Chromatography-Mass Spectrometry (LC-MS) Analysis**

222 For tumor enriched analyses, a timsTOF Ultra 2 MS connected to an EvoSep LC system
223 (EvoSep, Odense, Denmark) via a CaptiveSpray 2 source (Bruker, Billerica, MA) was used for

224 dia-PASEF LC-MS/MS analysis. Peptide digests were separated using a C18 IonOpticks column
225 (particle size 1.7 µm, 75 µM µm ID, 15 cm length) at a flow rate of 0.2 µL/min; solvent A (0.1%
226 formic acid in water), solvent B (0.1% formic acid in acetonitrile), using the Evosep Whisper Zoom
227 40 SPD method. For dia-PASEF analysis, the scan range was *m/z* 400-1000 with a mobility range
228 of 0.64 to 1.37 V·s/cm² with eight MS/MS ramps per MS scan. Three MS/MS scans were collected
229 per ramp starting at *m/z* 400, 600, and 800 with a width of 25 which increased by 25 Th for each
230 successive ramp. The capillary voltage was set at 1600 V, and the ramping collision energy was 20
231 eV at 0.6 mobility to 59 eV at 1.6 mobility. dia-PASEF scan range *m/z* 100–1700 in positive mode,
232 and IMS service ramp time of 100 ms.

233 For unbiased spatial proteomic analysis, an Orbitrap Astral (ThermoFisher Scientific, Inc)
234 MS was connected to a Vanquish Neo (ThermoFisher Scientific, Inc) using an Easy-Spray source
235 (ThermoFisher Scientific, Inc.). Peptides were loaded onto an Acclaim PepMap 100 pre-column
236 (75 µM µm ID x 2 cm) and separated on an IonOpticks Elite XT analytical column (1.7 µm, 75
237 µM µm ID x 15 cm) at a flow rate of 0.35 µL/min; solvent A (0.1% formic acid in water), solvent
238 B (0.1% formic acid in acetonitrile) from 0-25 %B over 25 min, 25-35 %B over 5 min, 35-99 %B
239 over 5 min followed by column washing at 99 %B for 6.3 min before equilibrating (5 column
240 volumes) for the next injection. For data-independent (DIA) analyses, full MS data were collected
241 in profile mode from *m/z* 380-980 at a resolution of 240k. Voltage was set to 1900; normalized
242 AGC target was 500% (5e6 absolute), RF Lens was 40%, with a maximum injection time of 100
243 ms using Advanced Peak Determination and lock mass correction (*m/z* 445.12003; 10 ppm mass
244 tolerance). DIA data were collected in the Astral detector over the same precursor mass range with
245 isolation windows of 2 Th and maximum injection time of 3 ms. The normalized AGC target was

500% (5e4 absolute), HCD collision energy was 27, an RF Lens of 45% and a scan range of *m/z* 150-2000 using a loop control time of 0.5 s.

Proteomic Data Analysis

Mass spectrometry RAW files were searched with Spectronaut v19.9 (Biognosys AG, Switzerland) in directDIA mode against a UniProtKB/Swiss-Prot database (Homo sapiens, downloaded May 24, 2024). Default Spectronaut settings were used, except minimum peptide length was set to 6, methionine oxidation was the only variable modification, and precursors passing FDR < 0.01 were considered in downstream analysis. Spectronaut parquet report outputs were imported into R v4.2.33 via the arrow package v16.1.04. The columns Stripped.Sequence, PEP.AllOccuringProteinAccessions, and FG.MS2Quantity were extracted as peptide sequence, mapped protein accessions, and precursor MS2 intensity, respectively. Precursors mapping to multiple proteins were assigned to the protein quantified with the most unique peptides. The sum of sample-level precursor MS2 intensities were normalized to the lowest sum-intensity observed across all samples. Proteins observed in ≤50% of samples were removed and remaining missing values were imputed by k-nearest neighbors (k = 11) via the impute package v1.72.05. Resulting protein abundances were median centered per sample and $\log_2$ transformed. From the resulting protein-level matrix, tumor-enriched proteome data from 10mm (Fig. 2B) or 5mm (Fig. 2C) regions were subset and standardized by row-wise z-scoring. Global proteomics data generated for unbiased spatial ROIs for a representative USC tumor tissue section underwent protein-wise quantile standardization across $\log_2$ fold-change ROI abundance ranges to scale between ± 2.0. Standardized protein abundances were then mapped to corresponding XML coordinates and rendered into tissue section-level proteome abundance maps using in-house tools that included R

v4.4.0 and ggplot2 (3.5.2). Tumor and admixture (average of stroma and immune scores) enrichment for unbiased proteome data was assessed using the ProteoMixture tool as previously described (7). Hierarchical clustering was performed with pheatmap v1.0.126 using Canberra distance and Ward.D linkage**.** Differential analyses of proteins in whole tumor collections and averaged protein abundances across enriched tumor collections was performed using a z-test and proteins significantly altered (z-test p-value < 0.05, FC $\pm$ 2.0) were included in downstream Hallmark and Canonical Pathway analyses (FDR < 0.05) using MSigDB (8).

292 **Results**

293 **Automated Generation of LMD optimized ROIs with ShapeUpLMD**

294 We developed ShapeUpLMD, an open-source, Python-based tool that generates optimized
295 and spatially defined ROIs for LMD from digital histopathology annotations (Figure 1A). To
296 demonstrate proof of concept, we analyzed fresh-frozen thin sections from 11 USC tissue
297 specimens. Expert pathology review of digitally scanned H&E images confirmed disease diagnosis
298 and provided initial annotations of tumor and non-tumor cell populations. These annotations were
299 manually transferred to digital images of PEN membrane slides pre-cut with fiducials, enabling
300 XML coordinate alignment on LMD microscopes as previously described (4). Tumor and adjacent
301 non-tumor regions were used to train a classifier for each patient tumor section using the HALO
302 digital pathology platform (see Supplementary Figure 1 for additional details). The accuracy of
303 automated ROI masks for tumor cell populations was assessed by direct comparison against
304 manual annotations (Supplementary Figure 2A). Co-registration and direct comparison across
305 eleven patient samples demonstrated high concordance between automated and manual tumor
306 annotations (Figure 1B, average accuracy = 0.87 ± 0.05, Supplementary Table 1). Predicted tumor
307 ROIs also exhibited a significant increase in tumor purity over whole tissue areas (Figure 1C,
308 $86 \pm 6\%$, MWU $p < 0.05$). Annotated regions were exported from HALO as XML coordinate files
309 and used by ShapeUpLMD for ROI optimization. ShapeUpLMD optimizes ROIs with user-
310 defined ROI area constraints (in mm^2) and distances between ROI areas, such that collection with
311 LMD is possible (see Methods). Optimized ROIs and fiducial coordinates are then exported as
312 LMD-compatible XML files. Once imported into LMD microscope software, fiducial coordinates
313 facilitate the alignment of optimized ROIs with corresponding PEN-mounted tissue sections.
314 Using this workflow, we harvested ShapeUpLMD-optimized tumor cell ROIs from eleven USC

315 tissue specimens and then compared pre-collection optimized-ROIs with classifier-determined
316 collected regions to assess performance. Post-collection assessment showed high fidelity of tissue
317 capture and boundary alignment (Figure 1D, average Accuracy = 0.9 ± 0.05, Supplementary Figure
318 2B, Supplementary Table 2).
319
320 **Spatially-Resolved Tumor ROI Collections with ShapeUpLMD**
321 Once ShapeUpLMD has created collectable regions, ROIs can be split into smaller,
322 spatially defined regions to enable high-resolution spatial analyses. To demonstrate this feature,
323 we subdivided tumor ROIs from a representative USC tissue section into three 10 mm^2 regions
324 and six 5 mm^2 regions (Figure 2A). For comparison, we collected whole tissue, enriched tumor
325 ROIs, and spatially defined ROIs of interest. All samples underwent PCT-assisted digestion, and
326 peptides were analyzed by diaPASEF mass spectrometry, resulting in the quantification of 5,294
327 total proteins across whole tumor, enriched tumor and spatially resolved, ShapeUpLMD-optimized
328 tumor cell ROIs. Unsupervised hierarchical clustering revealed distinct protein expression profiles
329 between whole tumor collections compared to enriched tumor ROIs (Figure 2B & 2C). Spatially
330 defined ROIs clustered according to local proximity, reflecting pronounced regional proteomic
331 heterogeneity within individual tumors (Figure 2B & 2C). Proteins elevated in whole versus tumor
332 specific ROI collections (z-test p-value < 0.05, FC ± 2.0) were enriched for pathways regulating
333 epithelial-mesenchymal transition and extracellular matrix dynamics and included multiple
334 collagen isoforms (COL1A2 & COL3A1) (Supplementary Tables 3 & 4). These findings are
335 consistent with greater stromal cell admixture and lower purity in whole tissue collections, as we
336 have previously reported (2).
337

338 **Spatially-Resolved Unbiased ROI Collections with ShapeUpLMD**

339 To evaluate the utility of ShapeUpLMD for unbiased spatial proteomics, we generated
340 whole tissue equivalent ROIs (~1 mm^2 each) spanning the entire tissue area from a representative
341 USC tumor section (Supplementary Figure 3). ShapeUpLMD generated fifty-five uniformly sized
342 ROI that were independently collected by LMD, which were processed and analyzed by DIA mass
343 spectrometry on an Orbitrap Astral mass spectrometer, resulting in 6,536 proteins that were co-
344 quantified across all tissue ROIs.

345 Tumor, stroma, non-tissue / necrotic tissue regions were predicted using the workflow
346 detailed above (Supplementary Figure 1, Supplementary Table 5). Tumor purity and non-tumor
347 (stroma and immune) cell admixture scores were inferred from global proteome data using
348 ProteoMixture (7) Supplementary Table 6). Classifier-based spatial predictions showed strong
349 concordance with proteome-derived estimates of tumor content ($r = 0.63$, P<1E-4) and non-tumor
350 cell admixture ($r = 0.76$, P<1E-4) ProteoMixture scores, demonstrating that ShapeUpLMD enables
351 spatially-resolved sampling that retains underlying biological composition.

352 Unsupervised hierarchical clustering of the top most variably abundant proteins revealed
353 that spatially unbiased ROIs segregated according to predicted tumor and non-tumor composition
354 (Figure 3A). Well-established epithelial (e.g., E-cadherin, CDH1 and (epithelial cellular adhesion
355 molecule, EPCAM) and stromal markers (e.g. fibroblast activating protein, FAP and collagen 1A1,
356 COL1A1) exhibited distinct regional localization patterns (2), with elevated abundance profiles in
357 ROIs enriched for their respective cellular compartments (Figure 3B, Supplementary Table 6).
358 These findings resonate prior studies demonstrating proteomic divergence between tumor and non-
359 tumor (stroma and immune) cell populations in gynecologic malignancies (1,2,9,10) and confirm
360 that ShapeUpLMD enables spatially resolved proteome mapping without prior ROI bias.

Discussion

Recent developments in spatial genomics and proteomics have underscored the importance of precisely and reproducibly characterizing cellular subpopulations in heterogeneous tumor microenvironments. Laser microdissection has emerged as a key enabling technology for this purpose, yet limitations in standardization, reproducibility, and throughput continue to restrict its widespread adoption. A central challenge to scaling LMD lies in the manual generation of regions of interest (ROIs) for enrichment of cellular subpopulations. This process is time-intensive, may expose tissue to ambient conditions for extended periods, and is subject to inter-operator variability.

Our group recently addressed several of these challenges through the development of Malleator (4), a software application that enables direct import of ROI coordinate files generated from digital slide images onto LMD microscopes to support automated tissue collections. Recent methods focused on increasing reproducibility for LMD-based experiments include pyLMD (5) and QuPath-to-LMD (6) which provide similar solutions for working with annotations and LMD scopes. QuPath-to-LMD provides an easy-to-use application online and largely focuses on exporting QuPath annotations to LMD XML files. pyLMD provides exporting functionality, but with additional annotation management, and extensive documentation with examples. In the present work, we extend these efforts in our development of ShapeUpLMD, an open-source, Python-based solution that provides flexible control for LMD ROI optimization parameters, including shape, size, and spacing. ShapeUpLMD further introduces the ability to generate spatially resolved ROIs to support spatial multi-omics workflows. Together, these capabilities provide a comprehensive solution for transforming digital pathology annotations into regionally resolved, LMD-optimized coordinate files compatible with automated collection workflows.

384 We framed the performance of ShapeUpLMD in a cohort of USC tumors, leveraging expert
385 pathologist annotations of tumor and non-tumor cell populations to train robust tumor cell-specific
386 classifiers using the HALO digital pathology platform. Classifier performance was validated by
387 direct comparison with manual annotations, demonstrating high concordance between manual and
388 automated annotation strategies. Notably, these findings indicate that our workflow for training
389 cellular subpopulation-specific ROI classifiers (Supplementary Figure 1) can be readily extended
390 to other digital pathology platforms capable of performing automated classification and ROI
391 export, such as QuPath (11). We further show that ROIs generated by third party software tools
392 like HALO can be directly processed by ShapeUpLMD to generate optimized regions which yield
393 high fidelity LMD collections, confirmed by our post-collection spatial concordance analysis. In
394 summary, our results demonstrate that ShapeUpLMD-generated ROIs generated are highly
395 concordant with manual annotation and, with our prior automation methods (2).

396 ShapeUpLMD also enables generation of spatially defined ROIs tailored to user defined
397 tissue areas and parameters. In our proof-of-concept application, we subdivided a representative
398 USC tumor into multiple spatially defined regions and performed global proteomic profiling. The
399 data revealed distinct global protein profiles between whole tumor and enriched tumor regions.
400 Proteins elevated in whole tumor regions included multiple collagen isoforms consistent with
401 admixture of non-tumor cells (i.e. stroma cell populations) (1,2). The proteomic profiles from
402 spatially resolved ROIs clustered according to local proximity, revealing region-specific proteomic
403 consistency within tumors and underscoring heterogeneity across tumor cell populations. To
404 further assess ShapeUpLMD for unbiased spatial proteomics, we generated evenly sized ~1 mm^2
405 ROIs spanning an entire USC tumor tissue section. Tumor and non-tumor classifications derived
406 from digital pathology strongly correlated with tumor and non-tumor cell admixture scores

407 inferred directly from companion spatial proteomics data (7), confirming that spatially separated
408 ROIs retain the unique protein abundance profiles common to tumor and non-tumor cell
409 populations (1,2,9,10). These findings illustrate that ShapeUpLMD not only improves the
410 precision of LMD-based enrichment but also facilitates spatially resolved proteomic analyses that
411 uncover regional molecular variation otherwise masked in bulk tumor profiling.

412 Given the demonstrated recovery of high-quality DNA, RNA, and protein from LMD-
413 harvested tissues (1), ROIs generated by ShapeUpLMD should prove useful for multi-omic
414 workflows incorporating next-generation sequencing. Notably, ShapeUpLMD provides flexibility
415 to users to define ROI area size and count, making it adaptable to diverse experimental designs
416 and tissue types.

417 Limitations of this study include the modest cohort size and focus on a single malignancy
418 (USC). ROI precision may vary across different LMD microscope platforms, laser parameters and
419 imaging magnification. However, ShapeUpLMD is designed for broad adaptability through user-
420 defined parameters, while also being readily refinable for diverse tissues and digital pathology
421 ecosystems.

422 In summary, ShapeUpLMD provides an open-source solution for automated generation of
423 LMD-optimized and regionally resolved ROIs. By bridging digital pathology, machine learning,
424 and automated LMD, this pipeline enables reproducible, high-throughput, and spatially resolved
425 tissue enrichment. As spatial multi-omics evolves toward comprehensive tissue atlases and
426 clinically integrated diagnostics, ShapeUpLMD provides a scalable foundation for high-fidelity
427 sampling across laboratories, diseases, and analytical platforms – advancing precision pathology
428 and accelerating the next generation of spatial systems biology. ShapeUpLMD is available at:
429 https://github.com/GYNCOE/ShapeUpLMD.

Declarations

Ethics approval and consent to participate

Fresh-frozen tumor tissues were selected from patients diagnosed with uterine serous carcinoma enrolled in a WCG IRB approved #20110222 Tissue and Data Acquisition Study of Gynecologic Disease who underwent surgery at Inova Fairfax Medical Campus (Falls Church, VA, USA), University of Oklahoma Health Sciences (Oklahoma City, OK, USA) or at University of Virginia (Charlottesville, VA, USA)

Consent for publication

Not applicable

Availability of data and materials

ShapeUpLMD is available at: https://github.com/GYNCOE/ShapeUpLMD; please contact batemann@whirc.org or conrads@whirc.org for access. Representative input image for ShapeUpLMD testing is available here: https://lmdomics.org/ShapeUpLMD/. Proteomic data generated in this study is accessible within the Mass Spectrometry Interactive Virtual Environment (MassIVE) resource under accession MSV000100553 and the ProteomeXChange (PRIDE) resource under accession PXD073411.

Competing interests

Nothing to disclose.

Funding

Funding for this project was provided from the Uniformed Services University of the Health Sciences from the Defense Health Program to the Henry M Jackson Foundation (HJF) for the Advancement of Military Medicine Inc. Gynecologic Cancer Center of Excellence Program,

including awards HU0001-20-2-0033, HU0001-21-2-0027, HU0001-22-2-0016, and HU0001-23-2-0038.

Authors' contributions

J.S., D.M., T.P.C and N.W.B. lead the study. D.M. and S.T. performed identification of clinical specimens, D.M., K.N.W. and K.A.C., performed sample collections and molecular extraction. B.L.H, J.L., N.W.B., J.O. generated and analyzed proteomics data. P.M.F. performed pathology review. J.S., D.M., T.P.C and N.W.B wrote the manuscript, S.T., K.N.W., K.A.C., J.L., B.L.H, T.A., A.L.H., A. L. K.M.D., C.M.T., G.L.M reviewed the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgements

We would like to acknowledge Emily A. Pennington for her review and feedback on the ShapeUpLMD software code.

Disclaimer

The views expressed herein are those of the authors and do not reflect the official policy of the Uniformed Services University of the Health Sciences, the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Inova Health System, the Department of Army/Navy/Air Force, Department of War, or U.S. Government. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

References

1. Bateman, N. W., Abulez, T., Soltis, A. R., McPherson, A., Choi, S., Garsed, D. W., Pandey, A., Tian, C., Hood, B. L., Conrads, K. A., Teng, P. N., Oliver, J., Gist, G., Mitchell, D., Litzi, T. J., Tarney, C. M., Crothers, B. A., Mhawech-Fauceglia, P., Dalgard, C. L., Wilkerson, M. D., Pierobon, M., Petricoin, E. F., Yan, C., Meerzaman, D., Bodelon, C., Wentzensen, N., Lee, J. S. H., Network, A. R., Huntsman, D. G., Shah, S., Shriver, C. D., Phippen, N. T., Darcy, K. M., Bowtell, D. D. L., Conrads, T. P., and Maxwell, G. L. (2024) Proteogenomic analysis of enriched HGSOC tumor epithelium identifies prognostic signatures and therapeutic vulnerabilities. *NPJ Precis Oncol* **8**, 68
2. Hunt, A. L., Bateman, N. W., Barakat, W., Makohon-Moore, S. C., Abulez, T., Driscoll, J. A., Schaaf, J. P., Hood, B. L., Conrads, K. A., Zhou, M., Calvert, V., Pierobon, M., Loffredo, J., Wilson, K. N., Litzi, T. J., Teng, P. N., Oliver, J., Mitchell, D., Gist, G., Rojas, C., Blanton, B., Darcy, K. M., Rao, U. N. M., Petricoin, E. F., Phippen, N. T., Maxwell, G. L., and Conrads, T. P. (2024) Mapping three-dimensional intratumor proteomic heterogeneity in uterine serous carcinoma by multiregion microsampling. *Clin Proteomics* **21**, 4
3. Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovacs, F., Brunner, A. D., Migh, E., Schweizer, L., Santos, A., Bzorek, M., Naimy, S., Rahbek-Gjerdrum, L. M., Dyring-Andersen, B., Bulkescher, J., Lukas, C., Eckert, M. A., Lengyel, E., Gnann, C., Lundberg, E., Horvath, P., and Mann, M. (2022) Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat Biotechnol* **40**, 1231-1240
4. Mitchell, D., Hunt, A. L., Conrads, K. A., Hood, B. L., Makohon-Moore, S. C., Rojas, C., Maxwell, G. L., Bateman, N. W., and Conrads, T. P. (2022) Industrialized, Artificial Intelligence-guided Laser Microdissection for Microscaled Proteomic Analysis of the Tumor Microenvironment. *J Vis Exp*
5. Schmacke, N. A., Mädler, S. C., Wallmann, G., Metousis, A., Bérouti, M., Harz, H., Leonhardt, H., Mann, M., and Hornung, V. (2023) SPARCS, a platform for genome-scale CRISPR screening for spatial cellular phenotypes. *bioRxiv*, 2023.2006.2001.542416
6. Makhmut, A., Qin, D., Hartlmayr, D., Seth, A., and Coscia, F. (2024) An Automated and Fast Sample Preparation Workflow for Laser Microdissection Guided Ultrasensitive Proteomics. *Mol Cell Proteomics* **23**, 100750
7. Teng, P. N., Schaaf, J. P., Abulez, T., Hood, B. L., Wilson, K. N., Litzi, T. J., Mitchell, D., Conrads, K. A., Hunt, A. L., Olowu, V., Oliver, J., Park, F. S., Edwards, M., Chiang, A., Wilkerson, M. D., Raj-Kumar, P. K., Tarney, C. M., Darcy, K. M., Phippen, N. T., Maxwell, G. L., Conrads, T. P., and Bateman, N. W. (2024) ProteoMixture: A cell type deconvolution tool for bulk tissue proteomic data. *iScience* **27**, 109198
8. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550
9. Hunt, A. L., Bateman, N. W., Barakat, W., Makohon-Moore, S., Hood, B. L., Conrads, K. A., Zhou, M., Calvert, V., Pierobon, M., Loffredo, J., Litzi, T. J., Oliver, J., Mitchell, D., Gist, G., Rojas, C., Blanton, B., Robinson, E. L., Odunsi, K., Sood, A. K., Casablanca, Y., Darcy, K. M., Shriver, C. D., Petricoin, E. F., Rao, U. N. M., Maxwell, G. L., and Conrads, T. P. (2021) Extensive three-dimensional intratumor proteomic heterogeneity

revealed by multiregion sampling in high-grade serous ovarian tumor specimens. *iScience* **24**, 102757

10. Eckert, M. A., Coscia, F., Chryplewicz, A., Chang, J. W., Hernandez, K. M., Pan, S., Tienda, S. M., Nahotko, D. A., Li, G., Blazenovic, I., Lastra, R. R., Curtis, M., Yamada, S. D., Perets, R., McGregor, S. M., Andrade, J., Fiehn, O., Moellering, R. E., Mann, M., and Lengyel, E. (2019) Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723-728

11. Bankhead, P., Loughrey, M. B., Fernandez, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., and Hamilton, P. W. (2017) QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878

544 **Figure Legends**

545 **Figure 1: Automated generation, optimization, and laser microdissection of tumor cell**
546 **regions of interest (ROI) using ShapeUpLMD. A:** Tumor and non-tumor cell populations are
547 annotated in digital histopathology images by expert pathology review and inform annotation of
548 tumor and non-tumor ROIs on adjacent polyethylene naphthalate (PEN) slide images that include
549 fiducials enabling feature alignment (also See Supplementary Figure 1). Annotated ROIs are used
550 to train a tumor-specific classifier in HALO. Tumor ROIs are then merged with an annotation layer
551 including fiducial coordinates to undergo shape optimization using ShapeUpLMD. Optimized
552 tumor ROIs and fiducial coordinates are then imported onto the laser microdissection microscope
553 for automated sample collection. B: Comparison of manual annotations and automated classifier
554 generated tumor ROIs revealed a high degree of overlap and accuracy between these regions, also
555 with high precision and recall metrics. C: Line plot comparing relative tumor purity in whole slide
556 tissue areas by comparison of automated tumor ROIs to whole tissue areas for 11 USC patient
557 tumors; p-value reflects MWU testing. D: Comparison of automated versus post-collected tumor
558 ROIs revealed a high degree of overlap and accuracy between these regions, also with high
559 precision and recall metrics.

560 **Figure 2: Spatially-resolved proteomics of laser microdissected tumor regions collected using**
561 **ShapeUpLMD. A:** Tumor ROIs can be further compartmentalized by the ShapeUpLMD tool into
562 discrete areas to enable spatial collections. Workflow shown here demonstrates annotation of three
563 tumor ROI areas at 10.0 mm^2 and six tumor ROI areas at 5.0 mm^2 areas for a single USC tissue
564 sample. Tumor ROIs are then collected by LMD and collected tissues undergo pressure-assisted
565 digestion with trypsin followed by liquid chromatography-data independent acquisition mass
566 spectrometry. B: Unsupervised, hierarchical cluster analysis of proteins quantified across all
567 samples comparing whole (all_tissue), enriched tumor (all_tumor) and three 10.0 mm^2 tumor ROI
568 areas. C: Unsupervised, hierarchical cluster analysis of proteins quantified across all samples
569 comparing whole (all_tissue), enriched tumor (all_tumor) and six 5.0 mm^2 tumor ROI areas.

570 **Figure 3: Unbiased spatial proteomic mapping of a uterine serous carcinoma tissue section**
571 **using ShapeUpLMD.** A: Unsupervised hierarchical clusters analysis of the top 1,000 most
572 variably abundant proteins (median absolute deviation) quantified across fifty-five ~1.0mm^2 ROIs
573 spanning a single representative tissue section from a uterine serous carcinoma tumor. Metadata
574 reflects predicted tumor or stroma & immune admixture scores calculated using ProteoMixture (7)
575 and is categorized as tertiles, Q3 = increased score, Q1 = decreased score. B: Representative
576 hematoxylin and eosin (H&E) stained tissue section, admixture predicted from automated
577 classifications defined as tumor, stroma and other (necrotic and tissue-free space) ROIs, as well as
578 spatial maps of representative tumor protein markers (E-Cadherin, CDH1 and epithelial cellular
579 adhesion molecule, EPCAM) as well as representative stroma markers (fibroblast activating
580 protein, FAP and collagen 1A1, COL1A1), derived from spatial proteomic analyses.

581 **Supplementary Figure 1.** Workflow for generating automated classifiers of tumor cell
582 populations in the HALO Image Analysis Platform.

583 **Supplementary Figure 2.** A: Representative USC tumor tissue section with matched tumor
584 regions notated manually or using our automated classifier method. B: Representative USC tumor

585 tissue section with matched tumor regions annotated using ShapeUpLMD-optimized tumor ROIs
586 (classifier) or following collection of ROIs of interest by LMD.

587 **Supplementary Figure 3.** Representative USC tumor H&E stained tissue section and fifty-five
588 ShapeUpLMD-optimized, unbiased tissue ROIs ± ~1.0 mm^2 spanning the entire tissue section.

589 **Supplementary Table Legends**

590 **Supplementary Table 1:** Performance metrics comparing manual versus automated
591 classifications of tumor ROIs for n11 USC tumor tissue samples.

592 **Supplementary Table 2:** Performance metrics comparing estimated collected areas versus
593 automated classifications of tumor ROIs for n11 USC tumor tissue samples.

594 **Supplementary Table 3:** Z-test results comparing proteome alterations in average enriched tumor
595 versus whole regions.

596 **Supplementary Table 4:** Pathway analysis of proteins elevated in enriched tumor versus whole
597 tissue collections.

598 **Supplementary Table 5:** Tissue section ROIs for unbiased spatial proteomics analyses; details
599 tissue regions, ROI area (mm^2) as well as percent tumor, stroma or other admixture predicted from
600 automated classifiers generated using the workflow in Supplementary Figure 1 informed from
601 review of companion H&E tissue sections.

602 **Supplementary Table 6:** ProteoMixture scores calculated from spatial proteomics data and
603 CDH1, EPCAM, FAP and COL1A1 protein abundance from unbiased spatial proteomics analyses.

604

605

606

607

Figure 1

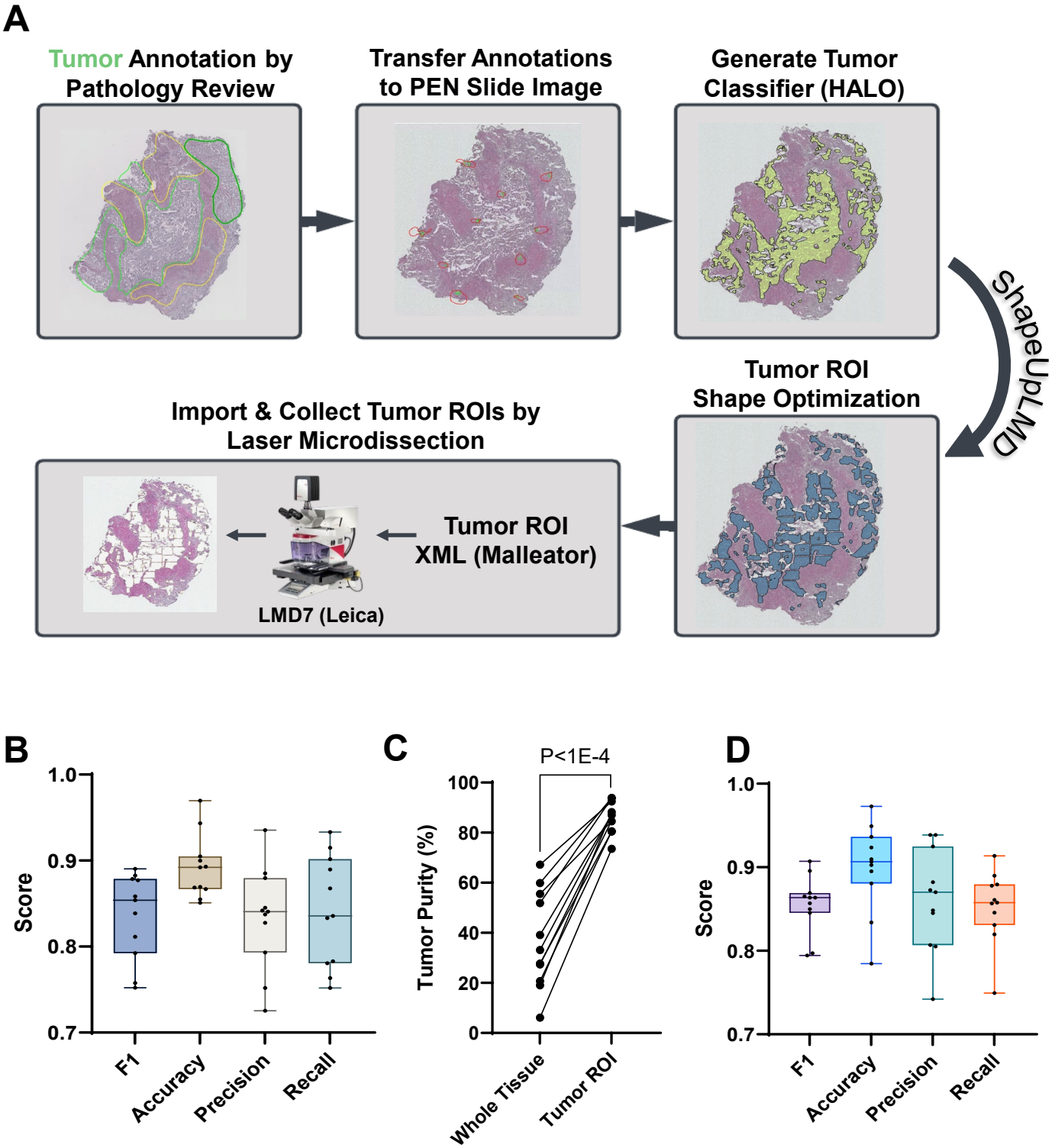

A
Tumor Annotation by Pathology Review
Transfer Annotations to PEN Slide Image
Generate Tumor Classifier (HALO)
shapeUpLMD
Tumor ROI Shape Optimization
Import & Collect Tumor ROIs by Laser Microdissection
Tumor ROI XML (Malleator)
LMD7 (Leica)
B
Score
1.0
0.9
0.8
0.7
F1
Accuracy
Precision
Recall
C
P<1E-4
Tumor Purity (%)
100
80
60
40
20
0
Whole Tissue
Tumor ROI
D
Score
1.0
0.9
0.8
0.7
F1
Accuracy
Precision
Recall

Figure 2

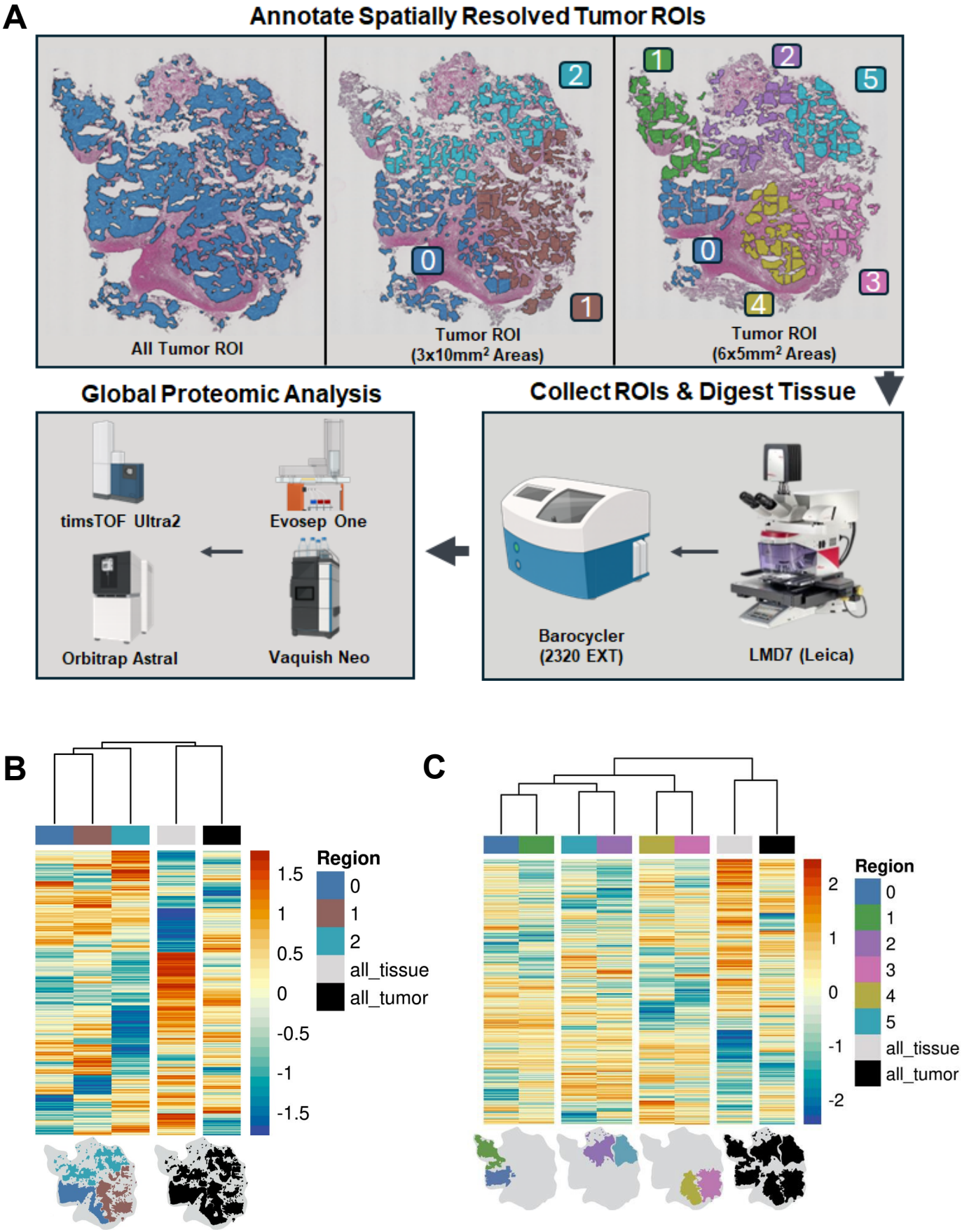
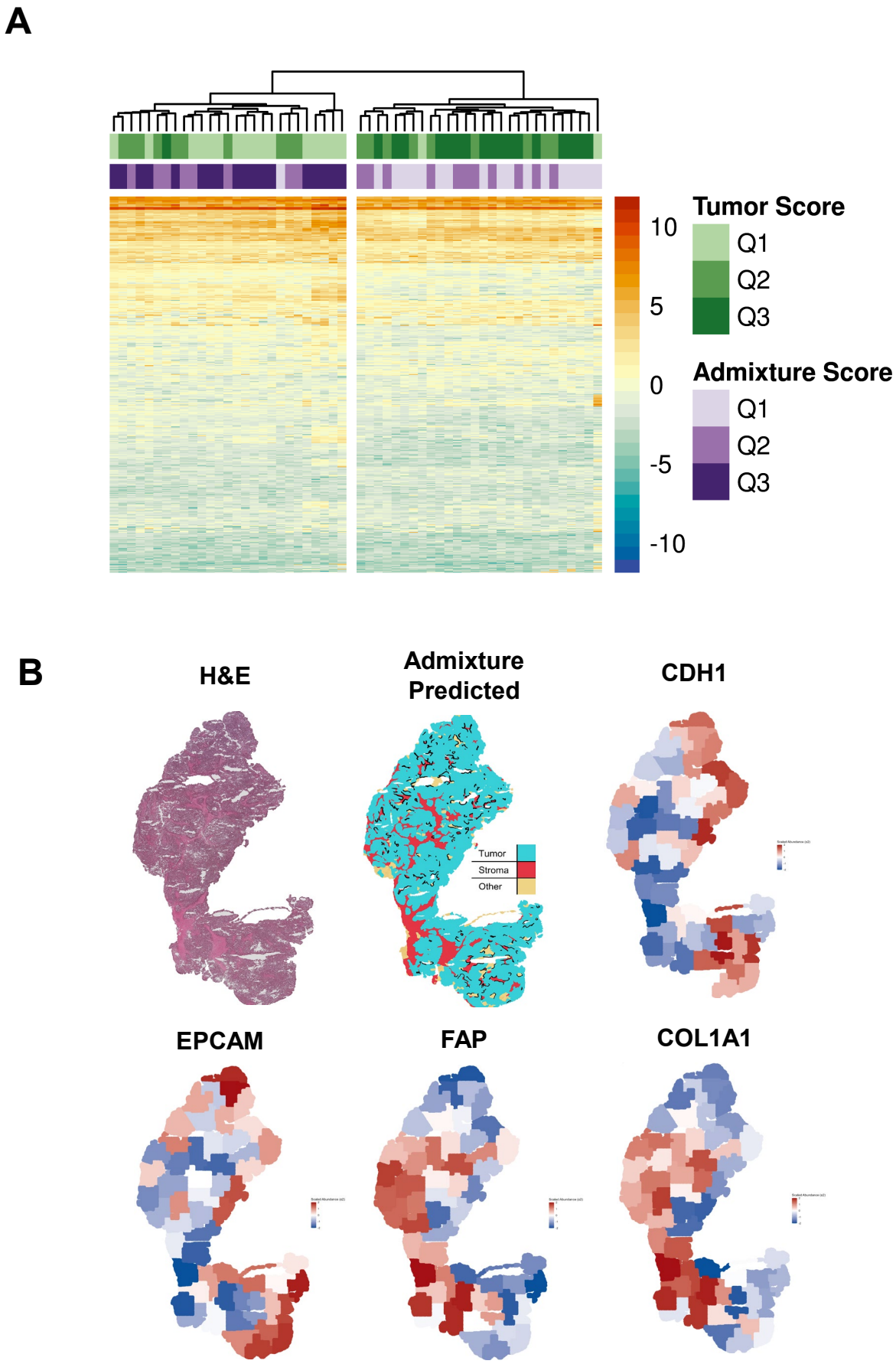
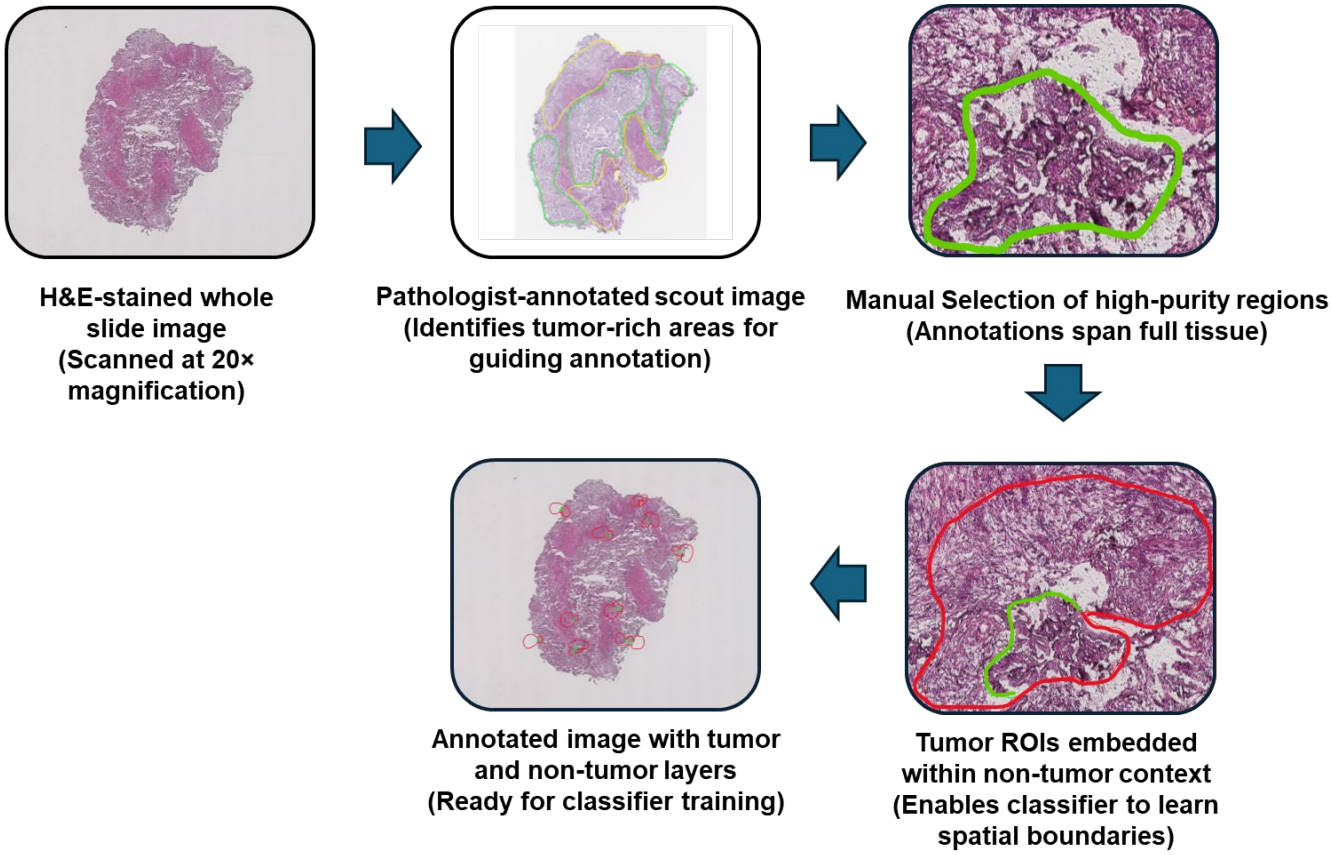

A
Annotate Spatially Resolved Tumor ROIs
All Tumor ROI
Tumor ROI
(3x10mm² Areas)
Tumor ROI
(6x5mm² Areas)
Global Proteomic Analysis
Collect ROIs & Digest Tissue
timsTOF Ultra2
Evosep One
Orbitrap Astral
Vaquish Neo
Barocycler
(2320 EXT)
LMD7 (Leica)
B
Region
0
1
2
all_tissue
all_tumor
1.5
1
0.5
0
-0.5
-1
-1.5
C
Region
0
1
2
3
4
5
all_tissue
all_tumor
2
1
0
-1
-2

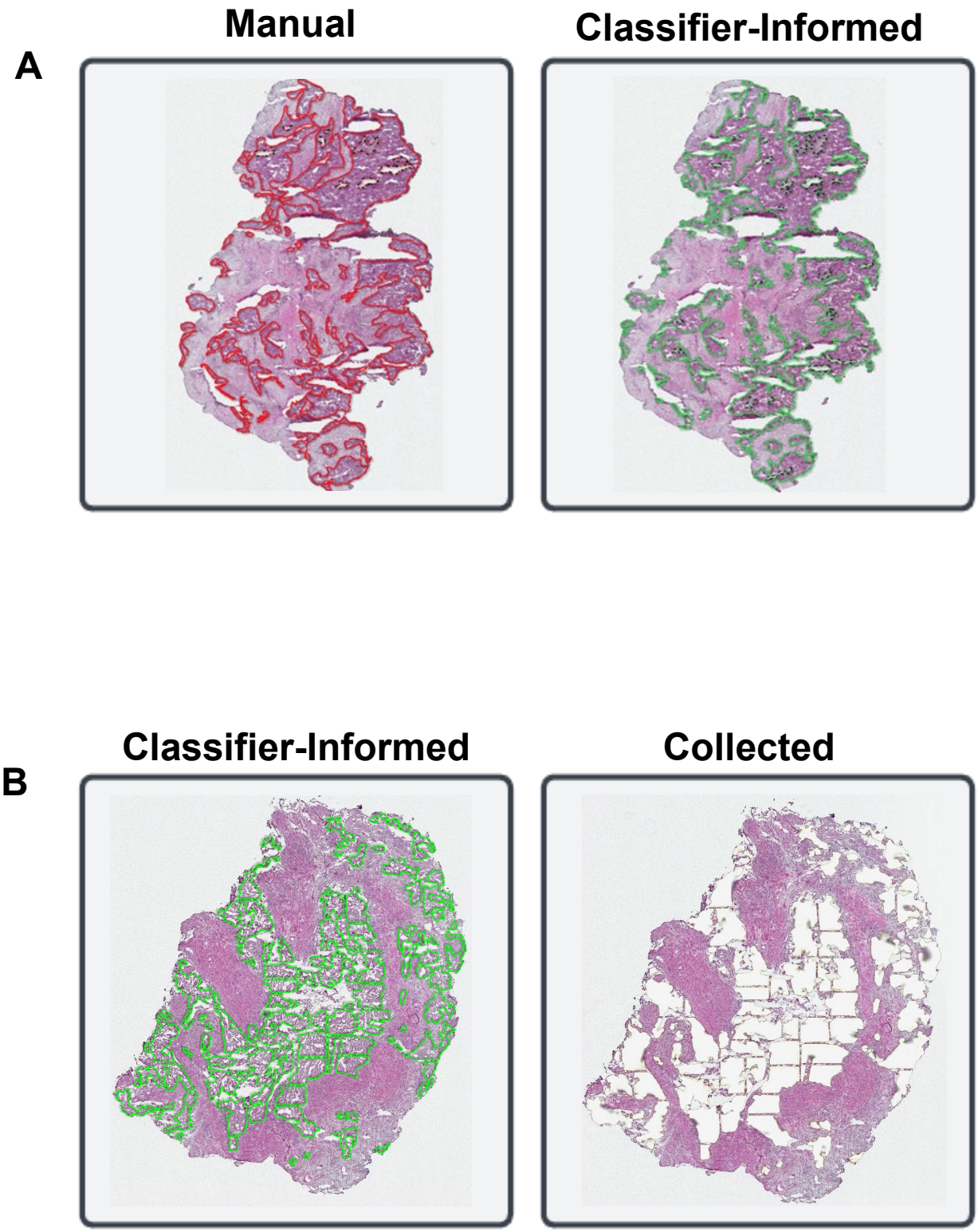
Figure 3


A
10
5
0
-5
-10
Tumor Score
Q1
Q2
Q3
Admixture Score
Q1
Q2
Q3
B
H&E
Admixture Predicted
CDH1
Tumor
Stroma
Other
EPCAM
FAP
COL1A1

Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3

H&E

Unbiased Tissue ROI (55 x ~1.0mm^2)