nature genetics

Technical Report

https://doi.org/10.1038/s41588-026-02610-1

Accurate, scalable and cross-platform cell identification for high-resolution spatial transcriptomics

Received: 16 May 2024

Accepted: 22 April 2026

Published online: 20 May 2026

Check for updates

Dongqing Sun[1,2,3,6], Lele Zhang[4,6], Tong Han[1,2], Qiu Wu[1,2,3], Peng Zhang[5] ✉ & Chenfei Wang[1,2,3] ✉

Recent advances in spatial transcriptomics (ST) have brought unprecedented insights into cellular diversity and cell–cell interactions within their spatial context. High-resolution ST techniques, including barcoding-based and imaging-based platforms, have achieved remarkable subcellular resolution. However, precise cell segmentation remains a major challenge, hampering effective single-cell spatial analysis. Existing methods are often platform specific and lack scalability for datasets with large fields of view. Here we introduce Cellist, a new, multi-modal, cell-segmentation method that combines image and expression information, enabling comprehensive cell-level analyses. Applied to mouse brain Stereo-seq data, Cellist improves within-cell transcriptomic coherence compared to existing approaches. It further enhances spatial domain identification and cell-type annotation. Importantly, Cellist is compatible with various ST techniques including Seq-Scope, seqFISH+, STARmap and 10x Xenium, exhibiting robust performance and high computational efficiency across diverse ST platforms and biological systems. Finally, application to post-neoadjuvant immunotherapy, nonsmall cell lung-cancer samples reveals the spatial heterogeneity of tumor clones and identifies therapy response-related myeloid subtypes and structures. These findings highlight the potential of Cellist in enhancing the power of high-resolution ST techniques for characterizing intricate tissue architectures. Cellist is publicly available at https://github.com/wanglabtongji/Cellist.

Spatial transcriptomics (ST) has emerged as a transformative technology, revolutionizing our ability to measure gene expression while preserving crucial spatial information. Earlier platforms, such as 10x Visium[1] and Slide-seq[2], have substantially advanced our understanding of cellular diversity and interactions, uncovering spatial regulatory mechanisms in various biological systems, including the neural system[3–5], organ development[6,7] and the tumor microenvironment (TME)[8–10]. However, their limited resolution and restricted field of view (FOV) constrain broader applications. To overcome these limitations, recent ST technologies have been developed to achieve subcellular resolution. One major class involves high-plex RNA imaging with predesigned probes, such as MERFISH[11] and STARmap[12]. These techniques offer single-molecule resolution through in situ hybridization or in situ sequencing and are implemented in commercial platforms including Vizgen MERSCOPE and 10x Xenium[13]. Another class employs denser molecular identifiers to capture transcripts in an unbiased manner at subcellular levels, exemplified by Seq-scope[14] and Stereo-seq[15].

A full list of affiliations appears at the end of the paper. ✉e-mail: zhangpeng1121@tongji.edu.cn; 08chenfeiwang@tongji.edu.cn

https://doi.org/10.1038/s41588-026-02610-1

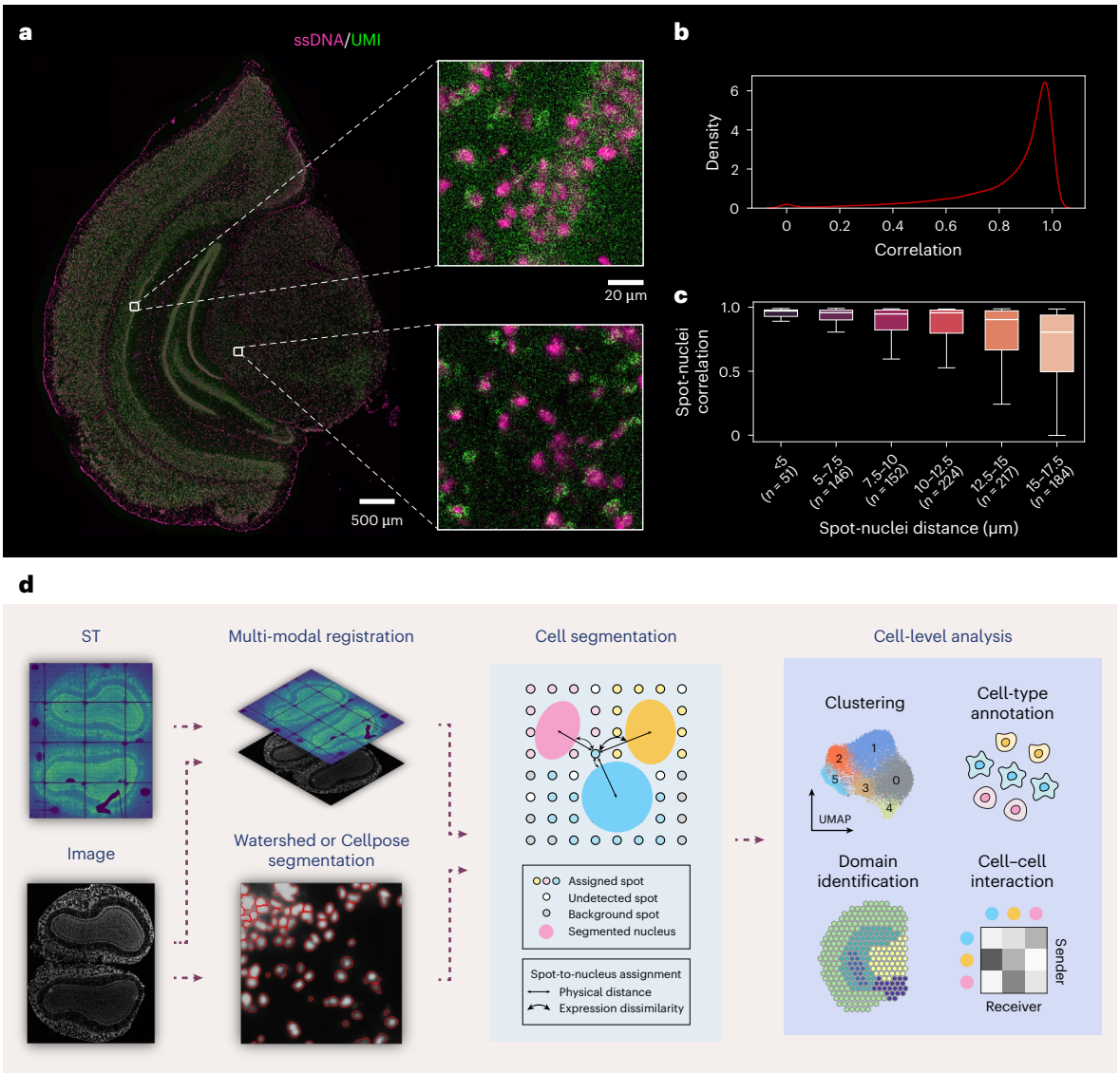


Fig. 1 | Characteristics of the Stereo-seq data and workflow of Cellist. **a**, Illustration showing the issue of transcript diffusion in the Stereo-seq mouse brain data. Stained ssDNA and sequenced unique molecule identifiers are depicted in magenta and green, respectively. **b**, Distribution of expression correlation between nucleus and cytoplasm in the Stereo-seq mouse brain data. This is visualized using a kernel density estimate, which provides a smoothed approximation of the histogram of correlation values. Nucleus and cytoplasm regions for each cell are defined using the expansion method. Transcripts within each region are aggregated to represent their expression and Pearson's correlation of expression between nucleus and cytoplasm was calculated in each cell. **c**, Changes in spot-nucleus correlation across distance bins in a representative cell from the Stereo-seq mouse brain dataset. The *x* axis represents binned distances between each outside spot and the nucleus center, whereas the *y* axis represents Pearson's correlation between each spot's augmented expression and the nuclear expression. The sample size (*n*), defined as the number of spatial spots within each distance bin, is indicated below each distance range on the *x* axis. Each data point represents one spatial spot. Box plots throughout the figure indicate the median (center line) and interquartile range (IQR) (25th–75th percentiles; box) and the whiskers extend to the most extreme values within 1.5× IQR. **d**, Illustration of the Cellist workflow. With spatial expression and a paired ssDNA-staining image as input, registration is performed to align information from the two modalities. Watershed or Cellpose segmentation is performed to identify all nuclei in the image, which serves as the potential cells. Each spot outside nuclei is assigned as part of either a cell or a background by combining the physical distance and expression dissimilarity between the spot and nearby nuclei. Last, cell-level analyses can be performed to understand the spatial organization of cells, including cell clustering, cell-type annotation, spatial domain identification and cellcell interaction analysis.

Despite their advances, high-resolution techniques pose substantial computational challenges. Each measured spot or pixel captures only a few transcripts, providing limited information about the sources of cell types. Segmentation-free approaches address this by assigning each pixel to a specific cell type using neighborhood information[16–18]. However, these methods fail to investigate within-cell-type heterogeneity and intercellular interactions. Consequently, considerable efforts have focused on cell segmentation in imaging-based methods to enable single-cell-level analysis. Early methods relied on auxiliary staining, such as DAPI[12,19] or poly(A)[20], to delineate cell boundaries without utilizing expression information. More recent approaches, such as pciSeq[21], Baysor[17] and Sparcle[22], treat staining-based cell segmentation as initial cell approximations and refine cell segmentation by modeling the spatial distribution of transcripts. These methods have shown promise in imaging-based platforms such as MERFISH or STARmap.

However, segmentation methods designed for imaging-based ST often fail to generalize to barcoding-based techniques. Key differences between these modalities include spatial resolution, transcript diffusion and gene panel size (Supplementary Fig. 1). Barcoding-based technologies like Stereo-seq profile the whole transcriptome, but they are more susceptible to excessive transcript diffusion, leading to blurred cellular boundaries (Fig. 1a). Moreover, barcoding-based platforms

depend on predefined spatial grids rather than optical imaging, introducing distinct noise and spatial distributions. These differences challenge the foundational assumptions of many existing methods[23]. For example, Baysor assumes ellipsoidal cell shapes via Gaussian priors—an assumption often violated in barcoding-based data, where cell shapes are irregular and transcript distributions are noisy. Furthermore, imaging-based ST methods typically operate on smaller FOVs and limited gene panels. When scaled to large barcoding-based datasets with millions of spots and tens of thousands of genes, they often suffer from prohibitive computational burden (Extended Data Fig. 1 and Supplementary Table 1).

Several methods attempt to bridge these gaps. SCS[24] and UCS[23] combine imaging and sequencing data using deep learning. However, as these methods emphasize precise morphological boundaries, they may inadvertently include extracellular or neighboring cell transcripts, leading to mixed or biologically misleading expression profiles. Moreover, both SCS and UCS impose substantial memory and runtime requirements, particularly for datasets with large FOVs (Extended Data Fig. 1 and Supplementary Table 1). StereoCell[25] scales to large barcoding-based ST techniques but is restricted to that modality. Currently, there is a lack of scalable and crossplatform solutions that can unify segmentation across both imaging-based and barcoding-based, high-resolution ST datasets.

Here we present Cellist, a new multi-modal method that integrates image and expression information to enhance cell segmentation in high-resolution spatial transcriptomics. Unlike previous methods that emphasize precise morphological boundaries, Cellist is designed to preserve transcriptomic integrity within each segmented cell. Using mouse brain Stereo-seq data, we demonstrated that Cellist improves within-cell expression consistency, thereby enhancing downstream single-cell analyses such as cell-type annotation. Moreover, Cellist generalizes beyond Stereo-seq, flexibly adapting to other high-resolution, barcoding-based or imaging-based ST techniques, including Seq-Scope, seqFISH+, STARmap and Xenium. This versatility highlights its scalability and applicability across diverse biological systems and experimental platforms. Finally, we applied Cellist to post-neoadjuvant immunotherapy, nonsmall cell lung-cancer (NSCLC) samples. Our findings provide valuable insights into the intricate spatial structures and cell types within the TME, offering a deeper understanding of the dynamics and therapy response-related features.

Results

Cellist is a multi-modal cell segmentation method for high-resolution spatial transcriptomics

Cell segmentation in traditional images typically classifies each pixel as part of either a cell or a background based on its intensity and local neighboring features. Analogously, cell segmentation in high-resolution ST aims to assign each captured molecule or spot to a specific cell relying on the spatial distribution of transcripts. However, compared to images, ST data generally exhibit a lower signal-to-noise ratio, making segmentation based solely on spatial expression information highly challenging. To address this, we proposed combining both image and expression information to achieve cell-level transcript quantification.

As a proof of concept, we designed a Watershed-based expansion strategy (Supplementary Fig. 2), which revealed a high expression correlation between the nucleus and the cytoplasm (Fig. 1b). Motivated by this finding, we assumed that the nucleus and cytoplasm share similar expression profiles in high-resolution ST data. We validated the robustness of this assumption across five representative ST platforms, as well as synthetic datasets with varying degrees of subcellular RNA localization (Supplementary Note 1). Although some genes exhibit localized enrichment patterns, most tend to be broadly distributed throughout the cell. To further characterize the spatial continuity of expression, we calculated the expression correlation between each nucleus and its surrounding spots. As expected, the nucleus–spot correlation is high when the spot is close to the nucleus and gradually decreases with increasing distance (Fig. 1c and Supplementary Fig. 3). This observation inspired us to incorporate expression similarity to guide segmentation.

Building on our previous findings, we introduced Cellist (cell identification in high-resolution spatial transcriptomics), a multi-modal

Fig. 2 | Application of Cellist and downstream analyses in the Stereo-seq mouse brain dataset. a, Number of cells generated by each segmentation method. **b**, Distribution of the number of detected genes per segmented cell generated by different segmentation methods. The sample size, defined as the number of segmented cells for each method, is shown in **a**. **c**, Spatially agnostic random correlation as a measure of within-cell expression consistency. Top: schematic illustrating random partitioning of all transcripts within a cell into two subsets, ignoring spatial position. Pearson's correlation coefficient between the pseudo-bulk expression profiles of the two subsets was computed for each cell. Bottom: box plots showing the distribution of correlation coefficients across segmented cells generated by different methods. Only cells containing >100 spatial spots were included in the analysis. Each data point represents one segmented cell. Statistical significance between methods was assessed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons (Expansion, $n = 62{,}572$; StereoCell, $n = 66{,}367$; SCS, $n = 50{,}218$; UCS, $n = 69{,}803$; and Cellist, $n = 53{,}088$). *P* values smaller than the smallest positive number that can be represented in the double-precision floating-point format (that is, $P < 2.23 \times 10^{-308}$) are reported as 0 due to computational precision limits. **d**, Directional random correlation for evaluating expression consistency across spatial halves of a cell. Top: schematic illustrating bisection of each cell along a randomly oriented line passing through the cell centroid. Correlation analysis is performed as in **c** to assess expression consistency between the two halves. Bottom: box plots showing the distribution of correlation coefficients across segmented cells generated by different methods. Statistical significance between methods was assessed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons. Cell numbers are the same as in **c**. **e**, Top: schematic of the variance-based purity score calculation. The purity score was computed as the ratio of the normalized variance of gene expression across spots within each predicted cell to that of its surrounding neighborhood. Bottom: cross-correlation strategy comparing segmentation concordance between methods. For any two segmentations from methods A and B, the intersection region and the difference regions are computed. Pearson's correlation between each difference region and the intersection region is calculated. **f**, Evaluation of intracellular expression uniformity using the metric of purity score across segmentation methods. Analysis was restricted to 29,780 nuclei shared across all methods. Statistical significance between each method and Cellist was assessed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons. **g**, Evaluation of the agreement between Cellist segmentation and other methods using the metric of cross-correlation, with Cellist as the target segmentation. Statistical significance between methods was assessed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons (Expansion versus Cellist, $n = 55{,}462$; StereoCell versus Cellist, $n = 49{,}658$; SCS versus Cellist, $n = 38{,}109$; and UCS versus Cellist, $n = 49{,}475$). **h**, Cell-type annotation using SingleR based on cell segmentation from Cellist. More details about the cell-type abbreviations are listed in Supplementary Table 2. **i**, Consistency between Stereo-seq and reference scRNA-seq. For each cell type identified in Stereo-seq based on each segmentation method, Pearson's correlation of averaged gene expression was calculated between Stereo-seq and scRNA-seq. Each dot represents one annotated cell type from **h**. Statistical significance between methods was assessed using paired, one-sided Wilcoxon's signed-rank tests without adjustment for multiple comparisons ($n = 34$ cell types). **j**, Marker specificity of annotated cell types, measured by the $\log_2$(FC) of the top DE genes in each cell type versus all other cell types. Only common genes shared across methods are used for fair comparison. Higher $\log_2$(FC) indicates more specific expression and thus more accurate segmentation. Each dot represents one DE gene in a cell type from **h**. The statistical significance was assessed using paired, one-sided Wilcoxon's signed-rank tests without adjustment for multiple comparisons ($n = 128$ genes). Box plots throughout the figure indicate the median (center line) and IQR (25th–75th percentiles; box) and the whiskers extend to the most extreme values within 1.5× the IQR.

https://doi.org/10.1038/s41588-026-02610-1

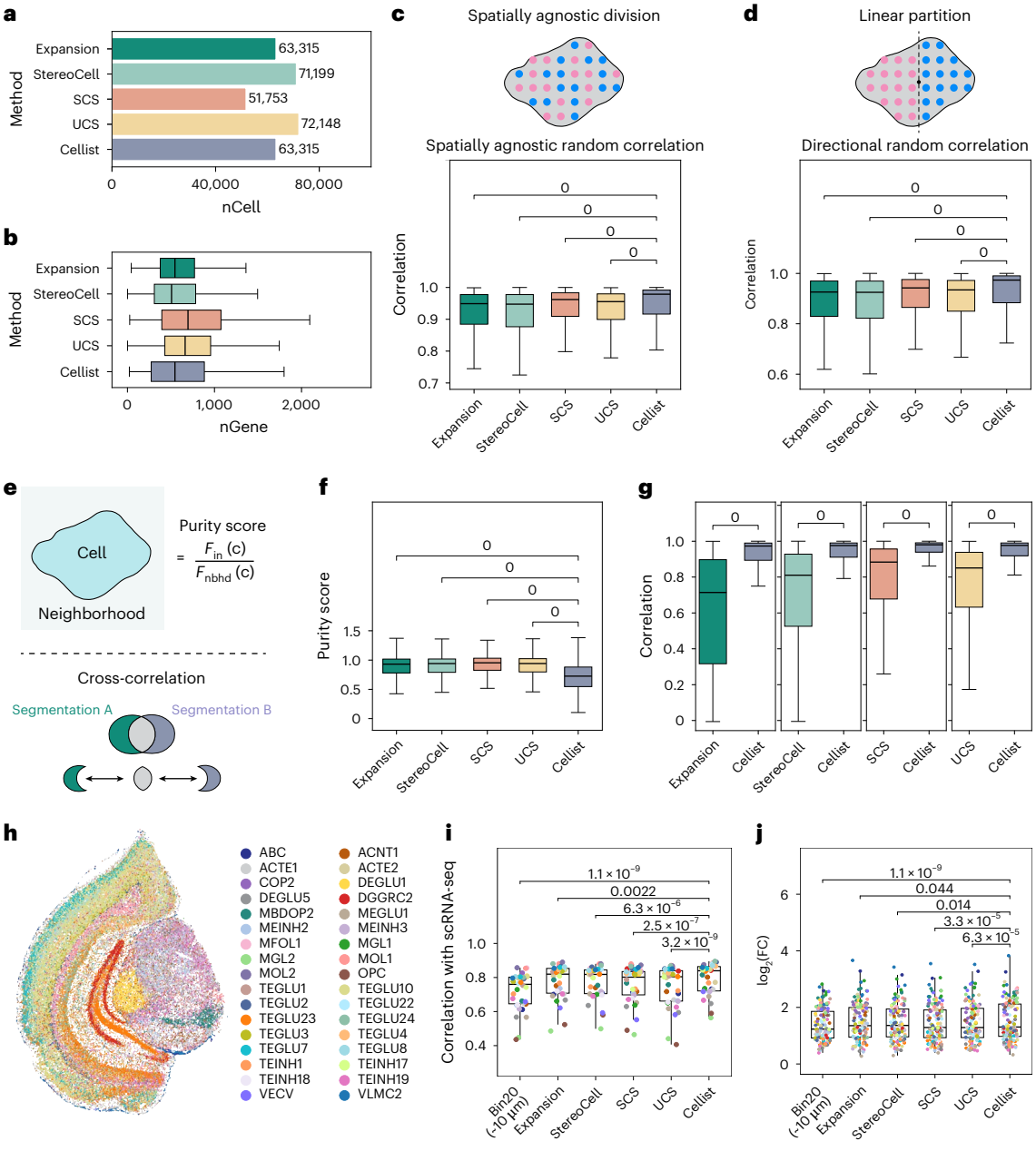
cell-segmentation method (Fig. 1d and Supplementary Notes 2 and 3). The Cellist workflow begins with image-based cell or nucleus segmentation, supporting both the Watershed algorithm[26] and Cellpose[27,28]. Subsequently, spatial transcriptomic data are then aligned to staining images through multi-modal registration to define initial cellular units. Spots outside these regions are assigned to cells via a probabilistic strategy that minimizes both expression dissimilarity and physical distance. This extends cells to include transcripts overlooked by image-based segmentation. The resulting single-cell spatial transcriptomic profiles facilitate diverse downstream cell-level tasks, including cell clustering, cell-type annotation, spatial domain identification and cell–cell interaction analysis.

Cell segmentation by Cellist achieves higher intracellular expression consistency in the Stereo-seq mouse brain data

We applied Cellist to a Stereo-seq mouse brain dataset with single-strand DNA (ssDNA) nucleus staining[15]. To evaluate its performance, we compared Cellist with two barcoding-specific methods (SCS[24] and StereoCell[25]), one unified method (UCS[23]) and the nucleus-based Expansion baseline (Supplementary Fig. 2).

We first examined basic cell-level characteristics, including the number of identified cells and the counts of covered spots and genes per cell (Fig. 2a,b and Supplementary Fig. 4). Remarkably, all other methods identified a higher number of cells compared to SCS (Fig. 2a), resulting in relatively smaller gene and spot coverage per cell (Fig. 2b).

To assess segmentation accuracy without ground truth, we analyzed within-cell expression purity. Accurate segmentation maintains transcriptomic coherence within cellular regions, whereas mis-segmentation compromises this by mixing signals from neighboring cells. To quantify this, we introduced random correlation, which calculates Pearson's correlation between pseudo-bulk profiles of randomly split transcript subsets (Fig. 2c), serving as a spatially unbiased proxy for consistency. A directional variant splitting cells along a random centroidal axis was also implemented (Fig. 2d). Across both variants, Cellist consistently achieved higher correlations than alternative methods, suggesting superior internal purity.



To complement correlation-based metrics, we further introduced a variance-based purity score (Fig. 2e, top), comparing normalized variance of gene expression across spots within each predicted cell to its surrounding neighborhood. A lower score indicates greater internal consistency relative to the local environment. Cellist exhibited significantly lower purity scores across all methods (Fig. 2f), reflecting minimal transcript contamination.

In addition to these internal consistency metrics, we employed a widely used strategy[17,24] to facilitate cross-method comparison. Specifically, we evaluated segmentation agreement by computing gene expression correlations between the overlapping region of matched cell pairs and their nonoverlapping portions (Fig. 2e, bottom). The underlying premise is that, if segmentation preserves transcriptomic coherence, the overlap should resemble the rest of the cell. In bidirectional assessments (using Cellist as both target and source), Cellist consistently demonstrated higher correlations than alternative methods (Fig. 2g and Supplementary Fig. 5), indicating more accurate segmentation.

Collectively, Cellist outperforms existing approaches across multiple complementary evaluations, including random and directional correlation, variance-based purity score and cross-method overlap analysis. Notably, Cellist achieves the highest within-cell transcriptomic consistency compared to the baseline Expansion approach and other existing methods.

Cellist enhances cell-type annotation and transcriptomic fidelity in Stereo-seq mouse brain data

Accurate cell segmentation is critical for resolving tissue composition. We investigated the potential of fine-grained cell segmentation to facilitate cell-type annotation. Given the diversity of cell types in the mouse brain, we integrated information from a well-annotated single-cell RNA sequencing (scRNA-seq) dataset[29] using SingleR[30] (Fig. 2h).

To quantify annotation accuracy, we calculated Pearson's correlation of average log(normalized gene expression) for each cell type between Stereo-seq and scRNA-seq. Notably, Cellist displayed a high level of concordance with the corresponding scRNA-seq data across various cell types (Supplementary Fig. 6). We benchmarked this against other methods and a coarse-grained 'bin 20' (~10 µm) baseline, which roughly matches the size of a typical mammalian cell (Supplementary Fig. 7). This binning approach offers comparable spatial resolution but lacks explicit cell boundaries, resulting in transcript mixtures from neighboring cells. Consequently, bin 20 serves as a nonsegmented baseline for evaluating the added value of true segmentation in recovering single-cell profiles. Although all segmentation methods exhibited improved correlations compared to bin 20, Cellist consistently outperformed all alternatives, achieving the highest correlation across nearly all cell types (Fig. 2i). These results suggest that Cellist more effectively separates mixed signals and preserves cell-type-specific gene expression, even in dense regions.

To further validate biological specificity, we performed differential expression (DE) analysis, computing $\log_2$(fold-changes) ($\log_2$(FC)) of marker genes for each cell type against all others. Cellist yielded higher $\log_2$(FC) values than other methods, indicating improved cell-type separation and clearer transcriptomic identities (Fig. 2j). In summary, these results demonstrate that Cellist segmentation not only enhances cell-type annotation but also improves biological interpretability by recovering more accurate cell-type-specific gene expression.

Cellist is scalable to different techniques and tissues

To demonstrate the scalability of Cellist to other platforms, we applied it to a mouse liver Seq-scope dataset[14] with matched hematoxylin and eosin (H&E) staining (Fig. 3a and Supplementary Fig. 8). Across four tiles, Cellist consistently demonstrated the highest intracellular expression consistency and improved separation of cell types among all methods (Fig. 3b,c and Extended Data Fig. 2).

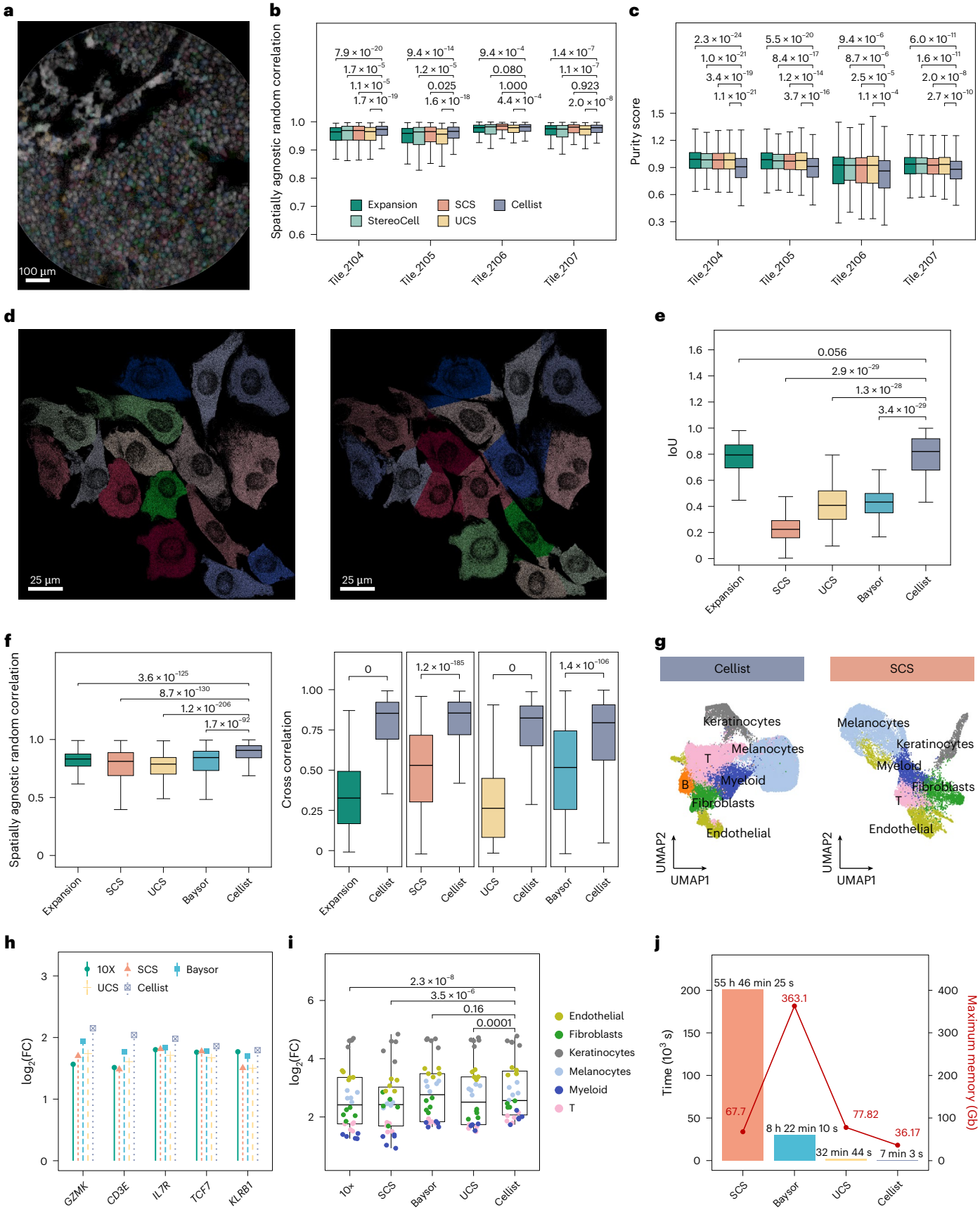
We further extended Cellist to imaging-based approaches, such as seqFISH+[19] and STARmap[12], by aggregating individual transcripts into fixed-size grids. For benchmarking, we replaced the barcode-specific StereoCell[25] with Baysor[17], a method tailored for imaging-based ST.

In the NIH/3T3 seqFISH+ dataset, using manual labels as ground truth (Fig. 3d), Cellist demonstrated comparable intersection over union (IoU) with Expansion and was significantly higher than other state-of-the-art methods, including Baysor, SCS and UCS (Fig. 3e and Supplementary Fig. 9). Similarly, in the mouse cortex STARmap dataset, Cellist outperformed these methods, exhibiting higher within-cell expression purity and higher marker specificity (Fig. 3f and Extended Data Fig. 3c–g).

To evaluate the performance on large FOV datasets, we applied Cellist to a human melanoma dataset profiled by 10x Xenium (Extended Data Fig. 4a). We compared it with the segmentation generated by 10x Xenium Ranger (Extended Data Fig. 4a), which is similar to the Expansion method, but employs a neural network for nucleus segmentation

Fig. 3 | Application of Cellist on different high-resolution ST platforms and tissues. a, Cellist cell segmentation results for tile 2,104 in the Seq-Scope mouse liver dataset. The background is the grayscaled H&E image and the foreground dots depict measured ST spots, colored by assigned cell labels. **b**, Intracellular, spatially agnostic, random correlation across segmentation methods in four tiles from the Seq-Scope mouse liver dataset. Only cells containing >100 spatial spots were included in the analysis. Statistical significance between each method and Cellist in each tile was assessed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons. Exact sample sizes for each comparison in each tile are provided in Source data. **c**, Within-cell purity score across segmentation methods in four tiles from the Seq-Scope mouse liver dataset. Analysis was restricted to nuclei shared among all methods in each tile. Statistical significance between each method and Cellist was assessed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons (tile 2,104, $n = 888$; tile 2,105, $n = 865$; tile 2,106, $n = 448$; and tile 2,107, $n = 709$). **d**, Manual cell segmentation from the original study (left) and cell segmentation generated by Cellist (right) in the FOV 0 of the seqFISH+ NIH/3T3 dataset. **e**, Consistency of different segmentation results with manual labels in cells from all seven FOVs of the seqFISH+ NIH/3T3 dataset. For each segmented cell from each method, the IoU was calculated between it and the manual segmentation label. Statistical significance between each method and Cellist was assessed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons (Expansion, $n = 103$; SCS, $n = 71$; UCS, $n = 102$; Baysor, $n = 93$; and Cellist, $n = 103$). **f**, Evaluation of the intracellular expression consistency in the STARmap mouse primary visual cortex dataset. Left: spatially agnostic random correlation across methods. Analysis was restricted to cells with >50 detected genes (Expansion, $n = 1{,}550$; SCS, $n = 1{,}529$; UCS, $n = 1{,}539$; Baysor, $n = 1{,}572$; and Cellist, $n = 1{,}544$). Right: cross-correlation with Cellist as the target segmentation (Expansion versus Cellist, $n = 1{,}546$; SCS versus Cellist, $n = 1{,}200$; UCS versus Cellist, $n = 1{,}547$; and Baysor versus Cellist, $n = 1{,}413$). Statistical significance between methods was assessed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons. **g**, Uniform Manifold Approximation and Projection (UMAP) plots showing the cell-type annotation based on Cellist (left) or SCS (right) segmentation in the 10x Xenium human melanoma dataset. **h**, Lollipop plots showing the $\log_2$(FC) of top marker genes in T cells versus all other cell types in the 10x Xenium human melanoma dataset. **i**, Marker specificity in annotated cell types in the 10x Xenium human melanoma (10x) dataset, measured by $\log_2$(FC) of the top DE genes in each cell type versus all other cell types. Each dot represents one DE gene in a cell type. The statistical significance between methods was assessed using paired, one-sided Wilcoxon's signed-rank tests without adjustment for multiple comparisons ($n = 30$ genes). **j**, Computational efficiency of Cellist in the 10x Xenium human melanoma dataset. The bar plot shows the runtime of different methods and the red line the memory usage. Box plots throughout the figure indicate the median (center line) and IQR (25th–75th percentiles; box) and the whiskers extend to the most extreme values within 1.5× the IQR.

https://doi.org/10.1038/s41588-026-02610-1



instead of Watershed. Although Ranger-derived nuclei were used as priors for all methods, SCS identified notably fewer cells (~20% of other methods) (Extended Data Fig. 4b). Consequently, the corresponding gene coverages were incomparable (Extended Data Fig. 4c), rendering correlation-based metrics unsuitable (Extended Data Fig. 4d,e). Instead, we evaluated segmentation performance based on downstream TME characterization, employing unsupervised clustering and marker-based annotation to discern cell-type compositions (Fig. 3g–i and Extended Data Fig. 5). Among all methods, only SCS failed to identify B cells or distinguish between basal and differentiated keratinocytes (Fig. 3g and Extended Data Fig. 5a). In DE analysis, Cellist yielded higher fold-changes compared to other methods, indicating more accurate cell-level expression estimates (Fig. 3h,i and Extended Data Fig. 5b). In addition, Cellist demonstrated superior computational efficiency, requiring less processing time and memory than SCS, UCS and Baysor in this large dataset (Fig. 3j, Extended Data Fig. 1 and Supplementary Table 1).

Overall, Cellist is versatile in handling both barcoding-based and imaging-based ST data across various tissue types. Its efficient runtime and memory usage make it scalable for large-scale datasets.

Cellist facilitates cell-type identification in spatial profiling of post-treatment NSCLCs

High-resolution spatial transcriptomics has emerged as a powerful tool in cancer research, allowing for the dissection of the spatial TME[31,32]. However, accurately segmenting cells in the complex TME poses major challenges. In the previous section, we demonstrated the robustness of Cellist in the 10x Xenium human melanoma dataset. We next examined how Cellist advanced our understanding of the spatial organization of cells within the TME after therapies.

We generated Stereo-seq data from six patients with NSCLC, comprising five cases of lung squamous cell carcinoma (LUSC) and one case of lung adenocarcinoma (LUAD) after neoadjuvant immune checking blockade (ICB) chemotherapy (Fig. 4a and Supplementary Table 3). Among these patients, two were classified as major pathological responders (MPRs, with ≤10% residual viable tumor cells in resected tumor specimens) and four were classified as nonmajor pathological responders (NMPRs, with >10% residual viable tumor cells) based on pathological assessment. Matched scRNA-seq profiles were generated to provide additional information and validate the findings.

After segmenting cells with Cellist, we conducted unsupervised cell clustering and marker-based cell-type annotation, revealing a diverse cellular composition within the TME. These encompassed epithelial cells, endothelial cells and fibroblasts, as well as various immune cells such as lymphocytes, plasma cells and myeloid cells (Fig. 4b and Supplementary Fig. 10). Benchmarking against other segmentation methods showed that Cellist yielded consistently higher fold-changes of marker genes between clusters, indicating more accurate single-cell expression estimates (Fig. 4c and Extended Data Fig. 6). In summary, Cellist markedly enhanced the accuracy of cell-type identification in NSCLC tumor sections at the single-cell and spatial levels.

Cellist enables the detection of spatially localized tumor clones in patients with NSCLC post-neoadjuvant ICB chemotherapy

Distinguishing malignant epithelial cells in LUAD is challenging due to the lack of clear markers[33]. To address this, we leveraged genomic signals inferred from single-cell transcriptomes using CopyKAT to estimate chromosomal copy number variations (CNVs). Based on hierarchical clustering of the inferred CNV profiles, we classified clusters exhibiting extensive chromosomal alterations as malignant epithelial cells, whereas the remaining diploid-like clusters were considered normal (Fig. 4d).

We validated the robustness of CNV inference from Stereo-seq data by comparing with matched scRNA-seq. The CNV profiles of predicted malignant cells showed strong concordance between the two modalities (Pearson's correlation: 0.78), indicating that Cellist-derived gene expression can support reliable genomic inference (Fig. 4d). Despite overall high agreement, Stereo-seq detected a gain in chromosome 1q absent in scRNA-seq. We attributed this discrepancy to sampling differences: Stereo-seq comprises more cells with broader spatial coverage, likely enabling the detection of rare or spatially localized subclones missed by scRNA-seq. This observation underscores the advantage of spatial transcriptomics in resolving clonal heterogeneity and low-frequency events within the tumor ecosystem.

By integrating the expression and CNV profiles, we localized malignant cells within the post-treatment TME of LUAD (Fig. 4e,f). We applied the same CNV-based approach across all six NSCLCs (Supplementary Figs. 11 and 12). To gain further insights into intratumoral heterogeneity, we performed hierarchical clustering on the inferred CNV profiles of malignant epithelial cells, revealing two genomically distinct tumor subclones: clone A with genomic gains in chromosome 6 and clone B with genomic gains in chromosome 1 (Fig. 4g). Spatially, clone A was dispersed throughout the entire slide, whereas clone B showed focal enrichment (Fig. 4h).

To explore their functional differences, we performed DE analysis between the two clones, followed by Hallmark pathway enrichment. Clone B exhibited significant upregulation of pathways related to adipogenesis and epithelial–mesenchymal transition (Fig. 4i), both of which have been implicated in tumor progression, immune evasion and therapy resistance in lung cancer and other tumor types[34–37].

Among the upregulated genes, *ALDH2*, encoding a key enzyme involved in fatty acid metabolism, was notably elevated in clone B (Supplementary Fig. 13). Importantly, members of the aldehyde dehydrogenase (ALDH) family, including *ALDH2*, are widely recognized as markers of cancer stem-like cells (CSCs) and have been associated with tumorigenicity and resistance to therapy[38]. Based on a curated CSC signature gene set[39], clone B consistently displayed higher CSC scores than clone A (Fig. 4j), supporting the notion that it may harbor more stem-like properties. Taken together, the enrichment of epithelial–mesenchymal transition, metabolic and stemness programs suggests that clone B represents a more aggressive or treatment-refractory phenotype.

Overall, our application of Cellist to NSCLC samples demonstrates its capability of discerning malignant cells and reveals the spatial heterogeneity of tumor subclones with distinct hallmark characteristics, highlighting Cellist's utility in resolving challenging clinical samples.

Cellist unveils conserved cellular spatial distribution in the TME of NSCLC

Understanding the spatial architecture of the TME is crucial for elucidating tumor progression, immune infiltration and therapeutic response. To dissect the spatial complexity of the post-treatment TME in NSCLC, we applied Spado[40] to identify spatial domains based on neighborhood cell compositions. We identified ten spatial domains across six samples (Fig. 5a and Extended Data Fig. 7). In an illustrative example from an LUSC sample (ST1), these domains displayed distinct cell-type enrichments (Fig. 5a and Supplementary Fig. 14). To systematically characterize the identified domains, we performed hierarchical clustering on the domain-level cell-type compositions across all samples (Fig. 5b). This analysis revealed eight conserved spatial ecotypes: tumor-enriched, airway-associated, three stroma-enriched (normal epithelial, endothelial and fibroblast) and three immune-enriched (lymphocyte, myeloid and plasma) ecotypes (Fig. 5b). These conserved spatial arrangements suggest shared biological organization across patients despite intersample variability.

Consistent with the pathological examination, the tumor-enriched ecotype was exclusively detected in NMPR samples. Lymphocyte

https://doi.org/10.1038/s41588-026-02610-1

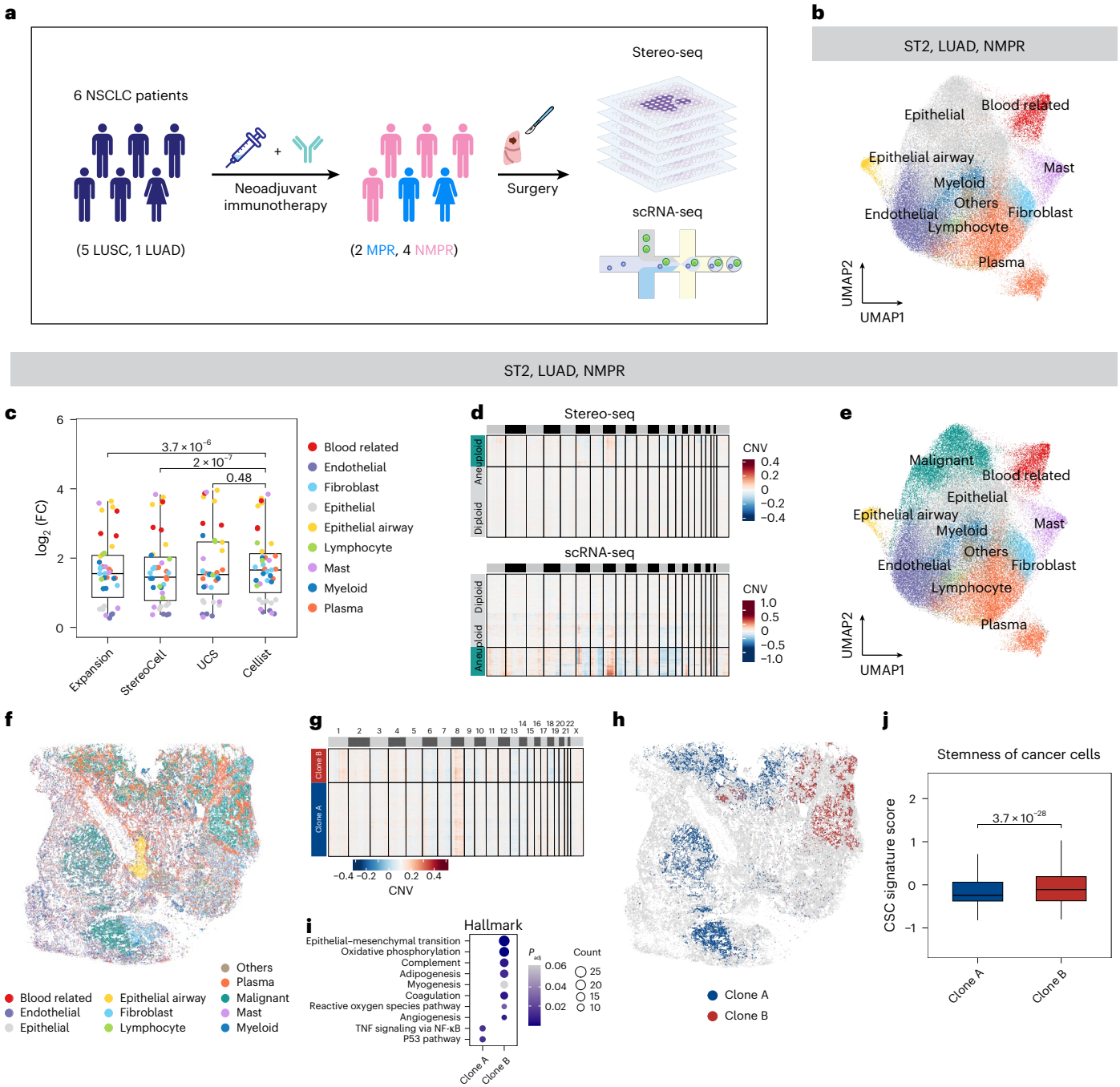


Fig. 4 | Application of Cellist in the characterization of post-therapy TME of NSCLC. a, Workflow of the research design in the study of TME in NSCLC. Six tumor specimens were collected from six NSCLC patients (five LUSCs and one LUAD) after neoadjuvant immunotherapy, which were subjected to Stereo-seq and matched scRNA-seq profiling. **b**, UMAP visualization of the ST2 Stereo-seq data colored by cell types annotated using a marker-based method. **c**, Marker specificity of annotated cell types in the ST2 sample from the Stereo-seq human NSCLC dataset, measured by $\log_2$(FC) of the top DE genes in each cell type versus all other cell types. Each dot represents one DE gene in a cell type shown in **b**. Statistical significance was assessed using paired, one-sided Wilcoxon's signed-rank tests across genes without adjustment for multiple comparisons ($n = 39$ genes). **d**, Heatmap showing inferred CNV profiles in the ST2 sample estimated from Stereo-seq data (top) and matched scRNA-seq data (bottom) using copyKAT. The color scale indicates relative copy number gain or loss. **e**, UMAP visualization of the ST2 Stereo-seq data colored by refined cell-type annotations integrating marker expression and CNV-based malignancy classification. **f**, Spatial distribution of cell types identified in **e**. **g**, Clonal substructure of malignant cells delineated by hierarchical clustering based on CNV profiles. The color scale indicates the relative copy number gain or loss. **h**, Spatial distribution of tumor clones identified in **g**. **i**, Hallmark enrichment analysis using upregulated genes in each of the tumor clones. Enrichment against MSigDB Hallmark gene sets was assessed using a one-sided hypergeometric test with Benjamini–Hochberg correction for multiple testing. The dot size represents the number of overlapping genes and the dot color indicates the adjusted *P* value. **j**, The CSC signature scores in clone A and clone B malignant cells. Statistical significance between the two clones was assessed using one-sided Wilcoxon's rank-sum tests (clone A, $n = 9{,}067$ cells; and clone B, $n = 4{,}182$ cells). Box plots throughout the figure indicate the median (center line) and IQR (25th–75th percentiles; box) and the whiskers extend to the most extreme values within 1.5× the IQR. NF-κB, nuclear factor-κB; TNF, tumor necrosis factor.

ecotypes, consisting of aggregated B and T cells, exhibited spatial variability, ranging from localization at the tumor–stroma interface (ST1, ST4 and ST5) to infiltration into the tumor core (ST2 and ST3). Notably, lymphocyte ecotypes were present in both responders (ST4) and nonresponders (ST1, ST2, ST3 and ST5), suggesting that lymphocytic infiltration alone is insufficient to predict therapeutic outcome and that spatial context is critical. Together, Cellist proves to be invaluable in identifying biologically relevant spatial domains, revealing conserved TME ecotypes and providing new insights into the spatial heterogeneity underlying therapeutic responses in NSCLC.

Cellist enables decoding of the functionalities and spatial locations of myeloid subtypes associated with therapy response

After our analysis of lineage-level cell composition and spatial distribution within the TME, we next investigated the spatial distribution of cell subtypes. Emerging studies highlighted the critical role of myeloid cells in establishing specialized niches that contribute to both antitumor and protumor activities within the TME[41–46]. To delve deeper into this, we isolated myeloid cells from each sample and integrated them using Harmony[47] (Supplementary Fig. 15). Reclustering unveiled five distinct macrophage subtypes, including tissue-specific alveolar macrophages (AMs) and several tumor-associated macrophage subtypes, such as *TREM2*$^+$ and *SPP1*$^+$ macrophages[42–44,48] (Fig. 5c,d). Notably, among these, *CXCL9*$^+$ macrophages exhibited potential antitumor activity, evidenced by high M1 signature scores. AMs and *SLPI*$^+$ macrophages displayed heightened phagocytosis potential, whereas *SPP1*$^+$ macrophages displayed elevated angiogenesis signatures, suggesting a potential role in tumor metastasis (Fig. 5e).

Utilizing the spatial ecotypes defined earlier, we examined subtype distributions across different ecotypes. *TREM2*$^+$ macrophages were concentrated in the plasma cell-enriched ecotype and AMs were enriched in stromal ecotypes (Fig. 5f). Furthermore, *SPP1*$^+$ and *CXCL9*$^+$ macrophages were co-enriched in lymphocyte-enriched and plasma-enriched ecotypes (Fig. 5f). Spatial visualization revealed that these two subtypes were co-localized at the tumor boundary (Fig. 5g,h and Extended Data Fig. 7). Consistent with prior reports showing that macrophage-derived CXCL9 promotes antitumor T cell responses[41], we observed enrichment of cytotoxic signals at the tumor boundary (Fig. 5i), indicating an association of *CXCL9*$^+$ macrophages with inflammatory reactions and antitumor immunity. Concerning *SPP1*$^+$ macrophages, prior studies have unveiled that their co-localization with *FAP*$^+$ fibroblasts facilitated the formation of immune-exclusive barriers to limit T cell infiltration[42]. We noted higher expression of *FAP*$^+$ fibroblast markers at the tumor boundary and reduced cytotoxic signals within the tumor (Fig. 5i and Extended Data Fig. 8). These findings suggest that, although cytotoxic T cells are recruited by *CXCL9*$^+$ macrophages to the boundary, their infiltration into the tumor core is impeded by *SPP1*$^+$ macrophages and cancer-associated fibroblasts, potentially contributing to residual tumors.

However, the spatial pattern was not universally conserved across patients in our cohort. In an NMPR patient with 40% residual tumor (ST5), macrophages were less abundant at the tumor boundary, with only modest enrichment of *SLPI*$^+$ and *CXCL9*$^+$ macrophages (Fig. 5h, Extended Data Fig. 7 and Supplementary Fig. 16). Unlike ST1, cytotoxic signals were generally low throughout the slide of ST5 (Fig. 5i). Considering low major histocompatibility complex (MHC) class II expression (Fig. 5j), we speculated that *SLPI*$^+$ macrophages might help to shape an immunosuppressive environment in ST5, potentially underlying the limited therapeutic efficacy.

Taken together, Cellist-enabled segmentation facilitates the identification of diverse myeloid cell subtypes in NSCLC and, when combined with spatial domain analysis, further reveals their spatial distribution and functional heterogeneity within the TME.

Discussion

High-resolution ST has provided unprecedented insights into tissue organization and cellular interactions. Cell segmentation is an indispensable step, particularly for subcellular ST data, because it directly determines the granularity and accuracy of downstream single-cell analyses. However, existing segmentation methods are often platform specific and struggle with scalability across large, high-resolution datasets.

To address these challenges, we introduced Cellist, a new multimodal-based cell segmentation method that integrates histological and transcriptomic information. Cellist enables robust cell segmentation across a wide range of high-resolution ST platforms. Benchmarking across multiple datasets demonstrates its computational efficiency and improved segmentation accuracy (Extended Data Fig. 1 and Supplementary Table 1), facilitating downstream analyses.

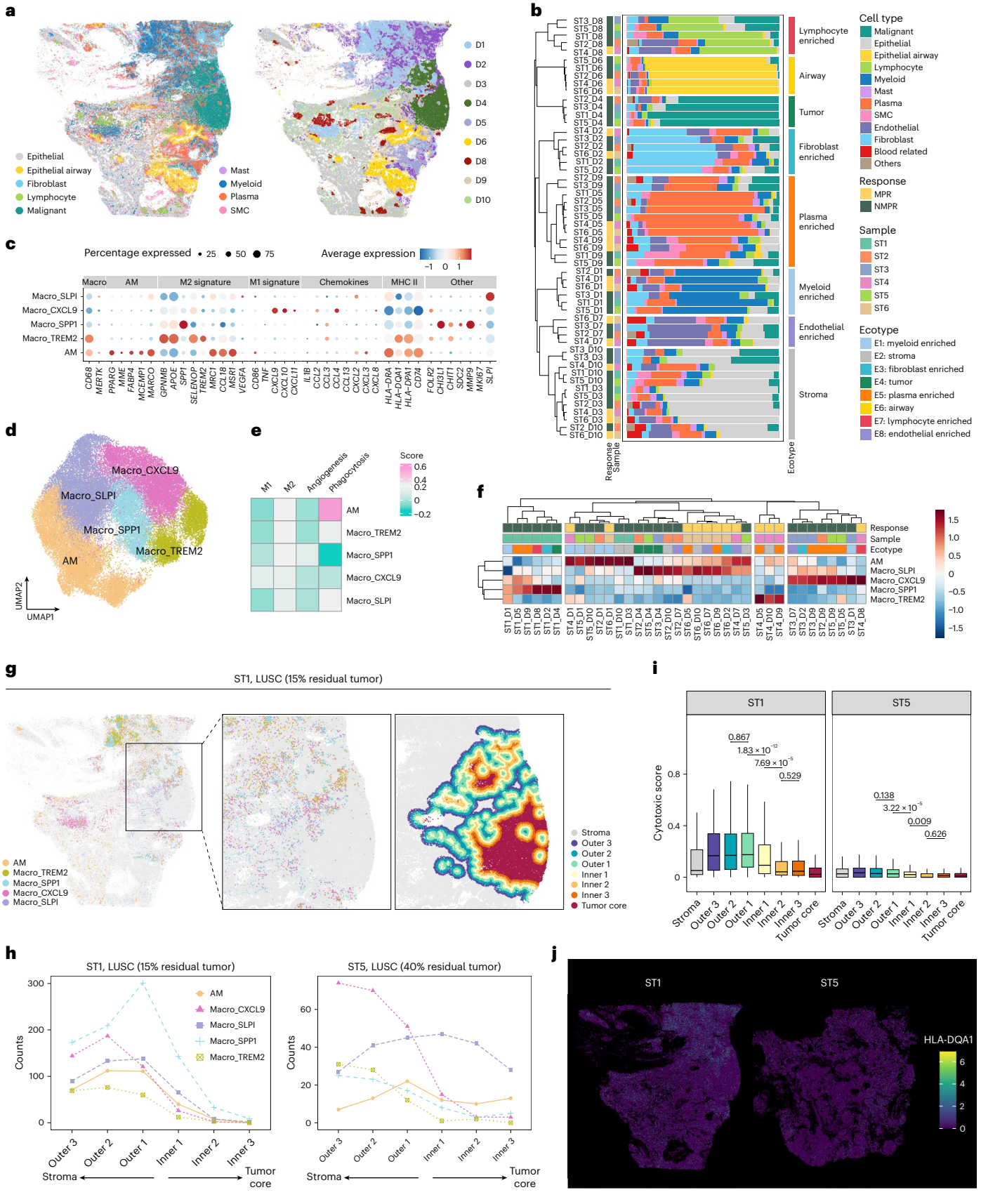
We applied it to post-neoadjuvant immunotherapy NSCLC samples profiled by Stereo-seq, enabling single-cell-level characterization of the post-treatment TME. This analysis revealed distinct myeloid subtypes with specific spatial organizations and functions. Notably, we identified two types of macrophage-involved niches at the tumor–stroma boundary, each potentially linked to limited therapeutic efficacy but in different ways. In ST1, *CXCL9*$^+$ and *SPP1*$^+$ macrophages co-accumulated at the boundary. The higher enrichment of cytotoxic signals at the boundary indicates an overall 'hot' immune environment contributed by *CXCL9*$^+$ macrophages, whereas reduced T cell infiltration within the tumor suggests potential exclusion mediated by *SPP1*$^+$ macrophages. In contrast, ST5 exhibited co-localization of *CXCL9*$^+$ and *SLPI*$^+$ macrophages at the boundary, as well as low macrophage abundance, reduced MHC class II expression and consistently weak cytotoxic signals, collectively revealing a relatively 'cold' immune environment. Yet, the mechanisms through which *SLPI*$^+$ macrophages contribute to immunosuppression remain to be elucidated and require further experimental validation.

Fig. 5 | Spatial distribution of myeloid subtypes in the TME of NSCLC. **a**, Spatial distribution of cell types (left) and spatial domains identified using SpaDo (right) in the sample of ST1. **b**, Cross-sample ecotype definition based on cell-type compositions. The bar plot displays the cell-type compositions in each domain from all samples. Hierarchical clustering on the cell-type enrichment profile classifies all domains into eight ecotypes: tumor, airway, three stromal ecotypes (enriched in epithelial, endothelial or fibroblast cells) and three immune ecotypes (enriched in lymphocyte, myeloid or plasma cells). **c**, Dot plot showing the marker genes for each myeloid subtype. The color of the dot represents the scaled expression and the size represents the percentage of cells expressing the marker. **d**, UMAP plot of myeloid cells extracted from all samples after Harmony integration, colored by marker-defined subtypes. **e**, Heatmap showing the signature score of different functional pathways in different myeloid subtypes. **f**, Heatmap showing the enrichment of different myeloid subtypes in different domains. Only domains with >100 myeloid cells were included in this analysis. Raw cell numbers are centered and scaled within each column. **g**, Spatial distribution of myeloid subtypes in the ST1 sample. Left: entire slide. Center, tumor–stroma boundary. Right: illustration of outer and inner cell layers determined according to the distance to the tumor domain. Each layer is ~75 µm in width. **h**, Line plot showing cell numbers of different myeloid subtypes in different cell layers of ST1 (left) and ST5 (right). **i**, Box plot showing the cytotoxic scores of lymphocytes located in different cell layers of ST1 and ST5. Each data point represents one lymphocyte. Statistical comparisons were performed using one-sided Wilcoxon's rank-sum tests without adjustment for multiple comparisons. Sample sizes were: ST1: outer 2 (n = 430 cells), outer 1 (n = 606 cells), inner 1 (n = 322 cells), inner 2 (n = 131 cells), Inner 3 (n = 89 cells); ST5: outer 2 (n = 430 cells), outer 1 (n = 569 cells), inner 1 (n = 294 cells), inner 2 (n = 181 cells), inner 3 (n = 141 cells). Box plots throughout the figure indicate the median (center line) and IQR (25th–75th percentiles; box) and the whiskers extend to the most extreme values within 1.5× the IQR. **j**, The spatial expression of *HLA-DQA1* in the samples of ST1 and ST5. SMC, smooth muscle cell.

https://doi.org/10.1038/s41588-026-02610-1

We note that biological findings from the NSCLC Stereo-seq dataset may be influenced by technical noise, limited sample size and the inherent heterogeneity of post-ICB tumor tissues. Nevertheless, these results demonstrate Cellist's robustness and practical utility in handling noisy, clinically relevant ST data.

Cellist also has certain limitations. First, Cellist prioritizes accurate recovery of cell-level expression rather than precise morphological delineation. This limitation is contingent on the strategies adopted by different ST techniques. In situ hybridization-based or in situ sequencing-based methods retain in situ transcript positions,



whereas barcoding-based approaches are more susceptible to transcript diffusion during tissue permeabilization, reducing the reliability of morphology inference. This challenge could be mitigated through improved experimental protocols that better preserve transcript localization. Second, although high-resolution ST techniques enable exploration of subcellular variability, preliminary analyses are constrained in distinguishing between nuclear and cytoplasmic compartments[15,24]. Consequently, detailed investigations into RNA distribution at the organelle level such as mitochondria or extracellular regions remain challenging. Advances in in situ techniques, protein labeling and auxiliary organelle staining hold promise for a deeper understanding of sophisticated RNA localization in an unbiased manner. Last, Cellist does not yet account for cell-type-specific differences in morphology or RNA distribution. Future extensions could integrate prior knowledge to enable more refined, cell-type-aware segmentation.

In conclusion, Cellist offers a valuable solution for precise cell segmentation in high-resolution imaging-based or barcoding-based ST platforms. Its ability to obtain high-purity expression profiles and facilitate downstream single-cell analyses—even under noisy, clinically relevant conditions—highlights its broad applicability. Cellist is publicly available at https://github.com/wanglabtongji/Cellist and we anticipate that it will play a pivotal role in preprocessing high-resolution ST data to advance understanding of tissue architecture at single-cell resolution.

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/s41588-026-02610-1.

References

1. Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
2. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
3. Maniatis, S. et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89–93 (2019).
4. Ortiz, C. et al. Molecular atlas of the adult mouse brain. *Sci. Adv.* **6**, eabb3446 (2020).
5. Chen, W. T. et al. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* **182**, 976–991 (2020).
6. Asp, M. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660 (2019).
7. Fawkner-Corbett, D. et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826 (2021).
8. Ji, A. L. et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 497–514 (2020).
9. Wu, S. Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347 (2021).
10. Wu, R. et al. Comprehensive analysis of spatial architecture in primary liver cancer. *Sci. Adv.* **7**, eabg3750 (2021).
11. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl Acad. Sci. USA* **116**, 19490–19499 (2019).
12. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* https://doi.org/10.1126/science.aat5691 (2018).
13. Janesick, A. et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).
14. Cho, C. S. et al. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184**, 3559–3572 (2021).
15. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).
16. Park, J. et al. Cell segmentation-free inference of cell types from in situ transcriptomics data. *Nat. Commun.* **12**, 3545 (2021).
17. Petukhov, V. et al. Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* **40**, 345–354 (2022).
18. Si, Y. et al. FICTURE: scalable segmentation-free analysis of submicron-resolution spatial transcriptomics. *Nat. Methods* **21**, 1843–1854 (2024).
19. Eng, C. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
20. Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
21. Qian, X. et al. Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* **17**, 101–106 (2020).
22. Prabhakaran, S. Sparcle: assigning transcripts to cells in multiplexed images. *Bioinform. Adv.* **2**, vbac048 (2022).
23. Chen, Y., Xu, X., Wan, X., Xiao, J. & Yang, C. UCS: a unified approach to cell segmentation for subcellular spatial transcriptomics. *Small Methods* **9**, e2400975 (2025).
24. Chen, H., Li, D. & Bar-Joseph, Z. SCS: cell segmentation for high-resolution spatial transcriptomics. *Nat. Methods* **20**, 1237–1243 (2023).
25. Li, M. et al. StereoCell enables high accuracy single cell segmentation for spatial transcriptomic dataset. Preprint at *bioRxiv* https://doi.org/10.1101/2023.02.28.530414 (2023).
26. Beucher, S. & Lantuéjoul, C. Use of watersheds in contour detection. In *Proc. International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*, 2.1–2.12 (CCETT/IRISA, 1979).
27. Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. *Nat. Methods* **22**, 592–599 (2025).
28. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
29. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
30. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
31. Wu, L. et al. An invasive zone in human liver cancer identified by Stereo-seq promotes hepatocyte-tumor cell crosstalk, local immunosuppression and tumor progression. *Cell Res.* **33**, 585–603 (2023).
32. Zhao, R. et al. Clonal dynamics and Stereo-seq resolve origin and phenotypic plasticity of adenosquamous carcinoma. *npj Precis. Oncol.* **7**, 80 (2023).
33. Yan, Y. et al. Multi-omic profiling highlights factors associated with resistance to immuno-chemotherapy in non-small-cell lung cancer. *Nat. Genet.* **57**, 126–139 (2025).
34. Yan, R. et al. The ZNF263/CPT1B axis regulates fatty acid beta-oxidation to affect cisplatin resistance in lung adenocarcinoma. *Pharmacogenomics J* **24**, 33 (2024).
35. Hoy, A. J., Nagarajan, S. R. & Butler, L. M. Tumour fatty acid metabolism in the context of therapy resistance and obesity. *Nat. Rev. Cancer* **21**, 753–766 (2021).

36. Yu, H. A. et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin. Cancer Res.* **19**, 2240–2247 (2013).
37. Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. *Cell* **166**, 21–45 (2016).
38. Lavudi, K., Nuguri, S. M., Pandey, P., Kokkanti, R. R. & Wang, Q. E. ALDH and cancer stem cells: pathways, challenges, and future directions in targeted therapy. *Life Sci.* **356**, 123033 (2024).
39. Zheng, Y. et al. Lung cancer stem cell markers as therapeutic targets: an update on signaling pathways and therapies. *Front. Oncol.* **12**, 873994 (2022).
40. Duan, B., Chen, S., Cheng, X. & Liu, Q. Multi-slice spatial transcriptome domain analysis with SpaDo. *Genome Biol.* **25**, 73 (2024).
41. House, I. G. et al. Macrophage-derived CXCL9 and CXCL10 are required for antitumor immune responses following immune checkpoint blockade. *Clin. Cancer Res.* **26**, 487–504 (2020).
42. 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).
43. Liu, Y. et al. Identification of a tumour immune barrier in the HCC microenvironment that determines the efficacy of immunotherapy. *J. Hepatol.* **78**, 770–782 (2023).
44. Bill, R. et al. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. *Science* **381**, 515–524 (2023).
45. Cheng, S. et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792–809 (2021).
46. Hu, J. et al. Tumor microenvironment remodeling after neoadjuvant immunotherapy in non-small cell lung cancer revealed by single-cell RNA sequencing. *Genome Med.* **15**, 14 (2023).
47. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
48. Zhang, L. et al. Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell* **181**, 442–459 (2020).

[1]Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of Ministry of Education, Department of Orthopedics, Tongji Hospital, School of Life Science and Technology, Tongji University, Shanghai, China. [2]Sycamore Research Institute of Life Sciences, Shanghai, China. [3]Frontier Science Center for Stem Cells, School of Life Sciences and Technology, Tongji University, Shanghai, China. [4]Central Laboratory, Innovation and Incubation Center, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China. [5]Department of Thoracic Surgery, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China. [6]These authors contributed equally: Dongqing Sun, Lele Zhang. ✉e-mail: zhangpeng1121@tongji.edu.cn; 08chenfeiwang@tongji.edu.cn

Methods

Nucleus segmentation for Cellist initialization

To initialize the segmentation process, Cellist supports two alternative nucleus- segmentation backends: a traditional Watershed-based approach[26] and the deep learning-based Cellpose algorithm[27,28]. This dual-backend design allows users to flexibly choose based on data characteristics and computational resources.

By default, we employed a Watershed-based segmentation pipeline due to its simplicity, computational efficiency and independence from graphics processing unit (GPU) acceleration. Although Watershed has known limitations in resolving tightly packed nuclei in dense tissues, it remains effective in moderately complex contexts and is well suited for large-scale datasets where GPU access is restricted.

In this pipeline, the ssDNA-stained images are first converted to grayscale. A global threshold is computed using the multi-Otsu method[49] to generate a coarse binary mask identifying background regions. To mitigate uneven illumination and enhance local contrast, an adaptive thresholding step with Gaussian filtering is applied to produce a second mask. The final nucleus binary mask is obtained by intersecting these two masks. To address common imaging artifacts, such as small holes within nuclei caused by noise or low signal, we performed a morphological closing operation on the binary mask. This operation helped preserve the integrity of nuclear shapes and ensured a more accurate distance transform.

Subsequently, a Euclidean distance map is computed for the foreground pixels (nuclei) and local maxima within this map are identified to mark potential nuclear centers. These markers guide the Watershed algorithm in delineating individual nuclei, which are then used as seeds for the downstream Cellist cell-segmentation process.

To address limitations of the Watershed method in densely packed or heterogeneous tissues, we also integrated Cellpose as an alternative nucleus-segmentation option. Cellpose leverages a deep convolutional neural network trained on diverse biological images and demonstrates improved accuracy in complex tissue environments. Users can switch to Cellpose when higher segmentation precision is needed and sufficient computational resources (for example, GPU) are available.

In our benchmarking analyses, we employed different nucleus-segmentation strategies depending on the dataset characteristics and available annotations. Specifically, Cellpose is used for the Stereo-seq mouse brain and human NSCLC datasets to better handle their dense and heterogeneous tissue structures. For the Seq-Scope mouse liver and seqFISH+ NIH/3T3 datasets, which exhibit more regular and well-separated nuclei, the default Watershed pipeline is applied. In addition, for the STARmap mouse cortex and Xenium human melanoma datasets, we directly utilized the nucleus segmentations provided in the original publications as initialization for Cellist.

Registration between staining and spatial expression profile

To accurately assign transcriptomic molecules to segmented nuclei or cells, Cellist performs multi-modal registration between spatial transcriptomic coordinates and the corresponding staining images. Although commercial platforms such as Stereo-seq provide preliminary alignment between transcriptomic data and staining images, minor misalignments can still occur and negatively impact segmentation accuracy. To address this, we integrated a robust image registration module based on the 'refine_alignment' function from the Spateo toolkit[50], which is fully automated and does not require manual landmark selection.

This approach is based on the biological rationale that unspliced RNA transcripts are predominantly localized in the nucleus. Therefore, it treats the registration problem as an image-to-image alignment task: the RNA spatial coordinate map (typically derived from unspliced RNA molecules) is used as the 'target', whereas the corresponding nuclear-stained image (for example, DAPI or ssDNA) is used as the 'source'. For H&E-stained samples, a preprocessing step is applied by using color deconvolution to extract the hematoxylin (nuclear) signal before registration.

The alignment is achieved via rigid affine transformation, which optimizes normalized cross-correlation between the two modalities at the global level. After registration, Cellist proceeds with pixel-level nucleus segmentation and quantifies gene expression by aggregating unique molecular identifiers within the refined nuclear region, yielding accurate nucleus-level expression profiles for downstream probabilistic spot assignment.

Overview of Cellist model

With nucleus segmentation completed, the next step is to expand the nucleus labels to include the cytoplasm. Based on the assumption that spots within a cell share similar expression profiles, spots should be assigned to nuclei or cells with high expression similarity. In addition, the proximity between the spot and the assigned nucleus should be minimized. This problem is reformulated as finding an optimal probabilistic assignment from each spot to neighboring nuclei, minimizing both expression dissimilarity and physical distance between the spot and the assigned nucleus. For each spot i, the problem is expressed as follows:

$$\begin{aligned} &\arg\min \sum_k \alpha R_{ik} D(\mathbf{X}_i, \mathbf{X}_k) + (1-\alpha) R_{ik} ||\mathbf{Z}_i - \mathbf{Z}_k||_2 \\ &\text{s.t.} \forall_i \forall_k R_{ik} \in \{0,1\} \end{aligned} \tag{1}$$

Here, s.t. denotes 'subject to' and R_{ik} denotes the assignment from spot i to cell k, with 1 representing membership. $||\mathbf{Z}_i - \mathbf{Z}_k||_2$ represents the Euclidean distance (in pixels) between spot i and the centroid of cell k. We also include a hyperparameter α to control the relative contributions of transcriptional dissimilarity and spatial distances. $\mathbf{X}_i$ and $\mathbf{X}_k$ indicate the gene expression of spot i and cell k, respectively, with the expression dissimilarity between them measured by Pearsons dissimilarity function D, where $D(\mathbf{A}, \mathbf{B}) = 1 - \text{corr}(\mathbf{A}, \mathbf{B})$, with $\text{corr}(\mathbf{A}, \mathbf{B})$ denoting the Pearson correlation coefficient between $\mathbf{A}$ and $\mathbf{B}$. Given the sparsity of gene expression in one spot, the spot's expression is augmented by borrowing information from neighboring spots within a radius r, with weights decreased by distance. The radius r is set to 2.5 µm by default in Cellist. This value is selected to balance between capturing sufficient neighboring information and avoiding oversmoothing.

$$x_{ig} = \sum_{j=1}^{J} w_{ij} x^*_{jg}$$

$$w_{ij} = \begin{cases} 1, & ||\mathbf{Z}_i - \mathbf{Z}_j||_2 = 0 \\ \frac{1}{||\mathbf{Z}_i - \mathbf{Z}_j||_2^2}, & 0 < ||\mathbf{Z}_i - \mathbf{Z}_j||_2 \le r \\ 0, & ||\mathbf{Z}_i - \mathbf{Z}_j||_2 > r \end{cases}$$

In the equation above, x_{ig} represents the augmented expression of gene g in spot i, x^*_{jg} the original expression of gene g in spot j and w_{ij} the weight between spot i and spot j.

To enable soft assignment between spots and nuclei, we added an entropy regularization term over R, with a weight hyperparameter σ, following the soft k-means clustering[51]. The formulation is:

$$\begin{aligned} &\arg\min \sum_k [\alpha R_{ik} D(\mathbf{X}_i, \mathbf{X}_k) + (1-\alpha) R_{ik} ||\mathbf{Z}_i - \mathbf{Z}_k||_2 + \sigma R_{ik} \log R_{ik} + \beta R_{ik} P_{ik}] \\ &\text{s.t.} \forall_i \forall_k R_{ik} > 0, \forall_i \sum_k R_{ik} = 1 \end{aligned} \tag{2}$$

A penalty term P_{ik}, weighted by a hyperparameter β, was included to prevent excessive distance between spots and assigned nuclei, with a maximum allowed distance $d_{\max}$:

$$P_{ik} = \begin{cases} \left(||\mathbf{Z}_i - \mathbf{Z}_k||_2 - d_{\max}\right)^2, & ||\mathbf{Z}_i - \mathbf{Z}_k||_2 \ge d_{\max} \\ 0, & ||\mathbf{Z}_i - \mathbf{Z}_k||_2 < d_{\max} \end{cases}$$

The value of $D(\mathbf{X}_i, \mathbf{X}_k)$ is within $[0,2]$. To ensure that all four terms in equation (2) are dimensionless and comparable, we scaled all physical distance-related terms relative to the average cell diameter, specified by the user-provided $d_{\max}$ mutiplied by 2, that is:

$$\arg\min \sum_k \left[\alpha R_{ik} D(\mathbf{X}_i, \mathbf{X}_k) + (1-\alpha) R_{ik} \frac{||\mathbf{Z}_i - \mathbf{Z}_k||_2}{2d_{\max}} + \sigma R_{ik} \log R_{ik} + \beta R_{ik} \frac{P_{ik}}{4{d_{\max}}^2} \right]$$
$$\text{s.t.} \forall_i \forall_k R_{ik} > 0, \forall_i \sum_k R_{ik} = 1 \tag{3}$$

Solution to the Cellist model

To solve the optimal assignment R_i, the Lagrangian function of equation (3) is derived with Lagrange multiplier λ:

$$L(R_i, \lambda) = \sum_{k=1}^{K} \alpha R_{ik} D(\mathbf{X}_i, \mathbf{X}_k) + (1-\alpha) R_{ik} \frac{||\mathbf{Z}_i - \mathbf{Z}_k||_2}{2d_{\max}} + \sigma R_{ik} \log R_{ik}$$
$$+ \beta R_{ik} \frac{P_{ik}}{4{d_{\max}}^2} + \lambda \left[\left(\sum_{k=1}^{K} R_{ik} \right) - 1 \right]$$

Then the partial derivative with respect to R_{ik} is calculated:

$$\frac{\delta L(R_i, \lambda)}{\delta R_{ik}} = \alpha D(\mathbf{X}_i, \mathbf{X}_k) + (1-\alpha) \frac{||\mathbf{Z}_i - \mathbf{Z}_k||_2}{2d_{\max}} + \sigma + \sigma \log R_{ik} + \beta \frac{P_{ik}}{4{d_{\max}}^2} + \lambda.$$

To make $\frac{\delta L(R_i, \lambda)}{\delta R_{ik}} = 0$, we could get R_{ik}:

$$\log(R_{ik}) = -\frac{4\alpha {d_{\max}}^2 D(\mathbf{X}_i, \mathbf{X}_k) + 2(1-\alpha) d_{\max} ||\mathbf{Z}_i - \mathbf{Z}_k||_2 + \beta P_{ik}}{\sigma} - \frac{\lambda}{\sigma} - 1$$
$$R_{ik} = \exp\left(-\frac{4\alpha {d_{\max}}^2 D(\mathbf{X}_i, \mathbf{X}_k) + 2(1-\alpha) d_{\max} ||\mathbf{Z}_i - \mathbf{Z}_k||_2 + \beta P_{ik}}{\sigma} \right) \exp\left(-\frac{\lambda}{\sigma} - 1 \right) \tag{4}$$

By combining the constraint for $\sum_{k=1}^{K} R_{ik} = 1$, we could solve for $\exp(-\frac{\lambda}{\sigma} - 1)$.

$$\sum_k \exp\left(-\frac{4\alpha {d_{\max}}^2 D(\mathbf{X}_i, \mathbf{X}_k) + 2(1-\alpha) d_{\max} ||\mathbf{Z}_i - \mathbf{Z}_k||_2 + \beta P_{ik}}{\sigma} \right) \exp\left(-\frac{\lambda}{\sigma} - 1 \right) = 1$$
$$\exp\left(-\frac{\lambda}{\sigma} - 1 \right) = \frac{1}{\sum_k \exp\left(-\frac{4\alpha {d_{\max}}^2 D(\mathbf{X}_i, \mathbf{X}_k) + 2(1-\alpha) d_{\max} ||\mathbf{Z}_i - \mathbf{Z}_k||_2 + \beta P_{ik}}{\sigma} \right)}$$

We substituted $\exp(-\frac{\lambda}{\sigma} - 1)$ in equation (4) to remove the parameter λ and finally got the normalized R_{ik}:

$$R_{ik} = \frac{\exp\left(-\frac{4\alpha {d_{\max}}^2 D(\mathbf{X}_i, \mathbf{X}_k) + 2(1-\alpha) d_{\max} ||\mathbf{Z}_i - \mathbf{Z}_k||_2 + \beta P_{ik}}{\sigma} \right)}{\sum_{k'} \exp\left(-\frac{4\alpha {d_{\max}}^2 D(\mathbf{X}_i, \mathbf{X}_{k'}) + 2(1-\alpha) d_{\max} ||\mathbf{Z}_i - \mathbf{Z}_{k'}||_2 + \beta P_{ik'}}{\sigma} \right)} \tag{5}$$

According to the assignment probability R_{ik}, spot i would be assigned to the cell with the highest probability, thus $R_i = \max_k R_{ik}$. Spots with low probability to all cells will be regarded as background, defined as those with $R_i < R_{\text{cut}}$. However, due to variability in the distribution of R_i across different platforms or datasets, selecting an appropriate value for R_{cut} can be challenging (Supplementary Note 2). To address this, we introduced the parameter P_{noise}, which specifies the expected proportion of spots to be classified as background noise. Specifically, spots with R_i values in the bottom P_{noise} quantile are considered background. A larger P_{noise} value corresponds to more spots being discarded as noise.

Implementation of Cellist

A typical Stereo-seq slide is ~1 cm × 1 cm (~20,000 spots × 20,000 spots) in size, which poses great challenges in terms of memory and runtime efficiency. Besides, it makes no sense to measure the similarity between nucleus and spots that are far apart. Based on the above two considerations, the slide was divided into smaller patches using a sliding-window approach for parallel computation. Patches were generated with a 550 μm × 550 μm (1,100 spots × 1,100 spots) sliding square and a 500-μm stride length. Within each patch, highly variable genes were identified based on the Watershed-segmented cells, and used for later refined Cellist segmentation. The sliding-window strategy might result in inconsistent cell assignment for spots located in the overlapped region of adjacent patches. The final assignments of these spots were determined according to the maximal probability.

Parameter selection

In our model, four parameters played crucial roles: α balances the transcriptional dissimilarity versus physical distance, σ represents the weight for the regularization term, β is the weight for the penalty term and $d_{\max}$ defines the maximum spatial threshold beyond which assignments are penalized. To determine appropriate values for the hyperparameters, we conducted a benchmark analysis across various parameter combinations using simulated data. Based on the benchmark results, we set $\alpha = 0.9$, $\sigma = 1.0$, $\beta = 5.0$ and $d_{\max} = 25$ as default values. Further details of the data simulation and the parameter selection are available in Supplementary Notes 1, 2 and 4.

Evaluation of segmentation performance

To comprehensively evaluate Cellist's segmentation performance across diverse spatial transcriptomic datasets, we employed a multifaceted strategy encompassing four distinct evaluation perspectives: spatial accuracy, intracellular consistency, neighborhood separation and biological interpretability. The choice of metrics was guided by the nature of each dataset (for example, presence of ground truth) and the specific biological questions addressed (Supplementary Table 5).

Segmentation accuracy with ground truth (IoU)

For datasets with available expert-annotated segmentation masks (for example, the NIH/3T3 seqFISH+ dataset), we evaluated spatial accuracy using the IoU metric. IoU quantifies the overlap between predicted cell regions and ground-truth cell masks, reflecting morphological fidelity without relying on gene expression patterns. This shape-based metric provides an unbiased benchmark to compare Cellist against other methods.

Intracellular expression consistency (random correlation)

For most datasets lacking ground truth, we prioritized evaluating whether segmented cells capture internally consistent gene expression. To this end, we introduced the random correlation metric. In this metric, transcripts assigned to each cell were randomly split into two subsets and Pearson's correlation between their pseudo-bulk gene expression profiles was computed. As this partitioning is random and spatially agnostic, it avoids biases from polarized or localized RNA distributions, offering a robust proxy for expression purity within cells. A higher random correlation indicates more internally consistent segmentation.

Moreover, we introduced an extension to this approach by splitting cells along a randomly oriented axis passing through each cell's centroid. This directional split naturally samples across spatial configurations (for example, left or right, top or bottom, diagonal), enhancing the robustness of the metric.

Neighborhood purity (purity score)

To complement random correlation and provide a more stable and interpretable boundary quality measure, we implemented a variance-based purity score. For each gene g in a predicted cell, we computed the normalized variance of gene expression across all spots $s \in c$:

$$F_{g,c} = \frac{\text{Var}(x_{g,s} | s \in c)}{\mu_{g,c}}$$

where $x_{g,s}$ is the imputed expression of gene g in spot s and $\mu_{g,c}$ is the mean expression of gene g in cell c. The gene set G includes highly variable genes across the patch. The within-cell variance score of cell c is then defined as the median across all genes in G:

$$F_{\text{in}}(c) = \text{Median}_{g \in G}(F_{g,c}).$$

To define a cell's neighborhood, we considered a square window of side length r centered at the centroid of cell c, where r is set slightly larger than the average cell diameter to include nearby cells. We collected all spots within this window and computed the same normalized variance score across these neighboring spots, denoted as $F_{\text{nbhd}}(c)$ (where the subscript 'nbhd' stands for the neighborhood). The final purity score is then defined as:

$$P(c) = \frac{F_{\text{in}}(c)}{F_{\text{nbhd}}(c)}.$$

A lower purity score indicates greater internal uniformity of the predicted cell relative to its surrounding neighborhood, suggesting well-captured cellular expression with limited transcript contamination.

Cross-method evaluation (cross-correlation)

To compare Cellist with other methods in the absence of ground truth, we adopted a cross-correlation strategy, commonly used in prior segmentation benchmarks[17,24]. This approach evaluates how well transcriptomic coherence is preserved across overlapping regions between different segmentation results.

Given two segmentation results A and B, we first identified all overlapping cell pairs. A single cell from segmentation B (treated as the source) may overlap with multiple cells from segmentation A (treated as the target), and vice versa. For each source cell, we matched it to the target cell with which it shares the largest overlapping area. We then computed the gene expression correlation between the overlap region and the nonoverlapping region of either the source or the target segmentation. If the source segmentation is more accurate, the transcriptomic profile of the overlap region will resemble that of the remaining part of the source cell, yielding a higher correlation. Conversely, if the target segmentation is more accurate, the overlap region will better match the transcriptomic profile of the target cell.

To ensure the robustness and fairness of this comparison, we performed the analysis in both directions. In the first pass, we treated Cellist as the target segmentation and each of the other methods as the source. In the second pass, we reversed the roles, using Cellist as the source and each other method as the target. By comparing the correlation values in both directions, we could infer which segmentation more faithfully preserves expression consistency across cell boundaries.

Biological fidelity (scRNA-seq concordance and marker specificity)

To assess whether segmentation results preserve meaningful cell-type-specific transcriptomic signals, we employed two biological benchmarks:

(1) Cell-type concordance with scRNA-seq: for datasets with matched scRNA-seq references, we compared pseudo-bulk gene expression profiles (averaged within each cell type) between spatial cells and their scRNA-seq counterparts. A higher Pearson's correlation suggested that segmentation better preserves true cell identity.

(2) Marker gene specificity: for each common cell type identified across methods, we constructed pseudo-bulk expression profiles by averaging gene expression within that cell type. We then compared each cell type to all other cell types using Wilcoxon's rank-sum test, identifying DE genes with Benjamini–Hochberg-adjusted $P < 0.01$. To enable fair comparison across segmentation methods, we selected the top DE genes for each method and focused on the shared subset of DE genes identified across all methods. These shared genes are used to evaluate and compare the $\log_2$(FC) across segmentation strategies. Higher fold-change values are interpreted as indicative of cleaner separation between cell types, thus reflecting more accurate cell-level expression estimates.

The above two biological metrics reflect how well segmentation captures cell-type-specific expression.

Application of Cellist on Stereo-seq mouse brain dataset

Cell segmentation. The Stereo-seq mouse brain dataset includes a high-resolution image of single-stranded DNA (ssDNA) staining, which provides valuable nuclear localization information for cell segmentation. To segment nuclei, we first applied the Cellpose algorithm (v.3.1.1.1) using the pretrained 'nuclei' model, which is optimized for nuclear morphology. This step generated nuclear masks that serve as initial candidate cell locations. After nucleus segmentation, Cellist was run with the parameters detailed in Supplementary Table 6.

Cell-type annotation. Considering the cellular diversity in the brain, we annotated the segmented cells using the reference-based method SingleR[30] implemented by Stereopy[52] (v.1.1.0). The reference data are from a scRNA-seq atlas of the adolescent mouse brain[29]. Due to the large number of cells in the original reference, we used a downsampled version provided by Stereopy. To evaluate the annotation performance on different segmentation results, we measured the expression consistency between Stereo-seq and scRNA-seq. In the scRNA-seq data, for each cell type, we identified DE genes and selected the top five as marker genes. All marker genes from all cell types were pooled together and only genes detected by both scRNA-seq and Stereo-seq were used for subsequent correlation calculation. Pearson's correlation of the averaged expression was computed between Stereo-seq and scRNA-seq for each cell type. A higher correlation indicated more accurate gene expression captured by the segmented cells.

To evaluate how well segmentation methods reconstruct single-cell expression profiles (Fig. 2i,j), we also added bin 20 (~10 µm) into the comparison, which more closely approximates the average size of a mammalian cell. This allowed for a more realistic comparison against segmentation-based approaches by serving as a nonsegmented baseline with comparable spatial resolution. Importantly, as square bins inherently include transcripts from multiple nearby cells, the bin 20 expression profile represented a local mixture rather than a clean single-cell signal. As such, it provided a meaningful benchmark for assessing how well segmentation methods—including Cellist—recover cleaner, cell-type-specific gene expression profiles from mixed spatial transcriptomics data.

Application of Cellist on Seq-scope mouse liver dataset

Cell segmentation. The Seq-Scope mouse liver dataset contains four tissue tiles[14], each with a matched H&E-staining image. The H&E-staining images were manually aligned with the RNA assay in the original study. Besides, Seq-Scope had a center-to-center resolution of ~0.6 µm, comparable to the resolution of Stereo-seq. Thus, no further transformations were applied to the images or the spatial expression data. Instead, they were directly input into Cellist. The parameters used in this dataset are detailed in Supplementary Table 6.

Cell-type annotation. To annotate cell types, we integrated data from all four Seq-scope tiles for joint analysis. Initially, low-quality cells were filtered out by removing those with <20 detected genes, ensuring the reliability of downstream analyses. The remaining data were then

normalized and subjected to dimensionality reduction using principal component analysis (PCA). After this, unsupervised graph-based Louvain clustering was applied and the resulting cell clusters were then automatically annotated using a marker-based method provided by MAESTRO[53]. The marker gene list is derived from the original publication. In particular, pericentral hepatocytes (Hep_PCs) were identified using markers such as *Mup9*, *Mup17*, *Cpy2a5*, *Cyp2c29*, *Cpy2e2*, *Mup11* and *Hamp*, whereas periportal hepatocytes (Hep_PPs) were marked by *Serpina1e*, *Cpy2f2*, *Pigr*, *Hsd17b13*, *Mup3* and *Mup20*.

Application of Cellist on the seqFISH+ NIH/3T3 dataset

The seqFISH+ NIH/3T3 dataset[19] profiling 10,000 genes contains 7 FOVs with the size of 200 µm × 200 µm. In addition, auxiliary DAPI staining was provided to localize the nucleus. Cellist was initially designed for the barcoding-based, high-resolution spatial transcriptomics where each spot captured multiple transcripts. But, in the seqFISH+ data of single-molecule resolution, each pixel was around 103 nm. To make the seqFISH+ data suitable for Cellist, transcripts were aggregated within each bin (5 pixels × 5 pixels), treating each bin as a spot akin to Stereo-seq. This process resulted in binned data with a comparable resolution (~0.5 µm) to Stereo-seq, suitable for Cellist. The original publication provided manual segmentation results that could serve as ground truth for evaluation. For each segmentation result, the agreement with the ground truth was measured by IoU. The parameters used in this dataset are listed in Supplementary Table 6.

Application of Cellist on STARmap mouse cortex dataset

Cell segmentation. The STARmap dataset profiles 1,020 genes for a mouse primary visual cortex tissue[12], with each pixel measuring ~78 nm. Paired DAPI nucleus-staining images were also available. Similar to the seqFISH+ data, the raw STARmap data were transformed into grid-shaped data with a bin size of 7 pixels × 7 pixels and all genes were used to run Cellist. To evaluate the performance of different cell-segmentation methods in this dataset, the random correlation and cross-correlation metrics were calculated as in the mouse brain Stereo-seq dataset. The parameters used in this dataset are listed in the Supplementary Table 6.

Cell-type annotation. Cell-type annotation is performed following the same workflow described for the Seq-Scope mouse liver dataset, including quality control, normalization, dimensionality reduction and clustering. To identify major cell types in this dataset, we used markers provided by the original publication. Marker genes included those for excitatory neurons (*Slc17a7*, *Cux2*, *Sema3e*, *Plcxd2*, *Tpbg*, *Syt6* and *Ctgf*), inhibitory neurons (*Vip*, *Reln*, *Sst*, *Lhx6*, *Npy*, *Pvalb*, *Gad1* and *Gad2*), astrocytes (*Aqp4* and *Slc1a3*), oligodendrocytes (*Enpp2*, *Mbp* and *Plp1*) and endothelial cells (*Cldn5* and *Flt1*).

Application of Cellist on Xenium human melanoma

Cell segmentation. The Xenium human melanoma dataset profiles 382 genes, including 282 from the Xenium Human Skin Gene Expression Panel and 100 additional genes such as chemokines, cytokines and immune cell subtype markers. DAPI nucleus-staining images and the corresponding results of nucleus and cell segmentation are provided by 10x Xenium Ranger. To make the dataset compatible with Cellist, the raw spatial expression data were transformed into grid-shaped data with a bin size of 1 µm × 1 µm. The nucleus-segmentation results provided by 10x Xenium Ranger serve as the initial segmentation for all methods, including Cellist. Supplementary Table 6 details the parameters to run Cellist on this dataset.

Cell-type annotation. Cell-type annotation was performed following the same workflow as described for the Seq-Scope mouse liver dataset. Specifically, annotation relied on the marker gene list provided by 10x Genomics (Supplementary Table 4).

Implementation of other segmentation methods

Expansion segmentation. The initial Watershed segmentation helped to determine the cell or nucleus locations and labels. Then the labeled regions could be expanded outward by a fixed distance without overflowing into neighboring cells, which is termed an 'Expansion' method. The expansion is implemented by the 'expand_labels' function from scikit-image (v.0.20.0). In our analysis, we specified the distance as ~4.5–5.0 µm.

SCS segmentation. The SCS[24] package was applied to both barcoding-based and imaging-based data, including the Stereo-seq mouse brain, Seq-Scope mouse liver, seqFISH+ NIH/3T3, STARmap mouse cortex and Xenium human melanoma datasets. SCS relies on a key user-defined parameter 'r_estimate', which approximates the expected cell radius and directly impacts the segmentation results. To ensure fair evaluation, we tested a range of 'r_estimate' values for each dataset and selected those yielding optimal segmentation performance based on spatially independent random correlation metric and cell count consistency (Supplementary Note 5). Other parameters such as 'patch_size', 'bin_size' and alignment settings were chosen according to the resolution and data type of each platform, following the original study's guidelines. For imaging-based datasets, binned expression matrices were generated as in Cellist to enable compatibility with the SCS framework. All selected parameter configurations are summarized in Supplementary Table 7.

StereoCell segmentation. StereoCell[25] segmentation was run only on both barcoding-based datasets, including the Stereo-seq mouse brain and Seq-Scope mouse liver. As StereoCell is designed as a turnkey solution specifically optimized for BGI's Stereo-seq platform, it exposes few tunable parameters. Following the developers' recommendations, we used its default configuration across all runs.

Baysor segmentation. Baysor (v0.6.2) was applied to all imaging-based datasets (that is, seqFISH+ NIH/3T3, STARmap mouse cortex and Xenium human melanoma) using its Docker environment. Baysor offered flexibility in parameterization, especially for 'min-molecules-per-cell', the required number of molecules for a real cell and 'prior-segmentation-confidence', the confidence of prior segmentation. After its official documentation and prior publications, we tested multiple configurations of these parameters and selected the best-performing ones based on segmentation quality metrics (for example, IoU or random correlation) (Supplementary Note 5). The full list of chosen values is available in Supplementary Table 8.

UCS segmentation. UCS is a unified deep learning-based segmentation method that integrates spatial gene expression and nuclear morphology to delineate accurate cell boundaries[23]. We applied UCS segmentation to both imaging-based and barcoding-based spatial transcriptomics datasets. Following the original publication, we generated gene maps and nuclear masks as required inputs, using platform-provided molecule coordinates and nucleus segmentation obtained via either Watershed or Cellpose. UCS was run with default configurations for most settings, with minor tuning of the learning rates for the cell and foreground networks to ensure training convergence across datasets. Details of parameter settings and dataset-specific inputs are provided in Supplementary Note 5 and summarized in Supplementary Table 9.

Stereo-seq experiment of human NSCLC samples

Patient cohorts and sample collection. Patients were recruited at the Shanghai Pulmonary Hospital between 2021 and 2023 as part of a prospective observational study of treatment response in NSCLC. Eligible participants were adult patients diagnosed with resectable NSCLC, without EGFR or ALK driver mutations, who were scheduled

to receive two to four cycles of neoadjuvant ICB (anti-programmed cell death protein-1) combined with platinum-based chemotherapy, followed by surgical resection. Patients were consecutively screened based on clinical eligibility criteria and willingness to provide informed consent for research use of tumor tissue.

Six patients meeting these criteria were enrolled and fresh tumor tissues were collected at the time of surgical resection. Detailed clinicopathological information is provided in Supplementary Table 3.

As participation required informed consent and eligibility for surgery after neoadjuvant therapy, potential sources of bias include self-selection bias (patients willing to participate in research) and clinical selection bias toward individuals with operable disease and sufficient performance status to undergo combined ICB chemotherapy and surgery. In addition, the exclusion of tumors harboring EGFR or ALK mutations restricts the cohort to a molecularly defined subset of NSCLC. These factors may limit the generalizability of findings to broader NSCLC populations, particularly patients with unresectable disease, poor performance status or alternative oncogenic drivers. However, as the study focused on spatial characterization of the TME after neoadjuvant ICB chemotherapy, the selected cohort was appropriate for addressing the specific biological questions investigated.

The study was approved by the Institutional Review Board of Shanghai Pulmonary Hospital (number K23-310) and written informed consent was obtained from all participants before tissue acquisition and genomic analyses.

Spatial transcriptomic profiling by Stereo-seq. Spatial transcriptomics of NSCLC samples were obtained through Stereo-seq following the protocol provided by BGI (https://en.stomics.tech/). In brief, optimal cutting temperature-embedded tissues were sectioned into 10-µm-thick cryosections using a Leica CM1950 cryostat. These sections were then affixed to the surface of the Stereo-seq chip (manufactured by BGI) and incubated. Subsequently, the sections were fixed in methanol, followed by incubation and staining with a nucleic acid dye before imaging. Afterward, the tissue sections were washed and permeabilized to release RNA. After dye removal, the released RNA molecules were captured by DNA nanoballs and reverse transcribed into complementary DNA. The purified cDNA was then subjected to PCR amplification. The resulting products were used to construct sequencing libraries, involving fragmentation, amplification and purification steps. Finally, the prepared cDNA libraries were sequenced using the MGI DNBSEQ-Tx sequencer. For more detailed procedures and reagents, please refer to the Stereo-seq technique method[15].

Application of Cellist on Stereo-seq human NSCLC dataset

Analysis of scRNA-seq data. Matched scRNA-seq data of the six samples were obtained from another study. Beginning with the count matrix, standard single-cell analysis was performed on each sample using MAESTRO (v.1.5.1). This included quality control, dimensionality reduction, cell clustering, DE analysis and cell-type annotation. Cell types were annotated using the following marker genes: epithelial cells (*EPCAM* and *KRT6A*), fibroblasts (*COL1A1*, *POSTN* and *LUM*), endothelial cells (*PECAM1*, *VWF* and *ACKR1*), lymphocytes (*LTB* and *CXCL14*), myeloid cells (*LYZ*, *CD68*, *C1QA* and *C1QB*), plasma cells (*JCHAIN* and *MZB1*), mast cells (*MS4A2* and *KIT*) and erythrocytes (*HBA1*, *HBA2* and *HBB*). For each cell type, cells from six samples were integrated using the Harmony algorithm[47] based on 2,000 highly variable genes, followed by data dimensionality reduction, clustering analysis and differential gene identification.

Stereo-seq raw data processing. The ImageStudio software (v.1.1.3) was employed to process the ssDNA images, including image quality assessment, image stitching, tissue segmentation and cell segmentation. Subsequently, the SAW pipeline (v.6.0.0), developed by BGI Research, was utilized for read filtering and mapping with the reference genome GRCh38.92. SAW also facilitated image registration based on the track lines in the Stereo-seq chip. Through the integration of image registration and read quantification, we were able to derive the spatial gene expression matrices.

Single-cell-level analysis of Stereo-seq. After cell segmentation by Cellist, we employed the conventional single-cell analysis workflow on the single-cell spatial expression data using Seurat (v.4.1.3). At first, low-quality cells with <50 genes and >5,000 genes were filtered out and the gene expression was normalized by the total expression in each cell, multiplied by a scale factor of 10,000 and log(transformed). Next, 2,000 highly variable genes were selected for downstream data scaling and dimensional reduction. PCA was performed on the scaled data, followed by unsupervised graph-based Louvain clustering on the first 20 PC dimensions with resolution set to 0.6. Last, we annotated the cell clusters based on DE genes and canonical cell-type-specific gene markers used in the above scRNA-seq analysis.

Spatially aware imputation of segmented cells. In the NSCLC data, post-treatment tissue sections exhibited reduced transcript coverage due to technical challenges in Stereo-seq. To enable downstream sublineage analysis, such as CNV inference and fine-grained myeloid subtyping, we performed spatially aware expression imputation on the segmented cells. Our imputation strategy was inspired by the concept of data diffusion in similar cells as proposed in MAGIC[54], a method originally designed to impute gene expression in scRNA-seq. However, different from MAGIC, we incorporated physical distance alongside expression similarity to define 'similar' cells, ensuring that only cells exhibiting both high expression similarity and close physical proximity were utilized for imputing the gene expression of the target cell. To further reduce the risk of crosslineage contamination during imputation, we applied this procedure independently within each major annotated cell lineage (for example, epithelial cells and myeloid cells), ensuring that expression signals were recovered from biologically relevant and phenotypically coherent neighbors.

In the original MAGIC, cell–cell similarity was quantified by an affinity matrix calculated based on distances between cells in a low-dimension PCA space. Specifically, the Euclidean distance between cell a and cell b in the PCA embedding space is computed as:

$$\mathrm{Dist}(a,b) = ||\mathbf{Y}_a - \mathbf{Y}_b||_2.$$

Here $\mathbf{Y}_a$ and $\mathbf{Y}_b$ represent coordinates of cell a and cell b in the PCA embedding space. The distance matrix was then transformed to an affinity matrix with an adaptive Gaussian kernel:

$$A_{ab} = e^{-\left(\frac{\mathrm{Dist}(a,b)}{\gamma}\right)^2}. \quad (6)$$

The kernel width γ is adapted for each cell following MAGIC[55], depending on the k_ath nearest neighbor, $\mathrm{neigh}(a, k_a)$:

$$\gamma_a = \mathrm{Dist}(a, \mathrm{neigh}(a, k_a)). \quad (7)$$

According to equation (6), too long a distance will lead to a tiny affinity, which makes little contribution to the imputation of the target cell. We limited the affinity calculation within the nearest k cells, where $k = 3k_a$, in other words, affinities of cells with distances exceeding $\mathrm{Dist}(a, \mathrm{neigh}(a, k))$ will be set to zero.

The utilization of an adaptive kernel width results in asymmetry in the affinity matrix ($A_{ab} \neq A_{ba}$). To symmetrize the affinity matrix, we took the additive approach by averaging the affinities:

$$A_{\mathrm{exp}}(a,b) = (A_{ab} + A_{ba})/2.$$

Likewise, we calculated the physical affinity $A_{\mathrm{phy}}(a,b)$ by substituting PC dimensions with the physical space in the above equations.

For each cell, we retained only neighbors with both positive expression affinity and positive physical affinity to the target cell, namely, $A_{\text{exp}}(a,b)$ will be set to 0 if $A_{\text{phy}}(a,b)=0$, and vice versa. Subsequently, the Markov transition matrices based on expression affinity and physical affinity could be derived as follows:

$$M_{\text{exp}}(a,b) = \frac{A_{\text{exp}}(a,b)}{\sum_{n=1}^{K} A_{\text{exp}}(a,n)}$$

$$M_{\text{phy}}(a,b) = \frac{A_{\text{phy}}(a,b)}{\sum_{n=1}^{K} A_{\text{phy}}(a,n)}$$

The final transition matrix is obtained by combining the two Markov transition matrices with a weight hyperparameter ω:

$$M_{\text{comb}}(a,b) = \omega M_{\text{exp}}(a,b) + (1-\omega) M_{\text{phy}}(a,b).$$

Then the expression of gene g in cell a can be imputed by transferring information from neighbor cells with specific transition probabilities:

$$E_{\text{imputed}}(a,g) = \sum_{n=1}^{K} M_{\text{comb}}(a,n) \times E_{\text{raw}}(n,g).$$

This imputation process enabled us to recover missing gene expression in high-resolution ST data while considering both expression similarity and physical proximity between cells.

CNV analysis in Stereo-seq and scRNA. To discriminate malignant cells from normal epithelial cells, we utilized CopyKAT[56] (v.1.1.0) to estimate the CNVs of segmented cells based on the imputed gene expression. CopyKAT is run on all epithelial cells with lymphocytes as normal reference and default parameters within each individual sample. Malignant cells with more CNV events could be separated from normal ones through hierarchical clustering. Based on the hierarchical clustering result, we performed cell-type annotation at the cluster level; the epithelial cluster with a higher percentage of malignant cells would be annotated as a malignant cluster. For each LUSC sample, we also considered the expression of malignant signatures for final normal versus malignant cell assignment (Supplementary Fig. 11). Notably, if the number of input cells (that is, epithelial cells and lymphocytes) exceeds 65,000, we downsampled them to 65,000 to ensure that CopyKAT works without memory overflow. Similar CNV analysis was performed on paired scRNA-seq data for comparison. Consistency between Stereo-seq and scRNA-seq was evaluated using Pearson's correlation of averaged CNV profiles of predicted malignant cells in both techniques.

Tumor clone detection and function analysis. To investigate functional differences between tumor subclones, we first identified clones by performing hierarchical clustering on inferred CNV profiles of malignant cells. For each clone, we aggregated raw gene expression across all cells (pseudo-bulk) to enable robust DE analysis. DE genes between clones were identified using Wilcoxon's rank-sum test implemented in the Seurat package. Significantly upregulated genes (adjusted P (P_{adj}) < 0.01) were subjected to pathway enrichment analysis using the clusterProfiler (v.4.6.2) package[57] against the Hallmark gene sets from the Molecular Signatures Database (MSigDB v.7.2)[58].

To further assess cancer stemness features, we calculated CSC signature scores using a curated gene set derived from a lung-cancer review[39]. Specifically, the CSC gene set included markers commonly associated with cancer stem-like properties: *CD44*, *ALCAM*, *EPCAM*, *SLC22A3*, *POU5F1* and *ALDH1A1*. We used the 'AddModuleScore' function in Seurat to compute the enrichment scores of these markers across tumor clones.

Spatial domain identification in lung cancer. To enable joint spatial domain analysis across multiple NSCLC samples, we adopted the SpaDo algorithm (v.1.2.0)[40], which is specifically designed for cross-sample spatial domain detection based on the composition of cell types within each cell's neighborhood. In our analysis, the 30 nearest neighbors of each cell were used to characterize its neighborhood composition. To improve the scalability of SpaDo for large-scale spatial transcriptomics datasets, we modified the original pipeline by replacing the default hierarchical clustering step with k-means clustering, with the number of clusters (k) set to 10.

Using the previously annotated cell types, we computed the cell-type enrichment for each spatial domain. Subsequently, hierarchical clustering was performed on the cell-type enrichment profiles across spatial domains from all six samples. This analysis yielded eight domain clusters, which we defined as eight TME ecotypes: tumor, airway, epithelial-enriched, endothelial-enriched and fibroblast-enriched stromal ecotypes, as well as lymphocyte-enriched, myeloid-enriched and plasma cell-enriched immune ecotypes.

Myeloid cell subtype identification. Myeloid cells were extracted from each sample and integrated together using highly variable genes detected in the paired scRNA-seq myeloid reference. Harmony was employed to remove batch effects from samples, followed by the reclustering to identify myeloid subpopulations[47]. These subpopulations were annotated according to DE genes and canonical gene markers. Notably, due to the potential technical limits, one cell might capture signals from the neighboring cells. We manually removed these cell clusters expressing marker genes of other cell types, such as a cluster with high *COL1A1* expression.

To better characterize different myeloid subtypes, we calculated the gene signature score using the 'AddModuleScore' function in Seurat[59] to understand their functions in M1 or M2 polarization, angiogenesis or phagocytosis, proinflammation or anti-inflammation. The detailed signature genes are listed in Supplementary Table 10.

Spatial distribution analysis of myeloid cell subtypes. To understand the distribution of different myeloid subtypes in the TME, we calculated their enrichment in previously defined spatial domains. Furthermore, we focused on their distribution with respect to tumor and tumor–stoma interface, with tumor and stroma regions identified through spatial domain analysis. To be specific, domain D4 was designated as the tumor region, given its high enrichment of inferred malignant cells, whereas all other regions were considered stromal.

To define the tumor–stroma boundary, we calculated the Euclidean distance from each stromal cell to the nearest malignant cell within domain D4. Stromal cells were then stratified into spatial layers based on their proximity to the tumor, using distance intervals of 75 µm. Specifically, stromal cells located within 0–75 µm, 75–150 µm and 150–225 µm from the tumor boundary were defined as outer 1, outer 2 and outer 3, respectively. Stromal cells located at ≥225-µm distance were grouped as the distal stroma.

Similarly, for malignant cells in the tumor region, we computed their nearest distances to stromal cells located in outer 1 and defined tumor-side layers following the same 75-µm interval scheme. Cells within 0–75 µm, 75–150 µm, and 150–225 µm from the stromal interface were categorized as inner 1, inner 2 and inner 3, respectively, whereas those ≥225 µm away were assigned to the tumor core. With these layers defined, we could count the number of myeloid cells in each layer.

To explore the potential association between myeloid cells and lymphocytes, we calculated cytotoxic scores for lymphocytes in each layer. For each cell, the cytotoxic score was defined as the mean expression of the following cytotoxic markers: *GZMK*, *CCL4L2*, *CCL4*, *CCL3L1*, *NKG7*, *CST7*, *GZMA*, *CCL5*, *CMC1*, *SLAMF7*, *GZMH*, *TRGC2*, *KLRG1*, *CD8A*, *IFNG*, *EOMES*, *DTHD1*, *CD8B*, *CCL3* and *AC243829.1*.

Benchmark of computational efficiency. To evaluate the computational efficiency of different segmentation methods, we benchmarked their runtime and memory usage across multiple spatial transcriptomics

datasets. All experiments were conducted on a high-performance computing server equipped with a single NVIDIA A800 GPU and a 64-core central processing unit.

For each dataset, we recorded the total execution time (wall-clock time) and peak GPU memory usage required by each method. These metrics are summarized in Supplementary Table 1 and visually compared in Extended Data Fig. 1. This benchmarking provided a practical perspective on the scalability and computational demands of each method, particularly in the context of large-scale or high-resolution spatial transcriptomic datasets.

Reporting summary

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

Data availability

The raw Stereo-seq data of human NSCLC generated in this study have been deposited in the Genome Sequence Archive for Human (GSA-Human) of the National Genomics Data Center (NGDC), China National Center for Bioinformation and Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/gsa-human), under accession number HRA007118 (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA007118). The matched scRNA-seq data are available under accession number HRA006493 (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA006493). These datasets are subject to controlled access to protect patient privacy. Access can be requested through the GSA-Human portal and will be granted to researchers for non-commercial use upon approval by the Data Access Committee. Public datasets used to evaluate the segmentation methods are listed as follows: (1) the Stereo-seq mouse brain data with ssDNA staining are available in the MOSTA data portal (https://db.cngb.org/stomics/datasets/STDS0000058/sample) under sample ID 'STSP0004903'. The Stereo-seq mouse olfactory bulb data used for parameter selection are available from https://db.cngb.org/data_resources/project/CNP0009482. The downsampled scRNA-seq reference for cell-type annotation was obtained from Stereopy following the instructions provided at https://stereopy.readthedocs.io/en/latest/Tutorials/SingleR.html. The specific reference file used in this study is titled 'Mouse_brain_ref.anndata075.h5ad'. (2) The Seq-Scope mouse liver expression data can be accessed from the Gene Expression Omnibus database through accession number GSE169706. Spatial coordinates for high-definition map coordinate identifier barcodes are generated using the scripts from https://github.com/leeju-umich/Cho_Xi_Seqscope. The manually aligned H&E-staining images were obtained from https://github.com/chenhcs/SCS/tree/main/data. Tiles 2,104, 2,105, 2,106 and 2,107 were used for performance evaluation in our study. (3) The seqFISH+ NIH/3T3 cell line data were downloaded from Zenodo (https://zenodo.org/records/2669683)[60]. (4) The STARmap mouse cortex data are available at https://kangaroo-goby.squarespace.com/data. (5) The Xenium human melanoma data are available at the 10x Genomics website: https://www.10xgenomics.com/datasets/human-skin-preview-data-xenium-human-skin-gene-expression-panel-add-on-1-standard. The predesigned gene panel that we used for marker-based cell-type annotation was downloaded from https://cdn.10xgenomics.com/raw/upload/v1699308550/software-support/Xenium-panels/hSkin_panel_files/Xenium_hSkin_v1_metadata.csv. All processed data necessary to reproduce the analyses and figures in this study are available via Zenodo at https://doi.org/10.5281/zenodo.18638251 (ref. 61). Source data are provided with this paper.

Code availability

Cellist is an open-source Python package with source code freely available at https://github.com/wanglabtongji/Cellist. The exact code and analysis workflows used to generate the results reported in this study are available via Zenodo at https://doi.org/10.5281/zenodo.18638251 (ref. 61).

References

49. Liao, P.-S., Chen, T.-S. & Chung, P. C. A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.* **17**, 713–727 (2001).
50. Qiu, X. et al. Spatiotemporal modeling of molecular holograms. *Cell* **187**, 7351–7373 (2024).
51. Mao, Q., Wang, L., Goodison, S. & Sun, Y. Dimensionality reduction via graph structure learning. In *Proc. 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (eds Cao, L. et al.) 765–774 (ACM, 2015).
52. Fang, S. et al. Stereopy: modeling comparative and spatiotemporal cellular heterogeneity via multi-sample spatial transcriptomics. *Nat. Commun.* **16**, 3741 (2025).
53. Wang, C. et al. Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol.* **21**, 198 (2020).
54. Wang, Y. et al. Sprod for de-noising spatially resolved transcriptomics data based on position and image information. *Nat. Methods* **19**, 950–958 (2022).
55. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729 (2018).
56. Gao, R. et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **39**, 599–608 (2021).
57. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
58. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
59. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
60. Eng, C. L. & Cai, L. NIH3T3_point_locations for RNA seqFISH+ experiments. *Zenodo* https://doi.org/10.5281/zenodo.2669683 (2019).
61. Sun, D., Zhang, L., Zhang, P. & Wang, C. Reproducibility code and source data for the evaluation of Cellist. *Zenodo* https://doi.org/10.5281/zenodo.18638251 (2026).

Acknowledgements

This work was supported by the National Key R&D Program of China (grant numbers 2025YFA1805500 and 2022YFA1106000 to C.W.), the National Natural Science Foundation of China (grant numbers 32222026 and 32170660 to C.W., 82521002 and 82125001 to P.Z. and 82430053 to L.Z.); Shanghai Rising Star Program (grant number 21QA1408200 to C.W.); Natural Science Foundation of Shanghai (grant number 24ZR1492800 to C.W.); Innovation Program of Shanghai Municipal Education Commission (grant number 2023ZKZD33 to P.Z.); the Fundamental Research Funds for the Central Universities (grant number 22120240435 to C.W.); Shanghai Pilot Program for Basic Research (to C.W.); and Foundation of Shanghai Pulmonary Hospital (grant numbers FKYQ2308 to L.Z. and FKYXJC003 to P.Z.). We thank BGI for providing the Stereo-seq data of adult mouse brain and olfactory bulb, as well as the Stereo-seq chips used in this study. We thank Y. Zhang and other colleagues from BGI for the helpful discussion.

Author contributions

C.W. conceived the study and supervised the project. D.S. designed the Cellist methods and performed computational analyses. L.Z., T.H., Q.W. and P.Z. facilitated NSCLC sample collection and clinical data curation. P.Z. provided administrative support. D.S. and C.W. wrote the manuscript with input from all authors. All authors contributed to the review of the manuscript and provided final approval of the work.

Competing interests
The authors declare no competing interests.

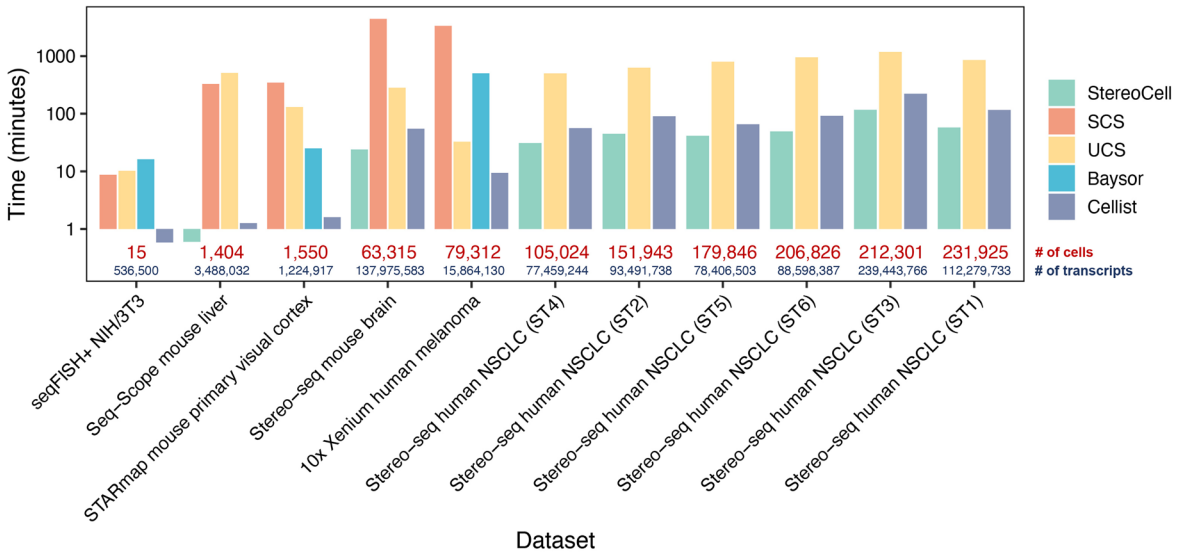
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41588-026-02610-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-026-02610-1.

Correspondence and requests for materials should be addressed to Peng Zhang or Chenfei Wang.

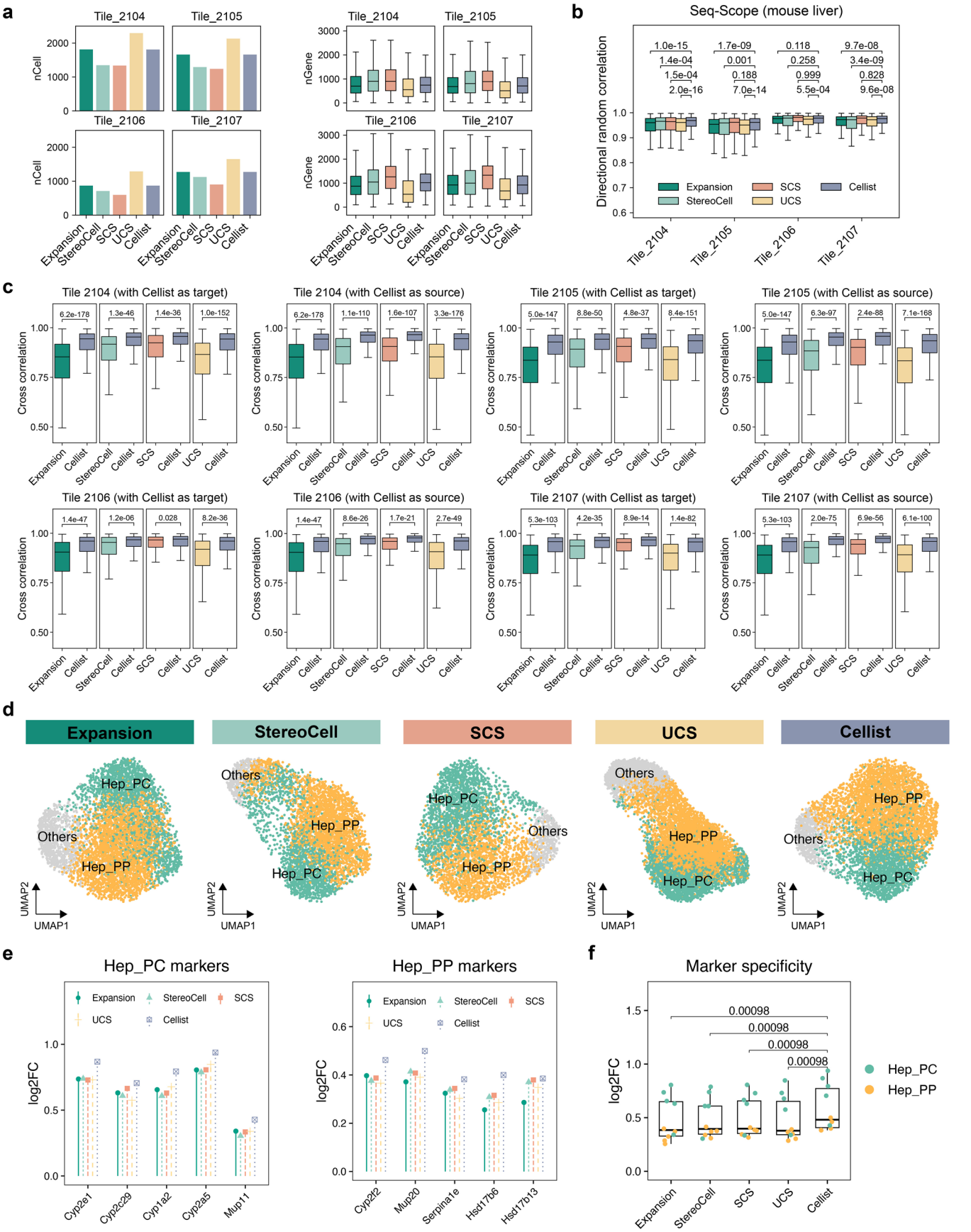
Peer review information *Nature Genetics* thanks Léo Guignard, Shyam Prabhakar and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-026-02610-1



Extended Data Fig. 1 | Runtime comparison of different segmentation methods across datasets of varying sizes. Runtime of each segmentation method on multiple datasets, ordered by the number of cells identified by Cellist. SCS failed to run on the Stereo-seq human NSCLC datasets due to memory limitations (NVIDIA A800 GPU). Red text indicates the number of cells identified by Cellist; blue text indicates the total number of transcripts in each dataset.

https://doi.org/10.1038/s41588-026-02610-1

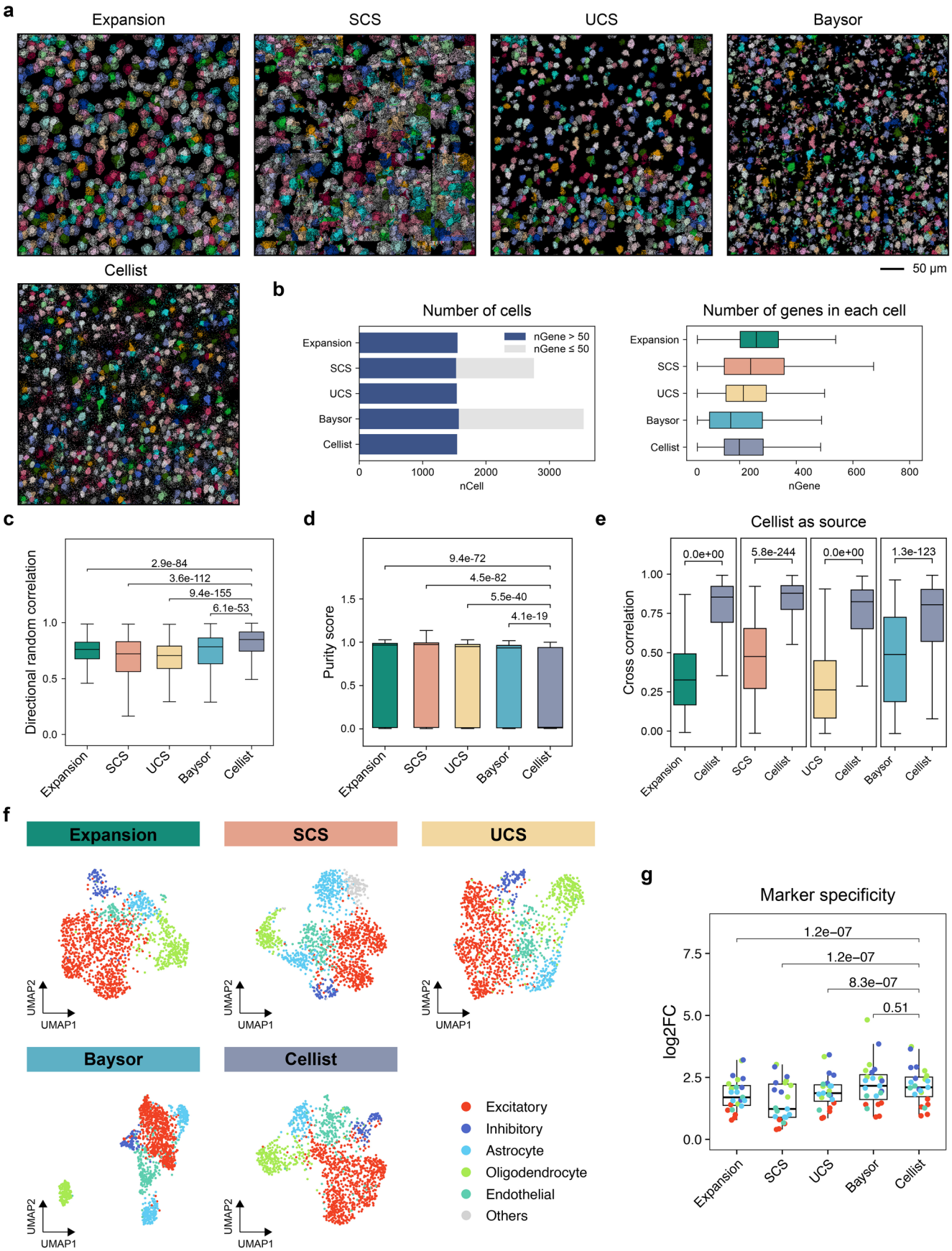


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41588-026-02610-1

Extended Data Fig. 2 | Comparison of cell segmentation results in the Seq-Scope mouse liver dataset. (**a**) Number of cells (left) and gene coverage (right) generated by different segmentation methods in each tile from the Seq-Scope mouse liver dataset. Exact sample sizes are provided in the Source Data. (**b**) Evaluation of intracellular expression homogeneity using the metric of directional random correlation. Each cell is bisected along a randomly oriented line passing through its centroid, and the Pearson correlation coefficient between the two subsets' pseudo-bulk expression profiles was computed to assess expression consistency across spatial halves. Analysis is restricted to cells containing more than 100 spatial spots. Statistical significance between each method and Cellist in each tile was assessed using one-sided Wilcoxon rank-sum tests. Exact sample sizes for each comparison are provided in the Source Data. (**c**) Evaluation of consistency between segmentations using the cross-correlation metric. Cross correlation was computed using Cellist segmentation as the target segmentation (left) and as the source segmentation (right), respectively. Statistical significance between methods was assessed in each tile using one-sided Wilcoxon rank-sum tests without adjustment for multiple comparisons. Exact sample sizes for each comparison are provided in the Source Data. (**d**) UMAP plots showing the cell type annotation based on cell segmentation by different methods in the Seq-Scope mouse liver dataset. (**e**) Lollipop plots showing the log2FC of top marker genes in each cell type versus all other cell types. The top 10 DE genes with highest log2FC are selected per method per cell type, and only common genes shared across methods are used for visualization. (**f**) Marker specificity of annotated cell types, measured by the log2FC of top DE genes in each cell type versus all other cell types. Only common genes shared across methods are used for fair comparison. Each dot represents one DE gene in a cell type from (**e**). The cell-type 'Others' is excluded from comparison. Statistical significance was assessed using paired one-sided Wilcoxon signed-rank tests without adjustment for multiple comparisons (n = 10 genes). Box plots throughout the figure indicate the median (center line), interquartile range (25^{th}-75^{th} percentiles; box), and whiskers extending to the most extreme values within 1.5× the IQR.

https://doi.org/10.1038/s41588-026-02610-1

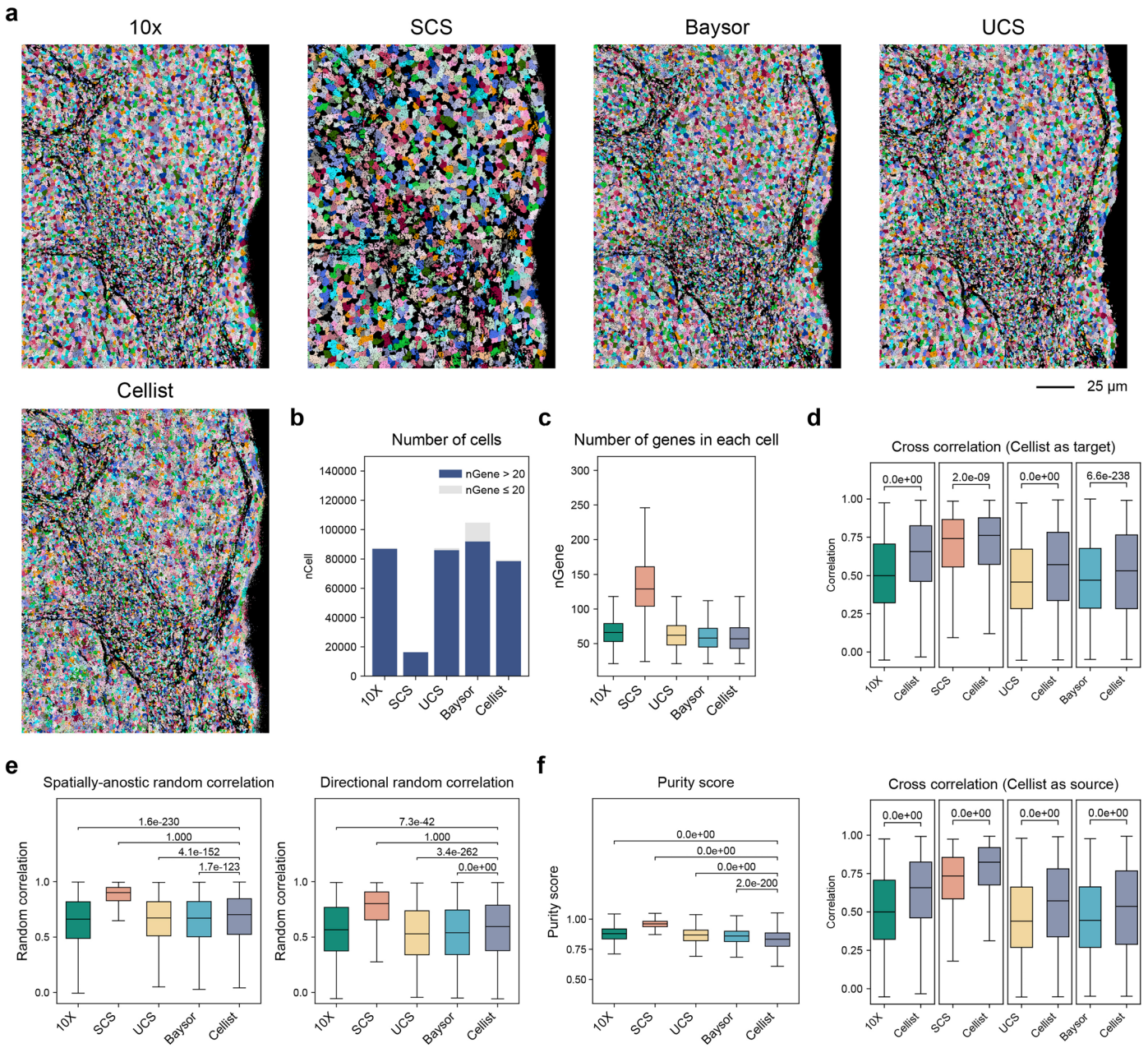


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41588-026-02610-1

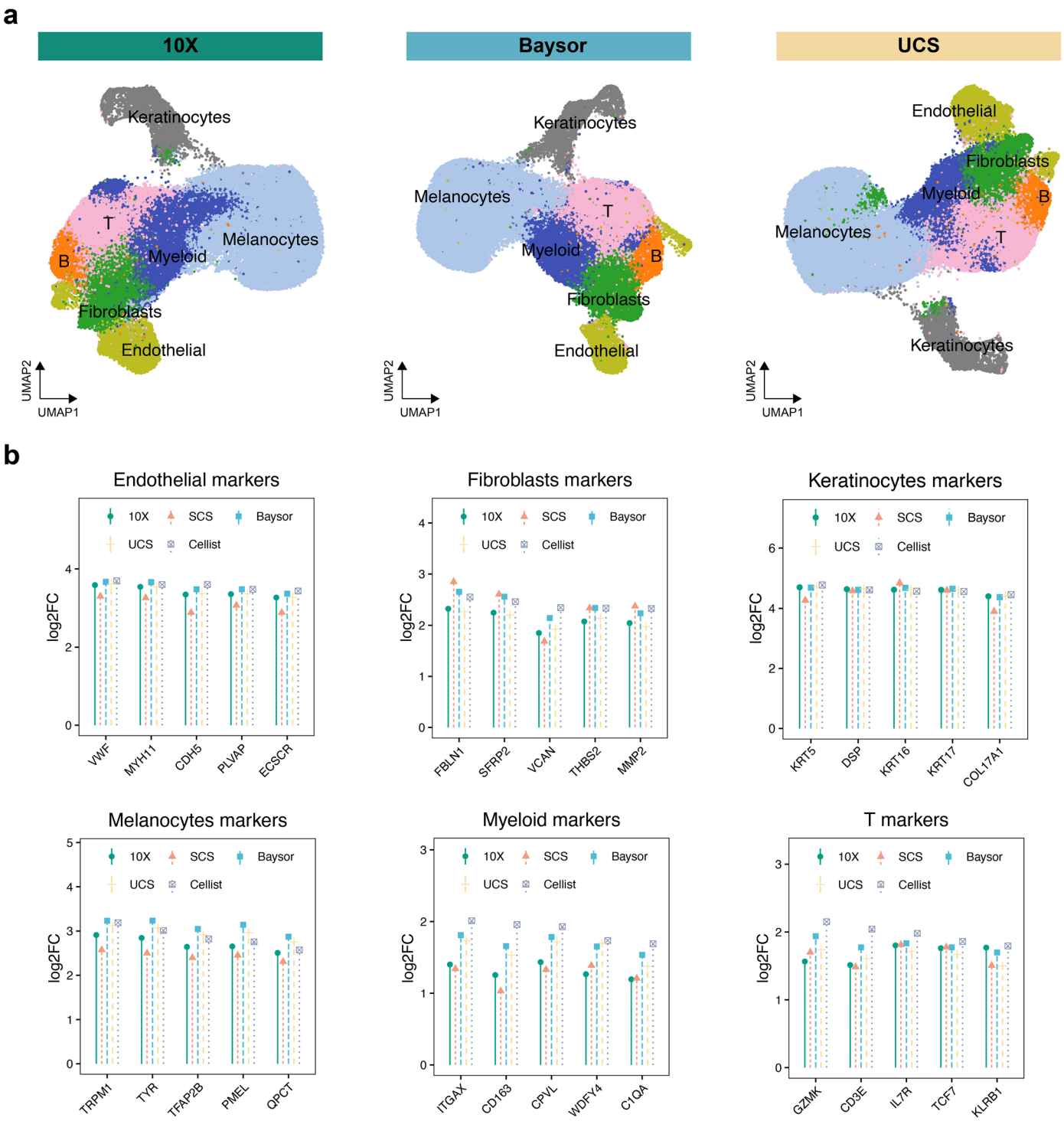
Extended Data Fig. 3 | Comparison of cell segmentation results in the STARmap mouse primary visual cortex dataset. (**a**) Cell segmentation by Expansion, SCS, UCS, Baysor and Cellist in the same FOV of the STARmap mouse primary visual cortex dataset. (**b**) Left: number of cells generated by each segmentation method. Cells with more than and less than 50 genes are colored in blue and gray, respectively. Right: distribution of the number of covered genes in cells generated by different segmentation methods. Only cells containing more than 50 genes are included in the analysis (Expansion, n = 1,550; SCS: n = 1,529; UCS: n = 1,539; Baysor: n = 1,572; Cellist: n = 1,544). (**c**) Evaluation of intracellular expression homogeneity using the metric of directional random correlation. Each cell is bisected along a randomly oriented line through its centroid. Pearson's correlation between the two subsets' pseudo-bulk expression profiles was computed to evaluate expression consistency across the two halves. Analysis is restricted to cells containing more than 50 detected genes. Statistical significance between methods was assessed using one-sided Wilcoxon rank-sum tests without adjustment for multiple comparisons. Cell numbers are the same as in (**b**). (**d**) Evaluation of intracellular expression homogeneity using the metric of purity score across different segmentation methods. Analysis is restricted to 973 nuclei shared among all methods. Statistical significance between each method and Cellist was assessed using one-sided Wilcoxon rank-sum tests without adjustment for multiple comparisons. (**e**) Evaluation of consistency between segmentations using the cross-correlation metric. Cross correlation was computed with Cellist segmentation as the source and each of other methods as the target. Statistical significance between methods was assessed using one-sided Wilcoxon rank-sum tests without adjustment for multiple comparisons (Expansion versus Cellist: n = 1,546; SCS versus Cellist: n = 10,000; UCS versus Cellist: n = 1,546; Baysor versus Cellist: n = 1,261). (**f**) UMAP plots showing the cell type annotation based on cell segmentation by different methods in the STARmap mouse primary visual cortex dataset. (**g**) Marker specificity of annotated cell types, measured by the log2FC of top DE genes in each cell type versus all other cell types. Only common genes shared across methods are used for fair comparison. Each dot represents one DE gene in a cell type from (**f**). The cell-type 'Others' is excluded from comparison. Statistical significance was assessed using paired one-sided Wilcoxon signed-rank tests without adjustment for multiple comparisons (n = 23 genes). Box plots throughout the figure indicate the median (center line), interquartile range (25^{th}-75^{th} percentiles; box), and whiskers extending to the most extreme values within 1.5× the IQR.

https://doi.org/10.1038/s41588-026-02610-1



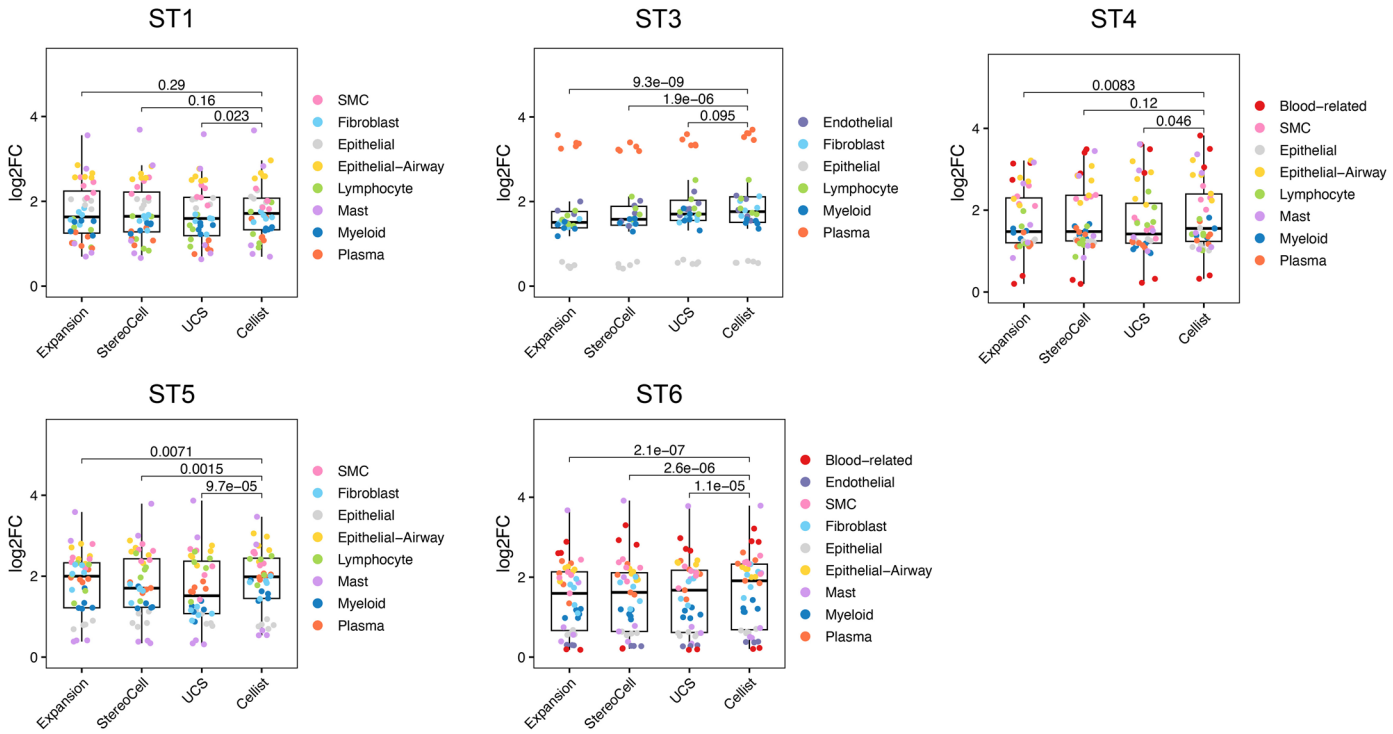
Extended Data Fig. 4 | Cell segmentation results of the 10x Xenium human melanoma dataset. (**a**) Cell segmentation by 10x Xenium Ranger, SCS, Baysor, UCS, and Cellist in the 10x Xenium melanoma dataset. (**b**) Number of cells generated by each segmentation method. Cells with more than and less than 20 genes are colored in blue and gray, respectively. (**c**) Distribution of the number of covered genes in cells generated by different segmentation methods. Only cells containing more than 20 genes are included in the analysis (10x: n = 87,000; SCS: n = 16,274; UCS: n = 86,086; Baysor: n = 91,875; Cellist: n = 78,563). (**d**) Evaluation of consistency between segmentations using the cross-correlation metric. Cross correlation was computed using Cellist segmentation as the target segmentation (top) and as the source segmentation (bottom), respectively. When Cellist is used as the target, sample sizes (number of intersecting cells) were: 10x versus Cellist (n = 72,977), SCS versus Cellist (n = 14,529), UCS versus Cellist (n = 72,992), and Baysor versus Cellist (n = 76,097). When Cellist is used as the source, sample sizes were: 10x versus Cellist (n = 72,977), SCS versus Cellist (n = 15,811), UCS versus Cellist (n = 74,277), and Baysor versus Cellist (n = 77,203). Statistical significance between methods was assessed using one-sided Wilcoxon rank-sum tests without adjustment for multiple comparisons. (**e**) Evaluation of intracellular expression homogeneity using the metrics of spatially-agnostic random correlation (left) and directional random correlation (right). Statistical significance between methods was assessed using one-sided Wilcoxon rank-sum tests without adjustment for multiple comparisons. Only cells containing more than 20 genes are included in the analysis. Cell numbers are the same as in (**c**). (**f**) Evaluation of intracellular expression homogeneity using the metric of purity score across different segmentation methods. Analysis is restricted to 16,068 nuclei shared among all methods. Statistical significance between each method and Cellist was assessed using one-sided Wilcoxon rank-sum tests without adjustment for multiple comparisons. Box plots throughout the figure indicate the median (center line), interquartile range (25^{th}-75^{th} percentiles; box), and whiskers extending to the most extreme values within 1.5× the IQR.

https://doi.org/10.1038/s41588-026-02610-1



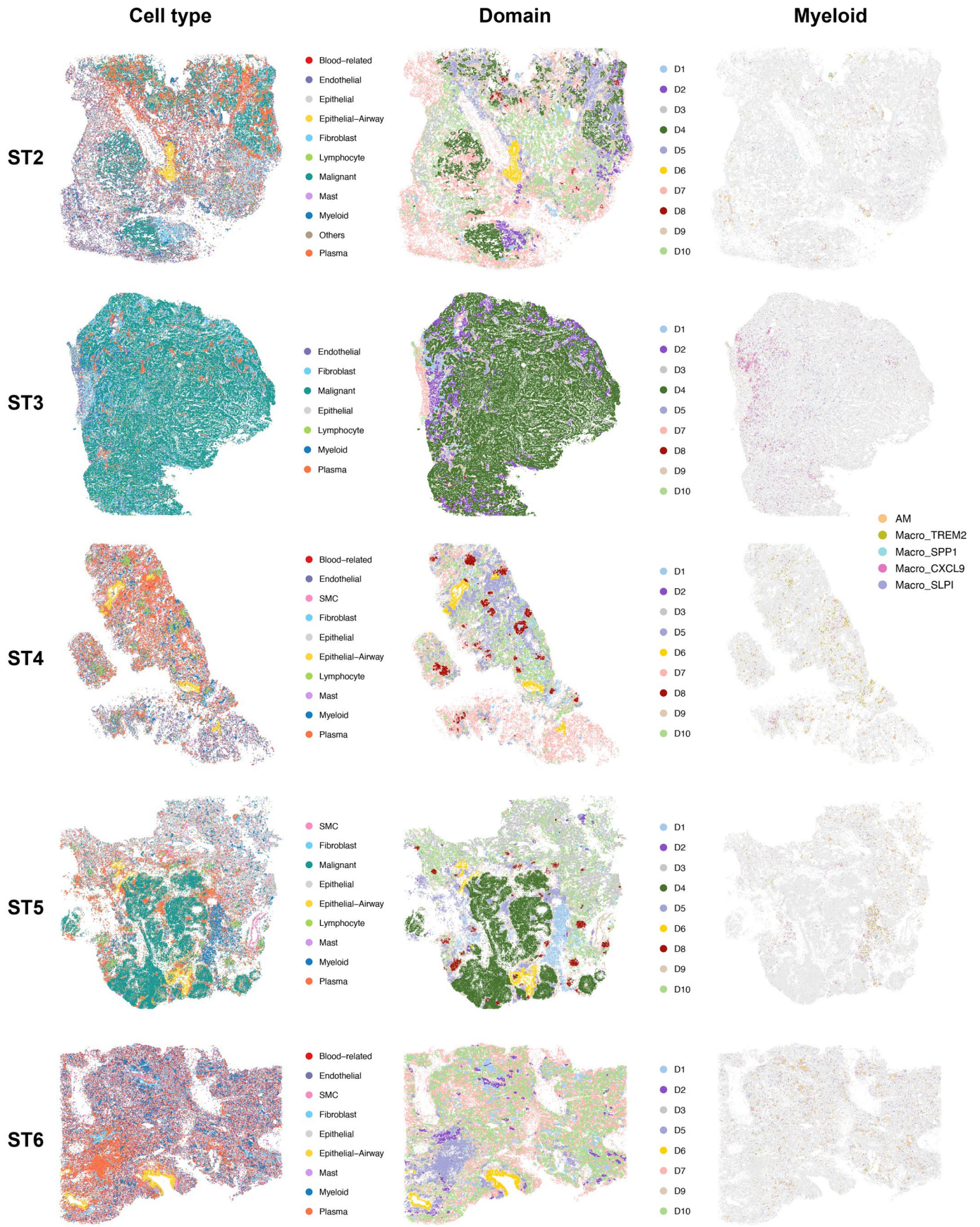
Extended Data Fig. 5 | Cell type annotation based on cell segmentation in the 10x Xenium human melanoma dataset. (**a**) UMAP plots showing the cell type annotation based on cell segmentation by 10x, Baysor and UCS in the 10x Xenium human melanoma dataset. (**b**) Lollipop plots showing the log2FC of top marker genes in each cell type versus all other cell types. Higher log2FC indicates more differential expression and thus more accurate segmentation. The top 10 DE genes with highest log2FC are selected per method per cell type, and only common genes shared across methods are used for visualization. Statistical significance between methods is shown in Main Fig. 3i.

https://doi.org/10.1038/s41588-026-02610-1



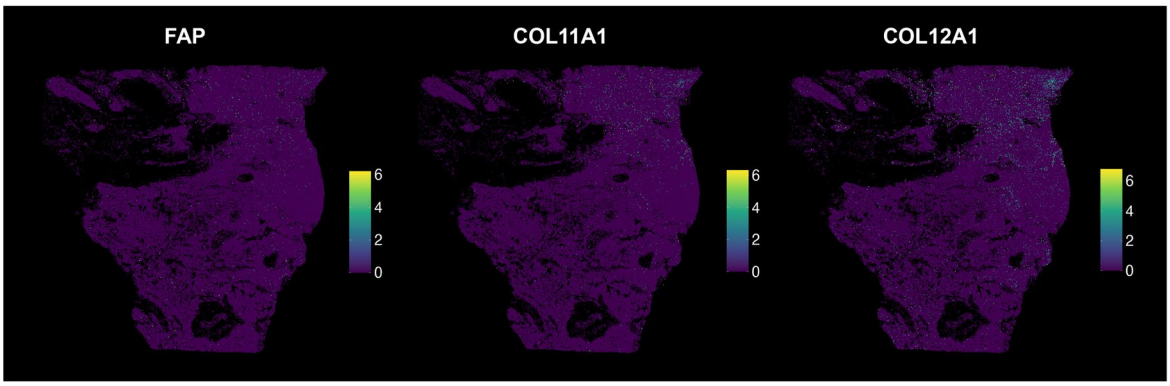
Extended Data Fig. 6 | Evaluation of cluster separation using marker specificity in each sample of the Stereo-seq human NSCLC dataset. Marker specificity of annotated cell types in each sample from the Stereo-seq human NSCLC dataset, measured by the log2-transformed fold change (log2FC) of top DE genes in each cell type versus all other cell types. Each dot represents one DE gene in a cell type. Statistical significance between methods was assessed using paired one-sided Wilcoxon signed-rank tests without adjustment for multiple comparisons (ST1: n = 40 genes; ST3: n = 30 genes; ST4: n = 40 genes; ST5: n = 40 genes; ST6: n = 43 genes). Box plots indicate the median (center line), interquartile range (25^{th}-75^{th} percentiles; box), and whiskers extending to the most extreme values within 1.5× the IQR.

https://doi.org/10.1038/s41588-026-02610-1



Extended Data Fig. 7 | Characterization of TME in post-immunotherapy NSCLC. Spatial distribution of cell types (left), spatial domains identified using SpaDo (center) and myeloid subtypes (right) in samples ST2-ST6.

https://doi.org/10.1038/s41588-026-02610-1



Extended Data Fig. 8 | Spatial expression of FAP+ fibroblast markers in ST1. The spatial expression of FAP (left), COL11A1 (center) and COL12A1 (right) in ST1.

nature portfolio

Corresponding author(s): Chenfei Wang and Peng Zhang

Last updated by author(s): Feb 14, 2026

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	No software was used for data collection.
Data analysis	The source code of Cellist (software reported in this manuscript, v1.1.1) is publicly available at https://github.com/wanglabtongji/Cellist. The following open-source packages were used to build Cellist: h5py (3.8.0), tables (3.7.0), matplotlib (3.7.3), seaborn (0.11.2), numpy (1.23.0), pandas (2.3.1), scanpy (1.9.4), scikit-learn (1.3.1), scikit-misc (0.2.0), scikit-image (0.20.0), scipy (1.9.1), spateo-release (1.0.2), pyvista (0.42), and leidenalg (0.10.2). Cellpose (v3.1.1.1) was employed to segment nuclei. SCS (version information not provided), Baysor (0.6.2), UCS (version information not provided), and StereoCell (version information not provided) were applied to segment cells and compared with Cellist. Stereopy (v1.1.0) was used for reference-based cell-type annotation in the Stereo-seq mouse brain data. For the analysis of NSCLC Stereo-seq data, ImageStudio (1.1.3) was employed to process the ssDNA images, and the SAW pipeline (6.0.0) was utilized for read filtering and mapping. CopyKAT (1.1.0) was used to estimate the copy number variations of cells in scRNA-seq and Stereo-seq. Seurat (4.1.3) and MAESTRO (1.5.1) were employed to perform conventional single-cell analysis and marker-based cell-type annotation, respectively. clusterProfiler (4.6.2) was used to perform functional enrichment analysis. Spado (1.2.0) was employed for spatial clustering.

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

The raw Stereo-seq data of human NSCLC generated in this study have been deposited in the Genome Sequence Archive for Human (GSA-Human) of the National Genomics Data Center (NGDC), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/gsa-human), under accession number HRA007118. The matched scRNA-seq data are available under accession number HRA006493. These datasets are subject to controlled access to protect patient privacy. Access can be requested through the GSA-Human portal and will be granted to researchers for non-commercial use upon approval by the Data Access Committee (DAC).
Public datasets used to evaluate the segmentation methods are listed as follows.
1. The Stereo-seq mouse brain data with ssDNA staining are available in the MOSTA data portal (https://db.cngb.org/stomics/mosta/download/) with files named 'Mouse_brain_Adult_GEM_bin1.tsv.gz' and 'Mouse_brain_Adult.tif'. The down-sampled scRNA-seq reference for cell type annotation was obtained from Stereopy following the instructions instructions provided at https://stereopy.readthedocs.io/en/latest/Tutorials/SingleR.html. The specific reference file used in this study is titled 'Mouse_brain_ref.anndata075.h5ad'.
2. The Seq-Scope mouse liver expression data can be accessed from the Gene Expression Omnibus database through the accession number GSE169706. Spatial coordinates for high-definition map coordinate identifier (HDMI) barcodes are generated using the scripts from https://github.com/leeju-umich/Cho_Xi_Seqscope. The manually-aligned H&E staining images were obtained from https://github.com/chenhcs/SCS/tree/main/data. Tiles 2104, 2105, 2106 and 2107 were used for performance evaluation in our study.
3. The seqFISH+ NIH/3T3 cell line data were downloaded from Zenodo (https://zenodo.org/records/2669683).
4. The STARmap mouse cortex data are available at https://kangaroo-goby.squarespace.com/data.
5. The Xenium human melanoma data are available at the 10x Genomics website https://www.10xgenomics.com/datasets/human-skin-preview-data-xenium-human-skin-gene-expression-panel-add-on-1-standard. The pre-designed gene panel which we used for marker-based cell-type annotation was downloaded from https://cdn.10xgenomics.com/raw/upload/v1699308550/software-support/Xenium-panels/hSkin_panel_files/Xenium_hSkin_v1_metadata.csv.
The Stereo-seq mouse olfactory bulb data used for parameter selection are requested from BGI and are available from the corresponding authors upon reasonable request. All processed data necessary to reproduce the analyses and figures in this study have been deposited in Zenodo at https://doi.org/10.5281/zenodo.18638251. Source data are provided with this paper.

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	The study collected six non-small-cell lung cancer (NSCLC) patients, including five males and one female. However, the study aims to characterize the tumor microenvironment (TME) post immuno-chemotherapy, not to identify TME differences between males and females. Thus, no gender-related analyses were performed.
Reporting on race, ethnicity, or other socially relevant groupings	Socially relevant categorization variables were not considered in the study design, and no related analyses were performed.
Population characteristics	Six patients who were diagnosed with NSCLC devoid of EGFR/ALK mutations were enrolled in this study, including five males and one female. Their ages ranged from 61 to 69 years. Among the participants, two were diagnosed at stage IIB and four at stage IIIB. All participants were treatment-naïve before recruitment and received 2-4 cycles of neoadjuvant ICB-chemotherapy (anti-PD1 and platinum-based chemotherapy), followed by surgical resection of NSCLC. Additional information about the participants is provided in Supplementary Table 3.
Recruitment	Treatment-naïve patients with stage IB-III NSCLC lacking EGFR/ALK mutations were identified in the Department of Thoracic Surgery, Shanghai Pulmonary Hospital, and recruited for this study during 2021-2023. Informed consent was obtained from all human research subjects. Because participation required informed consent and eligibility for surgery following neoadjuvant therapy, potential sources of bias include self-selection bias (patients willing to participate in research) and clinical selection bias toward individuals with operable disease and sufficient performance status to undergo combined ICB-chemotherapy and surgery. In addition, the exclusion of tumors harboring EGFR/ALK mutations restricts the cohort to a molecularly defined subset of NSCLC. These factors may limit the generalizability of findings to broader NSCLC populations, particularly patients with unresectable disease, poor performance status, or alternative oncogenic drivers. However, as the study focuses on spatial characterization of the tumor microenvironment following neoadjuvant ICB-chemotherapy, the selected cohort is appropriate for addressing the specific biological questions investigated.
Ethics oversight	All human studies were performed in accordance with ethical regulations and were pre-approved by the Ethics Committee of Shanghai Pulmonary Hospital (No. K23-310).

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	No statistical method was used to determine the patient sample size. The number of patient samples was primarily determined by the availability of tumor samples in the trials and the completion of Stereo-seq experiments.
Data exclusions	For the 10x Xenium human melanoma dataset, cells with fewer than 20 genes were excluded from subsequent analyses. For the Stereo-seq NSCLC data, cells with fewer than 50 genes were filtered out before downstream analyses. When identifying myeloid cell subtypes, cell clusters expressing marker genes of other cell types were regarded as doublets and excluded.
Replication	Stereo-seq experiments were conducted on six samples collected from six NSCLC patients, without biological or technical replicates. This was primarily due to the limited tissue samples available from clinical specimens, as well as funding limitations, which is typical in this field. All the findings can be reproduced using the provided codes.
Randomization	Previously untreated patients with stage IB-III NSCLC, lacking EGFR/ALK mutations, were recruited for this study. Previous studies have shown that the efficacy of PD-1 pathway blockade is unlikely to be associated with covariates such as age and gender. Thus, all participants received ICB-chemotherapy without randomization or grouping.
Blinding	Blinding was not applicable to this study because no sample group allocation was performed during patient treatment. Response groups (i.e., MPR or NMPR) were determined based on pathologic assessment post immuno-chemotherapy.

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	No plant data was collected or analyzed as part of this study.
Novel plant genotypes	No plant data was collected or analyzed as part of this study.
Authentication	No plant data was collected or analyzed as part of this study.