Ultra-resolution Deconvolution of Spatial Transcriptomics Unveils Spatiotemporal Cellular Dynamics in Complex Microenvironments

Authors: Yin Xu[1,2]†, Zurui Huang[1,2]†, Yawei Zhang[1,2]†, Zhenghang Wang[3]†, Peijin Guo[4], Feifan Zhang[1,2], Wenxuan Gong[1,2], Guanghao Liang[1,2], Boyuan Mei[4], Lihui Dong[1,2], Renbao Chang[1,2], Minghui Gong[1,2], Yu Xia[4], Haochen Ni[1,2], Jing Yang[1,2], Yuan Gao[1,2], Zhaoqi Liu[1,2], Lin Shen[3], Jian Li[3], Meng Michelle Xu[4]*, and Dali Han[1,2]*

Affiliations:

[1] Department of Computational Biology, Key Laboratory of RNA Science and Engineering, China National Center for Bioinformation, Beijing Institute of Genomics, Chinese Academy of Sciences; Beijing 100101, China.

[2]College of Future Technology, Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100049, China.

[3]Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Beijing 100142, China.

[4]State Key Laboratory of Molecular Oncology, School of Basic Medical Sciences, Institute for Immunology, Tsinghua University; Beijing 100871, China.

† These authors contributed equally.

*Corresponding author. Email: michellexu@mail.tsinghua.edu.cn (M.M.X.);
handl@big.ac.cn (D. H.)

Abstract:

The current challenge in analyzing sequencing-based spatial transcriptomics (ST) is to precisely resolve the spatial distribution of cell subpopulations with low abundance and detect context-dependent variation in cell states within complex tissue microenvironments. Here, we introduce an ultra-resolution ST deconvolution algorithm (UCASpatial) that improves the resolution of mapping cell subpopulations to spatial locations by leveraging the contribution of genes indicative of cell identity through entropy-based weighting. Using both in silico and real ST datasets, we demonstrate that UCASpatial improves the robustness and accuracy in identifying

low-abundant cell subpopulations and distinguishing transcriptionally heterogeneous cell subpopulations. Applying UCASpatial to murine wound healing, we recapitulate known spatiotemporal dynamics of multiple cell subpopulations involved in wound healing and reveal the spatial segregation of distinct macrophage subpopulations embedded within different cellular communities at the wound center. In human colorectal cancer (CRC), we link genomic alterations in individual cancer clones to multi-cellular characteristics of the tumor immune microenvironment (TIME) and reveal the co-evolution of tumor cells and TIME within the same tumor. We show that the copy number gain on chromosome 20q (chr20q-gain) in tumor cells orchestrates a T cell-excluded TIME, indicative of resistance to immunotherapy in CRC, and is associated with tumor-intrinsic endogenous retrovirus silencing and impaired type I interferon response. Our findings present UCASpatial as a versatile tool for deciphering ultra-resolution cellular landscapes in ST and exploring intercellular mechanisms in complex and dynamic microenvironments.

Introduction

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 technologies offer promising avenues for achieving this; however, current spot-based techniques face a significant limitation in lacking single-cell resolution. Methods like Visium[3], Tomo-Seq[4], Slide-seq[5], and HDST[6] 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), such as SPOTlight[9], RCTD[10], cell2location[11], CARD[12], and stereoscope[13]. 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[11,12,14-16].

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 (TME)[19-23]. However, several immune cells, particularly those infiltrating into tumors, are known for their sparse presence and low transcriptional activity[24-27]. Their discontinuous spatial distributions in these microenvironments limit the effectiveness of algorithms relying on spot-neighbor information[27,28]. Additionally, the low abundance and the context-dependent phenotypic variation of heterogeneous immune cells, particularly T cells, pose challenges for conventional deconvolution algorithms[24]. Furthermore, their scarcity makes them difficult to capture accurately using histological images, thus constraining the applicability of image-based algorithms[29,30]. Consequently, the precise characterization of complex and dynamic microenvironments remains a significant challenge.

To address the aforementioned challenge, we developed UCASpatial, an ultra-resolution cellular deconvolution algorithm in spatial transcriptomics 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 for indicating 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 precise and unbiased deconvolution of the cellular landscape. Through simulations and applications to murine wound healing and human CRC ST datasets, we illustrate UCASpatial's improvements in deconvolution and demonstrate its utility in revealing spatiotemporal transitions and multicellular communities among cell subpopulations.

Result

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. 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 (Methods). Third, UCASpatial employs Weighted Non-Negative Least Square (WNNLS) to deconvolute the mRNA counts of each spot in the ST data using the

weighted cGEPs, thereby estimating the cellular composition. 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 TME. 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. 1a-c, and Methods). The cellular composition of simulated spots reflects the actual cellular composition identified in typical tumor architectures within the TME[16,31] (Supplementary Fig. 1d, and Supplementary Table. 1).

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 cell-type 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. 1d). We compared UCASpatial with several recently established methods, including SPOTlight, RCTD, cell2location, CARD, and stereoscope, using Root Mean Square Error (RMSE) to first quantify the deviation between estimations and ground truth at the spot level. UCASpatial exhibited superior performance across all complexity levels (median RMSE = 0.051) compared to the aforementioned methods, with 10%, 21%, 31%, 36%, and 35% improvement in terms of RMSE compared to RCTD (0.058), SPOTlight (0.066), CARD (0.075), cell2location (0.081), and stereoscope (0.080), respectively (Fig. 2b). Additionally, UCASpatial demonstrated a comparable Pearson Correlation Coefficient (PCC) to cell2location and outperformed other methods, when comparing estimations with the ground truth (Supplementary Fig. 2b). Moreover, we also assessed the ability for identifying the existence of cell subpopulations within spots using the F1 score. UCASpatial consistently outperformed other methods across all complexity levels with the highest F1 score (with a 6-25% improvement) (Fig. 2b, and Supplementary Fig. 2c).

We next evaluated the precision and robustness of UCASpatial in deconvoluting different cell subpopulations. Examination of PCC between estimations and ground truth for each method 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

132 subpopulations (with a 39-187% improvement) (Fig. 2d). By categorizing the PCC values of
133 each cell subpopulation into four groups, we noticed UCASpatial exhibited PCC values
134 surpassing the medium-high correlation level (PCC > 0.5) for 19 out of 20 cell subpopulations,
135 with the sole exception $CD4^+$ Tfh (PCC = 0.47) while remained the best compared to alternative
136 methods (Fig. 2c, e). In contrast, other methods yielded PCC values below the medium-high
137 correlation level for at least 30% of cell subpopulations, with some even falling below 0.25 (Fig.
138 2e), underscoring UCASpatial's capability in discriminating cell subpopulations, even with
139 subtle transcriptomic differences (Supplementary Fig. 3a, b). Furthermore, assessments using
140 the PCC and F1 score reaffirmed UCASpatial's precise and robust decomposition across each
141 cell subpopulation (Supplementary Fig. 3c, d).

142 Considering the challenge of ST analysis in precisely mapping the spatial location of low-
143 abundant cell subpopulations, we next investigated UCASpatial's sensitivity in annotating cell
144 subpopulations with low frequencies within spots. We categorized the proportion of each cell
145 subpopulation within simulated spots into ten groups, forming a gradient from 0-10% to 0-100%
146 (Fig. 2f). Subsequently, we calculated the RMSE across various proportions to evaluate the
147 deviation of cell-subpopulation proportions. We observed that UCASpatial exhibited the lowest
148 overall RMSE (averaged by each cell subpopulation, medium RMSE = 0.051) across different
149 proportions with a 7-31% improvement over the other methods (Fig. 2g). Moreover, we
150 specifically examined the RMSE of different cell subpopulations, revealing that UCASpatial
151 notably outperformed other methods for cell subpopulations with subtle transcriptomic
152 differences such as $CD8^+$ Tex and $CD8^+$ Tem (with 25-63% and 13-26% improvements,
153 respectively) (Fig. 2g and Supplementary Fig. 4). Consistent results were observed in
154 assessments using the PCC and F1 score (Supplementary Figs. 5-6), highlighting UCASpatial's
155 superiority in accurately annotating cell subpopulations with low frequencies within spots, even
156 with subtle transcriptomic differences.

157 Furthermore, we illustrated how UCASpatial's entropy-based weighting and meta-purity
158 filtering designs enhance the performance of cellular deconvolution (Supplementary Fig. 7).
159 Specifically, the entropy-based weighting design showed improvements of 22%, 50%, and 17%
160 in terms of RMSE, PCC, and F1 score, respectively, compared to UCASpatial without
161 weighting. The meta-purity filtering design exhibited improvements across different parameters
162 with 4-11%, 1-8%, and 1-3% in terms of RMSE, PCC, and F1 score. In addition, the same
163 analyses as before showed that UCASpatial, without entropy-weighting and meta-purity
164 filtering, lost the previously mentioned improvements (Supplementary Figs. 2-6). Collectively,
165 our in-silico benchmarking demonstrates that UCASpatial enables precise and sensitive spatial

deconvolution, excelling in discerning cell subpopulations with subtle transcriptomic differences, even when they exist in low proportions.

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

To demonstrate UCASpatial's utility, we applied it to real ST datasets. We primarily investigated murine wound healing, a well-organized multicellular process characterized by high complexity and spatiotemporal dynamics[32-34]. Specifically, we employed a 2 mm ear punch wound model in MRL/MpJ mice (hereafter denoted as MRL), a strain with super-healing capacity that completely regenerates the ear tissue post-wounding[35-37]. Tissue samples were collected from MRL mice at various stages of wound healing: unwounded (day 0), inflammatory stage (day 3), return-to-unwounded stage (day 7), and regeneration stage (day 15)[33] (Fig. 3a). The ST dataset was generated using the 10x Genomics Visium platform. For reference, we generated scRNA-seq data from day 7 ear tissue, annotated with 20 cell subpopulations (Fig. 3b, c, and Supplementary Fig. 8a). We then utilized UCASpatial for the deconvolution of these 20 cell subpopulations within the murine wound healing ST dataset (Supplementary Fig. 8b).

To delineate the spatial distribution of cell subpopulations in different regions, we divided the sections into three subdomains: the wound center (2 mm center diameter), wound proximal (0-1 mm from the wound edge), and wound distal (1-2 mm from the wound edge) (Supplementary Fig. 8c). We firstly examined the infiltrating cell subpopulations within the wound center. We noticed the structural cell types, including keratinocytes, endothelial cells, and fibroblasts, revealing alignment with their respective anatomical positions (Fig. 3c, and Supplementary Figs. 8b and 10). Notably, keratinocytes demonstrated enrichment in regions of regenerative epithelium from the return-to-unwounded stage, signifying successful re-epithelialization[38] (Fig. 3c, and Supplementary Figs. 8b and 10). Additionally, endothelial cells were observed proximal to blood vessels, while 4 fibroblast subpopulations were enriched at the wound center during the return-to-unwounded and regeneration stages (Fig. 3c, and Supplementary Figs. 8b and 10). For example, *Lrrc15*$^{+}$ fibroblasts, which exhibited the characteristics of mature myofibroblasts, were accumulated at the wound center at the return-to-unwounded stage. Notably, *Lrrc15*$^{+}$ fibroblasts comprised over 33% of all cell subpopulations at that location, consistent with their function in promoting wound contraction[39] (Fig. 3b, and Supplementary Figs. 8a, d, and 10). The above observations recapitulate the spatial distribution and dynamics of structural cells during tissue repair[33,38,39].

200 We also investigated whether UCASpatial could capture the spatial dynamics of wound-
201 associated immune cells over time. We observed the enrichment of myeloid cells, including
202 neutrophils and macrophages, at the wound center during the inflammatory stage, in line with
203 previous studies demonstrating their pivotal role in orchestrating the early immune response to
204 tissue injury[40,41] (Supplementary Figs. 8b, d, and 10). Interestingly, while neutrophils decreased
205 during the return-to-unwounded stage, macrophages persisted in high abundance within the
206 repaired tissues, consistent with previous findings[33,40,42] (Supplementary Figs. 8d and 10).
207 Similar to neutrophils, basophils were enriched at the wound center in the inflammatory stage
208 and decreased during the return-to-unwounded stage, indicating that they are the first
209 responders to injury[43] (Supplementary Figs. 8d and 10).
210 Subsequently, we conducted a comparative analysis of UCASpatial alongside other
211 methods in murine wound healing. We calculated the PCC between the spatial distribution of
212 cell-subpopulation proportion and marker-gene score (Supplementary Table. 2 and Methods),
213 which was used to quantify the performance of deconvolution without any ground truth[15].
214 UCASpatial, along with cell2location, SPOTlight, and RCTD, demonstrated high PCC between
215 the spatial distribution of cell-subpopulation proportions and marker gene expression, while
216 UCASpatial still showed an improvement of 5%-20% in comparison (Supplementary Fig. 8e,
217 f). While all considered methods performed well at annotating structural cell subpopulations,
218 some methods failed to identify several immune cell subpopulations. For example, RCTD and
219 stereoscope overlooked the spatiotemporal dynamics of neutrophils, which was supposed to be
220 enriched at the inflammatory stage and declined at the return-to-unwounded stage
221 (Supplementary Figs. 8g and 10-15). Additionally, RCTD and Cell2location exhibited limited
222 capability in detecting macrophages on day 0, with proportions less than or equal to 1%
223 (Supplementary Figs. 8h and 10-15). We also noticed that UCASpatial without entropy
224 weighting and meta-purity filtering exhibited lower PCC compared with UCASpatial
225 (Supplementary Figs. 8e and 16). These results demonstrate the superior performance of
226 UCASpatial in delineating the spatiotemporal dynamics of cell subpopulations.
227

228 **UCASpatial reconstructs spatiotemporal dynamics of macrophage subpopulations during**
229 **murine wound healing**

230 Recent studies suggest that macrophages infiltrate healing tissues and respond to various
231 microenvironmental signals, leading to their differentiation into specialized subpopulations
232 with distinct functions and locations[33]. To examine whether UCASpatial could detect the
233 spatiotemporal dynamics of macrophage subpopulations including *Cd163*$^{+}$ macrophages, *Spp1*$^{+}$

macrophages, *MHCII*$^{+}$ macrophages (encoded by *H2-Aa* and *H2-Ab1*), and *Il1b*$^{+}$ macrophages, we focused on the wound center at the return-to-unwounded stage (day 7) for further analysis (Supplementary Fig. 9a, b). We subdivided the wound center into two layers: the epidermis layer and the dermis layer, based on hematoxylin and eosin (H&E) staining (Fig. 3c).

All four macrophage subpopulations annotated by UCASpatial exhibited distinct spatial distributions at the wound center, which were consistent with previous reports[33,44] (Fig. 3d). Specifically, *Cd163*$^{+}$ and *Il1b*$^{+}$ macrophages were predominantly found around the epidermis layer, whereas *Spp1*$^{+}$ and *MHCII*$^{+}$ macrophages were confined to the dermis layer, aligning with their marker-expression patterns (Supplementary Fig. 9c). Notably, while *Cd163*$^{+}$ macrophages were evenly distributed across the entire wound center, *Il1b*$^{+}$ macrophages exhibited spatial exclusivity, preferentially clustering at the migrating epithelial tongue (Fig. 3d). Functional enrichment analysis indicated that *Cd163*$^{+}$ macrophages may play roles in mesenchymal cell differentiation, endocytosis, and efferocytosis, while *Il1b*$^{+}$ macrophages were enriched for functions related to leukocyte activation and epithelial cell differentiation (Supplementary Fig. 9d). Intriguingly, multicellular community analysis also confirmed that *Il1b*$^{+}$ macrophages and keratinocytes co-occurred in the same community, indicating the function of *Il1b*$^{+}$ macrophages in re-epithelialization[44] (Supplementary Fig. 9e).

Although all considered methods displayed exceptional proficiency in elucidating *Cd163*$^{+}$ macrophages, UCASpatial achieved higher accuracy for annotating other macrophage subpopulations. Particularly, CARD, cell2location, and RCTD failed to identify *Il1b*$^{+}$ macrophages in the migrating epithelial tongue, while SPOTlight inaccurately annotated the unique spatial distribution patterns of *Spp1*$^{+}$ macrophages in the dermis (Fig. 3d).

Subsequently, we examined the spatiotemporal dynamics of macrophages throughout the wound-healing process. We observed that *Il1b*$^{+}$ macrophages appeared in the wound center at the inflammatory stage but gradually decreased in abundance at the return-to-unwounded stage (Fig. 3e). In contrast, the other three macrophage subpopulations were predominantly located on the wound proximal and distal, and had distinct temporal dynamics. While *MHCII*$^{+}$ macrophages were present at all measured time points, *Cd163*$^{+}$ macrophages were detected at the inflammatory stage and were subsequently replaced by *Spp1*$^{+}$ macrophages, which peaked in abundance at the return-to-unwounded stage and declined at the regeneration stage. These spatial segregations and temporal dynamics of macrophage subpopulations annotated by UCASpatial indicate their role in coordinating tissue repair towards an unwounded state[40] (Fig. 3e). Collectively, UCASpatial reconstructs an ultra-resolution spatiotemporal cellular

landscape in murine wound healing and reveals the unique spatial segregation and temporal dynamics of macrophage subpopulations that served in different steps of wound healing.

UCASpatial accurately annotates cell subpopulations in the human CRC ST dataset

Spatial omics has provided insights into the intricate interplay between tumor alterations and the co-evolved immune microenvironment surrounding tumors[21,45,46]. However, due to the low abundance and transcriptional output of tumor-infiltrating immune cells[24,26,27], characterization of immune cell compositions in the vicinity of tumor clones remains a challenge[21,47]. To investigate whether UCASpatial can precisely resolve the tumor immune microenvironment, we curated an ST dataset from eight tumor sections of seven patients with human CRC using the 10x Genomics Visium platform[48] (Fig. 4a, Supplementary Fig. 17a, and Supplementary Table. 3). We leveraged an integrated scRNA-seq dataset encompassing 34 cell subpopulations as a reference[48] (Fig. 4b, Supplementary Fig. 17b, c, and Methods).

We initially annotated the spatial distribution of 34 cell subpopulations in the human CRC ST dataset using UCASpatial (Supplementary Figs. 18-25). Notably, stromal cell subpopulations showed consistent distribution with anatomical locations as expected from prior literature, such as smooth muscle cells of the muscular layer and endothelial cells around blood vessels, along with pericytes[49,50]. Similarly, epithelial cell subpopulations displayed enrichment in the epithelial region, consistent with corresponding H&E staining and pathological annotations (Supplementary Fig. 17a). Upon further examination of immune subpopulations, we noticed B cell and various T cell and dendritic cell (DC) subpopulations, including $CD4^+$ Tn, $CD4^+$ Tfh, $CD8^+$ Tex, $CD8^+$ Tem, cDC2, and cDC3, formed a multicellular community resembling a tertiary lymphoid structure (TLS)[51,52] (Supplementary Figs. 18-25, see Supplementary Table. 4 for abbreviation information). Additionally, we noticed the spatial exclusivity of *C1QC*$^+$ macrophages and *PDPN*$^+$ macrophages, indicating their distinct functions in TME (Supplementary Figs. 18-25).

For comparative analysis, we also annotated this ST dataset using other methods (Supplementary Figs. 26-65). Remarkably, UCASpatial achieved the highest average PCC with an 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. 74a, Supplementary Table. 5, and Methods). In alignment with in-silico benchmarking results, all evaluated methods correctly annotated structural cell subpopulations, for example, *FAP*$^+$ myofibroblast, yet UCASpatial showed a higher sensitivity in annotating immune cell subpopulations (Fig. 4c and Supplementary Fig. 74a, b). For instance, UCASpatial

301 showed superior performance over other methods in identifying the spatial distribution of T cell
302 subpopulations in almost all five TLS regions (Supplementary Fig. 74c). While UCASpatial
303 depicted 4-6 T cell subpopulations consistent with their marker gene expressions
304 (Supplementary Fig. 74c, d), all compared methods only identified 0-2 T cell subpopulations.
305 Moreover, only UCASpatial captured distinct spatial distribution patterns of three $CD8^+$ T cell
306 subpopulations infiltrating tumors: $CD8^+$ Tex and $CD8^+$ Trm were proximal to the tumor,
307 whereas $CD8^+$ Tem was distal (Fig. 4c). Additionally, we noticed UCASpatial without entropy-
308 weighting and meta-purity filtering were unable to either identify T cell subpopulations in TLS
309 regions or discriminate the spatial distribution patterns of three $CD8^+$ T cell subpopulations,
310 exhibiting a comparable overall PCC with cell2location and SPOTlight (Supplementary Figs.
311 66-74). Collectively, these results demonstrate that UCASpatial can accurately resolve cell
312 subpopulations in the complex tumor immune microenvironment.
313
314 **UCASpatial spatially links the tumor genomics with diverse TIME at a clonal resolution.**
315 Tumor-intrinsic genomic alterations have collateral effects that modify the spatial organization
316 of the immune microenvironment[53-55]. To achieve the spatial linking of tumor genomics with
317 diverse immune microenvironments at a detailed clonal resolution, we generated an integrated
318 immunogenomics platform termed 'UCASpatial-clonalTIME'. This platform enables the
319 identification of tumor clones within tissue sections by inferring spatial copy number variations
320 (CNVs)[46] and concurrently estimating the composition of the TIME for each clone (clonalTIME)
321 using UCASpatial (Fig. 4c, Supplementary Fig. 75a, and Methods).
322 By assessing the proportions of $CD4^+$ and $CD8^+$ T cell subpopulations, we categorized the
323 clonalTIME into two phenotypes: T cell-rich and T cell-excluded (Supplementary Fig. 75b, c,
324 and Methods). We detected three out of thirteen tumor clones displaying the T cell-rich
325 clonalTIME phenotype (Fig. 4d, and Supplementary Fig. 75c, d). Within these T cell-rich
326 clonalTIMEs, all three subpopulations of $CD8^+$ T cells-$CD8^+$ Tem, $CD8^+$ Trm, and $CD8^+$ Tex-
327 were notably enriched (Fig. 4d). Analysis of the $CD4^+$ T cell compartment indicated significant
328 enrichment of $CD4^+$ Tm and Th1/Th17 within the T cell-rich clonalTIMEs (Fig. 4d).
329 Additionally, $CD4^+$ Tn and $CD4^+$ Tfh showed enrichment within these T cell-rich clonalTIMEs,
330 while $CD4^+$ Treg did not exhibit a significant enrichment pattern between these two
331 clonalTIME phenotypes in our ST dataset (Fig. 4d).
332 We next explored local cues that may influence the infiltration of cytotoxic T cells and the
333 transition of clonalTIMEs. Given that antigen-presenting cells (APCs) such as DCs and
334 macrophages can instruct T cell response[56-60], we assumed that these APC subpopulations with

335 functional complexity might exhibit spatial specifications that either support T cell infiltration
336 or hinder it. To this end, we identified potential macrophage-T multicellular communities. By
337 analyzing the PCC between various cell subpopulations across thirteen tumor clones, we
338 identified a community (community3) comprising *C1QC*$^{+}$ macrophages, CD8^{+} Tem, and CD8^{+}
339 Tex, which displayed a significant enrichment in T cell-rich clonalTIMEs (Fig. 4e, f, and
340 Supplementary Fig. 76a). This enrichment pattern of community3 was observed during the
341 progression of intra-tumoral clonalTIME transitions, indicating the spatiotemporal dynamics of
342 this community during tumor evolution (Fig. 4g).
343 Intriguingly, *C1QC*$^{+}$ macrophages exhibited upregulated expression of genes encoding
344 migration-inducing chemokines (*CCL8*, *CCL13*, and *CXCL12*), major histocompatibility
345 complex (MHC) molecules (*HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, and *HLA-*
346 *DQB2*), and complement components (*C1QA*, *C1QB*, and *C1QC*) (Supplementary Fig. 76b, c).
347 Furthermore, co-expression module analysis also suggested that *C1QC*$^{+}$ macrophages support
348 T cell activation in CRC[57,61,62] (Supplementary Fig. 65d-f). However, employing alternative
349 methods for ultra-resolution characterization of the clonalTIME revealed significant limitations.
350 For instance, cell2location, CARD, SPOTlight, and RCTD failed to capture the intra-tumor
351 heterogeneity across the distribution of T cell subpopulations at a clonal resolution
352 (Supplementary Fig. 77). Moreover, stereoscope, SPOTlight, and CARD failed to detect the co-
353 occurrence pattern between *C1QC*$^{+}$ macrophages and CD8^{+} Tem or CD8^{+} Tex (Supplementary
354 Figs. 26-33, 50-57, 78). Together, these data show that UCASpatial links the tumor genomics
355 with diverse immune microenvironments at a clonal resolution, and reveals the spatiotemporal
356 dynamics of multicellular communities during clonal evolution in CRC.
357

358 **UCASpatial identifies chr20q-gain as a tumor genomic feature associated with the**
359 **formation of T cell-excluded clonalTIME in CRC**

360 To investigate whether genetic alterations are associated with immune phenotypic transitions
361 in tumor clones, we conducted a comparison of CNVs across clones using our 'UCASpatial-
362 clonalTIME' platform (Supplementary Fig. 79a, and Supplementary Table. 6). We found that T
363 cell-excluded clones displayed specific CNVs compared to T cell-rich clones (Fig. 4d, and Fig.
364 5a). Notably, we observed a frequent copy number gain on chromosomes 20q and 20p (chr20q-
365 gain and chr20p-gain) in T cell-excluded clones. We hypothesized that these copy number gains
366 on chr20q and 20p emerged during clonal evolution and were co-opted with clonalTIME
367 transitions.
368 To further investigate this, we focused on a particular section (CRC4) containing three

clones with sequential clonal events. In this section, clone1, with chr1p-loss-2 as a common ancestor, gave rise to clone2 with chr8p-gain and chr8q-gain, alongside clone3 with chr20q-gain, chr20p-gain, and three additional CNVs (Fig. 5b, c). Interestingly, as clone1 progressed to clone2, the clonalTIME remained T cell-rich (Fig. 4d, Fig. 5a). However, upon progression from clone1 to clone3, we observed a significant decrease in $CD8^+$ T cell infiltration, especially $CD8^+$ Tex, and myeloid infiltration, especially *C1QC*$^+$ macrophages within their clonalTIMEs (Fig. 4d, e).

For further validation, we conducted an analysis of the PCC between the CNVs detected in CRC4 and the signature expression of distinct cell subpopulations using the TCGA-COAD dataset (Supplementary Fig. 79b). Our examination revealed significant correlations between clone3-specific CNVs, encompassing chr20q-gain, chr20p-gain, chr7p-gain, chr7q-gain, and chr17p-loss, and reduced signature expression of $CD8^+$ Tex, $CD8^+$ Tem, $CD8^+$ Trm, $CD4^+$ Tm, $CD4^+$ Tfh, and $CD4^+$ Treg cells (Fig. 5d). Notably, these five CNVs exhibited a concurrent occurrence pattern within the TCGA-COAD dataset (Fig. 5e). To ascertain which CNV primarily contributes to diminished T cell infiltration, we isolated the individual impacts of each CNV in the TCGA-COAD dataset based on the status of these five CNVs (Methods). Our analysis revealed that only the group harboring chr20q-gain, prevalent in 57.9% of CRC patients, exhibited a significant reduction in the expression levels of T cell marker genes (*CD3G*, *CD8A*, and *CD4*) compared to the wild-type (WT) group (Fig. 5f and Supplementary Fig. 79c). These findings suggest that the emergence of chr20q-gain may contribute to the transition of the clonalTIME phenotype from T cell-rich to T cell-excluded.

Molecular characteristics of clones with T cell-excluded clonalTIME

We proceeded to investigate the spatial molecular characteristics of these T cell-excluded tumor clones. Functional enrichment analysis unveiled an enrichment for pathways related to translation machinery among the upregulated genes in clone3 as compared to clone1 and clone2. Conversely, pathways related to type I interferon (IFN) production were notably enriched among the downregulated genes (Fig. 5g, and Supplementary Fig. 80a-d). 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*, exhibited reduced expression[63,64] (Fig. 5g, and Supplementary Fig. 80a-c). This observation potentially elucidates the T cell-excluded phenotype of clone3.

Research has demonstrated that the induction of the type I IFN response in tumors can be attributed to the innate sensing of cytosolic RNA derived from TEs[65]. Considering that TEs

403 serve as internal adjuvants eliciting T cell responses against tumors, we hypothesized that
404 differences in TE levels might underlie the immune phenotypic transition from T cell-rich to T
405 cell-excluded. To examine this hypothesis, we quantified the spatially-resolved expression of
406 TEs[66] (Fig. 5h and Methods). Our analysis revealed that two CRC-specifically upregulated TEs-
407 the internal sequence of human endogenous retrovirus subfamily H (*HERV-H*) and its
408 associated long terminal repeat *LTR7Y*-were downregulated in clone3 with chr20q-gain
409 compared to clone1 and clone2[67,68] (Fig. 5i, j, Supplementary Fig. 80d, and Methods). These
410 findings suggest that *HERV-H* downregulation could be employed by tumor cells as a
411 mechanism to evade immune recognition in CRC. Correspondingly, we also noted decreased
412 expression levels of *HERV-H* and attenuated signaling activity in viral transcription and type I
413 IFN production in tumor samples with chr20q-gain relative to chr20q-WT in TCGA-COAD
414 (Fig. 5k, l, Supplementary Fig. 80e, and Methods).
415 Considering the association of chr20q-gain with T cell excluded clonalTIME in CRC, and
416 the potential impact of T cells on clinical response to immune checkpoint blockade (ICB)
417 treatment, we hypothesized that chr20q-gain might influence the response to ICB treatment in
418 CRC patients. To examine this, we evaluated survival post-ICB treatment by categorizing
419 patients from Memorial Sloan Kettering Cancer Center (MSKCC) IMPACT (MSK-IMPACT)
420 cohorts (comprising 71 patients) into two groups based on their chr20q status[69,70]. Notably, we
421 observed a correlation between the presence of chr20q-gain and poorer survival in patients
422 undergoing ICB, which was not evident in CRC patients who did not receive ICB treatment
423 (TCGA-COAD dataset) (Fig. 5m, and Supplementary Fig. 80f, g). To rule out the potential
424 impact of tumor mutation burden (TMB), we categorized patients based on both CNV status
425 and TMB. Our analysis indicated that the correlation between chr20q-gain and diminished
426 survival in the CRC ICB cohort remained irrespective of TMB levels (Supplementary Fig. 69h).
427 Taken together, we identified chr20q-gain, a prevalent genomic defect in CRC, as a genomic
428 marker for T cell exclusion, suggesting its utility as an immunogenomic parameter for
429 predicting responsiveness to ICB. Furthermore, we propose a potential model wherein tumor
430 cells orchestrate a T cell-excluded clonalTIME by suppressing *HERV-H* expression during
431 clonal evolution.
432
433 **Discussion**
434 Most current spatial deconvolution methodologies excel in decomposing well-structured tissues
435 like the mouse brain[9-13,16]. However, unlike these structured tissues, sublayers within complex
436 microenvironments typically exhibit subtle differences, often with low abundance and high

437 heterogeneity. Here, we introduce UCASpatial, a robust, sensitive, and versatile algorithm
438 designed to deconvolute ultra-resolution spatiotemporal cellular composition. Additionally, we
439 introduce clonalTIME, a downstream function tailored to analyze UCASpatial results within
440 the TIME. These tools are seamlessly integrated into the Seurat framework, facilitating a broad
441 range of applications.

442 UCASpatial shares conceptual similarities with existing methodologies. For instance,
443 SPOTlight utilizes a seeded topic model and a modified non-negative least squares (NNLS)
444 approach to estimate cell type compositions[9], while stereoscope employs a model-based
445 probabilistic method to estimate both cell type signatures and relative cell abundance across
446 spots. RCTD[10] and cell2location[11] utilize statistical models (Poisson distribution and Bayesian
447 model, respectively) to estimate cell abundance across spots, while CARD[12] incorporates a
448 conditional autoregressive (CAR) assumption into the NMF model, accommodating spatial
449 correlation in cell-type composition across spots. However, specific meta-purity filtering and
450 entropy-based weighting mechanisms in UCASpatial offer distinct advantages, as evidenced by
451 both our in-silico benchmarking results and analysis of real ST datasets (Supplementary Figs.
452 2-7, 16, 66-74). The meta-purity filtering design in UCASpatial filters out intermediate cells
453 with transcriptomes similar to other cell types, thereby enhancing the representativeness of each
454 cell subpopulation[71-73]. Furthermore, through entropy-based weighting, UCASpatial enhances
455 the discrimination of fine cell subpopulations with subtle transcriptomic differences by
456 quantifying the ability of each gene to indicate cell identity and incorporating this information
457 in the deconvolution process.

458 Moreover, UCASpatial demonstrates versatility across platforms for spatial transcriptomic
459 sequencing. To illustrate, we applied UCASpatial to mouse hippocampus data from the Slide-
460 seqV2 platform[12,74], exemplifying its adaptability. As anticipated, UCASpatial accurately
461 delineated the three anatomical structures of the hippocampus: the cornu ammonis 1
462 (CA1)/CA2 region, the CA3 region, and the dentate gyrus (Supplementary Fig. 81). Specifically,
463 CA1 cells predominantly localized in the CA1 region, CA3 cells exhibited high enrichment in
464 CA3, while dentate cells were situated in the dentate gyrus forming a C-shaped ring.
465 Additionally, ependymal cells, characterized by their irregular and columnar shape[75], lined the
466 brain ventricles, whereas choroid cells resided beneath the ependymal cells and were found in
467 the choroid plexus[76] along with Cajal-Retzius cells[77] (Supplementary Fig. 81).

468 We notice that UCASpatial does not incorporate neighboring or image information, a
469 strategy commonly employed previously to enhance performance[12,29,30]. However, integrating
470 such information might not necessarily enhance the annotation of infiltrated immune cell

subpopulations with low proportions in complex microenvironments. This is due to the discrete distribution of these subpopulations resulting from recruitment dynamics. Moreover, the efficacy of UCASpatial is contingent upon the quality of the referenced scRNA-seq data. By leveraging unpaired scRNA-seq data as a reference, this limitation could potentially be mitigated. Additionally, establishing a high-quality scRNA-seq database encompassing various common tissues could address this concern.

Spatially inferred CNVs within ST provide insights into the tumor genomics landscapes at a clonal resolution. However, linking tumor genomics and diverse immune landscapes at such detailed clonal resolution remains constrained by the deciphering of the TIME. UCASpatial enables ultra-resolution analyses of TIMEs and provides an immunogenetic platform for addressing this challenge. By applying UCASpatial in human CRC, we identified a prevalent CNV, chr20q-gain, correlating with a T cell-excluded immune phenotype and poor responsiveness to ICB, which had not been previously reported. Additionally, considering the significant correlation of the chr20q-gain with the p53 mutation in CRC, and the potential effect of the p53 mutation in TE activation[78-80], we hypothesize that patients with a concurrent p53 mutation and chr20q-WT might have better responsiveness to ICB in CRC. Furthermore, extending UCASpatial to other cancer types could help identify novel CNV markers that affect clonalTIMEs or predict the efficacy of specific immunotherapies.

In future iterations, enhancing UCASpatial could involve integrating whole exome sequencing or whole genome sequencing to offer a more comprehensive elucidation of the co-option between tumor mutations and spatially-resolved clonalTIMEs. Additionally, employing spatial cellular deconvolution alongside cell interaction analysis holds promise for unraveling the spatiotemporal dynamics across multiple biological processes and understanding how they are modulated by external cues. Similarly, gaining insights into the perturbations of the microenvironment in response to disease can deepen our comprehension of pathological mechanisms and aid in the development of therapeutic strategies.

In conclusion, we anticipate that UCASpatial will find myriad applications in deciphering complex 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 spatial transcriptomics 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 $g = \{1, .., G\}$ at spots $s = \{1, .., S\}$ 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 $\boldsymbol{x}_s$ represents the proportions of the K cell types on spot s. By default, this matrix is estimated using an entropy-weighted non-negative least square 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 R package 'Seurat' with a default resolution of 100. We recommend the resolution parameter by 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 $pu(MC_i) = \max_{j \le K}(\frac{n_{ct_j}}{n_{MC_i}})$, where MC_i denotes the ith 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.

Down sampling

Another optional preprocessing for referenced scRNA-seq is down sampling. 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 $\widetilde{\mathbf{B}} \in \mathbb{R}^{G \times \widetilde{N}}$ has been established, in which $\widetilde{N} = 100 \times K$ by default.

Optimized non-negative matrix factorization (NMF).

In order to efficiently extract the cGEP matrix $\mathbf{E} \in \mathbb{R}^{G \times K}$, we utilize non-negative matrix factorization (NMF) on the processed scRNA-seq expression matrix $\widetilde{\mathbf{B}}$. Based on the provided

cell-type labels, we initialize $\mathbf{H} \in \mathbb{N}^{K \times \tilde{N}}$ 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 $\mathbf{E}$:

$$\tilde{\mathbf{B}} \sim \mathbf{E} \times \mathbf{H} \tag{1}$$

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 heterogenous 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 $\tilde{\mathbf{B}}$:

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

where $\mathbf{PCT}_{k,g}$ denotes the element of $\mathbf{PCT}$ representing cell type k for gene g, $\mathrm{num}(\tilde{\mathbf{B}}_{g,n \in k})$ represents the number of cells in cell type k in $\tilde{\mathbf{B}}$, $\mathrm{num}(\tilde{\mathbf{B}}_{g,n \in k} > 0)$ 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 $\mathbf{PCT}$ being divided by the sum of the corresponding row:

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

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

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

For better understanding, UCASpatial provides an adjusted entropy using the maximum entropy minus the entropy for each gene g_0:

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

where $Ent.maxm = \sup(\text{Entropy}(g)) = -\sum_1^K \frac{1}{K} \times \log(\frac{1}{K})$ 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:

$$\omega(g) = \max_{i \in \{1,2,\ldots,K\}} \left(\frac{2 \times \text{adj.ent}(g)}{Ent.maxm} \times 2 \times \widetilde{\mathbf{PCT}}_{i,g} \times \text{avg_logFC}(i,g)\right) \tag{6}$$

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

where $\text{avg_logFC}(i,g)$ 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 cosine similarity method[81] (version 0.9.0), for optional gene filtering in cGEPs, which can further improve the indicative capacity of genes in cGEPs.

Entropy-weighted non-negative least square 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[5,82]. Commonly, this step will utilize the non-negative least square (NNLS) model for each spot *s*:

$$\mathbf{P}_s = \mathbf{X}_s \times \mathbf{E} + \varepsilon_s \tag{8}$$

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

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

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 *j* 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 $\mathbf{P}_{s,g}$

where $j \in \{1,2,\ldots,G\}$. By minimizing the loss function $L(\varepsilon) = \|\varepsilon_s\|_2^2$, we can obtain the $\mathbf{X}_{s,i}$ which 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-weighted Non-Negative Least Squares (ENNLS) model. We first extend the weight vector $\omega(g)$ to a diagonal matrix $\mathbf{\Omega} \in \mathbb{R}^{G\times G}$:

$$\mathbf{\Omega} = \begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_G \end{pmatrix} \tag{10}$$

the corresponding ENNLS model has been extended into:

$$\mathbf{P}_s\sqrt{\mathbf{\Omega}} - \mathbf{X}_s\mathbf{E}\sqrt{\mathbf{\Omega}} \tag{11}$$

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

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

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

By minimizing the weighted loss function $\min \tilde{L}(\varepsilon) = \min \left\|\tilde{\varepsilon}_s\right\|_2^2$, 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 $\mathbf{X}$.

620

621 **Simulation strategy and evaluation metrics.**

622 **Simulation strategy**

623 We generated simulated spatial transcriptomic data sets from a reference CRC scRNA-seq data
624 set (Supplementary Figs. 1a-c and 4, see in the *Human specimens (in-house cohort)* part). To
625 simulate different situations of tumor microenvironment, we generated three different regions
626 including tumor core (TC), tumor margin (TM), and tumor stroma (TS)[16,54]. Each region has
627 five replicates with each replicate containing N_{spot} = 50 spots, generated based on random seeds
628 (19980205, 19990911, 19960330, 20000502, 20230509) The number of cells in each spot is
629 restricted to less than 20. To mimic the true situation as much as possible, we also imposed
630 restrictions on the types of cells present in different regions, so that spots from different regions
631 have different cell type co-localization patterns. Detailed restrictions for regional simulation are
632 listed in Supplementary Table 1. To better mimic the real capture, if the synthetic spot had over
633 25,000 UMI counts, it would be randomly down-sampled into 20,000 UMI counts.

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

641 **Construct complexity hierarchy**

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

647 **Evaluation metrics**

648 In each simulation replicate, we calculated the true cell-subpopulation proportions on each
649 spatial location as the number of cells in each cell subpopulation divided by the total number
650 of cells in the location. $T_{s,k}$ and $P_{s,k}$ represent the true and predicted proportion of cell
651 subpopulation *k*, respectively, in the *s*th spot through the number of total cell subpopulations *K*.
652 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:

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

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:

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

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:

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

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

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

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.

$$\text{F1 score} = \frac{2 \cdot \frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \tag{17}$$

Compared methods.

We compared UCASpatial with six 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 (v 0.3.1). Results of the stereoscope were computed with parameters "-sce 25000 -n 5000 -ste 25000 -stb 500 -scb 500 -gp". For other methods, we followed the tutorial on the corresponding GitHub pages and used the recommended default parameter settings for

681 deconvolution analysis.

682

683 **Experimental and Analysis Approach to Mouse Ear**

684 **Ear wound creation**

685 All mice utilized in the wounding experiments were aged between 8-10 weeks. Mice were
686 subjected to short-term anesthesia induced by isoflurane prior to wound generation. The dorsal
687 and ventral aspects of mouse ears were sanitized with cotton balls saturated in 75% alcohol.
688 Mouse ears were delicately flattened using forceps, and a tissue-penetrating ear puncher (2 mm)
689 was used to create perforating injuries within the vascularized regions of the ear. Hemostatic
690 powder was applied to the wounds using disposable cotton swabs. All instruments were
691 sterilized under high pressure before use.

692 **Single-cell RNA-seq**

693 Single-cell suspensions were obtained from MRL mouse ear tissue at day 7 post-wounding.
694 Specifically, mouse ear tissues were mechanically minced and then incubated at 37°C for 5-10
695 minutes with 40 μL of 50 mg/mL collagenase II and 40 μL of 50 mg/mL collagenase IV.
696 Following a brief low-speed centrifugation to precipitate undigested tissue fragments, the
697 supernatant was transferred into a stop solution composed of fetal bovine serum, EDTA, and
698 D-HBSS. Additional digestion enzyme was added to the partially digested tissue fragments, and
699 the digestion process was repeated until complete tissue digestion was achieved. The digested
700 mixture was filtered through a 70 μm membrane and centrifuged at 4°C at 1200 rpm for 10
701 minutes. After discarding the supernatant, the cell pellet was resuspended in FACS buffer for
702 cell sorting. Calcein and 7-AAD were freshly added before cell sorting. Then $FITC^{+}$ $7\text{-}AAD^{-}$
703 single cells were sorted for library construction of scRNA-seq. The libraries were performed
704 using Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit V3 (PN-1000075)
705 purchased from 10x Genomics according to the protocols provided by manufacturers. The
706 libraries were sequenced on an Illumina Novaseq 6000 platform.

707 **Spatial gene expression assay**

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

platform.

Single-cell RNA-seq data processing

10x Genomics scRNA-seq data were aligned and quantified using the Cell Ranger Single-Cell Software Suit (version 3.0.2) against the mm10 reference genome. Genes with less detected cell numbers (less than 3) and cells with less detected genes (less 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 5,000; (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 2,000 variable genes identified by the vst method were used to perform principal component analysis. 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. DoubletFinder was utilized to predict potential doublets and filter them out.

Spatial transcriptomics data processing

Using the Space Ranger software (https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/tutorials), the 10x Visium spatial transcriptomics data was 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 spatial transcriptomics 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 using default. Parameters for other methods: Default. Using the threshold of 3% to filter the noise after decomposition by each method (Supplementary Figs. 10-16).

Spatial transcriptomics 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 spatial transcriptomics 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 col, and each column replaced by the row. We next transformed the coordinates of each spot's cols abiding by the time points. ggplot2 was used for visualizing spatial transcriptomics data after coordinate transformation.

749

750 **Experimental and Analysis Approach to Human CRC**

751 **Tissue dissociation of patient samples**

752 Fresh human tumor specimens and patient-matched adjacent normal tissues were washed twice
753 with DPBS (gibco) and then mechanically disrupted into approximately 1 mm^3 pieces in the
754 RPMI-1640 (gibco) and further enzymatically digested with Human Tumor Dissociation Kit
755 (Miltenyi Biotec, 130-095-929) in C-tube in combination with gentleMACS™ Dissociator
756 (Miltenyi Biotec, 130-093-235) presetting program h_tumor_01, h_tumor_02 and h_tumor_03
757 sequentially according to the manufacturer's instruction. Between each program, samples were
758 incubated for 30 minutes at 37°C under continuous rotation using a rotator. The dissociated cells
759 were subsequently passed through a 70-μm cell strainer (Biologix, 15-1070) and centrifuged at
760 300g for 7 min. After the supernatant was removed, the cell pellet was washed with DPBS. The
761 pelleted cells were resuspended in a cell cryopreservation medium (FBS with 10% DMSO) at
762 a concentration of 1×10^6 cells/ml and further cryopreserved in liquid nitrogen for long-term
763 storage.

764 **Single-cell RNA-seq for patient samples**

765 On the day of scRNA-seq droplet capture, cells were thawed rapidly in a 37°C-water bath and
766 washed off DMSO in DPBS. Palleted cells were resuspended in a sorting buffer (PBS with 2%
767 FBS and 2 mM EDTA). Cells were stained with CD45-FITC (Biolegend, 368508) for 20
768 minutes at 4°C in the dark. 7-AAD was freshly added prior to FACS sorting performed on the
769 BD FACS Aria III instrument. Based on FACS analysis, 7-AAD^-CD45^+ cells and 7-AAD^-CD45^-
770 cells were sorted into 1.5ml tubes (ExCell Bio, CS015-0041) and counted manually under the
771 microscope (Leica), respectively. Libraries were generated using Chromium Next GEM Single
772 Cell 5' Library & Gel Bead Kit v1.1 (10x Genomics, PN-1000165) according to the
773 manufacturer's protocol. Purified libraries were sequenced by the Illumina NovaSeq 6000
774 platform with 150-bp paired-end reads.

775 **10x Visium spatial transcriptomics**

776 CRC samples were frozen and embedded in OCT (Sakura Tissue-Tek). All tissue blocks were
777 sectioned using the Leica CM1950 cryostat and were cut at 10 μm. Samples were selected on
778 the basis of morphology, orientation (based on H&E staining), and RNA integrity number (RIN)
779 that was obtained using Agilent 2100 Bioanalyzer. The RIN of selected samples ranged from
780 4.5 to 8.3. The sections were treated as recommended by 10x Genomics and the optimization
781 procedure showed the optimal permeabilization time determined as 12 min for CRC samples
782 using the Visium Spatial Tissue Optimization Kit (10x Genomics, PN-1000193). Spatial gene

783 expression slides were equilibrated inside the cryostat at -20°C before sectioning and all of the
784 reagents used for spatial transcriptomics were from the Visium Spatial Gene Expression Kit
785 (10x Genomics, PN-1000187). According to the manufacturer's instruction, the tissue sections
786 were fixed in chilled methanol at -20°C for 30 min and stained with H&E. Brightfield
787 histological images were performed on the Olympus Microscopes BX53 at ×20 magnification
788 and then taken by cellSens Dimension (version 1.18) software. Images were exported as tiled
789 tiffs for analysis. After imaging, tissue sections were permeabilized at 37°C for optimal
790 permeabilization time of release of mRNA as mentioned above. Following permeabilization,
791 mRNA was reverse-transcribed and underwent second-strand synthesis, denaturation, cDNA
792 amplification, and cleanup by SPRIselect beads sequentially (Visium, 10x Genomics, User
793 Guide CG000239). cDNA amplification cycles were determined by quantitative PCR (qPCR).
794 Library quality was evaluated using TapeStation (Agilent Technologies). Sequencing was
795 performed on the Illumina Hiseq X Ten platform.

796 **Human specimens (in-house cohort)**

797 One male (CRC3) and two female (CRC1 and CRC2) patients were enrolled and pathologically
798 diagnosed with colon cancer at Peking University Cancer Hospital & Institute (Supplementary
799 Table. 3). All patients provided written informed consent. This study was approved by the
800 Institutional Ethics Committee of Beijing Cancer Hospital (2023YJZ16). Among these patients,
801 CRC1 was diagnosed as stage II, and CRC3 and CRC2 were diagnosed as stage IV according
802 to the guidance of AJCC version 8. Fresh tumor tissues and patient-matched normal adjacent
803 colon tissues were collected from surgery resection, which was never treated with
804 chemotherapy or radiation therapy prior to the study. The resected samples were stored in RPMI
805 1640 (gibco) added with Amphotericin B (Shelleck, S1636) and Gentamicin Sulfate (gibco,
806 15710-064) and transported at 4 ˚C immediately. A small part of each tissue was frozen and
807 embedded in OCT (Sakura Tissue-Tek) and the remainder of the same tissue was dissociated to
808 a single-cell suspension.

809 **Single-cell RNA-seq data processing and annotation**

810 In-house single-cell RNA-seq data from 10x Genomics were processed using Cell Ranger
811 (version 6.0.1), aligning to the hg38 reference genome. Low-quality cells were filtered out if
812 they expressed fewer than 500 or more than 7,000 genes. Cells with mitochondrial gene
813 expression exceeding 20% were also excluded. Batch effects of in-house and public scRNA-
814 seq datasets were corrected using the integration workflow in the Seurat R package[83] (version
815 3.1.4). The top 2,000 variable features were used to identify anchors using the
816 FindIntegrationAnchors function, and data from in-house and public scRNA-seq datasets were

817 integrated using the IntegrateData function.

818 To identify major cell types, the top 50 PCs were calculated using the RunPCA function, and
819 the top 20 PCs were used as input to the uniform manifold approximation and projection
820 (UMAP) algorithm (RunUMAP). Nearest neighbor graphs were constructed using the
821 FindNeighbors function, and clusters were identified using the Louvain function (resolution =
822 0.6). Cell types were annotated based on well-known marker genes.

823 A second round of clustering was performed to further characterize subpopulations within
824 each cell type. Due to the variability in cell number and features among cell types, T&NK,
825 myeloid, and epithelial cells were re-integrated using the top 2,000 variable features and re-
826 clustering using different parameters, respectively. NK and T cells were clustering (resolution
827 = 0.4) on the basis of the top 15 PCs. Myeloid cells were clustering (resolution = 0.4) on the
828 basis of the top 30 PCs. Epithelial cells were clustering (resolution = 0.05) on the basis of the
829 top 30 PCs. One T cell and myeloid cluster with few specific markers but exhibiting high
830 expression of cell cycle-related genes was removed, as these genes could not be fully regressed
831 out after cell cycle score scaling. In total, we defined 34 cell types in CRC, and this CRC
832 scRNA-seq reference was used throughout the manuscript to maintain consistency in presenting
833 spatial mapping results (Fig. 4b, Supplementary Fig. 17b, c).

834 **Spatial transcriptomics data processing**

835 Raw sequencing reads of spatial transcriptomics were quality-checked and mapped by Space
836 Ranger v1.3.0. The gene-spot matrices generated after ST data processing from ST and Visium
837 samples were analyzed with the Seurat package in R. The processed data were used for
838 downstream analysis of UCASpatial (meta.resolution = 300, meta.purity = 0.9, other
839 parameters using default) and other methods (using default parameters). Using the threshold of
840 3% to filter the noise after decomposition by each method (Supplementary Figs. 18-73).

841 **Gene signature score calculating in ST**

842 Select the top 20 genes of each cell subpopulation provided in Supplementary Table. 2 or
843 Supplementary Table. 5 as the signature gene list for murine wound healing ST dataset and
844 human CRC ST dataset, respectively. Use the "AddModuleScore" method in the Seurat R
845 package to calculate the signature score of each cell subpopulation in the ST dataset using the
846 signature gene list.

847 **Gene co-expression module analysis**

848 To analyze the gene program pattern in macrophage subpopulations, the correlations of the top
849 2000 variance genes obtained from single-cell clustering step by the Seurat R package were
850 calculated with the Pearson method. To identify the gene co-expression module, a hierarchical

clustering method with complete distance was used for the correlation matrix of top variance genes. Seven co-expression modules were obtained (Supplementary Fig. 76d-f). Gene expression data were normalized by the "LogNormalize" method in the Seurat R package. The z-score transformed normalization data were then used to calculate the correlations of top variance genes. The score was calculated by averaging the normalized expression of genes in each set.

Tumor clonal inference and clonalTIME quantification using UCASpatial-clonalTIME

To infer tumor clonal diversity within each spatial transcriptomics section, 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[46]. 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 spatial transcriptomics section 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. 4c). 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. 5a-c, and Supplementary Fig. 79a). CNV regions locating on the same chromosome with similar variation patterns in the TCGA-COAD dataset were merged if their PCC (log_2FC_CNV) > 0.8, where log_2FC_CNV denotes the log2 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. 6).

ClonalTIME phenotyping and identification of multi-cellular community

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).

885 ClonalTIME were classified as T cell-riched or T cell-excluded based on the proportion of T
886 cell subpopulation within each clustering group (Fig. 4d, and Supplementary Fig. 75b-d). To
887 determine whether a given cell subpopulation was enriched in a specific clonalTIME phenotype,
888 we calculated the ratio of observed cell number to expected cell number (Ro/e) for T cell-rich
889 clonalTIME, where the expected proportion of cell subpopulation was determined from the chi-
890 squared test. We assumed that a subpopulation was enriched in T cell-rich clonalTIME if Ro/e >
891 1.2 and enriched in T cell-excluded clonalTIME if Ro/e < 0.8. To identify the cell subpopulation
892 co-change pattern at a clonal resolution, the correlations of the average proportion of cell
893 subpopulation in each clone were calculated with the Pearson method. To identify the multi-
894 cellular community, a hierarchical clustering method with complete distance was used for the
895 correlation matrix. Four cellular communities were obtained (Fig. 4e, and Supplementary Fig.
896 76a).

897 **Spatial quantification of TE expression**

898 We quantify the expression of transposable elements (TEs) using spatial transcriptomic data
899 based on a previously reported algorithm[66]. This algorithm allocates TE reads to TE metagenes
900 based on the TE type-specific sequence (Fig. 5h). In brief, the genes/TEs and cell count matrix
901 were generated according to default parameters and normalized using the NormalizeData
902 function. Clonal-specific TE expression programs were identified by comparing each tumor
903 clone region using the FindAllMarkers function. The list of CRC-specifically upregulated TEs
904 was downloaded from previous literature[67]. The *HERV-H* expression level was calculated as the
905 average expression of previously reported *HERV-H* elements (*HERVH-int*, *LTR7*, *LTR7B*, and
906 *LTR7Y*) (Fig. 5j).

907 **Differential expression analysis using spatial transcriptomics**

908 Differentially expressed genes across each clone were analyzed using the Seurat R package
909 (version 4.0.5). Space Ranger output files were imported, and the data were normalized and
910 scaled using the default NormalizeData and ScaleData functions. Differential gene expression
911 analyses were performed comparing groups using the FindMarkers function, with the following
912 parameters: $\log_2$(fold change) ≥ 0.1 and adjusted P-value ≤ 0.05 (Fig. 5g). The list of CHIP-
913 seq-based STAT1 target genes was downloaded from previous literature[84].

914 **Functional enrichment analysis**

915 The signatures of macrophage subpopulations in murine wound healing and human CRC were
916 obtained in scRNA-seq data by FindAllMarkers function in the Seurat package with min.pct =
917 0.25, min.diff.pct = 0.25, logfc.threshold = 0.25 and other default parameters. Metascape was
918 used to perform functional enrichment analysis for each macrophage subpopulation[85]

919 (Supplementary Fig. 76f). For differentially expressed genes in the CRC4 section, Gene
920 Ontogeny (GO) enrichment analysis was performed by clusterProfiler (v4.6.0)[86] (Fig. 5g).

921 **TCGA-COAD data analysis**

922 Gene expression and clinical data for Colon Adenocarcinoma (COAD) were downloaded from
923 The Cancer Genome Atlas (TCGA). Average copy number data were downloaded from the
924 UCSC Xena browser (https://xena.ucsc.edu). For CNV analysis, only the samples with CNV
925 information were obtained. The average log_2FC_CNV > 0.35 was defined as copy number gain,
926 while the average log_2FC_CNV < -0.35 was defined as copy number loss (Fig. 5f). The
927 signatures of each cell subpopulation were obtained in CRC single-cell data by the
928 FindAllMarkers function in the Seurat package with default parameters. The normalized
929 expression of the top 10 signatures in each cell subpopulation was used to calculate the average
930 signature score. The reads per million (RPM) of known human endogenous retroviruses (HERV)
931 in TCGA-COAD were downloaded from a previous report[87,88]. For HERV analysis, samples
932 without HERV expression information were excluded. Overall *HERV-H* expression was
933 calculated as the average expression of known transcribed *HERV-H* loci (*HERV-H1-7*)[87] (Fig.
934 5k). The pathway activity of each sample was calculated using GSVA with default parameter
935 settings[89] (Fig. 5l). Survival analyses were then performed with the survival R package (version
936 3.4-0) (Fig. 5m, and Supplementary Fig. 80g).

937 **MSK-IMPACT data analysis**

938 The genomic and clinical data of 71 CRC patients with Immune Checkpoint Blockade (ICB)
939 therapy were downloaded from published studies[69,70]. The copy number status of each loci was
940 determined using the GISTIC v2.0 function in GenePattern platform[90,91]. The average
941 log_2FC_CNV > 0.35 was defined as copy number gain, while the average log_2(copy number/2)
942 < -0.35 was defined as copy number loss. Survival analyses were then performed with the
943 survival R package (Fig. 5m, and Supplementary Fig. 80f, h).

944

945 **Analysis of mouse hippocampus Slide-seq V2 data.**

946 To test UCASpatial's ability in Slide-seq V2 platform, we used UCASpatial (cos.filter.threshold
947 = 0.1, ent.filter.threshold = 0.2, logfc.threshold = 0.1,min.diff.pct = 0.05, meta.assay = 'RNA')
948 to spatially annotate cell subpopulations in Slide-seq V2 data of the mouse hippocampus using
949 published scRNA-seq datasets of mouse hippocampus[10,92] (Supplementary Fig. 81).

950

951 **Data availability**

952 The sequencing data of both scRNA-seq and spatial transcriptomics of murine wound healing

have been deposited in the Genome Sequence Archive (GSA) repository with the accession number CRA016052. The sequencing data of human CRC have been deposited in the Genome Sequence Archive for human (GSA-human) repository with the accession number HRA007218. The public datasets used in this study include scRNA-seq dataset of human CRC (HRA000979), 10x Visium spatial transcriptomic dataset of human CRC (HRA000979), mouse hippocampus Slide-seqV2 dataset (https://singlecell.broadinstitute.org/single_cell/study/SCP948/robust-decomposition-of-cell-type-mixtures-in-spatial-transcriptomics)

Code availability

The UCASpatial package and code to reproduce the results in this study are available in the GitHub repositories at https://github.com/BIGHanLab/UCASpatial.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (NSFC) (32121001 to D.H.), Strategic Priority Research Program of the Chinese Academy of Science (XDB0570101 to D.H.), Beijing Natural Science Foundation (Z200023 to D.H.)

Author contributions

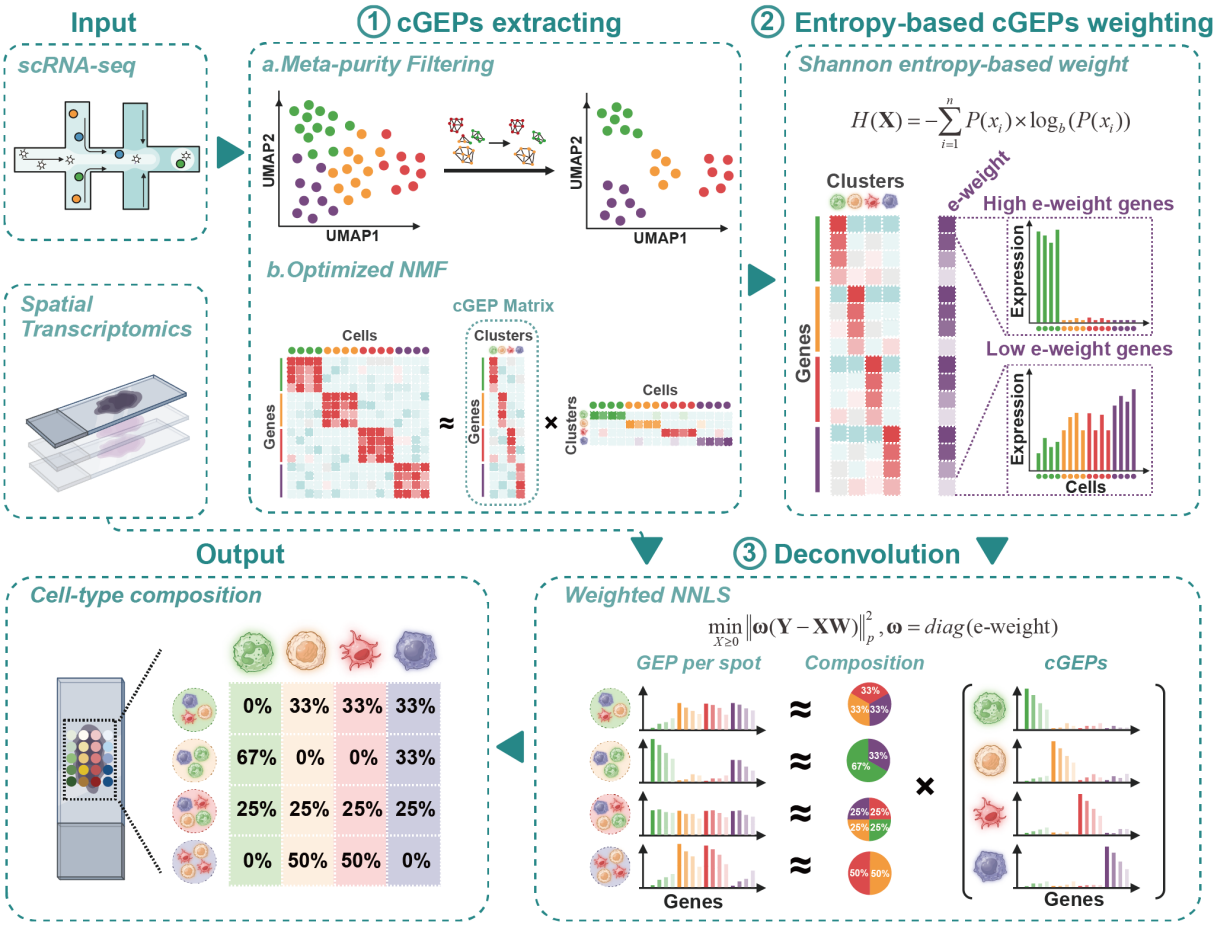
D.H. and M.M.X. conceived the study and provided overall supervision of the study. D.H., M.M.X., and Yin Xu designed the study. Yin Xu designed and developed UCASpatial. Yin Xu conducted benchmark analyses with assistance from F.Z.. R.C. conducted library construction for murine scRNA-seq. M.G. and R.C.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

981 **Main Figures**

Figure 1



982

983 **Fig. 1 | Schematic overview of UCASpatial.** UCASpatial takes both referenced single-cell
984 RNA sequencing (scRNA-seq) data with cluster information and spatial transcriptomic data as
985 input. After three-step processing involving cell type-specific GEPs (cGEPs) extracting,
986 entropy-based cGEPs weighting, and deconvolution, UCASpatial decomposes spatially
987 resolved, multi-cell RNA count matrices (i.e., spatial transcriptomic data) into a referenced cell-
988 type composition. GEP, gene expression profile. NMF, non-negative matrix factorization. e-
989 weight, Shannon entropy-based weight. NNLS, non-negative least square. Figure adapted from
990 images created with BioRender.com.

Figure 2

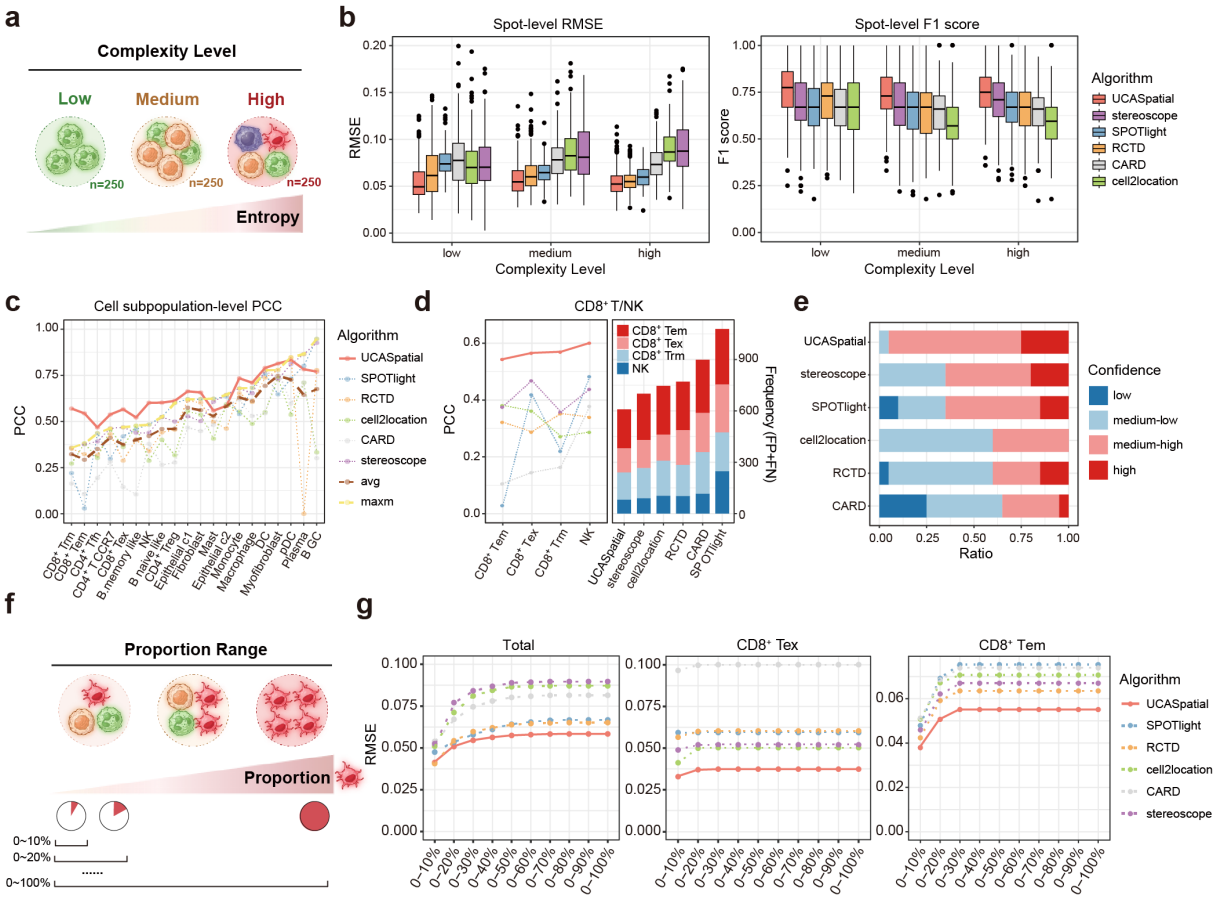


Fig. 2 | Validation of UCASpatial using simulated data. a, Graphic model showing the classification of 750 simulated spots into three groups, each representing a different complexity level - low, medium, and high, respectively. The classification is based on the entropy of cell-type composition for each spot, with each group containing 250 spots. **b**, Box plot showing spot-level root mean square error (RMSE) (left) and F1 score (right) for each method across three complexity levels. Each box plot ranges from the third and first quartiles with the median as the horizontal line, while whiskers represent 1.5 times the interquartile range from the lower and upper bounds of the box. **c**, Line charts showing cell-subpopulation-level Pearson correlation coefficient (PCC) for each method across 20 cell subpopulations. 'avg' and 'maxm' represent the average and maximum PCC, respectively, of the five 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 cells (NK) for each method. Right, stacked

1005 barplot showing the accumulated count of false predictions of these cell subpopulations for each
1006 method. Colors represent different cell subpopulations. FP, false positive; FN, false negative. **e**,
1007 The cell-subpopulation-level PCC values for each method were binned into four confidence
1008 groups: 0-0.25 (low), 0.25-0.5 (medium-low), 0.5-0.75 (medium-high), and 0.75-1 (high).
1009 Stacked barplot showing the ratio of four confidence groups for each method. **f**, Graphic model
1010 representing the classification based on the proportion range of a specific cell subpopulation
1011 within the simulated spots. All simulated spots were classified into ten groups from proportion
1012 0-10% to proportion 0-100% for each cell subpopulation. **g**, Left, line charts representing the
1013 average RMSE of 20 cell subpopulations across various proportions for each method. Middle,
1014 line charts showing an example of the RMSE for $CD8^+$ Tex across different proportions. Right,
1015 line charts showing an example of the RMSE for $CD8^+$ Tem across different proportions.
1016 Figures (**a** and **f**) adapted from images created with BioRender.com.

Figure 3

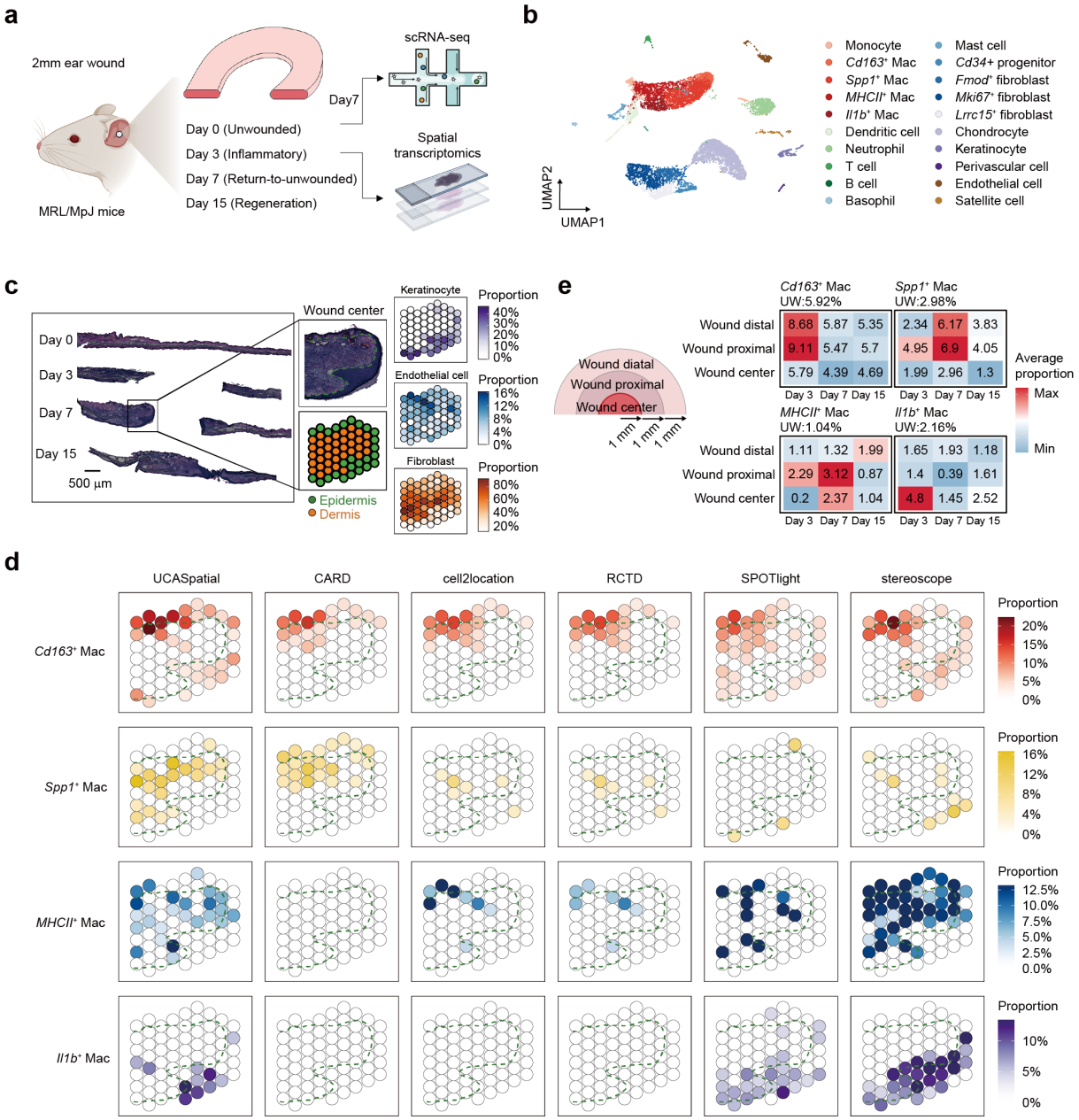


Fig. 3 | UCASpatial reconstructs spatiotemporal micro-anatomy dynamic of diverse macrophage subpopulations during murine wound healing. a, Graphical overview of experimental design and analysis workflow. scRNA-seq was performed using total cells from MRL mice ear tissue on Day 7 post-wounding. Spatial transcriptomics data were generated from MRL mice ear tissue on Day 0 (unwounded), Day 3, Day 7, and Day 15 post-wounding. Figure adapted from images created with BioRender.com. **b**, Uniform Manifold Approximation and Projection (UMAP) plot showing 20 cell subpopulations of all cells from scRNA-seq. **c**,

1025 Left, hematoxylin and eosin (H&E) staining of mouse ear tissues. Scale bar, 500 μm. Middle,
1026 micro-anatomic annotation of wound center on Day 7, the green dashed line represents the edge
1027 of the dermis-epidermis, and the red dashed line represents the blood vessel. Right, Spatial
1028 visualization of keratinocytes, endothelial cells, and fibroblasts at the wound center on Day 7.
1029 Colors represent the proportion of these three cell types. The proportion of fibroblast was
1030 calculated by the sum of 4 fibroblast subpopulations' proportion. **d**, Spatial visualization of the
1031 distribution of indicated macrophage subpopulations at the wound center on Day 7 based on
1032 UCASpatial, CARD, Cell2location, RCTD, SPOTlight, and stereoscope. Colors represent the
1033 proportion of each macrophage subpopulation. **e**, Left, schematic representation of the spot
1034 grouping criteria. Spots were divided into three subdomains: wound center (1 mm radius),
1035 wound proximal (0-1 mm from the wound center), and wound distal (1-2 mm from the wound
1036 center). Right, Heatmap showing the spatiotemporal micro-anatomy dynamics of macrophage
1037 subpopulations of three subdomains. The numbers in the heatmap represent the average
1038 proportion of macrophage subpopulations at specific subdomains and time points. Colors
1039 represent the average proportion for each macrophage subpopulation. The number of UW
1040 (above the heatmap) represents the average proportion of each macrophage subpopulation on
1041 Day 0, respectively. UW, unwounded.

Figure 4

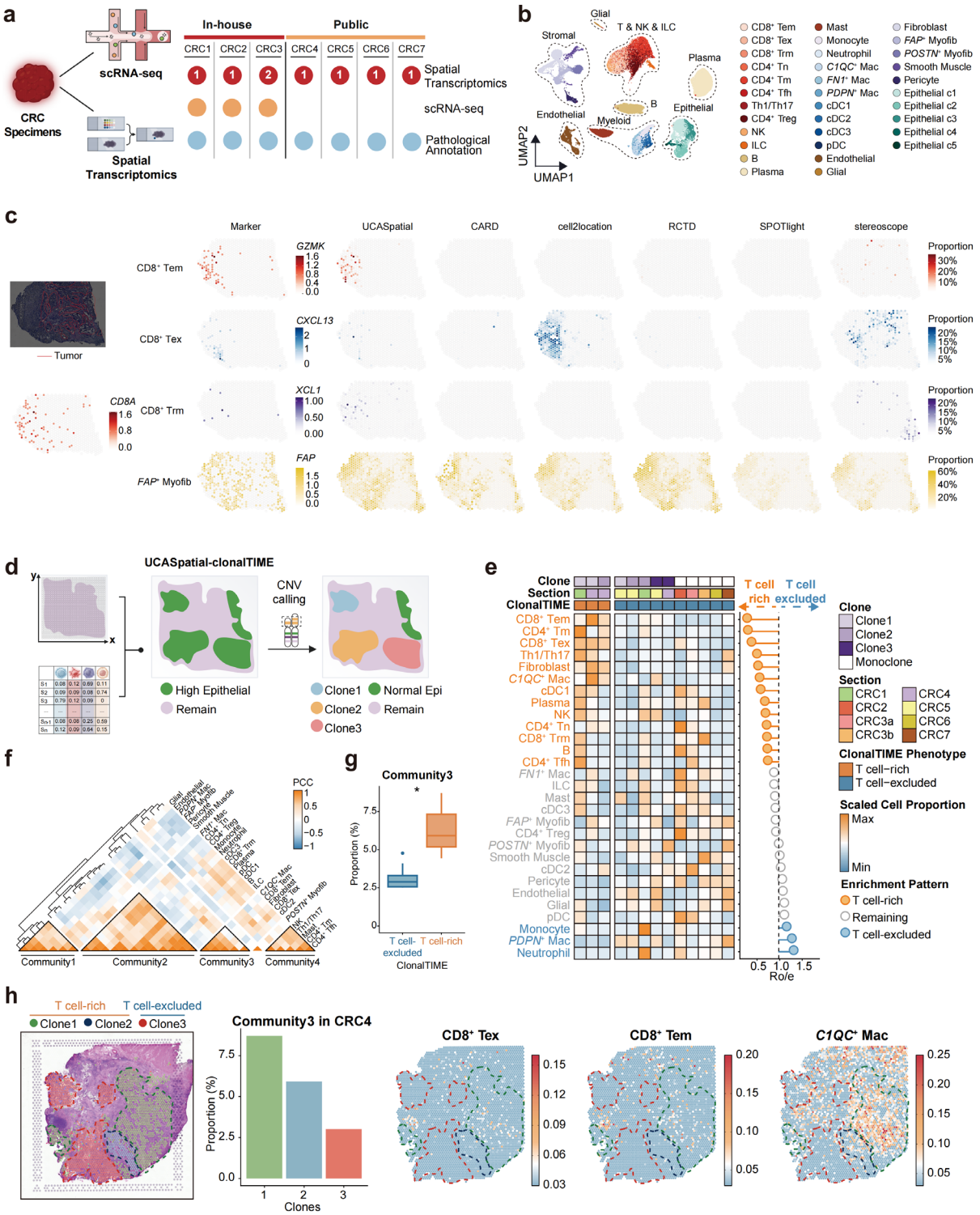


Fig. 4 | Characteristics of clonal-resolution landscape with two TIME phenotypes in human CRC. a, Left, experimental design of in-house colorectal cancer (CRC) dataset. Right, an overview of the CRC dataset used for UCASpatial analysis, with samples organized by

1046 resource. The red, orange, and blue dots represent the presence of spatial transcriptomics,
1047 scRNA-seq, or pathological annotation, respectively. The number in red dots indicates the count
1048 of spatial transcriptomic sections per patient. **b**, UMAP plot displaying 34 cell subpopulations
1049 identified from scRNA-seq reference. **c**, Left, the pathological annotations and indicating
1050 markers' expression (*CD8A*, *GZMK*, *CXCL13*, *XCL1*, and *FAP*) in the CRC3b section. Colors
1051 represent the expression level of each gene. Right, Spatial visualization of the distribution of
1052 indicated $CD8^+$ T cell subpopulations at the CRC3b section based on UCASpatial, CARD,
1053 cell2location, RCTD, SPOTlight, and stereoscope. Colors represent the proportion of each
1054 $CD8^+$ T cell subpopulation. **d**, Schematic representation of UCASpatial-clonalTIME functions.
1055 From left to right: identification of epithelial regions via UCASpatial; delineation of tumor
1056 clone regions and normal epithelial regions using spatialCNV. clonalTIME, clonal tumor
1057 microenvironment. **e**, Heatmap showing the distribution of cell subpopulations across clones
1058 grouped by clonalTIME phenotypes. Colors reflect scaled cell subpopulation proportions
1059 estimated by UCASpatial. Ro/e scores represent the enrichment pattern for each cell
1060 subpopulation between 2 clonalTIME phenotypes. Columns in the heatmap were annotated by
1061 cloneID, sectionID, and clonalTIME phenotype. Ro/e, the ratio of observed over expected cell
1062 subpopulation proportions. **f**, Heatmap showing the PCC between the proportion of each two
1063 cell subpopulations across 13 tumor clones. Communities were identified using hierarchical
1064 clustering. **g**, Boxplot showing the proportion of community3 between T cell-excluded and T
1065 cell-rich clonalTIMEs. Boxplot elements: center line, median; box limits, upper and lower
1066 quartiles; whiskers, 1%–99%. Statistical analysis was performed using a two-tailed unpaired *t*-
1067 test. $*0.01 \leq p < 0.05$. **h**, Left, the spatial distribution of each clone in the CRC4 section. Spots
1068 filled with green, blue, and red represent the tumor region of clone1, clone2, and clone3,
1069 respectively. Dashed lines mark the general contour of each tumor clone region. Middle, barplot
1070 showing the proportion of community3 in each tumor clone of CRC4. Right, spatial
1071 visualization of the proportions for 3 cell subpopulations ($CD8^+$ Tex, $CD8^+$ Tem, and *C1QC*$^+$

1072 macrophage) in community3 estimated by UCASpatial. Figures (**a** and **c**) adapted from images

1073 created with BioRender.com.

Figure 5

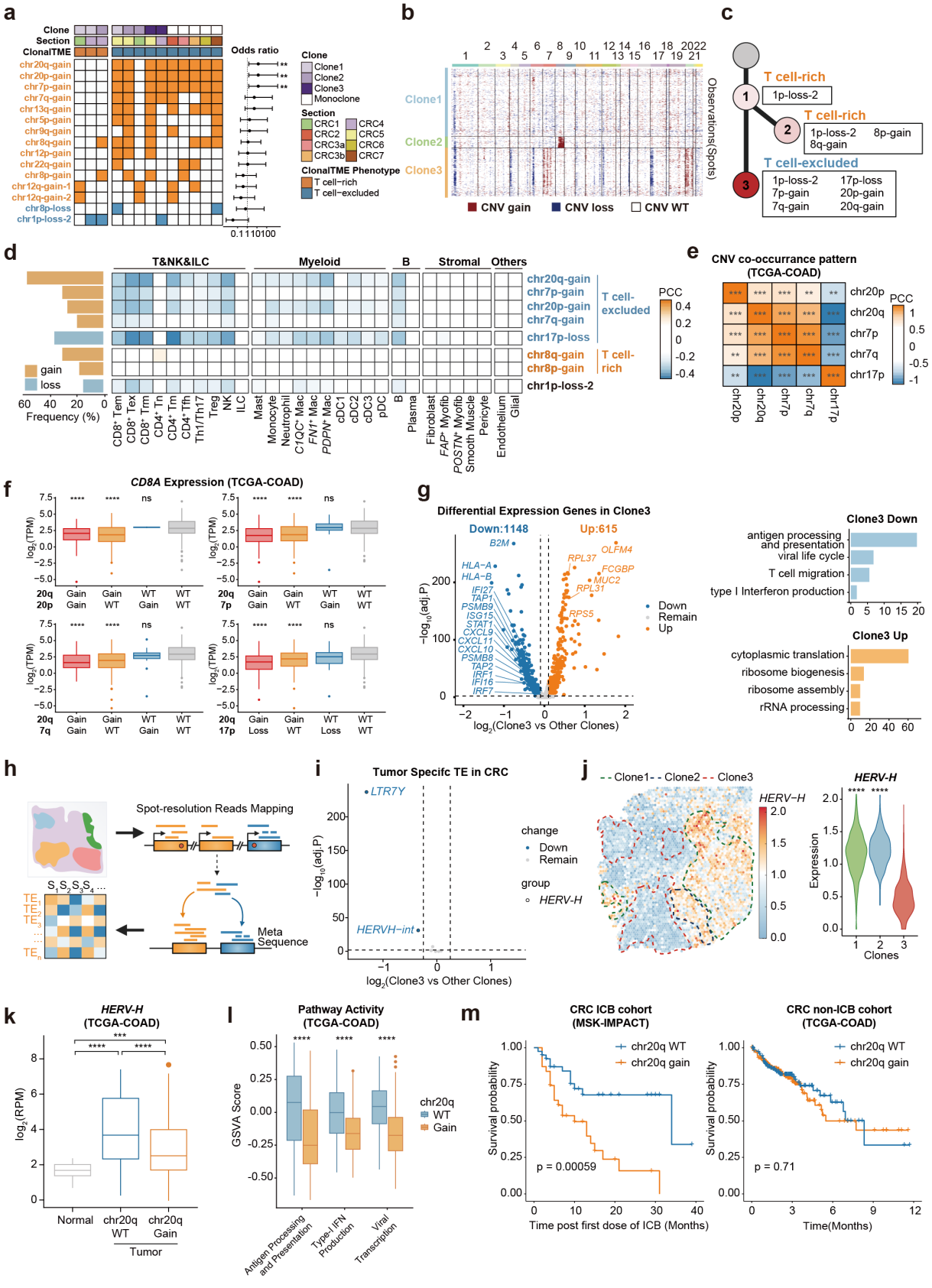


Fig. 5 | UCASpatial reveals clonalTIME transition orchestrated by chr20q-gain-

1076 **associated endogenous retrovirus silencing. a**, Heatmap showing copy number variation
1077 (CNV) landscape of each clone grouped by clonalTIME phenotypes. The color indicates the
1078 CNV status (Orange indicates CNV gain, and blue indicates CNV loss). Columns in the
1079 heatmap were annotated by cloneID, sectionID, and clonalTIME phenotype. Forest plots
1080 showing the odds ratio scores associated with clonalTIME phenotypes of each CNV. Dots
1081 indicate the calculated odds ratio and whiskers indicate 95% confidence intervals (CIs). Only
1082 CNVs occurring in at least 2 clones were shown in the heatmap. **b**, Heatmap showing the CNV
1083 profile inferred from spatial transcriptomics data from the CRC4 section. Clonal groupings of
1084 spots were determined by hierarchical clustering. **c**, Phylogenetic tree of three clones in the
1085 CRC4 section. The corresponding clonalTIME phenotype of each clone is listed alongside. The
1086 CNV status of each clone is listed (summarized by the chromosome number where the event
1087 was located, p/q arm, and gain/loss; full genomic locations of each CNV are provided in
1088 Supplementary Table 6). Gray dots represent benign cells, while red dots signify tumor clones.
1089 The intensity of red corresponds to the number of CNVs. **d**, Left, bar plot showing the frequency
1090 of each CNV status in the TCGA-COAD dataset. Right, heatmap showing the PCC between
1091 the copy number of indicated CNV status and signature scores of each cell subpopulation in the
1092 TCGA-COAD dataset. Colors represent the PCC value. Blank squares indicate non-significant
1093 associations (PCC < 0.2). The color of CNV names represents associated clonalTIME
1094 phenotypes in the CRC4 section. **e**, Heatmap displaying the PCC between the copy numbers of
1095 different CNVs status within the TCGA-COAD dataset. Colors represent the PCC value.
1096 Statistical analysis was performed using a two-tailed unpaired *t*-test. **f**, Boxplot showing *CD8A*
1097 expression in samples with indicated CNV status in the TCGA-COAD dataset. **g**, Gene
1098 expression difference between clone3 and other clones in CRC4 section. Left, volcano plot
1099 displaying differentially expressed genes in clone3 versus other clones in CRC4 section.
1100 Differentially expressed genes were defined by adjusted p-value < 0.01 and $|\log_2(\text{fold change})|$
1101 ≥ 0.1. P-values were calculated using the Wilcoxon rank-sum test and adjusted by the
1102 Bonferroni correction method. Right, functional enrichment analysis of upregulated (n=615)

1103 and downregulated genes (n=1148) of clone3. **h**, Schematic representation of transposable
1104 element (TE) quantification in spatial transcriptomics. The resulting assignments are condensed
1105 into a matrix of read counts for each cell, versus each TE. Figure adapted from images created
1106 with BioRender.com. **i**, Volcano plot showing the differentially expressed TEs between clone3
1107 and other clones in CRC4 section. Only TEs previously identified as being specifically
1108 expressed in CRC tumors were considered in this study. Differentially expressed TEs of clone3
1109 were defined by adjusted p-value < 0.01 and a $|\log_2(\text{fold change})| \geq 0.25$. P-values were
1110 calculated using the Wilcoxon rank-sum test and adjusted by the Bonferroni correction method.
1111 **j**, *HERV-H* expression of each clone in CRC4 section. Left, spatial visualization of *HERV-H*
1112 expression in CRC4 section. Dashed lines mark the general contour of each tumor clone region.
1113 Right, violin plot displaying the *HERV-H* expression across 3 clones in CRC4 section.
1114 Statistical analysis was performed using a two-tailed unpaired *t*-test. **k**, Boxplot representing
1115 the *HERV-H* expression across normal tissue, chr20q-WT, and chr20q-gain tumor samples in
1116 the TCGA-COAD dataset. **l**, Boxplot displaying the distribution of GSVA activity of the
1117 indicated pathways (antigen processing and presentation, type-I IFN production, viral
1118 transcription) between chr20q-WT and chr20q-gain tumor samples in the TCGA-COAD dataset.
1119 **m**, Prognostic significance of chr20q-gain in ICB-treated CRC patients. Left, overall survival
1120 (Kaplan-Meier curves) of CRC patients post ICB in the MSK-IMPACT cohort stratified by the
1121 levels of chr20q CNV status. Right, overall survival (Kaplan-Meier curves) of CRC patients
1122 without ICB treatment in the TCGA-COAD dataset, stratified by the levels of chr20q CNV
1123 status. Log-Rank p-values are shown. Boxplot elements: center line, median; box limits, upper
1124 and lower quartiles; whiskers, 1%–99%. Statistical analysis was performed using a two-tailed
1125 unpaired *t*-test. *$0.01 \leq p < 0.05$ **$p < 0.01$ ***$p < 0.001$ ****$p < 0.0001$.

1126 **Reference**

1127 1 Brbic, M. *et al.* Annotation of spatially resolved single-cell data with STELLAR. *Nat Methods* **19**,
1128 1411-1418, doi:10.1038/s41592-022-01651-8 (2022).
1129 2 Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell
1130 and spatial multi-omics. *Nat Rev Genet* **24**, 494-515, doi:10.1038/s41576-023-00580-2 (2023).
1131 3 Stahl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial
1132 transcriptomics. *Science* **353**, 78-82, doi:10.1126/science.aaf2403 (2016).
1133 4 Junker, J. P. *et al.* Genome-wide RNA Tomography in the zebrafish embryo. *Cell* **159**, 662-675,
1134 doi:10.1016/j.cell.2014.09.038 (2014).
1135 5 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression
1136 at high spatial resolution. *Science* **363**, 1463-1467, doi:10.1126/science.aaw1219 (2019).
1137 6 Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods*
1138 **16**, 987-990, doi:10.1038/s41592-019-0548-y (2019).
1139 7 Liao, J., Lu, X., Shao, X., Zhu, L. & Fan, X. Uncovering an Organ's Molecular Architecture at
1140 Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends Biotechnol* **39**, 43-58,
1141 doi:10.1016/j.tibtech.2020.05.006 (2021).
1142 8 Rao, A., Barkley, D., Franca, G. S. & Yanai, I. Exploring tissue architecture using spatial
1143 transcriptomics. *Nature* **596**, 211-220, doi:10.1038/s41586-021-03634-9 (2021).
1144 9 Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression
1145 to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res*
1146 **49**, e50, doi:10.1093/nar/gkab043 (2021).
1147 10 Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat*
1148 *Biotechnol* **40**, 517-526, doi:10.1038/s41587-021-00830-w (2022).
1149 11 Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics.
1150 *Nat Biotechnol* **40**, 661-671, doi:10.1038/s41587-021-01139-4 (2022).
1151 12 Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat*
1152 *Biotechnol* **40**, 1349-1359, doi:10.1038/s41587-022-01273-7 (2022).
1153 13 Andersson, A. *et al.* Single-cell and spatial transcriptomics enables probabilistic inference of cell
1154 type topography. *Commun Biol* **3**, 565, doi:10.1038/s42003-020-01247-y (2020).
1155 14 Li, B. *et al.* Benchmarking spatial and single-cell transcriptomics integration methods for
1156 transcript distribution prediction and cell type deconvolution. *Nat Methods* **19**, 662-670,
1157 doi:10.1038/s41592-022-01480-9 (2022).
1158 15 Li, H. *et al.* A comprehensive benchmarking with practical guidelines for cellular deconvolution
1159 of spatial transcriptomics. *Nat Commun* **14**, 1548, doi:10.1038/s41467-023-37168-7 (2023).

1160 16 Sun, D., Liu, Z., Li, T., Wu, Q. & Wang, C. STRIDE: accurately decomposing and integrating
1161 spatial transcriptomics using single-cell RNA sequencing. *Nucleic Acids Res* **50**, e42,
1162 doi:10.1093/nar/gkac150 (2022).
1163 17 Kuppe, C. *et al.* Spatial multi-omic map of human myocardial infarction. *Nature* **608**, 766-777,
1164 doi:10.1038/s41586-022-05060-x (2022).
1165 18 Ackerman, J. E. *et al.* Defining the spatial-molecular map of fibrotic tendon healing and the
1166 drivers of Scleraxis-lineage cell fate and function. *Cell Rep* **41**, 111706,
1167 doi:10.1016/j.celrep.2022.111706 (2022).
1168 19 Anderson, A. C. *et al.* Spatial transcriptomics. *Cancer Cell* **40**, 895-900,
1169 doi:10.1016/j.ccell.2022.08.021 (2022).
1170 20 Foster, D. S. *et al.* Multiomic analysis reveals conservation of cancer-associated fibroblast
1171 phenotypes across species and tissue of origin. *Cancer Cell* **40**, 1392-1406 e1397,
1172 doi:10.1016/j.ccell.2022.09.015 (2022).
1173 21 Heiser, C. N. *et al.* Molecular cartography uncovers evolutionary and microenvironmental
1174 dynamics in sporadic colorectal tumors. *Cell* **186**, 5620-5637 e5616,
1175 doi:10.1016/j.cell.2023.11.006 (2023).
1176 22 Matusiak, M. *et al.* Spatially Segregated Macrophage Populations Predict Distinct Outcomes In
1177 Colon Cancer. *Cancer Discov*, doi:10.1158/2159-8290.CD-23-1300 (2024).
1178 23 Kersten, K. *et al.* Spatiotemporal co-dependency between macrophages and exhausted CD8(+)
1179 T cells in cancer. *Cancer Cell* **40**, 624-638 e629, doi:10.1016/j.ccell.2022.05.004 (2022).
1180 24 van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8(+) T cell states in human cancer:
1181 insights from single-cell analysis. *Nat Rev Cancer* **20**, 218-232, doi:10.1038/s41568-019-0235-
1182 4 (2020).
1183 25 Kirschenbaum, D. *et al.* Time-resolved single-cell transcriptomics defines immune trajectories
1184 in glioblastoma. *Cell* **187**, 149-165 e123, doi:10.1016/j.cell.2023.11.032 (2024).
1185 26 Benotmane, J. K. *et al.* High-sensitive spatially resolved T cell receptor sequencing with SPTCR-
1186 seq. *Nat Commun* **14**, 7432, doi:10.1038/s41467-023-43201-6 (2023).
1187 27 Liu, S. *et al.* Spatial maps of T cell receptors and transcriptomes reveal distinct immune niches
1188 and interactions in the adaptive immune response. *Immunity* **55**, 1940-1952 e1945,
1189 doi:10.1016/j.immuni.2022.09.002 (2022).
1190 28 Delfini, M., Stakenborg, N., Viola, M. F. & Boeckxstaens, G. Macrophages in the gut: Masters in
1191 multitasking. *Immunity* **55**, 1530-1548, doi:10.1016/j.immuni.2022.08.005 (2022).
1192 29 Xun, Z. *et al.* Reconstruction of the tumor spatial microenvironment along the malignant-
1193 boundary-nonmalignant axis. *Nat Commun* **14**, 933, doi:10.1038/s41467-023-36560-7 (2023).

1194 30 Zhang, D. *et al.* Inferring super-resolution tissue architecture by integrating spatial
1195 transcriptomics with histology. *Nat Biotechnol*, doi:10.1038/s41587-023-02019-9 (2024).
1196 31 Fu, T. *et al.* Spatial architecture of the immune microenvironment orchestrates tumor immunity
1197 and therapeutic response. *J Hematol Oncol* **14**, 98, doi:10.1186/s13045-021-01103-4 (2021).
1198 32 Eming, S. A., Martin, P. & Tomic-Canic, M. Wound repair and regeneration: mechanisms,
1199 signaling, and translation. *Sci Transl Med* **6**, 265sr266, doi:10.1126/scitranslmed.3009337
1200 (2014).
1201 33 Hu, K. H. *et al.* Transcriptional space-time mapping identifies concerted immune and stromal
1202 cell patterns and gene programs in wound healing and cancer. *Cell Stem Cell* **30**, 885-903 e810,
1203 doi:10.1016/j.stem.2023.05.001 (2023).
1204 34 Mascharak, S. *et al.* Multi-omic analysis reveals divergent molecular events in scarring and
1205 regenerative wound healing. *Cell Stem Cell* **29**, 315-327 e316, doi:10.1016/j.stem.2021.12.011
1206 (2022).
1207 35 McBrearty, B. A., Clark, L. D., Zhang, X. M., Blankenhorn, E. P. & Heber-Katz, E. Genetic
1208 analysis of a mammalian wound-healing trait. *Proc Natl Acad Sci U S A* **95**, 11792-11797,
1209 doi:10.1073/pnas.95.20.11792 (1998).
1210 36 Clark, L. D., Clark, R. K. & Heber-Katz, E. A new murine model for mammalian wound repair
1211 and regeneration. *Clin Immunol Immunopathol* **88**, 35-45, doi:10.1006/clin.1998.4519 (1998).
1212 37 Heydemann, A. The super super-healing MRL mouse strain. *Front Biol (Beijing)* **7**, 522-538,
1213 doi:10.1007/s11515-012-1192-4 (2012).
1214 38 Pastar, I. *et al.* Epithelialization in Wound Healing: A Comprehensive Review. *Adv Wound Care*
1215 *(New Rochelle)* **3**, 445-464, doi:10.1089/wound.2013.0473 (2014).
1216 39 desJardins-Park, H. E., Foster, D. S. & Longaker, M. T. Fibroblasts and wound healing: an
1217 update. *Regen Med* **13**, 491-495, doi:10.2217/rme-2018-0073 (2018).
1218 40 Wynn, T. A. & Vannella, K. M. Macrophages in Tissue Repair, Regeneration, and Fibrosis.
1219 *Immunity* **44**, 450-462, doi:10.1016/j.immuni.2016.02.015 (2016).
1220 41 Kim, M. H. *et al.* Dynamics of neutrophil infiltration during cutaneous wound healing and
1221 infection using fluorescence imaging. *J Invest Dermatol* **128**, 1812-1820,
1222 doi:10.1038/sj.jid.5701223 (2008).
1223 42 Phillipson, M. & Kubes, P. The Healing Power of Neutrophils. *Trends Immunol* **40**, 635-647,
1224 doi:10.1016/j.it.2019.05.001 (2019).
1225 43 Raziyeva, K. *et al.* Immunology of Acute and Chronic Wound Healing. *Biomolecules* **11**,
1226 doi:10.3390/biom11050700 (2021).
1227 44 Liu, S. *et al.* A tissue injury sensing and repair pathway distinct from host pathogen defense.

1228 *Cell* **186**, 2127-2143 e2122, doi:10.1016/j.cell.2023.03.031 (2023).
1229 45 Seferbekova, Z., Lomakin, A., Yates, L. R. & Gerstung, M. Spatial biology of cancer evolution.
1230 *Nat Rev Genet* **24**, 295-313, doi:10.1038/s41576-022-00553-x (2023).
1231 46 Erickson, A. *et al.* Spatially resolved clonal copy number alterations in benign and malignant
1232 tissue. *Nature* **608**, 360-367, doi:10.1038/s41586-022-05023-2 (2022).
1233 47 Lomakin, A. *et al.* Spatial genomics maps the structure, nature and evolution of cancer clones.
1234 *Nature* **611**, 594-602, doi:10.1038/s41586-022-05425-2 (2022).
1235 48 Qi, J. *et al.* Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+)
1236 macrophages in colorectal cancer. *Nat Commun* **13**, 1742, doi:10.1038/s41467-022-29366-6
1237 (2022).
1238 49 Trimm, E. & Red-Horse, K. Vascular endothelial cell development and diversity. *Nat Rev Cardiol*
1239 **20**, 197-210, doi:10.1038/s41569-022-00770-1 (2023).
1240 50 Li, G., Gao, J., Ding, P. & Gao, Y. The role of endothelial cell-pericyte interactions in
1241 vascularization and diseases. *J Adv Res*, doi:10.1016/j.jare.2024.01.016 (2024).
1242 51 Fridman, W. H. *et al.* B cells and tertiary lymphoid structures as determinants of tumour immune
1243 contexture and clinical outcome. *Nat Rev Clin Oncol* **19**, 441-457, doi:10.1038/s41571-022-
1244 00619-z (2022).
1245 52 Munoz-Erazo, L., Rhodes, J. L., Marion, V. C. & Kemp, R. A. Tertiary lymphoid structures in
1246 cancer - considerations for patient prognosis. *Cell Mol Immunol* **17**, 570-575,
1247 doi:10.1038/s41423-020-0457-0 (2020).
1248 53 Danenberg, E. *et al.* Breast tumor microenvironment structures are associated with genomic
1249 features and clinical outcome. *Nat Genet* **54**, 660-669, doi:10.1038/s41588-022-01041-y (2022).
1250 54 Dhainaut, M. *et al.* Spatial CRISPR genomics identifies regulators of the tumor
1251 microenvironment. *Cell* **185**, 1223-1239 e1220, doi:10.1016/j.cell.2022.02.015 (2022).
1252 55 Ramirez, C. F. A. *et al.* Cancer cell genetics shaping of the tumor microenvironment reveals
1253 myeloid cell-centric exploitable vulnerabilities in hepatocellular carcinoma. *Nat Commun* **15**,
1254 2581, doi:10.1038/s41467-024-46835-2 (2024).
1255 56 DeNardo, D. G. & Ruffell, B. Macrophages as regulators of tumour immunity and immunotherapy.
1256 *Nat Rev Immunol* **19**, 369-382, doi:10.1038/s41577-019-0127-6 (2019).
1257 57 Dong, L. *et al.* The loss of RNA N(6)-adenosine methyltransferase Mettl14 in tumor-associated
1258 macrophages promotes CD8(+) T cell dysfunction and tumor growth. *Cancer Cell* **39**, 945-957
1259 e910, doi:10.1016/j.ccell.2021.04.016 (2021).
1260 58 Cassetta, L. *et al.* Human Tumor-Associated Macrophage and Monocyte Transcriptional
1261 Landscapes Reveal Cancer-Specific Reprogramming, Biomarkers, and Therapeutic Targets.

1262 *Cancer Cell* **35**, 588-602 e510, doi:10.1016/j.ccell.2019.02.009 (2019).
1263 59 Han, D. *et al.* Anti-tumour immunity controlled through mRNA m(6)A methylation and YTHDF1
1264 in dendritic cells. *Nature* **566**, 270-274, doi:10.1038/s41586-019-0916-x (2019).
1265 60 Kloosterman, D. J. & Akkari, L. Macrophages at the interface of the co-evolving cancer
1266 ecosystem. *Cell* **186**, 1627-1651, doi:10.1016/j.cell.2023.02.020 (2023).
1267 61 Zhang, L. *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in
1268 Colon Cancer. *Cell* **181**, 442-459 e429, doi:10.1016/j.cell.2020.03.048 (2020).
1269 62 Li, N. *et al.* Mapping and modeling human colorectal carcinoma interactions with the tumor
1270 microenvironment. *Nat Commun* **14**, 7915, doi:10.1038/s41467-023-43746-6 (2023).
1271 63 Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses.
1272 *Nature* **434**, 772-777, doi:10.1038/nature03464 (2005).
1273 64 Liu, S. Y., Sanchez, D. J., Aliyari, R., Lu, S. & Cheng, G. Systematic identification of type I and
1274 type II interferon-induced antiviral factors. *Proc Natl Acad Sci U S A* **109**, 4239-4244,
1275 doi:10.1073/pnas.1114981109 (2012).
1276 65 Chiappinelli, K. B. *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer
1277 via dsRNA Including Endogenous Retroviruses. *Cell* **162**, 974-986,
1278 doi:10.1016/j.cell.2015.07.011 (2015).
1279 66 He, J. *et al.* Identifying transposable element expression dynamics and heterogeneity during
1280 development at the single-cell level with a processing pipeline scTE. *Nat Commun* **12**, 1456,
1281 doi:10.1038/s41467-021-21808-x (2021).
1282 67 Kong, Y. *et al.* Transposable element expression in tumors is associated with immune infiltration
1283 and increased antigenicity. *Nat Commun* **10**, 5228, doi:10.1038/s41467-019-13035-2 (2019).
1284 68 Sexton, C. E., Tillett, R. L. & Han, M. V. The essential but enigmatic regulatory role of HERVH
1285 in pluripotency. *Trends Genet* **38**, 12-21, doi:10.1016/j.tig.2021.07.007 (2022).
1286 69 Samstein, R. M. *et al.* Tumor mutational load predicts survival after immunotherapy across
1287 multiple cancer types. *Nat Genet* **51**, 202-206, doi:10.1038/s41588-018-0312-8 (2019).
1288 70 Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical
1289 sequencing of 10,000 patients. *Nat Med* **23**, 703-713, doi:10.1038/nm.4333 (2017).
1290 71 Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly
1291 reconstructs lineage branching. *Nat Methods* **13**, 845-848, doi:10.1038/nmeth.3971 (2016).
1292 72 MacLean, A. L., Hong, T. & Nie, Q. Exploring intermediate cell states through the lens of single
1293 cells. *Curr Opin Syst Biol* **9**, 32-41, doi:10.1016/j.coisb.2018.02.009 (2018).
1294 73 Zhou, P., Wang, S., Li, T. & Nie, Q. Dissecting transition cells from single-cell transcriptome data
1295 through multiscale stochastic dynamics. *Nat Commun* **12**, 5609, doi:10.1038/s41467-021-

1296 25548-w (2021).
1297 74 Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-
1298 seqV2. *Nat Biotechnol* **39**, 313-319, doi:10.1038/s41587-020-0739-1 (2021).
1299 75 Del Bigio, M. R. Ependymal cells: biology and pathology. *Acta Neuropathol* **119**, 55-73,
1300 doi:10.1007/s00401-009-0624-y (2010).
1301 76 Ramachandran, V. S. *Encyclopedia of the Human Brain*. (Elsevier, 2002).
1302 77 Meyer, G. Building a human cortex: the evolutionary differentiation of Cajal-Retzius cells and
1303 the cortical hem. *J Anat* **217**, 334-343, doi:10.1111/j.1469-7580.2010.01266.x (2010).
1304 78 Ptashkin, R. N. *et al.* Chromosome 20q Amplification Defines a Subtype of Microsatellite Stable,
1305 Left-Sided Colon Cancers with Wild-type RAS/RAF and Better Overall Survival. *Mol Cancer Res*
1306 **15**, 708-713, doi:10.1158/1541-7786.MCR-16-0352 (2017).
1307 79 Shah, N. M. *et al.* Pan-cancer analysis identifies tumor-specific antigens derived from
1308 transposable elements. *Nat Genet* **55**, 631-639, doi:10.1038/s41588-023-01349-3 (2023).
1309 80 Liang, Y., Qu, X., Shah, N. M. & Wang, T. Towards targeting transposable elements for cancer
1310 therapy. *Nat Rev Cancer* **24**, 123-140, doi:10.1038/s41568-023-00653-8 (2024).
1311 81 Dai, M., Pei, X. & Wang, X. J. Accurate and fast cell marker gene identification with COSG. *Brief*
1312 *Bioinform* **23**, doi:10.1093/bib/bbab579 (2022).
1313 82 Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity
1314 with single-cell RNA-Seq. *Elife* **8**, doi:10.7554/eLife.43803 (2019).
1315 83 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic
1316 data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420,
1317 doi:10.1038/nbt.4096 (2018).
1318 84 Satoh, J. & Tabunoki, H. A Comprehensive Profile of ChIP-Seq-Based STAT1 Target Genes
1319 Suggests the Complexity of STAT1-Mediated Gene Regulatory Mechanisms. *Gene Regul Syst*
1320 *Bio* **7**, 41-56, doi:10.4137/GRSB.S11433 (2013).
1321 85 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-
1322 level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).
1323 86 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological
1324 themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
1325 87 Mayer, J., Blomberg, J. & Seal, R. L. A revised nomenclature for transcribed human endogenous
1326 retroviral loci. *Mob DNA* **2**, 7, doi:10.1186/1759-8753-2-7 (2011).
1327 88 Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic
1328 properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61,
1329 doi:10.1016/j.cell.2014.12.033 (2015).

1330 89 Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray
1331 and RNA-seq data. *BMC Bioinformatics* **14**, 7, doi:10.1186/1471-2105-14-7 (2013).
1332 90 Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of
1333 focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41, doi:10.1186/gb-
1334 2011-12-4-r41 (2011).
1335 91 Reich, M. *et al.* GenePattern 2.0. *Nat Genet* **38**, 500-501, doi:10.1038/ng0506-500 (2006).
1336 92 Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse
1337 Brain. *Cell* **174**, 1015-1030 e1016, doi:10.1016/j.cell.2018.07.028 (2018).
1338