

METHOD

Open Access



SpottedPy quantifies relationships between spatial transcriptomic hotspots and uncovers environmental cues of epithelial-mesenchymal plasticity in breast cancer

Eloise Withnell¹ and Maria Secrier^{1*}

*Correspondence:
m.secrier@ucl.ac.uk

¹ Department of Genetics,
Evolution and Environment, UCL
Genetics Institute, University
College London, London WC1E
6BT, UK

Abstract

Spatial transcriptomics is revolutionizing the exploration of intratissue heterogeneity in cancer, yet capturing cellular niches and their spatial relationships remains challenging. We introduce SpottedPy, a Python package designed to identify tumor hotspots and map spatial interactions within the cancer ecosystem. Using SpottedPy, we examine epithelial-mesenchymal plasticity in breast cancer and highlight stable niches associated with angiogenic and hypoxic regions, shielded by CAFs and macrophages. Hybrid and mesenchymal hotspot distribution follows transformation gradients reflecting progressive immunosuppression. Our method offers flexibility to explore spatial relationships at different scales, from immediate neighbors to broader tissue modules, providing new insights into tumor microenvironment dynamics.

Keywords: Spatial transcriptomics, Epithelial-to-mesenchymal transition, Tumor microenvironment, Cell plasticity, Hotspot analysis

Background

The spatial organization of cancer cells and their interactions with immune and stromal cells in their environment are instrumental in dictating tumor behavior and progression [1]. Recent developments in sequencing technologies that are able to profile large areas of the entire tissue in minute detail, such as spatial transcriptomics [2, 3], are increasingly enabling us to gain a more comprehensive understanding of the complex ecosystem of tumor microenvironments (TME) [4, 5]. Profiling the cancer tissue spatially allows us to explore the tumor architecture and its heterogeneity in detail, paving the way to deciphering the crosstalk between tumor cells and their surroundings and opening up new therapeutic opportunities [6, 7].

Several studies to date have demonstrated the advantages of spatial transcriptomics in delineating major tissue domains with distinct cell composition [8], cancer hallmarks



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[9], immunosuppressive hubs [10, 11], entire tumor ecotypes with divergent clinical outcomes [12], or the impact of specific drugs on inhibiting tumor progression [13]. However, focusing on areas of the tissue that are relevant for a specific biological question and surveying relationships between cell populations at the right scale remains a challenge in these datasets. To delineate biologically meaningful tissue areas using spatial transcriptomics, some of the current analytical methods, such as SpaGCN [14] and BayesSpace [15] focus on unsupervised clustering of gene expression. Approaches like NeST [16] or GASTON [17] take this one step further and incorporate a nested structure or topography metrics to outline hierarchically organized co-expression hotspots aligning with tissue histology. Given that similar cells often cluster together [18, 19], methods that can reliably detect statistically significant spatial clusters are important in reinforcing the accuracy of cell states determined from continuous signatures. Cell-Charter [20] capitalizes on this concept and uses Gaussian mixture models to identify stable clusters representing spatial niches exhibiting distinct shapes and cell plasticity. More targeted clustering approaches employing user-defined signatures or cell types have recently been implemented in Voyager [21] and Monkeybread [22]. Such methods can play a significant role in enhancing the interpretation of cell types that are inferred through spatial transcriptomic deconvolution techniques, but flexibly exploring spatial units at different scales remains difficult.

When it comes to assessing the spatial proximity of different clusters, methods that infer this through co-enrichment within the immediate neighborhood have been implemented in packages like SquidPy [23]. However, there is a lack of methods that calculate a differential spatial relationship between cell types or signatures of interest, e.g., hypoxia. Additionally, current approaches lack analytical methods to define and compare shorter and longer-range interactions between specific areas or cell populations of interest. Given that the scale of certain biological processes in cancer, such as hypoxia, remains elusive, relying solely on conventional spot neighborhood-centric methods might obscure complex, spatial interactions. This is known as the modifiable areal unit problem (MAUP) in geostatistics [24], where spatial data patterns are observed to shift contingent on the size and shape of the spatial analysis units. While a growing number of methods address the need for multi-scale analysis [11, 23–26], the effect of changing spatial units has generally been underexplored in spatial biology [18].

Here, we build on the use of key ideas in the geostatistics field within spatial biology as previously demonstrated by Voyager [21] to devise an analytical method tailored to interrogate spatial relationships at various scales within 10x Visium transcriptomic datasets. Our approach defines areas densely inhabited by particular cell types or marked by user-defined gene signatures (hotspots) and areas depleted of cell types or gene signatures (coldspots) and statistically evaluates the proximity of such areas to other predefined hotspots or coldspots. We additionally compare the spatial relationships detected using the hotspot approach to those observed when looking only at the immediate neighborhood of individual spots. We assess how the relationships between these variables change when varying the hotspot size or the neighborhood size surrounding a spot. Importantly, we allow users to perform differential spatial analysis between two signatures or cell types of interest. For example, we can answer questions such as “which immune hotspots are significantly closer to mesenchymal hotspots compared to

epithelial hotspots?” We have implemented this method in the Python package SpottedPy, available at <https://github.com/secrierlab/SpottedPy>.

To highlight the potential of our method, we focus on a key process underlying cancer progression, the epithelial-to-mesenchymal transition (EMT). During EMT, polarized epithelial cells undergo multiple molecular changes and lose their identity to acquire a mesenchymal phenotype [27]. The interplay between EMT and the tumor microenvironment (TME) is multifaceted: while the TME is believed to be an inducer of EMT, mesenchymal tumor cells potentially influence the TME [28, 29]. This dynamic is further complicated by the nature of EMT, which is not merely a dichotomous event. Current research suggests that EMT is a spectrum, varying from a continuous gradient to distinct, discrete stages [30, 31]. Depending on the context, cells undergoing this transition can be locked in an EMT state, or alternate between a large landscape of EMT states, a phenomenon called epithelial-mesenchymal plasticity (EMP) [32]. We have previously shown that the tumor spatial organization follows EMP gradients and that hybrid and mesenchymal cells establish distinct interactions with the TME in small datasets of spatially profiled breast tumors [33]. However, the spatial organization and interactions of cells across this EMP spectrum within the tumor milieu remain largely undefined at a larger scale and could offer great potential in developing therapies that exploit cell intrinsic or microenvironmental vulnerabilities linked with this process. Here, we showcase the capability of SpottedPy to unveil new relationships between EMT hotspots and the TME in breast cancer spatial transcriptomics data. By facilitating multi-scale comparisons of tumor and TME relationships, we yield rigorous evidence of spatial dynamics and offer an interpretable and intuitive measure of interactions between tumor cells and their environment at flexible scales.

Results

SpottedPy: a tool to investigate biological modules and spatial relationships at different scales

From direct cell–cell interactions to immediate neighborhoods and even across larger modules, cancer cell evolution is impacted at different scales by its environment (Fig. 1a). However, exploring this landscape flexibly and determining the areas within the tissue where these effects are most prominent is not straightforward. While neighborhood enrichment is commonly used in the field [23, 25, 28], analyzing continuous expression signatures and examining how the size of the neighborhood influences spatial relationships are areas that remain underdeveloped. Furthermore, inspecting the immediate neighborhood of cells of interest versus broader hotspots within the tissue will yield different insights into the tissue architecture and organization, and this warrants further investigation.

We introduce SpottedPy, a Python package that enables a statistically principled interrogation of biological modules and cell–cell relationships at different scales in spatially profiled tissue (Fig. 1b).

This method encompasses:

- Neighborhood enrichment analysis: Our method introduces a function designed to examine correlations between cell states, populations, or processes within individual

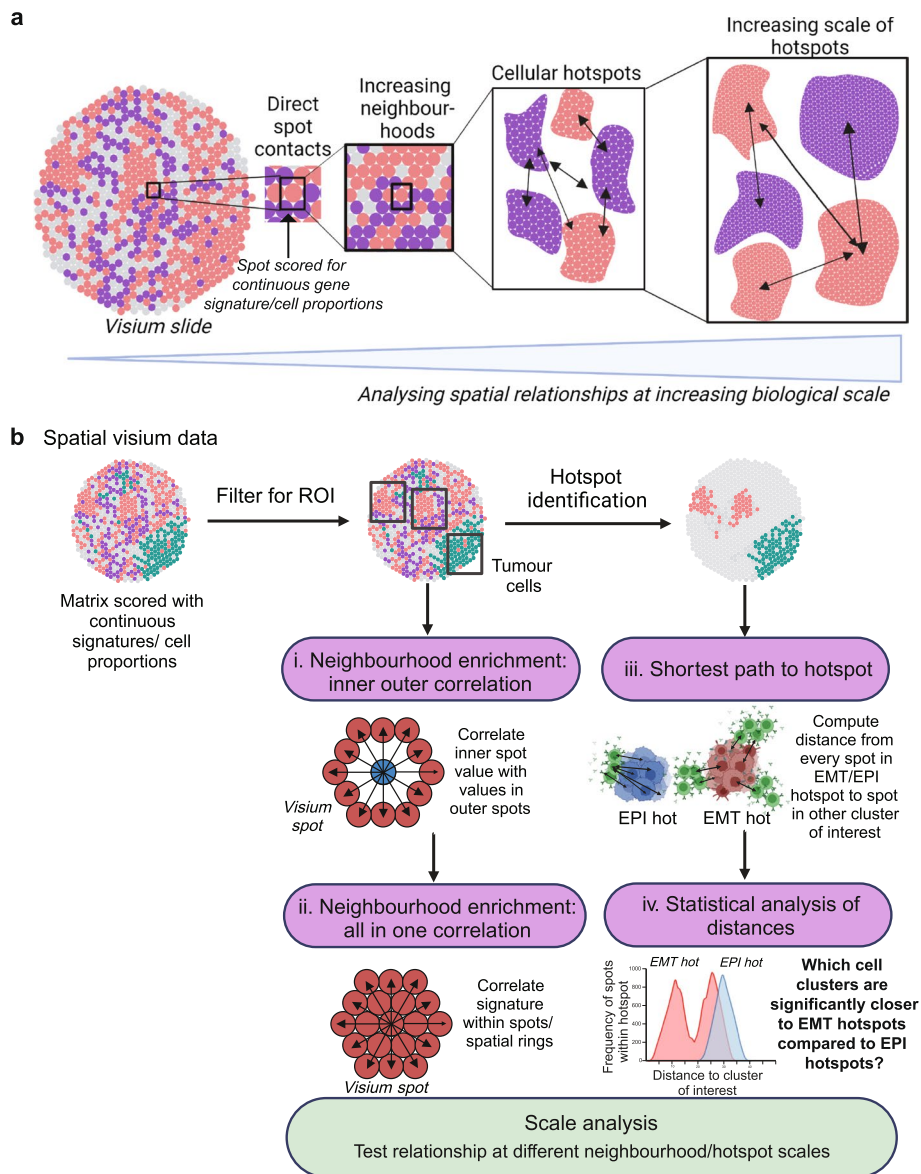


Fig. 1 SpottedPy provides a multi-scale approach to analyze spatial transcriptomic relationships. **a** Overview of spatial scales captured in the SpottedPy workflow, from direct cellular contacts to broader cellular hotspots. Figure created with BioRender.com. **b** SpottedPy workflow overview. Visium spatial transcriptomics data is loaded as a pre-processed *AnnData* object where there is the option to select the region of interest (ROI) within the slide, e.g., *AnnData.obs* column labeled with tumor cells. The default spatial analytics include the following: (i) neighborhood enrichment: inner-outer correlation, which correlates cell prevalence or signatures in individual spots with their immediate neighborhood; (ii) neighborhood enrichment: all-in-one correlation, which correlates cell prevalence of signatures within a spot or spatial unit; (iii) shortest path to hotspot, which calculates the minimum distance between each spot within a selected hotspot and the nearest spot in other hotspots; (iv) statistical analysis of distances, which compares distances from a reference hotspot to another hotspot of interest, and assesses the statistical significance of the relationships. Scale analysis allows us to compare relationships defined at different scales in both approaches, either by increasing the number of rings included for neighborhood enrichment or increasing the hotspot size. The outputs for the different modules include various plots to highlight the relationships. Figure created with BioRender.com

spatial transcriptomics spots and their immediate neighborhood (Fig. 1b (i–ii)). In this context, a “neighborhood” is delineated as a ring encompassing six Visium spots around a designated central spot, which we calculate by treating the spatial spots as a network. We offer functionality to test how a signature affects its direct neighborhood (inner-outer correlation), or within neighborhood units (all-in-one correlation).

- Hotspot identification: We have implemented the Getis-Ord G_i^* statistic to identify spatial clusters of continuous gene signatures across the tumor slide (Fig. 1b). Users can flexibly filter for specific regions within the slide to focus on when creating hotspots. By comparing defined regions with high or low gene expression or cell type signatures against a null hypothesis of random distribution, this analysis reveals statistically significant “hotspots” or “coldspots.” Hotspots represent areas with a high concentration of a particular cell type or signature, suggesting an aggregation that is unlikely to be due to chance. Conversely, coldspots indicate regions where the target cells or signatures are scarce, also beyond what would be expected in a random distribution. We provide functionality to test enrichment of specific signatures within hotspots and coldspots.
- Distance statistics: SpottedPy includes analytical tools to measure and interpret the distances between identified clusters, such as between tumors and immune hotspots. The primary method calculates the *shortest path to a hotspot*, defined as the minimum distance from any point within a defined hotspot to the nearest point within a specified comparison hotspot (Fig. 1b (iii)). Importantly, SpottedPy allows the user to compare distance distributions to key hotspots, for example, finding the hotspots that are significantly closer to mesenchymal hotspots than epithelial hotspots (or other areas that can be considered as a reference) (Fig. 1b (iv)). Importantly, SpottedPy assigns a statistical significance to these proximity measures to determine if observed patterns are likely to occur just by chance. To statistically analyze the relationships across multiple slides, we use generalized estimating equations. SpottedPy allows the user to test either the minimum, mean, or median distance from each hotspot or assess all distances from each spot within a hotspot.
- Scale/sensitivity analysis: We provide the ability to systematically evaluate how cell–cell relationships evolve within the tissue as we vary the size of the neighborhood or of the hotspot of interest. For the neighborhood enrichment approach, this can be assessed by varying the number of concentric rings around the central spot. For the hotspot approach, SpottedPy recalculates the Getis-Ord G_i^* statistic with varying neighborhood sizes, enabling the identification of clusters at different spatial scales. By analyzing how the distances between hotspots change with neighborhood size, the package can illuminate shifting spatial relationships, providing insights into how biological entities interact across different scales. Additionally, SpottedPy allows users to examine how cluster relationships change when modifying the significance threshold for identifying hotspots with the Getis-Ord G_i^* statistic.

To highlight the potential of our method, we employ SpottedPy to investigate the relationships between tumor cells undergoing EMT and the TME in 12 breast cancer slides profiled using the 10xGenomics Visium spatial transcriptomics platform, integrated

from Wu et al. [12], Barkley et al. [28] and the 10x Genomics website [34]. To infer individual cell types within the slides, we performed cell deconvolution using the Cell2location method [35] and a scRNA-seq reference of annotated breast cancer cell population profiles from 123,561 cells [12]. We scored the tumor cells in the scRNA-seq dataset with a defined epithelial (EPI) and an epithelial-to-mesenchymal transition (EMT) signature (see the “[Methods](#)” section) and used Gaussian mixture modeling to assign a state to the tumor cells [36, 37]. To more precisely identify tumor cells within the spatial transcriptomic data, which tend to show vast expression variability, we employed the copy number inference tool STARCH [38] and only kept spots that showed evidence of copy number changes, which are likely to be tumor-specific. We validated the STARCH results by comparing them to publicly available pathologist-annotated slides [39, 40] (Additional file 1: Fig. S1a–b). Furthermore, to explore the heterogeneity of EMT stable states established during the development and progression of breast cancer, we employed the discrete EMT states recently defined by Brown et al. [41], encompassing an epithelial phenotype, two intermediate (hybrid) states (EM2 and EM3), a late intermediate quasi-mesenchymal state (M1) and a fully mesenchymal state (M2).

The spatial landscape of EMT and associated tumor hallmarks

Firstly, we aimed to explore the cellular environment that is conducive to E/M progression. To do so, we focused solely on tumor hotspots and two key cellular hallmarks that have been previously linked with EMT, hypoxia, and angiogenesis. Hypoxia, characterized by low oxygen levels, has been long recognized as a key enabler of tumorigenic processes [42]. Under hypoxic conditions, tumor cells stabilize hypoxia-inducible factors (HIFs), primarily HIF-1 α , which promotes angiogenesis [43], the formation of new blood vessels from pre-existing vasculature, to re-establish oxygen supply. Hypoxia has been shown to induce EMT and resistance to therapy [44], and therefore understanding how such relationships develop spatially within the tissue can help devise localized therapies that can interrupt these interactions in breast cancer.

Within individual spatial transcriptomics slides, we used SpottedPy to delineate tumor areas, and further identified EMT hotspots within these areas using the EMT state as assigned using Cell2location (Fig. 2a). A certain degree of heterogeneity in the number and spatial distribution of EMT hotspots could be observed across the cohort (Additional file 1: Fig. S1c). Notably, slides 3, 7, and 10 exhibited the greatest dispersion, which was independent of the cancer subtype (TNBC and ER+HER2+). This dispersion did not influence overall EMT relationships with other cell types (Additional file 1: Fig. S2a). In contrast, slides 1 and 5 showed minimal dispersion, also without affecting EMT relationships with other cell types. The distributions of individual EMT and EPI signature scores per slide are highlighted in Additional file 1: Fig. S1c–d. The EPI signature typically followed a normal distribution across spatial tumor spots, whereas the EMT signature exhibited a positive skew, in line with this state being expected to be rarer within the primary tumors. Slides 0, 1, and 9 display more pronounced skewness for the EPI signature, suggesting a more advanced stage of EMT transformation within these tumors, which was not an effect of the cancer subtype as these slides present different breast cancer pathologies.

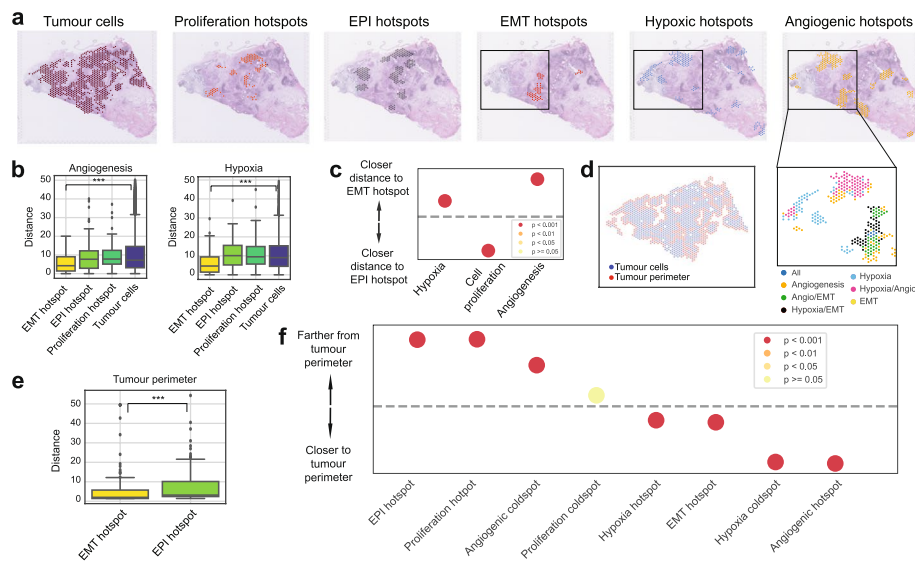


Fig. 2 The spatial interplay between EMT progression and cancer hallmarks. **a** A spatial transcriptomics slide (slide 0) highlighting from left to right: tumor spots, proliferation hotspots, EPI hotspots, EMT hotspots, hypoxic hotspots, and angiogenic hotspots identified by SpottedPy. The black square indicates a representative area where the close proximity of EMT, angiogenic, and hypoxic hotspots is depicted. **b** Distances from angiogenic (left) and hypoxic (right) hotspots to EMT hotspots, EPI hotspots, proliferative hotspots, and the average tumor cell, respectively, averaged across all 12 samples ($*** p < 0.001$). **c** Differences in proximity between EMT hotspots/EPI hotspots and hypoxic, proliferative, and angiogenic regions, summarized across the 12 slides. The dots situated above the dashed line indicate hallmarks that are significantly closer to EMT hotspots or EPI hotspots. The colors indicate the p -value ranges obtained from the Student's t -test for differences in distance to EMT hot/cold areas. **d** Spatial plot depicting the tumor perimeter in red and the tumor cells in blue. **e** Distance from the tumor perimeter to EMT hotspots and EPI hotspots, respectively ($*** p < 0.001$). **f** Distances from selected hotspots to the tumor perimeter, ordered by increasing proximity, across the 12 cases. The dashed line represents no significant difference. The colors depict p -value ranges obtained from Student's t -tests for differences in distance to the tumor perimeter

To confirm and further explore the emergence of additional cancer hallmarks in the context of EMT, we also defined proliferative, hypoxic, and angiogenic hotspots within the same slides (Fig. 2a). To check that SpottedPy accurately measures hotspot distance using the shortest path approach, we simulated a hypoxia hotspot moving away from a mesenchymal hotspot of interest (Additional file 1: Fig. S1e). As the hypoxia hotspot moves farther away from the mesenchymal hotspot the calculated distance between the hotspots increases, as expected.

When visually inspecting the slides, we find angiogenesis and hypoxia frequently accompanying EMT hotspots (Fig. 2a). When quantifying hotspot distances using SpottedPy, we confirm that EMT hotspots tend to be closer to angiogenic and hypoxic hotspots compared to EPI hotspots, proliferative hotspots, or the average tumor cell (Fig. 2b–c). In contrast, proliferative hotspots were significantly closer to EPI hotspots ($p < 0.001$, Fig. 2c). These relationships were consistently observed across breast cancer slides (Additional file 1: Fig. S2a).

To further grasp the positioning of these EMT and EPI areas within the tumor, we used SpottedPy to determine the tumor perimeter (Fig. 2d) and calculated distances to it. We conducted a visual benchmark of our tumor perimeter calculation against the Cottrazm method [40], confirming that our algorithm accurately captures similar perimeter trends

(Additional file 1: Fig. S2b). We find EMT hotspots closer to the tumor perimeter compared to EPI hotspots, suggesting a state with significant interaction with the surrounding microenvironment (Fig. 2e). As expected, angiogenesis hotspots were located closest to the tumor perimeter, followed by hypoxia hotspots (Fig. 2f). The spatial localization of angiogenesis near the tumor perimeter aligns with its function in supplying nutrients and oxygen to rapidly growing tumors [45]. The prominence of hypoxic regions succeeding angiogenic zones is consistent with the understanding that rapid tumor growth often outpaces its vascular supply, leading to hypoxic conditions [42]. These hypoxic conditions are alleviated in the angiogenic areas, as we find hypoxic coldspots are closest to the tumor perimeter (Fig. 2f). Cell proliferation hotspots were observed farthest from the tumor perimeter, and located at spatially distinct locations to EMT hotspots, in line with tumor growth studies outlining a proliferative epithelial core and EMT transformation at the periphery facilitating cancer cell intravasation and migration [46, 47].

EMT hotspots are immunosuppressed and shielded by myCAFs and macrophages

Having confirmed that SpottedPy is able to recapitulate expected spatial hallmarks of EMT within the breast cancer tissue, we next expanded our analysis to dissect the interplay between tumor cells undergoing EMT and other immune and stromal cells in the microenvironment. Alongside EMT hotspots, we calculated hotspots for 41 cell types in the TME, including different lymphocyte, myeloid, and fibroblast populations, as defined by Wu et al. [12] (Fig. 3a–c). When visually inspecting these hotspots, we observed myofibroblastic CAF (myCAF) hotspots and EMT hotspots tended to co-localize (Fig. 3a–c). Quantifying hotspot distances using SpottedPy allowed us to confirm that tumor EMT hotspots were indeed closer to myCAF hotspots (Fig. 3d). The relationship is particularly highlighted when we look at the cellular niches that are significantly closer to EMT hotspots compared to EPI hotspots, revealing a predominance of various CAF subtypes. This is well in line with existing studies, as myCAFs have been shown to produce TGF- β , which is a well-known EMT trigger [48]. They have also been linked to ECM deposition and suppression of antitumor immunity [49–52].

Additionally, monocytes and particularly tumor-associated macrophages (TAMs) (LAM2APOE+ macrophages and SIGLEC1+ macrophages) were prominently closer to EMT hotspots as compared to EPI hotspots. Monocytes, and TAMs derived from them, are known to modulate the environment of tumor cells undergoing EMT, usually by promoting immune suppression in the TME, which facilitates tumor progression and metastasis [53]. Natural Killer (NK), NK T-cells, and CD8+ T-cells, the immune cells that can directly kill transformed cells, were among the least associated with EMT hotspots, potentially reflecting a mechanism of immune evasion employed by tumor cells that have undergone EMT (Fig. 3f). The T-cell sub-population closest to EMT hotspots when compared to EPI were LAG3+ CD8+ T-cells, an exhausted population suggestive of immune evasion capacity in these EMT areas [54].

The cohort-level association between EMT hotspots and the myCAF s5 population was maintained in individual tumors, suggesting that this is a universal pattern of EMT transformation in breast cancer and not subtype-specific (Additional file 1: Fig. S2a). In contrast, levels of inter-patient heterogeneity, often even within the same breast cancer subtype, were evident for a variety of cells including macrophages, memory B-cells,

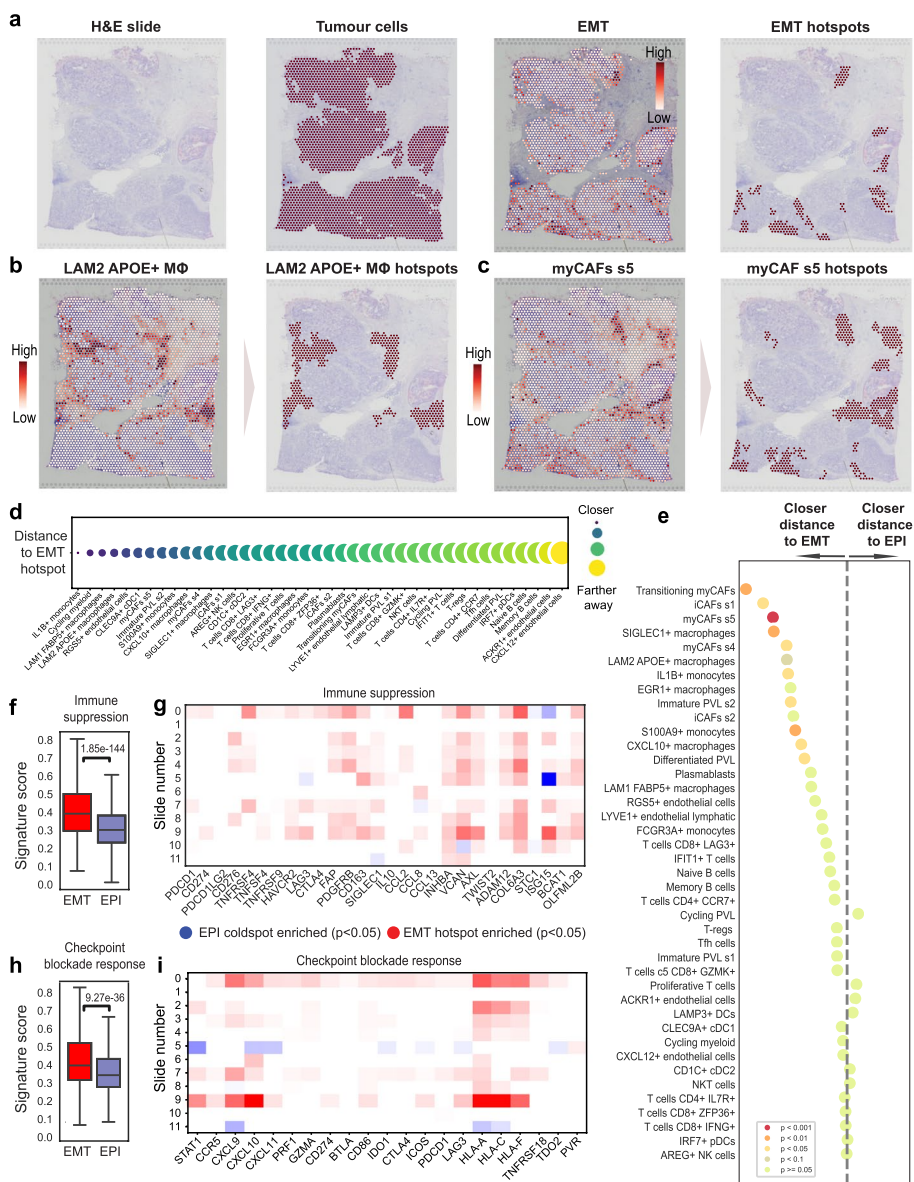


Fig. 3 The spatial interplay between EMT progression and the TME. **a** Spatial transcriptomics plots highlighting tumor cell spots (left), the EMT gradient through these tumor spots (middle), and EMT hotspots identified by SpottedPy (right) in slide 5. **b** Spatial localization of macrophage-enriched spots (left) and SpottedPy-defined LAM2 APOE+ macrophage hotspots (right) in slide 5. **c** Spatial localization of myCAF s5-enriched spots (left) and SpottedPy-defined myCAF hotspots in slide 5. **d** Distance between EMT hotspots and different TME cell hotspots, ranked by proximity. Smaller, darker bubbles represent shorter distances to EMT hotspots. Results are averaged over 12 slides. **e** Distances from various cells in the TME to EMT/EPI hotspots. The dashed line represents no difference in proximity to either EMT hotspots or EPI hotspots. The dots situated to the left of the dashed line indicate cell populations that are significantly closer to EMT hotspots, ordered by decreasing proximity. The colors indicate the p -value ranges obtained from the GEE fit for differences in distance to EMT hot/ EPI hot areas. Results are across 12 slides. **f** Barplots showing signature scores of immune suppression scored within EMT hotspots and EPI hotspots [12, 55]. **g** Differences in the average expression of genes in the immune suppression signature between EMT and EPI hotspots for each slide (row). Red depicts genes significantly upregulated in EMT hotspots and blue indicates genes significantly upregulated in EPI hotspots (Student's t -test $p < 0.05$, adjusted for multiple testing using the Bonferroni correction). White indicates a non-significant relationship [12, 55]. **h** Similar to **(f)** but for checkpoint inhibitor response [61, 62]. **i** Similar to **(g)** but focusing on the genes in the checkpoint inhibitor response signature

naïve B-cells, iCAFs, NK cells, NKT cells, CD4+ T-cells, and CD8+ T-cells. However, distances to EMT hotspots were consistent across subgroups of cells (Additional file 1: Fig. S2c), suggesting that, within individual patients, these cells share common response patterns irrespective of the broader heterogeneity observed across the patient cohort.

Due to the close relationship with potential immunosuppressive factors, we next sought to test whether EMT hotspots were indeed likely to be immunosuppressed. We found that EMT hotspots displayed a significantly increased expression of immunosuppression and exhaustion markers [12, 55] compared to EPI hotspots (Fig. 3f–g). Highly expressed suppressive genes included *FAP*, which has been shown to activate immune suppressive cells such as regulatory T-cells (Tregs) and myeloid-derived suppressor cells (MDSCs) [56, 57], *INHBA*, which fosters a switch in macrophage polarization towards a tumor-promoting state [58], *VCAN*, which has been shown to inhibit T-cell proliferation [59] and *COL6A3*, linked to the increased recruitment of macrophages [60]. Key immune checkpoints *B7-H3* (*CD276*), *OX40* (*TNFRSF4*) and *TIM3* (*HAVCR2*) also displayed significantly higher expression ($p < 0.05$) in some tumor slides (Fig. 3g). Indeed, EMT hotspots presented increased exhaustion (Additional file 1: Fig. S2d–e). Thus, it appears that the chronic nature of immune activation nearby EMT hotspots leads to the exhaustion of these cells, potentially suggesting opportunities for immune reactivation through checkpoint blockade strategies. To verify this hypothesis, we examined the expression of an interferon-gamma signature that has been associated with response to immunotherapy [61, 62] (Fig. 3h–i). We found that EMT hotspots had significantly increased expression of genes within the signature, most notably of *HLA-A*, and *HLA-C*, often associated with the activation of immune responses [63]. The hotspots also had increased expression of *HLA-F*, which has immune suppressive functions [64]. The expression of interferon-gamma-related genes, especially those involved in antigen presentation like HLA molecules, is a favorable prognostic marker in the context of checkpoint blockade therapy [65]. These findings suggest that while EMT hotspots are areas of significant immunosuppression and immune cell exhaustion, they also retain elements of immune activity that could be enhanced through targeted therapies such as checkpoint inhibitors.

EMT hotspots display intra- and inter-patient heterogeneity

We next sought to interrogate spatial relationships at a more granular level and analyzed the association of EMT hotspots with other immune and stromal areas within the same slide and across the different patient samples (Fig. 4a). While the cells that had the strongest relationship with EMT hotspots when averaged over the slides, such as SIGLEC+ and LAM2APOE+ macrophages and CAFs, displayed the most consistent trends across the slides, it was evident that these relationships were still heterogeneous. For example, in slide 4, while seven EMT hotspots were closer to LAM2APOE+ macrophages than the median EPI hotspots, two were not (Fig. 4a). This is further illustrated through visual inspection of the individual EMT hotspots and comparison with the LAM2APOE+ macrophage hotspots, enabled through SpottedPy functions (Fig. 4b). As expected, stronger associations with myCAFs and macrophage subtypes were pervasive in most tumors and hotspots (Fig. 4a). T-cells demonstrated noticeable heterogeneity across patients, but clustered together,

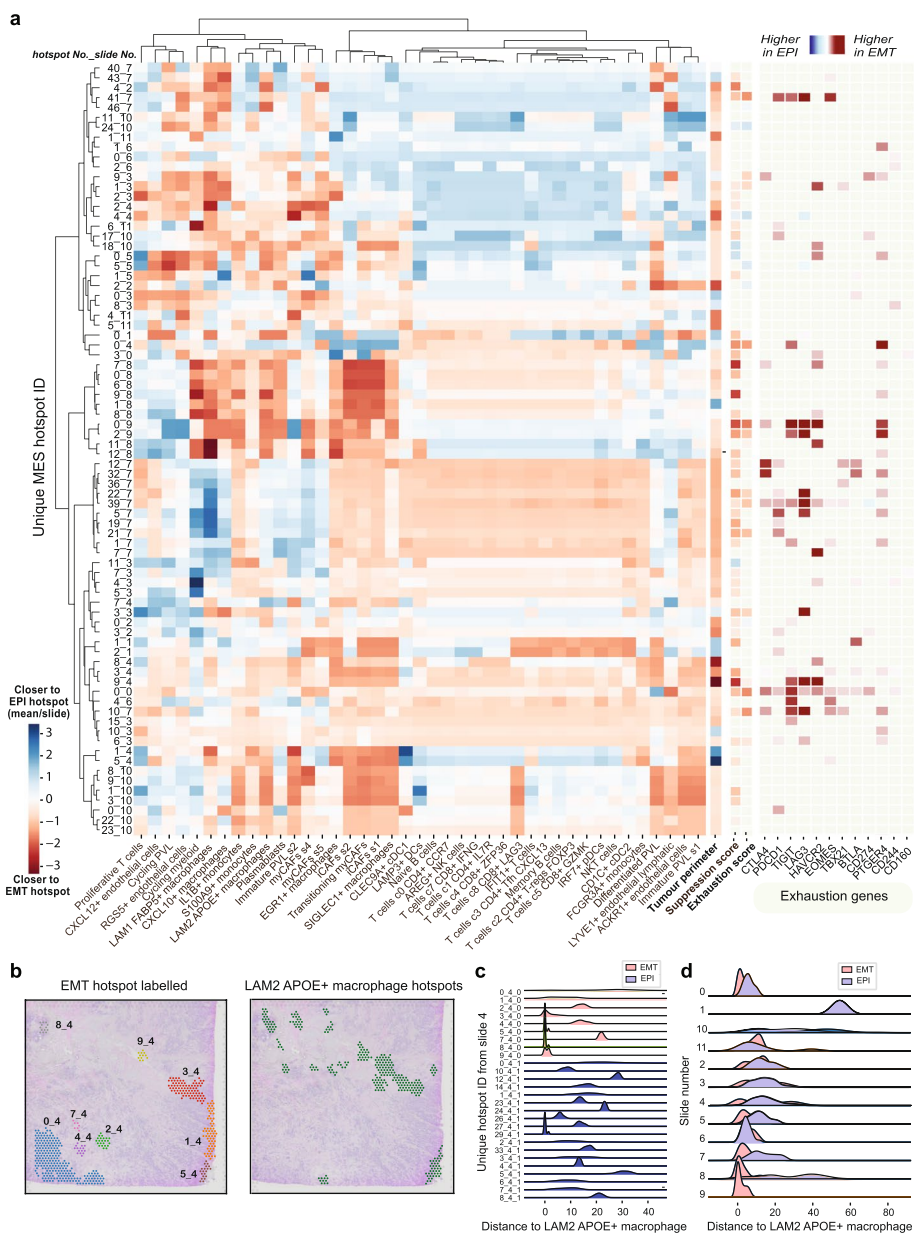


Fig. 4 Inter- and intratumor heterogeneity of EMT hotspots. **a** Dendrogram highlighting the proximity of EMT and EPI hotspots to TME cell types. The dendrogram is clustered according to the distances from EMT/EPI hotspots to regions enriched in immune/stromal cells. Red indicates that an EMT hotspot is closer to a cell type, while blue suggests that the EPI hotspots in that slide are on average closer. The x-axis displays individual EMT hotspots (label indicates hotspot number and slide number). To the right of the dendrogram, distances to the tumor perimeter, suppression, and exhaustion signature scores are illustrated. Red indicates that the hotspot is significantly enriched in these signatures compared to the average EPI hotspot in the slide ($p < 0.05$), and blue indicates EPI hotspots are significantly enriched ($p < 0.05$). Further to the right, individual genes associated with the exhaustion signature are shown, with red indicating the gene expression is higher in EMT hotspots ($p < 0.05$) and blue indicating the gene expression is higher in EPI hotspots ($p < 0.05$). **b** Slide 4 with individual EMT hotspots labeled (left) and LAM2 APOE + macrophage hotspots highlighted (right). **c** Distance distributions for each EMT hotspot in slide 4 to LAM2 APOE + macrophage hotspots. **d** Distance distributions of EMT hotspots to LAM2 APOE + macrophages across all 12 slides in the cohort

reinforcing the idea that these cells share common response patterns. The EMT hotspots that were closer to T-cells were more likely to be enriched in exhaustion markers (Fig. 4a, right panel), suggestive of chronic immune activation. We also show that EMT hotspots show a consistent trend of displaying higher suppressive scores than EPI hotspots (Fig. 4a). Additionally, nearly all EMT hotspots were closer to the tumor perimeter (Fig. 4a), corroborating the overall trends reported in Fig. 2e–f. Overall, the inter-patient heterogeneity seemed to supersede the intra-patient heterogeneity.

SpottedPy offers functionality to inspect the distance distributions for hotspots within a slide (Fig. 4c), and across each slide (Fig. 4d). Visualizing these distributions highlights that while LAM2 APOE + macrophages are on average closer to EMT hotspots compared to EPI hotspots, there is heterogeneity within each slide and across different slides.

Overall, these results showcase the range of hotspot analyses enabled by the SpottedPy package and the potential to uncover useful biological insights.

Sensitivity analysis of hotspots

Determining how the hotspot size, governed by the parameter specifying the number of nearest neighbors for the Getis-Ord G_i^* metric, affects spatial relationships is crucial for meaningful spatial analysis. We systematically increased hotspot dimensions (Fig. 5a–b) to assess the consistency and robustness of identified spatial associations. We find that associations between EMT hotspots, hypoxia, and angiogenesis, as well as mutual exclusivity with proliferative hotspots, are robust and consistent features of the tumor microenvironment, with such relationships remaining remarkably stable across a range of hotspot dimensions (Fig. 5c). The cell populations that we previously identified as having the nearest proximity to EMT hotspots at a fixed parameter size (myCAFs, macrophages and monocytes) also maintained this relationship when varying hotspot sizes. Cells that were farther apart presented less stable associations, such as CD8 + LAG3 + T-cells where the relationship broke down at a hotspot size of 250, and naive B-cells where the relationship changed multiple times with increasing hotspot size (Fig. 5c, Additional file 1: Fig. S3a). These findings suggest that interactions with certain cells in the TME may be more pronounced and relevant at a smaller scale. We found that proliferative hotspots were the most consistently adjacent to EPI hotspots at various hotspot sizes. Adjusting the p -value cut-off used to detect spatial clusters using the Getis-Ord G_i^* highlighted similar relationships (Additional file 1: Fig. S3b). To evaluate the stability of the spatial relationships, we introduced Gaussian noise and spot reshuffling into the dataset and examined the persistence of these relationships (see the “[Methods](#)” section). This approach demonstrated that the method is robust to low levels of noise, but also effectively discriminates between biologically meaningful signals from those arising from random spatial distributions (Fig. 5e–f). While random noise simulated through spot reshuffling can mimic some aspects of structured data (Additional file 1: Fig. S3c), the hotspots are significantly smaller than those that are biologically relevant (Additional file 1: Fig. S3d). Crucially, the loss of specific associations among particular cell types when noise is introduced (Fig. 5e–f) contributes to the reduction of false positives even if hotspots are identified [66].

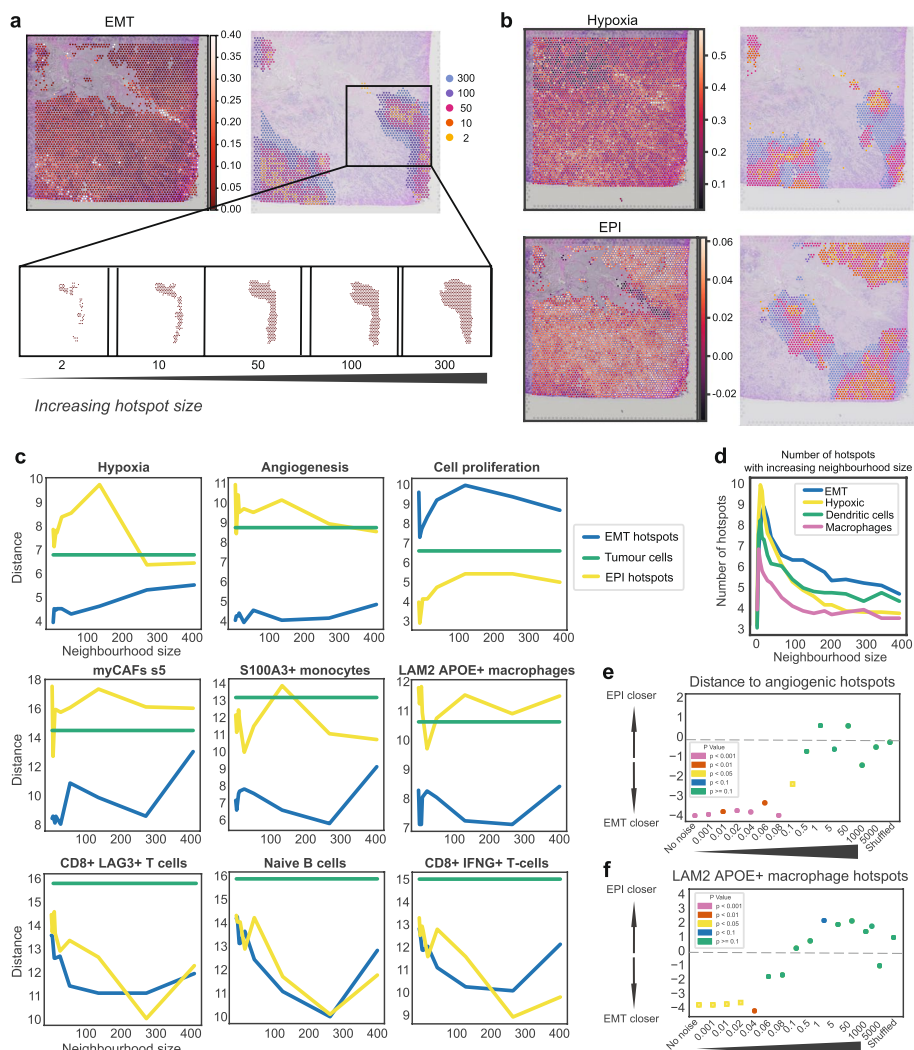


Fig. 5 Sensitivity analysis of hotspot relationships. **a** EMT hotspot generation using a hotspot neighborhood parameter of 2, 10, 50, 100, and 300, respectively. Increasingly larger neighborhoods are highlighted in different colors as indicated in the legend. **b** Hypoxia and epithelial hotspot generation using a hotspot neighborhood parameter of 2, 10, 50, 100, and 300, respectively. **c** Sensitivity plots highlighting the distance from EMT hotspots (blue) and EPI hotspots (yellow) to regions enriched in various cancer hallmarks and TME components as the hotspot size increases. The distances to the average of all tumor cells are used as a reference (green). Distances are averaged over all 12 slides. **d** Number of distinct hotspots identified as the hotspot neighborhood parameter is increased, averaged over 12 slides. **e** Evaluating the impact of increased noise or spot shuffling on the association between angiogenic hotspots and EMT/EPI hotspots. **f** Evaluating the impact of increased noise or spot shuffling on the association between LAM2 APOE+ macrophage hotspots and EMT/EPI hotspots

Other distance metrics

We note that there are alternative methods to assess hotspot distances. The “centroid to centroid” methodology offers a practical and straightforward way to approximate the distances between hotspots, but it is crucial to acknowledge its simplicity. As depicted in Additional file 1: Fig. S3e, the size of a hotspot significantly influences its centroid location. Consequently, a larger hotspot, despite being physically closer, may appear farther away when measuring centroid distances, due to its centroid being located farther from

the point of interest compared to a smaller, neighboring hotspot. Thus, the centroid approach may miss the local variation that the shortest path method can capture. This could suggest that applying it on our breast cancer slides could potentially miss the more complex relationship observed between EMT hotspots and macrophage or monocyte-enriched areas.

Spatial EMT relationships in other cancer types

We further investigated whether the relationships observed for EMT hotspots in breast cancer were consistent across other cancer types. We assessed these relationships in publicly available datasets from basal cell carcinomas (BCC) [67], pancreatic ductal adenocarcinomas (PDAC) [68], and colorectal cancers (CRC) [69].

Within the BCC slides, angiogenic and hypoxic hotspots were closer to EMT hotspots (Additional file 1: Fig. S4a–b). Interestingly, proliferative hotspots were also closer to EMT hotspots, suggesting an alternative relationship compared to breast cancer. POSTN+ fibroblasts were closer to EMT hotspots, while there were no significant spatial relationships with T-cells or NK cells, paralleling the findings in breast cancer.

We next assessed the relationships within one available PDAC slide (Additional file 1: Fig. S4c). Both angiogenesis and fibroblasts displayed a spatial relationship like that observed with EMT hotspots in breast cancer. In contrast, immune cells were located nearer to EMT coldspots, and there was no significant association between EMT hotspots and hypoxic environments, diverging from the patterns observed in breast cancer.

In CRC, myofibroblasts, angiogenesis, and hypoxia showed comparable spatial relationships to those seen in breast cancer (Additional file 1: Fig. S4d–e). Notably, regulatory T-cells, T-helper cells, and NK cells were significantly closer to EMT hotspots, potentially indicating enhanced immune recognition around EMT hotspots compared to other cancer types.

While limited in breadth and cell type resolution, these analyses suggest that the interplay between tumor cells undergoing EMT and other immune and stromal cells within the TME is likely to be tissue-specific, and future work should explore this in more detail.

Neighborhood enrichment analysis

The neighborhood enrichment technique captures more localized, shorter-range relationships with the TME. Additionally, it can assess spatial relationships of phenotypes that would be considered scattered (states that do not occur spatially clustered and therefore might be overlooked by a hotspot-based approach). We experimented with two approaches, ensuring a robust analysis that is less sensitive to the MAUP (Fig. 1b (i–ii)). We first assessed how the spatial relationships change by correlating phenotypes across a central tumor spot and the direct neighborhood surrounding it (a ring encompassing six Visium spots). We then assessed how the phenotypes are linked within a spot and then expanding what is considered a spatial spot. Varying the method and the number of rings in both cases enables us to assess whether the observed hotspot relationships shift with the unit of analysis and indicates how large of an influence the EMT regions have on surrounding spots.

Our analysis revealed that angiogenesis, myCAFs, macrophages, and monocytes exhibited the most significant correlation, in descending order, with cells undergoing EMT ($p < 0.001$) across the 12 slides (Additional file 1: Fig. S5a). This finding reinforces the spatial relationships that we previously identified using the hotspot method. A weaker association was evident for naive B-cells, T-cells, NK cells, and NKT cells, in accordance with the hotspot approach. We also found that these spatial relationships were stable across various neighborhood sizes (Additional file 1: Fig S5b).

The methods show broadly similar trends, suggesting the cellular relationships observed occur both due to colocalization in a spot as well as diffusing influence around the spot.

EMT state fluctuations shape distinct immune niches within the same tumor

As mentioned previously, EMT is not a binary process—instead, cells are found to occupy multiple hybrid states during the E/M transition. We sought to investigate the spatial distribution of tumor hotspots occupying epithelial (EPI), early intermediate (EM2, EM3), late intermediate quasi-mesenchymal (M1), and fully mesenchymal (M2) states using our multi-scale approach. We captured distinct gene programs representing these states using Non-Negative Matrix Factorisation (NMF) via the CoGAPs workflow [70] (Additional file 1: Fig. S6a). The corresponding hotspots occupied distinct spatial locations within the tissue (Additional file 1: Fig. S6b). When visually inspecting these hotspots, we detected a progressive transformation in the tumor, as highlighted in slide 4 (Fig. 6a). This transition was marked by a spatial shift from EPI into the M1 state, with EM3 serving as an intermediate stage. EM2 displayed volatility in this progression, while M2 was predominantly co-localized with the EPI state. The experimental study by Brown et al. [41] detected that M2 cell clones gained integrin $\beta 4$ (a key epithelial marker) when cultured, which might have played a significant role in steering these cells towards adopting characteristics more akin to an epithelial phenotype. This would possibly explain the co-localization of these two states within the spatial transcriptomics slide.

To investigate the relationship between these states, we correlated them with each other and with the generic EMT hallmark signature employed previously for each tumor spot (Additional file 1: Fig. S7a). The lack of a significant positive correlation between EPI, EM2, EM3, M1, and M2 suggests that these are discrete EMT states. We found a significant correlation between the EMT hallmark signature and the quasi-mesenchymal M1 state, suggesting these are likely capturing a similar state. The EPI state was negatively correlated with the EMT hallmark signature, as expected. In terms of spatial distribution, the EMT hallmark hotspots were located closest to the M1 hotspots, and furthest away from the EPI hotspots (Additional file 1: Fig. S7b), in line with the correlation analysis and confirming the hypothesized identities of these states.

When investigating how tumor cells occupying distinct EMT states relate to their microenvironment, we found that the EPI state has negative correlations with immune and stromal cells within the TME, suggestive of a state that is not directly being shaped by the TME (Fig. 6b). Interestingly, the M1 state had stronger associations overall with a wide range of cells within the TME, with the strongest relationships established with myCAFs, macrophages, and monocytes. We observed similar but weaker correlations

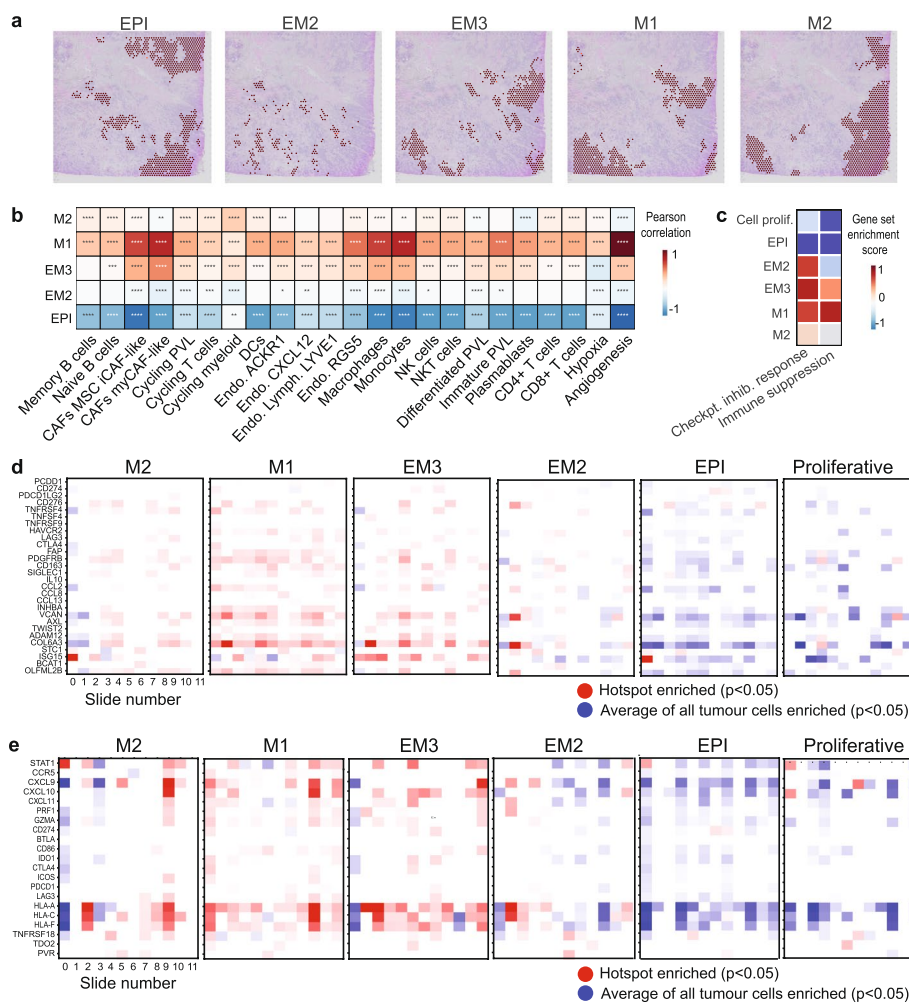


Fig. 6 EMT state dynamics uncovered from the spatial exploration of breast cancer tissue. **a** Spatial plots depicting epithelial (EPI), intermediate (EM2, EM3), quasi-mesenchymal (M1), and fully mesenchymal (M2) hotspots in slide 4. **b** Neighborhood enrichment analysis depicting the association between tumor cells occupying distinct EMT states and other cells in the immediate TME, summarized across all 12 slides. Red indicates a significant positive correlation (Pearson, $p < 0.05$), blue a significant negative correlation ($p < 0.05$), and white a non-significant correlation ($p > 0.05$). **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **c** Scaled immune suppression [12] and immunotherapy response signature [65] scores calculated using Gene Set Enrichment Analysis (GSEA) for each EMT state hotspot and proliferative hotspot, summarized across the 12 samples. **d** Enrichment and depletion of expression for genes in the immune suppression signature within EMT state hotspots for each slide (column). Red depicts genes significantly upregulated in EMT state hotspots compared to the average of all tumor cells and blue represents genes significantly downregulated in the EMT state hotspot (Student's t -test $p < 0.05$). White indicates a non-significant relationship. P -values were adjusted for multiple testing using the Bonferroni correction. **e** Similar to **(d)**, focusing on genes in the checkpoint inhibitor response signature

with the EM3 state and considerably weaker correlations with the EM2 state. The progressive loss of association with cells in the EM3 and EM2 states is in line with the idea of these states representing intermediate, more plastic states preceding the apparently more stable M1 state. The M1 state is observed in proximity to natural killer (NK) cells, which is a distinct deviation from the EMT hallmark signature. This observation suggests that while there are resemblances between the M1 and EMT hallmark signatures,

the M1 state is likely representative of a unique cellular phenotype with the potential to recruit cells capable of directly eliminating cancer cells.

Furthermore, the quasi-mesenchymal M1 state presented an enrichment of markers linked with immunosuppression and positive response to checkpoint inhibitors (Fig. 6c, Additional file 1: Fig. S7c), with a subset of genes, most notably *COL6A3*, *VCAN*, *HLA-A*, *HLA-C*, *CXCL9* and *CXCL10*, driving these relationships (Fig. 6d–e). The intermediate states are to a certain extent on the way to adopting this immune suppressive phenotype, with weaker relationships observed with the EM2 state, a slightly stronger enrichment with EM3, and the strongest score with M1 (Fig. 6c–e, Additional file 1: Fig. S7d). In contrast, the M2 state has a unique phenotype and displays both positive and negative relationships with genes within these signatures. We compared these states to the proliferative signature and found that proliferative hotspots mirror the relationship of the EPI state, suggesting that it is a tumor state that is not linked to immunosuppression.

Overall, this analysis sheds light on the changing landscape of tumor-TME interactions during EMT progression in breast cancer, highlighting both intratumor heterogeneity and universal interactions that could be exploited for therapy.

Discussion

In this study, we introduce SpottedPy, a Python package that identifies tumor hotspots in spatial transcriptomics slides and explores their interplay with the TME at varying scales. We show that the Getis-Ord G_i^* statistic can be successfully applied to delineate cellular hotspots and provide meaningful biological insights into the spatial organization of the tumor tissue in its immune and stromal contexture. While various studies have recently applied “hotspot”-type of analysis to spatial transcriptomic data [21, 71, 72], these methods do not offer a way to assess the confidence level in the identification of specific clusters/hotspots in a manner that can be tailored to the biological question and the scale at which the process is expected to act, whereas our method assigns a p -value to each hotspot which can be flexibly tuned to according to the user’s stringency requirements. Furthermore, other available methods do not extensively analyze the distances between hotspots. We build on these approaches to analyze the spatial relationships between hotspots in a statistically principled manner, with the additional ability of anchoring hotspot identification to specific regions of interest, such as tumor cells or non-tumor cells, enabling us to interrogate the spatial dynamics within the TME via a more targeted approach. By computing and statistically comparing distances, we offer an interpretable and intuitive measure of relationship between spatial variables. This approach further allows differential spatial analysis between a hotspot of interest and a reference region which other methods do not include. Our method additionally assesses the effect of hotspot size on spatial relationships and compares hotspot spatial trends to the relationships captured using neighborhood approaches, more frequently applied in spatial transcriptomic analysis, to build layers of spatial evidence.

By adopting our SpottedPy methodology to explore the spatial dynamics of tumor plasticity phenotypes in breast cancer, we have uncovered key differences between tumor regions undergoing EMT and those lacking evidence for this transformation. Our approach illuminates the pronounced spatial correlations of EMT with key cancer hallmarks, notably hypoxia and angiogenesis, in line with findings by He et al. [11] in breast

cancer spatial transcriptomics detecting these signatures overlapping certain niches. As tumor cells undergo EMT in response to hypoxic stimuli, they are likely to gain a survival advantage in a nutrient-deprived environment and be better equipped to invade and migrate towards regions with better oxygenation, potentially following angiogenic gradients [44, 47, 73].

We find a strong relationship between EMT and CAFs across all slides. CAFs have been linked to tumor cells undergoing EMT in our previous bulk and spatial transcriptomic analyses [33], and have been shown to induce EMT in endometrial cancer cells [74] and hepatocellular carcinoma [75]. It is worth noting that CAFs share similar genes with EMT signatures and therefore differentiating between these two cell types can be challenging, notably in bulk tumor settings [76]. Here, we have leveraged scRNA-seq for deconvolution, alongside detecting copy number aberrations to confirm the presence of tumor cells, which adds a further layer of confidence to the accurate delineation of these cell populations. However, the accuracy of this separation cannot be fully guaranteed and future research investigating this relationship using single-cell resolved spatial transcriptomic data would allow us to confirm this relationship more confidently. We further note that while we have tried to ensure EMT is captured only within the tumor cells themselves by investigating signatures only within the tumor spots identified by STARCH and by using EMT reference datasets from pure tumor populations, further validation of tumor regions, e.g., by staining with specific markers, would be beneficial for users who wish to employ our method in their spatial transcriptomics experiments.

In addition to expected CAF associations, we observe a strong relationship with macrophages and monocytes across multiple spatial scales. We particularly observed relationships with SIGLEC+ macrophages, LAM2APOE + macrophages, and EGR1 + macrophages, which are analogous to M2-like, tumor-promoting, macrophages [12]. Macrophages secrete TGF- β , TNF- α , IL-6 and IL-8, which are well-characterized EMT stimuli [77, 78]. The relationship has been observed in bulk transcriptomics [79], in spatial analysis of mouse models of skin carcinoma where depletion of macrophages inhibited EMT progression [30], and in specific niches within breast cancer spatial transcriptomics slides [11]. While the relationships with CAFs and macrophages were displayed across the majority of tumor slides, within each slide there were EMT hotspots where this relationship was less clear. This indicates that other factors which we have not accounted for in our analyses could play a significant role in driving EMT within local niches.

We uncovered heterogeneous relationships across breast tissue slides with other key immune cells such as NK cells, NKT cells, and T-cells, highlighting the multifaceted interplay between these components. T-cells have been shown to induce EMT in breast cancer [80, 81], and this relationship has further been highlighted in bulk transcriptomics [82] and smaller-scale spatial transcriptomic analyses [33, 83] however there is also evidence showing the exclusion of these cells, linked to the relationship between EMT, macrophages and CAFs promoting an immune suppressed environment [52, 77, 84]. Indeed, our findings uncover notable associations between EMT hotspots and immune suppression, alongside signatures indicative of a response to checkpoint therapy, building upon evidence that EMT may offer crucial insights for existing strategies in immunotherapy [11, 85].

Our analysis reveals that EMT occurs in discrete spatial locations distinct from proliferative signatures. This finding is in line with a recent analysis of breast cancer by Barkley et al. [28] utilizing a more focused spatial transcriptomic dataset, previous research by Tsai et al. [86] demonstrating that a departure from a mesenchymal-like state is a prerequisite for tumor cell proliferation in mouse models, and a recent study by Chen et al. [87] investigating EMT states in scRNA-seq data. Such spatial characterizations at various scales were largely unexplored.

Delving deeper into EMT, we observed that hybrid EMT states exhibit more heterogeneous and weaker associations with tumor-promoting populations in the TME in comparison to the quasi-mesenchymal M1 state. This disparity might be indicative of the inherent plasticity of hybrid EMT states [88, 89], complicating our ability to delineate clear relationships, but might also suggest a directed trajectory towards an M1 state. The M2 state demonstrates more similar distribution and TME associations to the EPI state, which may be attributed to the activation of integrin $\beta 4$ (a key epithelial marker) when cultured, a limitation mentioned in the original study which potentially transformed the state towards a more epithelial phenotype [41].

These results point to a highly dynamic and plastic nature of tumor cells in navigating the complexities of their microenvironment. The interactions likely extend beyond a linear framework. Hypoxia, a known catalyst for both angiogenesis and EMT [90], can set off a cascade of events that not only amplify these processes but also create a conducive environment for the recruitment of immunosuppressive cells such as macrophages [91, 92]. These cells can in turn bolster angiogenesis, thereby fuelling a self-perpetuating cycle that further complicates the tumor landscape. These insights provide a further understanding of the cellular interactions and environmental factors that underpin tumor progression and metastasis, and could in the future pave the way for the development of targeted interventions aimed at disrupting these complex networks for therapeutic benefit.

The consistency of the relationships we observed across different hotspot sizes and neighborhood scales further strengthens our confidence in the findings. The neighborhood ring approach predominantly captured TME cells that have infiltrated the tumor, offering insights into the immediate cellular interplay at the tumor periphery. In contrast, the hotspot methodology provided a broader view, encompassing interactions at more distal locations. By pinpointing statistically significant cellular hotspots, we bolster the reliability of our observations, especially considering the inherent inaccuracies that can arise from deconvolution algorithms applied to non-single-cell transcriptomic datasets such as those from the Visium platform.

To the best of our knowledge, there are no direct comparisons available for our SpottedPy methodology due to the unique nature of focusing on discrete spatial clusters of user-defined continuous signatures at expanding scales and performing differential spatial relationships compared to a reference for downstream analysis.

Our insights into the spatial organization of tumors during EMT progression are limited by the significant amount of uncertainty surrounding EMP programs [31, 32] and their incomplete characterization in different types of breast cancer and other neoplasms. In the future, integrating further hybrid states characterized in other breast cancer studies [87] will help expand our understanding of this complex process alongside

its multiple locally stable peaks and valleys. The heterogeneous relationships observed with CD8+ and CD4+ T-cells and NK cells require further experimental validation and exploration. As spatial transcriptomics datasets become more widely available, expanding this analysis beyond the current 12 slides could help clarify the perceived spatial heterogeneity and better distinguish universal relationships from local, patient-specific effects.

When investigating to what extent these spatial EMT relationships are maintained or differ across cancer types other than breast cancer, we were limited by the availability and size of such datasets, as well as the differences in cell composition between tissues. Ultimately, any uncovered differences are likely attributed to the unique TME and genetic basis of each cancer type, and in the future a more in-depth analysis in larger datasets once these become more widely available will shed light on the heterogeneity of these relationships. Additionally, extending these analyses to single cell-resolved spatial datasets and incorporating ligand-receptor signaling information into the evaluation of spatial effects on cell populations will increase the confidence in the identified relationships.

Overall, our findings confirm the expected spatial effects of EMT progression in tumors, demonstrating that SpottedPy can capture complex associations between tumor cells and their microenvironment. Such insights can help unveil the local effects of the TME and linked tumor cell vulnerabilities that could ultimately be exploited for therapeutic benefit. While the analyses presented here primarily illustrate insights into breast cancer tissue organization, we note that SpottedPy can be applied to discern spatial relationships in other cancer types (as briefly demonstrated) as well as other diseases and even within healthy tissue. SpottedPy has been developed on spatial transcriptomics data from the 10x Visium platform; however, we note it can be easily extended to other spatially-resolved platforms and future releases will provide further functionality to enable this.

Conclusions

In conclusion, SpottedPy provides a detailed and multifaceted analysis of the spatial dynamics within individual spatially profiled tumors. By rigorously investigating the proximities of various cellular components, we have underscored the significant influence exerted by cells undergoing EMT in sculpting the TME and highlight SpottedPy as a package that can be applied to answer other spatial biology questions.

Methods

The SpottedPy package

SpottedPy package is compatible with Python 3.9 and depends on scanpy, libpysal, and esda packages.

- GitHub: <https://github.com/secrierlab/SpottedPy>
- Tutorials: `spottedpy_multiple_slides.ipynb` (this tutorial walks through using SpottedPy with multiple spatial slides, highly recommended for downstream statistical analysis). `spottedpy_tutorial_sample_dataset.ipynb` tutorial walks through using SpottedPy with a single slide.

- Sample data: <https://zenodo.org/records/10392317>

The key functions are outlined in the relevant sections below.

Spatial transcriptomic datasets

Breast cancer Visium slides were obtained from Barkley et al. [28] (slides 0–2), from 10xGenomics (slides 3–5) [34] and Wu et al. [12] (slides 6–12). Slide annotations, if available, are displayed in Additional file 1: Fig. S1b. We combined the three datasets of breast cancer 10X Genomics Visium spatial transcriptomic datasets into a common *anndata* Python format for analysis. Pre-processing and normalization were conducted using the ScanPy (Single-Cell Analysis in Python) package [93]. We analyzed a total of 32,845 spatially profiled spots, and retained spots if they exhibited at least 100 genes with at least 1 count in a cell, had more than 250 counts per spot, and less than 20% of total counts for a cell which are mitochondrial. Pre-processed BCC slides were obtained from Gania et al. [67], PDAC slides were obtained from Ma et al. [68] and CRC slides were obtained from Valdeolivas et al. [69]. We used the deconvolution results provided in each of the source studies.

Spatial data deconvolution

Due to the imperfect near-single cell resolution of current spatial transcriptomic methods, we require a method to deconvolve each spot in order to infer the cellular populations enriched in each spot. Cellular deconvolution was carried out using Cell2location [35]. Cell2location decomposes the spatial count matrix into a predefined set of reference cell signatures by modeling the spatial matrix as a negative binomial distribution, given an unobserved gene expression level rate and gene- and batch-specific over-dispersion. A scRNA-seq breast cancer dataset containing 100,064 cells from 26 patients and 21 cell types from Wu et al. [12] was chosen to perform the deconvolution. Cell types in the chosen breast dataset consisted of cancer epithelial cells (basal, cycling, Her2, LumA, LumB), naïve and memory B-cells, myCAF-like and iCAF-like cancer-associated fibroblasts, perivascular-like cells (PVL), including immature, cycling and differentiated, cycling T-cells, cycling myeloid cells, dendritic cells (DCs), endothelial cells expressing *ACKR1*, *CXCL12* or *RGS5*, endothelial lymphatic *LYVE1*-expressing cells, luminal progenitors and mature luminal cells, macrophages, monocytes, myoepithelial cells, natural killer (NK) cells, natural killer T (NKT) cells, plasmablasts, CD4 + T-cells, and CD8 + T-cells. We scored the scRNA-seq cancer epithelial cells with EPI and EMT signatures [36, 37] and used Gaussian mixture modeling to assign the cells to EPI and EMT clusters.

The scRNA regression model was trained with 500 epochs, and the spatial transcriptomic model was trained with 20,000 epochs using a GPU. To delineate the tumor cells within our spatial transcriptomics dataset, we used the STARCH Python package designed to infer copy number alterations (CNAs) [38]. STARCH identifies tumor clones (setting $K=2$ clones) and non-tumor spots. It confirms the identification of normal spots by clustering the first principal component into two clusters using K-means. Changing the value of K alters the number of identified tumor clones, but the number of cells labeled as tumor cells remains the same. This approach is based on the principle

that the direction of maximum variance in the expression data typically reflects the division between non-cancerous and cancerous spots. Only tumor cell spots were considered for EMT analysis. The EPI and EMT spots identified using Cell2location were used to define the EPI and EMT hotspots in the breast cancer downstream analysis.

EMT state and hallmark signature scoring

To identify more granular distinct EMT states, we employed data from Brown et al. [41], consisting of seven RNA-seq sequenced cell clones, derived from SUM149PT inflammatory breast cancer cell line with 3 repeats spanning the EMT spectrum from epithelial-like (EPI), quasi-mesenchymal (M1), fully mesenchymal (M2) and three distinct intermediates (EM1, EM2, EM3). We used these data to derive a weighted gene signature to represent the EMT states. We captured EMT gene patterns from this data using non-negative matrix factorization (NMF) by applying the CoGAPs workflow [70]. We used ProjectR's implementation of *Imfit* R function to map the captured EMT patterns onto the spatial transcriptomic spots [94]. This transfer learning approach assumes that if datasets share common latent spaces, a feature mapping exists between them and can measure the extent of relationships between the datasets. The final states were captured with 20 patterns and 10,000 training iterations. The number of patterns was chosen based on capturing the discrete states with the highest accuracy. The EM1 state was not distinguishable from the EPI state, so we merged the two states. Thus, overall we obtained scores for one epithelial, two intermediate, a quasi-mesenchymal, and a fully mesenchymal state for each spot.

Hypoxia and angiogenesis were defined based on signatures deposited at MSigDB [95]. The proliferative signature was compiled from Nielsen et al. [96]. The immunosuppression signature was compiled from Wu et al. [12] and Cui et al. [55]. The checkpoint blockade response signature was compiled from Johnson et al. [61] and Liu et al. [62]. The exhaustion signature comprised classical exhaustion markers: *CTLA4*, *PDCD1*, *TIGIT*, *LAG3*, *HAVCR2*, *EOEMT*, *TBX21*, *BTLA*, *CD274*, *PTGER4*, *CD244*, and *CD160* [97]. All these signatures were scored using *scanpy.tl.score_genes* function. EMT hotspots and coldspots were identified in the BCC, CRC, and PDAC slides using the EMT hallmark signature [95].

Graph construction

The SquidPy [23] (Spatial Single-Cell Analysis in Python) package was used for graph construction using *sq.gr.spatial_neighbors* and slide visualization of the Visium spatial slides. NetworkX [98] was used for further analysis of the networks derived from the spatial transcriptomic spots. The deconvolved spot results were used to assign node labels. Edges were assigned based on the spot neighbors.

Neighborhood enrichment analysis

We calculated neighbors for each spot by summing the deconvolution results in a ring surrounding the spot of interest and normalizing by the number of spots assigned as a neighbor, using the adjacency matrix of the graph to calculate the interacting cells.

Two methods were developed to assess neighborhood enrichment. *Inner-outer* correlation (with the function *sp.calculate_inner_outer_correlations*) was calculated by

correlating signatures across a central spot of interest and the direct neighborhood of spots surrounding it (a ring encompassing six Visium spots), after filtering for tumor spots only. To perform the sensitivity analysis, we increased the number of rings surrounding a spatial transcriptomic spot (setting *rings_range* parameter in *sp.calculate_inner_outer_correlations* function) to consider as spot neighbors and compared the change in the correlation coefficient. The first ring consists of 6 spots, and the second ring includes 18 spots (combined from the 1st and 2nd rings). Subsequent rings follow this pattern. The number of rings selected for sensitivity analysis reflects a balance between spatial coverage and resolution. Using a smaller number of rings (e.g., 1, 2, 3) allows the analysis to focus on the immediate microenvironment around the central spot, providing high resolution. As more rings are added, the spatial coverage increases, capturing broader interactions but potentially diluting local-specific signals. Correlations were calculated using Pearson's correlation coefficient.

An *all-in-one* correlation (*sp.calculate_neighbourhood_correlation* function) was calculated by correlating phenotypes with cells within a spot, and then incrementally increasing the number of rings to correlate across progressively larger spatial units. The functions *sp.correlation_heatmap_neighbourhood* and *sp.plot_overall_change* plot the neighborhood results.

Hotspot analysis

Hotspots were calculated using The Getis-Ord G_i^* statistic as implemented using the PySAL package [99], using 10 as the neighborhood size parameter by default (number of spot neighbors surrounding the central Visium spot) and a p -value cut-off of 0.05, unless otherwise stated.

The Getis-Ord G_i^* equation is defined as follows:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}x_j - \bar{x} \sum_{j=1}^n w_{ij}}{s \sqrt{\frac{n \sum_{j=1}^n w_{ij}^2 - \left(\sum_{j=1}^n w_{ij}\right)^2}{n-1}}}$$

where w_{ij} is the spatial weight between location i and j , \bar{x} is the mean of the variable of interest across all locations, s is the standard deviation of the variable of interest across all locations and n is the total number of locations.

A high positive value at location i suggests a hotspot for the attribute, while a negative value indicates a coldspot. The significance of G^* is determined by comparing the observed G_{obs}^* to a distribution of G^* values generated under the assumption of spatial randomness. This distribution is obtained by permuting the attribute values across locations and recalculating G^* for each permutation. The p -value for a hotspot (when G^* is positive) or a coldspot (when G^* is negative) is then derived from this distribution. This approach provides a non-parametric method to evaluate the significance of spatial clusters, offering a robust measure against potential spatial randomness in the data.

Hotspots can be identified by calling *sp.create_hotspots* function, and specifying in the *filter_columns* parameter what region within the spatial slide to calculate the hotspot from, e.g., tumor cells. The *neighbourhood_parameter* can be altered here (default = 10). We encourage the user to choose the parameter most relevant for their biological

question, e.g., whether they are interested in local interactions of the signature, or broader tissue modules. SpottedPy allows the user to perform the sensitivity analysis to observe how the parameters affect downstream analysis. For the 10x Visium platform, we would recommend starting with parameter $k=10$ as this captures all the spots surrounding the central spot. The variable with the most stable relationships across a range of parameters (and therefore scales) is likely one of most interest for further investigation. However, specific short-range relationships defined locally rather than across scales could also be of interest in certain circumstances depending on the user's biological questions. Coldspots are automatically created when `sp.create_hotspots` is called, and hotspots are labeled in the `anndata` object by appending “_hot”, and coldspots by appending “_cold” to the original column name. When an appropriate contrasting signature is available for comparison, e.g., EPI compared to EMT we do not need to use the coldspots for comparison. The `relative_to_batch` parameter ensures hotspots are calculated across each slide, otherwise they are calculated across multiple slides. Importantly, if multiple slides are used (highly recommended for statistical power), these should be labeled using `.obs['batch']` within the `anndata` object. Additionally, the library ID in the `.uns` data slot should be labeled with the `.obs['batch']` value. Hotspots can be plotted using `sp.plot_hotspots`.

Hotspots and coldspots for EMT states and cell proliferation were calculated after filtering for tumor cells as labeled by STARCH, as we aimed to specifically capture these processes within the tumor cells themselves. EMT hotspots are the regions with a high proportion of mesenchymal tumor cells within the tumor-labeled spots. Therefore, they only include a subset of tumor spots. Similarly, cell proliferation hotspots are regions with high fractions of proliferating tumor cells. All other hotspots (deconvolved cell proportion data and angiogenic and hypoxia signatures) were calculated using all the spots within the spatial transcriptomic slide.

Distance metrics

After calculating the hotspots and coldspots, we then assessed the distances from hotspots of interest (EPI and EMT) to other cell types and signature hotspots and coldspots. We used the *shortest path* approach to calculate distances between hotspots as follows:

Let H represent the set of coordinates of spots in the hypoxia hotspot.

Let M represent the set of coordinates of spots in the mesenchymal tumor hotspot.

Let E represent the set of coordinates of spots in the epithelial tumor hotspot.

For a spot m in M and a spot e in E , the shortest path to any point h in H was determined:

$$d_{\min}(m, H) = \min_{h \in H}(d(m, h))$$

$$d_{\min}(e, H) = \min_{h \in H}(d(e, h))$$

where $d(m, h)$ represents the Euclidean distances from a spot m in M . After obtaining the minimum distances for each spot in M and E we calculated the median (with

the additional functionality to choose min or mean) to provide a summary statistic that reflects the general proximity of each hotspot (M and E) to H . The function *sp.calculateDistances* calculates this.

To then infer the impact of cellular hotspots on distance to EMT compared to EPI hotspots, we employed generalized estimating equations (GEE). This model enables us to estimate population-average effects involving repeated measurements across multiple spatial transcriptomic slides. The model estimates the coefficient (β_{mes}) for the transition from reference hotspots (E) to primary hotspot variables (M). A positive β_{mes} would indicate that mesenchymal hotspots are, on average, located further from hypoxic areas compared to epithelial hotspots, while a negative value suggests a closer proximity. *sp.plot_custom_scatter*, setting *compare_distance_metric* to *min*, *mean* or *median* to compare the summary statistics for each hotspot across each slide. Setting it to *None* calculates the statistical significance of all distances from each hotspot.

The *centroid approach* is calculated as follows. The centroid C_H of a set of spots H with coordinates x_h, y_h is the arithmetic mean of the coordinates. This point represents the center of the mass of the points in the set H .

For set H :

$$C_H = \left(\frac{\sum_{h \in H} x_h}{|H|}, \frac{\sum_{h \in H} y_h}{|H|} \right)$$

Similar calculations are employed for M and E . We then calculated the Euclidian distance between the centroids.

Tumor perimeter calculation

Any spot was considered part of the tumor perimeter if it had more than one neighboring spot (nodes in the graph) that were not classified as tumor spots. A spot $s \in S$ is considered part of the tumor perimeter, P , if:

$$s \in P \Leftrightarrow |N(s) \cap (S \setminus T)| > 1$$

where S denotes the set of all spots, T denotes the set of tumor spots, and $N(s)$ represents the neighboring spots of spot s . Additionally, we applied a filtering step to remove isolated perimeter spots. This involved eliminating any identified perimeter spots that had no neighboring perimeter spots, thereby excluding isolated perimeter spots caused by a single non-tumor labeled spot within the tumor. This approach helped us to delineate the boundary of the tumor accurately by focusing on the transitional area where tumor and non-tumor spots meet (called using *sp.calculate_tumour_perimeter*).

To quantify the number of tumor hotspots, we calculated the number of connected components within the graph that were labeled as hotspots. This calculation was crucial for understanding the distribution and clustering of tumor cells.

Sensitivity analysis

The sensitivity analysis to evaluate the impact of varying hotspot sizes on the spatial relationships was achieved by incrementally adjusting the neighborhood parameter for the Getis-Ord statistic, which directly influenced the size of identified hotspots

(*sp.sensitivity_calcs*). As we expanded the neighborhood parameter, we compared the distances between the newly defined hotspots and other existing hotspots of interest.

To assess the robustness of the spatial relationships between cell types and gene signatures, we systematically introduced Gaussian noise into our cell type proportion data and gene signature matrix. Gaussian noise, characterized by a mean of zero and varying standard deviations, was added to mimic experimental and technical variability. This approach allows us to evaluate the stability of detected EMT hotspots under different noise conditions. We defined a range of sigma values to represent varying levels of noise intensity. To further test the robustness of the spatial relationships, we randomly shuffled the cell proportion data and gene signature values and assessed how this affected downstream analysis.

Statistical analysis

Groups were compared using a two-sided Student's *t*-test. Multiple testing correction was performed where appropriate using the Bonferroni method. Graphs were generated using the seaborn and Matplotlib Python packages.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-024-03428-y>.

Additional file 1: Supplementary figures S1-S7 supporting the analyses in this study.

Additional file 2: Review history.

Authors' contributions

MS designed the study and supervised the analyses. EW developed the SpottedPy package and performed all the analyses.

Authors' X handles

X handles: @mariasecrier (Maria Secrier)

Review history

The review history is available as Additional File 2.

Peer review information

Veronique van den Berghe was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Funding

EW was supported by a studentship award from the Health Data Research UK-The Alan Turing Institute Wellcome PhD Programme in Health Data Science (218529/Z/19/Z). MS was supported by a UKRI Future Leaders Fellowship (MR/T042184/1). Work in MS's lab was supported by a BBSRC equipment grant (BB/R01356X/1) and a Wellcome Institutional Strategic Support Fund (204841/Z/16/Z).

Data availability

Breast cancer Visium slides were obtained from Barkley et al. (GEO accession number: GSE203612) [28, 100], Wu et al. (<https://doi.org/10.5281/zenodo.4739738>) [12, 101] and the 10xGenomics website [34]. The breast cancer scRNA-seq reference dataset used for deconvolution was obtained from Wu et al. (GEO accession number: GSE176078) [102]. The RNA-seq data used to define the five EMT states (EPI-M2) was obtained from Brown et al. [41, 103] (GEO accession number: GSE172613). SpottedPy is implemented as a Python package available at <https://github.com/secrierlab/SpottedPy> [104], released under a GNU General Public License v3.0 with an accompanying tutorial. Processed spatial transcriptomics data and the SpottedPy code have been deposited at Zenodo: <https://doi.org/10.5281/zenodo.13907274> [105], released under a GNU General Public License v3.0.

Declarations

Ethics approval and consent to participate

All datasets employed in this study are publicly available and comply with ethical regulations, with approval and informed consent for collection and sharing already obtained by the relevant organizations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 15 December 2023 Accepted: 24 October 2024

Published online: 11 November 2024

References

1. Yuan Y. Spatial heterogeneity in the tumor microenvironment. *Cold Spring Harb Perspect Med*. 2016;6(8):a026583.
2. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. 2016;353(6294):78–82.
3. Asp M, Bergenstråhle J, Lundeberg J. Spatially resolved transcriptomes—next generation tools for tissue exploration. *BioEssays*. 2020;42(10):1900221.
4. Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods*. 2022;19(5):534–46.
5. Marx V. Method of the Year: spatially resolved transcriptomics. *Nat Methods*. 2021;18(1):9–14.
6. Zhang L, Chen D, Song D, Liu X, Zhang Y, Xu X, et al. Clinical and translational values of spatial transcriptomics. *Sig Transduct Target Ther*. 2022;7(1):1–17.
7. Andersson A, Larsson L, Stenbeck L, Salmén F, Ehinger A, Wu SZ, et al. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat Commun*. 2021;12(1):6012.
8. Kumar T, Nee K, Wei R, He S, Nguyen QH, Bai S, et al. A spatially resolved single-cell genomic atlas of the adult human breast. *Nature*. 2023;620(7972):181–91.
9. Sibai M, Cervilla S, Grases D, Musulen E, Lazcano R, Mo CK, et al. The spatial landscape of Cancer Hallmarks reveals patterns of tumor ecology. *bioRxiv*. 2023;2022.06.18.496114.
10. Ji AL, Rubin AJ, Thrane K, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*. 2020;182(2):497–514.e22.
11. He S, Jin Y, Nazaret A, Shi L, Chen X, Rampersaud S, et al. Starfish reveals heterogeneous spatial dynamics in the breast tumor microenvironment. *bioRxiv*; 2022. p. 2022.11.21.517420. Available from: <https://www.biorxiv.org/content/10.1101/2022.11.21.517420v1>. Cited 2023 Mar 16.
12. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet*. 2021;53(9):1334–47.
13. Lengrand J, Pastushenko I, Vanuytven S, Song Y, Venet D, Sarate RM, et al. Pharmacological targeting of netrin-1 inhibits EMT in cancer. *Nature*. 2023;620(7973):402–8.
14. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods*. 2021;18(11):1342–51.
15. Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol*. 2021;39(11):1375–84.
16. Walker BL, Nie Q. NeST: nested hierarchical structure identification in spatial transcriptomic data. *Nat Commun*. 2023;14(1):6554.
17. Chitra U, Arnold BJ, Sarkar H, Ma C, Lopez-Darwin S, Sanno K, et al. Mapping the topography of spatial gene expression with interpretable deep learning. 2023. Preprint at: <https://doi.org/10.1101/2023.10.10.561757>.
18. Palla G, Fischer DS, Regev A, Theis FJ. Spatial components of molecular tissue biology. *Nat Biotechnol*. 2022;40(3):308–18.
19. Seferbekova Z, Lomakin A, Yates LR, Gerstung M. Spatial biology of cancer evolution. *Nat Rev Genet*. 2023;24(5):295–313.
20. Varrone M, Tavernari D, Santamaria-Martínez A, Walsh LA, Ciriello G. Cell Charter reveals spatial cell niches associated with tissue remodeling and cell plasticity. *Nat Genet*. 2023;8:1–11.
21. Moses L, Einarsson PH, Jackson K, Luebbert L, Boeshaghi AS, Antonsson S, et al. Voyager: exploratory single-cell genomics data analysis with geospatial statistics. *bioRxiv*. 2023. p. 2023.07.20.549945. Available from: <https://www.biorxiv.org/content/10.1101/2023.07.20.549945v2>. Cited 2023 Oct 23.
22. Bernstein MN, Scott D, Hession CC, Nieuwenhuis T, Gerritsen J, Tabrizi S, et al. Monkeybread: A Python toolkit for the analysis of cellular niches in single-cell resolution spatial transcriptomics data. *bioRxiv*. 2023. p. 2023.09.14.557736. Available from: <https://www.biorxiv.org/content/10.1101/2023.09.14.557736v1>. Cited 2023 Oct 23.
23. Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable framework for spatial omics analysis. *Nat Methods*. 2022;19(2):171–8.
24. Fotheringham AS, Wong DWS. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environ Plan A*. 1991;23(7):1025–44.
25. Dries R, Zhu Q, Dong R, Eng CHL, Li H, Liu K, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol*. 2021;22(1):78.
26. Li Z, Zhou X. BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. *Genome Biol*. 2022;23(1):168.
27. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for research on epithelial–mesenchymal transition. *Nat Rev Mol Cell Biol*. 2020;21(6):341–52.
28. Barkley D, Moncada R, Pour M, Liberman DA, Dryg I, Werba G, et al. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat Genet*. 2022;54(8):1192–201.
29. Aggarwal V, Montoya CA, Donnenberg VS, Sant S. Interplay between tumor microenvironment and partial EMT as the driver of tumor progression. *iScience*. 2021;24(2):102113.
30. Pastushenko I, Blanpain C. EMT Transition States during tumor progression and metastasis. *Trends Cell Biol*. 2019;29(3):212–26.

31. Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. *Nat Commun.* 2020;11(1):2142.
32. Haerincx J, Goossens S, Bex G. The epithelial–mesenchymal plasticity landscape: principles of design and mechanisms of regulation. *Nat Rev Genet.* 2023;24(9):590–609.
33. Malagoli Tagliazucchi G, Wiecek AJ, Withnell E, Secrier M. Genomic and microenvironmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer. *Nat Commun.* 2023;14(1):1–20.
34. Home Page - 10x Genomics. Available from: <https://www.10xgenomics.com/>. Cited 2023 Mar 6.
35. Kleshchevnikov V, Shmatko A, Dann E, Aivazidis A, King HW, Li T, et al. Cell 2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol.* 2022;40(5):661–71.
36. Sahoo S, Ramu S, Nair MG, Pillai M, Juan BPS, Milioli HZ, et al. Multi-modal transcriptomic analysis unravels enrichment of hybrid epithelial/mesenchymal state and enhanced phenotypic heterogeneity in basal breast cancer. 2023. Preprint at: <https://doi.org/10.1101/2023.09.30.558960>.
37. Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY, et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med.* 2014;6(10):1279–93.
38. Elyanow R, Zeira R, Land M, Raphael BJ. STARCH: copy number and clone inference from spatial transcriptomics data. *Phys Biol.* 2021;18(3):035001.
39. Pirrotta S, Masatti L, Corrà A, Pedrini F, Esposito G, Martini P, et al. signifinder enables the identification of tumor cell states and cancer expression signatures in bulk, single-cell and spatial transcriptomic data. 2023. Preprint at: <https://doi.org/10.1101/2023.03.07.530940>.
40. Xun Z, Ding X, Zhang Y, Zhang B, Lai S, Zou D, et al. Reconstruction of the tumor spatial microenvironment along the malignant-boundary-nonmalignant axis. *Nat Commun.* 2023;14(1):1–16.
41. Brown MS, Abdollahi B, Wilkins OM, Lu H, Chakraborty P, Ognjenovic NB, et al. Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. *Sci Adv.* 2022;8(31):eabj8002.
42. Muz B, de la Puente P, Azab F, Azab AK. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia (Auckl).* 2015;11(3):83–92.
43. Weidemann A, Johnson RS. Biology of HIF-1 α . *Cell Death Differ.* 2008;15(4):621–7.
44. Tam SY, Wu VWC, Law HKW. Hypoxia-induced epithelial-mesenchymal transition in cancers: HIF-1 α and beyond. *Front Oncol.* 2020;8(10):486.
45. Nishida N, Yano H, Nishida T, Kamura T, Kojiro M. Angiogenesis in Cancer. *Vasc Health Risk Manag.* 2006;2(3):213–9.
46. Almagro J, Messal HA, Elosegui-Artola A, van Rheenen J, Behrens A. Tissue architecture in tumor initiation and progression. *Trends Cancer.* 2022;8(6):494–505.
47. Saxena K, Jolly MK, Balamurugan K. Hypoxia, partial EMT and collective migration: Emerging culprits in metastasis. *Translational Oncology.* 2020;13(11):100845.
48. Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, et al. IL-1-induced JAK/STAT signaling is antagonized by TGF- β to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Discov.* 2019;9(2):282.
49. Ghahremanifard P, Chanda A, Bonni S, Bose P. TGF- β Mediated immune evasion in cancer—spotlight on cancer-associated fibroblasts. *Cancers.* 2020;12(12):3650.
50. Principe DR, Timbers KE, Atia LG, Koch RM, Rana A. TGF β Signaling in the Pancreatic Tumor Microenvironment. *Cancers (Basel).* 2021;13(20):5086.
51. Xu J, Lamouille S, Derynck R. TGF- β -induced epithelial to mesenchymal transition. *Cell Res.* 2009;19(2):156–72.
52. Monteran L, Erez N. myCAFs are better than yours: targeting myofibroblasts potentiates immunotherapy. *Trends Cancer.* 2023;9(1):1–2.
53. Amer HT, Stein U, El Tayebi HM. The monocyte, a Maestro in the Tumor Microenvironment (TME) of breast cancer. *Cancers.* 2022;14(21):5460.
54. Grosso JF, Kelleher CC, Harris TJ, Maris CH, Hipkiss EL, Marzo AD, et al. LAG-3 regulates CD8 $^{+}$ T cell accumulation and effector function in murine self- and tumor-tolerance systems. *J Clin Invest.* 2007;117(11):3383–92.
55. Cui C, Xu C, Yang W, Chi Z, Sheng X, Si L, et al. Ratio of the interferon- γ signature to the immunosuppression signature predicts anti-PD-1 therapy response in melanoma. *NPJ Genom Med.* 2021;6(1):7.
56. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat Rev Immunol.* 2015;15(11):669–82.
57. Yang X, Lin Y, Shi Y, Li B, Liu W, Yin W, et al. FAP promotes immunosuppression by cancer-associated fibroblasts in the tumor microenvironment via STAT3–CCL2 Signaling. *Cancer Res.* 2016;76(14):4124–35.
58. Zhao K, Yi Y, Ma Z, Zhang W. INHBA is a prognostic biomarker and correlated with immune cell infiltration in cervical cancer. *Front Genet.* 2022;4(12):705512.
59. Johnson KA, Emmerich P, Matkowskyj KA, Deming DA. Predicting CD8 $^{+}$ T-cell infiltration in colorectal cancer using versican proteolysis across molecular profiles. *JCO.* 2020;38(4_suppl):189–189.
60. Chen P, Cescon M, Bonaldo P. Collagen VI in cancer and its biological mechanisms. *Trends Mol Med.* 2013;19(7):410–7.
61. Johnson AM, Bullock BL, Neuwelt AJ, Poczebott JM, Kaspar RE, Li HY, et al. Cancer Cell-Intrinsic Expression of MHC Class II Regulates the Immune Microenvironment and Response to Anti-PD-1 Therapy in Lung Adenocarcinoma. *J Immunol.* 2020;204(8):2295–307.
62. Liu J, Ling Y, Su N, Li Y, Tian S, Hou B, et al. A novel immune checkpoint-related gene signature for predicting overall survival and immune status in triple-negative breast cancer. *Transl Cancer Res.* 2022;11(1):181–92.
63. Hazini A, Fisher K, Seymour L. Deregulation of HLA-I in cancer and its central importance for immunotherapy. *J Immunother Cancer.* 2021;9(8):e002899.
64. Wuerfel FM, Huebner H, Häberle L, Gass P, Hein A, Jud SM, et al. HLA-G and HLA-F protein isoform expression in breast cancer patients receiving neoadjuvant treatment. *Sci Rep.* 2020;10(1):15750.
65. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest.* 2017;127(8):2930–40.

66. Zormpas E, Queen R, Comber A, Cockell SJ. Mapping the transcriptome: Realizing the full potential of spatial data analysis. *Cell*. 2023;186(26):5677–89.
67. Ganier C, Mazin P, Herrera-Oropeza G, Du-Harpur X, Blakeley M, Gabriel J, et al. Multiscale spatial mapping of cell populations across anatomical sites in healthy human skin and basal cell carcinoma. *Proc Natl Acad Sci*. 2024;121(2):e2313326120.
68. Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat Biotechnol*. 2022;40(9):1349–59.
69. Valdeolivas A, Amberg B, Giroud N, Richardson M, Gálvez EJC, Badillo S, et al. Profiling the heterogeneity of colorectal cancer consensus molecular subtypes using spatial transcriptomics. *NPJ Precis Onc*. 2024;8(1):1–16.
70. Sherman TD, Gao T, Fertig EJ. CoGAPS 3: Bayesian non-negative matrix factorization for single-cell analysis with asynchronous updates and sparse data structures. *BMC Bioinformatics*. 2020;21(1):453.
71. Spatial and single-nucleus transcriptomic analysis of genetic and sporadic forms of Alzheimer's Disease | bioRxiv. Available from: <https://www.biorxiv.org/content/10.1101/2023.07.24.550282v1>. Cited 2023 Nov 27.
72. Ospina OE, Wilson CM, Soupir AC, Berglund A, Smalley I, Tsai KY, et al. spatialGE: quantification and visualization of the tumor microenvironment heterogeneity using spatial transcriptomics. *Bioinformatics*. 2022;38(9):2645–7.
73. Emami Nejad A, Najafgholian S, Rostami A, Sistani A, Shojaeifar S, Esparvarinha M, et al. The role of hypoxia in the tumor microenvironment and development of cancer stem cell: a novel approach to developing treatment. *Cancer Cell Int*. 2021;21(1):62.
74. Wang X, Zhang W, Sun X, Lin Y, Chen W. Cancer-associated fibroblasts induce epithelial-mesenchymal transition through secreted cytokines in endometrial cancer cells. *Oncol Lett*. 2018;15(4):5694–702.
75. Jia C, Wang G, Wang T, Fu B, Zhang Y, Huang L, et al. Cancer-associated Fibroblasts induce epithelial-mesenchymal transition via the Transglutaminase 2-dependent IL-6/IL6R/STAT3 axis in Hepatocellular Carcinoma. *Int J Biol Sci*. 2020 Jul 19;16(14):2542–58.
76. Tyler M, Tirosh I. Decoupling epithelial-mesenchymal transitions from stromal profiles by integrative expression analysis. *Nat Commun*. 2021;12(1):2592.
77. Li X, Chen L, Peng X, Zhan X. Progress of tumor-associated macrophages in the epithelial-mesenchymal transition of tumor. *Front Oncol*. 2022;28(12):911410.
78. Brabletz S, Schuhwerk H, Brabletz T, Stemmle MP. Dynamic EMT: a multi-tool for tumor progression. *EMBO J*. 2021 Sep 15;40(18):e108647.
79. Pang MF, Georgoudaki AM, Lambut L, Johansson J, Tabor V, Hagikura K, et al. TGF- β 1-induced EMT promotes targeted migration of breast cancer cells through the lymphatic system by the activation of CCR7/CCL21-mediated chemotaxis. *Oncogene*. 2016;35(6):748–60.
80. Kmiecik M, Knutson KL, Dumur CI, Manjili MH. HER-2/neu antigen loss and relapse of mammary carcinoma are actively induced by T cell-mediated anti-tumor immune responses. *Eur J Immunol*. 2007;37(3):675–85.
81. Santisteban M, Reiman JM, Asiedu MK, Behrens MD, Nassar A, Kalli KR, et al. Immune-induced epithelial to mesenchymal transition in vivo generates breast cancer stem cells. *Cancer Res*. 2009;69(7):2887–95.
82. Galsky MD, Wang L, Saci A, Szabo PM, Gong Y, Zhu J. 850PD - Epithelial-mesenchymal transition (EMT), T cell infiltration, and outcomes with nivolumab (nivo) in urothelial cancer (UC). *Ann Oncol*. 2017;1(28):v297.
83. Chockley PJ, Chen J, Chen G, Beer DG, Standiford TJ, Keshamouni VG. Epithelial-mesenchymal transition leads to NK cell-mediated metastasis-specific immunosurveillance in lung cancer. *J Clin Invest*. 2018;128(4):1384–96.
84. Datar I, Schalper KA. Epithelial-mesenchymal transition and immune evasion during lung cancer progression: the chicken or the egg? *Clin Cancer Res*. 2016;22(14):3422–4.
85. Terry S, Savagner P, Ortiz-Cuaran S, Mahjoubi L, Saintigny P, Thierry JP, et al. New insights into the role of EMT in tumor immune escape. *Mol Oncol*. 2017;11(7):824–46.
86. Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell*. 2012;22(6):725–36.
87. Cheng YC, Zhang Y, Tripathi S, Bv H, Jolly MK, Schiebinger G, et al. Reconstruction of single cell lineage trajectories and identification of diversity in fates during the epithelial-to-mesenchymal transition. 2023. Preprint at: <https://doi.org/10.1101/2023.09.19.558325>.
88. Sinha D, Saha P, Samanta A, Bishayee A. Emerging concepts of hybrid epithelial-to-mesenchymal transition in cancer progression. *Biomolecules*. 2020;10(11):1561.
89. Subhadarshini S, Markus J, Sahoo S, Jolly MK. Dynamics of epithelial-mesenchymal plasticity: what have single-cell investigations elucidated so far? *ACS Omega*. 2023;8(13):11665–73.
90. Zhang Y, Zhang H, Wang M, Schmid T, Xin Z, Kozhuharova L, et al. Hypoxia in breast cancer-scientific translation to therapeutic and diagnostic clinical applications. *Front Oncol*. 2021;11:652266.
91. Fu Z, Mowday AM, Smail JB, Hermans IF, Patterson AV. Tumour Hypoxia-Mediated Immunosuppression: mechanisms and therapeutic approaches to improve cancer immunotherapy. *Cells*. 2021;10(5):1006.
92. Lee CT, Mace T, Repasky EA. Hypoxia-Driven Immunosuppression: A new reason to use thermal therapy in the treatment of cancer? *Int J Hyperthermia*. 2010;26(3):232–46.
93. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. 2018;19(1):15.
94. Stein-O'Brien GL, Clark BS, Sherman T, Zibetti C, Hu Q, Sealton R, et al. Decomposing Cell Identity for Transfer Learning across Cellular Measurements, Platforms, Tissues, and Species. *Cell Systems*. 2019;8(5):395–411.e8.
95. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–25.
96. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A Comparison of PAM50 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor-Positive Breast Cancer. *Clin Cancer Res*. 2010;16(21):5222–32.
97. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol*. 2015;15(8):486.
98. Proceedings of the Python in Science Conference (SciPy): Exploring Network Structure, Dynamics, and Function using NetworkX. Available from: https://conference.scipy.org/proceedings/SciPy2008/paper_2/. Cited 2023 Mar 6.

99. Rey SJ, Anselin L. PySAL: A Python library of spatial analytical methods. *Rev Reg Stud.* 2007;37(1):5–27.
100. Barkley D, Yanai I. Cancer cell states recur across tumor types and form specific interactions with the tumor micro-environment. *Gene Expression Omnibus*; 2022. <https://identifiers.org/geo:GSE203612>.
101. Wu S, Swarbrick A. A single-cell and spatially resolved atlas of human breast cancers | spatial transcriptomics data. *Zenodo*; 2021. <https://zenodo.org/records/4739739>.
102. Wu S, Swarbrick A. A single-cell and spatially resolved atlas of human breast cancers. *Gene Expression Omnibus*; 2021. <https://identifiers.org/geo:GSE176078>.
103. Brown MS, Pattabiraman D. Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. *Gene Expression Omnibus*; 2022. <https://identifiers.org/geo:GSE172613>.
104. Withnell E, Secrier M. SpottedPy. *GitHub*; 2024. Available from: <https://github.com/secrierlab/SpottedPy>.
105. Withnell E, Secrier M. Sample SpottedPy dataset for tutorial. 2024. *Zenodo*. <https://doi.org/10.5281/zenodo.13907274>.

Publisher's Note

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