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

Eloise Withnell[1], Maria Secrier[1,*]

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

* To whom correspondence should be addressed (m.secrier@ucl.ac.uk)

Keywords: spatial transcriptomics, epithelial-to-mesenchymal transition, tumour microenvironment, cell plasticity, hotspot analysis

ABSTRACT

Spatial transcriptomics is revolutionising our ability to explore intratissue heterogeneity in cancer, but methods that can effectively capture cancer cell niches and explore their relationships with the tumour microenvironment at various spatial scales remain limited. Here we present SpottedPy, a Python package designed to identify tumour hotspots and map spatial interactions within the cancer ecosystem. We employ SpottedPy to examine epithelial-mesenchymal plasticity in breast cancer and highlight locally stable niches associated with angiogenic and hypoxic regions, and shielded by myCAFs, macrophages and perivascular cell populations. Hybrid and mesenchymal hotspot distribution followed transformation gradients within the tissue reflecting progressive immunosuppression. Our method offers the flexibility to explore spatial relationships at different scales, from immediate neighbours to broader tissue modules, providing new insights into the spatial dynamics of the tumour microenvironment.

BACKGROUND

The spatial organisation of cancer cells and their interactions with immune and stromal cells in their environment are instrumental in dictating tumour behaviour 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 tumour microenvironments (TMEs)[4,5]. Profiling the cancer tissue spatially allows us to explore the tumour architecture and its heterogeneity in detail, paving the way to deciphering the crosstalk between tumour 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[9], immunosuppressive hubs[10,11], entire tumour ecotypes with divergent clinical outcomes[12] or the impact of specific drugs on inhibiting tumour 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 organised co-expression hotspots aligning with tissue histology. Given that similar cells often cluster together[18,19], methods that can reliably detect statistically significant clusters are important in reinforcing the accuracy of cell states determined from continuous signatures. CellCharter[20] capitalises 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 neighbourhood have been implemented in packages like Squidpy[23]. However, these approaches lack analytical methods to define and compare distances between specific areas or cell populations of interest that might exert influence on each other across different ranges. Given that the scale of certain biological processes in cancer, such as hypoxia, remains elusive, relying solely on conventional spot neighbourhood-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. Whilst a growing number of methods address the need for multi-scale analysis [25,23,26,27], 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 neighbourhood of individual spots. We assess how the relationships between these variables change when varying the hotspot size or the neighbourhood size surrounding a spot. 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, polarised epithelial cells undergo multiple molecular changes and lose their identity to acquire a mesenchymal phenotype[28]. The interplay between EMT and the tumour microenvironment (TME) is multifaceted: while the TME is believed to be an inducer of EMT, mesenchymal tumour cells potentially influence the TME[29,30]. 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[31,32]. 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)[33]. We have previously shown that the tumour spatial organisation follows EMP gradients and that hybrid and mesenchymal cells establish distinct interactions with the TME in small datasets of spatially-profiled breast tumours[34]. However, the spatial organisation and interactions of cells across this EMP spectrum within the tumour 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 tumour and TME relationships, we yield rigorous evidence of spatial dynamics and offer an interpretable and intuitive measure of interactions between tumour 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 neighbourhoods 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. Whilst neighbourhood enrichment is commonly used in the field[23,25,29], analysing continuous expression signatures and examining of how the size of the neighbourhood influences spatial relationships are areas that remain underdeveloped. Furthermore, inspecting the immediate neighbourhood of cells of interest versus broader hotspots within the

tissue will yield different insights into the tissue architecture and organisation, 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:

- **Neighbourhood enrichment analysis:** Our method introduces a function designed to examine correlations between cell states, populations or processes within individual spatial transcriptomics spots and their immediate neighbourhood (**Fig. 1bi**). In this context, a 'neighbourhood' 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.

- **Hotspot identification:** We have implemented the Getis-Ord G* statistic to identify spatial clusters of continuous signatures across the tumour slide. Users can flexibly filter for specific regions within the slide to focus on creating hotspots. By comparing defined regions with high or low 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 hot and coldspots.

- **Distance metrics:** SpottedPy includes analytical tools to measure and interpret the distances between identified clusters, such as between tumour and immune hotspots. The primary method for this 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. 1bii**). Additionally, SpottedPy introduces two supplementary methods for distance analysis. The *centroid to centroid* approach compares the centroids of hotspots (**Fig. 1biii**) for a macro-level view of spatial distribution. The *gene signature correlation* method correlates tumour signatures within a defined region (therefore without hotspot creation) with the distance to other hotspots of interest, allowing an investigation of continuous tumour state signatures (**Fig. 1 iv**). Importantly, SpottedPy assigns a statistical significance to these proximity measures to determine if observed patterns are likely to occur just by chance.

- **Sensitivity analysis:** We provide the ability to systematically evaluate how cell-cell relationships evolve within the tissue as we vary the size of the neighbourhood or of the hotspot of interest. For the neighbourhood 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* statistic with varying neighbourhood sizes, enabling the identification of clusters at different spatial scales. By analysing how the distances between hotspots change with neighbourhood 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* statistic.

To highlight the potential of our method, we employ SpottedPy to investigate the relationships between tumour cells undergoing EMT and the TME in 12 breast cancer slides profiled using the 10x Genomics Visium spatial transcriptomics platform, integrated from Wu et al[18], Barkley et al[24] and the 10x Genomics website[25]. To infer individual cell types within the slides, we performed cell deconvolution using the Cell2location method[36] and a scRNA-seq reference of annotated breast cancer cell population profiles from 123,561 cells[13]. To more precisely identify tumour cells, which tend to show vast expression variability, we employed the copy number inference tool STARCH[37] and only kept spots that showed evidence of copy number changes, which are likely to be tumour-specific. Next, we delineated tumour hotspots using SpottedPy and scored them using an EMT hallmark signature which has been linked with patient outcomes in the clinic[38,39]. 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[40], 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 tumour hallmarks

Firstly, we aimed to explore the cellular environment that is conducive to E/M progression. To do so, we focused solely on tumour hotspots and two key cellular hallmarks that have been previously linked with EMT, hypoxia and angiogenesis. Hypoxia, characterised by low oxygen levels, has been long recognized as a key enabler of tumorigenic processes[41]. Under hypoxic conditions, tumour cells stabilise hypoxia-inducible factors (HIFs), primarily HIF-1α, which promotes angiogenesis[42], 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[43], and therefore understanding how such relationships develop spatially within the tissue can help devise localised therapies that can interrupt these interactions in breast cancer. Within individual spatial transcriptomics slides, we used SpottedPy to delineate tumour areas, and further identified EMT hotspots within these areas using using the EMT hallmark signature (**Fig. 2a**). 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)**. 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 EMT coldspots, proliferative hotspots, or the average tumour cell (**Fig. 2b-c**). In contrast, proliferative hotspots were significantly closer to EMT coldspots ($p<0.001$, **Fig. 2c**). These relationships were consistently observed across all breast cancer slides (**Supplementary Fig. 1a**).

To further grasp the positioning of these EMT “hot” and “cold” areas within the tumour, we used SpottedPy to determine the tumour perimeter (**Fig. 2d**) and calculated distances to it. We find EMT hotspots closer to the tumour perimeter compared to EMT coldspots, suggesting a state with significant interaction with the surrounding microenvironment (**Fig. 2e**). As expected, angiogenesis

hotspots were located closest to the tumour perimeter, followed by hypoxia hotspots (**Fig. 2f**). The spatial localisation of angiogenesis near the tumour perimeter aligns with its function in supplying nutrients and oxygen to rapidly growing tumours[44]. The prominence of hypoxic regions succeeding angiogenic zones is consistent with the understanding that rapid tumour growth often outpaces its vascular supply, leading to hypoxic conditions[41]. These hypoxic conditions are alleviated in the angiogenic areas, as we find hypoxic coldspots are closest to the tumour perimeter (**Fig. 2f**). Cell proliferation hotspots were observed furthest from the tumour perimeter, and located at spatially distinct locations to EMT hotspots, in line with tumour growth studies outlining a proliferative epithelial core and EMT transformation at the periphery facilitating cancer cell intravasation and migration[45,46].

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 tumour cells undergoing EMT and other immune and stromal cells in the microenvironment. Alongside EMT hotspots, we calculated hotspots for 21 cell types in the TME, including different lymphocyte, myeloid and fibroblast populations (**Fig. 3a-c**). When visually inspecting these hotspots, we observed myCAF hotspots and EMT hotspots tended to co-localise (**Fig. 3d**). Quantifying hotspot distances using SpottedPy allowed us to confirm that tumour EMT hotspots were indeed closest to myCAF hotspots (**Fig. 3e**). This is well in line with existing studies, as myCAFs have been shown to produce TGF-β, which is a well-known EMT trigger[47]. They have also been linked to ECM deposition and suppression of antitumor immunity[48,49,50,51]. In contrast, iCAFs appeared to be the most distant cell type to EMT hotspots (**Fig. 3e**). This is in keeping with observations in pancreatic cancer by Ohlund et al[52], showing that direct juxtracrine interactions with cancer cells are required for myCAF formation, whereas iCAFs are induced by secreted factors from cancer cells and thus are not found in their proximity.

myCAF hotspots appeared significantly closer to EMT hotspots compared to EMT coldspots, followed by macrophages, cycling perivascular-like cells (PVL), monocytes and dendritic cells (DCs) ($p<0.001$, **Fig. 3f**). Monocytes, and particularly tumour-associated macrophages (TAMs) derived from them, are known to modulate the environment of tumour cells undergoing EMT, usually by promoting immune suppression in the TME, which facilitates tumour progression and metastasis[53]. PVL cells, a type of stromal cell characterized by a perivascular-like profile yet not associated with blood vessels, represent a recently identified cell type whose functions are largely unknown[54]. Differentiated PVL cells have been linked to immune evasion in breast cancer[54]. Natural Killer (NK) and NK T cells, the immune cells that can directly kill transformed cells, were the least associated with EMT hotspots when averaged across slides, potentially reflecting a mechanism of immune evasion employed by tumour cells that have undergone EMT (**Fig. 3f**).

The cohort-level associations between EMT hotspots, myCAFs and macrophages were maintained in individual tumours, suggesting that these are universal patterns of EMT transformation in breast cancer and not subtype specific (**Supplementary Fig. 1b**). In contrast, levels of inter-patient heterogeneity, often even within the same breast cancer subtype, were evident for a variety of cells

including memory B-cells, naïve B-cells, iCAFs, NK cells, NKT cells, CD4+ T-cells and CD8+ T-cells. Distances to EMT hotspots were consistent across subgroups of cells (**Supplementary Fig. 1c**), 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 actors, 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[35] compared to EMT coldspots (**Fig. 3g-h**), with the key immune checkpoints *B7-H3* (*CD276)*, *OX40* (*TNFRSF4*) and *TIM3* (*HAVCR2*) displaying significantly higher expression ($p<0.05$) in the majority on tumour slides (**Fig. 3h**). 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[55] (**Fig. 3g**, **Fig. 3i**). We found that EMT hotspots had significantly increased expression of genes within the signature, most notably of *HLA-DRA*, a marker of improved response to anti-PD1 therapy[56].

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 neighbours for the Getis-Ord Gi* metric, affects spatial relationships is crucial for meaningful spatial analysis. We systematically increased hotspot dimensions (**Fig. 4a-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 tumour microenvironment, with such relationships remaining remarkably stable across a range of hotspot dimensions (**Fig. 4c**). The cell populations that we previously identified as having the nearest proximity to EMT hotspots at a fixed parameter size (myCAFs, macrophages, cycling PVL cells and monocytes) also maintained this relationship when varying hotspot sizes. Cells that were farther apart presented less stable associations, such as dendritic cells where the relationship broke down at a hotspot size of 200, and naive B-cells where the relationship changed multiple times with increasing hotspot size (**Fig. 4c**, **Supplementary Fig. 2a**). These findings suggest that interactions with certain cells in the TME may be more pronounced and relevant at a smaller scale. We found that NK cells were the only cells consistently adjacent to EMT coldspots at various hotspot sizes. Adjusting the p-value cut-off used to detect spatial clusters using the Getis-Ord Gi* highlighted similar relationships (**Supplementary Fig. 2b**). We also see that the number of hotspots decreases when increasing the hotspot parameter size, indicative of the hotspots becoming more clustered together as the neighbourhood expands, as expected (**Fig. 4d**).

291 **Other distance metrics**
292
293 All the results described so far have evaluated distances between hotspots using the shortest path
294 method (**Fig. 1bii**). We thus wondered whether the stability of these relationships would be affected
295 by the method chosen to measure hotspot distances. To investigate this, we used the "centroid to
296 centroid" approach to calculate the distance from centroids of each EMT hotspot to centroids for
297 cancer hallmarks and immune/stromal hotspots, with the resultant rankings displayed in
298 **Supplementary Fig. 2c-d**. Reassuringly, we observe a consistent relationship with tumour hallmark
299 signatures, with EMT hotspots located closer to hypoxic and angiogenic regions compared to EMT
300 coldspots and further away from proliferative regions. We also see cycling PVLs and myCAFs
301 remaining in the top cell types associated with EMT hotspots, while macrophages and monocytes
302 appear more distal to these hotspots compared to our previous shortest path distance approach.
303
304 Whilst the "centroid to centroid" methodology offers a practical and straightforward way to
305 approximate the distances between hotspots, it is crucial to acknowledge its simplicity. As depicted
306 in **Supplementary Fig. 2e**, the size of a hotspot significantly influences its centroid location.
307 Consequently, a larger hotspot, despite being physically closer, may appear farther away when
308 measuring centroid distances, due to its centroid being located farther from the point of interest
309 compared to a smaller, neighbouring hotspot. Thus, the centroid approach may miss local variation
310 that the shortest path method can capture. This could suggest that applying it on our breast cancer
311 slides could potentially miss the more complex relationship observed between EMT hotspots and
312 macrophage or monocyte-enriched areas.
313
314 **Neighbourhood enrichment analysis**
315
316 The neighbourhood enrichment technique captures more localised, shorter-range relationships with
317 the TME, by correlating phenotypes across a central tumour spot and the direct neighbourhood
318 surrounding it (a ring encompassing six Visium spots) (**Fig. 1bi**). Varying the number of rings
319 around the central spot enables us to assess whether the observed hotspot relationships shift with
320 the unit of analysis.
321
322 Our analysis revealed that angiogenesis, myCAFs and monocytes exhibited the most significant
323 correlation, in descending order, with EMT hallmark signatures ($p<0.001$) across the 12 slides (**Fig.**
324 **5a**). This finding reinforces the spatial association between these cell types and signatures with
325 EMT that we previously identified using the hotspot method. A significant association was not
326 evident for naive B-cells or CD4+ T-cells (**Fig. 5a**). This lack of a discernible relationship using the
327 neighbourhood enrichment approach for CD4+ T-cells, despite the hotspot approach capturing a
328 weak significant relationship, suggests that the relationship is more indirect, potentially occurring
329 over a broader spatial expanse than our neighbourhood enrichment approach can accurately
330 capture.
331
332 We found that these spatial relationships were stable across various neighbourhood sizes (**Fig 5b-**
333 **c**). Among the most consistent were the hypoxia and angiogenic signatures. Conversely,
334 relationships between EMT and plasmablasts, iCAFs or proliferative hotspots were the most

variable (**Fig 5c**), suggesting that this relationship might be more strongly determined by the environment in a local neighbourhood rather than at a broader scale.

EMT state fluctuations shape distinct immune niches within the same tumour

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 tumour 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[57] (**Supplementary Fig. 3a**). The corresponding hotspots occupied distinct spatial locations within the tissue (**Supplementary Fig. 3b**). When visually inspecting these hotspots, we detected a progressive transformation in the tumour, 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-localised with the EPI state. The experimental study by Brown et al[40] had detected that M2 cell clones gained integrin β4 (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 the generic EMT hallmarks signature employed previously for each tumour spot (**Supplementary Fig. 4a**). The lack of significant positive correlation between EPI, EM2, EM3, M1, 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 EMT hallmarks signature, as expected. EM2, EM3 and M2 are likely distinct states, as evidenced by the lack of correlations across the spots. In terms of spatial distribution, the EMT hallmark hotspots were located closest to the M1 hotspots, and furthest away from the EPI hotspots (**Supplementary Fig. 4b**), in line with the correlation analysis and confirming the hypothesised identities of these states.

When investigating how tumour 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 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**, **Supplementary Fig. 4c**), with a subset of genes, most notably *OX40 (TNFRSF4), TIM3* (*HAVCR2*), *HLA-DRA, 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**, **Supplementary Fig. 4d**). 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 tumour state that is not linked to immunosuppression.

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

DISCUSSION

In this study we introduce SpottedPy, a Python package that identifies tumour hotspots in spatial transcriptomics slides and explores their interplay with the TME at varying scales. We show that the Getis-Ord G* statistic can be successfully applied to delineate cellular hotspots and provide meaningful biological insights into the spatial organisation of the tumour tissue in its immune and stromal contexture. Whilst various studies have recently applied hotspot analysis to spatial transcriptomic data[58,59,21], these methods do not analyse the distances between hotspots. We build on these approaches to analyse the spatial relationships between hotspots, with the additional ability of anchoring hotspot identification to specific regions of interest, such as tumour cells or non-tumour cells, enabling us to delve into 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. Our method additionally assesses the effect of hotspot size on spatial relationships and compares hotspot spatial trends to the relationships captured using neighbourhood 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 tumour plasticity phenotypes in breast cancer, we have uncovered key differences between tumour 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 tumour 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[60,43,46].

We find a strong relationship between EMT and myCAFs across all slides. CAFs have been linked to tumour cells undergoing EMT in our previous bulk and spatial transcriptomic analyses[34], and have been shown to induce EMT in endometrial cancer cells[61] and hepatocellular carcinoma[62]. However, the relationship between the recently defined CAF subtypes, myCAFs and iCAFs, and EMT has been less well studied and here we show specific interactions with the former and not the latter. 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 tumour settings[63]. Here we have leveraged scRNA-seq for deconvolution, alongside detecting copy number aberrations (CNAs) to confirm the presence of tumour cells, which adds an additional 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 resolution spatial transcriptomic data would allow us to confirm this relationship more confidently.

In addition to the expected myCAF associations, we observe a strong relationship with macrophages and monocytes across multiple spatial scales. Macrophages secrete TGF-β, TNF-α, IL-6 and IL-8, which are all well characterised EMT stimuli[64,65]. The relationship has been observed in bulk transcriptomics[66], in spatial analysis of mouse models of skin carcinoma where depletion of macrophages inhibited EMT progression[31] and in specific niches within breast cancer spatial transcriptomics slides[11].

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[67,68], and this relationship has further been highlighted in bulk transcriptomics[69] and smaller scale spatial transcriptomic analyses[34,70] 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[64,51,71]. 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[72,11].

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 Jia et al[51] utilising a more focused spatial transcriptomic dataset, previous research by Tsai et al[73] demonstrating that a departure from a mesenchymal-like state is a prerequisite for tumour cell proliferation in mouse models, and a recent study by Chen et al[74] investigating EMT states in scRNA-seq data. Such spatial characterisations at various scales were largely unexplored.

Delving deeper into EMT, we observed that hybrid EMT states exhibit more heterogeneous and weaker associations with the TME in comparison to the quasi-mesenchymal M1 state. This disparity might be indicative of the inherent plasticity of hybrid EMT states[75,76], complicating our ability to

465 delineate clear relationships, but also of a directed trajectory towards an M1 state. The M2 state
466 demonstrates relatively weaker connections with the TME, which may be attributed to the
467 experimental gain of integrin β4 (a key epithelial marker) when cultured, a limitation mentioned in
468 the original study which potentially transformed the state towards a more epithelial phenotype[40]. We
469 found that the epithelial state exhibits a notable degree of independence from the TME which may
470 be due to the ability of cancer cells to successfully avoid immune recognition in the incipient stages
471 of neoplasia.
472
473 These results point to a highly dynamic and plastic nature of tumour cells in navigating the
474 complexities of their microenvironment. The interactions likely extend beyond a linear framework.
475 Hypoxia, a known catalyst for both angiogenesis and EMT[77], can set off a cascade of events that
476 not only amplify these processes but also create a conducive environment for the recruitment of
477 immunosuppressive cells such as macrophages[78,79]. These cells can in turn bolster angiogenesis,
478 thereby fuelling a self-perpetuating cycle that further complicates the tumour landscape. These
479 insights provide a further understanding of the cellular interactions and environmental factors that
480 underpin tumour progression and metastasis, and could in the future pave the way for the
481 development of targeted interventions aimed at disrupting these complex networks for therapeutic
482 benefit.
483
484 The consistency of the relationships we observed across different hotspot sizes and neighbourhood
485 scales further strengthens our confidence in the findings. The neighbourhood ring approach
486 predominantly captured TME cells that have infiltrated the tumour, offering insights into the
487 immediate cellular interplay at the tumour periphery. In contrast, the hotspot methodology provided
488 a broader view, encompassing interactions at more distal locations. By pinpointing statistically
489 significant cellular hotspots, we bolster the reliability of our observations, especially considering the
490 inherent inaccuracies that can arise from deconvolution algorithms applied to non-single cell
491 transcriptomic datasets such as those from the Visium platform. Furthermore, we offer three
492 methods to analyse hotspot distances, which allows flexibility in the analyses depending on the
493 biological questions that the user intends to answer.
494
495 In the future, a comprehensive benchmarking of different feature-based spatial clustering
496 approaches, such as kernel density estimation, and different statistical approaches to quantify the
497 distances between different hotspots would be valuable. This would enable us to determine how
498 sensitive spatial relationships are to methodological choice. Moreover, extending these analyses to
499 single cell-resolved spatial datasets and incorporating ligand-receptor signalling information into the
500 evaluation of spatial effects on cell populations will increase the confidence in the identified
501 relationships.
502
503 Our insights into the spatial organisation of tumours during EMT progression are limited by the
504 significant amount of uncertainty surrounding EMP programmes[32,33] and their incomplete
505 characterisation in different types of breast cancer and other neoplasms. In the future, integrating
506 further hybrid states characterised in other breast cancer studies[74] will help expand our
507 understanding of this complex process alongside its multiple locally stable peaks and valleys. The
508 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.

Overall, our findings confirm expected spatial effects of EMT progression in tumours, demonstrating that SpottedPy is an adequate tool to capture complex associations between tumour cells and their microenvironment. Such insights can help unveil local effects of the TME and linked tumour cell vulnerabilities that could ultimately be exploited for therapeutic benefit. While the analyses presented here illustrate insights into breast cancer tissue organisation, we note that SpottedPy can be applied to discern spatial relationships in other cancer types 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.

Conclusion

In conclusion, SpottedPy provides a detailed and multifaceted analysis of the spatial dynamics within individual spatially-profiled tumours. 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

Spatial transcriptomic datasets

Breast cancer Visium slides were obtained from Barkley et al.[29] (slides 0-2), from 10x Genomics (slides 3-5)[35] and Wu et al.[12] (slides 6-12). Slide annotations, if available, are displayed in Supplementary Fig. 1b. We combined the three datasets of breasts cancer 10X Genomics Visium spatial transcriptomic datasets into a common *anndata* Python format for analysis. Pre-processing and normalisation were conducted using the ScanPy (Single-Cell Analysis in Python) package[80]. We analysed 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.

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[36]. Cell2location decomposes the spatial count matrix into a predefined set of reference cell signatures by modelling 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

552 26 patients and 21 cell types from Wu et al[12] was chosen to perform the deconvolution. Cell types
553 in the chosen breast dataset consisted of cancer epithelial cells (basal, cycling, Her2, LumA, LumB),
554 naïve and memory B cells, myCAF-like and iCAF-like cancer-associated fibroblasts, perivascular-
555 like cells (PVL), including immature, cycling and differentiated, cycling T-cells, cycling myeloid cells,
556 dendritic cells (DCs), endothelial cells expressing *ACKR1*, *CXCL12* or *RGS5*, endothelial lymphatic
557 *LYVE1*-expressing cells, luminal progenitors and mature luminal cells, macrophages, monocytes,
558 myoepithelial cells, natural killer (NK) cells, natural killer T (NKT) cells, plasmablasts, CD4+ T cells,
559 and CD8+ T cells. The scRNA regression model was trained with 500 epochs, and the spatial
560 transcriptomic model trained with 20,000 epochs using a GPU. To delineate the tumour cells within
561 our spatial transcriptomics dataset, we used the STARCH Python package designed to infer copy
562 number alterations (CNAs)[37]. STARCH identifies tumour clones (setting *K*=3 clones) and non-
563 tumour spots. It confirms identification of normal spots by clustering the first principal component
564 into two clusters using K-means (setting K=2). This approach is based on the principle that the
565 direction of maximum variance in the expression data typically reflects the division between non-
566 cancerous and cancerous spots.
567
568 **EMT state and hallmark signature scoring**
569
570 To identify distinct EMT states, we employed data from Brown et al[40], consisting of seven RNA-seq
571 sequenced cell clones, derived from SUM149PT inflammatory breast cancer cell line with 3 repeats
572 spanning the EMT spectrum from epithelial-like (EPI), quasi-mesenchymal (M1), fully mesenchymal
573 (M2) and three distinct intermediates (EM1, EM2, EM3) . We used these data to derive a weighted
574 gene signature to represent the EMT states. We captured EMT gene patterns from this data using
575 non-negative matrix factorisation (NMF) by applying the CoGAPs workflow[57]. We used ProjectR's
576 implementation of *lmfit* R function to map the captured EMT patterns onto the spatial transcriptomic
577 spots[81]. This transfer learning approach assumes that if datasets share common latent spaces, a
578 feature mapping exists between them and can measure the extent of relationships between the
579 datasets. The final states were captured with 20 patterns and 10,000 training iterations. The number
580 of patterns were chosen based on capturing the discrete states with the highest accuracy. The EM1
581 state was not distinguishable from the EPI state, so we merged the two states. Thus, overall we
582 obtained scores for one epithelial, two intermediate, a quasi-mesenchymal and a fully mesenchymal
583 state for each spot.
584
585 The EMT, hypoxia and angiogenic hallmark signatures[38], response to immunotherapy signature[82],
586 immunosuppressive signature[12] and proliferative signature[83] were scored using
587 *scanpy.tl.score_genes* function.
588
589 **Graph construction**
590
591 The SquidPy[23] (Spatial Single-Cell Analysis in Python) package was used for graph construction
592 using *sq.gr.spatial_neighbors* and slide visualisation of the Visium spatial slides. NetworkX[84] was
593 used for further analysis of the networks derived from the spatial transcriptomic spots. The
594 deconvolved spot results were used to assign node labels. Edges were assigned based on the spot
595 neighbours.

Neighbourhood enrichment analysis

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

Neighbourhood enrichment was calculated by correlating signatures across a central spot of interest and the direct neighbourhood of spots surrounding it (a ring encompassing six Visium spots), after filtering for tumour spots only. Correlations were calculated using Pearson's correlation coefficient. To perform the sensitivity analysis, we increased the number of rings surrounding a spatial transcriptomic spot to consider as spot neighbours and compared the change in correlation coefficient.

We additionally compared this neighbourhood enrichment strategy to an approach solely correlating the central spot with the cellular composition in the neighbourhood ring around the spot, for which found less stable spatial relationships. Consequently, we kept the formed as default neighbourhood enrichment method, although the latter is also available to employ in the SpottedPy package.

Hotspot analysis

Hotspots were calculated using The Getis-Ord G* statistic as implemented using the PySAL package[85], using 10 as neighbourhood size parameter by default and a p-value of 0.05, unless otherwise stated.

The Getis-Ord G* equation is defined as follows:

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

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 for EMT states and cell proliferation were calculated after filtering for tumour cells as labelled by STARCH. 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 method to evaluate the significance of spatial clusters, offering a robust measure against potential spatial randomness in the data. Hotspots for EMT states and cell proliferation were calculated after filtering for tumour cells as labelled by STARCH.

Distance metrics

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

For a spot m in M and a spot e in E, we calculated the Euclidean distance to a spot h as $d(m,\ h)$. The minimum distances from M to H and E to$H$ are:

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

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

We can then compare the distribution of distances between the different spots in the hotspots.

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 centre 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.

We calculated the ***gene signature correlation*** metric by correlating each tumour signature value with the distance from that spot to the nearest hotspot of interest. We then compared how different signatures, using a continuous score to represent them, are correlated to distances to other hotspots of interest.

Tumour perimeter calculation

Any spot was considered part of the tumour perimeter if it had more than two neighbouring spots (nodes in the graph) that were not classified as tumour spots. This approach helped us to delineate the boundary of the tumour accurately by focusing on the transitional area where tumour and non-

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

Sensitivity analysis

The sensitivity analysis to evaluate the impact of varying hotspot sizes on the spatial relationships was achieved by incrementally adjusting the neighbourhood parameter for the Getis-Ord statistic, which directly influenced the size of identified hotspots. As we expanded the neighbourhood parameter, we compared the distances between the newly defined hotspots and other existing hotspots of interest.

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.

FIGURES

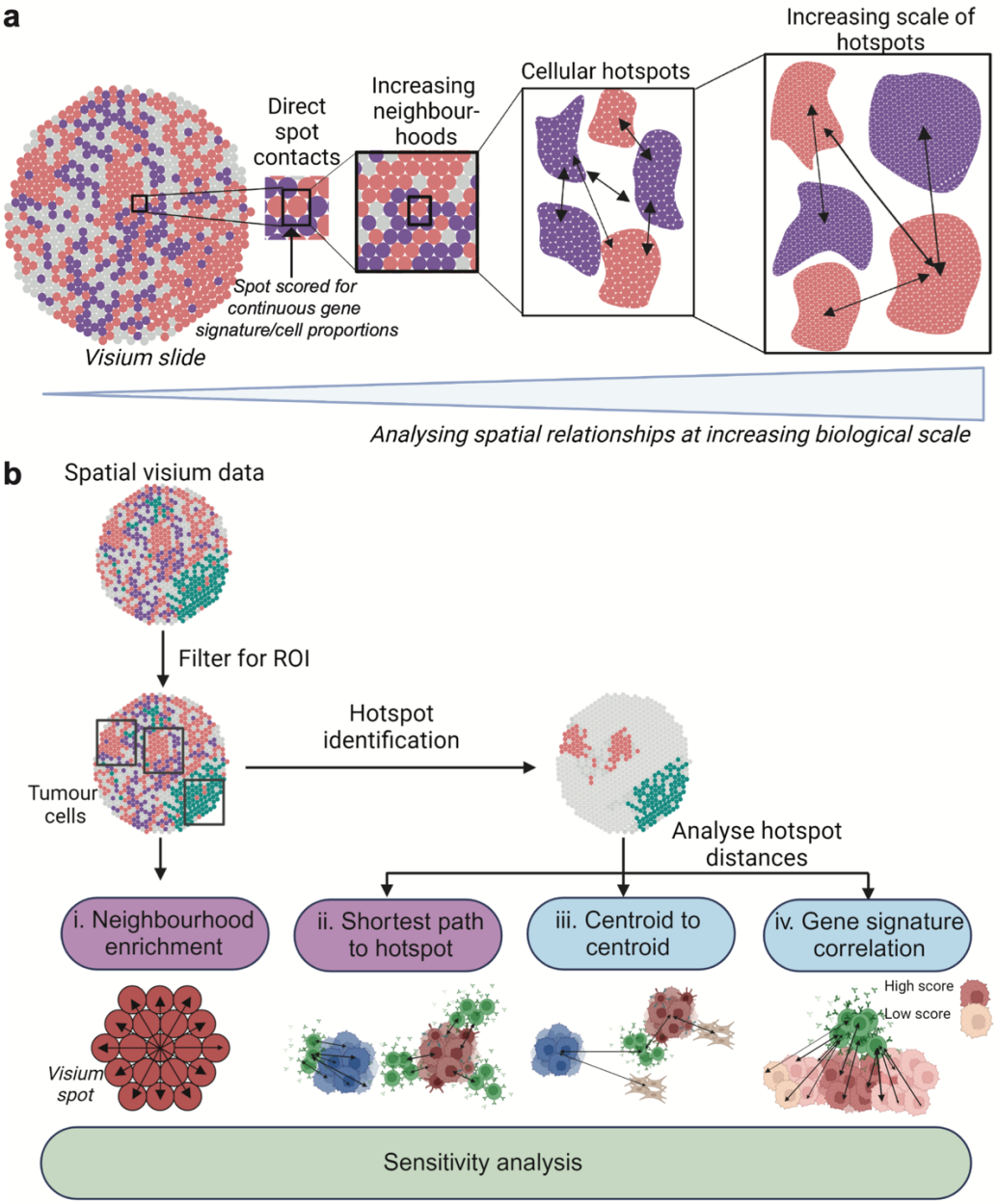

Figure 1: SpottedPy provides a multi-scale approach to analyse spatial transcriptomic relationships. (a) Overview of spatial scales captured in the SpottedPy workflow, from direct cellular contacts to broader cellular hotspots. Figure created with BioRender.com. **(b)** SpottedPy workflow overview. Visium spatial transcriptomic data is loaded as a pre-processed *AnnData* object where there is the option to select the region of interest within the slide e.g., *AnnData.obs* column labelled with tumour cells. Hotspots are identified based on transcriptional signatures or cell types. Key spatial measurements are highlighted in purple, with the additional functionality in blue. The default distance metrics include: (i) Neighbourhood enrichment, which correlates cell prevalence or signatures in individual spots with their immediate neighbourhood, and (ii) Shortest path to hotspot,

calculating the minimum distance between each spot within a selected hotspot and the nearest spot in other hotspots. Additional features for analysing hotspot distances in the package include: (iii) Centroid-to-centroid, which calculates distances between the centroids of two hotspots, and (iv) Gene signature correlation, correlating gene signatures with distances to hotspots. Sensitivity analysis allows us to compare relationships defined at different scales in both approaches, either by increasing the number of rings included for neighbourhood enrichment or increasing the hotspot size. The outputs for the different modules include various plots to highlight the relationships. Figure created with BioRender.com.

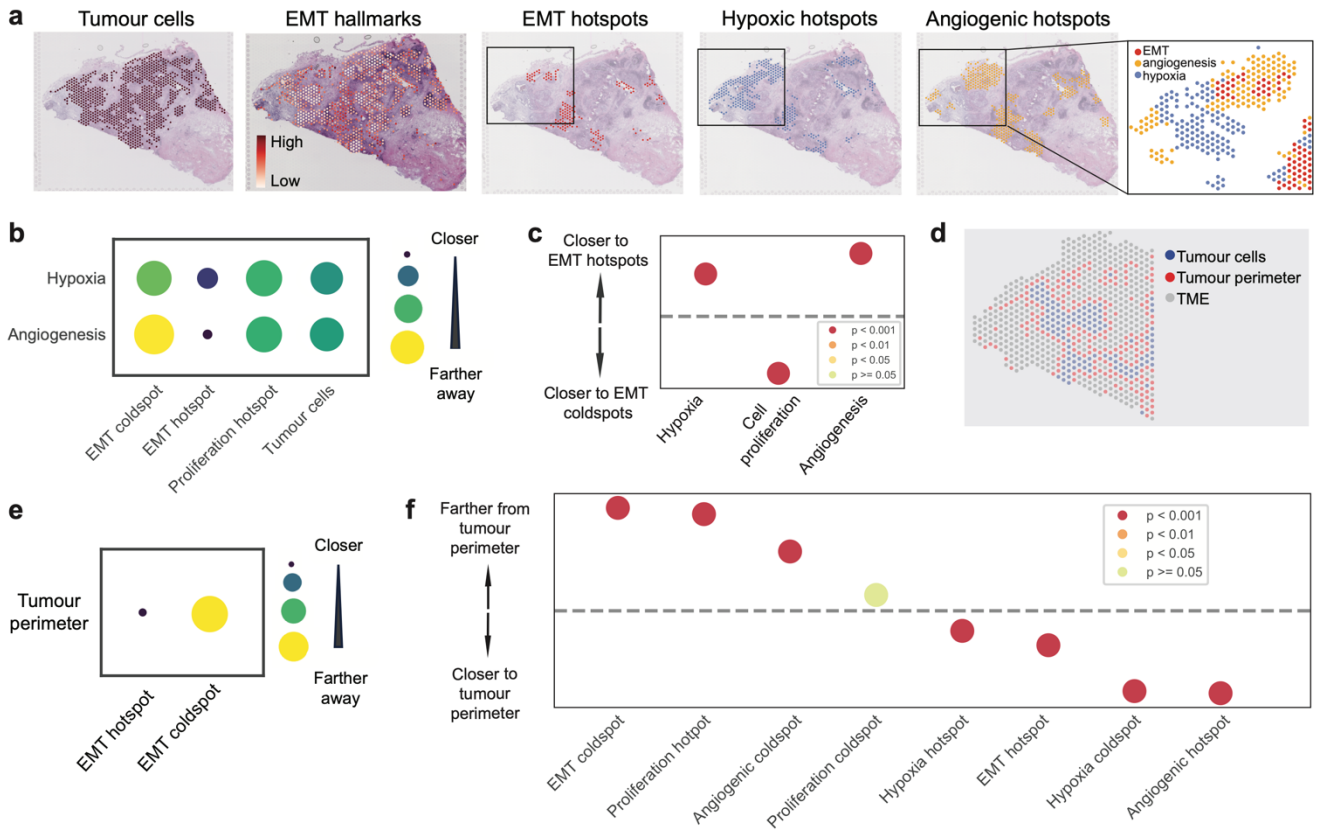

Figure 2: The spatial interplay between EMT progression and cancer hallmarks. (a) Spatial transcriptomics slide (Slide 0) highlighting from left to right: tumour spots, the EMT gradient across these tumour spots, EMT hotspots, hypoxic hotspots and angiogenic hotspots identified by SpottedPy. The black square indicates a representative area where the close proximity of EMT, angiogenic and hypoxic hotspots is depicted. **(b)** Bubble plot illustrating the distance from EMT hotspots, EMT coldspots, proliferative hotspots and the average tumour cell to angiogenic and hypoxic hotspots, averaged across all the 12 samples. Smaller bubbles represent increased vicinity. **(c)** Differences in proximity between EMT hotspots/coldspots and hypoxic, proliferative and angiogenic regions, summarised across the 12 slides. The dashed line represents no difference in relative distance to EMT hotspots or EMT coldspots. The dots situated above the dashed line indicate hallmarks that are significantly closer to EMT hotspots. The colours 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 tumour perimeter in red, tumour cells in blue and the microenvironment in grey. **(e)** Distance from the tumour perimeter to EMT hotspots and EMT coldspots, respectively. Smaller bubbles represent closer hotspots, and the data are summarised across all 12 slides. **(f)** Distances from selected hotspots to the tumour perimeter, ordered by increasing proximity, across the 12 cases. The dashed line represents no significant difference. The colours depict p-value ranges obtained from Student's t tests for differences in distance to the tumour perimeter.

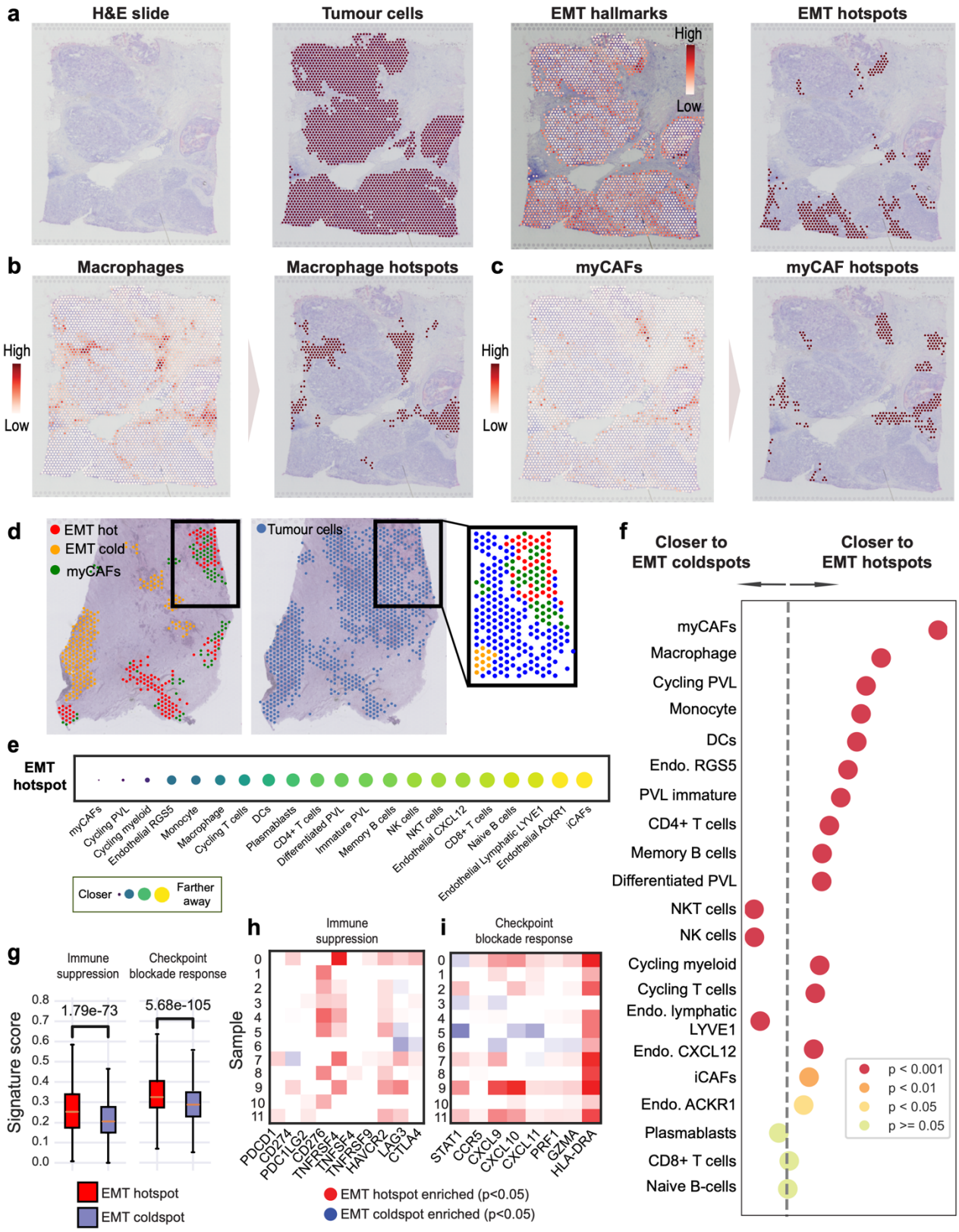

Figure 3: The spatial interplay between EMT progression and the TME. (a) Spatial transcriptomics plots highlighting tumour cell spots (left), the EMT gradient through these tumour

spots (middle) and EMT hotspots identified by SpottedPy (right) in Slide 5. **(b)** Spatial localisation of macrophage-enriched spots (left) and SpottedPy-defined macrophage hotspots (right) in Slide 5. **(c)** Spatial localisation of myCAF-enriched spots (left) and SpottedPy-defined myCAF hotspots in Slide 5. **(d)** EMT hotspots, EMT coldspots and myCAF hotspots overlayed on the same slide (left), and illustrated within the context of all tumour areas (right). The zoomed in area of interest (black square) highlights the close proximity of EMT hotspot to myCAFs. **(e**) Distance between EMT hotspots and different TME cell hotspots, ranked by proximity. Smaller bubbles represent shorter distances to EMT hotspots. Results are averaged over 12 slides. **(f)** Distances from various cells in the TME to EMT hotspots/coldspots. The dashed line represents no difference in proximity to either EMT hotspots or EMT coldspots. The dots situated to the right of the dashed line indicate cell populations that are significantly closer to EMT hotspots, ordered by decreasing proximity. The colours indicate the p-value ranges obtained from the Student's t test for differences in distance to EMT hot/cold areas. Results are averaged across 12 slides. **(g)** Barplots showing signature scores of immune suppression[12] and checkpoint inhibitor response[82] scored within EMT hotspots and coldspots. **(h)** Differences in the average expression of genes in the immune suppression signature[12] between EMT hotspots and EMT cold spots for each slide (row). Red depicts genes significantly upregulated in EMT hotspots and blue indicates genes significantly upregulated in EMT coldspots (Student's t test $p<0.05$, adjusted for multiple testing using the Bonferroni correction). White indicates a non-significant relationship. **(i)** Similar to (g) but focusing on the genes in the checkpoint inhibitor response signature[82].

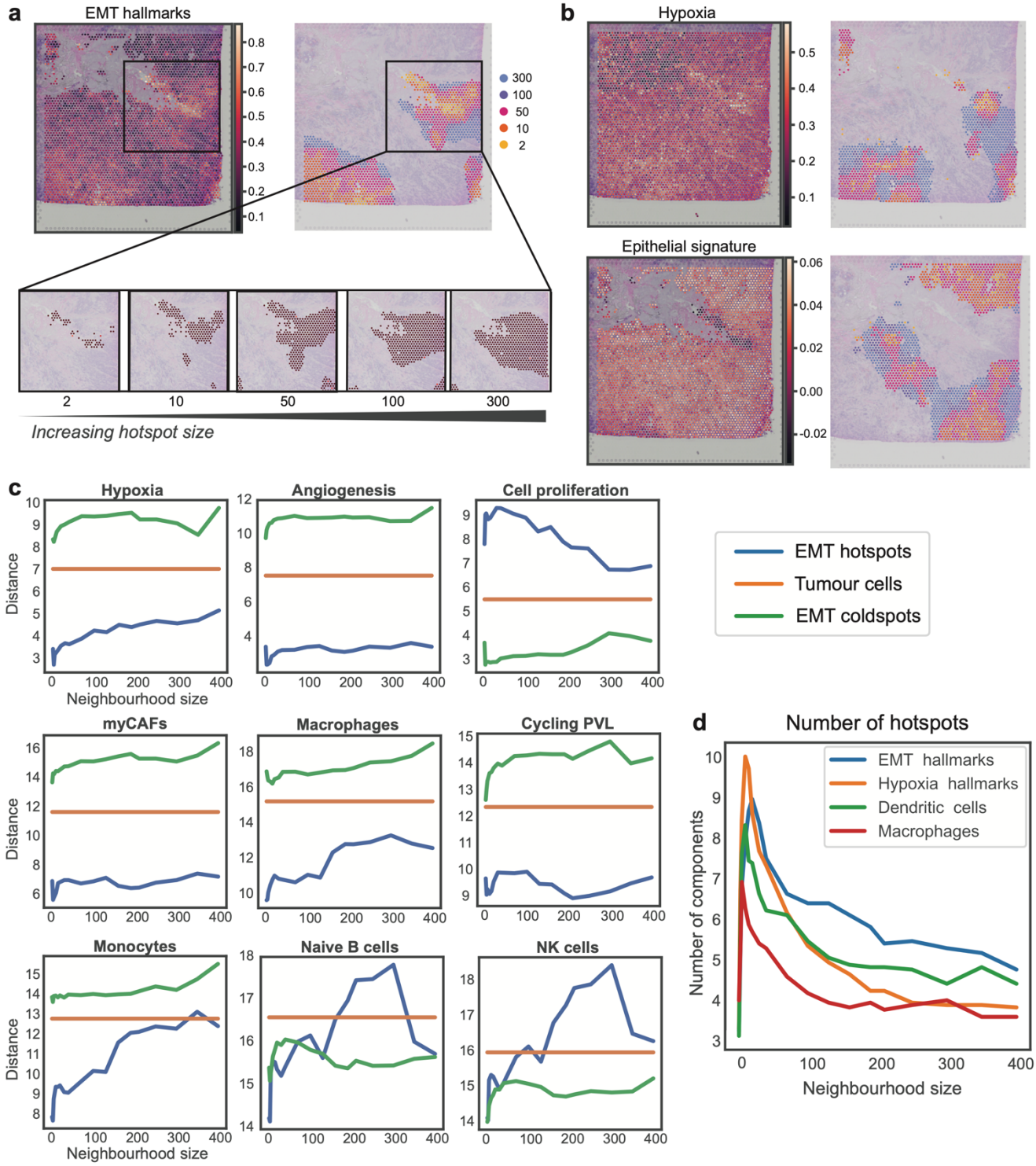

Figure 4: Sensitivity analysis of hotspot relationships. (a) EMT hotspot generation using a hotspot neighbourhood parameter of 2, 10, 50, 100 and 300, respectively. Increasingly larger neighbourhoods are highlighted in different colours as indicated in the legend. **(b)** Hypoxia and epithelial hotspot generation using a hotspot neighbourhood parameter of 2, 10, 50, 100 and 300, respectively. **(c)** Sensitivity plots highlighting the distance from EMT hotspots (blue) and EMT

coldspots (green) to regions enriched in various cancer hallmarks and TME components as the hotspot size increases. The distances to the average of all tumour cells are used as a reference (orange). Distances are averaged over all 12 slides. **(d)** Number of distinct hotspots identified as the hotspot neighbourhood parameter is increased, averaged over 12 slides.

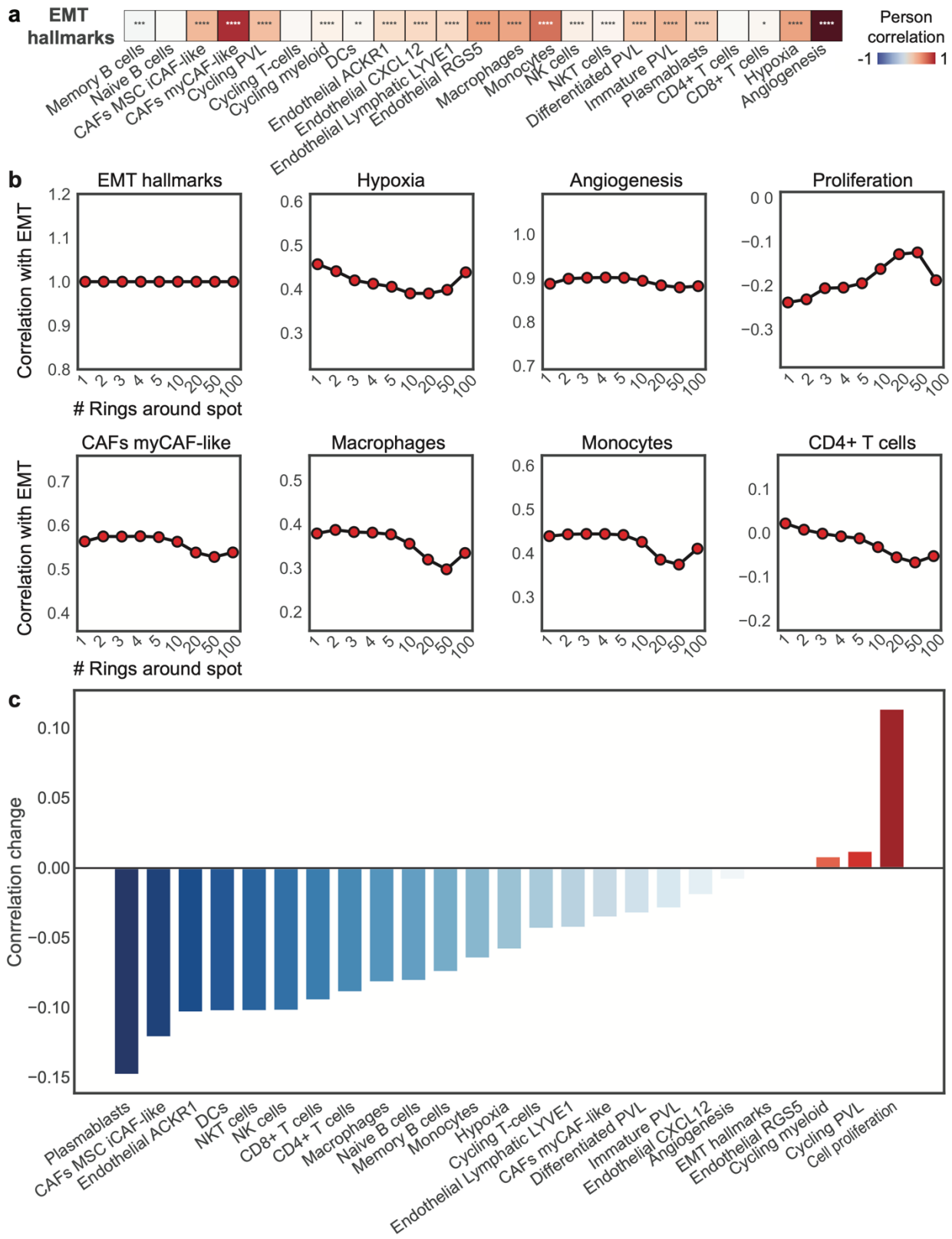

Figure 5: Neighbourhood enrichment analysis of EMT spatial dynamics. (a) Neighbourhood enrichment analysis results displaying correlations between EMT within tumour cell spots and the

775 abundance of various cell populations within the immediate TME (surrounding spots only). Red
776 indicates a positive correlation, blue a negative correlation and white a non-significant correlation
777 (Pearson $p>0.05$). 1 ring is used to define the neighbourhood. **** $p<0.0001$, *** $p<0.001$, ** $p<0.01$,
778 * $p<0.05$. **(b)** Line plots illustrating the impact of progressively expanding the number of concentric
779 rings - from 1 to 100- around a Visium spot on the correlation between the EMT signature and
780 various cells in the TME. **(c)** Bar plot highlighting the change in correlation strength as ring size is
781 increased from 1 to 100. Red indicates a positive correlation change, blue a negative change.
782

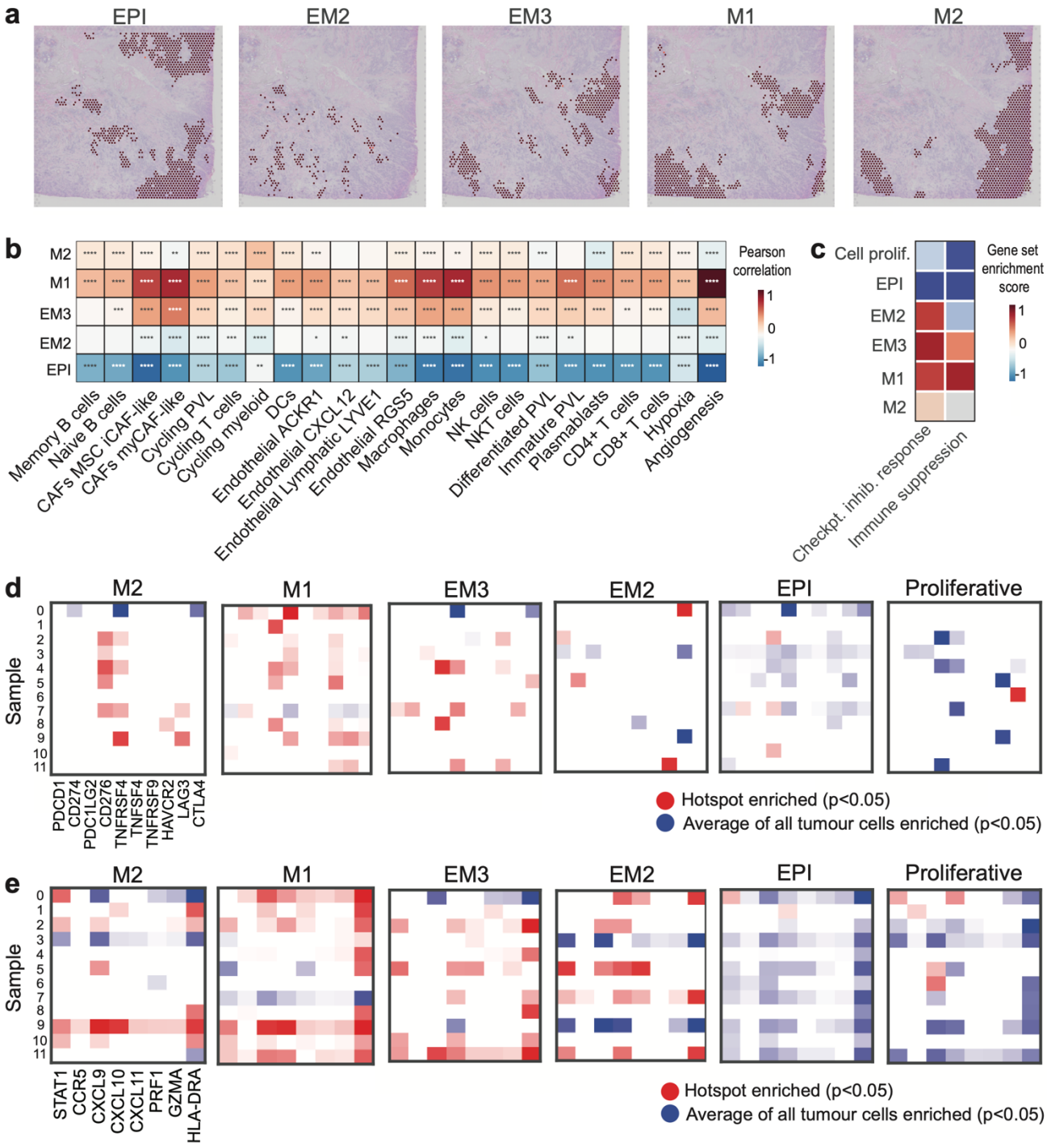

Figure 6: EMT state dynamics uncovered from the spatial exploration of breast cancer tissue. **(a)** Spatial plots depicting epithelial (EPI), intermediate (EM2, EM3), quasi-mesenchymal (M1) and fully mesenchymal (M2) hotspots in Slide 4. **(b)** Neighbourhood enrichment analysis depicting the association between tumour cells occupying distinct EMT states and other cells in the immediate TME, summarised 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 signature[82] scores calculated using Gene Set Enrichment Analysis (GSEA) for each EMT state hotspot and proliferative hotspot, summarised across the 12 samples. **(d)** Enrichment and depletion of expression for genes in the immune suppression signature within EMT state hotspots for each slide (row). Red depicts genes significantly upregulated in EMT state hotspots compared to the average of all tumour 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.

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.

Availability of data and materials

Breast cancer Visium slides were obtained from Barkley et al.[29], Wu et al.[12] and the 10x Genomics website[35].

SpottedPy is implemented as a Python package available at https://github.com/secrierlab/SpottedPy, released under a GNU General Public License v3.0 with an accompanying tutorial. Processed spatial transcriptomics data have been deposited at Zenodo: https://zenodo.org/records/10371890.

Competing interests

The authors declare no competing interests.

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).

Authors' contributions

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

845 **REFERENCES**

846 1. Yuan, Y. Spatial Heterogeneity in the Tumor Microenvironment. *Cold Spring Harb Perspect Med*
847 **6**, a026583 (2016).
848 2. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial
849 transcriptomics. *Science* (2016) doi:10.1126/science.aaf2403.
850 3. Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially Resolved Transcriptomes—Next
851 Generation Tools for Tissue Exploration. *BioEssays* **42**, 1900221 (2020).
852 4. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat Methods* **19**, 534–546 (2022).
853 5. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat Methods* **18**, 9–14 (2021).
854 6. Zhang, L. *et al.* Clinical and translational values of spatial transcriptomics. *Sig Transduct Target*
855 *Ther* **7**, 1–17 (2022).
856 7. Andersson, A. *et al.* Spatial deconvolution of HER2-positive breast cancer delineates tumor-
857 associated cell type interactions. *Nat Commun* **12**, 6012 (2021).
858 8. Kumar, T. *et al.* A spatially resolved single-cell genomic atlas of the adult human breast. *Nature*
859 **620**, 181–191 (2023).
860 9. Sibai, M. *et al.* The spatial landscape of Cancer Hallmarks reveals patterns of tumor ecology.
861 *bioRxiv* 2022.06.18.496114 (2023) doi:10.1101/2022.06.18.496114.
862 10. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell
863 Carcinoma. *Cell* **182**, 497-514.e22 (2020).
864 11. He, S. *et al.* Starfysh reveals heterogeneous spatial dynamics in the breast tumor
865 microenvironment. 2022.11.21.517420 Preprint at https://doi.org/10.1101/2022.11.21.517420
866 (2022).
867 12. Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet*
868 **53**, 1334–1347 (2021).
869 13. Lengrand, J. *et al.* Pharmacological targeting of netrin-1 inhibits EMT in cancer. *Nature* **620**,
870 402–408 (2023).

871 14. Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify
872 spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* **18**,
873 1342–1351 (2021).
874 15. Zhao, E. *et al.* Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol*
875 **39**, 1375–1384 (2021).
876 16. Walker, B. L. & Nie, Q. NeST: nested hierarchical structure identification in spatial
877 transcriptomic data. *Nat Commun* **14**, 6554 (2023).
878 17. Chitra, U. *et al.* Mapping the topography of spatial gene expression with interpretable deep
879 learning. *bioRxiv* 2023.10.10.561757 (2023) doi:10.1101/2023.10.10.561757.
880 18. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue
881 biology. *Nat Biotechnol* **40**, 308–318 (2022).
882 19. Seferbekova, Z., Lomakin, A., Yates, L. R. & Gerstung, M. Spatial biology of cancer evolution.
883 *Nat Rev Genet* **24**, 295–313 (2023).
884 20. Varrone, M., Tavernari, D., Santamaria-Martínez, A., Walsh, L. A. & Ciriello, G. CellCharter
885 reveals spatial cell niches associated with tissue remodeling and cell plasticity. *Nat Genet* 1–11
886 (2023) doi:10.1038/s41588-023-01588-4.
887 21. Moses, L. *et al.* Voyager: exploratory single-cell genomics data analysis with geospatial
888 statistics. 2023.07.20.549945 Preprint at https://doi.org/10.1101/2023.07.20.549945 (2023).
889 22. Bernstein, M. N. *et al.* Monkeybread: A Python toolkit for the analysis of cellular niches in single-
890 cell resolution spatial transcriptomics data. 2023.09.14.557736 Preprint at
891 https://doi.org/10.1101/2023.09.14.557736 (2023).
892 23. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat Methods* **19**, 171–
893 178 (2022).
894 24. Fotheringham, A. S. & Wong, D. W. S. The Modifiable Areal Unit Problem in Multivariate
895 Statistical Analysis. *Environ Plan A* **23**, 1025–1044 (1991).

896 25. Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression
897 data. *Genome Biology* **22**, 78 (2021).
898 26. He, Y. *et al.* ClusterMap for multi-scale clustering analysis of spatial gene expression. *Nat*
899 *Commun* **12**, 5909 (2021).
900 27. Li, Z. & Zhou, X. BASS: multi-scale and multi-sample analysis enables accurate cell type
901 clustering and spatial domain detection in spatial transcriptomic studies. *Genome Biol* **23**, 168
902 (2022).
903 28. Yang, J. *et al.* Guidelines and definitions for research on epithelial–mesenchymal transition. *Nat*
904 *Rev Mol Cell Biol* **21**, 341–352 (2020).
905 29. Barkley, D. *et al.* Cancer cell states recur across tumor types and form specific interactions with
906 the tumor microenvironment. *Nat Genet* **54**, 1192–1201 (2022).
907 30. Aggarwal, V., Montoya, C. A., Donnenberg, V. S. & Sant, S. Interplay between tumor
908 microenvironment and partial EMT as the driver of tumor progression. *iScience* **24**, 102113
909 (2021).
910 31. Pastushenko, I. & Blanpain, C. EMT Transition States during Tumor Progression and
911 Metastasis. *Trends in Cell Biology* **29**, 212–226 (2019).
912 32. Cook, D. P. & Vanderhyden, B. C. Context specificity of the EMT transcriptional response. *Nat*
913 *Commun* **11**, 2142 (2020).
914 33. Haerinck, J., Goossens, S. & Berx, G. The epithelial–mesenchymal plasticity landscape:
915 principles of design and mechanisms of regulation. *Nat Rev Genet* **24**, 590–609 (2023).
916 34. Malagoli Tagliazucchi, G., Wiecek, A. J., Withnell, E. & Secrier, M. Genomic and
917 microenvironmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer. *Nat*
918 *Commun* **14**, 1–20 (2023).
919 35. Home Page - 10x Genomics. https://www.10xgenomics.com/.
920 36. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics.
921 *Nat Biotechnol* **40**, 661–671 (2022).

922 37. Elyanow, R., Zeira, R., Land, M. & Raphael, B. J. STARCH: copy number and clone inference
923 from spatial transcriptomics data. *Phys Biol* **18**, 035001 (2021).
924 38. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection.
925 *Cell Syst* **1**, 417–425 (2015).
926 39. Vasaikar, S. V. *et al.* EMTome: a resource for pan-cancer analysis of epithelial-mesenchymal
927 transition genes and signatures. *Br J Cancer* **124**, 259–269 (2021).
928 40. Brown, M. S. *et al.* Phenotypic heterogeneity driven by plasticity of the intermediate EMT state
929 governs disease progression and metastasis in breast cancer. *Science Advances* **8**, eabj8002
930 (2022).
931 41. Muz, B., de la Puente, P., Azab, F. & Azab, A. K. The role of hypoxia in cancer progression,
932 angiogenesis, metastasis, and resistance to therapy. *Hypoxia (Auckl)* **3**, 83–92 (2015).
933 42. Weidemann, A. & Johnson, R. S. Biology of HIF-1α. *Cell Death Differ* **15**, 621–627 (2008).
934 43. Tam, S. Y., Wu, V. W. C. & Law, H. K. W. Hypoxia-Induced Epithelial-Mesenchymal Transition
935 in Cancers: HIF-1α and Beyond. *Front Oncol* **10**, 486 (2020).
936 44. Nishida, N., Yano, H., Nishida, T., Kamura, T. & Kojiro, M. Angiogenesis in Cancer. *Vasc Health*
937 *Risk Manag* **2**, 213–219 (2006).
938 45. Almagro, J., Messal, H. A., Elosegui-Artola, A., Rheenen, J. van & Behrens, A. Tissue
939 architecture in tumor initiation and progression. *Trends in Cancer* **8**, 494–505 (2022).
940 46. Saxena, K., Jolly, M. K. & Balamurugan, K. Hypoxia, partial EMT and collective migration:
941 Emerging culprits in metastasis. *Translational Oncology* **13**, 100845 (2020).
942 47. Biffi, G. *et al.* IL-1-induced JAK/STAT signaling is antagonized by TGF-β to shape CAF
943 heterogeneity in pancreatic ductal adenocarcinoma. *Cancer discovery* **9**, 282 (2019).
944 48. Ghahremanifard, P., Chanda, A., Bonni, S. & Bose, P. TGF-β Mediated Immune Evasion in
945 Cancer—Spotlight on Cancer-Associated Fibroblasts. *Cancers* **12**, 3650 (2020).
946 49. Principe, D. R., Timbers, K. E., Atia, L. G., Koch, R. M. & Rana, A. TGFβ Signaling in the
947 Pancreatic Tumor Microenvironment. *Cancers (Basel)* **13**, 5086 (2021).

948 50. Xu, J., Lamouille, S. & Derynck, R. TGF-β-induced epithelial to mesenchymal transition. *Cell*
949 *Res* **19**, 156–172 (2009).
950 51. myCAFs are better than yours: targeting myofibroblasts potentiates immunotherapy. *Trends in*
951 *Cancer* **9**, 1–2 (2023).
952 52. Öhlund, D. *et al.* Distinct populations of inflammatory fibroblasts and myofibroblasts in
953 pancreatic cancer. *Journal of Experimental Medicine* **214**, 579–596 (2017).
954 53. Amer, H. T., Stein, U. & El Tayebi, H. M. The Monocyte, a Maestro in the Tumor
955 Microenvironment (TME) of Breast Cancer. *Cancers* **14**, 5460 (2022).
956 54. Wu, S. Z. *et al.* Stromal cell diversity associated with immune evasion in human triple-negative
957 breast cancer. *The EMBO Journal* **39**, e104063 (2020).
958 55. Johnson, A. M. *et al.* Cancer Cell-Intrinsic Expression of MHC Class II Regulates the Immune
959 Microenvironment and Response to Anti-PD-1 Therapy in Lung Adenocarcinoma. *J Immunol*
960 **204**, 2295–2307 (2020).
961 56. Senosain, M.-F. *et al.* HLA-DR cancer cells expression correlates with T cell infiltration and is
962 enriched in lung adenocarcinoma with indolent behavior. *Sci Rep* **11**, 14424 (2021).
963 57. Sherman, T. D., Gao, T. & Fertig, E. J. CoGAPS 3: Bayesian non-negative matrix factorization
964 for single-cell analysis with asynchronous updates and sparse data structures. *BMC*
965 *Bioinformatics* **21**, 453 (2020).
966 58. Spatial and single-nucleus transcriptomic analysis of genetic and sporadic forms of Alzheimer's
967 Disease | bioRxiv. https://www.biorxiv.org/content/10.1101/2023.07.24.550282v1.
968 59. Ospina, O. E. *et al.* spatialGE: quantification and visualization of the tumor microenvironment
969 heterogeneity using spatial transcriptomics. *Bioinformatics* **38**, 2645–2647 (2022).
970 60. Emami Nejad, A. *et al.* The role of hypoxia in the tumor microenvironment and development of
971 cancer stem cell: a novel approach to developing treatment. *Cancer Cell Int* **21**, 62 (2021).

972 61. Wang, X., Zhang, W., Sun, X., Lin, Y. & Chen, W. Cancer-associated fibroblasts induce
973 epithelial-mesenchymal transition through secreted cytokines in endometrial cancer cells. *Oncol*
974 *Lett* **15**, 5694–5702 (2018).
975 62. Jia, C. *et al.* Cancer-associated Fibroblasts induce epithelial-mesenchymal transition via the
976 Transglutaminase 2-dependent IL-6/IL6R/STAT3 axis in Hepatocellular Carcinoma. *Int J Biol*
977 *Sci* **16**, 2542–2558 (2020).
978 63. Tyler, M. & Tirosh, I. Decoupling epithelial-mesenchymal transitions from stromal profiles by
979 integrative expression analysis. *Nat Commun* **12**, 2592 (2021).
980 64. Li, X., Chen, L., Peng, X. & Zhan, X. Progress of tumor-associated macrophages in the
981 epithelial-mesenchymal transition of tumor. *Front Oncol* **12**, 911410 (2022).
982 65. Brabletz, S., Schuhwerk, H., Brabletz, T. & Stemmler, M. P. Dynamic EMT: a multi-tool for
983 tumor progression. *The EMBO Journal* **40**, e108647 (2021).
984 66. Johansson, J., Tabor, V., Wikell, A., Jalkanen, S. & Fuxe, J. TGF-β1-Induced Epithelial–
985 Mesenchymal Transition Promotes Monocyte/Macrophage Properties in Breast Cancer Cells.
986 *Frontiers in Oncology* **5**, (2015).
987 67. Kmieciak, M., Knutson, K. L., Dumur, C. I. & Manjili, M. H. HER-2/neu antigen loss and relapse
988 of mammary carcinoma are actively induced by T cell-mediated anti-tumor immune responses.
989 *Eur J Immunol* **37**, 675–685 (2007).
990 68. Santisteban, M. *et al.* Immune-induced epithelial to mesenchymal transition in vivo generates
991 breast cancer stem cells. *Cancer Res* **69**, 2887–2895 (2009).
992 69. Galsky, M. D. *et al.* 850PD - Epithelial-mesenchymal transition (EMT), T cell infiltration, and
993 outcomes with nivolumab (nivo) in urothelial cancer (UC). *Annals of Oncology* **28**, v297 (2017).
994 70. Chockley, P. J. *et al.* Epithelial-mesenchymal transition leads to NK cell–mediated metastasis-
995 specific immunosurveillance in lung cancer. *J Clin Invest* **128**, 1384–1396.
996 71. Datar, I. & Schalper, K. A. Epithelial–Mesenchymal Transition and Immune Evasion during Lung
997 Cancer Progression: The Chicken or the Egg? *Clin Cancer Res* **22**, 3422–3424 (2016).

998 72. Terry, S. *et al.* New insights into the role of EMT in tumor immune escape. *Mol Oncol* **11**, 824–
999 846 (2017).
1000 73. Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal regulation of
1001 epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer*
1002 *Cell* **22**, 725–736 (2012).
1003 74. Cheng, Y.-C. *et al.* Reconstruction of single cell lineage trajectories and identification of diversity
1004 in fates during the epithelial-to-mesenchymal transition. *bioRxiv* 2023.09.19.558325 (2023)
1005 doi:10.1101/2023.09.19.558325.
1006 75. Sinha, D., Saha, P., Samanta, A. & Bishayee, A. Emerging Concepts of Hybrid Epithelial-to-
1007 Mesenchymal Transition in Cancer Progression. *Biomolecules* **10**, 1561 (2020).
1008 76. Subhadarshini, S., Markus, J., Sahoo, S. & Jolly, M. K. Dynamics of Epithelial–Mesenchymal
1009 Plasticity: What Have Single-Cell Investigations Elucidated So Far? *ACS Omega* **8**, 11665–
1010 11673 (2023).
1011 77. Zhang, Y. *et al.* Hypoxia in Breast Cancer—Scientific Translation to Therapeutic and Diagnostic
1012 Clinical Applications. *Frontiers in Oncology* **11**, (2021).
1013 78. Fu, Z., Mowday, A. M., Smaill, J. B., Hermans, I. F. & Patterson, A. V. Tumour Hypoxia-
1014 Mediated Immunosuppression: Mechanisms and Therapeutic Approaches to Improve Cancer
1015 Immunotherapy. *Cells* **10**, 1006 (2021).
1016 79. Lee, C.-T., Mace, T. & Repasky, E. A. Hypoxia-Driven Immunosuppression: A new reason to
1017 use thermal therapy in the treatment of cancer? *Int J Hyperthermia* **26**, 232–246 (2010).
1018 80. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data
1019 analysis. *Genome Biology* **19**, 15 (2018).
1020 81. Stein-O'Brien, G. L. *et al.* Decomposing Cell Identity for Transfer Learning across Cellular
1021 Measurements, Platforms, Tissues, and Species. *Cell Systems* **8**, 395-411.e8 (2019).
1022 82. Ayers, M. *et al.* IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin*
1023 *Invest* **127**, 2930–2940 (2017).

1024 83. Nielsen, T. O. *et al.* A Comparison of PAM50 Intrinsic Subtyping with Immunohistochemistry
1025 and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor–Positive Breast
1026 Cancer. *Clinical Cancer Research* **16**, 5222–5232 (2010).
1027 84. Proceedings of the Python in Science Conference (SciPy): Exploring Network Structure,
1028 Dynamics, and Function using NetworkX.
1029 https://conference.scipy.org/proceedings/SciPy2008/paper_2/.
1030 85. Rey, S. J. & Anselin, L. PySAL: A Python library of spatial analytical methods. *The Review of*
1031 *regional studies* **37**, 5–27 (2007).
1032