nature genetics

Article

https://doi.org/10.1038/s41588-025-02119-z

Quantitative characterization of tissue states using multiomics and ecological spatial analysis

Received: 24 June 2024

Accepted: 5 February 2025

Published online: 1 April 2025

Check for updates

Daisy Yi Ding[1,11], Zeyu Tang[2,11], Bokai Zhu[3,4,11], Hongyu Ren[5], Alex K. Shalek[3,4,6,7,8], Robert Tibshirani[1,9,12] & Garry P. Nolan[10,12]

The spatial organization of cells in tissues underlies biological function, and recent advances in spatial profiling technologies have enhanced our ability to analyze such arrangements to study biological processes and disease progression. We propose MESA (multiomics and ecological spatial analysis), a framework drawing inspiration from ecological concepts to delineate functional and spatial shifts across tissue states. MESA introduces metrics to systematically quantify spatial diversity and identify hot spots, linking spatial patterns to phenotypic outcomes, including disease progression. Furthermore, MESA integrates spatial and single-cell multiomics data to facilitate an in-depth, molecular understanding of cellular neighborhoods and their spatial interactions within tissue microenvironments. Applying MESA to diverse datasets demonstrates additional insights it brings over prior methods, including newly identified spatial structures and key cell populations linked to disease states. Available as a Python package, MESA offers a versatile framework for quantitative decoding of tissue architectures in spatial omics across health and disease.

The spatial organization of cells in tissues critically influences biological processes from embryonic development to disease progression. Because cells do not operate in isolation, their functionality and behaviors are intrinsically linked to their local microenvironments. Advances in spatial profiling technologies[1–6] have enhanced our ability to analyze cellular characteristics within native tissue context and derive mechanistic insights[7–10]. Computational algorithms have been developed for spatial omics to explore cellular interactions and their spatial dependencies, providing tools to analyze cellular neighborhoods and tissue niches[10–16].

However, it is still challenging to analyze spatial-omics data systematically and robustly. Methods are needed to quantitatively assess spatial distributions and structures across dynamic ranges (for example, 50 µm to 1 mm) and identify associations with functional outcomes. Current approaches focus on visualization in examining cellular diversity, with limited systematic metrics for characterizing spatial patterns related to phenotypic outcomes such as disease progression[10–16]. This limitation presents a gap in the analytical toolkit for spatial-omics data.

Inspired by ecology, we recognize an opportunity to enable an analogy between species biodiversity and cellular diversity. Ecologists have developed metrics to quantify species distributions across spatial scales and habitats[17,18], and recent studies have begun applying them to spatial-omics data[19–21]. Adapting ecology metrics for cellular

[1]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [2]Weill Cornell Graduate School of Medical Sciences, Weill Cornell Medicine, New York, NY, USA. [3]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Department of Computer Science, Stanford University, Stanford, CA, USA. [6]Institute for Medical Engineering and Science, MIT, Cambridge, MA, USA. [7]Department of Chemistry, MIT, Cambridge, MA, USA. [8]Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA, USA. [9]Department of Statistics, Stanford University, Stanford, CA, USA. [10]Department of Pathology, Stanford University, Stanford, CA, USA. [11]These authors contributed equally: Daisy Yi Ding, Zeyu Tang, Bokai Zhu. [12]These authors jointly supervised this work: Robert Tibshirani and Garry P. Nolan e-mail: gnolan@stanford.edu

environments allows us to quantify tissue spatial structures and identify their associations with functional outcomes.

By analyzing the biology across various omics views, we can uncover insights hidden in a single spatial-omics modality. A well-implemented multiomics approach can harness the wealth of available single-cell data, such as data from single-cell RNA sequencing (scRNA-seq) technologies, by integrating them with spatial omics to enrich the information captured. Recent data integration algorithms[22–31] enable high-throughput scRNA-seq measurements of thousands of genes to inform spatial-omics data, such of those from CODEX, which retains tissue spatial context. Fusing these approaches maps proteins and RNAs at the single-cell level within tissues. In silico-unified multiomics data, combined with multiscale perspectives on cellular diversity, provide a more holistic characterization of cellular landscape and would pave the way for an enhanced understanding of complex biological systems.

We present MESA (multiomics and ecological spatial analysis), a framework combining ecological principles and multiomics integration to decode tissue remodeling. Drawing inspiration from ecology, MESA adapts diversity metrics traditionally used to gauge biodiversity for spatial omics, creating tools for systematic quantification of cellular diversity. We introduce a multiscale diversity index (MDI), alongside global and local diversity indices, to capture not only tissue overarching diversity but also localized patterns and dependencies. Furthermore, MESA uses a multiomics approach to spatial-omics analyses. MESA in silico amalgamates cross-modality single-cell data to enrich the context of spatial-omics observations. With the additional layers of information brought to bear by multiomics, MESA facilitates an extended view of cellular neighborhoods and their spatial interactions within tissue microenvironments. MESA's approach, incorporating differential expression (DE), gene set enrichment and ligand-receptor interaction (LRI) analyses within these spatially defined cellular assemblies, further enhances a mechanistic understanding of tissue remodeling across disease states.

Applying MESA to datasets of CODEX human tonsil[32], mouse spleen[2], human intestine[33] and CosMx SMI human liver[34] revealed key cellular components, spatial structures and functionalities linked to tissue disease states, which were not discerned with prior techniques. Thus, this quantitative approach to spatial pattern analysis and multiomics integration advances our understanding of tissue remodeling and its functional implications in diseases and therapies. MESA (https://github.com/Feanor007/MESA) is available as a Python package.

Results

Overview of the MESA pipeline

MESA integrates spatial omics with corresponding single-cell datasets (for example, scRNA-seq) from the same tissue type and disease condition (Fig. 1a). It matches cells across modalities through MaxFuse[22], enriching spatial-omics data with single-cell information for spatial analyses (Fig. 1a(i)). The framework leverages available single-cell data resources, reducing the need to conduct all omics experiments directly. Based on in silico-created multiomics profiles, MESA characterizes the local neighborhood of each cell to identify conserved, distinct cellular neighborhoods (Fig. 1a(ii)). This process aggregates multiomics information from spatially determined neighbors (for example, 20 cells) to capture the cellular environment.

In prior publications, we determined cellular neighborhoods[10,33,35] using a sliding window that determines local cell compositions wherein cell types are defined using clustering or other approaches. In such approaches, cell types are binned as univariate entities (that is, $CD4^+$ $FOXP3^+$ regulatory T cells (T_{reg} cells) or $B220^+CD20^+$ B cells)[33,36], and the expression of any given protein on those cells can exist within a range. However, MESA does not use cell designations as such but goes beyond this by incorporating multiomics information, such as protein and mRNA expression levels (or any other omics data that pertain). Instead of using a cell type as a predesignated entity, MESA assesses average protein and mRNA levels of cells in a region to determine neighborhood compositions (Fig. 1a(ii)). This process allows MESA to take advantage of the dynamic range of proteins as a proxy for cell states that previously has been ignored by traditional clustering or gating algorithms and is therefore more sensitive to coregulated protein and mRNA levels across cells in a region.

Subsequently, MESA uses *k*-means clustering[37] to identify conserved neighborhood patterns, followed by DE analysis and gene set enrichment analysis (GSEA) to explore functional pathways and implications (Fig. 1a(iii)). These analyses are enabled by the multiomics approach and can uncover nuanced functional pathways within complex cellular environments.

The second component of MESA uses systematic metrics to quantify cellular diversity (Fig. 1b). First, we propose an MDI to evaluate diversity variations across spatial scales (Fig. 1b(i)). MDI functions by dividing tissue sections into patches of varying sizes, assessing diversity within each patch, and computing an average diversity score for the corresponding scale. MDI is derived as the slope of a linear regression line fitted to these diversity scores across scales, and therefore a measure of how cellular diversity fluctuates as scale changes. Lower MDI values indicate consistent diversity across scales, whereas higher values signal more pronounced shifts (Fig. 1b(i)). For instance, tissues with uniform cell-type distributions would maintain consistent diversity across scales. Conversely, tissues with disproportionately distributed cell types would exhibit more pronounced diversity fluctuations across scales. To represent this, MESA generates a diversity heatmap by calculating the diversity of local patches within a tissue sample (Fig. 1b(ii)). We use this to derive the global diversity index (GDI), which assesses whether patches of similar diversity are spatially adjacent, and the local diversity index (LDI), which distinguishes regions by their diversity patterns and identifies 'hot spots' (clusters of patches with high diversity) and 'cold spots' (clusters of patches with low diversity).

This approach parallels biodiversity hot spots and cold spots in geo-ecology. MESA assesses the hot spots and cold spots by analyzing the abundance and cohabitation of cell types in these regions, uncovering signals not evident in whole-tissue analysis. Of note, these spatial compartments differ from tissue neighborhoods/niches[14,15,38,39], which are conventionally defined by cellular identity instead of cellular diversity. Furthermore, the diversity proximity index (DPI) evaluates spatial relationships among hot/cold spots (Fig. 1b(iii)). This concept parallels ecological studies where habitat size and proximity affect species dynamics. As in ecology, where larger and proximate habitats enhance species interactions, higher DPI values suggest spots that are closer and larger and could indicate more dynamic cellular interactions. Conversely, lower DPI values reflect smaller, dispersed spots, which could suggest fewer cell–cell interactions. MESA's quantitative metrics enable the exploration of associations between cellular diversity and phenotypic outcomes like disease progression. We provide more details on the MESA framework in the Methods section.

Multiomics analysis enhances tonsil niche characterization

To demonstrate how multiomics enhances niche characterization, we applied MESA to a tonsil CODEX dataset[32] integrated with tonsil scRNA-seq data[40]. MESA facilitated the delineation of distinct cellular interactions and neighborhoods beyond those observable by traditional analysis techniques and revealed DE patterns of protein and mRNA between neighborhoods. Figure 2a shows the tonsil tissue sample with cells color-coded by cell type. Figure 2b presents neighborhood characterization based on cellular composition (commonly used in prior techniques[33,36]; left), average protein expression (middle) and mRNA expression (right) using in silico multiomics fusion.

We observed enhanced spatial delineation of neighborhoods (labeled as neighborhoods 0, 1 and 3 in Fig. 2b middle and right panels) that consistently emerge within germinal centers, which was initially

https://doi.org/10.1038/s41588-025-02119-z

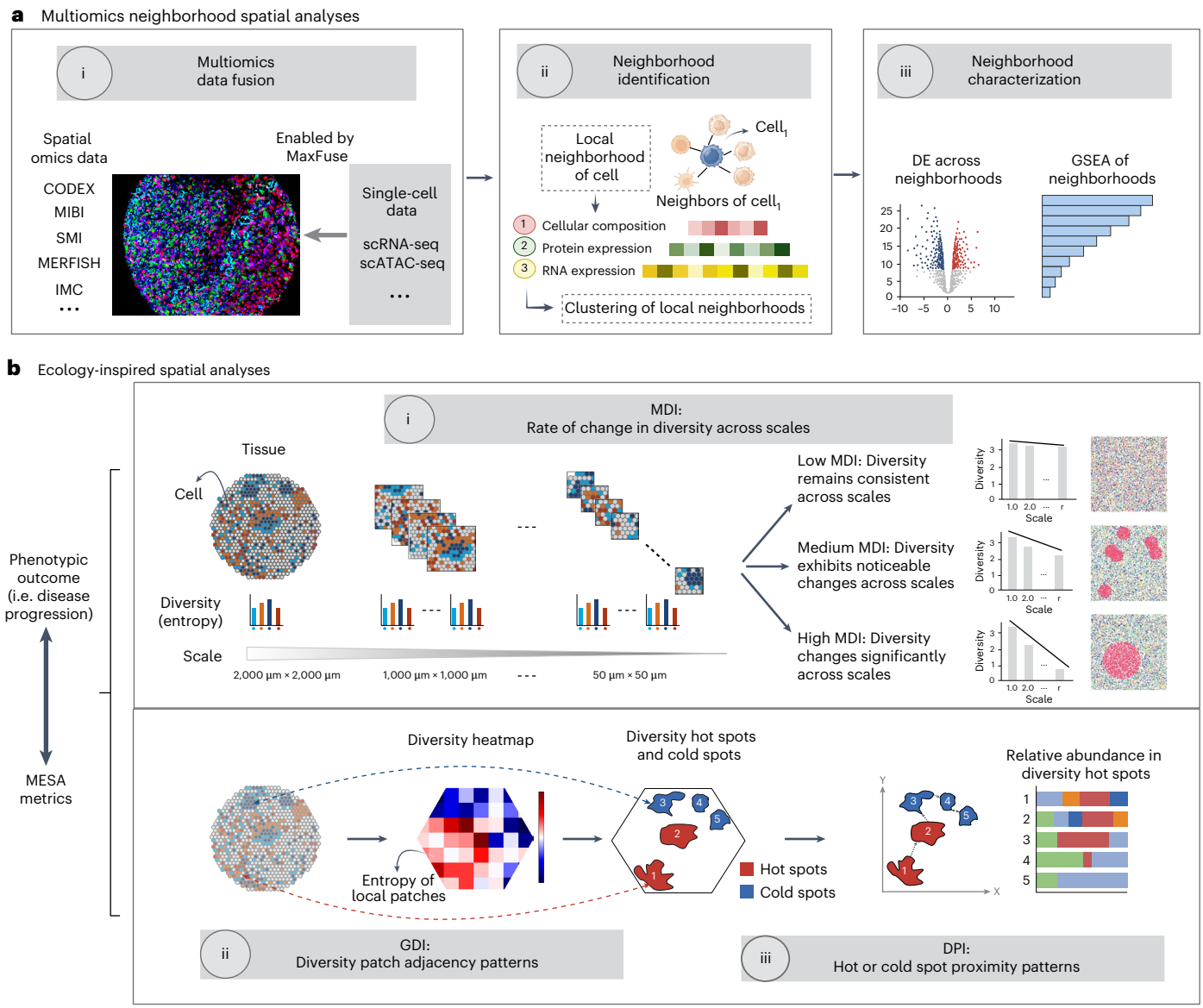


Fig. 1 | MESA overview. a, Multiomics neighborhood spatial analysis integrates spatial-omics data (for example, CODEX, MIBI, IMC and MERFISH) with single-cell datasets (for example, scRNA-seq and scATAC-seq) using algorithms like MaxFuse[22]. The framework is designed to be versatile, allowing the utilization of various other data integration methods to accommodate different analytical needs. This integration enriches spatial-omics data and creates an in silico multiomics spatial profiling for downstream analysis. Neighborhood identification follows, with multiomics information aggregated from each cell's *k*-nearest neighbor (*k*-NN) into Neighborhood Feature Vector (NFV), capturing the local cellular environment. The *k*-NN is determined based on spatial distance. Different types of NFVs are computed, including cellular composition, local average protein expression and local average RNA expression (by in silico matching of scRNA-seq data). These NFVs serve as the basis for clustering to identify distinct, conserved cellular neighborhoods. MESA conducts DE analysis and GSEA to gain functional insights into the identified cellular neighborhoods. **b**, Ecology-inspired spatial analyses use the MDI to quantify variations in diversity across spatial scales. It works by dividing a tissue sample into patches of varying sizes, evaluating diversity within each patch, and subsequently calculating a mean diversity score corresponding to each scale. The MDI measures the rate of change in diversity across scales: low MDI values indicate consistent cellular diversity across scales, and higher values signal more pronounced diversity shifts, which may imply a disproportionate distribution of certain cell types within the tissue. To represent this, MESA generates a diversity heatmap by computing the entropy of local patches. Based on the diversity heatmap, the GDI evaluates spatial adjacency of patches sharing similar diversity levels, whereas the LDI identifies diversity hot spots (regions characterized by high diversity) and cold spots (regions characterized by low diversity). MESA analyzes hot spots and cold spots by examining cell-type prevalence and cohabitation within these regions, with the potential to reveal patterns that might not be evident when considering the entire tissue. The DPI measures the spatial proximity and size relationships between hot spots and cold spots, with higher values indicating larger and more proximate diversity spots. Created in BioRender.

identified as neighborhood 1 (left panel). Figure 2c zooms into one such germinal center (circled by dotted lines) to illustrate these distinctions. Composed mainly of B cells (i), the germinal center corresponds to neighborhood 1 (orange; ii). MESA subdivides it into distinct neighborhoods (orange, red and blue; iii and iv), a pattern seen consistently across multiple germinal centers, revealing distinct subniches undetected by previous algorithms.

The granularity of niche characterization was assessed using Shannon entropy. Higher entropy values indicate finer delineation: protein and mRNA-based methods (3.1 and 3.0) outperformed cellular composition (2.7), supporting our visual observations and demonstrating MESA's capability to identify more nuanced neighborhoods. Although minor inconsistencies between protein and mRNA-based results occur at compartment boundaries or transitional regions (Fig. 2b,c and

Supplementary Fig. 1), overall compartment structures remain consistent. To demonstrate MESA's robustness, we integrated tonsil CODEX data with another tonsil scRNA-seq dataset[41], yielding highly consistent results that preserved key spatial structures (Extended Data Fig. 1).

To assess the generalizability of our proposed multiomics framework, we evaluated three additional clustering methods—BIRCH[42], hierarchical clustering[43] and FlowSOM[44]—alongside *k*-means. Results consistently showed improved neighborhood delineation using multiomics information (Supplementary Fig. 2). Comparison with marker intensity-based neighborhood characterization methods, including BANKSY[45], CellCharter[16] and UTAG[15], provided finer characterization than cellular composition-based approaches, but MESA more effectively delineated nuanced neighborhoods (Extended Data Fig. 2). Finally, while applying more granular cell-type annotations enhanced composition-based methods, MESA's integration of continuous multiomics expression and spatial context provided deeper insights (Supplementary Fig. 3).

The cellular composition, protein and mRNA expression across neighborhoods were visualized in Fig. 2d. All 46 CODEX protein markers and top 50 variable mRNA genes revealed distinct expression patterns despite similar cellular compositions across neighborhoods. For example, neighborhoods 1, 0 and 3 (red box) showed similar cellular makeups but different expression levels of proteins Ki67, CD21 and CD22 and mRNAs *CD74* and *TUBA1B*, consistent with known markers of distinct B cell subpopulations in germinal centers[46–48].

Furthermore, DE and GSEA revealed functional distinctions between the finer-grained neighborhoods (Fig. 2e and Supplementary Fig. 4). Neighborhood 0 showed significant enrichment for pathways including E2F targets, G2M checkpoint and DNA repair, whereas neighborhood 3 for TNFA signaling via nuclear factor κB (NF-κB), IL-2 STAT5 signaling and interferon-gamma response. These findings align with germinal center dark and light zones' distinct functions in B cell proliferation versus selection and antigen presentation[49]. MESA's analysis reflects this segregation, revealing molecular profiles consistent with literature expectations. This enhanced characterization demonstrated MESA's capability to uncover vital cellular dynamics within complex tissue structures. These types of insights will be useful for elucidating the mechanisms of immune surveillance and response within lymphoid tissues.

Eco-spatial analysis reveals autoimmune tissue remodeling

We applied MESA's eco-spatial framework to a CODEX spleen dataset of healthy and MRL/lpr (a murine model of systemic autoimmune disease) mice[2]. Tissues were segmented into 30 µm × 30-µm patches. Figure 3a shows the cell-type map (top), diversity heatmap (middle; red: higher, blue: lower) and diversity hot/cold spots (bottom).

High-diversity regions align with the marginal zone and red pulp, whereas low-diversity regions correspond to B follicles and periarteriolar lymphoid sheath (PALS). MESA revealed notable shifts in cellular diversity patterns during disease progression. MRL/lpr tissues show higher MDI versus healthy tissues (Fig. 3b, left). This difference reflects the spatial arrangement of homogeneous regions in healthy spleen tissues, which maintain consistent cell-type diversity across spatial scales. These regions, predominantly B follicles and PALS, are crucial for spleen immune functions. In MRL/lpr tissues, disease progression disrupts this organization through marginal zone dissipation, PALS disintegration and erythroblast invasion of the red pulp[2].

Via MESA, these latter changes are seen to disrupt previously organized tissue structures, producing mosaic-like patterns and diversity fluctuations across scales. Healthy spleen tissues have higher GDI (Fig. 3b, middle), indicating distinct segregation of high and low diversity regions and more organized compartments. In contrast, diseased tissues show lower GDI and more blended patterns, indicating disrupted spleen functionality. Furthermore, using LDI, regions identified as hot/cold spots showed differential spatial proximity in healthy versus diseased samples. Healthy tissues show higher DPI with more expansive and proximate hot spots (Fig. 3b, right). In healthy tissues, hot spots align with red pulp (containing erythrocytes, F4/80$^+$ macrophages, stromal cells), whereas cold spots associate with B follicles (B cells) and PALS (T cells). In diseased tissues, hot spots shift to restructured PALS and invasion fronts, with cold spots in remaining B follicles.

To demonstrate MESA's unique capability in discovering these important spatial compartments beyond prior neighborhood/niche methods, we compared it with Spatial-LDA[14], UTAG[15] and CellCharter[16]. MESA's diversity hot/cold spots did not align with regions identified by these conventional neighborhood/niche methods (Extended Data Fig. 3a), underscoring its distinct advantage.

Among limited quantitative metrics for spatial omics, CellCharter introduces shape metrics gaining traction in spatial analysis, including curl, elongation, linearity and purity[16]. MESA's metrics outperformed CellCharter's in distinguishing MRL/lpr from healthy tissues (Extended Data Fig. 3b). MESA also showed superior performance compared to established spatial ecology metrics like fractal dimension index[50] and shape index[51] (Supplementary Fig. 5). In addition, we incorporated global and local Getis-Ord Gi/Gi* statistics[52,53] into our benchmarking, with MESA demonstrating better performance (Supplementary Fig. 6), and as additional options in the MESA package.

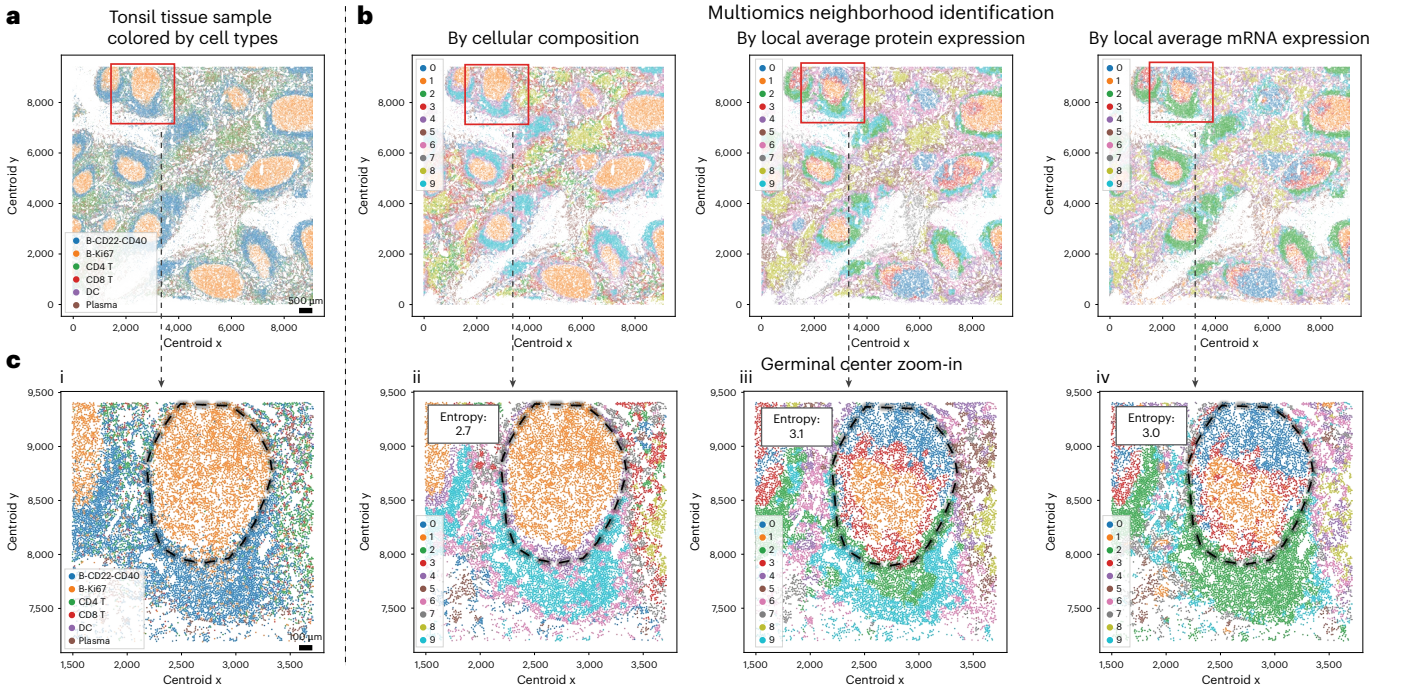
We analyzed the cellular composition and cohabitation in hot spots between healthy and diseased spleens to identify characteristic cell types or combinations associated with disease. B cells significantly declined in frequency in both whole tissue and hot spots of diseased spleens (Fig. 3c, left). However, hot spots uniquely showed increased B220$^+$, DN T cells, and CD106$^+$CD16/32$^+$CD31$^+$ stromal populations (Fig. 3c, middle/right and Supplementary Figs. 7–9), highlighting the utility of hot-spot-focused analysis in detecting nuanced changes.

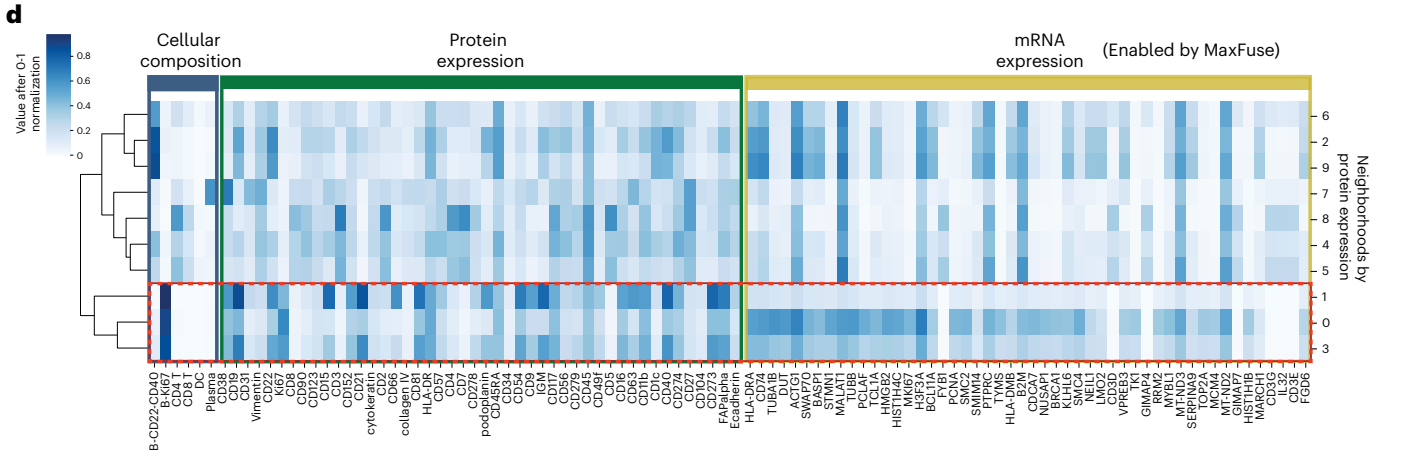
Fig. 2 | Multiomics spatial analysis enhances niche characterization in tonsil tissues. a, Tonsil tissue sample with cells color-coded by cell type. **b**, Niche characterization results based on cellular composition (left), local average protein expression (middle), and local average RNA expression (right, enabled by the fusion algorithm MaxFuse). We observed enhanced spatial delineation of neighborhoods, labeled as neighborhoods 0 (blue), 1 (orange) and 3 (red) in the middle and right panels, that consistently emerge within the germinal centers, which are identified only as neighborhood 1 (orange) in the left panel. **c**, A zoom-in visualization at one of the germinal centers (circled by dotted lines). The entropy values (calculated as the Shannon diversity index: $H = -\sum p(x) \log p(x)$, where $p(x)$ is the proportion of cells in each neighborhood) quantify the granularity of neighborhood characterization, with higher values indicating increased delineation. Primarily composed of B cells, as shown in subpanel i, this germinal center corresponds to the neighborhood labeled as 1 and colored orange in subpanel ii. MESA reveals that it can be subdivided into more granular neighborhoods, colored orange, red and blue in subpanels iii and iv, and this pattern emerges consistently across multiple germinal centers. **d**, Heatmap visualizing cellular composition versus protein and mRNA expression levels. The red box highlights that, despite comparable cellular makeups in neighborhoods 0, 1 and 3, notable variances exist in protein and mRNA expression profiles. The values are normalized to 0 and 1 for visualization. **e**, DE and GSEA were conducted on the newly emerged neighborhoods (0 and 3), revealing distinct B cell developmental states within the germinal center's dark and light zones. DE analysis was performed using two-sided exact test with *P* values adjusted for multiple testing using the Benjamini–Hochberg (BH) procedure. The adjusted *P* value threshold was set at 0.05, and the effect size (log fold-change) threshold at 0.1. In the GSEA plot, the pathways colored blue are significantly enriched in neighborhood 0, red in neighborhood 3 and gray indicates no statistical significance after false delivery rate (FDR) correction. GSEA was conducted using normalized enrichment scores with permutation-based two-sided tests and *P* values adjusted using the BH procedure. The adjusted *P* value threshold was set at 0.05.

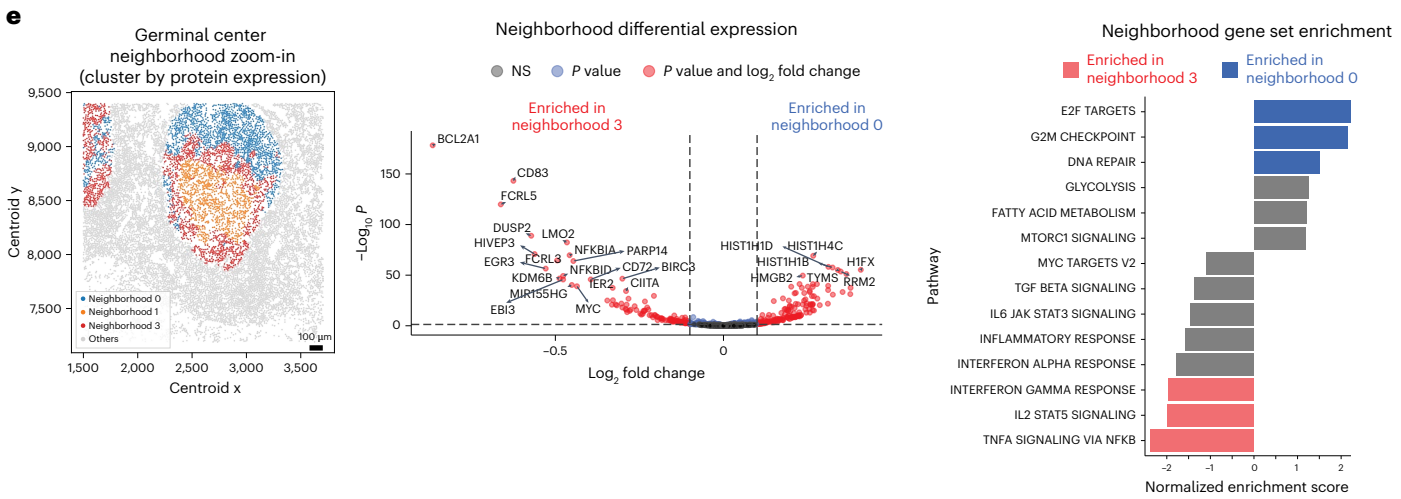
https://doi.org/10.1038/s41588-025-02119-z

Visualization of cell-type composition across hot/cold spots (Extended Data Fig. 5) revealed spatial heterogeneity within and across samples. In terms of cohabitation, diseased samples showed increased cohabitation between CD8$^+$ T cells and both CD106$^+$CD16/32$^+$CD31$^-$Ly6c$^-$ stromal cells (Fig. 3d, left) and ERTR7$^+$ stromal cells (Fig. 3d, middle).

Such patterns emerged only in hot spots, not whole tissues (Fig. 3d, Extended Data Fig. 4 and Supplementary Fig. 10). Enhanced spatial association between ERTR7$^+$ stromal cells and CD8$^+$ T cells suggests ERTR7$^+$ stromal cells actively facilitate T cell movement and possibly activation in diseased states[54]. In contrast, healthy samples showed







https://doi.org/10.1038/s41588-025-02119-z

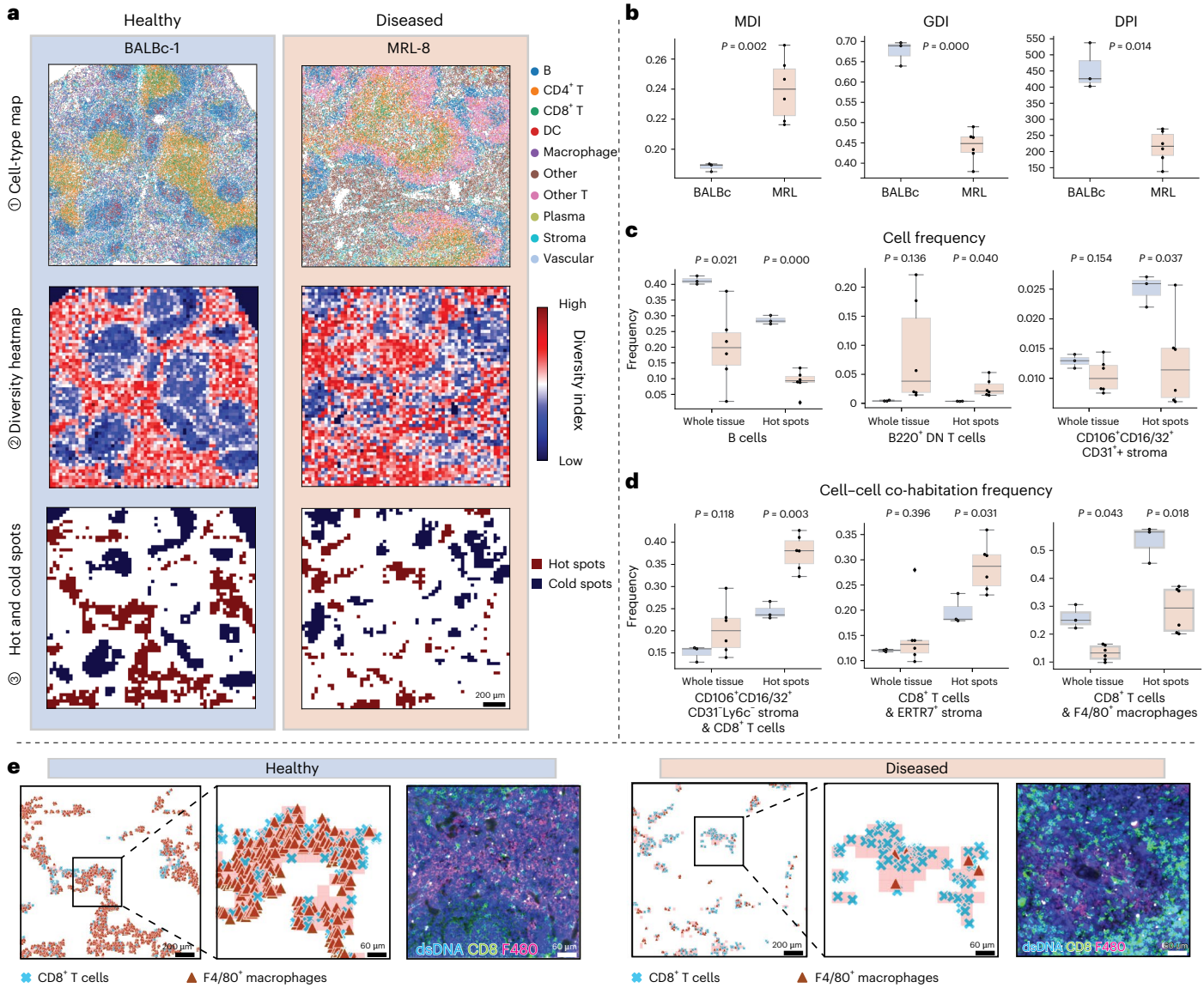


Fig. 3 | Eco-spatial analysis reveals distinct tissue remodeling in autoimmune disease. a, Side-by-side visualization of tissue sections from BALBc-1 (healthy control) and MRL-8 (MRL/lpr) mice showing (1) the cell-type map (top); (2) the diversity heatmap (middle), with varying shades indicating cellular diversity levels (red representing higher diversity and blue lower diversity); and (3) the map of the diversity hot spots and cold spots (bottom). DC, dendritic cell. **b**, Quantitative evaluation of MESA diversity metrics, including MDI, GDI and DPI, highlighting statistically significant differences in spatial diversity patterns between healthy ($n = 3$) and MRL/lpr tissues ($n = 6$; same sample size for **b**–**d**). Standard box plot metrics were used throughout: median (center line), quartiles (box), and whiskers (1.5 × IQR). Each point in **b**–**d** corresponds to an individual tissue sample. Healthy tissues show higher values of GDI ($P < 0.001$) and DPI ($P = 0.014$), whereas diseased MRL/lpr tissues exhibit a higher value of MDI ($P = 0.002$). Statistical comparisons were conducted across different tissue samples using two-sided Welch's *t*-test. **c**, Distribution of cell types as a percentage of total cell population in healthy and diseased spleen tissues. Comparisons were conducted using two-sided Welch's *t*-test and adjusted for FDR correction using the BH procedure. Notably, the differences in cell frequency between healthy and diseased tissues are more pronounced within the diversity hot spots than within the whole tissue. **d**, Cohabitation of different cell types in healthy and diseased tissues. Comparisons were conducted using a two-sided Welch's *t*-test and adjusted for FDR correction using the BH procedure. As before, the differences are more pronounced within the diversity hot spots than the entire tissue, with lower *P* values. **e**, Side-by-side visualization contrasting the cohabitation patterns of CD8$^+$ T cells and F4/80$^+$ macrophages in healthy and diseased tissues, with the corresponding CODEX images.

increased CD8$^+$ T cell and F4/80$^+$ macrophage cohabitation (Fig. 3d, right), consistent with F4/80$^+$ macrophages' role in immunoregulation and autoimmune prevention[55]. These differences were more pronounced in hot spots versus whole tissues across all cases studied, achieving statistical significance with lower *P* values. Fig. 3e shows CD8$^+$ T cell and F4/80$^+$ macrophage distribution in hot spots, with zoomed-in hot spot regions and corresponding CODEX images showing higher frequency in healthy versus diseased spleens. These alterations in cellular frequencies and interactions signify major restructuring of spleen tissue architecture and altered immune landscape, shedding light on mechanisms driving disease progression and immune dysregulation.

MESA improves prognostic capabilities for CRC

Prior research on colorectal cancer (CRC) identified spatial patterns linked to two subtypes: Crohn's-like reaction (CLR) and diffuse inflammatory infiltration (DII)[33]. These patterns, however, were previously not systematically quantified. Tissue samples from the previously analyzed CODEX patients with CRC were segmented into 50 µm × 50-µm patches.

Figure 4a shows the cell-type map (top), diversity heatmap (middle; red: higher, blue: lower) and diversity hot/cold spots (bottom).

Applying MESA metrics to the CRC dataset revealed notable differences between CLR and DII subtypes (Fig. 4b). CLR patients showed higher MDI with more dynamic variations in cellular diversity across scales (Fig. 4b, left), higher GDI indicating distinct segregation of high and low diversity regions (middle), and higher DPI reflecting larger, proximate hot spots (right). In contrast, DII patients displayed more intermixed patterns and smaller, dispersed hot spots. MESA's metrics quantitatively characterize CRC tissue organization and outperform CellCharter's metrics in distinguishing disease states (Supplementary Fig. 11).

We used MESA diversity metrics to model CRC patient survival outcomes, outperforming traditional CLR/DII annotations by pathologists. This demonstrates the value of these automatic metrics in patient stratification. A Cox proportional hazards model based on spatial diversity stratified patients into two groups with significantly different Kaplan–Meier curves ($P < 0.0001$). Notably, diversity-based stratification showed greater Kaplan–Meier curve separation than CLR/DII annotations (Fig. 4c, left/middle). Our model also achieved higher predictive accuracy, increasing the concordance index (C-index) from 0.717 to 0.734. Combining diversity metrics with CLR/DII annotations further improved prognostic power (C-index: 0.763; Fig. 4c, right).

We analyzed cellular composition and cohabitation within hot spots to identify cell types and their combinations associated with CLR and DII subtypes (Fig. 4d and Supplementary Figs. 11–14). Cell frequency shifts between CLR and DII patients matched findings from Schürch et al.[33], with more pronounced differences in hot spots versus whole tissue. For instance, B cell frequency was higher in CLR than DII, especially pronounced in hot spots versus whole tissue (Fig. 4d, left).

More interestingly, T_{reg} cells were more abundant in DII patients specifically within hot spots, supporting hypotheses from Schürch et al.[33] that lacked statistical backing in their original analysis. T_{reg} cell frequency differences between DII and CLR became significant only in hot spots, not in whole tissue (Fig. 4d, middle). These T_{reg} cells frequently co-occurred with $CD68^+CD163^+$ macrophages (M2-like) (Fig. 4d, right), aligning with the notion that T_{reg} cells, alongside M2-like macrophages, may promote immunosuppression[56,57] within tumor microenvironments, potentially driving poorer outcomes in DII patients.

Additionally, DII hot spots showed reduced cohabitation of B cells and $CD4^+$ T cells CD45RO (Supplementary Fig. 14), consistent with studies linking these cells to lower CRC risk[58]. Figure 4e illustrates cellular cohabitation in CLR and DII across whole tissues (left) and hot spots (right), where node size indicates cell-type abundance and color shows edge counts (redder indicates more). Edge thickness represents cohabitation frequency, with edges below 0.01 excluded for visual clarity. The cohabitation patterns with significant differences between CLR and DII—exclusively within hot spots (green edges)—include T_{reg} cells with M2-like macrophages and $CD8^+$ T cells, B cells with $CD4^+$ T cells $CD45RO^+$, and granulocytes with M2-like macrophages (pointed out by arrows in Fig. 4e, right). These localized cohabitation patterns revealed subtype differences not detected in whole-tissue analyses.

Lastly, using MESA's multiomics framework, we analyzed functional differences of cells based on spatial locations (Fig. 4f and Supplementary Fig. 15). Integrating CRC scRNA-seq[59] with CODEX using MaxFuse[22], we performed DE on $CD68^+CD163^+$ macrophages in hot versus cold spots (Fig. 4f, left). Cold spot macrophages expressed higher *LYVE1*, a marker for lymphatic/vessel-related M2-like macrophages[60–62]. Hot-spot macrophages showed elevated interferon-stimulated genes (*ISG20*, *ISG15*) and cytokines (*CXCL10*, *CCL5*), suggesting tumor-associated phenotypes[63–67]. GSEA showed hot-spot macrophages upregulated interferon responses, IL-2/TNF signaling, and complement pathways (Fig. 4f, right). MESA revealed distinct functional states of cells in hot spots versus cold spots.

MESA augments functional analysis for liver cancer

To demonstrate MESA's versatility, we applied it to a CosMx SMI spatial transcriptomics dataset of human liver tissues[34], analyzing LRIs in tissue remodeling across disease states. The dataset includes samples from healthy and hepatocellular carcinoma (HCC) liver tissues, each containing over 300,000 cells. Due to limited samples (one per condition), we implemented random subsampling to facilitate statistical comparisons, generating 10 representative subsamples (1,600 µm × 1,600 µm, ~12,000 cells) for each condition with preserved cellular composition (Supplementary Fig. 16). Analysis of the full dataset showed consistent patterns (Supplementary Fig. 17). Subsamples were divided into 50 µm × 50-µm patches. Figure 5a presents healthy and HCC samples showing cell-type map (top), diversity heatmap (middle; red: higher, blue: lower) and diversity hot/cold spots (bottom).

MESA metrics revealed statistically significant differences between healthy and HCC samples (Fig. 5b). Normal liver tissues showed higher MDI, reflecting heterogeneous cell-type diversity across scales due to their complex structure of lobules with hepatocytes, Kupffer cells, stellate cells and endothelial cells[68]. HCC tissues showed lower MDI due to homogeneous tumor regions with consistent diversity across scales (Fig. 5b, left). HCC tissues also showed higher GDI, indicating distinct segregation of high and low diversity areas, whereas the blended pattern in healthy tissues reflects heterogeneous zonation supporting diverse metabolic functions[69] (Fig. 5b, middle). Additionally, HCC tissues had higher DPI, with more expansive and proximate hot spots compared to healthy tissues (Fig. 5b, right).

Diversity hot spots showed distinct cellular composition and cohabitation patterns between healthy and HCC tissues (Fig. 5c, Extended Data Fig. 6 and Supplementary Figs. 18–20). HCC hot spots showed higher inflammatory macrophage abundance (Fig. 5c, upper left), and increased cohabitation with B cells (Fig. 5c, lower left) and non-inflammatory macrophages (Fig. 5c, lower right), as well

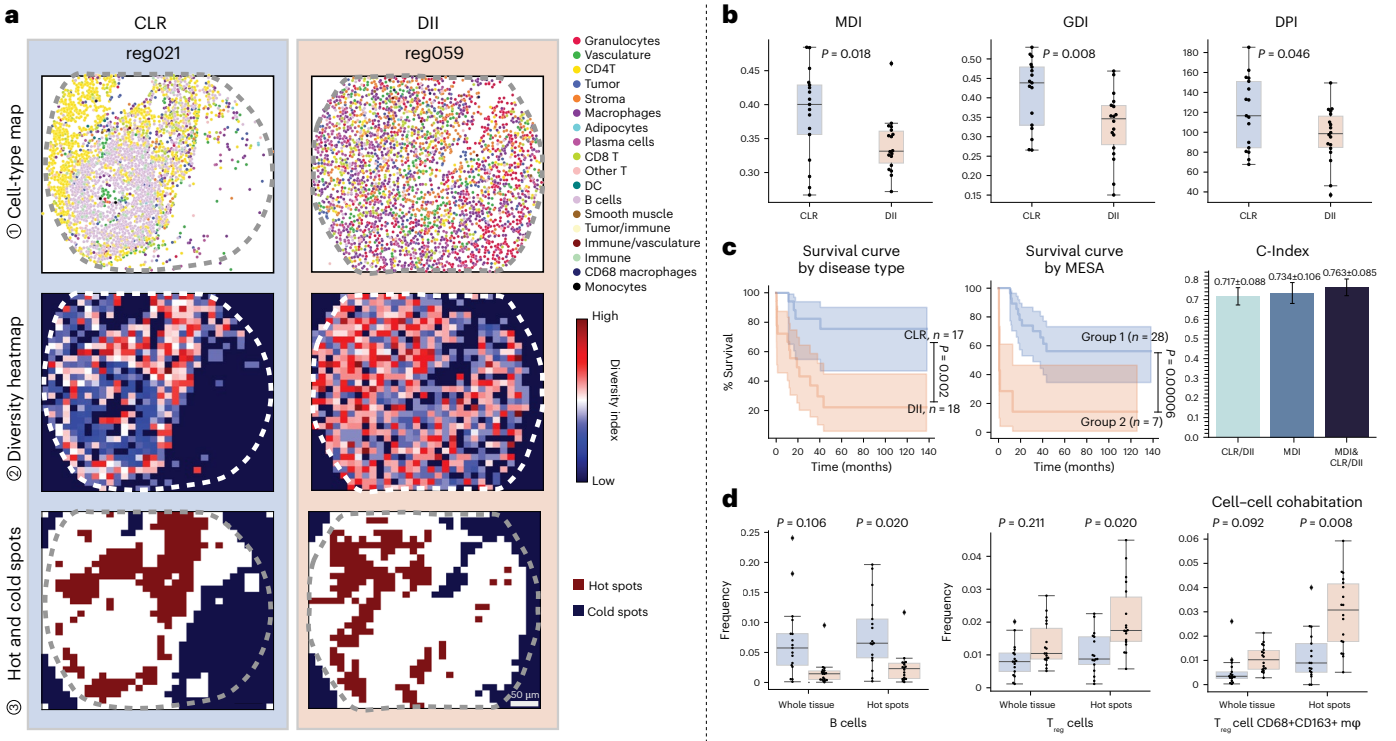
Fig. 4 | Multiomics and eco-spatial analysis improves prognostic capabilities for CRC. a, Visualization of tissue samples from CLR and DII patients: (1) the cell-type map (top), (2) the diversity heatmap (middle, red: higher diversity, blue: lower diversity) and (3) the map of the diversity hot spots and cold spots (bottom). **b**, Quantitative assessment of MESA diversity metrics in CLR ($n = 17$) and DII ($n = 18$) tissues (same sample size for **b** and **d**). Standard box plot metrics were used throughout. The analysis highlighted significant differences in spatial diversity patterns, with CLR tissues showing higher MDI ($P = 0.018$), GDI ($P = 0.006$) and DPI ($P = 0.046$), indicative of distinct spatial patterns between the two CRC subtypes. Statistical comparisons are conducted using two-sided Welch's *t*-test. **c**, Kaplan–Meier survival curves stratify CRC patients using pathologist-annotated CLR/DII classification and MESA diversity metrics. Two-sided log-rank tests show significant stratification for both approaches, with MESA metrics demonstrating a lower *P* value. Cox proportional hazards models demonstrate improved performance using MESA metrics based on the concordance index (C-index: 0.734 ± 0.106) over CLR/DII annotation (0.717 ± 0.088), with further improvement when combined (0.763 ± 0.085). **d**, Abundance of B cells, T_{reg} cells and T_{reg} cell cohabitation frequencies with $CD68^+CD163^+$ macrophages in CLR and DII tissues. Notably, the distinctions between CLR and DII are more pronounced within hot spots versus whole tissue, achieving lower *P* values based on two-sided Welch's *t*-test with BH FDR correction. **e**, Circos plots illustrate cellular abundance and cohabitation frequencies, where nodes represent distinct cell types and edges represent cohabitation relationships. Node size corresponds to cell-type abundance, whereas edge thickness denotes cohabitation strength. **f**, DE and gene set enrichment analyses reveal unique gene expression patterns distinguishing $CD68^+CD163^+$ macrophages located in hot spots versus cold spots. DE was performed with a two-sided exact test and GSEA with a two-sided permutation-based tests; all *P* values have been adjusted using the BH procedure. The adjusted *P* value threshold in DE was set at 0.05 and the effect size (log fold-change) threshold at 0.1.

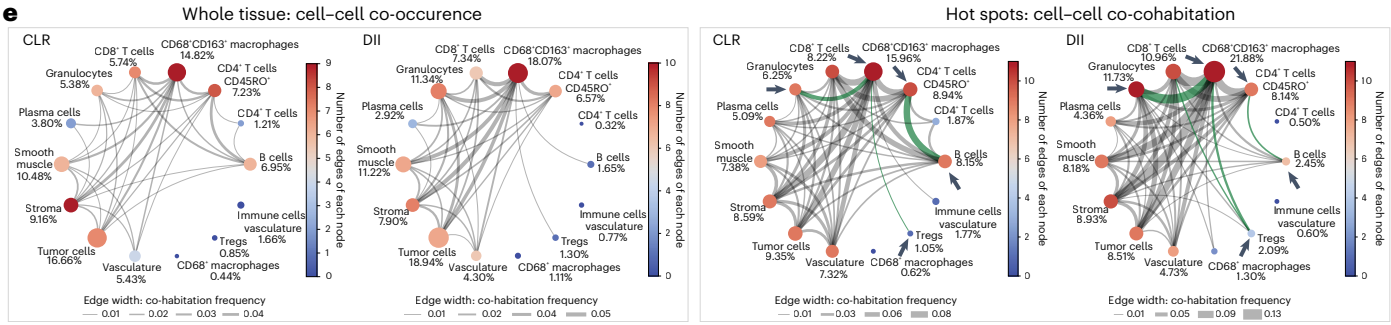
https://doi.org/10.1038/s41588-025-02119-z

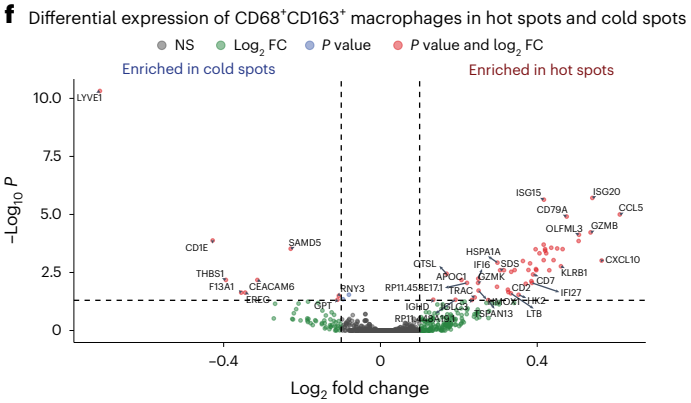
as increased T cell and noninflammatory macrophage cohabitation (Fig. 5c, upper right). These hot-spot patterns within hot spots were statistically significant and identified macrophages as keystone species in HCC. Notably, these patterns were not detected in whole-tissue analyses, highlighting the value of hot-spot-focused approaches.

To explore cell–cell communication, we performed LRI analysis using the mRNA modality with SpatialDM package[70]. Figure 5d shows a heatmap of LRI pathways (rows) across samples (columns). Normal and HCC samples are distinctly clustered, revealing differential LRI profile characteristics. Similar to hot-spot analysis, macrophage-related LRI pathways, including *SPP1*-related LRIs[71–73], were significantly different between healthy and HCC tissues, reinforcing the importance of macrophages in HCC. Additionally, we calculated an LRI communication score for each cell based on significant LRI counts, visualized by color intensity in Fig. 5e. Notably, high LRI-scoring regions overlapped with diversity hot spots, particularly in HCC tissues (highlighted in the last row of Fig. 5a,e).

However, analyzing LRIs with the 1,000-gene panel from CosMx SMI may not capture the full spectrum of interactions. MESA's multiomics component integrates the CosMx SMI data with scRNA-seq







https://doi.org/10.1038/s41588-025-02119-z

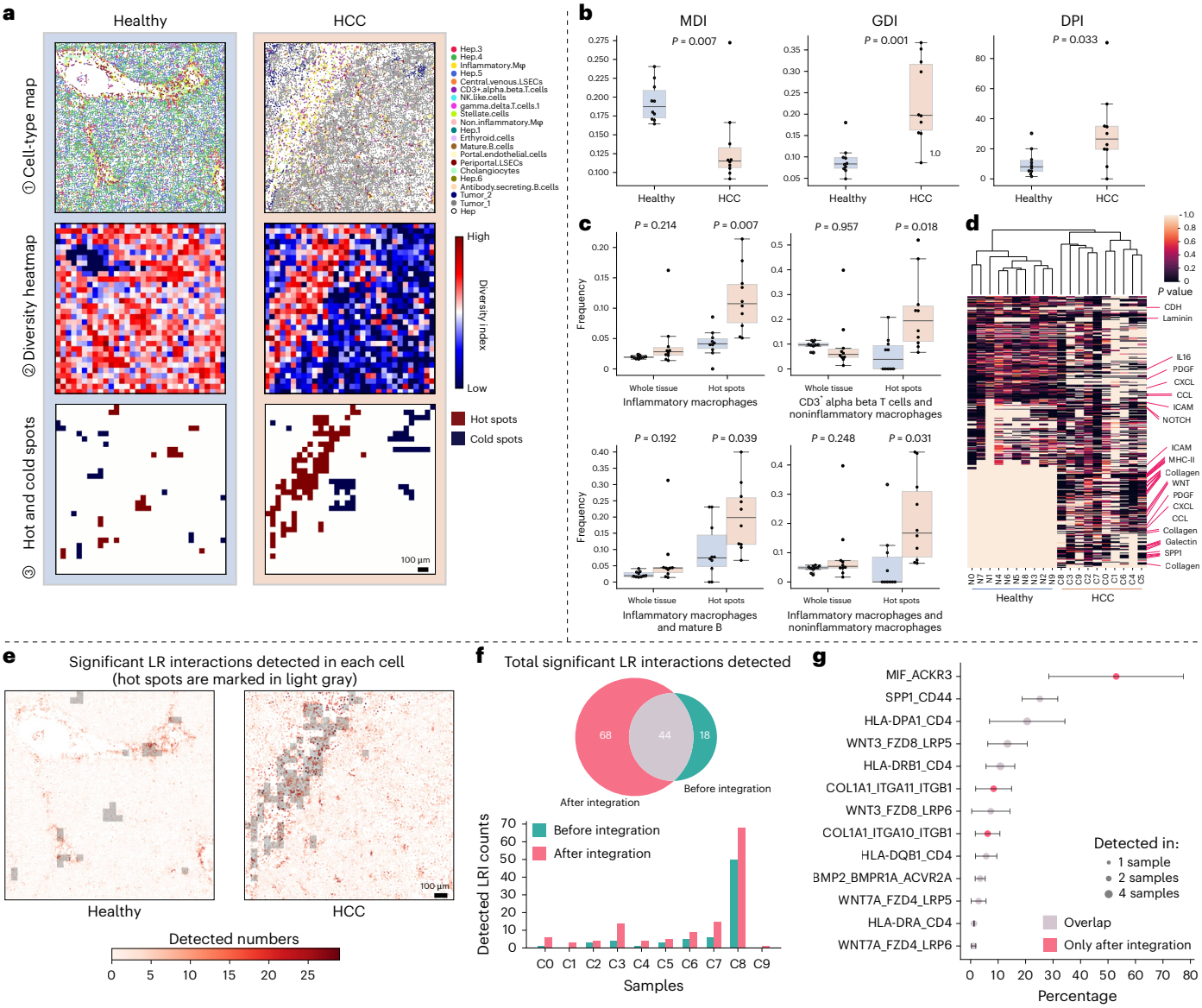


Fig. 5 | MESA extends to spatial transcriptomics and augments functional analysis of liver cancer. a, Visualization of healthy and HCC liver tissue subsamples showing: (1) the cell-type map (top), (2) the diversity heatmap (middle, red: higher diversity, blue: lower diversity) and (3) the map of the diversity hot spots and cold spots (bottom). **b**, Quantitative assessment of MESA diversity metrics, highlighting significant differences in spatial diversity patterns between healthy ($n = 10$) and HCC ($n = 10$) tissue subsamples (same sample size for panels **b**, **c** and **g**). Standard box plot metrics were used throughout. The HCC tissue shows higher values of GDI ($P = 0.001$) and DPI ($P = 0.033$), whereas the healthy tissue exhibits higher MDI value ($P = 0.007$). Statistical comparisons are conducted using two-sided Welch's t-test. **c**, Distribution of cell types and cellular cohabitation frequency in healthy and HCC liver tissues. The differences are more pronounced within hot spots versus whole tissue, achieving statistical significance with lower p-values (two-sided Welch's t-test with BH adjustment). **d**, Cluster heatmap showing differentially expressed LRI pathways (columns: samples, rows: pathways). Only pathways with differentially expressed LRIs in more than half of the samples are shown. Color intensity represents P values obtained from two-sided permutation tests, adjusted using BH procedure. **e**, Cells colored by their communication scores (number of detected LRIs), with hot spots marked in light gray. Notably, regions of high communication scores overlap with diversity hot spots. **f**, The Venn diagram illustrates the number of significant LRIs between tumor-associated macrophages and T cells, identified before and after the integration of the CosMx data with the scRNA-seq data using MaxFuse. Bar plot provides a detailed breakdown of LRIs identified in each subsample, demonstrating an improved analysis spectrum on LRIs with multiomics integration. **g**, Bubble plot showing detected LR pairs between tumor-associated macrophages and T cells (mean percentage ± 95% confidence interval). Bubble size indicates the number of subsamples in which the corresponding LRI was detected.

data from liver cancer samples[74] using MaxFuse to extend LRI analysis to the full transcriptome. Given the importance of macrophages and T cells in HCC tumor microenvironments, as identified in previous analyses, we focused on LRIs between these cell types. Figure 5f shows a Venn diagram of expanded LRI analysis from MESA's multiomics component, with 44 of 62 core LR pairs preserved across pre- and post-integration analyses.

Post-integration revealed 68 new LR pairs, extending the number of interactions beyond those captured pre-integration. The histogram in Fig. 5f illustrates the number of detected LRIs before and after multiomics integration across subsamples, highlighting the expanded scope of whole transcriptome analysis for characterizing cellular interactions. Figure 5g ranks top LR pairs by the percentage of cell pairs exhibiting specific interactions in each HCC subsample, with

pink dots representing pairs that are post-integration only and gray dots showing pairs detected both pre- and post-integration. The most prevalent LRI between macrophages and T cells was *MIF-ACKR3*. *ACKR3* (*CXCR7*) promotes macrophage migration and immune escape in HCC[75], whereas *MIF* correlates with tumor size and patient outcomes in HCC[76]. These findings underscore the critical role of T cell-macrophage interactions in HCC progression. Notably, *MIF-ACKR3* was undetectable in the original 1,000-gene panel, highlighting the importance of whole transcriptome integration.

Discussion

The development and implementation of MESA merges ecological principles with multiomics integration to decode the complex spatial arrangements of tissue environments. This analytical framework systematically quantifies cellular diversity, leveraging multiscale diversity indices to explore the association between cellular heterogeneity and phenotypic outcomes, such as disease progression. The application of MESA across various datasets was validated via its utility in uncovering previously unrecognized spatial structures and key cellular populations but also facilitated a deeper understanding of the relationships between cellular diversity and disease mechanisms.

In synthesizing the results, MESA's ability to enrich spatial-omics data with additional single-cell information has revealed granular insights into the spatial organization of cells. MESA enabled the identification of nuanced cellular neighborhoods, which were critically linked to functional phenotypic outcomes. The use of diversity indices to systematically assess spatial patterns underscores the intricacy of tissue microenvironments and the pivotal role of cellular heterogeneity in biological and pathological processes. Through integrated analysis of DE, gene set enrichment and LRI within these spatially defined cellular assemblies, MESA offers mechanistic insights on tissue remodeling across disease states.

MESA's two components (multiomics integration-enabled functional analysis of tissue compartments and ecology-inspired quantitative measurements of tissue states) can enhance spatial-omics analysis both independently and synergistically. For the multiomics component, we demonstrated its ability to facilitate delineation of distinct cellular interactions that were not observable using traditional techniques through the tonsil example, enhance functional understanding of tissue compartments through the CRC example and augment LRI analysis through the HCC example. Its effectiveness depends on the availability and quality of additional single-cell modalities, along with the integration method implemented. Modalities such as CycIF (typically containing 10–20 markers) may present greater integration challenges than CODEX or CosMx (containing more than 50 features)[77]. Additionally, while we used MaxFuse[22], alternative integration methods can be implemented[24–31,78]. Although computational integration expands analytical capabilities and generates hypotheses, experimental studies remain crucial for validation.

The ecological component is compatible with most single-cell spatial-omics technologies, as MESA metrics and hot/cold spot identification only require single-cell-level cell-type information and spatial locations. MESA's combination of multiomics and ecological components enables detailed mapping of cellular interactions and tissue environments, offering insights into tissue organization, disease mechanisms and therapeutic opportunities.

MESA application should extend beyond the datasets explored, with the potential to add further insights into the study of many tissue types or disease states characterized by complex cellular arrangements. MESA, coupled with the integration of emerging spatial profiling technologies and single-cell datasets, will likely expand its utility, enabling researchers to navigate the spatial complexities of biological systems with unprecedented precision and depth. As such, MESA offers a scalable and versatile toolset for studying the spatial architectures of tissues in health and disease.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-025-02119-z.

References

1. Angelo, M. et al. Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* **20**, 436–442 (2014).
2. Goltsev, Y. et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* **174**, 968–981.e15 (2018).
3. Giesen, C. et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
4. Lin, J.-R. et al. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018).
5. Chen, K. H. et al. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
6. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
7. Zhang, M. et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).
8. Greenbaum, S. et al. A spatially resolved timeline of the human maternal–fetal interface. *Nature* **619**, 595–605 (2023).
9. Hickey, J. W. et al. Organization of the human intestine at single-cell resolution. *Nature* **619**, 572–584 (2023).
10. Bhate, S. S., Barlow, G. L., Schürch, C. M. & Nolan, G. P. Tissue schematics map the specialization of immune tissue motifs and their appropriation by tumors. *Cell Syst.* **13**, 109–130 (2022).
11. Palla, G. et al. Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).
12. Edsgärd, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods* **15**, 339–342 (2018).
13. Svensson, V., Teichmann, S. A. & Stegle, O. Spatialde: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).
14. Chen, Z., Soifer, I., Hilton, H., Keren, L. & Jojic, V. Modeling multiplexed images with spatial-lda reveals novel tissue microenvironments. *J. Comput. Biol.* **27**, 1204–1218 (2020).
15. Kim, J. et al. Unsupervised discovery of tissue architecture in multiplexed imaging. *Nat. Methods* **19**, 1653–1661 (2022).
16. Varrone, M., Tavernari, D., Santamaria-Martínez, A., Walsh, L. A. & Ciriello, G. Cellcharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. *Nat. Genet.* **56**, 74–84 (2023).
17. Dale, M. & Fortin, M.-J. *Spatial Analysis* 2nd edn. (Cambridge Univ. Press, 2014).
18. Turner, M. G. and Gardner, R. H. *Landscape Ecology in Theory and Practice* (Springer, 2015).
19. Boyce, Hunter Bryan and Mallick, P. Geostatistical visualization of ecological interactions in tumors. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE, 2019).
20. Moses, L. et al. Voyager: exploratory single-cell genomics data analysis with geospatial statistics. Preprint at *bioRxiv* https://doi.org/10.1101/2023.07.20.549945 (2023).
21. Feng, Y. et al. Spatial analysis with spiat and spasim to characterize and simulate tissue microenvironments. *Nat. Commun.* **14**, 2697 (2023).
22. Chen, S. et al. Integration of spatial and single-cell data across modalities with weakly linked features. *Nat. Biotechnol.* **42**, 1096–1106 (2024).
23. Zhu, B. et al. Robust single-cell matching and multimodal analysis using shared and distinct features. *Nat. Methods* **20**, 304–315 (2023).

24. Cao, Zhi-Jie & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 (2022).
25. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887 (2019).
26. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
27. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
28. Rahimi, A., Vale-Silva, L. A., Fälth Savitski, M., Tanevski, J. & Saez-Rodriguez, J. Dot: a flexible multi-objective optimization framework for transferring features across single-cell and spatial omics. *Nat. Commun.* **15**, 4994 (2024).
29. Mages, S. et al. Tacco unifies annotation transfer and decomposition of cell identities for single-cell and spatial omics. *Nat. Biotechnol.* **41**, 1465–1473 (2023).
30. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat. Methods* **16**, 1289–1296 (2019).
31. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat. Methods* **18**, 1352–1362 (2021).
32. Kennedy-Darling, J. et al. Highly multiplexed tissue imaging using repeated oligonucleotide exchange reaction. *Eur. J. Immunol.* **51**, 1262–1277 (2021).
33. Schürch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* **182**, 1341–1359 (2020).
34. He, S. et al. High-plex imaging of rna and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).
35. Hickey, J. W. et al. T cell-mediated curation and restructuring of tumor tissue coordinates an effective immune response. *Cell Rep.* **42**, 113494 (2023).
36. Karimi, E. et al. Single-cell spatial immune landscapes of primary and metastatic brain tumours. *Nature* **614**, 555–563 (2023).
37. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).
38. Hu, Y. et al. Unsupervised and supervised discovery of tissue cellular neighborhoods from cell phenotypes. *Nat. Methods* **21**, 267–278 (2024).
39. Zhao, E. et al. Spatial transcriptomics at subspot resolution with bayesspace. *Nat. Biotechnol.* **39**, 1375–1384 (2021).
40. King, H. W. et al. Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci. Immunol.* **6**, eabe6291 (2021).
41. Massoni-Badosa, R. et al. An atlas of cells in the human tonsil. *Immunity* **57**, 379–399 (2024).
42. Zhang, T., Ramakrishnan, R. & Livny, M. Birch: an efficient data clustering method for very large databases. *ACM Sigmod. Rec.* **25**, 103–114 (1996).
43. Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* **32**, 241–254 (1967).
44. Van Gassen, S. et al. Flowsom: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A* **87**, 636–645 (2015).
45. Singhal, V. et al. Banksy unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nat. Genet.* **56**, 431–441 (2024).
46. Holmes, A. B. et al. Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. *J. Exp. Med.* **217**, e20200483 (2020).
47. Attaf, N., Baaklini, S., Binet, L. & Milpied, P. Heterogeneity of germinal center B cells: New insights from single-cell studies. *Eur. J. Immunol.* **51**, 2555–2567 (2021).
48. Glass, D. R. et al. An integrated multi-omic single-cell atlas of human B cell identity. *Immunity* **53**, 217–232 (2020).
49. De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* **15**, 137–148 (2015).
50. Mandelbrot, B. B. *The Fractal Geometry of Nature, Revised and Enlarged Edition* (W. H. Freeman, 1983).
51. Patton, D. R. A diversity index for quantifying habitat "edge". *Wildlife Soc. Bull.* **3**, 171–173 (1975).
52. Getis, A. & Ord, J. K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **24**, 189–206 (1992).
53. Ord, J. K. & Getis, A. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr. Anal.* **27**, 286–306 (1995).
54. Burrell, B. E. et al. Lymph node stromal fiber ER-TR7 modulates $CD4^+$ T cell lymph node trafficking and transplant tolerance. *Transplantation* **99**, 1119–1125 (2015).
55. Kurotaki, D. et al. CSF-1–dependent red pulp macrophages regulate CD4 T cell responses. *J. Immunol.* **186**, 2229–2237 (2011).
56. Sun, W. et al. A positive-feedback loop between tumour infiltrating activated T_{reg} cells and type 2-skewed macrophages is essential for progression of laryngeal squamous cell carcinoma. *Br. J. Cancer* **117**, 1631–1643 (2017).
57. Pan, Y. et al. Tumor-associated macrophages in tumor immunity. *Front. Immunol.* **11**, 583084 (2020).
58. Pagès, F. et al. Immune infiltration in human tumors: a prognostic factor that should not be ignored. *Oncogene* **29**, 1093–1102 (2009).
59. Pelka, K. et al. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734–4752 (2021).
60. Siret, C. et al. Deciphering the heterogeneity of the $LYVE1^+$ perivascular macrophages in the mouse brain. *Nat. Commun.* **13**, 7366 (2022).
61. Kieu, ThoaiQuoc et al. Kinetics of LYVE-1-positive M2-like macrophages in developing and repairing dental pulp in vivo and their pro-angiogenic activity in vitro. *Sci. Rep.* **12**, 5176 (2022).
62. Lim, HweeYing et al. Hyaluronan receptor LYVE-1-expressing macrophages maintain arterial tone through hyaluronan-mediated regulation of smooth muscle cell collagen. *Immunity* **49**, 326–341 (2018).
63. Peng, Y. et al. Integrated bioinformatics analysis and experimental validation reveal isg20 as a novel prognostic indicator expressed on m2 macrophage in glioma. *BMC Cancer* **23**, 596 (2023).
64. Wu, N. et al. The interferon stimulated gene 20 protein (ISG20) is an innate defense antiviral factor that discriminates self versus non-self translation. *PLoS Pathog.* **15**, e1008093 (2019).
65. Baldanta, S. et al. Isg15 governs mitochondrial function in macrophages following vaccinia virus infection. *PLoS Pathog.* **13**, e1006651 (2017).
66. Ardighieri, L. et al. Infiltration by CXCL10 secreting macrophages is associated with antitumor immunity and response to therapy in ovarian cancer subtypes. *Front. Immunol.* **12**, 690201 (2021).
67. Liu, C. et al. Macrophage-derived CCL5 facilitates immune escape of colorectal cancer cells via the p65/STATt3-CSN5-PD-L1 pathway. *Cell Death Differ.* **27**, 1765–1781 (2020).
68. Guilliams, M. et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379–396 (2022).
69. Ben-Moshe, S. & Itzkovitz, S. Spatial heterogeneity in the mammalian liver. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 395–410 (2019).
70. Li, Z. et al. SpatialDM for rapid identification of spatially co-expressed ligand-receptor and revealing cell-cell communication patterns. *Nat. Commun.* **14**, 3995 (2023).
71. Qi, J. et al. Single-cell and spatial analysis reveal interaction of FAP^+ fibroblasts and $SPP1^+$ macrophages in colorectal cancer. *Nat. Commun.* **13**, 1742 (2022).

72. Bill, R. et al. Cxcl9: Spp1 macrophage polarity identifies a network of cellular programs that control human cancers. *Science* **381**, 515–524 (2023).
73. Ouyang, J. F. et al. Systems level identification of a matrisome-associated macrophage polarisation state in multi-organ fibrosis. *eLife* **12**, e85530 (2023).
74. Ma, L. et al. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *J. Hepatol.* **75**, 1397–1408 (2021).
75. Chen, F. et al. Histone deacetylase 2 regulates STAT1-dependent upregulation of atypical chemokine receptor 3 to induce M2 macrophage migration and immune escape in hepatocellular carcinoma. *Mol. Immunol.* **151**, 204–217 (2022).
76. Wirtz, T. H. et al. Macrophage migration inhibitory factor predicts an unfavorable outcome after transarterial chemoembolization for hepatic malignancies. *Clin. Transl. Sci.* **14**, 1853–1863 (2021).
77. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science* **381**, eabq4964 (2023).
78. Wang, G. et al. scmodal: a general deep learning framework for comprehensive single-cell multi-omics data alignment with feature links. Preprint at *bioRxiv* https://doi.org/10.1101/2024.10.01.616142 (2024).

Methods

No new biological experiments were performed in this study, and thus no ethical regulations were required.

Multiomics spatial analyses

Multiomics data fusion with MaxFuse. We integrated spatial-omics data (for example, CODEX, MIBI, IMC and CosMx) with single-cell data (for example, scRNA-seq and scATAC-seq) using algorithms like MaxFuse. The detailed process of MaxFuse has been previously described in[22], and the relevant code used in this study has been deposited. In brief, for the given dataset input (single-cell spatial and sequencing data), we first identified 'shared' features across modalities (for example, CD3 protein in CODEX and CD3-related mRNAs in scRNA-seq). The shared features were used to calculate a cross-modality distance matrix based on Pearson correlation, and subsequently, a round of initial matching (via linear assignment) was employed to create first-round single-cell level pairing information. Afterward, Canonical Correlation Analysis (CCA) was implemented to produce a cross-modality distance matrix based on the CCA embeddings which incorporates the information from all features, and then a second round of matching was performed via linear assignment. When constructing the distance matrices in the initial and secondary matching, a smoothing process based on the global cell population structure was performed to boost the correlation between the weakly linked features. Finally, cell pairs with low scores (based on the CCA embeddings) were removed, so only high-quality matching pairs were used for downstream analysis. The details on hyperparameters used in the MaxFuse process for each dataset can be found in the deposited code.

Neighborhood identification. Local cellular neighborhoods were defined using the k nearest neighbors of each cell based on spatial distance, where k was set to 20 in our analyses. This value of k balances cellular resolution with sufficient spatial context of local interactions, consistent with previous studies[10]. Empirical validation across various tissue types and spatial profiling technologies suggests that k values between 10 and 30 yield robust results.

For each cell i, we aggregated information from its k nearest neighbors, focusing on three key types of information: cellular composition, protein expression and mRNA expression (via in silico multiomics fusion). This aggregation process involved computing the average of these features, resulting in a Neighborhood Feature Vector (NFV) for each cell, denoted as NFV_i. The NFV for a feature f in the neighborhood of cell i is calculated as:

$$\mathrm{NFV}_i(f) = \frac{1}{K}\sum_{k=1}^{K} f_k,$$

where f_k represents the feature vector of the k-th nearest neighbor. If the feature of interest is cellular composition, with M distinct cell types, f_i is an M-dimensional binary vector indicating cell-type identity. If the feature of interest is protein or mRNA expression, f_i is a p-dimensional vector representing expression level, where p is the number of protein or mRNA markers.

We identified conserved, distinct cellular neighborhoods by applying k-means clustering to these NFVs. The optimal number of clusters was determined using the elbow method based on within-cluster sum of squares[79], which quantifies cluster compactness through total squared distances between points and their respective cluster centroids.

Neighborhood characterization. To characterize molecular signatures and functional pathways in identified neighborhoods, we performed DE analysis and GSEA. We conducted DE using two-sided exact tests to identify differentially expressed genes between cellular neighborhoods (edgeR v.3.36.0), with P values adjusted for multiple testing using the Benjamini–Hochberg (BH) procedure. GSEA was performed using two-sided permutation-based tests on normalized enrichment scores (fgsea v.1.20.0) with hallmark gene sets[80] to identify enriched pathways. The integration of scRNA-seq data with spatial-omics data was critical for this step. Multiomics integration enriched the spatial data by providing more comprehensive gene expression profiles, thereby enabling deeper characterization of functional states of distinct tissue microenvironments.

Ecological spatial analyses

Tissue samples are dissected into gridded patches across a range of scales, denoted as a series $(1/2)^n$ ($n = 0, 1, 2, \ldots$), to facilitate multiscale analysis. At each scale, we compute diversity indices for individual patches and average them to obtain the scale's overall diversity. The MDI is then derived by calculating the slope of a linear regression line fitted to these scale-specific diversity indices, quantifying how diversity shifts across spatial scales. To examine the spatial structure of cellular diversity, we perform spatial autocorrelation analyses using a diversity heatmap. We generate this heatmap by partitioning samples into patches of approximately 50 µm × 50 µm in size, with each patch containing roughly 10 to 20 cells. The GDI and LDI are derived from this heatmap, capturing broad-scale and localized diversity patterns. Using the LDI, we identify regions of high diversity (hot spots) and low diversity (cold spots). This delineation enables an in-depth examination of cell-type composition and cohabitation patterns within these defined regions, providing a different perspective of cellular interactions and distribution dynamics beyond whole-tissue analyses. Each of these analytical steps is detailed in the following sections.

Spatial tessellation. Given a tissue sample S of dimensions $m \times n$, we partition it into a set of non-overlapping rectangles, denoted as $\{s_i\}_{i=1}^{t}$, at a predefined scale ϵ, with t representing the total number of rectangle patches. Each rectangle s_i has dimensions $\frac{m}{\epsilon} \times \frac{n}{\epsilon}$, where $s_i \cap s_{j,j\neq i} = \varnothing$, $\bigcup_i^t s_i = S$, ensuring complete coverage without overlap. Each rectangle s_i within the sample is defined as $s_i = \{(x,y) \in \mathbb{R}^2 : a \le x \le a + \frac{m}{\epsilon}, b \le y \le b + \frac{n}{\epsilon}\}$, where (a, b) represents the lower left corner coordinates in $\mathbb{R}^2$. The choice of scale depends on the spatial density of cells in the dataset: for the CODEX mouse spleen dataset, the scale is set to $\epsilon = 64$, resulting in a grid of patches of size around 30 µm × 30 µm; for the CODEX CRC dataset, the scale is set to $\epsilon = 32$, leading to patches with a larger size of around 50 µm × 50 µm; for the CosMx spatial transcriptomic dataset of human liver tissues, the scale is set to $\epsilon = 32$, yielding a grid of patches of size around 50 µm × 50 µm. Detailed guidance on spatial tessellation scale selection is available in MESA's tutorial: https://mesa-py.readthedocs.io/en/latest/index.html.

Shannon diversity index. The Shannon diversity index[81], denoted as H, is a measure originated in information theory but has been widely adopted across fields, including ecology. This index quantifies the uncertainty or entropy in predicting the category of an individual data point randomly sampled from a dataset, and is defined as:

$$H = -\sum_{i=1}^{S} p_i \log p_i \quad (1)$$

where, in the context of cellular diversity based on cell-type information, S represents the total number of distinct cell types and p_i denotes the proportion of cells belonging to the ith cell type. Higher H values indicate greater cell-type diversity and more heterogeneous cellular composition.

MDI. Let $\epsilon > 0$ be a scale parameter and $N(\epsilon)$ denote the number of patches at scale ϵ. For each patch $n = 1, 2, \ldots, N(\epsilon)$, we denote $C(n, \epsilon)$ as the number of distinct cell types in patch n at scale ϵ. Let $p_c(n, \epsilon)$ represent

the proportion of cell type c in patch n at scale ϵ, where $c = 1, 2, \ldots, C(n, \epsilon)$. We define the diversity index at scale ϵ as

$$I_D(\epsilon) = -\frac{1}{N(\epsilon)} \sum_{n=1}^{N(\epsilon)} \sum_{c=1}^{C(n,\epsilon)} p_c(n,\epsilon) \log(\, p_c(n,\epsilon)). \quad (2)$$

This formulation computes the average entropy across all patches at scale ϵ, quantifying cellular diversity at each spatial resolution. The MDI is derived as the slope of the linear regression between $I_D(\epsilon)$ and $\log(\epsilon)$, quantifying the rate of change in cellular diversity across spatial scales within tissues.

GDI and LDI. Following our spatial tessellation procedure, we partition tissue samples into nonoverlapping patches at scale ϵ. Within each patch, we calculate the Shannon diversity index, which is then used to assemble a diversity heatmap for the tissue sample. The GDI is derived by applying Moran's I statistic to this heatmap, which quantifies spatial autocorrelation as discussed in the next section. The LDI is computed using local Moran's I statistics and identifies significant spatial clustering patterns in cellular diversity. Using LDI, we identify diversity hot spots (regions of high diversity values surrounded by similarly high values) and cold spots (regions of low diversity values surrounded by low values).

Moran's *I*. Moran's I is a measure of spatial autocorrelation that quantifies the degree to which a variable, such as cellular diversity, correlates with itself across space[82]. For a variable x at location i (for example, cellular diversity), we calculate its deviation from the mean as $z_i = x_i - \bar{x}$, where $\bar{x}$ is the mean across all locations. Moran's I is defined as:

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} w_{ij} z_i z_j}{\sum_{i=1}^{N} z_i^2}, \quad (3)$$

where N is the number of patches and $W = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$ is the sum of all spatial weights w_{ij}, which quantify the spatial relationship between patch i and patch j. In our study, we set $w_{ij} = 1$ if patch i and patch j are spatial neighbors, and $w_{ij} = 0$ otherwise. The spatial weight matrix, W, is constructed by representing the diversity heatmap as a 2D grid, where connectivity between spatial units is typically established using a rook structuring element, which considers two units as neighbors if they share an edge. Alternative choices of structuring elements include bishop (sharing a vertex) or queen (sharing an edge or a vertex), depending on desired spatial contiguity criteria. Values range from −1 (complete dispersion) to 1 (perfect correlation), with positive values indicating spatial clustering, negative values dispersion, and zero no spatial autocorrelation.

Local Moran's *I*. Local Moran's I measures spatial autocorrelation at the location-specific level[83]. The Local Moran's I for patch i is defined as:

$$I_i = z_i \sum_j w_{ij} z_j, \quad (4)$$

where w_{ij} are elements of the spatial weight matrix W defining spatial connectivity between patch i and j, and z_i and z_j are the deviations from the mean as defined above. To assess the statistical significance of the observed local Moran's I values, a permutation approach is used, with the number of permutations set to 999 in our analyses.

Hot spots and cold spots—defined as the spatial clustering of high or low local Moran's I values surrounded by similar values—are identified based on a predetermined significance threshold: $P = 0.01$ for the spleen and liver datasets, and $P = 0.05$ for the CRC dataset due to the lower cell density observed in some CRC tissues. For practical usage, we recommend visualizing tissues using the significance map, where each patch is annotated by the significance level of local Moran's I value (Supplementary Fig. 21). Although a P value threshold of 0.01 generally yields robust results, the threshold can be relaxed to 0.05 for sparse cell distributions. LDI quantifies local spatial autocorrelation and enables the identification of diversity hot spots and cold spots, whereas GDI measures tissue-wide spatial autocorrelation to reveal broad-scale diversity patterns.

DPI. We identify contiguous 'islands' of diversity by grouping proximate hot spots or cold spots using connected components labeling with rook contiguity (ndimage.label, *scipy*, v.1.11.2). Drawing inspiration from island biogeography[84], we introduce the DPI:

$$DPI = \sum_{i=1}^{N} \frac{S_i}{d_i}, \quad (5)$$

where S_i quantifies the area of each identified island and d_i denotes the shortest distance from island i to its nearest neighboring island. Lower DPI values indicate smaller, more dispersed islands, whereas higher DPI values reflect larger, more proximate islands.

Bray–Curtis dissimilarity. We assess the cellular composition between two islands using the Bray–Curtis dissimilarity[85], which is defined as:

$$\mathrm{BCD} = \frac{\sum_{i=1}^{n} |x_i - y_i|}{\sum_{i=1}^{n} (x_i + y_i)}$$

where x_i represents the count of the i-th cell type in island x, and y_i represents the count of the i-th cell type in island y (spatial.distance.braycurtis, scipy v.1.11.2).

Cell-type frequency and cohabitation. We analyze cellular composition by quantifying the frequency of each cell type and the cohabitation frequencies of cell-type pairs within diversity hot spots and cold spots, and across the entire tissue. The frequency of a particular cell type i is calculated as N_i/N_{total}, where N_i is the cell count of type i, and N_{total} is the total cell count in the region of interest. Specifically, for the whole-tissue analysis, N_{total} is the sum of all cells across the tissue sample, whereas for the hot-spot and cold-spot analysis, N_{total} corresponds to the cell count within these delineated regions.

Cohabitation frequency for cell types i and j is defined as:

$$F_{ij,\text{total}} = \frac{N_{ij}}{N_{\text{patches,total}}}$$

for whole-tissue analysis, where N_{ij} is the number of patches containing both cell types and $N_{\text{patches,total}}$ is the total number of patches in the entire tissue. For the hot spots and cold spots analysis, $N_{ij,\text{ hot spots}}$ or $N_{ij,\text{ cold spots}}$ is the number of patches that contain both cell types i and j within those spots, respectively. We divide this value by the total count of patches identified as hot spots or cold spots, denoted as $N_{\text{patches,hotspots}}$ or $N_{\text{patches,coldspots}}$:

$$F_{ij,\text{hot spots or cold spots}} = \frac{N_{ij,\text{hot spots}} \text{ or } N_{ij,\text{cold spots}}}{N_{\text{patches,hot spots}} \text{ or } N_{\text{patches,cold spots}}}$$

This provides a normalized measure of cell cohabitation and enables standardized comparison across tissue regions with varying cell distributions. Statistical significance is assessed using Welch's t-test for comparing cell-type frequencies and cohabitation patterns between conditions, with P values adjusted using BH procedure for false discovery rate control.

Survival analysis. We assessed the prognostic value of diversity metrics for CRC patient survival using the Cox proportional hazards regression model (lifelines, v.0.27.7). The model was fitted using the

diversity indices across scales with fivefold cross-validation. Model performance was assessed using the concordance index (C-index), which quantifies predictive accuracy in ranking patient survival times. We compared C-index values across cross-validation folds for models based on pathologist-annotated CLR/DII subtypes, automatically calculated diversity indices, and a combination of both.

Spatial cell–cell communication. We identified spatially significant LRIs in the spatial transcriptomics dataset from human liver tissues[34] using SpatialDM (v.0.2.0)[70]. SpatialDM takes a curated LRI database as input, such as CellChatDB[86] by default, and extracts LR pairs that appear in the sample for downstream hypothesis testing. It involves computing the bi-variate Moran's *I* statistic for each LR pair. Statistical significance was assessed using a permutation test, with the number of permutations set to 500. LRIs are considered significant if they achieve a *P* value of less than 0.05 after FDR correction with the BH procedure.

Benchmarking analysis. To evaluate MESA's ability to discover spatial compartments, we compared it with three established methods: Spatial-LDA[14], UTAG[15] and CellCharter[16]. We implemented CellCharter (v.0.2.1), Spatial-LDA (v.0.1.3) and UTAG (v.0.1.1) following their official documentation (https://cellcharter.readthedocs.io/en/latest/, https://github.com/calico/spatial_lda, https://github.com/ElementoLab/utag/tree/main, respectively). Spatial-LDA was applied using cell-type annotations and cell spatial coordinates as inputs. CellCharter and UTAG analyses were conducted using protein expression data and cell spatial coordinates. All methods were implemented using default parameters as specified in their tutorials.

Moreover, we benchmarked MESA's metrics (MDI, GDI and DPI) against other quantitative metrics previously proposed for spatial-omics data, including linearity, curl, elongation and purity from CellCharter[16], on their effectiveness in differentiating tissue states. The shape metrics were calculated following CellCharter's default parameters (https://cellcharter.readthedocs.io/en/latest/). All analyses used cell-type annotations and cell spatial coordinates as inputs, with *P* values adjusted using the BH procedure.

Statistics and reproducibility. All data used in this study were generated from other studies, and therefore no statistical method was used to predetermine the sample size. Similarly, no data were excluded from the original dataset for analysis. Since no new experiment was performed, randomization of study design or blinding is irrelevant to this study. During data analysis, randomization was performed when relevant, and the researcher was blinded to the outcome until the visualization of results.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used in this manuscript are publicly available via the following links. CODEX tonsil data[32]: https://doi.org/10.1002/eji.202048891; scRNA-seq tonsil data[40]: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165860; CODEX mouse spleen data[2]: https://doi.org/10.1016/j.cell.2018.07.010; CODEX CRC data[33]: https://doi.org/10.1016/j.cell.2020.07.005; scRNA-seq CRC data[59]: https://doi.org/10.1016/j.cell.2021.08.003; CosMx liver data[34]: https://doi.org/10.1038/s41587-022-01483-z; scRNA-seq liver data[74]: https://doi.org/10.1016/j.jhep.2021.06.028; CellChatDB database[86]: https://doi.org/10.1038/s41467-021-21246-9.

Code availability

MESA code[87] can be accessed via https://mesa-py.readthedocs.io/en/latest/ and on GitHub via https://github.com/Feanor007/MESA.

References

79. Kodinariya, T. M. et al. Review on determining number of cluster in k-means clustering. *Int. J.* **1**, 90–95 (2013).
80. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
81. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
82. Moran, P. A. P. The interpretation of statistical maps. *J. R. Stat. Soc. Ser. B Methodol.* **10**, 243–251 (1948).
83. Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* **27**, 93–115 (1995).
84. Gustafson, E. J. & Parker, G. R. Using an index of habitat patch proximity for landscape design. *Landsc. Urban Plan.* **29**, 117–130 (1994).
85. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monog.* **27**, 325–349 (1957).
86. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun.* **12**, 1088 (2021).
87. Ding, Daisy Yi et al. Feanor007/MESA: MESA-py: Multiomics and Ecological Spatial Analysis beta release (v.0.1.0). *Zenodo* https://doi.org/10.5281/zenodo.14504057 (2024).

Acknowledgements

This work was funded in part by grants from, the National Institutes of Health 3U54HG010426 (G.P.N.), the US Food and Drug Administration Medical Countermeasures Initiative contracts HHSF223201610018C and 75F40120C00176 (G.P.N.), the Parker Institute for Cancer Immunotherapy (G.P.N.), Cancer Research UK C27165/A29073 (G.P.N.), and the Rachford and Carlota A. Harris Endowed Professorship (G.P.N.). NIH 1P01AI177687-01 (A.K.S.), and Bill & Melinda Gates Foundation INV-055706 (A.K.S.), and Stanford Graduate Fellowship (D.Y.D), and Stanford Data Science Scholarship (D.Y.D). This article reflects the views of the authors and should not be construed as representing the views or policies of the institutions that provided funding.

Author contributions

D.Y.D., Z.T., B.Z., H.R., A.K.S., R.T. and G.P.N. contributed to the conceptualization of the study. D.Y.D., Z.T. and B.Z. were responsible for algorithm development and implementation. Analysis was conducted by D.Y.D., Z.T. and B.Z. The initial draft of the manuscript was written by D.Y.D., Z.T., B.Z. and G.P.N., and D.Y.D., Z.T., B.Z., A.K.S., R.T. and G.P.N. participated in the review and editing process. H.R. and A.K.S. provided tools and expertise for the study. Supervision was provided by R.T. and G.P.N.

Competing interests

G.P.N. received research grants from Pfizer, Vaxart, Celgene and Juno Therapeutics during the time of and unrelated to this work. G.P.N. is a cofounder of Ionpath. G.P.N. is a cofounder of Akoya Biosciences, inventor on patent US9909167 and scientific advisory board member for Akoya Biosciences. A.K.S. reports compensation for consulting and/or scientific advisory board membership from Honeycomb Biotechnologies, Cellarity, Ochre Bio, Relation Therapeutics, Fogpharma, IntrECate Biotherapeutics, Bio-Rad Laboratories, Passkey Therapeutics, Senda Biosciences and Dahlia Biosciences unrelated to this work. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-025-02119-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02119-z.

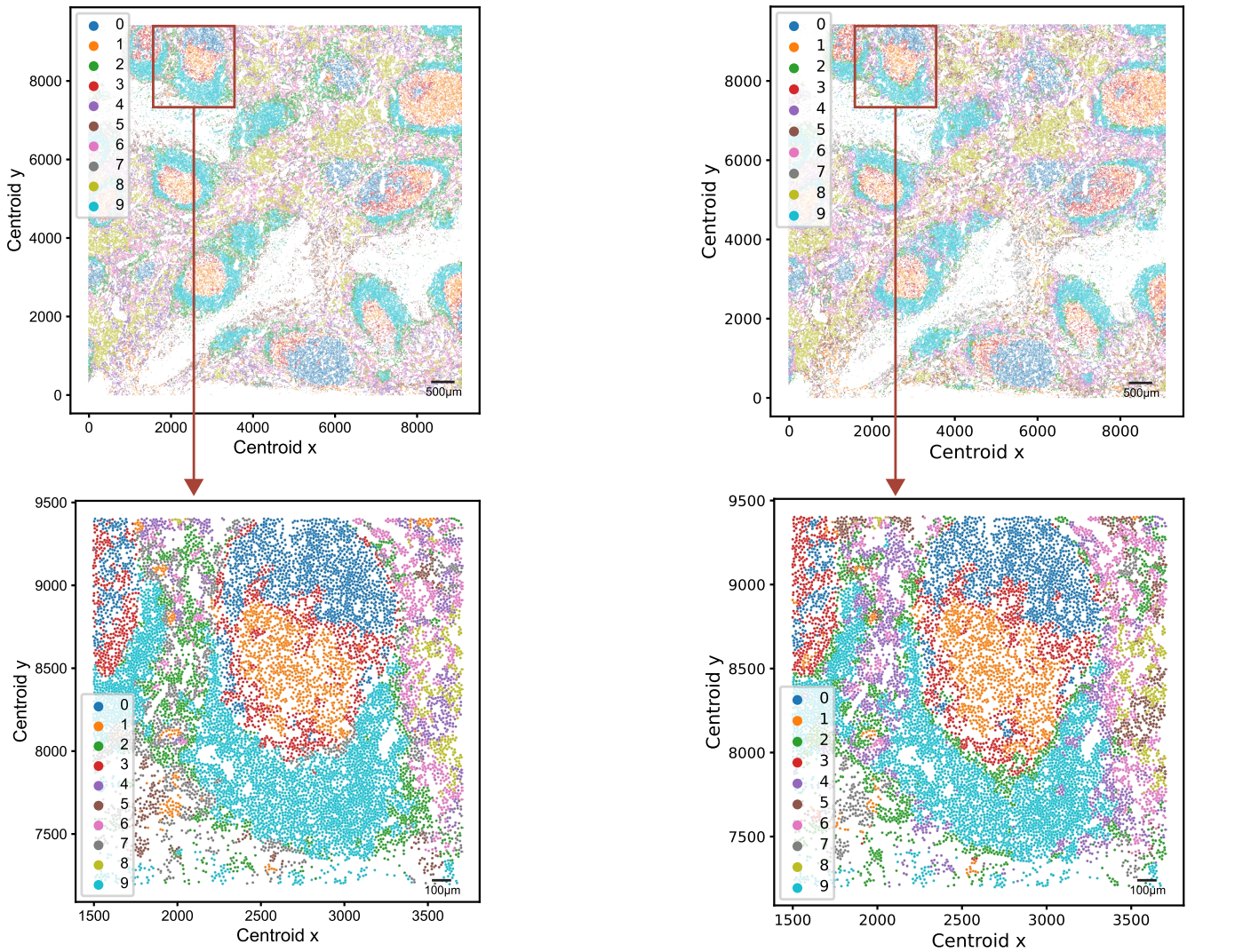
Correspondence and requests for materials should be addressed to Garry P. Nolan.

Peer review information *Nature Genetics* thanks Simon Cockell, Lambda Moses, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-025-02119-z

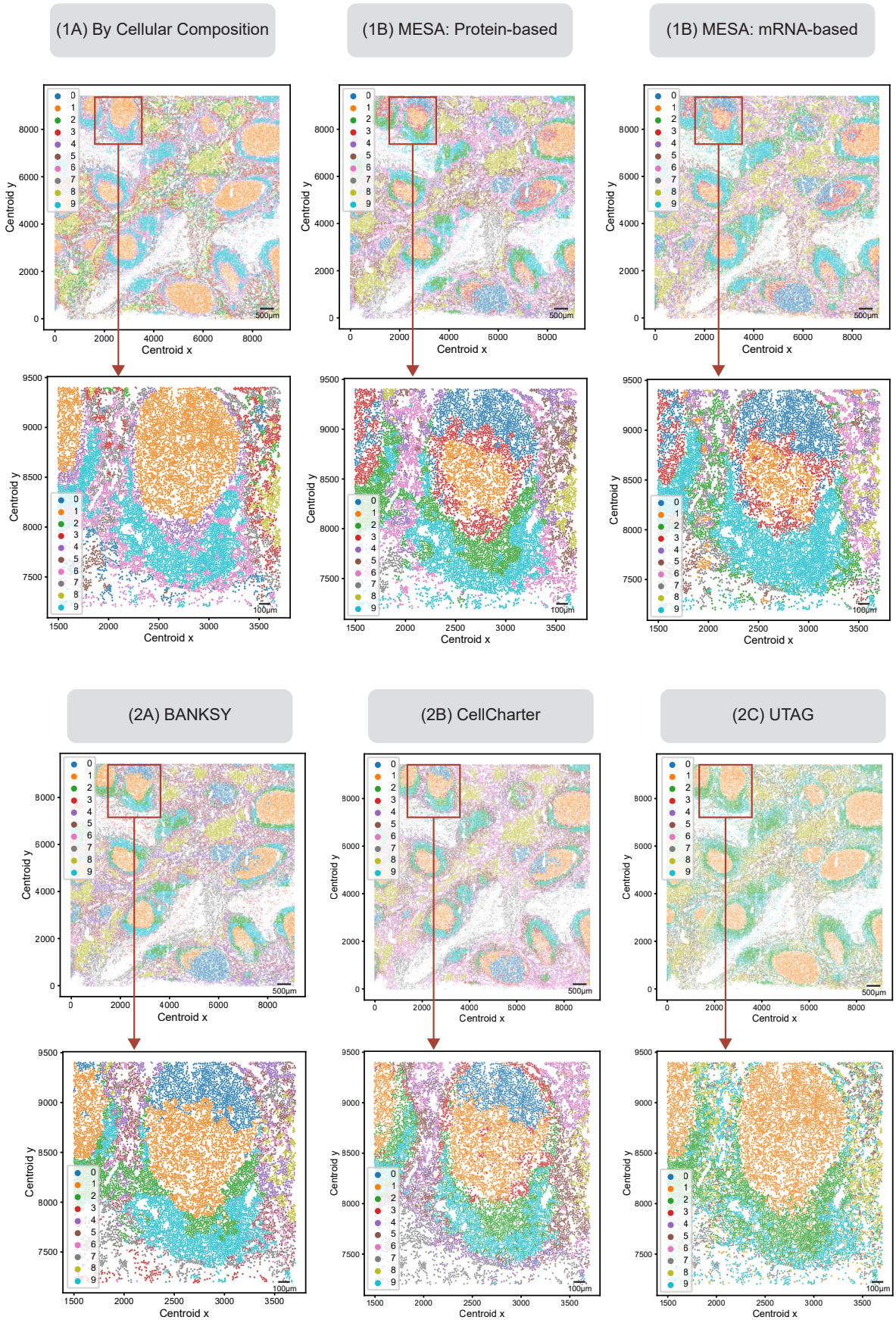
A. MESA with the tonsil scRNA-seq data from King et al., *Science Immunology*, 2021

B. MESA with the tonsil scRNA-seq data from Massoni-Badosa et al., *Immunity*, 2024



Extended Data Fig. 1 | MESA demonstrates robust neighborhood characterization results across distinct scRNA-seq datasets of tonsil tissues for multiomics integration. **(A)** mRNA-based neighborhood characterization results obtained from integrating the tonsil CODEX data with the scRNA-seq data from[40], as in the analysis shown in Fig. 2. The top panel shows the tonsil tissue with MESA-identified neighborhoods displayed in distinct colors. The bottom panel provides a magnified view of a representative germinal center region. **(B)** Neighborhood characterization results from integrating the tonsil CODEX data with an independent scRNA-seq data from[41]. While there are some minor differences between the two analyses, the consistency in key spatial patterns and structures demonstrates MESA's robustness across different scRNA-seq data sources.

https://doi.org/10.1038/s41588-025-02119-z

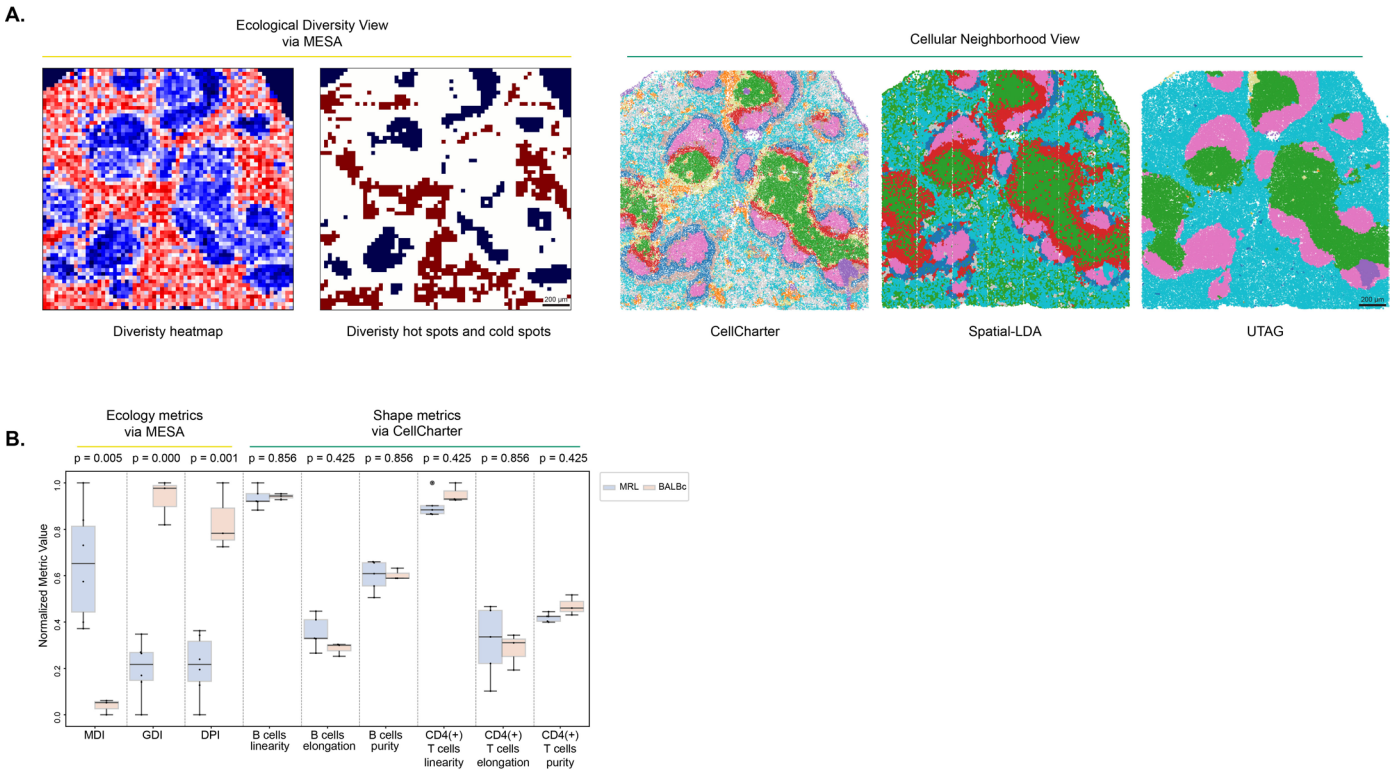


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41588-025-02119-z

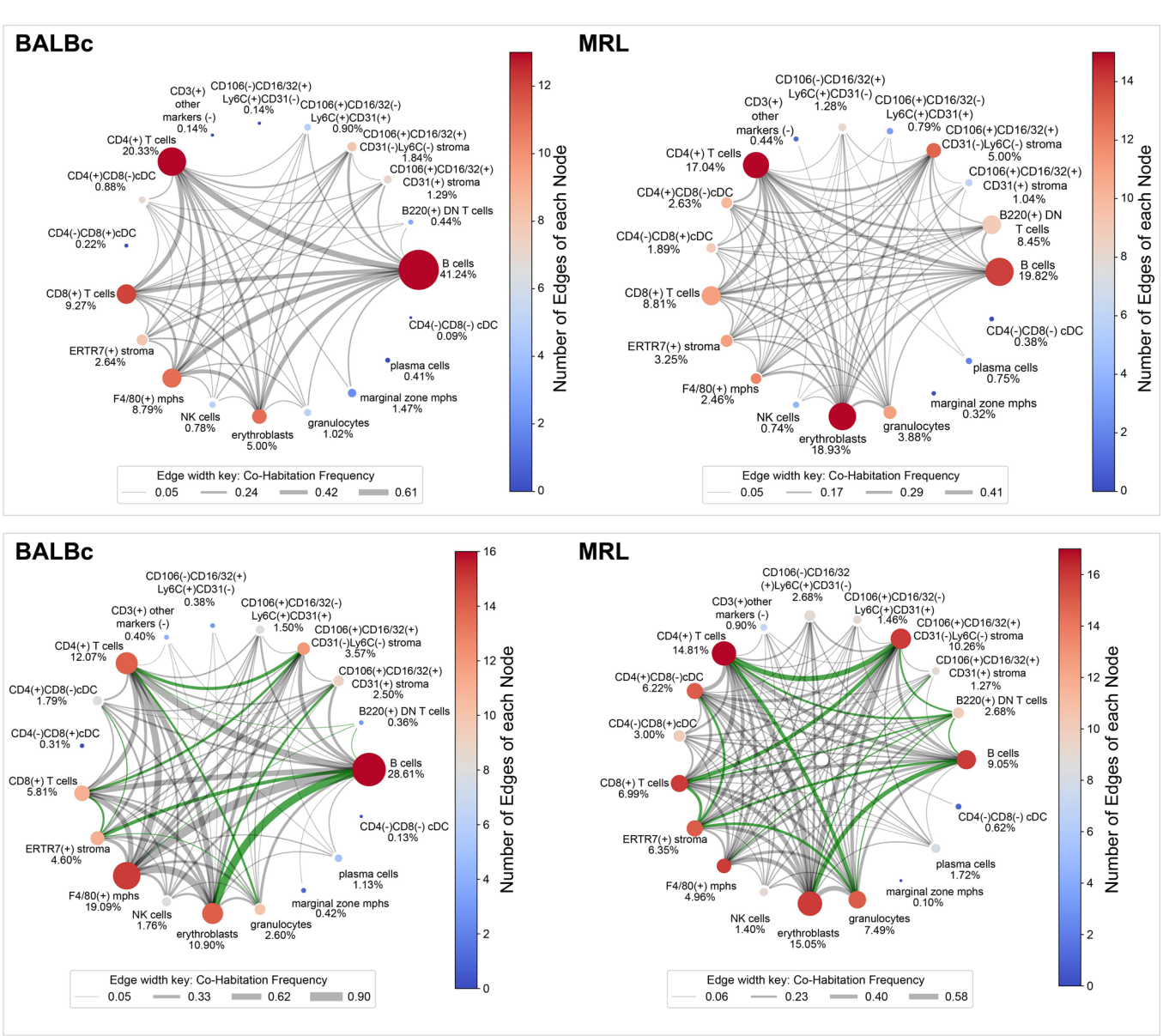
Extended Data Fig. 2 | Benchmarking results of neighborhood characterization methods in tonsil tissue. (1A) Cellular composition-based neighborhood characterization. **(1B)** MESA: Protein-based neighborhood characterization. **(1C)** MESA: mRNA-based neighborhood characterization. **(2A)** BANKSY neighborhood characterization[45]. **(2B)** CellCharter neighborhood characterization[16]. **(2C)** UTAG neighborhood characterization[15]. Each panel shows the full tonsil tissue sample (top) and a zoom-in view of a representative germinal center (bottom), with distinct colors denoting different neighborhoods. As compared to the approach based solely on cellular composition (1A), BANKSY (2A) and CellCharter (2B) showed finer-grained neighborhood characterization; however, MESA (1B and 1C) enables more granular neighborhood delineation. For example, in the germinal centers, BANKSY identified two distinct neighborhoods (labeled as 0 and 1); UTAG characterized a single neighborhood (labeled as 1); CellCharter produced results most similar to MESA, with three neighborhoods labeled as 0, 1, and 3. These comparative analyses highlight the advantage of incorporating multiomics information in niche identification. Methods leveraging the dynamic range of protein measurements as a proxy for cell states show enhanced sensitivity to spatially coregulated protein and mRNA expression patterns, enabling enhanced neighborhood identification.

https://doi.org/10.1038/s41588-025-02119-z



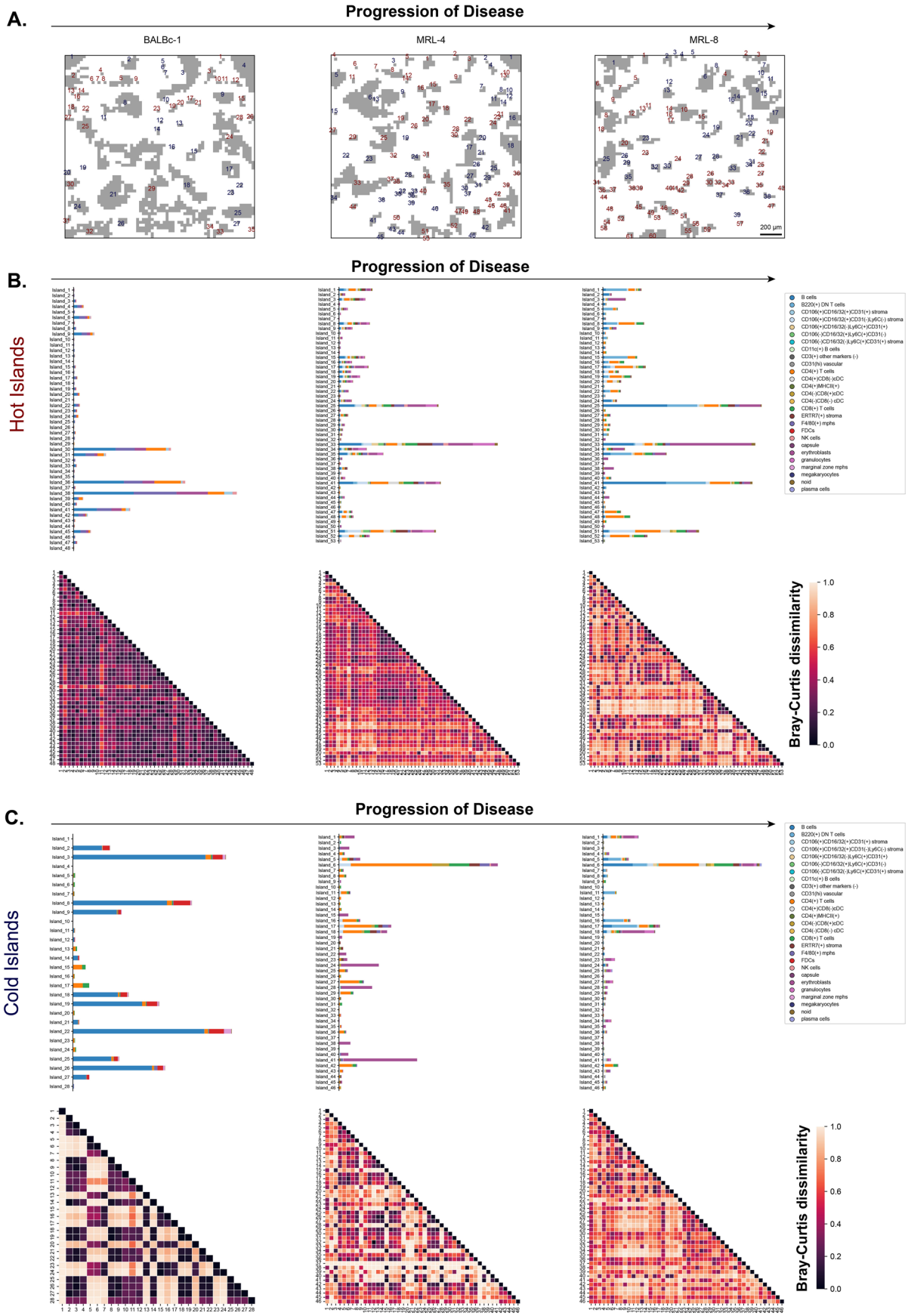
Extended Data Fig. 3 | MESA's ecological framework demonstrates enhanced performance in tissue state characterization compared to existing spatial analysis methods. (A) Comparison of MESA with other neighborhood identification methods in the mouse spleen CODEX data. Left panel: MESA-derived diversity heatmap showing spatial heterogeneity (red: high diversity; blue: low diversity) with the identified diversity hot spots (dark red) and cold spots (dark blue). Right panel: Cellular neighborhood/niche identified by CellCharter[16], Spatial-LDA[14] and UTAG[14] in the same tissue sample. Distinct colors denote different neighborhoods. MESA-identified hot spots and cold spots reveal diversity patterns not captured by conventional neighborhood/niche methodologies. **(B)** Comparative analysis of MESA's ecological metrics versus CellCharter's metrics in mouse spleen CODEX data of healthy ($n = 3$) and MRL/lpr ($n = 36$) samples. The CellCharter's metrics include a set of quantitative measures to characterize spatial patterns: (1) curl, which quantifies the degree of curvature or twisting in a shape; (2) elongation, measured as the proportion between the longest and shortest axes; (3) linearity, which evaluates how closely a shape follows a straight path; and (4) purity, assessing the homogeneity of cell types within a defined cluster region. Standard box plot metrics were used, with points representing individual tissue samples. Two-sided Welch's t-test is used to compute the P values, which have been adjusted using the BH procedure for FDR correction. The plot shows only valid CellCharter results for cell types and shape metrics (linearity, curl, elongation, and purity), excluding cases where default settings produced null values.

https://doi.org/10.1038/s41588-025-02119-z



Extended Data Fig. 4 | Visualization of cellular abundance and cohabitation patterns in healthy (BALBc) and MRL/lpr mouse spleen tissues. (A) Circos plots illustrate cellular abundance and cohabitation frequencies in whole tissue. Nodes represent distinct cell types and edges cohabitation relationships. Node size corresponds to cell-type abundance, while edge thickness denotes cohabitation strength. **(B)** Analysis focused on the identified diversity hot spots, with green edges highlighting cohabitation patterns unique to these regions. These cohabitation relationships, while not significantly different between healthy and MRL in whole-tissue analysis (two-sided Welch's t-test with BH correction for FDR control), emerged as distinctive features within hot spots.

https://doi.org/10.1038/s41588-025-02119-z

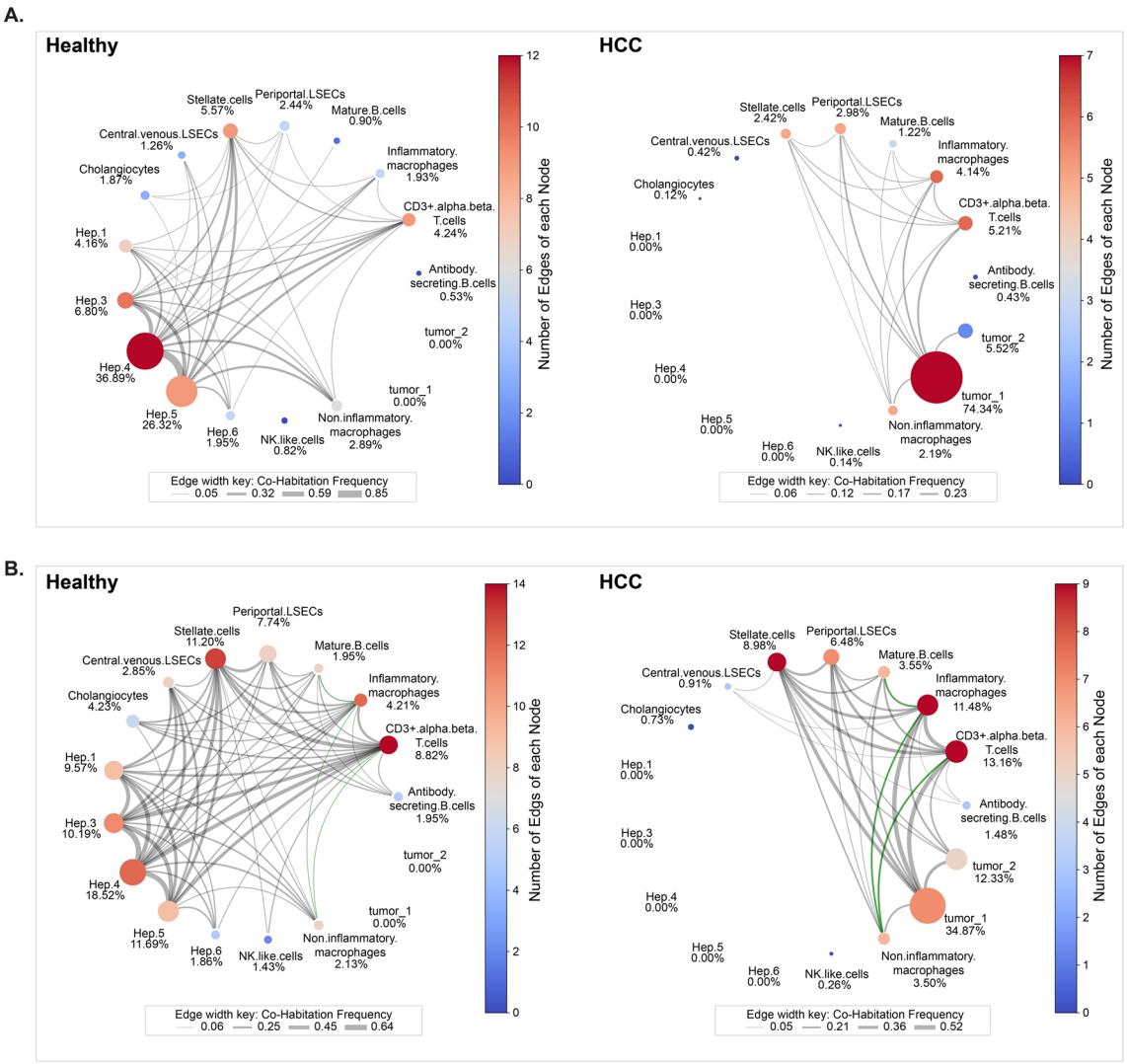


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41588-025-02119-z

Extended Data Fig. 5 | Spatial heterogeneity analysis of cellular composition in hot spots and cold spots during mouse spleen disease progression. **(A)** Spatial distribution of hot-spot (red) and cold-spot (blue) islands (connected components) in three representative samples spanning disease progression: healthy (BALBc-1), intermediate stage of disease (MRL-4), and late stage of disease (MRL-8). **(B)** Hot-spot island characterization. Top: Cell-type composition of individual hot-spot islands. Each row represents a non-contiguous hot-spot island, where colors denote different cell types and bar lengths indicate their counts. Bottom: Bray–Curtis dissimilarity[85] indices between hot-spot islands, quantifying compositional dissimilarity between islands, ranging from 0 (identical composition) to 1 (completely different composition). **(C)** Cold spot island characterization, following panel (B) format. Interestingly, hot-spot islands in healthy tissues (BALBc) show similar cellular compositions, reflected by relatively low Bray–Curtis dissimilarity indices. Disease progression correlates with increasing dissimilarity indices among hot-spot islands, indicating greater spatial heterogeneity. This analysis allows us to assess the degree of compositional variation across hot spots and cold spots within each sample, providing insights into systematic changes in tissue organization during disease progression.

https://doi.org/10.1038/s41588-025-02119-z



Extended Data Fig. 6 | Visualization of cellular abundance and cohabitation patterns in healthy and HCC liver tissues. (A) Circos plots illustrate cellular abundance and cohabitation frequencies in whole tissue. Nodes represent distinct cell types and edges cohabitation relationships. Node size corresponds to cell-type abundance, while edge thickness denotes cohabitation strength. **(B)** Analysis focused on the identified diversity hot spots, with green edges highlighting cohabitation patterns unique to these regions. These cohabitation relationships, while not significantly different between healthy and HCC in whole-tissue analysis (two-sided Welch's t-test with BH correction for FDR control), emerged as distinctive features within hot spots.

nature portfolio

Corresponding author(s): Garry P. Nolan

Last updated by author(s): 2024/12/16

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All data utilized in this study are obtained from previously published studies and have been processed in their data collection process, so no software tool was used for this purpose in our work.
Data analysis	All code used for data analysis in this study is available at: https://mesa-py.readthedocs.io/en/latest/, and on Github: https://github.com/Feanor007/MESA. Other software tools used include Python (version 3.11) and R (version 4.2) for general programming and analysis, FlowSOM (version 2.2.0) for niche characterization, scikit-learn (version 1.5.0) for clustering algorithms including K-means, hierarchical clustering, and BIRCH, Spatial-LDA (version 0.1.3), BANKSY (version 1.1.1), CellCharter (version 0.2.1) and UTAG (version 0.1.1) for spatial analysis, R package edgeR (version 3.36.0) for differential expression analysis, R package fgsea (version 1.20.0) for gene set enrichment analysis, scipy (version 1.11.2) for statistical tests and Bray-Curtis dissimilarity calculation, lifelines (version 0.27.7) for survival analysis, and SpatialDM (version 0.2.0) for cell-cell communication analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data utilized in this study are obtained from previously published studies and are all publicly accessible. The links are listed here:
CODEX tonsil data: https://onlinelibrary.wiley.com/doi/10.1002/eji.202048891;
scRNA-Seq tonsil data: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165860;
CODEX mouse spleen data: https://doi.org/10.1016/j.cell.2018.07.010;
CODEX CRC data: https://doi.org/10.1016/j.cell.2020.07.005;
scRNA-Seq CRC data: https://doi.org/10.1016/j.cell.2021.08.003;
CosMx liver data: https://doi.org/10.1038/s41587-022-01483-z;
scRNA-Seq liver data: https://doi.org/10.1016/j.jhep.2021.06.028;
CellChatDB database: https://doi.org/10.1038/s41467-021-21246-9.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This information is not available and not required in our manuscript, as it is not related to any analysis or investigation performed in this study.
Reporting on race, ethnicity, or other socially relevant groupings	This information is not available and not required in our manuscript, as it is not related to any analysis or investigation performed in this study.
Population characteristics	This information is not available and not required in our manuscript, as it is not related to any analysis or investigation performed in this study.
Recruitment	This information is not available and not required in our manuscript, as it is not related to any analysis or investigation performed in this study.
Ethics oversight	This information is not available and not required in our manuscript, as it is not related to any analysis or investigation performed in this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All data utilized in this manuscript are obtained from previously published studies, with datasets listed in the Data Availability section. Therefore, the sample sizes were predetermined by the original peer-reviewed studies. Our computational analysis utilized all available samples from these established datasets. For the liver CosMx dataset, with only one sample each for healthy and hepatocellular carcinoma tissues, we performed random subsampling to facilitate statistical comparisons. We generated 10 subsampled patches per condition, a number chosen to balance statistical power while preserving original tissue characteristics, ensuring that the cellular composition of each subsampled patch closely matched that of the original sample. The robustness of this approach was validated by confirming consistent patterns when analyzing the entire dataset.
Data exclusions	No data were excluded from the analyses. All data utilized in this manuscript are obtained from previously published studies. All analyses and comparisons were performed within tissue samples, using all available data points from the investigated tissue regions.

Replication	No replication was performed in the context of life science study design, as no experimental data was produced in our study. Replication was conducted in the context of computational comparisons between methods. All experimental findings can be reproduced using the same random seed. The computational analyses and their results are fully reproducible given the same input data and specified random seed.
Randomization	All data utilized in this manuscript are obtained from previously published studies. Randomization was not applicable as our study focused on computational analysis of these existing datasets. All the analyses were conducted consistently regardless of sample characteristics. Control of covariates was not applicable as we performed systematic computational analysis within each dataset independently, maintaining their original experimental designs. For the liver CosMx dataset, with only one sample each for healthy and diseased tissues, we performed random subsampling to facilitate statistical comparisons. Potential confounding factors were controlled by ensuring each subsampled patch maintained the original tissue characteristics and cellular composition of the original sample.
Blinding	All data utilized in this manuscript are obtained from previously published studies. Blinding was not relevant as our study focused on computational method development with automated analysis pipelines and objective metrics. All the analyses were performed systematically regardless of sample identity, so the process does not involve potential experimental bias.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	This information is not available and not required in our manuscript, as it is not related to any analysis or investigation performed in this study.
Novel plant genotypes	This information is not available and not required in our manuscript, as it is not related to any analysis or investigation performed in this study.
Authentication	This information is not available and not required in our manuscript, as it is not related to any analysis or investigation performed in this study.