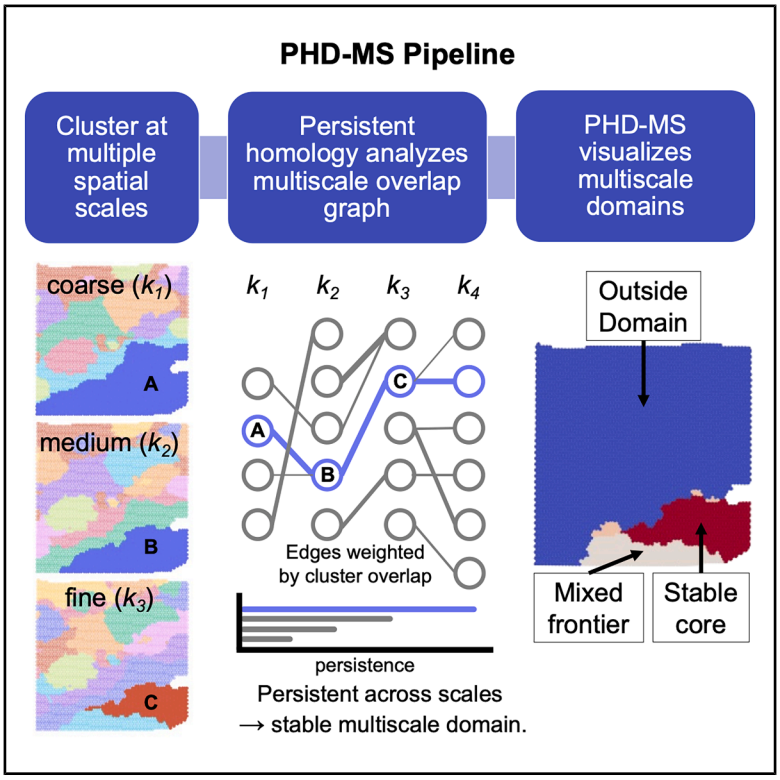
Cell Reports
Methods

Article

Multiscale domain identification for spatial transcriptomics via persistent homology

Graphical abstract



Authors

Perry Beamer, Zixuan Cang

Correspondence

zcang@ncsu.edu

In brief

Beamer et al. introduce PHD-MS, a tool for spatial transcriptomics that finds spatial patterns at multiple scales. By scanning clustering results from broad to fine substructures, it links large tissue domains to their subregions and reveals overlooked within-region heterogeneity in gene expression.

Highlights

- We present PHD-MS, a multiscale clustering tool for spatial transcriptomic data
- PHD-MS identifies interactions between clusters across spatial scales
- Regions of transcriptomic heterogeneity are identified lying between stable clusters
- PHD-MS complements existing methods, revealing multiscale and heterogeneous structure

Beamer & Cang, 2026, Cell Reports Methods 6, 101376
May 18, 2026 © 2026 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.crmeth.2026.101376

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Multiscale domain identification for spatial transcriptomics via persistent homology

Perry Beamer[1] and Zixuan Cang[1,2,3,*]
[1]Department of Mathematics, North Carolina State University, Raleigh, NC, USA
[2]Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, USA
[3]Lead contact
*Correspondence: zcang@ncsu.edu
https://doi.org/10.1016/j.crmeth.2026.101376

MOTIVATION Recent methods aim to segment a tissue into distinct spatial domains of related gene expression by clustering high-throughput spatial transcriptomics (ST) data. However, few methods examine spatial domains across spatial scales. In this study, we present a computational method to identify spatial patterns in gene expression that persist across scales, persistent homology for domains at multiple scales (PHD-MS). By integrating multiple scales, PHD-MS identifies morphological features that other methods do not.

SUMMARY

Spatial transcriptomics (ST) measures gene expression at a set of spatial locations in a tissue. Communities of nearby cells that express similar genes form spatial domains. Specialized clustering algorithms have been developed to identify spatial domains. These methods often locate spatial domains at a single morphological scale, and interactions across multiple scales are often overlooked. For example, large domains often contain smaller substructures and heterogeneous regions may lie between homogeneous domains. Topological data analysis (TDA) is an emerging mathematical toolkit that studies the underlying features of data at various geometric scales, especially useful for analyzing biological datasets with multiscale characteristics. Using TDA, we develop persistent homology for domains at multiple scales (PHD-MS) to locate tissue structures that persist across morphological scales. We apply PHD-MS to highlight multiscale spatial domains across tissue types and ST technologies. We compare PHD-MS domains against expert-annotated ground truth, where PHD-MS outperforms traditional clustering approaches.

INTRODUCTION

Spatial transcriptomics (ST) technologies map gene expression to precise tissue coordinates, generating expression measurements directly linked to spatial locations within a tissue section.[1] The resolution of these measurements varies across platforms, including multicellular spots, single-cell resolution, and subcellular compartments. By coupling gene expression with spatial context, ST enables the study of tissue architecture, microenvironment patterns, and cell-cell interactions with unprecedented spatial details. One major analysis task in ST is domain segmentation, which aims to partition the tissue into contiguous regions that are internally homogeneous in expression and consistent with morphological structure. Spatial domains provide a structural framework for characterizing tissue organization. Within these domains, domain-specific marker genes have been identified,[2] oncogenesis and cancer evolution have been examined,[3] and a variety of analyses in cell biology and anatomy have been conducted. Consequently, numerous clustering methods have been developed specifically for spatial domain segmentation using ST data.[4,5]

Current spatial domain segmentation methods mostly include three steps: a spatial neighborhood representation usually achieved by *k*-nearest neighbor graphs or distance cutoff graphs, a mechanism to integrate expression information within the spatial context such as graph neural networks, and a partitioning step to segment spatial domains by performing clustering on the features integrating expression and spatial information.[5] GraphST,[6] SCAN-IT,[7] SpaceFlow,[8] and STAGATE[9] derive low-dimensional embeddings of spots from graph neural networks and then cluster them using algorithms such as the Leiden clustering. SpaGCN[2] further fuses histology information into the deep learning models. Statistical approaches include BASS[10] and BayeSpace[11] that use the Bayesian framework and BANKSY[12] that derives interpretable augmented spatial features.

Despite strong empirical performance, these approaches often require a user-chosen scale or resolution parameter that controls domain granularity, and in the absence of ground truth, the segmentation results can be sensitive to that choice. In addition, spatial domains are often described as disjoint sets with

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

hard boundaries in current methods. However, biologically, spatial domains are naturally multiscale and heterogeneous. For example, a prominent large-scale domain could contain smaller ones such as the hippocampus containing Ammon's horn, which further contains the CA3 subfield, and so on. A single-scale approach reveals only one level of details. In another example, in the tumor microenvironment (TME), healthy and malignant cells mingle in the tumor frontier,[13] demonstrating the heterogeneous nature of the domain. Revealing the mixing patterns within domains, especially near domain boundaries, as well as uncovering their hierarchical structures, remains largely unexplored. In practice, researchers sweep parameters and visually reconcile multiple partitions. There is a lack of systematic approach to summarize the most stable structure across scales or to quantify how domains split, merge, or interact as the scale changes.

Motivated by this gap, NeST[14] adopts a multiscale hotspot strategy. It identifies local enrichments of individual genes at multiple spatial resolutions and aggregates co-localized enrichments across genes into coexpression hotspots. The hotspot detection is carried out over a range of scales, resulting in structures that vary in size and are naturally nested. This yields a hierarchical map of spatial coexpression hotspots. This framework highlights multiscale signal, but it is not formulated as domain segmentation and does not quantify how putative domains persist or interact across scales. It also does not delineate stable cores versus heterogeneous frontier regions. To address these limitations, we introduce an approach based on topological data analysis (TDA) that summarizes cross-scale cluster evolution through persistence and quantitatively characterizes mixing patterns, thereby delineating the core and frontier regions of heterogeneous tissues.

TDA[15] is a mathematical framework for describing the shape and organization of complex data across multiple scales and dimensions. Persistent homology (PH),[16,17] a central method in TDA, tracks how topological structures such as connected components, loops, and higher-dimensional cavities appear and disappear with varying geometric scales, quantifying their significance by how long each feature persists across scales. In PH, the underlying structures of data across various geometric scales are organized as a filtration, a nested sequence of simplicial complexes constructed by gradually adding connections among discrete data points in the form of edges, triangles, and other higher dimensional simplices. Through the filtration, PH identifies structural features in the form of different dimensional holes with connected components as 0-dimensional holes and loops as 1-dimensional holes. Together, PH provides a compact summary of structural features across scales and dimensions that is inherently robust to noise, which is common in biological datasets.

TDA has been used for robust clustering of omics data. The multiscale clustering filtration (MCF)[18] considers a sequence of clusterings ordered by resolution, from fine to coarse, and uses PH to represent the extent to which the clusterings are nested or hierarchical. However, the MCF is not specifically adapted for spatial data or for transcriptomic data. For single-cell transcriptomics data, HiDef[19] applies PH to study prominence of cellular communities at multiple resolutions. HiDef was developed to analyze non-spatial scRNA-seq data and so does not consider spatial relationships between cell communities. For spatial data, a recent method TopACT[20] employed multiparameter PH to annotate individual cell types using subcellular-resolution spatial transcriptomic data. MCIST[21] achieved improved clustering performance in ST data by encoding the multiscale interactions between cells using persistent Laplacians.

In this paper, we introduce the concept of a multiscale domain, which relates domains identified at small and large scales. A multiscale domain consists of a homogeneous core, where spots are often assigned to the same domain across scale parameters, and heterogeneous frontier regions, where spots are less reliably included. To identify multiscale domains, we introduce a TDA pipeline, called persistent homology for domains at multiple scales (PHD-MS). PHD-MS builds a spatial cluster filtration to identify connections between tissue domains at multiple spatial scales and computes PH to identify prominent multiscale patterns. Our multiscale approach (1) locates domains that remain stable across multiple spatial resolutions and (2) identifies interactions between domains by considering their similarity across scales. In addition, we have implemented a point-and-click visualization utility. When only interested in visualizing the PHD-MS domains around a single point in a tissue, one can utilize this utility to plot all domains containing this point.

We demonstrate the utility of the multiscale domain framework through several case studies. In the Visium mouse brain data, PHD-MS resolves the hierarchical organization of subregions and highlights nested structures within the tissue. When applied to a Visium breast cancer dataset, PHD-MS distinguishes stable and unstable regions of the TME, which correspond to mixed border regions between malignant and healthy tissue. Integrated with differential gene expression analysis, PHD-MS further reveals scale-dependent transcriptional patterns within tumors. Quantitative benchmarking across multiple Visium datasets with ground-truth annotations shows that PHD-MS consistently improves domain identification over single-scale analyses. To support this evaluation, we extend normalized mutual information (NMI) to multiscale clusters and introduce a Wasserstein distance-based metric that explicitly accounts for spatial context. Finally, we demonstrate that PHD-MS generalizes to single-cell-resolution ST, capturing coherent multiscale domains in datasets from MERFISH and osmFISH.

RESULTS

Overview of PHD-MS workflow

PHD-MS analyzes multiscale ST clusters in three stages, aiming to identify links between domains at multiple scales. First, the data are clustered at a sequence of resolution parameters, each of which assigns every cell or spot to a spatial domain. Next, these clustering results are organized as a weighted graph whose nodes represent clusters obtained at different resolutions. This graph serves as input for the final stage, in which PH is computed to identify domains that remain stable across scales. The resulting domain maps provide a multiscale visualization of tissue architecture and highlight how stable and

CellPress
OPEN ACCESS

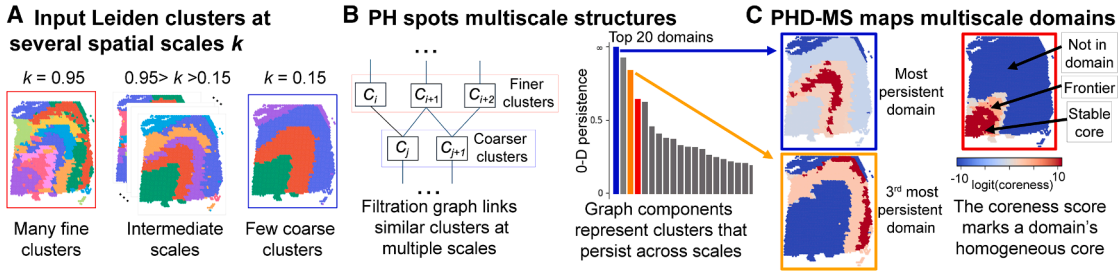


Figure 1. Method overview

(A) PHD-MS considers a sequence of domain segmentations using multiple scale parameters.

(B) A filtration function compares the overlap between clusters, creating a graph that encodes multiscale connectivity of tissue domains. Persistent homology identifies connected components, which represent prominent multiscale communities. These persistent communities are visualized using a persistence diagram.

(C) Using persistent homology, PHD-MS maps multiscale domains. Each point receives a coreness score, which represents its stability within the multiscale domain.

heterogeneous regions are organized across resolutions. A table of related technical terms is provided in Table S1.

In the first step, we consider a sequence of scale parameters $k_1, \ldots, k_n$. Each scale parameter yields a clustering as a collection of disjoint domains, whose sizes typically increase as the parameter decreases. In this way, small domains grow and merge into larger domains as the scale parameter decreases (Figure 1A).

In the second step, a weighted graph is constructed to represent the connections between domains across consecutive scales. To quantify these connections, we measure the pairwise overlap between domains at each pair of the neighboring scales k_i and k_{i+1}, assigning a dissimilarity score f to each pair of clusters (Figure 1B). Values of f lie between 0 and 1, where two identical domains receive $f = 0$, partially overlapping domains receive a score inversely proportional to the number of shared cells, and totally disjoint domains receive $f = 1$. In this graph, each node represents a domain, and edges connect domains that share many of the same cells, weighted by their dissimilarity score. The resulting structure encodes the correspondence between spatial domains across scales.

In the final step, we apply PH to this weighted graph to characterize the hierarchical connections among multiscale domains. In this filtration, all nodes (domains) are present from the start and edges are added sequentially according to their dissimilarity scores. Beginning with only edges of zero dissimilarity, additional connections are introduced in increasing order of dissimilarity. During this process, we record the values of dissimilarity score at which connected components merge. These merging events mark the scale at which previously distinct domains become unified. Components that persist over a wide range of dissimilarity values correspond to stable, recurrent domain structures that remain consistent across scales. A persistence diagram summarizes the persistence of each component (Figure 1B). We then project the persistent components back onto the original tissue to map multiscale domains. Each multiscale domain is represented as a set of per-spot scores, which we call a "coreness" score (Figure 1C). These scores range between 0 and 1 and measure how strongly a spot belongs to the domain's stable core. Spots with high coreness are assigned to the same underlying cluster across many scale parameters and form the domain's homogeneous core, whereas spots with lower coreness belong to the heterogeneous outer region. Each spot receives multiple PHD-MS coreness scores, one for every multiscale domain to which it is partially assigned.

Each multiscale domain is visualized as a spatial map illustrating its stability, where higher coreness values indicate the stable core and lower values correspond to the heterogeneous outer regions (Figure 1C). Note that this representation is not a new clustering of the tissue into distinct domains, because multiscale domains may overlap with one another. In this way, our results can be viewed as a soft clustering of the tissue, in which domains do not rigidly assign cells to unique regions but instead represent broad, overlapping patterns of tissue organization across scales.

Recapitulating hierarchical organization across scales in tissues

We demonstrate the utility of PHD-MS for multiscale domain representation on a Visium dataset of the coronal cross-section of mouse brain.[22] For anatomical reference, we use the Allen Brain Atlas (Figure 2A),[23,24] which provides annotated coronal sections of a mouse brain. For comparison, NeST coexpression hotspots provide an alternative multiscale perspective (Figure 2B). We present PHD-MS domains (Figure 2C), selected by cross-referencing persistent domains against the Allen atlas, identifying the most important fine- and coarse-scale structures in brain tissue.

PHD-MS captures the organization of the coronal mouse brain across scales. At the highest level, it separates the tissue into outer cortical and inner subcortical regions. At the intermediate scale, PHD-MS recovers all major anatomical divisions annotated in the Allen Brain Atlas, including the cerebral cortex, fiber tracts, hippocampus, thalamus, basal nuclei, amygdala, and hypothalamus, and resolves specialized subregions within most of these regions (Figure 2C). In addition to regional identities, PHD-MS distinguishes domains with diverse geometries, including compact, contiguous regions like the thalamus and narrow, elongated structures such as fiber tracts and laminar bands of the cerebral cortex. Together, these results show that PHD-MS recapitulates hierarchical tissue architecture while capturing both broad compartments and fine structures.

In comparison, NeST coexpression hotspots separate the brain into the outer cortex and inner subcortical regions and,

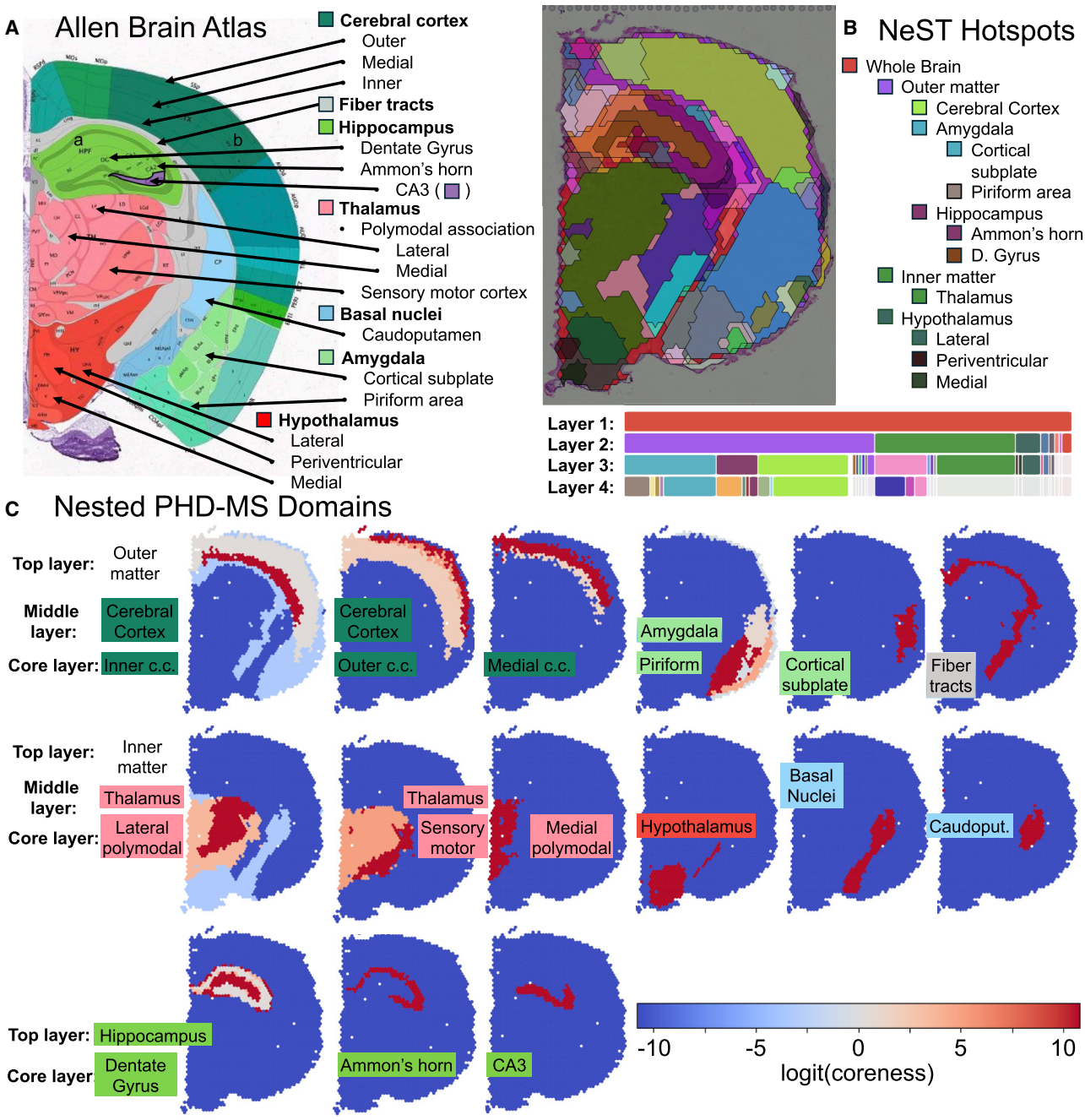


Figure 2. PHD-MS identifies stable brain structures in a Visium mouse brain

(A) An annotated coronal section from Allen Brain Atlas as the ground truth.

(B) NeST coexpression hotspots of the Visium mouse brain data. NeST identifies 5 relevant layers of structure, color coded by layer (bottom), matched to the annotated regions of Allen Atlas (right).

(C) Nested PHD-MS domains with layer labels, color coded with matched regions in Allen Atlas. Top layer represents the domain at its broadest scale, including low-coreness spots. Middle layer includes medium-coreness spots, and core layer forms the domain's stable core.

within these layers, align with major anatomical formations including the cerebral cortex, amygdala, hippocampus, thalamus, and hypothalamus. NeST further recovers several specialized subregions, especially within the hypothalamus, amygdala, and hippocampus. However, some regions do not neatly correspond to Allen Brain Atlas annotations, particularly along the interface between the outer cortex and inner brain. Relative to NeST, PHD-MS identifies several structures that

NeST does not at both the intermediate and core scales. At the intermediate scale, PHD-MS identifies every major annotated region, whereas NeST omits fiber tracts and basal nuclei. At the core scale, PHD-MS resolves additional subregions within the cerebral cortex and thalamus. NeST uniquely identifies certain hypothalamic subregions and most effectively characterizes compact, contiguous formations, but it tends to underrepresent long, thin domains that PHD-MS captures more clearly.

Multiscale domain stability reveals tumor heterogeneity

The TME is highly heterogeneous, particularly along boundaries between tumors and healthy tissue. These border regions may contain a diverse mixture of cell types, including healthy tissue, immune cells, invading malignant cells, and tumor-associated stromal cells.[25] Furthermore, the tumor boundary, often a thin band of cells surrounding a much larger cancerous mass, exists at a much finer spatial scale than the tumor body. Here, we apply PHD-MS to study the TME of an invasive ductal carcinoma (IDC), the most common form of invasive breast cancer. Like other TMEs, the IDC microenvironment is highly heterogeneous along tumor boundaries,[13] where immune cells and tumor-associated stromal cells mix with healthy tissue. This ST dataset was analyzed and annotated by the authors of BayesSpace,[11] who discovered distinct regions of tumor transcriptomic heterogeneity. We extend this analysis across multiple scales, revealing within-region heterogeneity that is not captured by single-scale clusterings.

In Figure 3, we highlight nine PHD-MS multiscale domains that represent the main structural characteristics of the IDC tissue. These nine domains are selected from the top 25 most persistent multiscale domains (weighted by the size of the domain to avoid outliers) to best characterize the distinct morphological structures of the TME. Figure 3A shows annotation of the original immunofluorescence image, where PHD-MS-identified regions are labeled with a white digit, corresponding to the PHD-MS domains in Figure 3B. For reference, the chosen PHD-MS regions achieve an NMI of 0.53 and a mean Wasserstein distance of 535.425 μm relative to the matched ground-truth domains, while the best GraphST clustering achieves an NMI of 0.64 and a mean Wasserstein distance of 382.345 μm. While the single-scale results perform better in quantitative benchmarking, PHD-MS uniquely reveals heterogeneous structures within the TME that are not captured by methods optimized for annotation agreement (Figure 3B). When PHD-MS domains are selected at the annotation scale, the method achieves an NMI of 0.604 and a mean Wasserstein distance of 259.406 μm. We discuss quantitative metrics in the next section, where we focus on quantitatively demonstrating the accuracy of PHD-MS.

In the top row of Figure 3B, we identify three homogeneous tumor regions corresponding to the tumors in the upper part of the tissue. Each of these regions exhibits uniformly high coreness with small low-coreness frontiers, suggesting local mixing at tumor edges. Tumor regions 1 and 2 correspond to IDC regions in immunofluorescent image, while 3 corresponds to regions of benign hyperplasia, suggesting these tumor formations are largely homogeneous in composition. Region 4 corresponds to the major regions of carcinoma *in situ*. Regions 5 and 6 represent mixed healthy/frontier regions, where high-coreness cores mark predominantly healthy tissue surrounded by less-stable border regions adjacent to tumors. Notably, region 5 contains the tumors depicted in region 4 as unstable regions. Similarly, region 9 contains tumor regions 1–3 as border domains. Such interface regions are expected to enrich for immune activity or cancer-associated stromal cells. In panels 7–9, we decompose the large lower IDC tumor into three zones with distinct cores within the larger tumor, delineating local transcriptomic variation within the same broader tumor. We also analyzed the IDC tissue using NeST (Figure S1). While NeST captures the major tumor structures, it does not distinguish between regions 1 and 3 and does not resolve the finer sub-regions labeled 7–9, which are identified by PHD-MS.

To elaborate on PHD-MS' morphological characterization of the IDC tissue, we conduct a differential gene expression (DGE) analysis using the regions identified in panels 1–9 of Figure 3. After identifying highly differentially expressed genes for each domain, we apply Gene Ontology (GO) enrichment analysis[26–28] to identify biological processes associated with these highly expressed genes. Tumor regions do not show significant enrichment for broad GO biological processes. Instead, each tumor exhibits a unique profile of oncogenic transcription. Regions 5 and 6, which contain tumor border regions, express genes associated with immune activity. According to GO analysis, highly expressed genes in regions 5–6 are associated with a wide array of immune processes, including B and T cell signaling, activation, and proliferation; complement activation; tumor necrosis factor production; and several others. Figure 3D shows mean immune expression across a set of known immune genes, which spatially coincide with regions 5 and 6. Consistent with the coreness map, the low-coreness border of region 6 shows higher immune activity than the stable core. Moreover, comparing the three subregions of the large IDC mass (regions 7–9) with DGE analysis, we identify 17 overexpressed genes in region 7 including major IDC oncogene APOE,[29] 144 overexpressed genes in region 8 including oncogene PGK1,[30] and 27 overexpressed genes in region 9 including oncogene MATLAT1.[31] Figure 3E shows the expression of these three major IDC oncogenes, noting that regions of highest expression largely correspond to each of the three identified tumor regions. Unlike the highly stable IDCs 1 and 2, these results suggest that the lower IDC region contains three distinct loci of transcriptomic variation, including differences in the expression of major oncogenes.

Combined with differential gene expression analysis, these results show that PHD-MS-generated multiscale domains decompose a complex TME into constituent parts, highlight relationships between distinct tumor regions, and identify areas of higher and lower heterogeneity within the sample and further within individual tumors. By introducing the coreness and heterogeneity scores, PHD-MS quantifies and characterizes transcriptional variation within the TME.

Multiscale integration improves single-resolution clustering accuracy

PHD-MS domains are unique in two ways: (1) each domain assigns a coreness score between 0 and 1 to each spot and (2)

domains may overlap. Thus, each multiscale domain is represented by a per-spot coreness score vector. In contrast, for domain segmentation at a single-scale (e.g., segmentations by GraphST, SCAN-IT, or Banksy), each transcriptomic spot is assigned to a unique domain. Such domains can be encoded as binary vectors indicating whether a spot belongs to the domain. Hereafter, we refer to these binary partitions as single-scale or binary domains, and to PHD-MS outputs as multiscale domains.

Binary domains are typically compared using indices like the Jaccard index. If ground-truth annotations are available, metrics like NMI or adjusted mutual information can be used to measure how accurately a set of binary domains matches the ground-truth labels. However, in their standard form, these metrics are not directly applicable to multiscale domains. To this end, we introduce two benchmarking metrics, one extending NMI to handle multiscale domains and another capturing spatial relationships. First, we generalize NMI to operate on soft and potentially overlapping domains. The detailed definition is described in the section STAR Methods, and a derivation in the supplemental information shows that the generalized NMI reduces to standard NMI when inputs are binary. Second, we use the Wasserstein distance to quantify the spatial-aware differences between ground-truth annotations and multiscale or single-scale domains. The Wasserstein metric complements NMI by providing a spatially interpretable distance, expressed in the units of the tissue coordinates (microns), that penalizes spatial displacement and enables per-domain evaluation. Equipped with these metrics, we quantitatively assess the accuracy of PHD-MS against ground-truth annotations and compare its performance against existing single-scale methods.

We assess the performance of PHD-MS on several Visium datasets from the LIBD human dorsolateral prefrontal cortex[32] with ground truth from expert annotations. For PHD-MS, we compute NMI and Wasserstein distance relative to the annotations for persistent multiscale domains and select a top-performing subset of domains for each slice. We assess these metrics against three leading single-scale segmentation methods, GraphST, SCAN-IT, and Banksy. Figure 4A shows the ground-truth segmentation alongside segmentations from each single-scale method and the top PHD-MS domains. As illustrated in Figure 4A, PHD-MS paired with SCAN-IT embeddings improves upon single-resolution results and more uniquely delineates the core regions of cortical layers.

We also conduct numerical experiments to demonstrate the robustness of PHD-MS to different types of input. Here, we obtain clusterings at different spatial scales by altering the resolution parameter in the Leiden clustering algorithm. Since PHD-MS results are derived by scanning clusterings generated by several scale parameters, performance depends on the choice of parameter set and the quality of the upstream embeddings. In Figure 4B, we test several parameter schemes on the human DLPFC datasets (using GraphST embeddings) and compare their NMI scores. The default parameter schedule achieves the highest median NMI and the least negative variance. Moreover, the results remain relatively stable across alternative schedules, demonstrating robustness to parameter choices. Figure 4C shows the NMI performance of PHD-MS across a series of perturbed input clusterings. We simulate perturbations by generating 100 unique GraphST embeddings of DLPFC slice 151673 from distinct random seeds. Each unique embedding produces a set of unique Leiden clusterings, which in turn produces a unique set of PHD-MS domains. Across these 100 simulations, the majority of results remain within 0.01 of the median NMI, indicating robustness to embedding variability.

Last, we demonstrate that PHD-MS improves upon single-scale approaches overall (Figures 4D and 4E). For each method, we select the single-scale clustering that best matches the number of domains in ground truth and compute its Wasserstein distance and NMI against the ground-truth annotations. We compare these against the top PHD-MS domains. On average, PHD-MS improves upon all three single-scale methods (GraphST, SCAN-IT, and Banksy), in both metrics. Importantly, PHD-MS results increase in accuracy as the input quality increases. GraphST, on average, produces the most accurate single-scale results, and consequently PHD-MS is also most accurate using GraphST embeddings as inputs. In practice, using the strongest available embeddings yields the best multiscale outcomes. Additionally, runtime and memory profiling show that PHD-MS is computationally efficient without excessive overhead (Figure 4D).

Identifying morphological domains at single-cell resolution

Several ST technologies, such as MERFISH[33] measure a tissue's transcriptomic profile at single-cell resolution, although with typical tradeoffs including limited number of assayed genes or smaller tissue areas. Here, we demonstrate that PHD-MS capably analyzes data at fine spatial resolution. In particular, we demonstrate PHD-MS' capabilities on two representative datasets, a MERFISH mouse hypothalamic preoptic region dataset[34] and an osmFISH mouse somatosensory cortex[35] dataset.

The MERFISH mouse hypothalamic preoptic region dataset consists of several slices, labeled by distance from bregma. Each slice is annotated with ground-truth tissue domains, taken from a previous study by BayesSpace.[11] Figure 5A depicts the bregma −0.24-mm slice, with several persistent PHD-MS domains. These persistent PHD-MS domains isolate the most prominent features of the preoptic region, clearly distinguishing the midline ventricle (V3) from the rest of the hypothalamus. At top right, the thalamus is isolated, whose core aligns with the fornix (fx), and beneath it, hypothalamic tissue is identified. At bottom left, PHD-MS locates the medial preoptic area and indicates that this region is closely associated with the periventricular hypothalamus. As with the Visium data, we compute the Wasserstein distance between the ground-truth domains and PHD-MS multiscale domains. Across slices, most ground-truth domains are identified with high accuracy (Figure 5B). The majority of ground-truth domains match an identified multiscale domain within 5–10 microns in Wasserstein distance, while the average distance between neighboring MERFISH spots is about 0.5 microns,[33] indicating a deviation of approximately 10–20 spot widths. When normalized by spot spacing, this performance is comparable to that observed on Visium data in the previous section.

We also compute Wasserstein distances for the osmFISH somatosensory cortex results (Figure 5C). We consistently observe

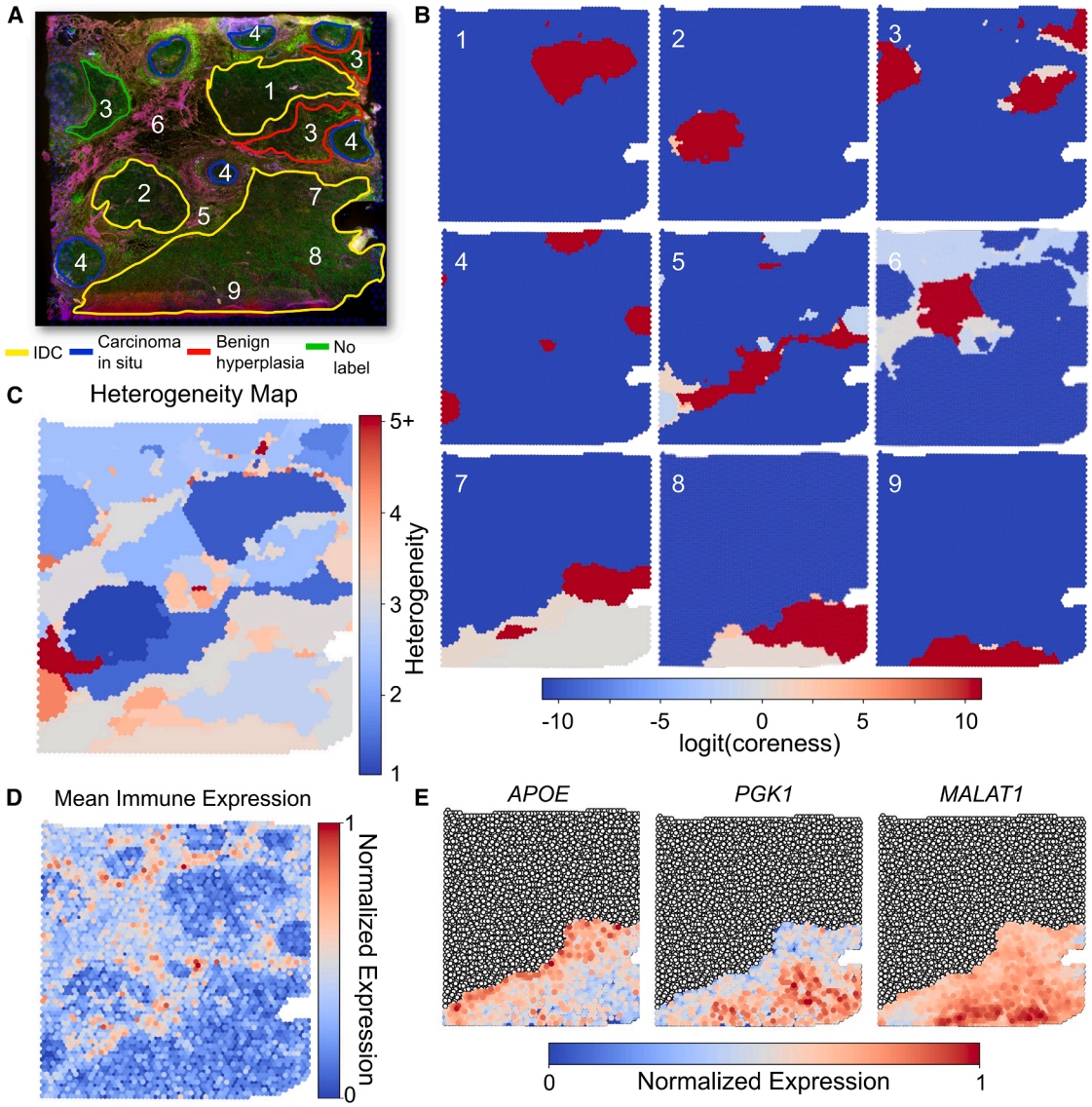


Figure 3. PHD-MS analyzes tumor microenvironment morphology

(A) Expert annotated tumors from immunofluorescent image, with anti-CD3 intensity in pink and DAPI intensity in green. IDC is outlined in yellow, carcinoma *in situ* is outlined in blue, benign hyperplasia is outlined in red, and unclassified tumor is in green.

(B) Prominent multiscale domains identified by PHD-MS, selected from the most persistent domains. Regions are numbered to correspond with the labels in (A).

(C) PHD-MS heterogeneity map highlights highly heterogeneous regions. The heterogeneity score approximates the number of persistent domains that contain each spot.

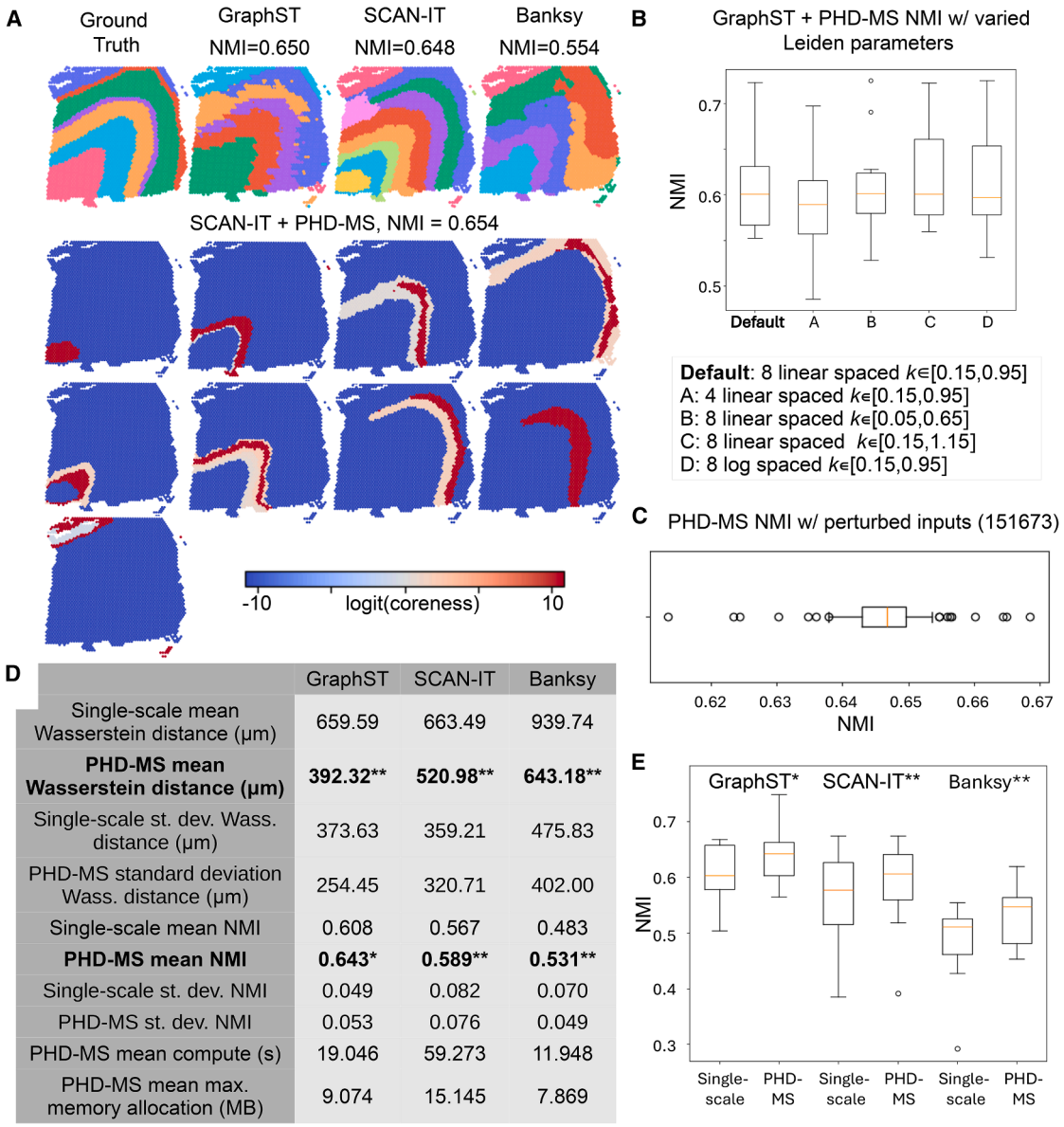
(D) Mean normalized immune expression from a set of immune genes (PTPRC, CD4, CD8A, CD14, CD68, and IGHG3).

(E) Normalized expression of key oncogenes APOE, PGK1, and MALAT1 in the lower IDC. From DGE analysis, APOE is overexpressed in region 7 relative to 8 and 9, PGK1 is overexpressed in 8, and MALAT1 is overexpressed in 9.

low Wasserstein distances, comparable to those from the MERFISH and Visium analyses. We also compare the ground-truth annotation with several PHD-MS domains (Figure 5D). At the top left, PHD-MS identifies white matter as the core of a region that also includes all other major white matter structures: the lateral ventricle (V), corpus callosum (ICC), and pia layer (Pia L). PHD-MS further isolates the entire somatosensory cortex, at the top center panel, and subdivides it into several layers visible in subsequent domains. These results demonstrate that PHD-MS effectively captures hierarchical and spatially coherent tissue organization at single-cell resolution.

DISCUSSION

We introduce PHD-MS, a tool for analyzing tissue domains across spatial scales. PHD-MS is based on PH, constructing a weighted cross-scale overlap graph that ranks domain overlaps and adds edges in order of dissimilarity to merge connected



D

	GraphST	SCAN-IT	Banksy
Single-scale mean Wasserstein distance (μm)	659.59	663.49	939.74
PHD-MS mean Wasserstein distance (μm)	**392.32****	**520.98****	**643.18****
Single-scale st. dev. Wass. distance (μm)	373.63	359.21	475.83
PHD-MS standard deviation Wass. distance (μm)	254.45	320.71	402.00
Single-scale mean NMI	0.608	0.567	0.483
PHD-MS mean NMI	**0.643***	**0.589****	**0.531****
Single-scale st. dev. NMI	0.049	0.082	0.070
PHD-MS st. dev. NMI	0.053	0.076	0.049
PHD-MS mean compute (s)	19.046	59.273	11.948
PHD-MS mean max. memory allocation (MB)	9.074	15.145	7.869

Figure 4. Quantitative benchmarking on Visium DLPFC

(A) Method comparison on slice 151673, including ground-truth annotations; single-scale segmentations from GraphST, SCAN-IT, and Banksy; and a collection of PHD-MS domains generated using SCAN-IT inputs across scales.

(B) Boxplots comparing NMI of PHD-MS results (constructed with GraphST embeddings) using several sets of input Leiden resolutions, where the default scheme produces the greatest median with the least negative variance. Each boxplot represents 12 samples (one for each DLPFC slice).

(C) Boxplot comparing PHD-MS NMI across 100 GraphST inputs generated from unique random seeds on slice 151673. Majority of results remain within 0.01 NMI across simulated perturbations.

(D) Table comparing PHD-MS performance across input clustering methods. Regardless of input type, PHD-MS reports statistically significant improvement over single-scale methods with low computational cost.

(E) Boxplot of PHD-MS NMI across methods. PHD-MS outperforms single-scale approaches across the board. Each boxplot represents 12 samples, one for each DLPFC slice.

In (D and E), $*p < 0.05$, $**p < 0.01$ represent statistically significant improvements over single-scale results in Wilcoxon signed-rank test. For boxplots in (B, C, and E), the box represents data between the first quartile (Q1) and third quartile (Q3), the red line represents the median, whiskers represent data within 1.5 inter-quartile range (IQR) of Q1 or Q3, and flier points signify data outside 1.5 IQR of Q1/Q3.

components. This reveals regions that remain stable across clustering results. We provide visualizations that decompose tissue into prominent multiscale structures and identify within-domain cores, yielding a more detailed picture of morphology than any single fixed-scale clustering. We introduce benchmarking metrics (a multiscale NMI and a spatial Wasserstein

CellPress
OPEN ACCESS

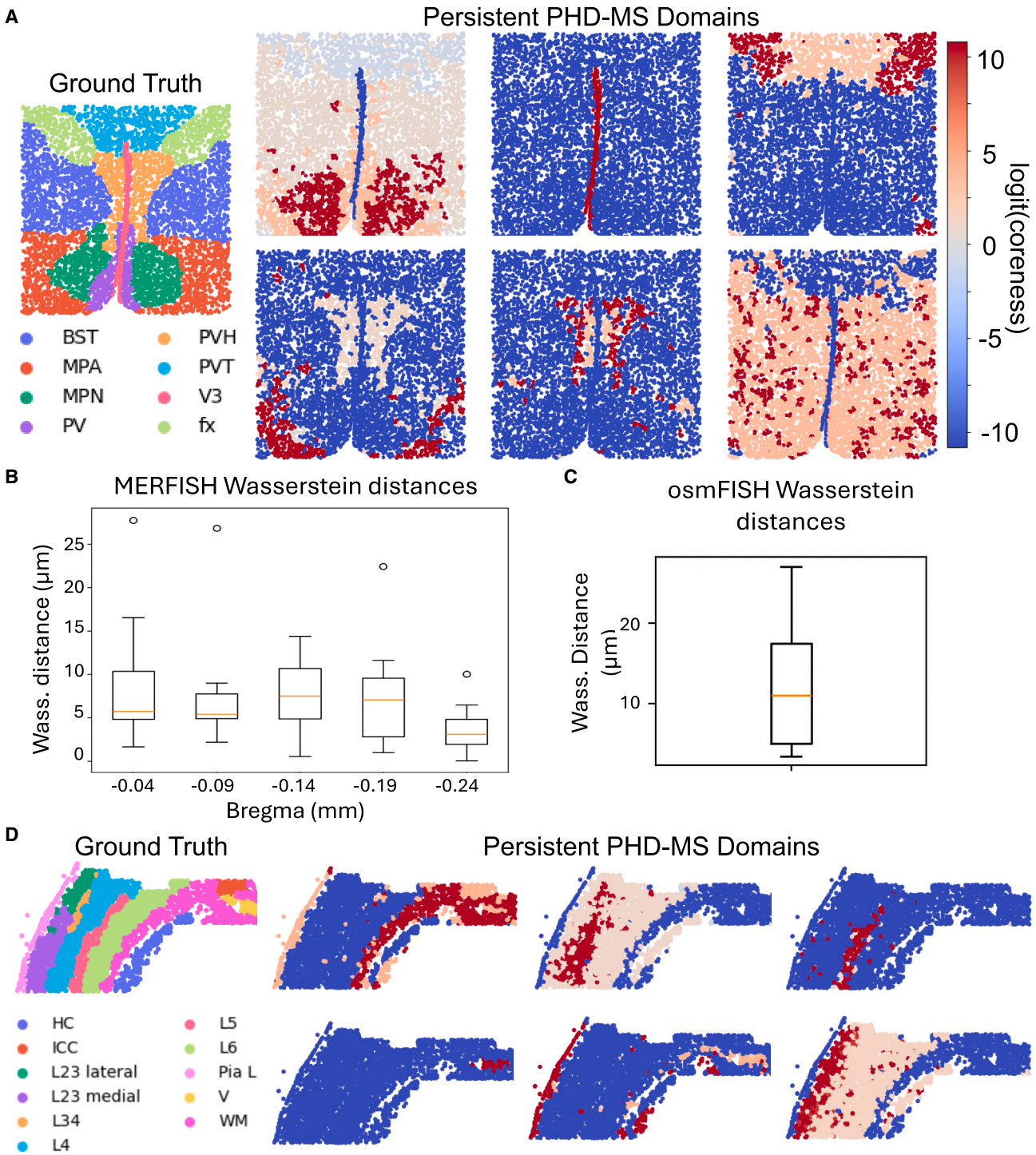


Figure 5. PHD-MS reveals patterns at single-cell scale

(A) Ground truth and persistent PHD-MS domains of the MERFISH mouse brain, bregma = −24.

(B) Boxplot of Wasserstein distance between MERFISH slices and their best-matching PHD-MS multiscale domain. Each boxplot includes one sample for each ground-truth domain; e.g., for bregma −0.24, there are 8 ground-truth domains, yielding 8 boxplot samples.

(C) Boxplot of Wasserstein distance between osmFISH ground-truth domains and their best-matching PHD-MS multiscale domain. The boxplot represents 11 samples, one for each ground-truth domain.

(D) Ground truth and persistent PHD-MS domains for osmFISH.

For boxplots in (B and C), the box represents data between the first quartile (Q1) and third quartile (Q3), the red line represents the median, whiskers represent data within 1.5 IQR of Q1 or Q3, and flier points signify data outside 1.5 IQR of Q1/Q3.

distance) to compare against ground-truth annotations. Finally, the tool is available as an open-source package with a user-friendly graphical interface for exploring multiscale tissue domains.

Because of the design of the method, low-coreness regions could also arise from technical noise rather than biological heterogeneity. These cases are challenging to distinguish without biological prior knowledge, but they can be partially assessed through quality-control (QC) metrics. Specifically, we compare coreness scores with standard QC metrics such as percentage of mitochondrial genes and total counts and inspect their spatial maps to evaluate whether low-coreness areas coincide with regions of low data quality. In the IDC example where low-coreness regions were used to interpret tumor borders, the QC maps show no localized low-quality regions, supporting the biological interpretation of low-coreness areas (Figure S2). Furthermore, PHD-MS outputs are stable under simulated perturbations (Figure 4C), indicating its robustness to mild technical variation. In future analyses, especially when biological knowledge is limited, coreness scores can be compared with QC metrics, and their spatial co-localization can be used to assess biological relevance.

PHD-MS advances multiscale domain analyses beyond recent frameworks such as NeST[14] and SCALE,[36] introducing several methodological innovations. As shown in Figure 2, PHD-MS identifies hierarchical tissue structures that are not recovered by NeST. Conceptually, NeST characterizes hierarchical organization by detecting co-expression hotspots, whereas PHD-MS integrates information across resolutions to capture both hierarchical and non-hierarchical relationships among domains. SCALE characterizes hierarchical structures while focusing on identifying a single optimal scale for clustering. In contrast, PHD-MS aggregates information from all scales to produce a coherent set of multiscale domains. Moreover, PHD-MS introduces a quantitative coreness score that measures the membership strength of each spot within a domain, enabling numerical comparison across domains and datasets. We also introduce the PHD-MS heterogeneity score to identify regions of unstable cluster assignment across scales. Unlike existing heterogeneity metrics, including entropy metrics[37] such as Entropy-based Local Indicator of Spatial Association,[38] the PHD-MS heterogeneity score is not limited to a single morphological scale by a variable scale parameter. The IDC analysis demonstrates how the coreness and heterogeneity scores facilitate a detailed examination of intra-domain heterogeneity.

At a broader level, the need for multiscale domain identification arises because a series of single-scale clusterings, while informative at each resolution, does not encode relationships across scales. Tissue regions interact across scales and are frequently nested. In the brain, for example, the CA3 subfield sits within Ammon's horn, which sits within the hippocampus, which in turn belongs to the limbic system. Separate single-scale analyses may recover each region but miss their hierarchical relations. By contrast, PHD-MS explicitly represents domain-subdomain linkage across resolutions and visualizes these cross-scale dependencies. Beyond hierarchy, many tissues exhibit both homogeneous regions with stable cell type composition and heterogeneous regions or region borders. In cancer microenvironments, core tumor regions dominated by malignant cells and healthy tissue are separated by frontier regions with mixed healthy and unhealthy cells, precluding strict boundaries. Traditional single-scale clustering imposes rigid partitions and underrepresents such interfaces, whereas PHD-MS recognizes overlapping domains across scales and marks the mixed healthy-malignant interface as an unstable subregion of both domains.

In summary, these considerations motivate the adoption of multiscale domain identification. PHD-MS complements existing spatial clustering methods by converting a collection of single-scale clustering results into an organized and interpretable map of multiscale domains. Because PHD-MS is a downstream analysis of a collection of clusterings, it can be paired with any spatial domain-clustering pipeline. Here, we illustrate it using GraphST, SCAN-IT, and Banksy as representatives. Future work could incorporate statistical investigations, such as spatial bootstraps, to quantify the uncertainty in persistences and coreness scores and to better distinguish biological heterogeneity. Further development could enable applications to volumetric data, explore subcellular patterns in high-resolution ST data, and incorporate additional modalities such as proteomics. Mathematically, PHD-MS analyzes 0-dimensional connected components of the cluster filtration, but further study of its higher-dimensional homological features such as 1-dimensional loops and 2-dimensional cavities could reveal higher-order relationships between domains. PHD-MS is computationally efficient in our experiments (Figure S3), and its cost is driven mainly by cross-scale overlap computation and 0D persistence. For larger fields of view, practical speedups may include omitting extremely fine clusterings and approximating overlap weights through spot subsampling.

Limitations of the study

Before applying PHD-MS to ST data, the user must generate input domain segmentations at several spatial scales. Because of this design, PHD-MS' accuracy is tied to the accuracy of its input. While PHD-MS remains relatively stable as input quality degrades, more accurate input clusterings generally produce more accurate PHD-MS domains.

Similarly, PHD-MS results depend on the quality of ST reads. For example, low coreness scores typically represent regions of higher transcriptomic heterogeneity, but low coreness could also arise from technical instability. Therefore, we recommend verifying the quality of ST data using standard QC metrics before applying PHD-MS.

PHD-MS produces overlapping "soft" domains as output, where each transcriptomic spot receives a coreness score rather than a unique domain assignment. For this reason, PHD-MS result must be adapted for any downstream analyses that require distinct, discrete clusters. For example, we must introduce a modified NMI score before evaluating the accuracy of PHD-MS domains against ground-truth annotations.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Zixuan Cang (zcang@ncsu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The Visium mouse brain data[39] and MERFISH data[34] were accessed via the Squidpy Python package. Visium invasive ductal carcinoma data are available at https://www.10xgenomics.com/datasets. Visium human dorsolateral prefrontal cortex data[32] are available from https://figshare.com/articles/dataset/Visium_DLPFC_preprocessed. OSMFISH data[35] are available from https://figshare.com/articles/dataset/osmFISH_datasets. The preprocessed input data for PHD-MS are available on FigShare at https://figshare.com/s/a2561a1090a252183479.
- The open-source implementation of PHD-MS (with example usage) is available on GitHub at https://github.com/pzbeamer/phd-ms, https://doi.org/10.5281/zenodo.18733158.
- Any additional information required to reanalyze the results reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by NSF grant DMS2151934 and NIH grant R01GM152494.

AUTHOR CONTRIBUTIONS

Conceptualization, P.B. and Z.C.; methodology, P.B.; investigation, P.B. and Z.C.; writing – original draft, P.B.; writing – review & editing, P.B. and Z.C.; funding acquisition, Z.C.; resources, Z.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

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

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Cluster filtration
  - Persistent homology results and analysis
  - Heterogeneity score
  - Wasserstein distance
  - Normalized mutual information
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Differential gene expression
  - Quantitative benchmarking

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2026.101376.

Received: July 25, 2025
Revised: November 3, 2025
Accepted: March 4, 2026
Published: March 30, 2026

REFERENCES

1. Moses, L., and Pachter, L. (2022). Museum of spatial transcriptomics. Nat. Methods *19*, 534–546.
2. Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D.J., Lee, E.B., Shinohara, R.T., and Li, M. (2021). Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat. Methods *18*, 1342–1351. https://doi.org/10.1038/s41592-021-01255-8.
3. Seferbekova, Z., Lomakin, A., Yates, L.R., and Gerstung, M. (2023). Spatial biology of cancer evolution. Nat. Rev. Genet. *24*, 295–313. https://doi.org/10.1038/s41576-022-00553-x.
4. Hu, Y., Xie, M., Li, Y., Rao, M., Shen, W., Luo, C., Qin, H., Baek, J., and Zhou, X.M. (2024). Benchmarking clustering, alignment, and integration methods for spatial transcriptomics. Genome Biol. *25*, 212.
5. Yuan, Z., Zhao, F., Lin, S., Zhao, Y., Yao, J., Cui, Y., Zhang, X.-Y., and Zhao, Y. (2024). Benchmarking spatial clustering methods with spatially resolved transcriptomics data. Nat. Methods *21*, 712–722. https://doi.org/10.1038/s41592-024-02215-8.
6. Long, Y., Ang, K.S., Li, M., Chong, K.L.K., Sethi, R., Zhong, C., Xu, H., Ong, Z., Sachaphibulkij, K., Chen, A., et al. (2023). Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. Nat. Commun. *14*, 1155. https://doi.org/10.1038/s41467-023-36796-3.
7. Cang, Z., Ning, X., Nie, A., Xu, M., and Zhang, J. (2021). Scan-it: Domain segmentation of spatial transcriptomics images by graph neural network. BMVC *32*, 406.
8. Ren, H., Walker, B.L., Cang, Z., and Nie, Q. (2022). Identifying multicellular spatiotemporal organization of cells with spaceflow. Nat. Commun. *13*, 4076.
9. Dong, K., and Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nat. Commun. *13*, 1739.
10. Li, Z., and Zhou, X. (2022). Bass: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. Genome Biol. *23*, 168. https://doi.org/10.1186/s13059-022-02734-7.
11. Zhao, E., Stone, M.R., Ren, X., Guenthoer, J., Smythe, K.S., Pulliam, T., Williams, S.R., Uytingco, C.R., Taylor, S.E.B., Nghiem, P., et al. (2021). Spatial transcriptomics at subspot resolution with bayesspace. Nat. Biotechnol. *39*, 1375–1384. https://doi.org/10.1038/s41587-021-00935-2.
12. Singhal, V., Chou, N., Lee, J., Yue, Y., Liu, J., Chock, W.K., Lin, L., Chang, Y.-C., Teo, E.M.L., Aow, J., et al. (2024). Banksy unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. Nat. Genet. *56*, 431–441. https://doi.org/10.1038/s41588-024-01664-3.
13. Kim, M., Chung, Y.R., Kim, H.J., Woo, J.W., Ahn, S., and Park, S.Y. (2020). Immune microenvironment in ductal carcinoma in situ: a comparison with invasive carcinoma of the breast. Breast Cancer Res. *22*, 32. https://doi.org/10.1186/s13058-020-01267-w.
14. Walker, B.L., and Nie, Q. (2023). Nest: nested hierarchical structure identification in spatial transcriptomic data. Nat. Commun. *14*, 6554. https://doi.org/10.1038/s41467-023-42343-x.
15. Wasserman, L. (2018). Topological data analysis. Annu. Rev. Stat. Appl. *5*, 501–532.
16. Edelsbrunner, L., and Zomorodian; and Zomorodian. (2002). Topological persistence and simplification. Discrete Comput. Geom. *28*, 511–533.
17. Zomorodian, A., and Carlsson, G. (2004). Computing persistent homology. In Proceedings of the twentieth annual symposium on Computational geometry, pp. 347–356.
18. Schindler, D.J., and Barahona, M. (2023). Persistent Homology of the Multiscale Clustering Filtration. arXiv. https://doi.org/10.48550/arXiv.2305.04281.
19. Zheng, F., Zhang, S., Churas, C., Pratt, D., Bahar, I., and Ideker, T. (2021). Hidef: identifying persistent structures in multiscale 'omics data. Genome Biol. *22*, 21. https://doi.org/10.1186/s13059-020-02228-4.
20. Benjamin, K., Bhandari, A., Kepple, J.D., Qi, R., Shang, Z., Xing, Y., An, Y., Zhang, N., Hou, Y., Crockford, T.L., et al. (2024). Multiscale topology classifies cells in subcellular spatial transcriptomics. Nature *630*, 943–949. https://doi.org/10.1038/s41586-024-07563-1.

21. Cottrell, S., and Wei, G.-W. (2025). Multiscale cell-cell interactive spatial transcriptomics analysis. Adv. Sci. *12*, e08358.
22. Mouse Brain Section (Coronal), Spatial Gene Expression Dataset Analyzed Using Space Ranger 1.1.0, 10x Genomics, (2020, June 23). https://www.10xgenomics.com/datasets/mouse-brain-section-coronal-1-standard-1-1-0
23. Allen Mouse. (2024). Brain Atlas [dataset] (Allen Institute for Brain Science). https://mouse.brain-map.org/search/index.
24. (2011). Allen Reference Atlas – Mouse Brain [brain atlas] (Allen Institute for Brain Science). https://mouse.brain-map.org/experiment/thumbnails/100048576?image_type=atlas.
25. Anderson, N.M., and Simon, M.C. (2020). The tumor microenvironment. Curr. Biol. *30*, R921–R925. https://doi.org/10.1016/j.cub.2020.06.081.
26. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.
27. Gene Ontology Consortium; Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., et al. (2023). The gene ontology knowledgebase in 2023. Genetics *224*, iyad031. https://doi.org/10.1093/genetics/iyad031.
28. Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.-P., and Mi, H. (2022). Panther: Making genome-scale phylogenetics accessible to all. Protein Sci. *31*, 8–22. https://doi.org/10.1002/pro.4218.
29. Zunarelli, E., Nicoll, J.A., Migaldi, M., and Trentini, G.P. (2000). Apolipoprotein E polymorphism and breast carcinoma: correlation with cell proliferation indices and clinical outcome. Breast Cancer Res. Treat. *63*, 193–198.
30. Cui, J., Chai, S., Liu, R., and Shen, G. (2024). Targeting PGK1: A new frontier in breast cancer therapy under hypoxic conditions. Curr. Issues Mol. Biol. *46*, 12214–12229.
31. Arun, G., and Spector, D.L. (2019). MALAT1 long non-coding RNA and breast cancer. RNA Biol. *16*, 860–863.
32. Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippani, M., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat. Neurosci. *24*, 425–436. https://doi.org/10.1038/s41593-020-00787-0.
33. Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed rna profiling in single cells. Science *348*, aaa6090. https://doi.org/10.1126/science.aaa6090.
34. Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., and Zhuang, X. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science *362*, eaau5324. https://doi.org/10.1126/science.aau5324.
35. Codeluppi, S., Borm, L.E., Zeisel, A., La Manno, G., van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmfish. Nat. Methods *15*, 932–935. https://doi.org/10.1038/s41592-018-0175-z.
36. Yousefi, B., Schaub, D.P., Khatri, R., Kaiser, N., Kuehl, M., Ly, C., Puelles, V.G., Huber, T.B., Prinz, I., Krebs, C.F., et al. (2026). Scale: unsupervised multiscale domain identification in spatial omics data. Nucleic Acids Res. *54*, gkaf1456.
37. Li, X., Ren, X., and Venugopal, R. (2025). Entropy measures for quantifying complexity in digital pathology and spatial omics. iScience *28*, 112765. https://doi.org/10.1016/j.isci.2025.112765.
38. Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., and Alibakhshi, S. (2019). Elsa: Entropy-based local indicator of spatial association. Spatial Stat. *29*, 66–88. https://doi.org/10.1016/j.spasta.2018.10.001.
39. Space Ranger 1.1.0 (2020). Mouse brain section (coronal). https://support.10xgenomics.com/spatial-gene-expression/datasets/1.1.0/V1_Adult_Mouse_Brain?10xGenomics.
40. Traag, V.A., Waltman, L., and Van Eck, N.J. (2019). From louvain to leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.
41. Flamary, R., Courty, N., Gramfort, A., Alaya, M.Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., et al. (2021). Pot: Python optimal transport. J. Mach. Learn. Res. *22*, 1–8.
42. Maria, C., Boissonnat, J.-D., Glisse, M., and Yvinec, M. (2014). The gudhi library: Simplicial complexes and persistent homology. In Mathematical Software – ICMS 2014, H. Hong and C. Yap, eds. (Springer Berlin Heidelberg)978-3-662-44199-2, pp. 167–174.
43. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.
44. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Visium mouse brain section (coronal)	10x Genomics	https://squidpy.readthedocs.io/en/stable/api/squidpy.datasets.visium_hne_adata.html#squidpy.datasets.visium_hne_adata
Visium Invasive Ductal Carcinoma Stained With Fluorescent CD3 Antibody	Zhao et al.[11]	https://www.10xgenomics.com/datasets/invasive-ductal-carcinoma-stained-with-fluorescent-cd-3-antibody-1-standard-1-2-0
Visium Human Dorsalateral Prefrontal Cortex	Maynard et al.[32]	https://figshare.com/articles/dataset/Visium_DLPFC_preprocessed
osmFISH Mouse Somatosensory Cortex	Codeluppi et al.[35]	https://figshare.com/articles/dataset/osmFISH_datasets
MERFISH Mouse Hypothalamic Preoptic Region	Moffitt et al.[34]	https://squidpy.readthedocs.io/en/stable/api/squidpy.datasets.merfish.html#squidpy.datasets.merfish
Software and algorithms		
Python 3.13.5	Python team	https://www.python.org
leidenalg v0.10.2	Traag et al.[40]	https://github.com/vtraag/leidenalg
SCAN-IT v0.1	Cang et al.[7]	https://github.com/zcang/SCAN-IT
GraphST v1.1.1	Long et al.[6]	https://github.com/JinmiaoChenLab/GraphST
Banksy v1.3.4	Singhal et al.[12]	https://github.com/prabhakarlab/Banksy_py
NeST v1.0.4	Walker et al.[14]	https://github.com/bwalker1/NeST
Python Optimal Transport v0.9.5	Flamary et al.[41]	https://github.com/PythonOT/POT
GUDHI v3.11.0	Maria et al.[42]	https://github.com/GUDHI
scanpy v1.11.2	Wolf et al.[43]	https://github.com/scverse/scanpy
scipy v1.16.0	Virtanen et al.[44]	https://github.com/scipy/scipy
PHD-MS	This paper	https://github.com/pzbeamer/phd-ms, doi: https://doi.org/10.5281/zenodo.18733158

METHOD DETAILS

Cluster filtration

The input of PHD-MS is a collection of spatial domain clustering results of varying resolution from any clustering method of spatial transcriptomics data. Here we demonstrate the method using the clusters generated by GraphST, SCAN-IT, and Banksy. More generally, PHD-MS operates only on the resulting domain partitions and therefore can be applied to clusterings produced by other spatial domain methods such as SpaGCN[2] and BayesSpace,[11] provided clusterings are available across multiple resolutions. To generate the input clusterings for the examples in this paper, we construct spatially-aware embeddings via GraphST, SCAN-IT, or Banksy. We then perform spatial domain clustering with these embeddings at a series of scale parameters using the Leiden algorithm,[40] computed using the leidenalg package in Python. For the presented examples, a sequence of 8 clustering results with increasing spatial scale (larger and fewer clusters) are generated using a uniform grid of decreasing resolution parameters between $r = 0.95$ and $r = 0.15$ in the Leiden algorithm. The scale parameters can be easily modified by users to suit different applications.

A graph of cluster results is constructed, where each node represents a cluster at a scale $k \in \{k_i\}_{i=1}^{K}$. We index the j-th cluster at scale k_i by $C_{i,j}$. Edges are drawn between all nodes from sequential scale parameters. In other words, for $i \in \{1, \ldots, K\text{-}1\}$, we include an edge between each pair of clusters $[C_{i,\cdot}; C_{i+1,\cdot}]$ at scale k_i and k_{i+1}.

To define a filtration on the vertex set of multi-resolution clusters, we construct a real-valued function f on pairs of clusters (or edges). This function defines the order in which edges are included in persistent homology analysis. (Full definitions of persistent

homology concepts are provided in the Method S1). Two related filtration functions are considered, derived from the *Jaccard index* and *Containment index*,[19] both of which quantify similarity between clusters. The containment index filtration is intended for a sequence of clusterings where clusters are largely nested or contained within others, while the Jaccard index filtration applies more generally to any set of clusterings. We use the containment index filtration for all tissues in the results, except for the IDC in Figure 3. We use the Jaccard index-based filtration in this case, because tumors are not necessarily organized into specialized subregions nested within larger anatomical features. The containment index is defined by

$$f_C\left(\left[C_{i,j}, C_{i+1,l}\right]\right) = 1 - \frac{\left|C_{i,j} \cap C_{i+1,l}\right|}{\left|C_{i,j}\right|}. \quad \text{(Equation 1)}$$

Intuitively, f_C represents the inverse proportion of nodes in $C_{i,j}$ contained in $C_{i+1,l}$. The Jaccard index, meanwhile, is defined by:

$$f_J\left(\left[C_{i,j}, C_{i+1,l}\right]\right) = 1 - \frac{\left|C_{i,j} \cap C_{i+1,l}\right|}{\left|C_{i,j} \cup C_{i+1,l}\right|}. \quad \text{(Equation 2)}$$

Here, f_J is the inverse fraction of the overlap between clusters over their union. In general, when f_C or f_J is ≈ 0, the two clusters are highly similar, while when these functions are ≈ 1, two clusters are nearly entirely disjoint. To define a filtration, we choose the Jaccard or containment index, and extend the function to individual vertices by simply stipulating that $f_*([C_{i,j}]) = 0$ for all i,j.

Intuitively, stable tissue domains across resolutions are recorded as connected components with f-values close to 0. In this way, the filtration captures information about tissue domains which highly overlap as the resolution parameter varies, representing the most important underlying domains in the data.

Persistent homology results and analysis

For detailed definitions of concepts in persistent homology, consult the supplementary materials. We compute persistent homology using the custom-built cluster filtration via the Python package GUDHI.[42] We apply a union-find algorithm to recover the cells/spots which participate in each persistent connected component of the cluster filtration. These persistent connected components represent multiscale tissue domains. A coreness score is assigned to each spot in the tissue, one minus the lowest filtration value at which the spot is included in the domain. Recall that the filtration value of a cluster is determined from the Jaccard or Containment index, and a cluster is added to the connected component at this filtration value. In formal language, let D denote the multiscale domain, a connected component of the filtration consisting of some set of clusters $\left\{C_{i,j}\right\}_{(i,j)\in I_D}$. Additionally, let X denote the set of all tissue spots. For some tissue spot $x \in X$, we say x belongs to D if there exists a distinct pair of clusters $C_{i,j}, C_{i+1,l}$ such that $x \in C_{i,j} \cup C_{i+1,l}$. For a spot x belonging to the domain D, its coreness score $c_D: X \to [0,1]$ is given by

$$c_D(x) = 1 - \min_{C_{i,j}, C_{i+1,l} \in D} \left\{ f_*\left(\left[C_{i,j}, C_{i+1,l}\right]\right) : x \in C_{i,j} \cup C_{i+1,l} \right\} \quad \text{(Equation 3)}$$

For a spot x not belonging to D, we set $c_D(x) = 0$. Intuitively, a low weight means x is an unstable frontier member of the domain, and a high weight means x is a stable core member of the domain. The multiscale domain thus consists of the pair (D, c_D), the set of clusters belonging to the domain and the corresponding coreness function, where each individual domain D is associated with a unique coreness function c_D. In practice, we typically normalize c_D between 0 and 1 for each domain, so that the domain's stable core always has coreness score 1:

$$\overline{c}_D(x) = \frac{c_D(x) - \min_y\{c_D(y)\}}{\max_y\{c_D(y)\} - \min_y\{c_D(y)\}} \quad \text{(Equation 4)}$$

Recall that each domain is generated from a persistent component of the cluster filtration, so that every domain has a persistence lifetime. For a transcriptomic spot x, the normalized coreness score $\overline{c}_D(x)$ then represents the proportion of D's lifetime where D contains x.

We visualize each persistent domain D by plotting normalized $\overline{c}_D$ as a scalar field on the spatial coordinates of the tissue. Spots with large $\overline{c}_D$ values are part of the domain's stable core, and spots with small $\overline{c}_D$ values are part of the domain's unstable frontier. An interactive point-and-click tool is provided for exploring the multiscale tissue morphology centered on a particular region. When the point-and-click option is selected, the user can click a point in a map of the tissue to manually visualize all the multiscale domains centered on this point.

Heterogeneity score

We define the heterogeneity score, which estimates the number of unique domains that contain each transcriptomic spot. Let $D_1, \ldots, D_M$ be the set of all PHD-MS domains. Recall that each D_i corresponds to a persistent component in the cluster filtration. Then let p_i denote the persistence lifetime (death minus birth) of D_i, (where we stipulate that if a domain lives forever, $p_i = 1$). The heterogeneity score h of transcriptomic spot x is the weighted sum

$$h(x) = \sum_{i=1}^{M} p_i \overline{c}_{D_i}(x) \quad \text{(Equation 5)}$$

CellPress
OPEN ACCESS

Intuitively, the heterogeneity score is high when x is a core member of many highly persistent components, and is minimized when x is a member of only one persistent component across scales. In this way, the heterogeneity score tracks whether x is consistently assigned to the same underlying domain across scales, or if x belongs to many different domains.

Wasserstein distance

The Wasserstein distance is used to compare the spatial relevance and similarity between spatial domains by representing both traditional binary domains and the multiscale domains as spatial distributions. To compute the Wasserstein distance between two domains, we first convert each domain to a probability over the spatial coordinates of X. For a multiscale domain D, we normalize $\overline{c}_D$ to probability distributions:

$$\widehat{c}_D(y) = \frac{\overline{c}_D(y)}{\sum_{x \in X} \overline{c}_D(x)} \quad \text{(Equation 6)}$$

In the output of existing spatial clustering algorithms and the ground truth annotations, spatial domains are often represented as traditional binary clusters, and cell-spots are not assigned a coreness score. Instead, cell-spots are either members of the binary cluster, or not. All ground-truth annotations are traditional binary clusters in this sense. In these cases, we simply consider a uniform distribution on the spots in a binary cluster D

$$\widehat{c}_D(y) = \begin{cases} \frac{1}{|D|} & y \in D \\ 0 & \text{elsewhere.} \end{cases} \quad \text{(Equation 7)}$$

Then for domains D_1 and D_2, the 2-Wasserstein metric

$$W_2\left(\widehat{c}_{D_1}, \widehat{c}_{D_2}\right) = \min_{P \in \Gamma\left(\widehat{c}_{D_1}, \widehat{c}_{D_2}\right)} \left(\sum_{ij} \| x_i - x_j \|^2 P_{ij} \right)^{\frac{1}{2}} \quad \text{(Equation 8)}$$

depicts a spatial-aware distance between distributions $\widehat{c}_{D_1}$ and $\widehat{c}_{D_2}$ where N is the number of transcriptomic spots in X and $\Gamma(\widehat{c}_{D_1}, \widehat{c}_{D_2}) = \left\{P \in \mathbb{R}_+^{N \times N} : P1_N = \widehat{c}_{D_1}, P^\top 1_N = \widehat{c}_{D_2}\right\}$, . We compute the Wasserstein distance using the POT library.[41]

Normalized mutual information

Normalized mutual information (NMI) measures the shared information between two clusterings, and the NMI of a clustering and a set of ground truth labels measures the accuracy of that clustering. The standard NMI applies to binary clusters, so we have generalized it to the multiscale context. For binary clusters, our generalized version is equivalent to the standard NMI. A full derivation is provided in supplements (Method S2).

Suppose we have $x_1, \ldots, x_N$ transcriptomic spots. Given a subset of multiscale domains, $D_1, \ldots, D_M$. We construct a matrix $\mathbf{D} \in [0,1]^{N \times M}$ of coreness scores, whose ij-th entry is the coreness score of the i-th transcriptomic spot in the j-th domain. We also normalize $\mathbf{D}$ so that its rows sum to 1:

$$D_{ij} = \frac{c_{D_j}(x_i)}{\sum_{k=1}^{m} c_{D_k}(x_i)} \quad \text{(Equation 9)}$$

If $D_1, \ldots, D_M$ are binary clusters, each spot x_i belongs to one and only one D_j. Therefore, rows of $\mathbf{D}$ will be identity vectors specifying which unique domain contains each x_i. Column $\mathbf{D}_j$ of $\mathbf{D}$ represents the coreness scores at each transcriptomic spot $\{x_i\}_{i=1}^N$ for domain D_j.

The mutual information between two sets of multiscale domains $\mathbf{D} = [\mathbf{D}_1 \ldots \mathbf{D}_{M_1}]$ and $\mathbf{E} = [\mathbf{E}_1 \ldots \mathbf{E}_{M_2}]$ is given by

$$\mathrm{MI}(\mathbf{D}, \mathbf{E}) = \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \frac{\mathbf{D}_i^\top \mathbf{E}_j}{N} \ln\left[\frac{N \mathbf{D}_i^\top \mathbf{E}_j}{\|\mathbf{D}_i\|_1 \, \|\mathbf{E}_j\|_1} \right] \quad \text{(Equation 10)}$$

where $\mathbf{D}_i$ denotes the i-th column of $\mathbf{D}$, and $\|*\|_1$ the standard ℓ_1-norm.

Then the NMI is the mutual information normalized by the arithmetic mean of the Shannon's entropy H of the columns of $\mathbf{D}$,$\mathbf{E}$, where

$$H(\mathbf{D}) = -\sum_{i=1}^{M_1} \frac{\|\mathbf{D}_i\|_1}{N} \ln\left(\frac{\|\mathbf{D}_i\|_1}{N} \right) \quad \text{(Equation 11)}$$

Then

$$\mathrm{NMI}(\mathbf{D}, \mathbf{E}) = \frac{\mathrm{MI}(\mathbf{D}, \mathbf{E})}{1/2[H(\mathbf{D}) + H(\mathbf{E})]} \quad \text{(Equation 12)}$$

QUANTIFICATION AND STATISTICAL ANALYSIS

Differential gene expression

We conduct a differential gene expression study on the Visium IDC dataset, using the 9 PHD-MS multiscale domains in Figure 3 as categories. We assign each spot to a unique region according to its maximum coreness score across all 9 domains. For each multiscale domain, we collect all genes differentially expressed with log-fold change >1.5 and $p < 0.001$ according to a standard t test (the rank-genes-group test implemented in the Scanpy[43] package). In each individual domain, the expression of each gene is tested against the expression of this gene across all other domains. These highly-expressed genes are fed to a Gene Ontology Enrichment[26] analysis, which identifies biological processes associated with these highly-expressed genes in each domain. Each significant biological process is statistically overrepresented in the set of highly-expressed genes according to Fisher's Exact test with $p < 0.05$.

We also conduct differential gene expression studies restricted to the set of regions annotated '7', '8', and '9'. These regions border one another, and differential gene expression identifies how much oncogenic expression differs between these continuous regions. We collect all genes differentially expressed with log-fold change >0.75 and $p < 0.001$ according to a t test. Overexpressed genes are cross-referenced with the OncoDB database to identify IDC-associated genes.

Quantitative benchmarking

In quantitative evaluations, we compute the distance $W_2(\widehat{c}_{D_1}, \widehat{c}_{D_2})$ where D_2 is a binary ground-truth domain. Then, W_2 measures how well a multiscale domain matches a ground-truth domain. In benchmarking, we also compute the Wasserstein distance between traditional single-scale clusters and ground-truth domains. For this purpose, we fix the resolution parameter to a single scale. In particular, we select the scale that best matches the ground-truth number of clusters. Statistically significant improvement is demonstrated by one-sided Wilcoxon signed-rank test computed using Scipy[44] in Python, with significance thresholds $p < 0.05$ and $p < 0.01$.

We compute NMI(**D**,**G**) where **G** is the matrix of binary ground-truth domains. Then, the NMI measures how well a set of multiscale domains matches the ground-truth domains. For benchmarking, we construct **D** from a subset of all possible PHD-MS domains, selecting a subset that maximizes the NMI. In benchmarking, we also compute the NMI between single-scale clusters and ground-truth domains. For this purpose, we fix the resolution parameter to a single scale. In particular, we select the scale that best matches the ground-truth number of clusters. Statistical significant improvement is demonstrated by one-sided Wilcoxon signed-rank test computed using Scipy in Python, with significance thresholds $p < 0.05$ and $p < 0.01$.