

Ultra-precision deconvolution of spatial transcriptomics decodes immune heterogeneity and fate-defining programs in tissues

Received: 25 September 2025

Accepted: 27 February 2026

Published online: 20 March 2026

 Check for updates

Yin Xu ^{1,2,3,7}, Zurui Huang ^{1,2,3,7}, Yawei Zhang^{1,2,3,7}, Minghui Gong^{1,2,3,7}, Zhenghang Wang^{4,5,7}, Peijin Guo⁶, Feifan Zhang^{1,2,3}, Jing Yang ^{1,2,3}, Guanghao Liang^{1,2,3}, Lihui Dong^{1,2,3}, Renbao Chang^{1,2,3}, Yu Xia⁶, Haochen Ni^{1,2,3}, Wenxuan Gong^{1,2,3}, Boyuan Mei⁶, Yuan Gao^{1,3}, Zhaoqi Liu ^{1,3}, Lin Shen ⁴, Jian Li ⁴, Meng Michelle Xu ⁶ ✉ & Dali Han ^{1,2,3} ✉

Elucidating the spatial organization and functional specialization of immune cells within complex tissues remains challenging. We present UCASpatial, an ultra-precision spatial transcriptomics deconvolution algorithm utilizing entropy-based weighting to accurately map cell subpopulations. Benchmarking confirms its superiority in identifying low-abundant cell subpopulations and distinguishing transcriptionally heterogeneous cell subpopulations. Applying UCASpatial to human colorectal cancer, we reveal that chromosome 20q gain in individual cancer clones orchestrates a T cell-excluded microenvironment, associated with *HERV-H* silencing and impaired type I interferon responses. In murine wound healing models, we reveal spatiotemporal dynamics distinguishing scarring from regenerative phenotypes. Specifically, we identify a pro-fibrotic community comprising *Igfbp5*⁺ chondrocytes, *Cd36*⁺ *Gpnmb*⁺ *Il1b*⁺ macrophages, and *Fmod*⁺ fibroblasts in scarring-healing mice (C57BL/6). We further demonstrate that IL11-IL11RA signaling within this triad drives the pro-fibrotic community formation and limits regeneration. Together, UCASpatial serves as a versatile tool for deciphering fine-grained cellular landscapes and exploring intercellular mechanisms in complex and dynamic microenvironments.

The ability to elucidate the spatial organization of cells within the microenvironment is crucial for understanding cellular functions and exploring intercellular mechanisms that govern development, physiology, and pathogenesis^{1,2}. Emerging spatial transcriptomics (ST) technologies offer promising avenues for achieving this; however, current spot-based techniques face a significant limitation in lacking single-cell resolution. Methods like Visium³, Tomo-Seq⁴, Slide-seq⁵, and HDST⁶ amalgamate RNA-seq measurements from multiple cells or cell

fractions at each spot. As a result, these techniques struggle with the precise identification of individual cell types within the microenvironment. Consequently, annotating cell types at each spot becomes imperative for accurately capturing spatial cell organization^{7,8}.

To address this challenge, various computational approaches have been proposed to deconvolute spatial transcriptomic profiles into cellular compositions based on gene expression profiles of each cell type identified through single-cell RNA sequencing (scRNA-seq),

A full list of affiliations appears at the end of the paper. ✉ e-mail: michelllexu@mail.tsinghua.edu.cn; handl@big.ac.cn

such as SPOTlight⁹, RCTD¹⁰, cell2location¹¹, CARD¹², stereoscope¹³, and Spotiphy¹⁴. These methods effectively delineate significant differences in cell type compositions across diverse anatomical or pathological layers in well-structured tissues such as the brain, heart, and lymphatic structures, even at fine-grained resolutions^{11,12,15,16}.

In recent years, a growing body of research has emphasized the importance of characterizing the complex microenvironment, particularly the influx of immune cells that orchestrate dynamic spatial patterns, varying between tissue homeostasis and non-homeostasis conditions. This complexity is evident in scenarios such as tissue damage^{17,18} and the tumor microenvironment^{19–23}. However, several immune cells, particularly those infiltrating into tumors, are known for their sparse presence and low transcriptional activity^{24,25}. Their discontinuous spatial distributions in these microenvironments limit the effectiveness of algorithms relying on spot-neighbor information. Additionally, the low abundance and the context-dependent phenotypic variation of heterogeneous immune cells, particularly T cells, pose challenges for conventional deconvolution algorithms²⁶. Furthermore, their scarcity makes them difficult to capture accurately using histological images, thus constraining the applicability of image-based algorithms^{27,28}. Consequently, the precise characterization of complex and dynamic microenvironments remains a significant challenge.

To address the aforementioned challenge, we developed UCASpatial, an ultra-precision cellular deconvolution algorithm for ST based on entropy-weighted regression. A unique advancement of UCASpatial is its formulation of the Shannon entropy-based weights to quantify the ability of each gene to indicate cell identity and incorporate this information in the deconvolution process. This approach improves the robustness and accuracy in identifying low-abundant cell subpopulations and distinguishing transcriptionally heterogeneous cell subpopulations in ST, thus allowing for more precise and unbiased deconvolution of the cellular landscape.

By applying UCASpatial to human colorectal cancer (CRC), UCASpatial has delineated the spatial patterns of immune cells within the tumor immune microenvironment (TIME). By integrating the fine-grained TIME landscape with spatially inferred copy number variations (CNVs), we characterize two functional multicellular communities associated with T cell-rich and T cell-excluded TIMEs at a clonal resolution. Further, we demonstrate that the chr20q-gain in CRC tumor cells is associated with reduced expression of the tumor-intrinsic human endogenous retrovirus subfamily H (*HERV-H*) and impaired type I interferon response, which orchestrates the T cell-excluded TIME phenotype. In the ear wound healing models of both regenerative MRL/MpJ (MRL) and scarring-healing C57BL/6 mice (B6 mice), we illustrate the spatiotemporal dynamics of immune cell subpopulations involved in the healing process and unveil the spatial segregation of macrophage subpopulations embedded within distinct cellular communities at the wound sites. Moreover, we identify a previously uncharacterized *Igfbp5*⁺ chondrocyte subpopulation that coordinates with *Cd36*⁺ *Gpnb*⁺ *Il1b*⁺ macrophage and *Fmod*⁺ fibroblast at the wound bed, establishing a three-cell-type community with pro-fibrotic functions linked to scarring in B6 mice. By blocking the *Fmod*⁺ fibroblast differentiation with the PPAR γ agonist rosiglitazone, or by inhibiting the IL11-IL11RA axis within the community in B6 mice, we observed a reduction in the formation of the pro-fibrotic community and an improvement in the regenerative ability of ear wounds in B6 mice.

Results

UCASpatial: an entropy-weighted regression model for spatial deconvolution of cellular landscapes

UCASpatial deconvolutes ST data into cellular compositions by leveraging cell identity-specific gene expression profiles (cGEPs) derived from either paired or unpaired scRNA-seq data (Fig. 1). First,

UCASpatial extracts cGEPs from the referenced scRNA-seq dataset with a set of user-provided cell-subpopulation labels (Supplementary Fig. S1A–C). It strategically selects cells meeting a meta-purity threshold to enhance discrimination of referenced cell identities, and then extracts cGEPs through optimized non-negative matrix factorization (NMF). Second, UCASpatial formulates Shannon entropy-based weights for each gene within the extracted cGEPs to enhance genes indicative of cell identities (see “Methods” section). Third, UCASpatial employs weighted non-negative least squares (WNNLS) to deconvolute the mRNA counts of each spot in the ST data using the weighted cGEPs, thereby estimating the cellular compositions. Additionally, UCASpatial offers a suite of downstream analysis tools. These include identifying tumor clones and their corresponding TIME, as well as spatially resolved quantification of transposable elements (TEs). To ensure versatility in applications, UCASpatial is seamlessly integrated with the Seurat framework.

Benchmarking the performance of UCASpatial in the simulated ST dataset

To assess UCASpatial's performance, we generated a simulated ST dataset comprising 750 ST spots designed to mimic the real cellular composition of the tumor microenvironment. Each spot was constituted by the transcripts of 5 to 20 cells sampled from scRNA-seq data of human CRC, encompassing 20 reference cell subpopulations (Supplementary Fig. S1A, B, and “Methods” section). The cellular composition of simulated spots was designed to recapitulate the spatially correlated cell-type distribution patterns observed in tumor microenvironments, particularly the relationship between epithelial, stromal, and immune cell frequencies across tumor core (TC), margin, and stromal regions (Supplementary Fig. S1D and Supplementary Table 1)^{29,30}.

To assess the performance of UCASpatial on the simulated ST dataset with varying complexity, we ranked the complexity of these simulated spots based on their cellular composition entropy. Subsequently, we categorized the simulated spots into three groups (low, medium, and high complexity), where higher entropy denotes increased complexity (Fig. 2A and Supplementary Fig. S1D). We compared UCASpatial with several recently established methods, including SPOTlight, RCTD, cell2location, CARD, stereoscope, and Spotiphy, using Root Mean Square Error (RMSE) to first quantify the deviation between estimations and ground truth at the spot level (see “Methods” section). UCASpatial exhibited significantly superior performance across all complexity levels (median RMSE = 0.051) compared to the aforementioned methods, with 10%, 21%, 31%, 31%, 36%, and 35% improvement in terms of RMSE compared to RCTD (0.058), SPOTlight (0.066), CARD (0.075), Spotiphy (0.075), cell2location (0.081), and stereoscope (0.080), respectively (Fig. 2B and Supplementary Fig. S1E). Additionally, UCASpatial demonstrated a comparable Pearson Correlation Coefficient (PCC) to cell2location and significantly outperformed other methods, when comparing estimations with the ground truth (Supplementary Fig. S1F). Moreover, we also assessed the ability to identify the existence of cell subpopulations within spots using the F1 score. Consistently, UCASpatial significantly outperformed other methods across all complexity levels with the highest F1 score (with a 6–25% improvement) (Fig. 2B and Supplementary Fig. S1G).

Additionally, to assess the robustness of UCASpatial across different cell numbers, we categorized spots into three groups based on the number of cells per spot: 5–9 cells, 10–12 cells, and 13–20 cells (Supplementary Fig. S1H). Our analysis revealed that UCASpatial maintained robust performance across these varying cell numbers, whereas the performance of cell2location and stereoscope dropped as cell number increased (Supplementary Fig. S1I). These results highlight UCASpatial's robustness in handling diverse cell densities, thereby confirming its adaptability and reliability.

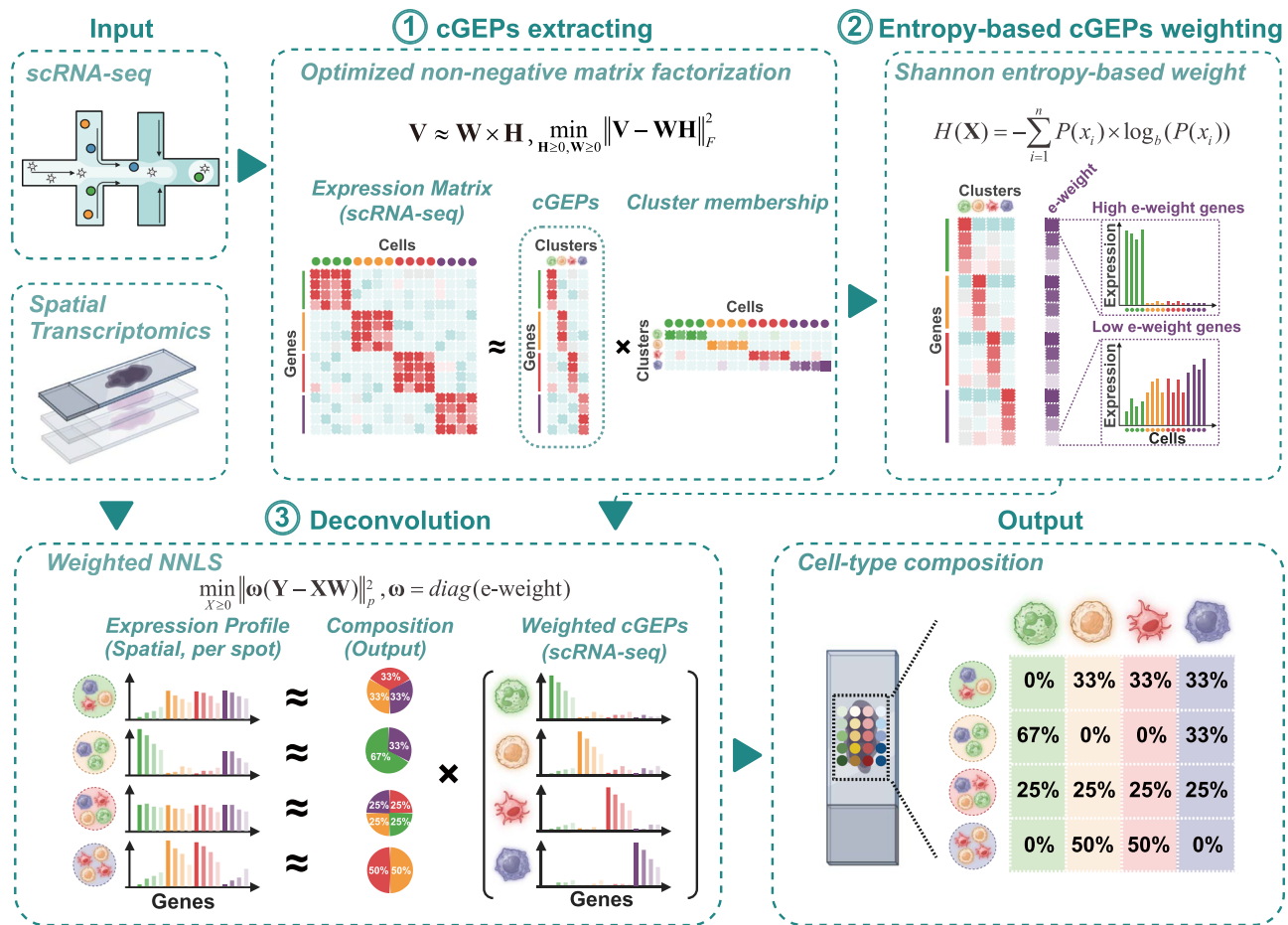


Fig. 1 | Schematic overview of UCASpatial. UCASpatial takes both the reference scRNA-seq data with cluster information and spatial transcriptomic data as input. After three-step processing involving cell-identity-specific GEP (cGEP) extraction, entropy-based cGEPs weighting, and deconvolution, UCASpatial decomposes

spatially resolved, multi-cell RNA count matrices (i.e., spatial transcriptomic data) into the corresponding cell-type composition. GEP, gene expression profile. e-weight, Shannon entropy-based weight. NNLS non-negative least squares. Created in BioRender. Zhang, Y. (2026) <https://BioRender.com/lund1gg>.

We next evaluated the precision and robustness of UCASpatial in deconvoluting different cell subpopulations. Through in-silico analyses using three distinct reference scRNA-seq labels with varying numbers of subpopulations (7, 11, and 20), we observed that the advantages of UCASpatial became more pronounced as the number of subpopulations increased compared with other methods (Supplementary Fig. S1J, K). With 20 subpopulations, UCASpatial exhibited superior performance across all metrics, including RMSE, PCC, and F1 score, highlighting its proficiency in deciphering fine-grained spatial landscapes (Supplementary Fig. S1K). Specifically, examination of PCC between estimations and ground truth for each method under 20 subpopulations conditions unveiled UCASpatial's superior performance, achieving the highest PCC in 15 out of 20 cell subpopulations (Fig. 2C), particularly notable for NK cell and CD8⁺ T cell subpopulations (with a 39–187% improvement) (Fig. 2D and Supplementary Fig. S1L). By categorizing the PCC values of each cell subpopulation into four groups, we noticed UCASpatial exhibited PCC values surpassing the medium–high correlation level (PCC > 0.5) for 19 out of 20 cell subpopulations, with the sole exception of CD4⁺ Tfh (PCC = 0.47), while remaining the best compared to alternative methods (Fig. 2C, E). In contrast, other methods yielded PCC values below the medium–high correlation level for at least 30% of cell subpopulations, with some even falling below 0.25 (Fig. 2E and Supplementary Fig. S1M). This comparison underscored UCASpatial's capability in discriminating cell subpopulations, even with subtle transcriptomic differences (Supplementary Fig. S1N). Furthermore, assessments using

the PCC and F1 score reaffirmed UCASpatial's precise and robust decomposition across each cell subpopulation (Supplementary Fig. S1O, P).

Considering the challenge of ST analysis in precisely mapping the spatial location of low-abundant cell subpopulations, we next investigated UCASpatial's sensitivity in annotating cell subpopulations with low frequencies within spots. We categorized the proportion of each cell subpopulation within simulated spots into ten groups, forming a gradient that spanned from 0–10% to 100% (Fig. 2F). Subsequently, we calculated the RMSE across various proportions to evaluate the deviation of cell-subpopulation proportions. We observed that UCASpatial exhibited the lowest overall RMSE (averaged by each cell subpopulation, medium RMSE = 0.051) across different proportions, with a 7–31% improvement over the other methods (Fig. 2G). Moreover, we specifically examined the RMSE of different cell subpopulations, revealing that UCASpatial notably outperformed other methods for cell subpopulations with subtle transcriptomic differences, such as CD8⁺ Tex and CD8⁺ Tem (with 25–63% and 13–26% improvements, respectively) (Fig. 2G and Supplementary Fig. S1Q). Consistent results were observed in assessments using the PCC and F1 score (Supplementary Fig. S1R), highlighting UCASpatial's superiority in accurately annotating cell subpopulations with low frequencies within spots, even with subtle transcriptomic differences.

Furthermore, we evaluated the performance enhancement afforded by UCASpatial's entropy-based weighting and meta-purity filtering

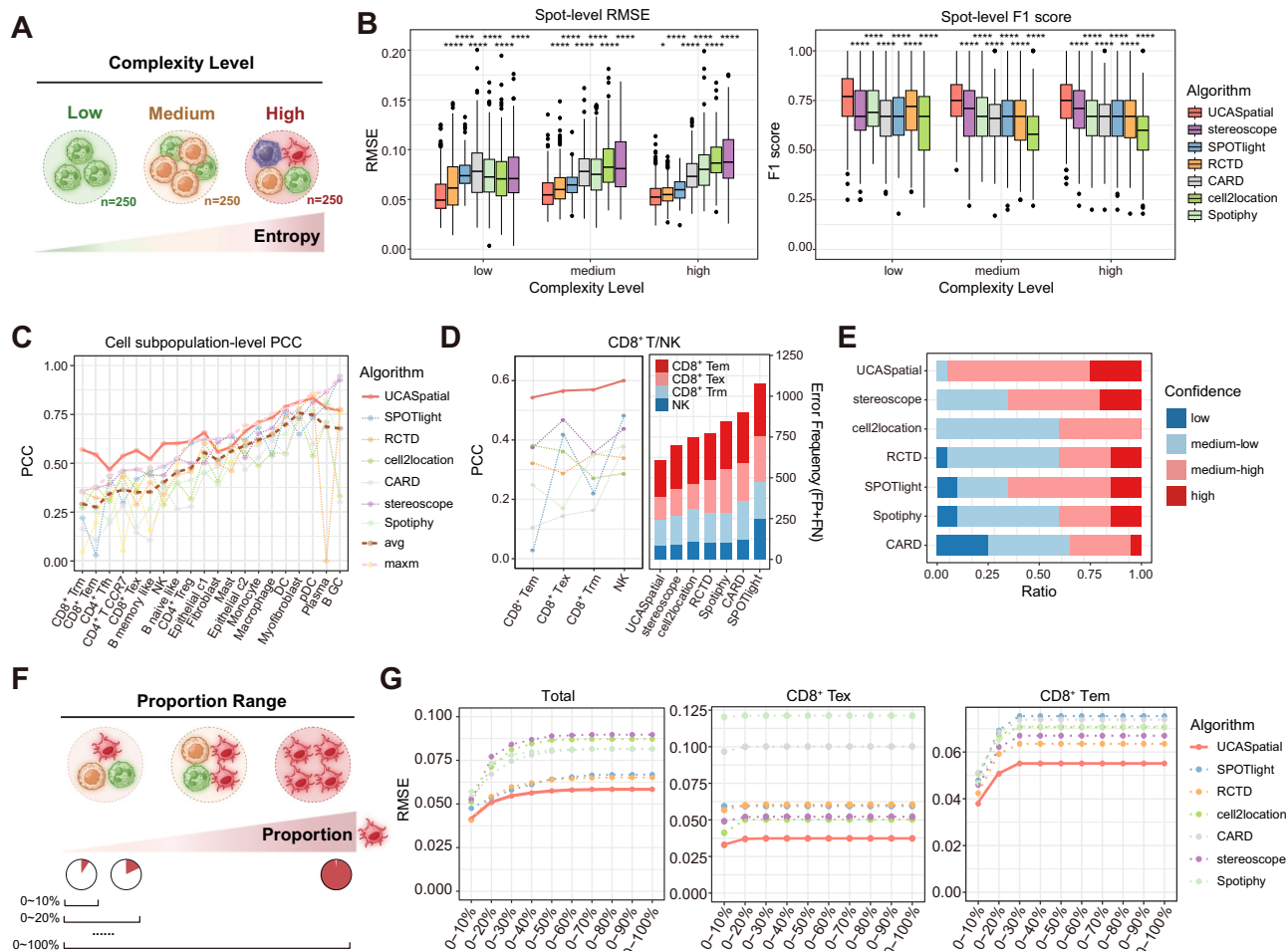


Fig. 2 | Validation of UCASpatial using simulated data. **A** Schematic showing 750 simulated spots stratified into three complexity levels (low, medium, high; $n = 250$ each) by cell-type composition entropy. Created in BioRender. Zhang, Y. (2026) <https://BioRender.com/9an18yc>. **B** Boxplots showing spot-level RMSE and F1 score for each method across three complexity levels ($n = 250$ spots per complexity level). Boxplot elements are defined as: center line, median; box limits, upper and lower quartiles; whiskers, $1.5 \times$ interquartile range; points, outliers. Statistical analysis: two-tailed paired t -test. All comparisons between UCASpatial and other methods yielded $p < 0.001$, except for RCTD in the high-complexity group for RMSE ($p = 0.026$). ns non-significant, $p > 0.05$; $0.01 \leq p < 0.05$; $^*p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$. **C** Line charts showing cell-subpopulation-level PCC for each method across 20 subpopulations. 'avg' and 'maxm' represent the average and maximum PCC, respectively, of the six methods excluding UCASpatial. **D** Left, line charts representing cell-subpopulation-level PCC of CD8⁺ effector memory T cells (CD8⁺ Tem), CD8⁺ exhausted T cells (CD8⁺ Tex), CD8⁺ tissue-resident memory T cells (CD8⁺ Trm), and natural killer (NK) cells for each method. Right, stacked bar plot showing the accumulated count of false predictions of these subpopulations for each method. Colors represent different subpopulations. FP false positive, FN false negative. **E** The cell-subpopulation-level PCC values for each method were binned into four confidence groups: 0–0.25 (low), 0.25–0.5 (medium–low), 0.5–0.75 (medium–high), and 0.75–1 (high). Stacked bar plot showing the ratio of four confidence groups for each method. **F** Schematic showing simulated spots stratified into ten groups by subpopulation proportions, spanning 0–100% in 10% increments. Created in BioRender. Zhang, Y. (2026) <https://BioRender.com/4992rj4>. **G** Left, line charts representing the average RMSE of 20 subpopulations across various proportions for each method. Middle/Right, line charts showing an example of the RMSE for CD8⁺ Tex and CD8⁺ Tem across different proportions. Each dot in the line charts represents the average RMSE from multiple simulations. All exact p -values and sample sizes are provided in the Source Data. Source data are provided as a Source Data file.

designs (Supplementary Fig. S1S–U). We confirmed that the meta-purity filtering process did not result in the loss of any predefined cell subpopulations, with PCC values before and after filtering consistently exceeding 0.9 (Supplementary Fig. S1V–X). We found that the entropy-based weighting design showed improvements of 22%, 50%, and 17% in terms of RMSE, PCC, and F1 score, respectively, compared to UCASpatial without weighting. The meta-purity filtering design exhibited improvements across different parameters with 4–11%, 1–8%, and 1–3% in terms of RMSE, PCC, and F1 score. In addition, the same analyses as before showed that UCASpatial, without entropy-weighting and meta-purity filtering, lost the aforementioned improvements (Supplementary Fig. S1E–G, I, K–M, O–R).

We evaluated the runtime and memory usage of UCASpatial and other methods (see “Methods” section), analyzing the impact of the

number of spots in spatial data and the number of cells in scRNA-seq data on computational resources. UCASpatial showed stable time costs, processing data in under 1 h as the number of spots increased (Supplementary Fig. S1Y). Cell2location encountered memory errors on our GPU platform with the largest dataset (50,000 spots and ~12,500 cells). Other methods, except the stereoscope, were minimally affected by the scRNA-seq reference size (Supplementary Fig. S1Y). Using a simulated dataset (50,000 spots, ~12,500 cells, 20 cell subpopulations), UCASpatial demonstrated the lowest memory usage (<26GB) and highest efficiency (Supplementary Fig. S1Z). Collectively, our in-silico benchmarking demonstrates that UCASpatial enables precise and sensitive spatial deconvolution, excelling in discerning cell subpopulations with subtle transcriptomic differences, even in low proportions.

UCASpatial demonstrates broad applicability and high precision across multi-resolution platforms

To test whether UCASpatial is applicable across platforms with varying resolutions, we applied UCASpatial to 10x Visium HD data from a CRC sample³¹, using the recommended bin size of $8\mu\text{m} \times 8\mu\text{m}$ (Supplementary Fig. S2A and see “Methods” section). We compared UCASpatial with RCTD, a method also suitable for processing 10x Visium HD data using PCC between estimated proportions of cell subpopulations and marker-gene scores in the absence of ground truth¹⁶. We observed that UCASpatial achieved the highest PCC, and exhibited over 20 times faster than RCTD on the runtime evaluation (Supplementary Fig. S2B–D, see “Methods” section).

We also analyzed Slide-seq V2 data from the mouse hippocampus to evaluate UCASpatial’s performance at single-cell resolution (Supplementary Fig. S2E–H). Since strictly defined ground-truth cell type labels are not available experimentally, we adopted the widely accepted evaluation strategy established by the CARD paper¹². This involves verifying whether the deconvolved cell types accurately reconstruct known anatomical structures (e.g., CA1, CA3, Dentate Gyrus). Specifically, we visualized the predicted top 1 cell type for each location (Supplementary Fig. S2F) and quantitatively evaluated the concordance between method-predicted cell-type proportions and known region-specific marker expression in three anatomical structures¹² by calculating PCC (Supplementary Fig. S2G). Our results show that UCASpatial is highly concordant with other methods, successfully captures key anatomical structures, and confirms a comparable level of accuracy and spatial resolution on this neurobiological dataset (Supplementary Fig. S2G, H).

Together, these results demonstrate UCASpatial’s broad applicability and high precision across platforms with varying resolutions, from bin resolution (10x Visium HD) to single-cell resolution (Slide-seq V2).

Benchmarking the performance of UCASpatial using Visium HD-derived pseudo-spots

To obtain synthetic data that better capture spatial correlation structures between neighboring spots and cell-subpopulation-specific capture efficiency variations, we generated synthetic spots by spatially aggregating segmented single-cell transcriptomes from three 10x Genomics Visium HD CRC tissue sections³¹ into $49\mu\text{m} \times 49\mu\text{m}$ bins (Supplementary Fig. S2I–K), closely matching standard Visium spot size. This approach preserves local spatial correlations and yields 39,813 pseudo-spots with ground-truth cell-subpopulation composition reflecting the tumor microenvironment’s complexity.

Benchmarking across this dataset recapitulated prior performance trends. UCASpatial achieved the highest average PCC (0.87) and F1 score (0.82), alongside the lowest RMSE (0.05) among other methods at the spot level (Supplementary Fig. S2L), even under scenarios of varying complexity (Supplementary Fig. S2M). Cell-subpopulation-level evaluation further confirmed UCASpatial’s superior accuracy in deciphering different cell subpopulations (Supplementary Fig. S2L, N). This data further supports UCASpatial’s accurate and stable performance on biologically realistic synthetic data.

UCASpatial accurately annotates cell subpopulations in human CRC TIMES

Spatial omics has provided a powerful tool for dissecting the intricate interplays between tumor alterations and the co-evolving immune microenvironment^{21,32,33}. However, the characterization of immune cell compositions, particularly those infiltrating into distinct tumor clone regions, remains a formidable challenge due to their low abundance and limited transcriptional output^{21,25,34}. To investigate whether UCASpatial can precisely resolve the TIME, we curated an ST dataset from eight tumor slices of seven patients with human CRC using the

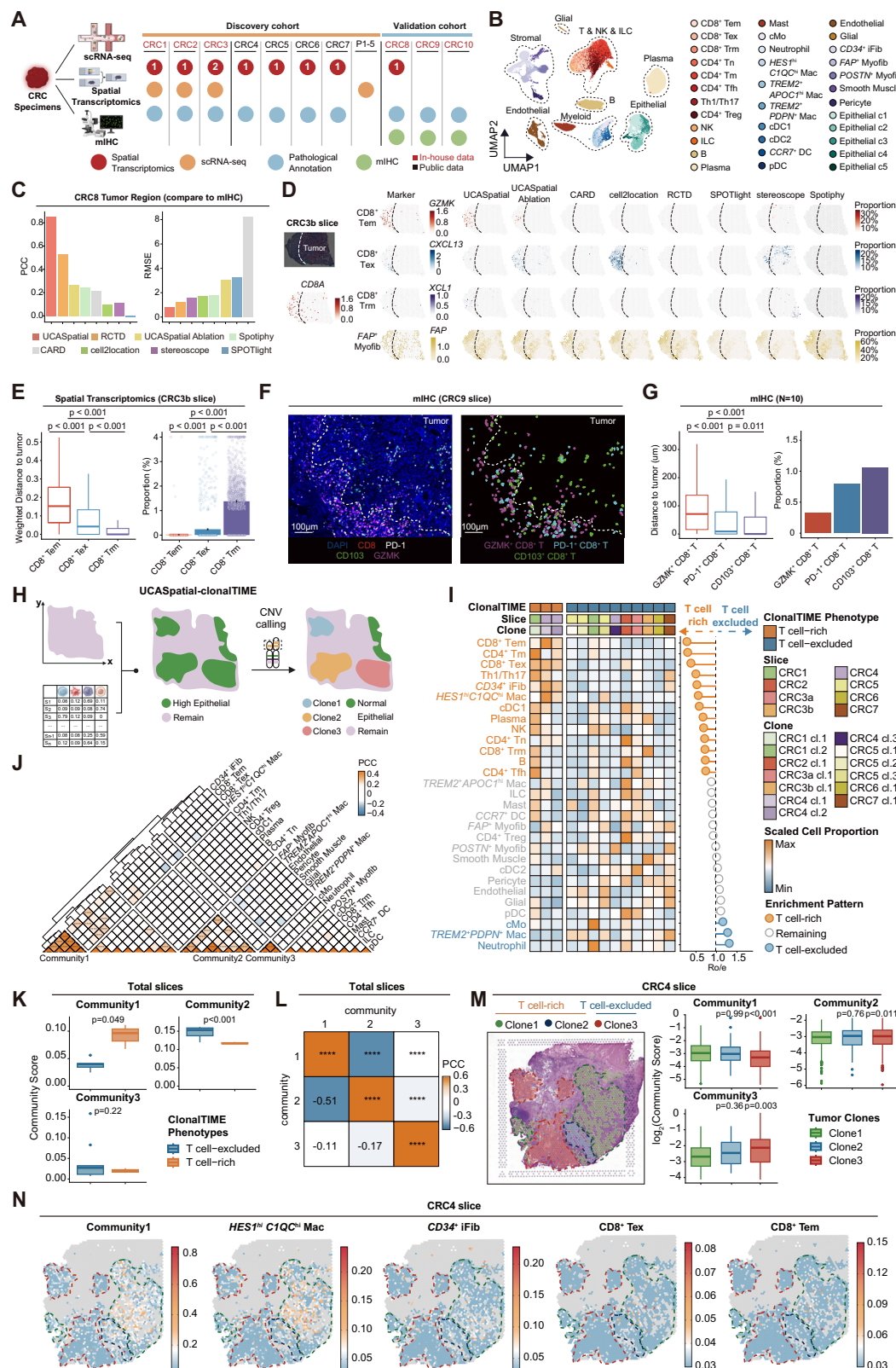
10x Genomics Visium platform as the discovery cohort³⁵ (Fig. 3A and Supplementary Fig. S3A, and Supplementary Table 2). We leveraged an integrated scRNA-seq dataset comprising an in-house scRNA-seq with three patients, and a public scRNA-seq dataset with five patients (P1–5)³⁵, encompassing 34 cell subpopulations as a reference (Fig. 3A, B and Supplementary Fig. S3B–D and see “Methods” section).

We annotated the spatial distribution of 34 cell subpopulations in the human CRC ST discovery cohort using UCASpatial. Notably, stromal cell subpopulations showed consistent distribution with anatomical locations as expected from prior literature^{36,37}, such as smooth muscle cells of the muscular layer and endothelial cells around blood vessels, along with pericytes (Supplementary Fig. S3E). Similarly, epithelial cell subpopulations displayed enrichment in the epithelial region, consistent with corresponding H&E staining and pathological annotations (Supplementary Fig. S3F). Upon further examination of immune subpopulations, we noticed B cell and various T cell and dendritic cell (DC) subpopulations, including $\text{CD4}^+ \text{Tn}$, $\text{CD4}^+ \text{Tfh}$, $\text{CD8}^+ \text{Tex}$, $\text{CD8}^+ \text{Tem}$, cDC2 , and $\text{CCR7}^+ \text{DC}$, formed a multicellular community resembling a tertiary lymphoid structure (TLS) (Supplementary Fig. S3G and see Supplementary Table 3 for abbreviation information)^{38,39}. Additionally, we noticed the spatial exclusivity of $\text{HES}^{\text{hi}} \text{CIQC}^{\text{hi}}$ macrophages and $\text{TREM2}^+ \text{PDPN}^+$ macrophages (Supplementary Fig. S3H), indicating their distinct functions in TIMES.

For comparative analysis, we also annotated this ST dataset using other methods. Remarkably, UCASpatial achieved the highest average PCC with a significant improvement of 17–104% over other methods between the cell-subpopulation proportions and signature expressions, especially in T cell subpopulations (with at least a 26% improvement) (Supplementary Fig. S3I, J, Supplementary Data 1, and “Methods” section). In alignment with in-silico benchmarking results, all evaluated methods correctly annotated structural cell subpopulations, for example, FAP^+ myofibroblast (Supplementary Fig. S3J), yet UCASpatial showed a higher sensitivity in annotating immune cell subpopulations (Supplementary Fig. S3I, J). For instance, UCASpatial performed better than other methods in identifying the spatial distribution of T cell subpopulations in almost all five TLS regions (Supplementary Fig. S3K). While UCASpatial depicted 4–6 T cell subpopulations consistent with their marker gene expressions, all compared methods only identified 0–2 T cell subpopulations (Supplementary Fig. S3L).

To assess the accuracy of immune cell annotation in the real dataset, we constructed the validation cohort. We first performed ST on an additional human CRC tumor sample (CRC8) and validated the ST results using mIHC (multiplexed immunohistochemistry) staining on consecutive slices from the same tissue block utilized for ST (Fig. 3A and Supplementary Fig. S4A–C, and see “Methods” section). The ST data were annotated using unpaired scRNA-seq data obtained from independent CRC datasets (Fig. 3B)³⁵. Quantitative analysis of cellular proportions via mIHC revealed that among 12 validated immune subpopulations, the majority of them had an average proportion of less than 10% within the pseudo-spots (see “Methods” section), with T cell subpopulations exhibiting the lowest proportions (Supplementary Fig. S4D). In alignment with the results from mIHC, UCASpatial correctly annotated the spatial distribution of these cell subpopulations (Fig. 3C and Supplementary Fig. S4E–H). We further compared the cellular proportions estimated by other computational algorithms with those quantified by mIHC in tumor regions. Our analysis demonstrated that UCASpatial consistently achieved the highest PCC and the lowest RMSE for these immune cell subpopulations, particularly for T cell subpopulations, in tumor regions (Fig. 3C and Supplementary Fig. S4E–H). These findings underscore the capability of UCASpatial in accurately identifying low-abundant cell subpopulations, even in complex tissue contexts.

We found that the spatial distribution patterns of three CD8^+ T cell subpopulations at the tumor-invasive front were distinct: $\text{CD8}^+ \text{Tex}$



and CD8⁺ Trm were more proximal to the tumor, whereas CD8⁺ Tem was relatively distal (Fig. 3D, E). Also, there was a higher proportion of CD8⁺ Tex and CD8⁺ Trm infiltrated into the tumor region than CD8⁺ Tem (Fig. 3E). To validate these findings, we performed miHC on human CRC samples from two additional patients (CRC9 and CRC10) in the validation cohort (Fig. 3A) and quantified the distance of each CD8⁺ T cell subpopulation to the tumor. This analysis revealed that PD-

1⁺ CD8⁺ T cells and CD103⁺ CD8⁺ T cells were positioned closer to the tumor compared to GZMK⁺ CD8⁺ T cells (Fig. 3F, G). Further calculation of the proportion of each CD8⁺ T cell subpopulation within the tumor region supported that PD-1⁺ CD8⁺ T cells and CD103⁺ CD8⁺ T cells exhibited higher tumor infiltration proportions compared to GZMK⁺ CD8⁺ T cells (Fig. 3F, G). Additionally, we noticed UCASpatial without entropy-weighting and meta-purity filtering was unable to either

Fig. 3 | Characteristics of clonal-resolution landscape with two TIME phenotypes in human CRC. **A** Experimental design and overview of the CRC dataset for UCASpatial analysis. Created in BioRender. Zhang, Y. (2026) <https://BioRender.com/jrmv6>. **B** UMAP of 34 cell subpopulations from scRNA-seq reference. **C** Bar plots displaying PCC and RMSE between each method and the mIHC results in the tumor regions of the CRC8 slice. **D** Pathological annotations, marker expression, and spatial distribution of subpopulations in CRC3b; dashed lines indicate the tumor-invasive front. **E** Left, boxplot of distances between CD8⁺ T cell subpopulations and tumor regions in CRC3b; $n = 72$ (CD8⁺ Tem), 103 (CD8⁺ Tex), and 164 (CD8⁺ Trm) spots. Right, proportions of CD8⁺ T cell subpopulations within CRC3b tumor regions ($n = 1320$ spots). Data shown as mean $\pm$ SEM. **F** Immunofluorescence (left) and spatial distribution (right) of CD8⁺ T cell subpopulations in CRC9. Scale bar, 100 μ m. **G** Left, boxplot showing distances between CD8⁺ T cell subpopulations and tumor regions in 10 ROIs (region of interest). Right, bar plot showing the proportion of CD8⁺ T cell subpopulations within tumor

regions in 10 ROIs. **H** Schematic of the UCASpatial-clonalTIME platform. Created in BioRender. Zhang, Y. (2026) <https://BioRender.com/Odc25h4>. **I** Cell subpopulation distribution across clones (heatmap) and Ro/e scores (lollipop plot) grouped by clonalTIME phenotypes. Ro/e, ratio of observed over expected cell subpopulation proportions. **J** Heatmap of PCC between subpopulation proportions across tumor spots after neighborhood processing. **K** Community scores compared between T cell-excluded ($n = 10$ clones) and T cell-rich ($n = 3$ clones) clonalTIMES. **L** Heatmap displaying the PCC between community scores of spots within tumor regions. **M** Spatial distribution of clones in CRC4 (Left; $n = 1089$ (clone1), 171 (clone2), and 749 (clone3)) and corresponding community scores per clone (Right; boxplots). **N** Spatial visualization of community1 scores and involved cell proportions in CRC4. Statistical tests: two-tailed Wilcoxon rank-sum (**E**, **G**); two-tailed unpaired t -test (**K**, **M**); Pearson correlation analysis (**J**, **L**). **** $p < 0.0001$. Boxplots: center, median; box, quartiles; whiskers, 1.5 $\times$ IQR; points, outliers. Source data are provided as a Source Data file.

identify T cell subpopulations in TLS regions or discriminate the spatial distribution of tumor-infiltrating CD8⁺ T cell subpopulations (Fig. 3D, Supplementary Fig. S3I–L, and S4E–H). These results demonstrate that UCASpatial can accurately resolve cell subpopulations in the complex TIME.

Identification of functional multicellular communities associated with T cell-rich and T cell-excluded clonalTIMES in CRC

Tumor-intrinsic genomic alterations have collateral effects that modify the spatial organization of the TIMES^{40–42}. To spatially characterize diverse TIMES at a detailed clonal resolution, we generated an integrated immunogenomics platform termed ‘UCASpatial-clonalTIME’ to spatially link tumor genomics with the TIMES. This platform enables the identification of tumor clones within tissue slices by inferring spatial CNVs³³ and concurrently estimating the composition of the TIME for each clone (clonalTIME) (Fig. 3H and Supplementary Fig. S4I, and “Methods” section). By assessing the proportions of CD4⁺ and CD8⁺ T cell subpopulations, we categorized the clonalTIMES into two phenotypes: T cell-rich and T cell-excluded (Fig. 3I and Supplementary Fig. S4J, and “Methods” section). We detected three out of thirteen tumor clones displaying the T cell-rich clonalTIME phenotype (Fig. 3I and Supplementary Fig. S4J–L). Within these T cell-rich clonalTIMES, all three subpopulations of CD8⁺ T cells—CD8⁺ Tem, CD8⁺ Trm, and CD8⁺ Tex—were notably enriched (Fig. 3I). Analysis of the CD4⁺ T cell compartment showed significant enrichments of CD4⁺ Tm and Th1/Th17 within these T cell-rich clonalTIMES (Fig. 3I). Conversely, within T cell-excluded clonalTIMES, we found a depletion of all T cell subpopulations. Instead, there was an enrichment of classical monocytes (cMo), neutrophils, and *TREM2*⁺ *PDPN*⁺ macrophages (Fig. 3I).

Immune cells within the TIMES engage in crosstalk with non-immune cells that coordinately form multicellular communities. To identify these communities associated with two clonalTIME phenotypes, we calculated the PCC between the neighborhood-averaged proportions of various cell subpopulations, including 22 immune subpopulations and 7 non-immune subpopulations, across all spots from the 13 clonalTIMES (see “Methods” section). Through hierarchical clustering followed by functional enrichment analysis (see “Methods” section), we observed three multicellular communities discriminated by different cellular compositions and specialized functionalities: (1) Community1 encompassed *HES1*^{hi} *CIQC*^{hi} macrophages, CD8⁺ Tem, CD8⁺ Tex and inflammatory *CD34*⁺ fibroblast (*CD34*⁺ iFib) (Fig. 3J and Supplementary Fig. S4M), with enriched pathways in leukocyte mediated cytotoxicity and homeostasis (Supplementary Fig. S4N); (2) Community2 comprised endothelial cells, pericytes, *TREM2*⁺ *PDPN*⁺ macrophages, *TREM2*⁺ *APOC1*^{hi} macrophages, *FAP*⁺ myofibroblasts, smooth muscle cells, and glial cells (Fig. 3J), with functional pathways related to angiogenesis, wound healing, and WNT signaling (Supplementary Fig. S4N); (3) Community3, including neutrophils and cMos

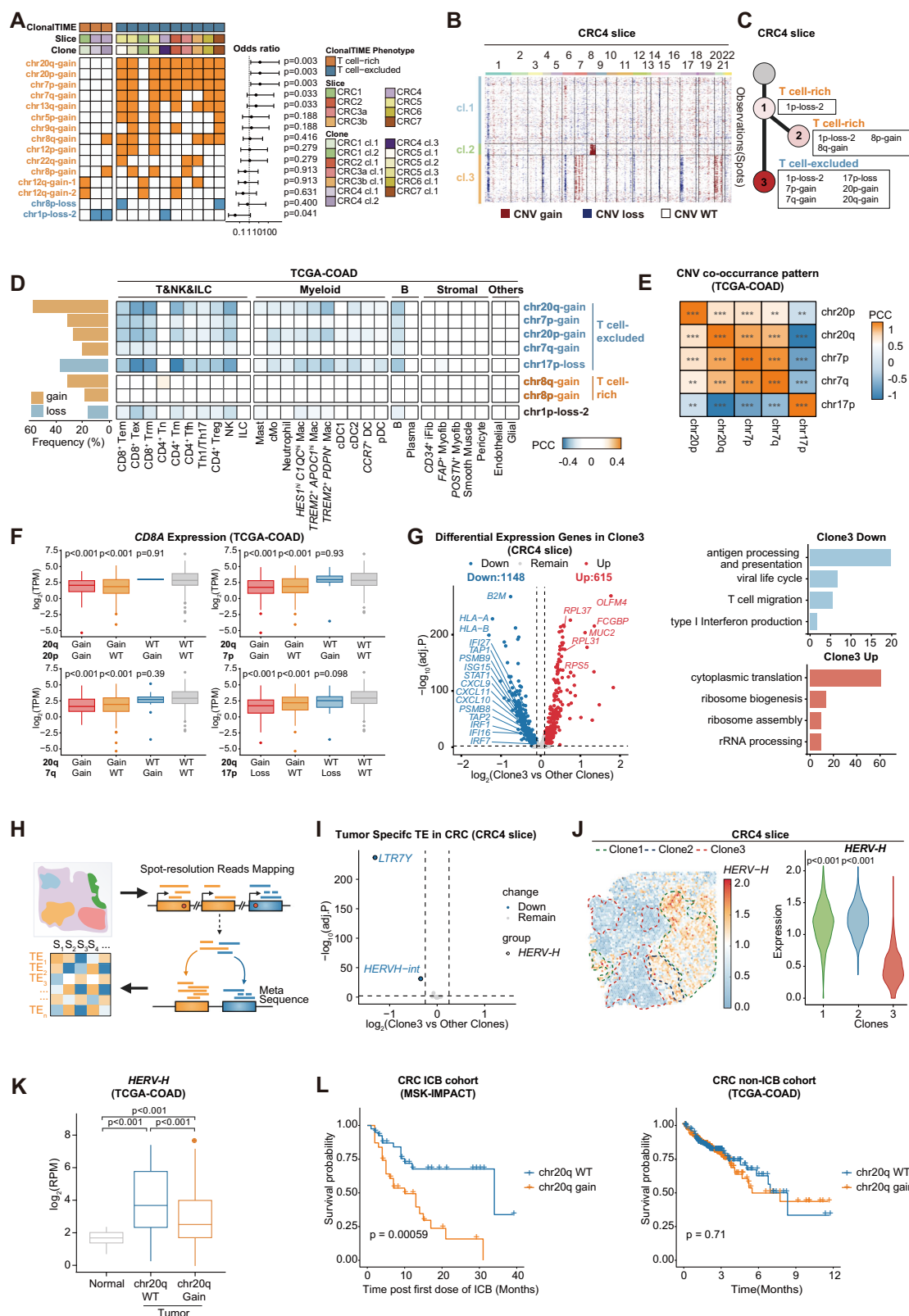
(Fig. 3J), exhibited enriched pathways in cytokine-mediated signaling pathway and acute inflammatory response (Supplementary Fig. S4N).

To link these communities with T cell-rich versus T cell-excluded clonalTIME phenotype, we examined the score of each community in the respective clonalTIME phenotypes (see “Methods” section). We observed a significantly higher score of community1 within T cell-rich clonalTIMES, while the score of community2 was higher in T cell-excluded clonalTIMES (Fig. 3K). The score of community3 showed no significant differences between the two clonalTIME phenotypes (Fig. 3K). Moreover, an opposite enrichment trend of community1 and community2 was detected across 13 clonalTIMES (PCC = -0.51) (Fig. 3L). We next closely examined community distributions in a particular ST slice (CRC4) containing both T cell-rich (clones 1 and 2) and T cell-excluded (clone 3) tumor clones. The result revealed a loss of community1 coincident with an increase of community2 and community3 as the clonalTIME transitioned from a T cell-rich to a T cell-excluded phenotype (Fig. 3M, N). Together, we spatially characterized the multicellular communities within T cell-rich and T cell-excluded TIME at a clonal resolution and identified two of these functional communities associated with T cell-rich versus excluded clonalTIME phenotypes in CRC.

Identification of tumor genetic alterations associated with clonalTIME transition in CRC

To investigate whether genetic alterations are associated with immune phenotypic transitions in tumor clones, we conducted a comparison of CNVs across clones (Supplementary Fig. S5A and Supplementary Table 4). Our findings revealed a distinct pattern among T cell-excluded clones, exhibiting unique CNVs that are absent in T cell-rich clones (Fig. 4A). Notably, we observed a frequent copy number gain on chromosomes 20q, 20p, and 7p (chr20q-gain, chr20p-gain, and chr7p-gain) within T cell-excluded clones (Fig. 4A). This observation prompted us to hypothesize that these chromosomal gains may arise during clonal evolution and promote the transition within the tumor microenvironment from a T cell-rich to a T cell-excluded state.

To further investigate the relationship between genetic alterations and the immunophenotype of tumor clones, we focused on the aforementioned CRC4 ST slice containing three clones exhibiting sequential clonal evolution. Clone1, characterized by chr1p-loss-2, served as a common ancestor (Fig. 4B, C). It gave rise to two distinct progenies: clone2 that acquired chr8p-gain and chr8q-gain, and clone3 with chr20q-gain, chr20p-gain, chr7p-gain, and two additional CNVs (Fig. 4B, C). Notably, while the evolution from clone1 to clone2 was accompanied by maintenance of a T cell-rich TIME (Fig. 4C), the evolution to clone3 was marked by a transition to the T cell-excluded phenotype, evidenced by a significant reduction in CD8⁺ Tex, CD8⁺ Tem, and *HES1*^{hi} *CIQC*^{hi} macrophages infiltration within their



clonalTIMes (Figs. 3I, N and 4C). This transition suggests that the genetic changes accumulated in clone3, particularly the chr20q-gain and associated CNVs, may be critical in shaping a T cell-excluded TIME.

For further validation of the relationship between the genetic changes specific to CRC4-clone3 and the T cell-exclusive immunophenotype, we conducted an integrative analysis utilizing the TCGA-COAD dataset, which encompasses DNA copy-number variation

(Affymetrix 6.0 microarrays) and bulk RNA sequencing data with 471 samples. We identified the existence of CRC4-identified CNVs and estimated the infiltration levels of various immune cell subpopulations based on their signature expressions derived from bulk RNA-seq for each patient. We then calculated the PCC between the presence of these CNVs and estimated cell subpopulations (Supplementary Fig. S5B), revealing significant correlations between

Fig. 4 | UCASpatial reveals clonalTIME transition orchestrated by chr20q-gain-associated endogenous retrovirus silencing. **A** Heatmap of clone-specific CNVs (orange, gain; blue, loss). Forest plots show odds ratios (ORs) with 95% CIs comparing T cell-excluded ($n = 10$ clones) and T cell-rich ($n = 3$ clones) phenotypes. Dots represent ORs; error bars represent 95% confidence intervals (CIs). **B** Inferred CNV profile from CRC4 (others in Supplementary Fig. S5A). **C** Phylogenetic tree of three CRC4 clones; gray dots indicate benign cells, and red dots represent tumor clones (intensity scales with CNV count). CNV status is summarized by chromosome arm/event (full genomic locations of each CNV are provided in Supplementary Table 4). **D** Left: frequency of CNVs in TCGA-COAD. Right: heatmap of PCC between CNV status and cell subpopulation signature scores in TCGA-COAD (blank squares: $PCC < 0.2$). **E** Heatmap displaying the PCC between the copy numbers of different CNV statuses within the TCGA-COAD dataset. **F** Boxplot showing *CD8A* expression in samples with indicated CNV status in the TCGA-COAD dataset. **G** Left: volcano plot of differentially expressed genes (clone3 versus others) in CRC4 (adj.

$p < 0.01$, $|\log_2FC| \geq 0.1$). Right: functional enrichment of upregulated ($n = 615$) and downregulated ($n = 1148$) genes in clone3. **H** Schematic representation of TE quantification in ST. **I** Volcano plot of differentially expressed TEs in clone3 (adj. $p < 0.01$, $|\log_2FC| \geq 0.25$). **J** Spatial visualization (left) and violin plot (right) of *HERV-H* expression across CRC4 clones. **K** Boxplot representing *HERV-H* expression across normal tissue, chr20q-WT, and chr20q-gain tumor samples in the TCGA-COAD dataset. **L** Kaplan–Meier curves for overall survival stratified by chr20q status in the MSK-IMPACT cohort (post-ICB, left) and TCGA-COAD (without ICB, right). Log-Rank p -values are shown. **H** created in BioRender. Zhang, Y. (2026) <https://BioRender.com/53h0h8g>. Statistical tests: Fisher's exact test (**A**); two-tailed Wilcoxon rank-sum with Bonferroni correction (**G**, **I**); log-rank test (**L**); two-tailed unpaired t -test (**F**, **J**, **K**); Pearson correlation analysis (**E**). ns non-significant, $p > 0.05$; $^*0.01 \leq p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$. Boxplots: center, median; box, quartiles; whiskers, $1.5 \times IQR$; points, outliers. Source data are provided as a Source Data file.

the presence of clone3-specific CNVs and the reduction in the signature expression of various T cell subpopulations (Fig. 4D).

Notably, all of these five clone3-specific CNVs (chr20q-gain, chr20p-gain, chr7p-gain, chr7q-gain, and chr17p-loss) exhibited a concurrent occurrence pattern within the TCGA-COAD dataset (Fig. 4E). To determine the primary CNV responsible for the diminished T-cell infiltration, we isolated the individual impacts of each CNV in the TCGA-COAD dataset based on their presence (see “Methods” section). Our analysis revealed that only the group harboring chr20q-gain, prevalent in 57.9% of CRC patients (Fig. 4D), exhibited a significant reduction in the expression levels of T cell marker genes (*CD3G*, *CD8A*, and *CD4*) when compared to the group without chr20q-gain (Fig. 4F and Supplementary Fig. S5C). Since chromosomal instability (CIN) has been previously reported to be associated with decreased immune cell infiltration⁴³, we stratified the samples by aneuploidy levels to control for its potential effects (see “Methods” section). Patients with chr20q-gain showed significantly lower T cell marker expression than chr20q-WT patients in both low and intermediate aneuploidy groups (Supplementary Fig. S5D), demonstrating a CIN-independent role of chr20q-gain in driving T cell exclusion. These findings suggest that the presence of chr20q-gain may play a pivotal role in driving the transition of the clonalTIME phenotype from T cell-rich to T cell-excluded.

Tumor-intrinsic endogenous retrovirus silencing correlates with chr20q-gain related T cell-excluded clonalTIME formation

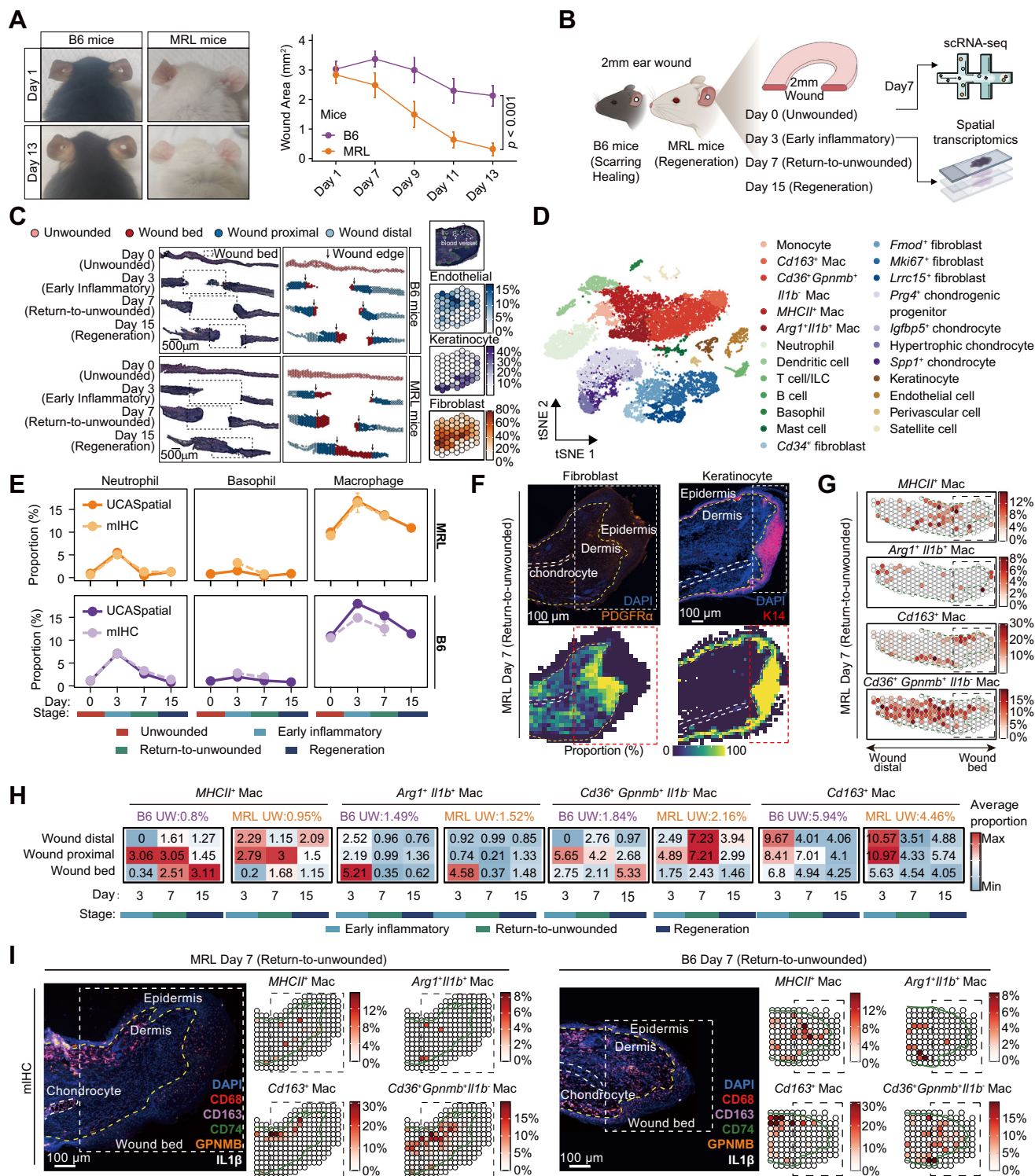
To investigate the potential mechanism of the immune phenotypic transitions in tumor clones, we performed differential gene expression analysis in clone3, which had acquired a chr20q gain, compared to clone1 and clone2 of CRC4. Through functional enrichment analysis, we identified enrichment for pathways related to translation machinery among the upregulated genes in clone3 (Fig. 4G). Conversely, pathways related to type I interferon (IFN) production were notably enriched among the downregulated genes, indicating a potential suppression of immune response (Fig. 4G and Supplementary Fig. S5E–G). Notably, chemokine-coding genes responsible for cytotoxic CD8⁺ T cell recruitment, such as *CXCL9* and *CXCL10*, and genes associated with antigen presentation, including *TAP1*, *TAP2*, *HLA-A*, *HLA-B*, and *B2M*^{44,45}, exhibited reduced expression (Fig. 4G and Supplementary Fig. S5E–G). This observation potentially elucidates the T cell-excluded phenotype of clone3.

It is well-established that the activation of the type I IFN response in tumors can be attributed to the innate sensing of cytosolic RNA derived from TEs (transposable elements)⁴⁶. Considering their role as internal adjuvants eliciting T cell responses against tumors, we hypothesized that differences in TE expression levels could directly influence the immune phenotypic transition from T cell-rich to T cell-excluded. To examine this hypothesis, we conducted a spatial analysis

of the expression of 40 TEs that were specifically overexpressed in CRC tumor cells (Fig. 4H and Methods)⁴⁷. Our analysis identified two CRC-specific TEs—the internal sequence of human endogenous retrovirus subfamily H (*HERV-H*) and its associated long terminal repeat *LTR7Y*^{48,49}—that were significantly downregulated in clone3 compared to clone1 and clone2 (Fig. 4I, J and Supplementary Fig. S5H, and “Methods” section). The statistical analysis across all slices verified the observations in CRC4 (Supplementary Fig. S5I). This downregulation suggests that tumor cells might suppress *HERV-H* expression to evade immune recognition in CRC.

To investigate whether the decreased *HERV-H* expression and attenuated antiviral response observed in clone3 were associated with chr20q-gain, we analyzed TCGA-COAD tumor samples. Notably, tumors with chr20q-gain showed significantly reduced *HERV-H* expression and diminished viral transcription and type I IFN signaling compared to chr20q-WT (Fig. 4K and Supplementary Fig. S5J–K, and “Methods” section). We also confirmed the attenuated type I IFN response in tumor cells with chr20q-gain using scRNA-seq data from our cohort (Supplementary Fig. S5L–O). These results collectively suggest a compelling correlation between the chr20q-gain in tumor cells and the downregulation of *HERV-H*, which may be implicated in the immune evasion strategies of CRC.

Given our discovery of a strong correlation between chr20q-gain and the consequent downregulation of *HERV-H*, which results in a T cell-excluded immune phenotype, coupled with the recognized influence of T cells on the clinical response to immune checkpoint blockade (ICB) treatment⁵⁰, we hypothesized that chr20q-gain might influence the response to ICB treatment in CRC patients. To examine this, we evaluated survival post-ICB treatment by categorizing patients from Memorial Sloan Kettering Cancer Center (MSKCC) IMPACT (MSK-IMPACT) cohorts (comprising 71 patients)^{51,52} into two groups based on their chr20q status. Our evaluation revealed that patients with chr20q-gain exhibited significantly poorer survival outcomes following ICB treatment (Fig. 4L and Supplementary Fig. S5Q). In contrast, this adverse survival trend was not observed in CRC patients who did not receive ICB treatment (TCGA-COAD dataset) (Fig. 4L and Supplementary Fig. S5R). To rule out the potential impact of tumor mutation burden (TMB), we categorized patients based on both CNV status and TMB. Our analysis indicated that the correlation between chr20q-gain and diminished survival in the CRC ICB cohort remained irrespective of TMB levels (Supplementary Fig. S5P). Taken together, we identified chr20q-gain, a prevalent genomic defect in CRC, as a genomic biomarker for T-cell exclusion, suggesting its utility as an immunogenomic parameter for predicting responsiveness to ICB. Furthermore, our findings lay the groundwork for a novel model in which tumor cells with chr20q-gain orchestrate a T cell-excluded clonalTIME by suppressing *HERV-H* expression during clonal evolution, thereby potentially evading immune surveillance.



UCASpatial accurately annotates cell subpopulations with spatiotemporal dynamics in the murine wound healing ST dataset

To demonstrate UCASpatial's utility, we next applied it to another real ST dataset, focusing on murine wound healing—a complex, highly organized multicellular process with distinct spatiotemporal dynamics^{53–55}. We employed a 2 mm ear punch wound model in MRL, a strain with super-healing capacity, which fully regenerates the ear punch wound without fibrosis^{56–58}. As a control, we used non-regenerating strain B6 mice, known to form scar tissue over ear wound edges (Fig. 5A)^{56–58}. To interrogate the tissue-wide spatiotemporal dynamics of various cell subpopulations during wound

healing, we collected ear tissue samples from MRL and B6 mice at the following stages: unwounded (day 0), early inflammatory (day 3), return-to-unwounded (day 7), and regeneration stage (day 15)⁵⁴. We performed ST sequencing using the 10x Genomics Visium platform (Fig. 5B, C). As a representative reference for capturing the cellular diversity present throughout the wound healing process, we utilized scRNA-seq data from day 7 ear tissues from both MRL and B6 mice, classifying them into 23 distinct cell subpopulations (Fig. 5D and Supplementary Fig. S6A, B).

UCASpatial successfully reconstructed the anatomical distributions of structural cell types, including keratinocytes, endothelial cells,

Fig. 5 | UCASpatial reconstructs the spatiotemporal dynamics of diverse macrophage subpopulations during murine wound healing. **A** Gross photographs and wound closure curves for B6 ($n=4$) and MRL ($n=4$) ear wounds. Data are mean $\pm$ SEM; two-tailed unpaired t -test. **B** Graphical overview of experimental design. Created in BioRender. Zhang, Y. (2026) <https://BioRender.com/ervvhv2>. **C** Left: H&E staining and spatial plots showing regions: unwounded (Day 0), wound bed (1 mm), proximal (0–1 mm), and distal (>1 mm). Scale bar, 500 μ m. Right: Day 7 wound bed micro-anatomy visualizing endothelial cells, keratinocytes, and fibroblasts. Green line: epidermis-dermis; white line: blood vessels. **D** The t -distributed Stochastic Neighbor Embedding (tSNE) plot showing 23 cell subpopulations from scRNA-seq. **E** Temporal dynamics of neutrophil, basophil, and macrophage proportions in MRL and B6 mice (UCASpatial: solid lines; mIHC: dashed lines). UCASpatial spots: B6 (D0: 182, D3: 210, D7: 196, D15: 240); MRL (D0: 228, D3: 170, D7: 286, D15: 283). mIHC sample sizes (n): Neutrophils: B6 (D0: 3, D3: 1, D7: 2, D15: 3), MRL

(D0: 3, D3: 2, D7: 1, D15: 3); Basophils: B6 (D3: 5, D7: 2), MRL (D3: 4, D7: 1); Macrophages: B6 (D0: 1, D3: 1, D7: 3), MRL (D0: 1, D3: 3, D7: 2). Data shown as mean $\pm$ SEM. **F** Upper: representative immunofluorescence of fibroblasts and keratinocytes in MRL mice at day 7 (scale bar: 100 μ m). Bottom: quantitative distribution (50 μ m-bin resolution) based on mIHC. **G** Day 7 spatial visualization of macrophage subpopulations. **H** Heatmap of spatiotemporal dynamics of macrophage subpopulations. UW, unwounded (Day 0 average, above heatmap). **I** Representative mIHC and the quantitative evaluation (50 μ m-bin resolution) of the distributions of indicated macrophage subpopulations at the return-to-unwounded stage (day 7 post-injury) in MRL mice ($n=2$) and B6 mice ($n=2$). Scale bar, 100 μ m. For mIHC (**E**, **I**), annotations indicate the wound bed (dashed rectangle), chondrocytes (white), and epidermis-dermis interface (yellow). For the quantitative evaluations (**G**, **I**), annotations indicate the wound bed (dashed rectangle) and epidermis-dermis interface (green line). Source data are provided as a Source Data file.

and fibroblasts (Fig. 5C and Supplementary Fig. S6C). Moreover, UCASpatial resolved the spatiotemporal dynamics of wound-associated immune cells. Specifically, an increase in myeloid cells—neutrophils, basophils, and macrophages—was observed at the early inflammatory stage in both MRL and B6 mice (Fig. 5E). This observation is consistent with the established notion that innate immune cells are the first responders to injury^{59–61}. At the return-to-unwounded stage, neutrophils and basophils exhibited a reduction in frequency, whereas macrophages maintained a high abundance (Fig. 5E), underscoring their roles in both the early injury response and the subsequent tissue repair and regeneration processes^{54,59,60,62,63}.

We experimentally validated the spatial distribution and proportion of six major cell types, including keratinocytes, fibroblasts, endothelial cells, neutrophils, macrophages, and basophils, using multiplexed immunohistochemistry (mIHC) (Supplementary Fig. S6D). The anatomical distributions of major cell types detected by mIHC were consistent with UCASpatial's deconvolution analysis (Fig. 5C, E, F and Supplementary Fig. S6E–G). Notably, the spatiotemporal dynamics of cell subpopulations quantified by UCASpatial were more closely aligned with the mIHC results than those obtained by other methods (Fig. 5E and Supplementary Fig. S6G). This consistency between the mIHC and UCASpatial's deconvolution results validates the accuracy of UCASpatial.

To compare the performance of UCASpatial with existing methods, we calculated the PCC between the spatial distribution of cell subpopulations and marker-gene scores (Supplementary Data 2 and “Methods” section), which allowed us to quantify the performance of deconvolution in the absence of ground truth¹⁶. Among all methods compared, UCASpatial exhibited the highest PCC across cell subpopulations (Supplementary Fig. S6H, I). Specifically, by comparing UCASpatial with “UCASpatial ablation”, which shares the same framework but lacks the core functions of entropy-based weighting and meta-purity filtering process, we noticed that the improved accuracy of UCASpatial is attributed to the incorporation of entropy weighting and meta-purity filtering (Supplementary Fig. S6H, I). Notably, existing methods lacked precision in identifying immune cell subpopulations during wound healing. For example, RCTD, stereoscope, and Spotiphy failed to accurately capture the spatiotemporal dynamics of neutrophils, which peaked at the early inflammatory stage in both strains (Supplementary Fig. S6G). Additionally, RCTD, cell2location, CARD, and Spotiphy showed limited ability to detect macrophages at the unwounded stage (Supplementary Fig. S6G).

Further, UCASpatial precisely captured the spatial segregation patterns of macrophage subpopulations. Specifically, *MHCII*⁺ macrophages, present at all measured time points, were predominantly located in the wound's proximal and distal regions in MRL mice, while predominantly located in the wound's bed and distal regions in B6 mice (Fig. 5G, H and Supplementary Fig. S6J–L). *Arg1*⁺ *Il1b*⁺ inflammatory macrophages, which appeared transiently in the wound bed at the early inflammatory stage (Fig. 5H), preferentially clustered at the

migrating epithelial tongue (Fig. 5G and Supplementary Fig. S6M). In contrast to *Arg1*⁺ *Il1b*⁺ macrophages, *Cd163*⁺ macrophages with enriched functions in endocytosis and lysosome (Supplementary Fig. S6N) predominantly resided along the epidermis layer outside the migrating epithelial tongue across all measured time points post-injury (Fig. 5G and Supplementary Fig. S6M). *Cd36*⁺ *Gpnmb*⁺ *Il1b*⁺ macrophages, functionally resembling lipid-associated macrophages implicated in fibrosis^{64–66} (Supplementary Fig. S6N), are located in the wound proximal and distal regions at the return-to-unwounded stage (Fig. 5G, H). Such spatial distribution patterns of four macrophage subpopulations at the return-to-unwounded stage were confirmed by mIHC (Fig. 5I).

We also observed a differential distribution pattern of *Cd36*⁺ *Gpnmb*⁺ *Il1b*⁺ macrophages in B6 and MRL mice. This subpopulation was initially located in the wound proximal and distal regions of both strains at the return-to-unwounded stage (Fig. 5G, H and Supplementary Fig. S6L). However, at the regeneration stage, *Cd36*⁺ *Gpnmb*⁺ *Il1b*⁺ macrophages in B6 mice accumulated within the wound bed, while in MRL mice, they remained distal to the wound bed (Fig. 5H and Supplementary Fig. S6O). This unique migration pattern observed in B6 mice might explain their scarring-healing phenotype. Collectively, UCASpatial reconstructs a fine-grained spatiotemporal cellular landscape in murine wound healing, revealing the unique spatial segregation and temporal dynamics of macrophage subpopulations that serve in different stages of wound healing.

Spatial characterization of molecularly distinct chondrocyte subpopulations linked to regenerative versus scarring-healing phenotypes

After demonstrating our capability to spatially map various cell subpopulations in wounded tissues, we proceeded to investigate the cellular events orchestrating regenerative versus scarring-healing phenotypes in MRL and B6 mice between 7 and 15 days post-injury. We identified a scarring-associated myofibroblast subpopulation *Lrrc15*⁺ fibroblasts, located within the wound bed at the return-to-unwounded stage in both MRL and B6 mice^{67–69} (Supplementary Fig. S7A, B). While a reduction in the frequency of *Lrrc15*⁺ fibroblasts was observed in both strains at the regeneration stage, these cells were significantly more enriched in B6 mice compared to MRL (Supplementary Fig. S7A, B). This differential enrichment may account for the scarring-healing phenotype observed in B6 mice.

We also delineated the spatiotemporal dynamics of keratinocytes and chondrocytes, two major cell populations contributing to regenerative healing^{70,71}. We observed a comparable distribution of keratinocytes in MRL and B6 mice within regions of the migrating epithelial tongue at the return-to-unwounded stage (Supplementary Fig. S7C–E), indicating complete re-epithelialization occurred in both strains. However, at the regeneration stage, while the proportion of keratinocytes remained high within the wound bed of MRL mice, a substantial decrease was observed in B6 mice, supporting the notion that

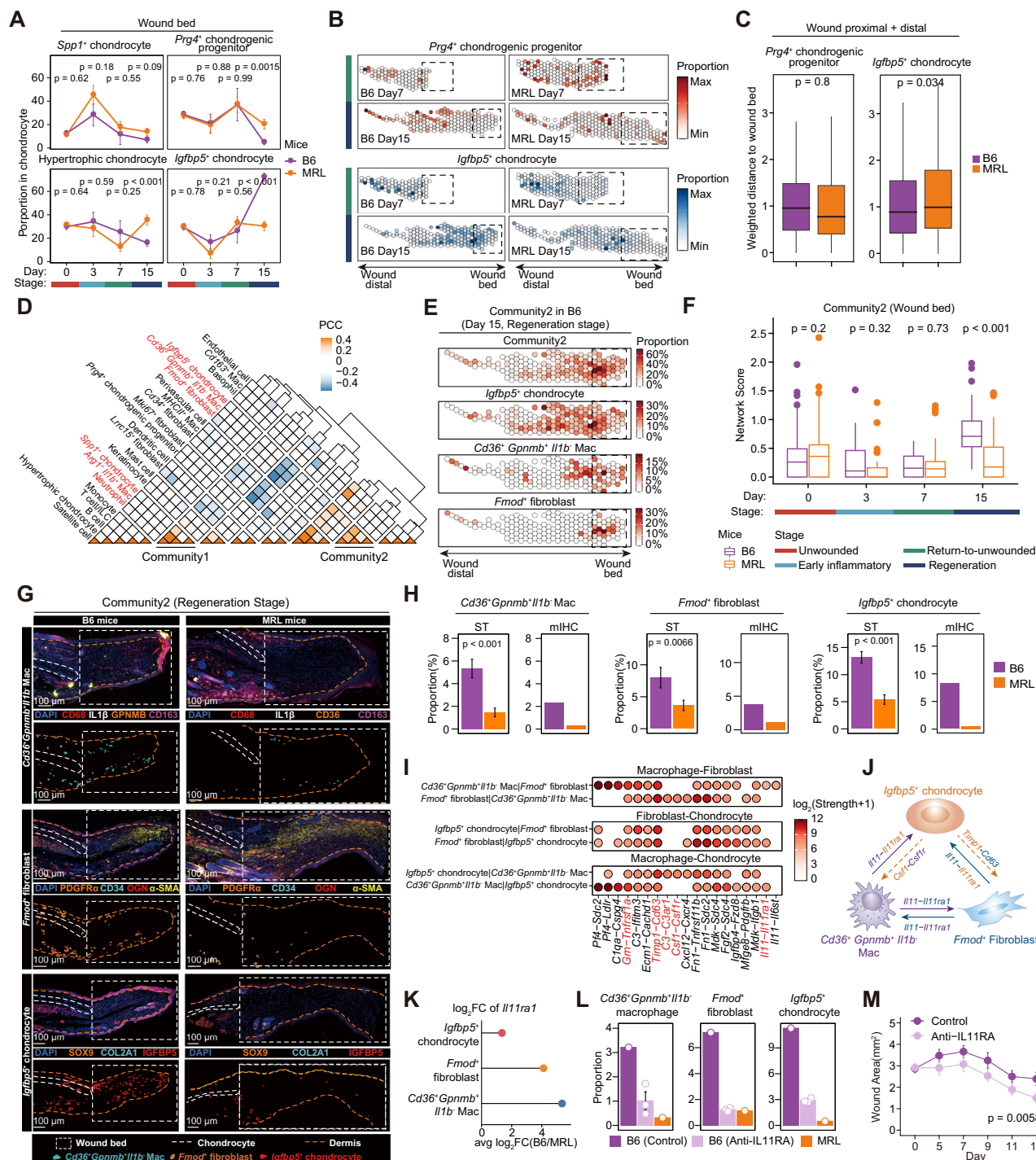


Fig. 6 | Disruption of the IL11–IL11RA axis diminishes the pro-fibrotic community and promotes regenerative healing. **A** Line plots showing the temporal dynamics of chondrocyte subpopulations. Spot numbers for days 0, 3, 7, and 15: B6 (182, 16, 13, 38); MRL (228, 29, 63, 67). **B** Spatial distribution of indicated subpopulations on day 7 and 15 in MRL and B6 mice; dashed boxes: wound bed. **C** Boxplots showing weighted distance of subpopulations in wound proximal and distal regions on day 15 in MRL ($n = 144$ spots) and B6 ($n = 132$ spots). Statistical analysis: one-tailed unpaired t -test. **D** Heatmap of PCC between subpopulation proportions across all spots after neighborhood processing. **E** Spatial distribution of community2 and indicated subpopulations at the wound bed on day 15 in B6 mice. The dashed rectangles represent the wound bed. **F** Boxplots of community2 scores. Spot numbers as in (A). **G** Representative immunofluorescence imaging of cell subpopulations within community2 at the regeneration stage. Annotations indicate the wound bed (dashed rectangle), chondrocytes

(white), and epidermis-dermis interface (yellow). Scale bar, 100 μ m. **H** Bar plots showing the proportion of indicated cell subpopulations in wound bed regions at the regeneration stage in ST data (B6: $n = 55$ spots, MRL: $n = 112$ spots) and mIHC ($n = 1$). **I** Bubble heatmaps showing the $\log_2$ strength for selected ligand-receptor pairs from the scRNA-seq of B6 mice. **J** Schematic diagram of community2. **K** Lollipop plot showing the $\log_2$ fold change of *Il11ra1* in indicated cell subpopulations across MRL and B6 mice. **L** Bar plots showing the proportion of indicated cell subpopulations in wound bed regions at the regeneration stage in mIHC. B6 control (anti-rabbit IgG), $n = 1$; B6 Anti-IL11RA, $n = 3$; MRL, $n = 1$. **M** Wound curves of B6 ear wounds in the control ($n = 5$) and Anti-IL11RA-treated ($n = 7$) groups. Statistical analysis in (A, F, H, M): two-tailed unpaired t -test. Boxplot elements: center line, median; limits, upper/lower quartiles; whiskers, $1.5 \times$ IQR. Data in (A, H, L, M) are mean $\pm$ SEM. Source data are provided as a Source Data file.

keratinocytes play a crucial role in the regeneration process (Supplementary Fig. S7C, D)^{72–74}.

Spatial mapping of four chondrocyte subpopulations (Supplementary Fig. S7F and G) during post-injury chondrogenesis revealed distinct regional and temporal patterns. Injury-induced, de-differentiated *Spp1*⁺ chondrocytes⁷⁵ were acutely present at the wounded bed, with comparable levels in MRL and B6 mice at the early inflammatory stage (Fig. 6A). *Prg4*⁺ chondrogenic progenitors, which give rise to hypertrophic chondrocytes^{76–78}, were transiently observed within the wound bed of B6 mice but persisted in MRL mice from the return-to-unwounded stage till the regeneration stage (Fig. 6A, B). Consistently, a higher proportion of mature hypertrophic chondrocytes was observed in MRL mice at the regeneration stage (Fig. 6A), suggesting that this progenitor is linked to regenerative healing in MRL mice.

In addition to the *Prg4*⁺ chondrogenic progenitor, we noticed an uncharacterized *Igfbp5*⁺ chondrocyte subpopulation that expressed *Cd34* and *Ogn* (Supplementary Fig. S7G). *Igfbp5*⁺ chondrocyte was closely related to *Prg4*⁺ chondrogenic progenitor and was maintained in a less-differentiated state, as evidenced by lower expression levels of *Col2a1* and *Acan* (Supplementary Fig. S7H–K)^{79,80}. Functional analysis of *Igfbp5*⁺ chondrocytes showed enriched pathways for extracellular matrix (ECM) organization and ECM-receptor interaction (Supplementary Fig. S7L), implying potential interaction between *Igfbp5*⁺ chondrocytes and ECM-producing cells like fibroblasts and macrophages⁸¹. Intriguingly, the spatiotemporal dynamics of *Igfbp5*⁺ chondrocytes differed in MRL and B6 mice. While these cells initially occupied regions proximal to the wound edge in both strains at the return-to-unwounded stage, they continued spreading into the wound bed, replacing the *Prg4*⁺ chondrogenic progenitor at the regeneration stage in B6 mice (Fig. 6A, B). In contrast, *Igfbp5*⁺ chondrocytes were spatially restricted in MRL mice, remaining distal to the wound bed at the regeneration stage (Fig. 6B, C). Therefore, the replacement of *Prg4*⁺ chondrogenic progenitors with *Igfbp5*⁺ chondrocytes is associated with the scarring healing phenotype in B6 mice.

***Igfbp5*⁺ chondrocyte, *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophage and *Fmod*⁺ fibroblast formed a three-cell-type cluster underlying scarring wound healing**

Next, we asked whether *Igfbp5*⁺ chondrocytes positioned within the wound bed could locally crosstalk with surrounding cells and form a niche that favors scarring healing in B6 mice. To test this, we calculated the PCC between 23 cell subpopulations and identified two chondrocyte-related communities by hierarchical clustering (see “Methods” section). The community1, comprising *Spp1*⁺ chondrocytes, neutrophils, and *Arg1*⁺ *Il1b*⁺ macrophages, formed a transient “inflammatory community” within the wound bed at the early inflammatory stage in both MRL and B6 mice (Fig. 6D and Supplementary Fig. S7M). The community2, which we termed the “pro-fibrotic community”, featured a stable three-cell-type cluster involving *Igfbp5*⁺ chondrocytes, ECM organization-related *Fmod*⁺ fibroblasts (Supplementary Fig. S7N), and *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophages (Fig. 6D–F). This cluster was particularly prominent in the wound bed at the regeneration stage in B6 mice, which has been validated by mIHC (Fig. 6G, H and Supplementary Fig. S7O).

Analysis of functional-enriched pathways across three cell subpopulations within community2 suggested their overlapping functions were responsible for fibrosis, including ECM organization, collagen fibril organization, and supramolecular fiber organization (Supplementary Fig. S7P and “Methods” section). To determine whether the scarring-healing phenotype observed in B6 mice could be attributed to the formation of this pro-fibrotic community, we conducted cellular-level intervention of the pro-fibrotic community by blocking *Fmod*⁺ fibroblast differentiation. Trajectory analysis using the diffusion map algorithm revealed that *Fmod*⁺ fibroblasts might

originate from *Cd34*⁺ fibroblasts (Supplementary Fig. S7Q). During this transition, PPAR γ signaling-pathway-related genes were down-regulated, which are known to be associated with fibrosis and regeneration failure (Supplementary Fig. S7R)^{82–84}. As expected, treating B6 mice with the PPAR γ agonist rosiglitazone inhibited the differentiation of *Cd34*⁺ fibroblasts into *Fmod*⁺ fibroblasts (Supplementary Fig. S7S, T, “Methods” section). We observed that the three-cell-type cluster in community2 was prominent in the wound bed of control B6 mice at the regeneration stage, whereas this cluster was absent in rosiglitazone-treated B6 mice (Supplementary Fig. S7U, V). Additionally, rosiglitazone treatment led to a significant reduction in the remaining ear hole area, indicating enhanced wound healing (Supplementary Fig. S7W). Together, these findings underscore the spatial coordination among *Igfbp5*⁺ chondrocyte, *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophage, and *Fmod*⁺ fibroblast, forming a three-cell-type cluster that contributes to the scarring-healing phenotype observed in B6 mice.

Disruption of the IL11–IL11RA axis within the pro-fibrotic community enhances wound healing in B6 mice

Next, we examined the potential cross-talks within the pro-fibrotic community. Analysis of specific ligand-receptor interactions revealed a stable interaction between *Fmod*⁺ fibroblasts and *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophages (Fig. 6I, J), supporting their known partnership⁸⁵. Additionally, we observed reciprocal interactions between *Igfbp5*⁺ chondrocytes and both *Fmod*⁺ fibroblasts and *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophages. Specifically, *Igfbp5*⁺ chondrocytes can crosstalk with *Fmod*⁺ fibroblasts via the ligand-receptor pair *Timp1*-*Cd63* (Fig. 6I), the ligand-receptor pair for the induction of collagen synthesis during fibrosis^{86,87}. Further, *Igfbp5*⁺ chondrocytes can potentially interplay with *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophages through the ligand-receptor pair of *C4*-*C3arl* and *Csf1*-*Csf1r* (Fig. 6I), signals required for macrophage survival and polarization^{88,89}. *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophages and *Fmod*⁺ fibroblasts exhibited interaction with *Igfbp5*⁺ chondrocytes via *Il11*–*Il11ra1*, the ligand-receptor pair that could promote fibrosis (Fig. 6I)⁹⁰. Besides, *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophages might also promote the proliferation of *Igfbp5*⁺ chondrocytes and *Fmod*⁺ fibroblasts via the interaction of *Grn*-*Tnfrsf1a* (Fig. 6I)⁹¹.

Among these ligand-receptor interactions, we identified an unexplored axis involving the *Il11*–*Il11ra1* signaling pathway in the pro-fibrotic community (Fig. 6I). IL-11 is a member of the IL-6 cytokine family and plays a pivotal role in driving organ fibrosis and dysfunction^{92–94}. Given that IL11RA is expressed in a wide array of cell types, including stromal cells, pericytes, and epithelial cells, IL-11 signaling can trigger diverse cellular responses, leading to varied outcomes depending on the affected cell type⁹². It remains unknown which specific cell types receive IL-11 signals during wound healing and whether this subsequently affects wound healing.

We noticed that *Il11ra1* expression was dramatically up-regulated in *Cd36*⁺ *Gpnmb*⁺ *Il1b*[−] macrophages, *Igfbp5*⁺ chondrocytes, and *Fmod*⁺ fibroblasts of B6 mice compared to MRL mice (Fig. 6K and Supplementary Fig. S7X). This strain-specific difference suggests that the IL11–IL11RA axis might be responsible for the formation of the pro-fibrotic community in B6 mice. To test this, we blocked IL11–IL11RA signaling using an anti-IL11RA neutralizing antibody (anti-IL11RA Ab) in B6 mice. We found that the formation of the pro-fibrotic community was prevented in the wound bed in mice receiving anti-IL11RA Ab compared to the control group, indicating that IL11–IL11RA-mediated interactions contribute to the formation of the pro-fibrotic community in B6 mice (Fig. 6L and Supplementary Fig. S7U). Furthermore, neutralization of IL11RA significantly reduced the remaining ear hole area, suggesting that IL-11 signaling limits regeneration in B6 mice (Fig. 6M). Together, these results reveal that the strain-specific difference in wound healing between B6 and MRL mice is driven by the IL11–IL11RA axis, underscoring the power of UCASpatial in dissecting the molecular mechanisms underlying complex phenotypes.

Discussion

Spatial deconvolution methodologies have made significant strides in deciphering the cell composition of well-structured tissues like the mouse brain and heart^{9–13,29}. However, these methodologies encounter challenges when applied to complex microenvironments, where cell subpopulations typically exhibit subtle differences, low abundance, and high heterogeneity. To address these challenges, we introduce UCASpatial, a robust, sensitive, and versatile algorithm designed to deconvolute fine-grained spatiotemporal cellular composition.

UCASpatial shares conceptual similarities with existing methodologies such as SPOTlight and CARD, which apply NMF-based models to estimate cell-type compositions^{9,12}. Yet, UCASpatial distinguishes itself through its meta-purity filtering and entropy-based weighting mechanisms, providing distinct advantages for discerning cell subpopulations with subtle transcriptomic differences, as evidenced by our ablation tests, in-silico benchmarking results, and analysis of real ST datasets. Moreover, UCASpatial does not incorporate neighboring or image information, a strategy commonly employed previously to enhance performance^{12,14,27,28}. Integrating such information might not be suitable for the annotation of infiltrated immune cell subpopulations with low proportions in complex microenvironments, given their discrete distribution patterns due to recruitment dynamics. Furthermore, the integration of the UCASpatial algorithm and its downstream functions into the Seurat framework significantly facilitates a wide range of applications, streamlining the process for researchers.

Beyond UCASpatial's strengths in resolving heterogeneous immune microenvironments and rare cell subsets, other deconvolution tools retain niche advantages that justify their continued use. Spotiphy uniquely fuses histological images with expression to deliver single-cell-resolution whole-transcriptome images. RCTD and cell2location perform comparably in benchmarks, yet RCTD's lightweight Bayesian model demands few dependencies and minimal memory, making it attractive for rapid, resource-limited projects. Conversely, cell2location requires GPUs and large memory, but its seamless integration with Scanpy offers Python-native users a familiar, reproducible pipeline. CARD, by embedding neighborhood or morphological priors, is preferable for highly ordered tissues such as cortex or cardiac muscle, where spatial continuity dominates. Thus, while UCASpatial maximizes performance in heterogeneous, immune-rich microenvironments, method selection should balance compute budget, programming ecosystem, and the value of image information to match each biological question.

Despite its strengths, UCASpatial has some limitations that should be considered. The performance of UCASpatial is partially dependent on the quality of the reference scRNA-seq data provided by the user. While UCASpatial's ability to leverage unpaired scRNA-seq data enhances its accessibility and utility, high-fidelity scRNA-seq data are crucial to ensure the accuracy and reliability of the deconvolution analysis. Additionally, we noticed that UCASpatial may not always outperform other methods for certain cell types, such as myofibroblasts and epithelial cells. This variability is likely due to the distinct features of these cell types, which allow most methods to annotate them with comparable accuracy. Despite these challenges, by employing entropy-based weighting in UCASpatial, we have noticed significant improvements in UCASpatial's ability to resolve immune subpopulations compared to other methods, particularly in real datasets.

Our study employs UCASpatial to decipher the fine-grained landscape of TIMEs, offering an immunogenetic platform for linking tumor genomics with the complexity of immune landscapes at clonal resolution. In our analysis of human CRC, we identified a prevalent CNV, chr20q-gain, which emerges during clonal evolution and correlates with a T cell-excluded immune phenotype, consistent with prior observations linking chr20q-gain to low immune infiltration in

tumors⁹⁵. By linking clonal-level cellular community transitions to genetic alterations in chr20q, we uncovered a potential molecular mechanism for T cell exclusion: the suppression of the endogenous retrovirus *HERV-H*, associated with impaired type I interferon response, during clonal evolution from chr20q-WT to chr20q-gain. Furthermore, we found that chr20q-gain serves as a novel biomarker for predicting poor outcomes in ICB therapy, a finding not previously reported. Although our analysis does not provide direct experimental evidence of causation, the consistency of these findings across multiple datasets suggests a biologically significant association. UCASpatial further extends its utility in revealing the spatiotemporal dynamics during the wound healing process, which has empowered us to dissect the previously obscure dynamics of cellular turnover and spatial organization within the regenerative process. This advancement has not only illuminated the complex interplay of cellular communities in murine wound healing, but has also provided a roadmap for future investigations into the spatial orchestration of tissue repair and regeneration. These discoveries highlight the potential of UCASpatial as a tool for uncovering clinically relevant biomarkers and deepening our understanding of the immunogenomic landscape.

As we envision the future development of UCASpatial, its integration with additional omics dimensions, such as whole-exome sequencing and spatial metabolomics, presents a promising avenue. This multi-omics approach could provide a more comprehensive elucidation of the spatiotemporal dynamics across a spectrum of biological processes. Expanding UCASpatial's application to other cancer types may uncover novel mechanisms influencing TIMEs and offer predictive insights into the efficacy of specific immunotherapies. Furthermore, the effectiveness of UCASpatial is dependent upon the quality of the referenced scRNA-seq data. To this end, the utilization and establishment of a robust scRNA-seq database, inclusive of a wide array of common tissues and tumor types, will be paramount. This resource would not only enhance the algorithm's accuracy but also broaden its applicability across diverse research and clinical settings.

In conclusion, we anticipate that UCASpatial will find myriad applications in deciphering complex immune microenvironments across various developmental stages, states of health, and disease conditions within a spatial framework.

Methods

UCASpatial model

In brief, UCASpatial is an entropy-weighted model that annotates the cell-identity (equal to cell-type) compositions on each spatial location by deconvoluting the spatial expression count matrix into a predefined set of reference cell types. The model requires both ST data and paired or unpaired scRNA-seq data from a given tissue as input. By default, the model takes an untransformed spatial-expression count matrix $\mathbf{P} \in \mathbb{R}^{S \times G}$ of genes $g = \{1, \dots, G\}$ at spots $s = \{1, \dots, S\}$ as input, as, for example, obtained from the 10x Space Ranger software (10x Visium data). The scRNA-seq data $\mathbf{B} \in \mathbb{R}^{G \times N}$ serve as a reference and consist of N cells with the same set of G genes with $\mathbf{P}$. All the N cells will be classified into K cell types, which need to be provided by the users.

We denote $\mathbf{X} \in \mathbb{R}^{S \times K}$ as the cell-type composition matrix, where each row x_s represents the proportions of the K cell types on spot s . By default, this matrix is estimated using an entropy-WNNLS model, which allows for accurate and robust annotation in complex situations (see below).

Meta-purity filtering

The existence of intermediate cells blurs cell-type borders due to their mixed identities from multiple cell types, which will disturb the recognition of cell-identity-specific gene expression profiles (cGEPs). To remove the effect brought by intermediate cells in the reference scRNA-seq data, we designed the meta-purity filtering strategy as an optional preprocessing.

Based on the input scRNA-seq reference matrix $\mathbf{B}$, we construct meta cells using the function 'FindClusters' from the R package 'Seurat' with a default resolution of 100. We recommend the resolution parameter based on an expectation of an average of 100 cells per meta-cell, which serves as a parameter selection criterion. Using pre-existing cell-type labels, we define the purity of meta-cells by $pu(MC_i) = \max_{j \leq K} \frac{n_{ct_j}}{n_{MC_i}}$, where MC_i denotes the i th meta-cell, $pu(MC_i)$ denotes the purity of MC_i , n_{ct_j} denotes the cell number of cell type j in MC_i , and n_{MC_i} denotes the cell number of MC_i . The cells contained in meta cells with purity below the given threshold (default 0.95) will be filtered out.

Notably, the meta-purity filtering process does not result in the loss of any predefined cell types (Supplementary Figs. S1V-X, S3B, and S6A). The meta-purity filtering process has specific designs to prevent the loss of specified cell subsets. We have implemented a stringent criterion for the meta-purity filtering process. If any loss of the specified cell subsets is detected, the process is immediately halted, and displays an error message prompting a review of input parameters. This ensures that there is no unintended loss of any cell types specified by the user.

Downsampling

Another optional preprocessing for referenced scRNA-seq is downsampling. By default, we will randomly downsample the scRNA-seq reference matrix $\mathbf{B}$ to 100 cells for each cell type based on a given random seed to improve the speed for downstream analysis. Thus, the processed scRNA-seq expression matrix $\tilde{\mathbf{B}} \in \mathbb{R}^{G \times N}$ has been established, in which $N = 100 \times K$ by default.

Optimized NMF

In order to efficiently extract the cGEP matrix $\mathbf{E} \in \mathbb{R}^{G \times K}$, we utilize NMF on the processed scRNA-seq expression matrix $\tilde{\mathbf{B}}$. Based on the provided cell-type labels, we initialize $\mathbf{H} \in \mathbb{N}^{K \times N}$ as a one-hot matrix representing the identity of each cell, thus guiding it towards a biologically relevant result. This initialization also aims at reducing variability and improving the consistency between runs. We iterated through NMF to estimate the cGEP matrix $\mathbf{E}$:

$$\tilde{\mathbf{B}} \sim \mathbf{E} \times \mathbf{H} \quad (1)$$

where each element in the matrix $\mathbf{E}$ and $\mathbf{H}$ is constrained to be non-negative.

Entropy-based cGEPs weighting

Considering that the indicative capacity for cell types of each gene is unequal, quantifying the indicative capacity of each gene and taking it into account in the decomposition step would benefit from distinguishing the heterogeneous cell states with subtle transcriptome differences. Shannon entropy for the expression percentage of each gene across different cell types can be a great quantification for the indicative capacity of each gene. Thus, we formulate Shannon entropy-based weights for each gene g of the extracted cGEP matrix $\mathbf{E}$ to highlight genes indicative of cell identities.

UCASpatial calculates the expression percentage matrix $\mathbf{PCT} \in \mathbb{R}^{K \times G}$ based on the processed scRNA-seq expression matrix $\tilde{\mathbf{B}}$:

$$\mathbf{PCT}_{k,g} = \frac{\text{num}(\tilde{\mathbf{B}}_{g,n \in k} > 0)}{\text{num}(\tilde{\mathbf{B}}_{g,n \in k})} \quad (2)$$

where $\mathbf{PCT}_{k,g}$ denotes the element of $\mathbf{PCT}$ representing cell type k for gene g , $\text{num}(\tilde{\mathbf{B}}_{g,n \in k})$ represents the number of cells in cell type k in $\tilde{\mathbf{B}}$, $\text{num}(\tilde{\mathbf{B}}_{g,n \in k} > 0)$ represents the number of cells in cell type k with gene g expression over 0. The expression percentage matrix $\mathbf{PCT}$ will be

further normalized into $\widetilde{\mathbf{PCT}}$ via each element $\mathbf{PCT}_{k,g}$ being divided by the sum of the corresponding row:

$$\widetilde{\mathbf{PCT}}_{k,g} = \frac{\mathbf{PCT}_{k,g}}{\sum_{i=1}^K \mathbf{PCT}_{i,g}} \quad (3)$$

Based on the $\widetilde{\mathbf{PCT}}$ matrix, UCASpatial further calculates the Shannon entropy for each gene:

$$\text{Entropy}(g) = - \sum_{i=1}^K \widetilde{\mathbf{PCT}}_{i,g} \times \log \widetilde{\mathbf{PCT}}_{i,g} \quad (4)$$

For better understanding, UCASpatial provides an adjusted entropy using the maximum entropy minus the entropy for each gene g :

$$\text{adj.ent}(g) = \text{Ent.maxm} - \text{Entropy}(g) \quad (5)$$

where $\text{Ent.maxm} = \sup(\text{Entropy}(g)) = - \sum_{i=1}^K \frac{1}{K} \times \log \frac{1}{K}$ represents the supremum of $\text{Entropy}(g)$. The maximum entropy Ent.maxm is only related to the number of cell types K . The adjusted entropy could excellently represent the expression specificity for genes, with a higher adjusted entropy representing a higher indicative capacity for cell types of a gene.

Considering that the adjusted entropy can only evaluate the expression specificity of each gene, we thus took not only the adjusted entropy but also the expression percentage and expression level into account, designing the entropy-based weight of each gene as:

$$\omega(g) = \max_{i \in \{1, 2, \dots, K\}} \left(\frac{2 \times \text{adj.ent}(g)}{\text{Ent.maxm}} \times 2 \times \widetilde{\mathbf{PCT}}_{i,g} \times \text{avg. log FC}(i, g) \right) \quad (6)$$

$$\omega(g) = \begin{cases} \omega(g), & \omega(g) \leq 1 \\ 1, & \omega(g) > 1 \end{cases} \quad (7)$$

where $\text{avg. log FC}(i, g)$ denotes the average of $\log_2$ fold-change expression for gene g in cell type i , $\omega(g)$ denotes the entropy-based weight of gene g , which is constrained to be non-negative. UCASpatial also combined common marker finding methods, including the FindAllMarkers function in Seurat (version 3.1.4) and the cosine similarity method⁹⁶ (version 0.9.0), for optional gene filtering in cGEPs, which can further improve the indicative capacity of genes in cGEPs.

Entropy-WNNLS model

Taking the untransformed spatial expression count matrix $\mathbf{P}$ as input, this step aims to decompose the RNA counts of each spot into referenced cell type composition $\mathbf{X} \in \mathbb{R}^{S \times K}$ based on $\mathbf{E}$. Unit variance normalization by gene is performed in $\mathbf{P}$ in order to standardize discretized gene expression levels^{5,97}. Commonly, this step will utilize the NNLS model for each spot s :

$$\mathbf{P}_s = \mathbf{X}_s \times \mathbf{E} + \epsilon_s \quad (8)$$

where $\mathbf{P}_s$ denotes the spatial expression count vector of spot s , $\epsilon_s \in \mathbb{R}^{S \times 1}$ represents the residual error of spot s which is restricted to being non-negative. By minimizing the loss function for each spot s :

$$\min(\epsilon) = \min \|\epsilon_s\|_2^2 = \min_{\mathbf{X}_{s,k}} \sum_{k=1}^K \sum_{g=1}^G (\mathbf{P}_{s,g} - \mathbf{X}_{s,k} \mathbf{E}_{g,k})^2 \quad (9)$$

where $\mathbf{X}_{s,k}$ denotes the cell type composition of spot s , with each element restricted to be non-negative; $\mathbf{P}_{s,g}$ denotes the spatial expression count of gene g in spot s ; ϵ_s denotes the residual error between the deconvoluted expression vector $\mathbf{X}_{s,k} \mathbf{E}_{g,k}$ and true spatial

expression $\mathbf{P}_{s,g}$. By minimizing the loss function $L(\epsilon) = \|\epsilon_s\|_2^2$, we can obtain the $\mathbf{X}_s$ that can best mimic the real cell type composition of spot s under the given cell identities.

To further take the entropy-based weight into account, we extended the common NNLS into an entropy-WNNLS model. We first extend the weight vector $\omega(g)$ to a diagonal matrix $\mathbf{\Omega} \in \mathbb{R}^{G \times G}$:

$$\mathbf{\Omega} = \begin{pmatrix} \omega(1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \omega(G) \end{pmatrix} \quad (10)$$

the corresponding WNNLS model has been extended into:

$$\mathbf{P}_s \sqrt{\mathbf{\Omega}} - \mathbf{X}_s \mathbf{E} \sqrt{\mathbf{\Omega}} \quad (11)$$

The loss function of the WNNLS model weighted by the diagonal matrix $\mathbf{\Omega}$ is:

$$\begin{aligned} \text{Min} \tilde{L}(\epsilon) &= \min \|\tilde{\epsilon}_s\|_2^2 = \min_{\mathbf{X}_{s,k}} \|\mathbf{P}_s \sqrt{\mathbf{\Omega}} - \mathbf{X}_s \mathbf{E} \sqrt{\mathbf{\Omega}}\|_2^2 \\ &= \min_{\mathbf{X}_{s,k}} \sum_{k=1}^K \sum_{g=1}^G \left(\mathbf{P}_{s,g} \sqrt{\mathbf{\Omega}_{g,g}} - \mathbf{X}_{s,k} \mathbf{E}_{g,k} \sqrt{\mathbf{\Omega}_{g,g}} \right)^2 \end{aligned} \quad (12)$$

where each element in $\mathbf{X}_{s,k}$ is restrained to be non-negative. Each row of the cell-type composition spectrum $\mathbf{X}$ is normalized into $[0, 1]$ by $\mathbf{X}_{s,k} = \frac{\mathbf{X}_{s,k}}{\sum_{i=1}^K \mathbf{X}_{s,i}}$.

By minimizing the weighted loss function $\text{min} \tilde{L}(\epsilon) = \min \|\tilde{\epsilon}_s\|_2^2$, we can highlight the contribution of genes indicative of cell identities, and finally calculate the cell-type composition spectrum $\mathbf{X}$ in a more biological way.

Referring to SPOTlight, by selecting all cells in $\tilde{\mathbf{B}}$ from the same cell type k , and computing the median of each column of $\mathbf{E}$ for a consensus cell type-specific topic signature, we use $\mathbf{H}$ to further correct $\mathbf{E}$.

Additionally, while UCASpatial is implemented in R, its deconvolution algorithm NMF is implemented with an efficient C++ code that is linked back to the main functions of UCASpatial through Rcpp, ensuring scalable computation.

Finally, considering that the number of cells in each spot is generally within a certain range, for example, 5–20 cells for 10x Visium, we believe that a proportion less than the threshold will be noise in the calculation process. By default, we use a less stringent filter of 0.01 to remove potential noise in the cell-type composition $\mathbf{X}$.

Optional other weighting strategies

In addition to an entropy-based weighting strategy, we implemented alternative weighting strategies to accommodate diverse user requirements and analytical contexts—including mutual information (MI) weighting and variance-based weighting—are provided and described as follows.

Variance-based weighting: for each gene, mean expression levels were computed across predefined cellular subpopulations (clusters) using normalized scRNA-seq data from the selected assay and data slot. The variance of these cluster-wise mean expressions served as a specificity metric, reflecting the extent of differential expression between subpopulations. These variances were linearly normalized to the $[0, 1]$ range to obtain variance-based weights. To integrate gene expression magnitude and percentage, these weights were combined multiplicatively with average log fold change and cluster-wise expression frequency metrics. The final combined weight was bounded to a maximum value of 1.

MI-based weighting: to capture nonlinear or complex dependencies between gene expression and subpopulation identity, gene expression was binarized (expressed versus non-expressed) across cells. MI between this binary expression state and cluster labels was computed using empirical entropy-based estimators, quantifying the reduction in uncertainty of cluster identity given gene expression. MI scores were rescaled to $[0, 1]$ and combined with log fold change and expression percentage as above. The resulting weight was capped at 1.

Notably, the weighting strategy can be conveniently adjusted via the *weight.strategy* parameter interface provided in UCASpatial.

Implementation of the reference diagnosis tool

To systematically assess the quality and specificity of the scRNA-seq reference prior to deconvolution, and to address potential biases arising from dissociation or cell type definition, we implemented a reference diagnosis tool within the UCASpatial framework. This tool evaluates whether the user-provided reference cell subpopulations possess distinct, recoverable transcriptional features under the algorithm's feature extraction mechanism.

Specifically, the tool leverages the correspondence between the extracted latent topics (cGEPs) and the user-annotated cell subpopulations. We visualize this relationship using a dot plot matrix, where the x-axis represents the reference cell subpopulations and the y-axis represents the extracted NMF topics. The size and color intensity of each dot reflect the normalized weight or contribution of a specific topic to a cell subpopulation.

This tool provides a practical and systematic guideline for users: if the diagnosis indicates a poor mapping, the reference requires refinement (e.g., re-clustering, removing low-quality cells, or acquiring an alternative reference).

Simulation strategy and evaluation metrics

Simulation strategy. We generated simulated spatial transcriptomic data sets from a reference CRC scRNA-seq data set (Supplementary Fig. S1A–D, see in the *Human specimens (in-house cohort)* part). To simulate different situations of tumor microenvironment, we generated three different regions, including TC, tumor margin (TM), and tumor stroma (TS)^{29,41}. Each region has five replicates, with each replicate containing $N_{\text{spot}} = 50$ spots, generated based on random seeds (19980205, 19990911, 19960330, 20000502, and 20230509). The number of cells in each spot is restricted to less than 20. To mimic the true situation as much as possible, we also imposed restrictions on the types of cells present in different regions, so that spots from different regions have different cell type co-localization patterns. Detailed restrictions for regional simulation are listed in Supplementary Table 1. To better mimic the real capture, if the synthetic spot had over 25,000 UMI counts, it would be randomly down-sampled into 20,000 UMI counts.

We applied different methods (see the “Compared methods” section) to deconvolute the above datasets. In each analysis, we provided identical simulated spatial data and reference scRNA-seq data as input for all methods. To ensure an equitable assessment of algorithm performance, we adopted the identical threshold of 0.05 (established on spots containing up to 20 cells) as UCASpatial for algorithms lacking default filtering thresholds, aiming to eliminate noise in assessing the ability to determine whether individual cell types exist at a spot (i.e., spots with non-zero cell abundance or not).

Construct a complexity hierarchy. To evaluate UCASpatial in microenvironments with different complexity, we rank the complexity hierarchy of these simulated spots by calculating their cell-type composition entropy: $\text{Entropy}(s) = -\sum_{i=1}^K \mathbf{X}_{s,i} \times \log \mathbf{X}_{s,i}$, where $\mathbf{X}_{s,i}$ denotes the proportion of cell type i in spot s . We classify them into three groups (low, medium, and high), with $N_{\text{spot}} = 250$ in each group, where higher entropy indicates a higher level of complexity.

Evaluation metrics. In each simulation replicate, we calculated the true cell-subpopulation proportions on each spatial location as the number of cells in each cell subpopulation divided by the total number of cells in the location. $T_{s,k}$ and $P_{s,k}$ represent the true and predicted proportion of cell subpopulation k , respectively, in the s th spot through the number of total cell subpopulations K . To evaluate the performance of the methods comprehensively, the main benchmark metrics used were PCC, RMSE, and F1 score. The definitions of these metrics are provided below.

PCC. The cell subpopulation-level PCC value was calculated between the predicted proportion of the specific cell subpopulation P_k and the ground truth proportion of the specific cell subpopulation T_k using the following equation:

$$\text{PCC}_{\text{cell}}(P_k, T_k) = \frac{E[P_k, T_k] - E[P_k]E[T_k]}{\sqrt{E[P_k^2] - (E[P_k])^2} \sqrt{E[T_k^2] - (E[T_k])^2}} \quad (13)$$

The spot level PCC value was calculated between the predicted spot composition of P_s and the ground truth spot composition T_s using the following equation:

$$\text{PCC}_{\text{spot}}(P_s, T_s) = \frac{E[P_s, T_s] - E[P_s]E[T_s]}{\sqrt{E[P_s^2] - (E[P_s])^2} \sqrt{E[T_s^2] - (E[T_s])^2}} \quad (14)$$

RMSE. The cell subpopulation-level RMSE was calculated between the predicted proportion of the specific cell subpopulation P_k and the ground truth proportion of the specific cell subpopulation T_k using the following equation:

$$\text{RMSE}_{\text{cell}}(P_k, T_k) = \sqrt{\frac{1}{S} \sum_{s=1}^S (T_{s,k} - P_{s,k})^2} \quad (15)$$

The spot level RMSE value was calculated between the predicted spot composition of P_s and the ground truth spot composition T_s using the following equation:

$$\text{RMSE}_{\text{spot}}(P_s, T_s) = \sqrt{\frac{1}{K} \sum_{k=1}^K (T_{s,k} - P_{s,k})^2} \quad (16)$$

F1 score: F1 score was selected to evaluate the ability of alternative methods to determine the presence of individual cell types at specific locations, i.e., with non-zero cell abundances, based on the confusion matrix. A confusion matrix is a table used in the field of machine learning and statistics to assess the performance of a classification model. It summarizes the performance of a classification algorithm by showing the counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions for each class in a dataset.

$$\text{F1score} = \frac{2 \times \frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \quad (17)$$

Compared methods. We compared UCASpatial with five deconvolution methods: (1) SPOTlight (version 0.1.7), (2) RCTD (version 1.2.0) (3) cell2location (version 0.8a0), (4) CARD (version 1.0), and (5) stereoscope (version 0.3.1). (6) Spotiphy (version 0.1.3). Results of the stereoscope were computed with parameters “-sce 25000 -n 5000 -ste 25000 -stb 500 -scb 500 -gp”. Results of the RCTD were used to test the ‘full model’. For other methods, we followed the tutorial on the corresponding GitHub pages and used the recommended default parameter settings for deconvolution analysis.

Computer platform for evaluating runtime and memory usage. We ran CPU tests of the seven methods on a computer cluster with an identical CPU platform (2.1 GHz, 60.5 MB L3 cache, 44 CPU cores) and 384 GB memory (DDR4 2666 MHz). The GPU tests for cell2location and stereoscope were performed on the same computer with an additional GPU (GeForce RTX 4090 with 24 GB memory). To assess the impact of various data attributes (including the number of cells in scRNA-seq data and the number of spots in spatial data) on the computing resources consumed by those 7 methods, we constructed simulated spatial datasets with: (1) varying number of spots (50, 500, 5000, and 50,000), 12,500 cells, and 20 cell types; (2) varying numbers of cells (approximately 1000, 2500, 5000, 10,000, 15,000, 25,000, and 35,000), 20 cell types, and 5000 spots. For the memory evaluation, we used the simulated dataset with 50,000 spots ST and about 12,500 cells, and 20 cell types. Then, we evaluated the impacts of these data attributes on computing resources consumed.

Analysis of human CRC 10x Visium HD data. To test UCASpatial’s ability on the 10x Visium HD platform, we applied UCASpatial to 10x Visium HD data from a human CRC sample³¹. We used the recommended bin size of $8 \mu\text{m} \times 8 \mu\text{m}$ for this analysis. With the scRNA-seq data shown in Fig. 3B as a reference, we spatially annotated 34 cell subpopulations using the hyper-parameter setting ‘meta.filter = F’ (Supplementary Fig. S2A). For comparison, we also performed RCTD using the ‘full’ model. Other aforementioned methods did not finish the task for the $8 \mu\text{m}$ bin annotation, possibly due to high memory demands, long run times, or software version incompatibility with 10x Visium HD. The high-resolution images of the deconvolution results have been uploaded to GitHub (<https://github.com/BIGHanLab/UCASpatial>) for better visibility.

In our testing environment (2.1 GHz, 60.5 MB L3 cache, 44 CPU cores), UCASpatial completed the task in just 1.97 h, while RCTD required 2.27 days for complete annotation. Furthermore, we evaluated UCASpatial on 10x Visium HD data with $2 \mu\text{m} \times 2 \mu\text{m}$ bin size, which was processed in 2.14 h, demonstrating its scalability to high-resolution data.

Analysis of mouse hippocampus Slide-seq V2 data. To test UCASpatial’s ability in Slide-seq V2 platform, we used UCASpatial (meta.resolution = 200, meta.purity = 0.98, cos.filter.threshold = 0.1, ent.filter.threshold = 0.2, logfc.threshold = 0.1, min.diff.pct = 0.05, meta.assay = ‘RNA’) to spatially annotate cell subpopulations in Slide-seq V2 data of the mouse hippocampus using published scRNA-seq datasets of mouse hippocampus^{10,98}.

It is noteworthy that for CARD, we utilized a downsampled scRNA-seq reference dataset consistent with the original CARD paper¹² to faithfully reproduce its results, whereas other methods were run using the full single-cell reference. Similar to the strategy in the CARD paper, we didn’t apply cell2location on this dataset due to its heavy computational burden. Results of the stereoscope were computed with parameters “-sce 25000 -n 5000 -ste 25000 -stb 500 -scb 500 -gp”. Results of the RCTD were used with the ‘full model’. For other methods, we followed the tutorial on the corresponding GitHub pages and used the recommended default parameter settings for deconvolution analysis. The visualization results of each method can be found in Supplementary Fig. S2H.

Simulated data generation and method benchmarking based on 10x Visium HD data. Public 10x Visium HD data from human CRC samples were obtained³¹. Cell segmentation and spatial alignment were performed using Space Ranger software (version 4.0.1), and reads were aligned to the hg38 reference genome. Cells were annotated according to their transcriptomic profiles (Supplementary Figs. S2J, K and 3A). To approximate the capture area of a standard Visium spot ($55 \mu\text{m}$ in diameter), we overlaid a grid of $49 \mu\text{m} \times 49 \mu\text{m}$ square bins

onto the coordinate system of each tissue section. For each square bin, we aggregated the transcriptomic counts (UMI counts) from all segmented cells whose centroids lay within its boundaries, thereby generating a composite gene expression profile representing a single “pseudo-spot” (Supplementary Fig. S2I).

To evaluate the performance of each deconvolution method on these pseudo-spots, we employed UCASpatial (with `meta.filter = FALSE`) to spatially annotate cell subpopulations within each pseudo-spot. We did not apply Cell2location to this dataset due to its substantial computational demands. Results from Stereoscope were obtained using the parameters “-sce 500 -n 5000 -ste 500 -stb 500 -scb 500 -gp.” For RCTD, we used the “full model.” For all other methods, we followed the tutorials available on the respective GitHub repositories and applied the recommended default parameter settings for deconvolution analysis.

Detailed deconvolution results of each method are publicly available in the Zenodo repository (see “Data availability”).

Experimental and analytical approach to human CRC

Human specimens (in-house cohort). This study complies with all relevant ethical regulations and was approved by the Institutional Ethics Committee of Beijing Cancer Hospital (approval number: 2023YJZ16). All patients were enrolled and pathologically diagnosed with colon cancer at Peking University Cancer Hospital & Institute (Supplementary Table 2). Written informed consent was obtained from all patients. Clinical and pathological stages were determined according to the guidelines of the AJCC version 8. Detailed clinical information was summarized in Supplementary Table 2. Fresh tumor tissues and patient-matched normal adjacent colon tissues were collected from surgical resection specimens, with no prior exposure to chemotherapy or radiation therapy. The resected samples were stored in RPMI 1640 (Gibco) supplemented with HEPES (Thermo Fisher, 15630080), Penicillin-Streptomycin (Thermo Fisher, 15140122), Amphotericin B (Sigma, S1636), and Gentamicin Sulfate (Gibco, 15710-064) and transported at 4 °C immediately. As for CRC1, CRC2, and CRC3, a small portion of each tissue was frozen and embedded in OCT (Sakura Tissue-Tek), while the remainder of the same tissue was dissociated into a single-cell suspension.

Single-cell RNA-seq for patient samples. Fresh human tumor specimens and patient-matched adjacent normal tissues were washed twice with DPBS (Gibco) and then mechanically disrupted into approximately 1 mm³ pieces in the RPMI-1640 (Gibco) and further enzymatically digested with Human Tumor Dissociation Kit (Miltenyi Biotec, 130-095-929) in the C-tube in combination with gentleMACS™ Dissociator (Miltenyi Biotec, 130-093-235) presetting program `h_tumor_01`, `h_tumor_02`, and `h_tumor_03` sequentially according to the manufacturer’s instruction. Between each program, samples were incubated for 30 min at 37 °C under continuous rotation using a rotator. The dissociated cells were subsequently passed through a 70-μm cell strainer (Biologix, 15-1070) and centrifuged at 300×g for 7 min. After the supernatant was removed, the cell pellet was washed with DPBS. The pelleted cells were resuspended in a cell cryopreservation medium (FBS with 10% DMSO) at a concentration of 1×10^6 cells/mL and further cryopreserved in liquid nitrogen for long-term storage.

On the day of scRNA-seq droplet capture, cells were thawed rapidly in a 37 °C-water bath and washed off DMSO in DPBS. Palleted cells were resuspended in a sorting buffer (PBS with 2% FBS and 1 mM EDTA). Cells were stained with CD45-FITC (BioLegend Cat# 368508, clone 2D1, RRID: AB_2566368, 100 μg/mL, 1:500) for 20 min at 4 °C in the dark. 7-amino-actinomycin D (7-AAD, Sigma-Aldrich Cat# 129935) was freshly added at a final concentration of 1 μg/mL prior to FACS sorting performed on the BD FACS Aria III instrument. Based on FACS analysis, 7-AAD⁺CD45⁺ cells, and 7-AAD⁺CD45[−] cells were sorted into 1.5 mL tubes (ExCell Bio, CS015-0041) and counted manually under the

microscope (Leica), respectively. Libraries were generated using Chromium Next GEM Single Cell 5' Library & Gel Bead Kit v1.1 (10x Genomics, PN-1000165) according to the manufacturer’s protocol. Purified libraries were sequenced by the Illumina NovaSeq 6000 platform with 150-bp paired-end reads.

10x Visium ST. CRC samples were frozen and embedded in OCT (Sakura Tissue-Tek). All tissue blocks were sectioned using the Leica CM1950 cryostat and were cut at 10 μm. Samples were selected on the basis of morphology, orientation (based on H&E staining), and RNA integrity number (RIN) that was obtained using Agilent 2100 Bioanalyzer. The RIN of selected samples ranged from 4.5 to 8.3. The sections were treated as recommended by 10x Genomics, and the optimization procedure showed the optimal permeabilization time determined as 12 min for CRC samples using the Visium Spatial Tissue Optimization Kit (10x Genomics, PN-1000193). Spatial gene expression slides were equilibrated inside the cryostat at −20 °C before sectioning, and all of the reagents used for ST were from the Visium Spatial Gene Expression Kit (10x Genomics, PN-1000187). According to the manufacturer’s instructions, the tissue sections were fixed in chilled methanol at −20 °C for 30 min and stained with H&E. Brightfield histological images were performed on the Olympus Microscopes BX53 at ×20 magnification and then taken by cellSens Dimension (version 1.18) software. Images were exported as tiled TIFFs for analysis. After imaging, tissue sections were permeabilized at 37 °C for the optimal permeabilization time of release of mRNA, as mentioned above. Following permeabilization, mRNA was reverse-transcribed and underwent second-strand synthesis, denaturation, cDNA amplification, and cleanup by SPRIselect beads sequentially (Visium, 10x Genomics, User Guide CG000239). cDNA amplification cycles were determined by quantitative PCR (qPCR). Library quality was evaluated using TapeStation (Agilent Technologies). Sequencing was performed on the Illumina HiSeq X Ten platform.

Single-cell RNA-seq data processing and annotation. In-house single-cell RNA-seq data from 10x Genomics were processed using Cell Ranger (version 6.0.1), aligning to the hg38 reference genome. Low-quality cells were filtered out if they expressed fewer than 500 or more than 7000 genes. Cells with mitochondrial gene expression exceeding 20% were also excluded. Batch effects of in-house and public scRNA-seq datasets were corrected using the integration workflow in the Seurat R package (version 3.1.4)⁹⁹.

To identify and refine cell subpopulations, we employed a multi-step approach. Initially, we employed log-normalization with a scaling factor of 10,000 to adjust for variations in library size among cells. We then identified the top 2000 highly variable genes utilizing the Seurat package and conducted principal component analysis (PCA) on these genes, retaining the top 20 principal components that determined by the elbow method on a Scree plot. Next, we proceeded with an iterative clustering approach to delineate clusters within the single-cell data. Generally, this iterative process initially identified clusters using the Leiden algorithm, followed by a merging of clusters from cell types that appeared to be identical. In the context of certain cell types characterized by heightened heterogeneity, such as T&NK, myeloid, and epithelial cells, we performed cell-type-specific reintegration based on the top 2000 variable features to capture more signature genes that can represent the heterogeneity of subpopulations. Then we performed sub-clustering to further identify subpopulations using Leiden clustering. This iterative clustering process allowed us to refine the initial clustering, enabling the splitting of initial clusters that contained multiple subpopulations into more specific lower-level subpopulations. One T cell and myeloid cluster with few specific markers but exhibiting high expression of cell cycle-related genes was removed, as these genes could not be fully regressed out after cell cycle score scaling. In total, we defined 34 cell types in CRC, and this

CRC scRNA-seq reference was used throughout the manuscript to maintain consistency in presenting spatial mapping results (Fig. 3B and Supplementary Fig. S3B–D).

To infer tumor clonal diversity from scRNA-seq data, CNVs were evaluated using the InferCNV v1.2.2R package (<https://github.com/broadinstitute/infercnv>). CNVs of epithelial cells were estimated using non-epithelial cells as a reference. We used a cutoff of 0.1 for the minimum average read counts per gene among reference cells, clustered each group of cells separately, and denoised our output (Supplementary Fig. S5L).

ST data processing. Raw sequencing reads of ST were quality-checked and mapped by Space Ranger v1.3.0. The gene-spot matrices generated after ST data processing from ST and Visium samples were analyzed with the Seurat package in R. The processed data were used for downstream analysis of UCASpatial (meta.resolution = 300, meta.purity = 0.9, other parameters using default) and other methods (using default parameters). Using the threshold of 3% to filter the noise after decomposition by each method.

Gene signature score calculated in ST. Selected the top 20 genes of each cell subpopulation provided in Supplementary Data 1 or Supplementary Data 2 as the signature gene list for the murine wound healing ST dataset and human CRC ST dataset, respectively. Used the “AddModuleScore” method in the Seurat R package to calculate the signature score of each cell subpopulation in the ST dataset using the signature gene list.

Cellular distance evaluation in ST. In order to quantify the distance relationships between different cell subpopulations in ST data, we developed an algorithm based on the deconvolution results of UCASpatial. To calculate the spatial distance between cell subpopulation A and cell subpopulation B in ST data, we first conduct a breadth-first search for each spot containing cell subpopulation A to identify the nearest spot containing cell subpopulation B, and record the distance, as well as the proportions of subpopulation A and B within their respective spots. Then we calculate a weighted average on each shortest distance, taking into account the proportions of subpopulation A and subpopulation B within the spots simultaneously. This yields the distribution of spatial distances between the two subpopulations.

Multi-color immunohistochemistry (mIHC). mIHC was performed manually using the Opal 7-Color Manual IHC Kit (AKOYA BIOSCIENCES, Cat# NEL81 1001KT), following the manufacturer’s protocol. Fresh human tumor specimens were excised and fixed in 10% formalin for 24 h at 4 °C. The fixed tissues were then embedded in paraffin and sectioned at 5 µm thickness for histological evaluation. Tissue sections were baked in the oven at 65 °C for 1 h, dewaxed with xylene, and rehydrated through a graded series of alcohol concentrations (100%, 95%, and 70%).

Endogenous Peroxidase Blocking Buffer (Beyotime Cat# P0100A) is used in the method to quench endogenous peroxidase activity. Microwave treatment is used for antigen retrieval and removing antibodies after a target has been detected. The slide was boiled in the microwave oven for 45 s at 100% power and then 15 min at 20% power, followed by blocking for 10 min at room temperature. Primary antibodies against protein CD14 (Proteintech Cat# 60253-1-Ig, clone 2CID9, RRID:AB_2881374; 1:200), C1qC (Proteintech Cat# 16889-1-AP, polyclonal, RRID:AB_2878328; 1:100), IGF1 (Proteintech Cat# 28530-1-AP, polyclonal, RRID:AB_2881164; 1:100), SPP1 (Proteintech Cat# 22952-1-AP, polyclonal, RRID:AB_2783651; 1:200), PDPN (Proteintech Cat# 11629-1-AP, polyclonal, RRID:AB_2162067; 1:100), ANXA1 (Abcam Cat# ab214486, clone EPR19342, RRID:AB_2890907; 1:4000), CXCL13 (Abcam Cat# ab246518, clone EPR23400-92, RRID:AB_2894862;

1:1000), PD-1 (Abcam Cat# ab137132, clone EPR4877(2), RRID:AB_2894867; 1:500), CD4 (Abcam Cat# ab133616, clone EPR6855, RRID:AB_2750883; 1:100), FOXP3 (Abcam Cat# ab20034, clone 236 A/E7, RRID:AB_445284, 1:1000), IL-17a (Proteintech Cat# 26163-1-AP, polyclonal, RRID:AB_2880409, 1:1000), HLA-DR (Abcam Cat# ab92511, clone EPR3692, RRID:AB_10563656, 1:400), CD20 (Abcam Cat# ab78237, clone EP459Y, RRID:AB_1640323, 1:250), CLEC9A (Proteintech Cat# 55451-1-AP, polyclonal, RRID:AB_2881338; 1:200), CD1c (Abcam Cat# ab246520, clone EPR23189-196, RRID:AB_2894854; 1:1000), CD208 (Abcam Cat# ab271053, clone EPR24265-8, RRID:AB_2943181; 1:500), GZMK (Abcam Cat# ab282703, clone EPR24601-164, RRID:AB_3107073; 1:500), CD8 (Abcam Cat# ab17147, clone C8/144B, RRID:AB_443686; 1:100), CD103 (Abcam Cat# ab224202, clone EPR22590-27, RRID:AB_2936238; 1:800), CTHRC1 (Abcam Cat# ab256458, clone EPR22851-145, RRID:AB_3096210; 1:400), and TCF1 (Cell Signaling Technology Cat# 2203 T, clone C63D9; 1:200) were sequentially applied at 4 °C overnight or at room temperature for 1 h.

The secondary antibody Polymer HRP Ms + Rb (Cat# ARH1001EA) was applied and incubated at room temperature for 10 min. Signal amplification was performed using the Opal Fluorophore Working Solution on the slides at room temperature for 10 min. Finally, nuclei were stained with DAPI (Invitrogen Cat# D1306, RRID: AB_2629482, 5 mg/mL, 1:2500). Multi-color IHC imaging was collected using the Vectra Polaris Imaging System (PerkinElmer)¹⁰⁰ at 20× magnification and integrated using the InForm Advanced Image Analysis Software (PerkinElmer, version 2.4.1).

Cell segmentation and quantification were carried out using QuPath¹⁰¹. In short, DAPI stains were used for cell detection with the following parameters: requestedPixelSize 1 µm, backgroundRadius 10 µm, sigma 0.5 µm, minArea 10 µm, maxArea 400 µm, threshold 0.1, cellExpansion 1.5 µm. Cell subpopulations were classified based on the combinatorial expression of key protein markers (Supplementary Fig. S4C). The distances between CD8⁺ T cell subpopulations and the tumor region were calculated using the “Distance to annotation 2D” function in QuPath. For the pseudo-spot visualization, we first divide the mIHC data into densely distributed bins, each with a width and length of approximately 50 µm. Then, we calculate the proportion of each cell subpopulation in the pseudo-spots by dividing the number of specified cell subpopulations by the total cell count defined by DAPI.

Tumor clonal inference and clonalTIME quantification using UCASpatial-clonalTIME. To infer tumor clonal diversity within each ST slice, CNVs were evaluated by ordering genes according to their chromosomal location and applying a moving average to the relative expression values. Genes were arranged according to their genomic localization using the SpatialInferCNV R package (version 0.1.0)³³. CNVs in high epithelial spots (overall epithelial proportion > 40%) were estimated using reference spots containing few or no malignant cells selected from UCASpatial annotated remaining spots (overall epithelial proportion < 40%). Considering the non-dispersed spatial distribution of epithelial cells, scattered high epithelial spots (neighboring same-type spots < 1, step = 1) were considered noise and defined as remaining regions. The spot-resolution CNV profiles for each ST slice were then estimated using the SpatialInferCNV workflow (cutoff = 0.1, cluster_by_groups = FALSE, denoise = TRUE, HMM = FALSE), excluding chrM, chrX, and chrY genes. To further define tumor regions and infer tumor clonal diversity based on CNV data, we constructed dendrogram trees to determine clonal groupings of each spot. Dendrogram nodes with at least one high-probability CNV region were defined as tumor clones and assigned a specific identity. Nodes without high-probability CNVs were assigned to normal epithelial or remaining regions based on their histological structure. The average cell subpopulation composition across all spots within a clonal region was defined as the cell subpopulation composition of that clonalTIME (Fig. 3H). To perform a comprehensive comparison of individual CNVs

across clones, we initially identified the CNV locations and statuses of each clone using an HMM model in SpatialInferCNV (Fig. 4A–C and Supplementary Fig. S5A). CNV regions located on the same chromosome with similar variation patterns in the TCGA-COAD dataset were merged if their PCC ($\log_2\text{FC}_{\text{CNV}}$) > 0.8 , where $\log_2\text{FC}_{\text{CNV}}$ denotes the $\log_2$ fold change of the copy number of the tumor over the copy number of normal. These CNVs were then separated based on the location of the centromere (Supplementary Table 4).

ClonalTIME phenotyping. To classify the clonalTIME phenotype of each tumor clone, the scaled proportion of the T cell subpopulation of each clone was clustered using the K-means function (centers = 2). ClonalTIMes were classified as T cell-rich or T cell-excluded based on the proportion of T cell subpopulation within each clustering group (Fig. 3I and Supplementary Fig. S4J, K). To determine whether a given cell subpopulation was enriched in a specific clonalTIME phenotype, we calculated the ratio of observed cell number to expected cell number (Ro/e) for T cell-rich clonalTIME, where the expected proportion of cell subpopulation was determined from the chi-squared test. We assumed that a subpopulation was enriched in T cell-rich clonalTIME if $\text{Ro/e} > 1.2$ and enriched in T cell-excluded clonalTIME if $\text{Ro/e} < 0.8$.

Multi-cellular community analysis. To identify cell subpopulation colocalization patterns, we first adjusted the cellular composition of each spot by averaging the proportions of cell subpopulations within that spot and its immediate neighbors (step = 1). Subsequently, we employed the Pearson method to calculate the correlations among the adjusted cell subpopulations across all spots within the tumor region. To identify distinct multicellular communities, we applied the ward.D2 hierarchical clustering algorithm to the correlation matrix. This analysis yielded three multicellular communities in the human CRC part, and two multicellular communities in the murine wound healing part (Figs. 3J and 6D).

The functionality of each multicellular community was characterized based on the shared enriched pathways exhibited by its constituent cell subpopulations. Shared enriched pathways were defined as those that were statistically significantly enriched in at least half of the subpopulations within each community, with a minimum of one subpopulation per community contributing to this enrichment. To address the issue of redundancy in pathway databases, the identified shared enriched pathways were subjected to hierarchical clustering using the pairwise Pearson correlation distance calculated across the gene space. Here, each pathway was represented by its associated differentially expressed genes. Pathway clusters were named manually.

Multi-cellular community score. The proportions of each cell subpopulation were first normalized by the proportion of non-epithelial cells, ensuring the sum of non-epithelial cell proportions of a spot equals one. To enable inter-cell type comparability, we scaled the proportion of each cell subpopulation using the 99.5th percentile across all tumor spots, resulting in a score for each cell subpopulation. The average score of all cell subpopulations within a community was then calculated as the community score.

Spatial quantification of TE expression. We quantify the expression of TEs using spatial transcriptomic data based on a previously reported algorithm⁴⁷. This algorithm allocates TE reads to TE metagenes based on the TE type-specific sequence (Fig. 4H). In brief, the genes/TEs and cell count matrix were generated according to default parameters and normalized using the NormalizeData function. Clonal-specific TE expression programs were identified by comparing each tumor clone region using the FindAllMarkers function. The list of CRC-specifically upregulated TEs was downloaded from previous literature⁴⁸. The

HERV-H expression level was calculated as the average expression of previously reported *HERV-H* elements (*HERVH-int*, *LTR7*, *LTR7B*, and *LTR7Y*) (Fig. 4J).

Differential expression analysis using ST. Differentially expressed genes across each clone were analyzed using the Seurat R package (version 4.0.5). Space Ranger output files were imported, and the data were normalized and scaled using the default NormalizeData and ScaleData functions. Differential gene expression analyses were performed comparing groups using the FindMarkers function, with the following parameters: $\log_2(\text{fold change}) \geq 0.1$ and adjusted p -value ≤ 0.05 (Fig. 4G). The list of CHIP-seq-based STAT1 target genes was downloaded from previous literature¹⁰². For differentially expressed genes in the CRC4 slice, gene ontology (GO) enrichment analysis was performed by clusterProfiler (v4.6.0) (Fig. 4G)¹⁰³.

TCGA-COAD data analysis. Gene expression and clinical data for Colon Adenocarcinoma (COAD) were downloaded from The Cancer Genome Atlas (TCGA). Average copy number data were downloaded from the UCSC Xena browser (<https://xena.ucsc.edu>). For CNV analysis, only the samples with CNV information were obtained. The average $\log_2\text{FC}_{\text{CNV}} > 0.35$ was defined as copy number gain, while the average $\log_2\text{FC}_{\text{CNV}} < -0.35$ was defined as copy number loss (Fig. 4F). The signatures of each cell subpopulation were obtained in CRC single-cell data by the FindAllMarkers function in the Seurat package with default parameters. The normalized expression of the top 10 signatures in each cell subpopulation was used to calculate the average signature score. The reads per million (RPM) of known human endogenous retroviruses (HERV) in TCGA-COAD were downloaded from a previous report^{104,105}. For HERV analysis, samples without HERV expression information were excluded. Overall *HERV-H* expression was calculated as the average expression of known transcribed *HERV-H* loci (*HERV-HI-7*) (Fig. 4K)¹⁰⁴. For CIN assessment, samples were stratified into tertiles by aneuploidy score, followed by comparative analysis of T cell marker expression across groups. The pathway activity of each sample was calculated using GSVA with default parameter settings (Supplementary Fig. S5K)¹⁰⁶. Survival analyses were then performed with the survival R package (version 3.5-3) (Fig. 4L and Supplementary Fig. S5R).

MSK-IMPACT data analysis. The genomic and clinical data of 71 CRC patients with ICB therapy were downloaded from published studies^{51,52}. The copy number status of each locus was determined using the GISTIC v2.0 function in the GenePattern platform^{107,108}. The average $\log_2\text{FC}_{\text{CNV}} > 0.35$ was defined as copy number gain, while the average $\log_2(\text{copy number}/2) < -0.35$ was defined as copy number loss. Survival analyses were then performed with the survival R package (Fig. 4L and Supplementary Fig. S5P–Q).

Experimental and analytical approach to mouse ear

Ear wound creation and treatment. All mouse experiments were conducted in accordance with all relevant ethical regulations and were approved by the Institutional Animal Care and Use Committees of Beijing Institute of Genomics, Chinese Academy of Sciences (approval number: 2019A010). MRL (RRID: IMSR_JAX:000486) and C57BL/6 (B6; RRID: MGI:2159769) mice of both sexes, aged 8–10 weeks, were used for ear wound creation. All mice were housed under specific pathogen-free (SPF) conditions in the barrier facilities of the Tsinghua University Laboratory Animal Center and the Laboratory Animal Center of the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences. Animals were maintained under a 12 h light/12 h dark cycle, at an ambient temperature of $22 \pm 1^\circ\text{C}$ and a relative humidity of $50 \pm 10\%$, with food and water provided ad libitum. For wound generation, mice were briefly anesthetized with isoflurane. The dorsal and ventral aspects of mouse ears were sanitized with cotton balls saturated in 75% alcohol. Mouse ears were delicately flattened using

forceps, and a tissue-penetrating ear puncher (2 mm) was used to create perforating injuries within the vascularized regions of the ear. Hemostatic powder was applied to the wounds using disposable cotton swabs. All instruments were sterilized under high pressure before use.

To block the differentiation of *Fmod*⁺ fibroblasts, mice were either locally injected with rosiglitazone (Selleck Cat# S2556, 3 mg/kg) every other day post-injury or smeared with rosiglitazone dissolved in petrolatum every day. For the injection administration, the control group received an equivalent volume of DMSO. For the topical administration, the control group received petrolatum containing DMSO.

To block IL11-IL11RA signaling, mice were locally injected with anti-IL11RA antibody (Invitrogen Cat# PA5-47667, 2.5 µg/wound) at day 3 post-injury. The control group received an equivalent volume of anti-rabbit IgG.

Flow cytometry. Single-cell suspensions were obtained from the ear tissues of B6 and MRL mice at day 7 post-injury. Specifically, mouse ear tissues were mechanically minced and then incubated at 37 °C for 5–10 min with 40 µL collagenase II (50 mg/mL) and 40 µL collagenase IV (50 mg/mL). Following a brief low-speed centrifugation (300×g) to precipitate undigested tissue fragments, the supernatant was transferred into a stop solution composed of fetal bovine serum, EDTA, and D-HBSS. Additional digestion enzymes were added to the partially digested tissue fragments, and the digestion process was repeated until complete tissue digestion was achieved. The digested mixture was filtered through a 70-µm membrane and centrifuged at 4 °C at 1200 rpm for 10 min. After discarding the supernatant, the cell pellet was washed in FACS buffer (PBS supplemented with 2% FBS and 1 mM EDTA). The cells were first stained with Zombie Aqua™ Fixable Viability Kit (BioLegend Cat# 423102, 1:1000) to exclude dead cells, followed by incubation with Fc Block (BioXCell Cat# BE0307, clone 2.4G2, RRID:AB_2736987) at a 1 µg/mL concentration for 10 min. Subsequently, the following antibodies were used for surface staining: CD45-BUV395 (BD Biosciences Cat# 564279, clone 30-F11, RRID:AB_2651134), CD31-FITC (Invitrogen Cat# 11-0311-81, clone 390, RRID:AB_465012), CD29-PE (Invitrogen Cat# 12-0291-81, clone HMB1-1, RRID:AB_657733), Sca-1-APC/Cy7 (BioLegend Cat# 108125, clone D7, RRID:AB_10639725), CD34-Pacific Blue (BioLegend Cat# 119321, clone MEC14.7, RRID:AB_10900980), CD55-PE/Cy7 (BioLegend Cat# 131813, clone RIKO-3, RRID:AB_2800633), and CD90.2-BV711 (BioLegend Cat# 105349, clone 30-H12, RRID:AB_2800564). Antibodies were applied at a 0.2 µg/mL concentration. Staining was performed for 30 min on ice in the dark. Flow cytometry was performed using a BD Symphony A5 cytometer for analysis and a BD Fusion cell sorter for sorting experiments. The flow cytometry results were analyzed using FlowJo™ v10 Software (BD Life Sciences).

mIHC. At predetermined time points, murine ears were excised and fixed in 10% formalin for 48 h at 4 °C. The fixed tissues were then embedded in paraffin and sectioned at 7 µm thickness for histological evaluation. Tissue sections were then deparaffinized, rehydrated, and stained using 7-color multispectral Opal reagents (AKOYA BIOSCIENCES, Cat# number NEL81 1001KT). Heat-mediated antigen retrieval was performed using a microwave for rehydrated tissue sections. Primary antibodies against CD11b (Abcam Cat# ab133357, clone EPR1344, RRID:AB_2650514; 1:500), IL-1β polyclonal antibody (Invitrogen Cat# P420B, RRID:AB_223478; 1:200), CD163 (Abcam Cat# ab182422, clone EPR19518, RRID:AB_2753196; 1:1000), CD68 (Proteintech Cat# 28058-1-AP, RRID:AB_2881049; 1:2000), CD74 (Abcam Cat# ab289885, clone EPR25399-94, 1:2000), GPNMB (Proteintech Cat# 66926-1-Ig, clone 2B10B8, RRID:AB_2882253; 1:200), α-SMA (Abcam Cat# ab124964, clone EPR5368, RRID:AB_11129103; 1:2000), PDGFR alpha (PDGFRα, Abcam Cat# ab96569, clone 16A1, RRID:AB_10687154; 1:500), CD34 (Invitrogen Cat# MA5-29674, clone 013,

RRID:AB_2785500; 1:2500), OGN (Proteintech Cat# 12755-1-AP, polyclonal, RRID:AB_10646479; 1:200), SOX9 (CST Cat# 82630T, clone D8G8H, 1:500), Collagen Type II (Proteintech Cat# 28459-1-AP, polyclonal, RRID:AB_2881147; 1:200), IGFBP5 (Proteintech Cat# 55205-1-AP, polyclonal, RRID:AB_2736835; 1:200), PRG4 (Proteintech Cat# 25732-1-AP, polyclonal, RRID:AB_2880215; 1:200), K14 (Santa Cruz Cat# sc-53253, clone LL001, RRID:AB_2134820; 1:300), CD31 (Abcam Cat# ab281583, clone RM1006, RRID:AB_3096925; 1:2000), FcεRIα (Proteintech Cat#10980-1-AP, polyclonal, RRID:AB_2103033; 1:50), Ly6G (Abclonal Cat# A22270, RRID:AB_3718562; 1:200), IL11RA (Proteintech Cat# 10264-1-AP, polyclonal, RRID:AB_2125660; 1:500) and Osteopontin (Proteintech Cat# 22952-1-AP, polyclonal, RRID:AB_2783651; 1:500) were individually applied.

The secondary antibody Polymer HRP Ms + Rb (Cat# ARH1 001EA) was applied and incubated at room temperature for 10 min. Signal amplification was performed using the Opal Fluorophore Working Solution on the slides at room temperature for 10 min. Finally, nuclei were stained with DAPI (Invitrogen Cat# D1306, RRID:AB_2629482, 5 mg/mL, 1:2500). mIHC imaging was collected using the Vectra Polaris Imaging System (PerkinElmer)¹⁰⁰ at 20× magnification and analyzed using the InForm Advanced Image Analysis Software (PerkinElmer, version 2.4.1).

Cell segmentation and quantification were carried out using QuPath¹⁰¹. In short, DAPI stains were used for cell detection with the following parameters: requestedPixelSize 0.5 µm, backgroundRadius 8 µm, sigma 0.5 µm, minArea 10 µm, maxArea 400 µm, threshold 0.5, cellExpansion 5 µm. Cell subpopulations were identified by training an object classifier based on the combination of the markers' expression. In the epidermis, keratinocytes were measured, while in the dermis, a comprehensive identification and quantification of other cellular subpopulations were carried out. For the pseudo-spot visualization, we first divide the mIHC data into densely distributed bins, each with a width and length of approximately 50 µm. Then, we calculate the proportion of each cell subpopulation in the pseudo-spots by dividing the number of specified cell subpopulations by the total cell count defined by DAPI.

Single-cell RNA-seq. Single-cell suspensions were obtained from B6 and MRL mouse ear tissue at day 7 post-wounding. Calcein (Invitrogen Cat# C3099) and 7-AAD (Sigma-Aldrich Cat# 129935) were freshly added at final concentrations of 1 µM and 1 µg/mL, respectively, before cell sorting. Then FITC⁺ 7-AAD⁻ single cells were sorted for library construction of scRNA-seq. The libraries were performed using Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit V3 (PN-1000075) purchased from 10x Genomics according to the protocols provided by manufacturers. The libraries were sequenced on an Illumina Novaseq 6000 platform.

Spatial gene expression assay. Fresh mouse ear tissues were embedded in the Optimal Cutting Temperature Compound (OCT, Sakura Tissue-TEK), and stored at -80 °C. The 10-µm section was placed on the pre-chilled Optimization slides (Visium, 10x Genomics, PN-1000193). After collecting histological morphological information, the tissue was processed within the optimal permeabilization time of 18 min to digest and release mRNA from tissue slides. The released mRNA was subsequently captured by proximal oligonucleotides within the tissue, culminating in the generation of the respective libraries. The libraries were sequenced on an Illumina HiSeq X Ten platform.

Single-cell RNA-seq data processing. The 10x Genomics scRNA-seq data were aligned and quantified using the Cell Ranger Single-Cell Software Suite (version 3.0.2) against the mm10 reference genome. Genes with fewer detected cell numbers (fewer than 3) and cells with fewer detected genes (fewer than 200) were removed. To further filter out low-quality data, the following criteria were applied to filter cells:

(1) the number of expressed genes greater than 500 or less than 5000 and (2) mitochondrial proportion less than 20%. To cluster single cells by their expression, the unsupervised graph-based clustering algorithm implemented in the Seurat R package (version 3.1.4) was used. The top 2000 variable genes identified by the vst method were used to perform PCA. The RunUMAP and RunTSNE functions were employed to generate UMAP and t-SNE dimensionality reduction results, while the FindNeighbors function was used to construct a shared nearest neighbor graph. The DoubletFinder was utilized to predict potential doublets and filter them out.

ST data processing. Using the Space Ranger software (<https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/tutorials>), the 10x Visium ST data were aligned and quantified against the mm10 version of the mouse reference genome. Downstream analysis was conducted using the R package Seurat. Integration of unpaired scRNA-seq data with ST sequencing data was performed for deconvolution analysis to obtain the proportion of each cell subtype within each spot of the spatial transcriptome. Parameters for UCASpatial: downsample_n = 1000, meta.resolution = 200, meta.filter = 0.95, cos.filter.threshold = 0.1, the others use the default. Parameters for other methods: Default. Using the threshold of 3% to filter the noise after decomposition by each method (Supplementary Fig. S6C).

ST data visualization. The original data placed mouse ear tissues on two slides: tissues from day 0, day 7, and day 15 were on the first slide, and tissues from day 3 were on the second slide. To better visualize the ST data of mouse ears, we merged all the tissues into one slide and aligned them based on the time point post-wounding. Specifically, we first extracted the coordinates of each spot in all 4 tissues. We then transformed the coordinates of spots from day 3, with each row replaced by the column, and each column replaced by the row. We next transformed the coordinates of each spot's columns, abiding by the time points. ggplot2 was used for visualizing ST data after coordinate transformation.

The murine ear tissues have been annotated into four groups of regions based on H&E staining: unwounded (day 0), wound bed (1 mm radius), wound proximal (0–1 mm from the wound bed), and wound distal (>1 mm from the wound bed).

Functional enrichment analysis in murine wound healing data. The signatures of macrophage subpopulations, *Fmod*⁺ fibroblast, and *Igf1bp5*⁺ chondrocyte in the murine wound healing dataset were obtained in scRNA-seq data by the FindAllMarkers function in the Seurat package with min.pct = 0.1, min.diff.pct = 0.1, logfc.threshold = 0.25, and other default parameters. Metascape was used to perform functional enrichment analysis for *Fmod*⁺ fibroblast and *Igf1bp5*⁺ chondrocyte (Supplementary Fig. S7L, N)¹⁰⁹. The enrichKEGG in the clusterProfiler package was used to perform functional enrichment analysis for macrophage subpopulations (Supplementary Fig. S6N)¹⁰³.

Trajectory interference for chondrocyte subpopulations. Monocle3 was used to infer the trajectory of chondrocytes in scRNA-seq data from the murine wound healing model (Supplementary Fig. S7I)^{110–113}. The top 20 PCA defined in scRNA-seq of chondrocytes were used. The reduce_dimension function was used to reduce the dimensions of chondrocytes (Parameters: umap.metric = “cosine”, umap.min_dist = 0.05, umap.n_neighbors = 20). cluster_cells function with parameters cluster_method = “Louvain”, resolution = 0.001, random_seed = 123 was used to cluster the cells. The learn_graph function with default parameters was used to learn the principal graph from the reduced dimension space using reversed graph embedding.

Ligand and receptor interaction analysis. To identify the significant and specific ligand-receptor pairs between *Igf1bp5*⁺ chondrocyte, *Fmod*⁺ fibroblast, and *Cd36*⁺ *Gpnmb*⁺ *Il1b*⁺ macrophage, the CellPhoneDB tool was used in scRNA-seq data from murine wound healing^{114,115}. Cell expression was normalized by dividing the total counts. The interaction pairs with *p*-value < 0.01 or the mean attraction strength > 0 were retained.

Statistics and reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. Mice were randomly assigned to experimental groups. The Investigators were not blinded to allocation during experiments and outcome assessment. Sample size and statistical tests are described in the figure legends. Data are expressed as mean ± SEM unless otherwise specified. Statistical tests were performed using R statistical software.

Reporting summary

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

Data availability

The raw data of single-cell RNA-seq and ST of human CRC generated in this study were deposited in the Genome Sequence Archive with accession ID [HRA007218](https://gsa.genomics.cn/HRA007218). The guidance for making a data access request to GSA for humans can be downloaded from https://ngdc.cncb.ac.cn/gsa-human/document/GSA-Human_Request_Guide_for_Users_us.pdf. The raw data of single-cell RNA-seq and ST of murine wound healing generated in this study were deposited in the Genome Sequence Archive with accession ID [CRA016052](https://gsa.genomics.cn/CRA016052). The public datasets used in this study include the scRNA-seq dataset and the 10x Visium spatial transcriptomic dataset of human CRC (<https://ngdc.cncb.ac.cn/gsa-human/browse/HRA000979>), the 10x Visium HD dataset of human CRC, and the mouse hippocampus Slide-seqV2 dataset (https://singlecell.broadinstitute.org/single_cell/study/SCP948/robust-decomposition-of-cell-type-mixtures-in-spatial-transcriptomics). The processed data and code used to reproduce the results in this study are available in the Zenodo repository at <https://doi.org/10.5281/zenodo.12634581>. The additional source data are available in the Zenodo repository at <https://doi.org/10.5281/zenodo.18482631>. Source data are provided with this paper.

Code availability

The UCASpatial is available in the GitHub repository <https://github.com/BIGHanLab/UCASpatial> under the GNU General Public License version 3 (GPL-3.0 license), and has been deposited in Zenodo with the identifier doi:10.5281/zenodo.18410641¹¹⁶. The data and code to reproduce the results in this study are available in the Zenodo repositories at .

References

1. Brbic, M. et al. Annotation of spatially resolved single-cell data with STELLAR. *Nat. Methods* **19**, 1411–1418 (2022).
2. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023).
3. Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
4. Junker, J. P. et al. Genome-wide RNA tomography in the zebrafish embryo. *Cell* **159**, 662–675 (2014).
5. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

6. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
7. Liao, J., Lu, X., Shao, X., Zhu, L. & Fan, X. Uncovering an organ's molecular architecture at single-cell resolution by spatially resolved transcriptomics. *Trends Biotechnol.* **39**, 43–58 (2021).
8. Rao, A., Barkley, D., Franca, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
9. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).
10. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).
11. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
12. Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat. Biotechnol.* **40**, 1349–1359 (2022).
13. Andersson, A. et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* **3**, 565 (2020).
14. Yang, J. et al. Spotiphy enables single-cell spatial whole transcriptomics across an entire section. *Nat. Methods* **22**, 724–736 (2025).
15. Li, B. et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670 (2022).
16. Li, H. et al. A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nat. Commun.* **14**, 1548 (2023).
17. Kuppe, C. et al. Spatial multi-omic map of human myocardial infarction. *Nature* **608**, 766–777 (2022).
18. Ackerman, J. E. et al. Defining the spatial-molecular map of fibrotic tendon healing and the drivers of Scleraxis-lineage cell fate and function. *Cell Rep.* **41**, 111706 (2022).
19. Anderson, A. C. et al. Spatial transcriptomics. *Cancer Cell* **40**, 895–900 (2022).
20. Foster, D. S. et al. Multiomic analysis reveals conservation of cancer-associated fibroblast phenotypes across species and tissue of origin. *Cancer Cell* **40**, 1392–1406 e1397 (2022).
21. Heiser, C. N. et al. Molecular cartography uncovers evolutionary and microenvironmental dynamics in sporadic colorectal tumors. *Cell* **186**, 5620–5637 e5616 (2023).
22. Matusiak, M. et al. Spatially segregated macrophage populations predict distinct outcomes in colon cancer. *Cancer Discov.* **14**, 1418–1439 (2024).
23. Kersten, K. et al. Spatiotemporal co-dependency between macrophages and exhausted CD8(+) T cells in cancer. *Cancer Cell* **40**, 624–638 e629 (2022).
24. Kirschenbaum, D. et al. Time-resolved single-cell transcriptomics defines immune trajectories in glioblastoma. *Cell* **187**, 149–165 e123 (2024).
25. Benotmane, J. K. et al. High-sensitive spatially resolved T cell receptor sequencing with SPTCR-seq. *Nat. Commun.* **14**, 7432 (2023).
26. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8(+) T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).
27. Xun, Z. et al. Reconstruction of the tumor spatial microenvironment along the malignant-boundary-nonmalignant axis. *Nat. Commun.* **14**, 933 (2023).
28. Zhang, D. et al. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nat. Biotechnol.* **42**, 1372–1377 (2024).
29. Sun, D., Liu, Z., Li, T., Wu, Q. & Wang, C. STRIDE: accurately decomposing and integrating spatial transcriptomics using single-cell RNA sequencing. *Nucleic Acids Res.* **50**, e42 (2022).
30. Fu, T. et al. Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response. *J. Hematol. Oncol.* **14**, 98 (2021).
31. Oliveira, M. F. et al. High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer. *Nat. Genet.* **57**, 1512–1523 (2025).
32. Seferbekova, Z., Lomakin, A., Yates, L. R. & Gerstung, M. Spatial biology of cancer evolution. *Nat. Rev. Genet.* **24**, 295–313 (2023).
33. Erickson, A. et al. Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* **608**, 360–367 (2022).
34. Lomakin, A. et al. Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature* **611**, 594–602 (2022).
35. Qi, J. et al. Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPPI(+) macrophages in colorectal cancer. *Nat. Commun.* **13**, 1742 (2022).
36. Trimm, E. & Red-Horse, K. Vascular endothelial cell development and diversity. *Nat. Rev. Cardiol.* **20**, 197–210 (2023).
37. Li, G., Gao, J., Ding, P. & Gao, Y. The role of endothelial cell-pericyte interactions in vascularization and diseases. *J. Adv. Res.* **67**, 269–288 (2024).
38. Fridman, W. H. et al. B cells and tertiary lymphoid structures as determinants of tumour immune contexture and clinical outcome. *Nat. Rev. Clin. Oncol.* **19**, 441–457 (2022).
39. Munoz-Erazo, L., Rhodes, J. L., Marion, V. C. & Kemp, R. A. Tertiary lymphoid structures in cancer—considerations for patient prognosis. *Cell Mol. Immunol.* **17**, 570–575 (2020).
40. Danenberg, E. et al. Breast tumor microenvironment structures are associated with genomic features and clinical outcome. *Nat. Genet.* **54**, 660–669 (2022).
41. Dhainaut, M. et al. Spatial CRISPR genomics identifies regulators of the tumor microenvironment. *Cell* **185**, 1223–1239 e1220 (2022).
42. Ramirez, C. F. A. et al. Cancer cell genetics shaping of the tumor microenvironment reveals myeloid cell-centric exploitable vulnerabilities in hepatocellular carcinoma. *Nat. Commun.* **15**, 2581 (2024).
43. Chen, X., Agustinus, A. S., Li, J., DiBona, M. & Bakhoum, S. F. Chromosomal instability as a driver of cancer progression. *Nat. Rev. Genet.* **26**, 31–46 (2025).
44. Honda, K. et al. IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772–777 (2005).
45. Liu, S. Y., Sanchez, D. J., Aliyari, R., Lu, S. & Cheng, G. Systematic identification of type I and type II interferon-induced antiviral factors. *Proc. Natl. Acad. Sci. USA* **109**, 4239–4244 (2012).
46. Chiappinelli, K. B. et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986 (2015).
47. He, J. et al. Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nat. Commun.* **12**, 1456 (2021).
48. Kong, Y. et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat. Commun.* **10**, 5228 (2019).
49. Sexton, C. E., Tillett, R. L. & Han, M. V. The essential but enigmatic regulatory role of HERVH in pluripotency. *Trends Genet.* **38**, 12–21 (2022).
50. Tumeh, P. C. et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
51. Samstein, R. M. et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
52. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).

53. Eming, S. A., Martin, P. & Tomic-Canic, M. Wound repair and regeneration: mechanisms, signaling, and translation. *Sci. Transl. Med.* **6**, 265sr266 (2014).
54. Hu, K. H. et al. Transcriptional space-time mapping identifies concerted immune and stromal cell patterns and gene programs in wound healing and cancer. *Cell Stem Cell* **30**, 885–903 e810 (2023).
55. Mascharak, S. et al. Multi-omic analysis reveals divergent molecular events in scarring and regenerative wound healing. *Cell Stem Cell* **29**, 315–327 e316 (2022).
56. McBrearty, B. A., Clark, L. D., Zhang, X. M., Blankenhorn, E. P. & Heber-Katz, E. Genetic analysis of a mammalian wound-healing trait. *Proc. Natl. Acad. Sci. USA* **95**, 11792–11797 (1998).
57. Clark, L. D., Clark, R. K. & Heber-Katz, E. A new murine model for mammalian wound repair and regeneration. *Clin. Immunol. Immunopathol.* **88**, 35–45 (1998).
58. Heydemann, A. The super super-healing MRL mouse strain. *Front. Biol.* **7**, 522–538 (2012).
59. Wynn, T. A. & Vannella, K. M. Macrophages in tissue repair, regeneration, and fibrosis. *Immunity* **44**, 450–462 (2016).
60. Kim, M. H. et al. Dynamics of neutrophil infiltration during cutaneous wound healing and infection using fluorescence imaging. *J. Invest. Dermatol.* **128**, 1812–1820 (2008).
61. Canedo-Dorantes, L. & Canedo-Ayala, M. Skin acute wound healing: a comprehensive review. *Int. J. Inflam.* **2019**, 3706315 (2019).
62. Phillipson, M. & Kubes, P. The healing power of neutrophils. *Trends Immunol.* **40**, 635–647 (2019).
63. Raziyeve, K. et al. Immunology of acute and chronic wound healing. *Biomolecules* **11**, 700 (2021).
64. Hill, D. A. et al. Distinct macrophage populations direct inflammatory versus physiological changes in adipose tissue. *Proc. Natl. Acad. Sci. USA* **115**, E5096–E5105 (2018).
65. Jaitin, D. A. et al. Lipid-associated macrophages control metabolic homeostasis in a Trem2-dependent manner. *Cell* **178**, 686–698 e614 (2019).
66. Henlon, Y. et al. Single-cell analysis identifies distinct macrophage phenotypes associated with prodisease and proresolving functions in the endometriotic niche. *Proc. Natl. Acad. Sci. USA* **121**, e2405474121 (2024).
67. Franzen, L. et al. Mapping spatially resolved transcriptomes in human and mouse pulmonary fibrosis. *Nat. Genet.* **56**, 1725–1736 (2024).
68. Abedini, A. et al. Single-cell multi-omic and spatial profiling of human kidneys implicates the fibrotic microenvironment in kidney disease progression. *Nat. Genet.* **56**, 1712–1724 (2024).
69. Buechler, M. B. et al. Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
70. Konieczny, P. et al. Interleukin-17 governs hypoxic adaptation of injured epithelium. *Science* **377**, eabg9302 (2022).
71. Gawriluk, T. R. et al. Comparative analysis of ear-hole closure identifies epimorphic regeneration as a discrete trait in mammals. *Nat. Commun.* **7**, 11164 (2016).
72. Pastar, I. et al. Epithelialization in wound healing: a comprehensive review. *Adv. Wound Care* **3**, 445–464 (2014).
73. Rousselle, P., Braye, F. & Dayan, G. Re-epithelialization of adult skin wounds: cellular mechanisms and therapeutic strategies. *Adv. Drug Deliv. Rev.* **146**, 344–365 (2019).
74. Pena, O. A. & Martin, P. Cellular and molecular mechanisms of skin wound healing. *Nat. Rev. Mol. Cell Biol.* **25**, 599–616 (2024).
75. Qu, Y. et al. A comprehensive analysis of single-cell RNA transcriptome reveals unique SP1+ chondrocytes in human osteoarthritis. *Comput. Biol. Med.* **160**, 106926 (2023).
76. Chau, M. et al. The synovial microenvironment suppresses chondrocyte hypertrophy and promotes articular chondrocyte differentiation. *NPJ Regen. Med.* **7**, 51 (2022).
77. Ignatyeva, N., Gavrilov, N., Timashev, P. S. & Medvedeva, E. V. Prg4-expressing chondroprogenitor cells in the superficial zone of articular cartilage. *Int. J. Mol. Sci.* **25**, 5605 (2024).
78. Massengale, M. et al. Adult Prg4+ progenitors repair long-term articular cartilage wounds in vivo. *JCI Insight* **8**, e167858 (2023).
79. Diekmann, B. O. et al. Cartilage tissue engineering using differentiated and purified induced pluripotent stem cells. *Proc. Natl. Acad. Sci. USA* **109**, 19172–19177 (2012).
80. Ono, N., Ono, W., Nagasawa, T. & Kronenberg, H. M. A subset of chondrogenic cells provides early mesenchymal progenitors in growing bones. *Nat. Cell Biol.* **16**, 1157–1167 (2014).
81. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J. & Werb, Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* **11**, 5120 (2020).
82. Burgess, H. A. et al. PPARgamma agonists inhibit TGF-beta induced pulmonary myofibroblast differentiation and collagen production: implications for therapy of lung fibrosis. *Am. J. Physiol. Lung Cell Mol. Physiol.* **288**, L1146–L1153 (2005).
83. Hua, Q. et al. PPARgamma mediates the anti-pulmonary fibrosis effect of icaritin. *Toxicol. Lett.* **350**, 81–90 (2021).
84. Kokeny, G., Calvier, L. & Hansmann, G. PPARgamma and TGFbeta-major regulators of metabolism, inflammation, and fibrosis in the lungs and kidneys. *Int. J. Mol. Sci.* **22**, 10431 (2021).
85. Zhou, X. et al. Circuit design features of a stable two-cell system. *Cell* **172**, 744–757 e717 (2018).
86. Takawale, A. et al. Tissue inhibitor of matrix metalloproteinase-1 promotes myocardial fibrosis by mediating CD63-integrin beta1 interaction. *Hypertension* **69**, 1092–1103 (2017).
87. Dong, J. & Ma, Q. TIMP1 promotes multi-walled carbon nanotube-induced lung fibrosis by stimulating fibroblast activation and proliferation. *Nanotoxicology* **11**, 41–51 (2017).
88. Zhang, Y. et al. An NFAT1-C3a-C3aR positive feedback loop in tumor-associated macrophages promotes a glioma stem cell malignant phenotype. *Cancer Immunol. Res.* **12**, 363–376 (2024).
89. Sehgal, A., Irvine, K. M. & Hume, D. A. Functions of macrophage colony-stimulating factor (CSF1) in development, homeostasis, and tissue repair. *Semin. Immunol.* **54**, 101509 (2021).
90. Cook, S. A. Understanding interleukin 11 as a disease gene and therapeutic target. *Biochem. J.* **480**, 1987–2008 (2023).
91. Sedger, L. M. & McDermott, M. F. TNF and TNF-receptors: from mediators of cell death and inflammation to therapeutic giants - past, present and future. *Cytokine Growth Factor Rev.* **25**, 453–472 (2014).
92. Cook, S. A. & Schafer, S. Hiding in plain sight: interleukin-11 emerges as a master regulator of fibrosis, tissue integrity, and stromal inflammation. *Annu. Rev. Med.* **71**, 263–276 (2020).
93. Widjaja, A. A. et al. Redefining IL11 as a regeneration-limiting hepatotoxin and therapeutic target in acetaminophen-induced liver injury. *Sci. Transl. Med.* **13**, eaba8146 (2021).
94. Widjaja, A. A. et al. Targeting endogenous kidney regeneration using anti-IL11 therapy in acute and chronic models of kidney disease. *Nat. Commun.* **13**, 7497 (2022).
95. Zhang, B., Yao, K., Zhou, E., Zhang, L. & Cheng, C. Chr20q amplification defines a distinct molecular subtype of microsatellite stable colorectal cancer. *Cancer Res.* **81**, 1977–1987 (2021).
96. Dai, M., Pei, X. & Wang, X. J. Accurate and fast cell marker gene identification with COSG. *Brief Bioinform.* **23**, bbab579 (2022).
97. Kotliar, D. et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife* **8**, e43803 (2019).
98. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 e1016 (2018).

99. 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).
100. Lee, C. W. et al. Multiplex immunofluorescence staining and image analysis assay for diffuse large B cell lymphoma. *J. Immunol. Methods* **478**, 112714 (2020).
101. Bankhead, P. et al. QuPath: open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
102. Satoh, J. & Tabunoki, H. A comprehensive profile of ChIP-Seq-based STAT1 target genes suggests the complexity of STAT1-mediated gene regulatory mechanisms. *Gene Regul. Syst. Bio.* **7**, 41–56 (2013).
103. 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).
104. Mayer, J., Blomberg, J. & Seal, R. L. A revised nomenclature for transcribed human endogenous retroviral loci. *Mob. DNA* **2**, 7 (2011).
105. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
106. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* **14**, 7 (2013).
107. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
108. Reich, M. et al. GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).
109. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
110. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
111. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
112. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
113. McInnes, L., Healy, J. & Saul, N. Lukas & Großberger UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
114. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
115. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
116. Xu, Y. et al. Ultra-precision deconvolution of spatial transcriptomics decodes immune heterogeneity and fate-defining programs in tissues. *GitHub, UCASpatial* <https://doi.org/10.5281/zenodo.18410641> (2026).

Acknowledgements

We thank Prof. Youqiong Ye (Shanghai Jiao Tong University School of Medicine) for kindly sharing scRNA-seq and ST data of human CRC. This work was funded by the National Key R&D Program of China (2024YFC3405901), the Strategic Priority Research Program of the Chinese Academy of Science (XDB0570101), the National Natural Science Foundation of China (NSFC) T2495272, 32121001, 22293052, the Beijing Natural Science Foundation (L244023, L259070, L259072), the

National Key R&D Program of China (2024YFA1802102), the CAS Youth Interdisciplinary Team, CNCB-initiative program (iCNCB2025001), the National Funded Postdoctoral Researcher program (GZC20241664), the Next-Generation Bioinformatics Algorithms (XDA0460302), and the Natural Science Foundation of Inner Mongolia (2024MS08010).

Author contributions

D.H. and M.M.X. conceived the study and provided overall supervision of the study. D.H. and Yin Xu designed and developed UCASpatial. Yin Xu conducted benchmark analyses with assistance from F.Z R.C. conducted library construction for murine scRNA-seq. M.G. and R.C. conducted library construction for murine ST. Y.Z. performed bioinformatics analyses for the murine dataset. M.G. and Yu Xia performed IHC staining of murine ears. Z.W., L.S., and J.L. provided human specimens. J.Y. performed IHC staining of human samples and organized and summarized the patient information. P.G. and L.D. conducted library construction for human scRNA-seq and ST. G.L. preprocessed the human scRNA-seq data. Z.H. and Yin Xu performed bioinformatics analyses for the human CRC dataset. Z.H. performed bioinformatics analyses of TEs. D.H., M.M.X., Yin Xu, Y.Z., and Z.H. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-70645-3>.

Correspondence and requests for materials should be addressed to Meng Michelle Xu or Dali Han.

Peer review information *Nature Communications* thanks Ning Xu Landen, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

¹RNA Center for Cross-disciplinary Omics and Intelligent Decoding, China National Center for Bioinformation, Beijing, China. ²State Key Laboratory of RNA Innovation, Science and Engineering, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China. ³College of Future Technology, Sino-Danish College, University of Chinese Academy of Sciences, Beijing, China. ⁴Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Beijing, China. ⁵Department of Digestive Oncology, Peking University Cancer Hospital (Inner Mongolia Campus), Affiliated Cancer Hospital of Inner Mongolia Medical University, Hohhot, China. ⁶State Key Laboratory of Molecular Oncology, School of Basic Medical Sciences, Institute for Immunology, Tsinghua University, Beijing, China. ⁷These authors contributed equally: Yin Xu, Zurui Huang, Yawei Zhang, Minghui Gong, Zhenghang Wang. ✉ e-mail: michellexu@mail.tsinghua.edu.cn; handl@big.ac.cn