SpottedPy quantifies relationships between spatial transcriptomic hotspots and uncovers new environmental cues of epithelial-mesesenchymal plasticity in breast 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 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 tumour 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 followed transformation gradients reflecting progressive immunosuppression. Our method offers flexibility to explore spatial relationships at different scales, from immediate neighbours to broader tissue modules, providing new insights into tumour microenvironment dynamics.

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 (TME)(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 spatial 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, 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 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 (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 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. 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, polarised epithelial cells undergo multiple molecular changes and lose their identity to acquire a mesenchymal phenotype(27). 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(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 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(33). 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),(28), 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-ii**). 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. We offer functionality to test how a signature effects its direct neighbourhood (inner-outer correlation), or within neighbourhood units (all-in-one correlation).

- **Hotspot identification:** We have implemented the Getis-Ord G* statistic to identify spatial clusters of continuous gene signatures across the tumour 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 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 tumour 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. 1biii**). 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. 1biv).** Importantly, SpottedPy assigns a statistical significance to these proximity measures to determine if observed patterns are likely to occur just by chance. To statistically analyse the relationships across multiple slides, we use generalised 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 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(35) and a scRNA-seq reference of annotated breast cancer cell population profiles from 123,561 cells[13]. We scored the tumour cells in the scRNA-seq dataset with a defined epithelial (EPI) and an epithelial-to-mesenchymal transition (EMT) signature (see Methods), and used Gaussian mixture modelling to assign a state to the tumour cells (36,37) . To more precisely identify tumour 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 tumour-specific. We validated the STARCH results by comparing to publicly available pathologist annotated slides (39,40) **(Supplementary Fig. 1a-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 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(42). Under hypoxic conditions, tumour cells stabilise hypoxia-inducible factors (HIFs), primarily HIF-1α, 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 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 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 **(Supplementary Fig. 1c).** Notably, Slides 3, 7 and 10 exhibited the greatest dispersion, which was independent of cancer subtype (TNBC and ER+

HER2+). This dispersion did not influence overall EMT relationships with other cell types **(Supplementary Fig. 2a)**. 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 **Supplementary Fig. 1c-d**. The EPI signature typically followed a normal distribution across spatial tumour spots, whereas the EMT signature exhibited a positive skew, in line with this state being expected to be rarer within the primary tumours. Slides 0, 1, and 9 display more pronounced skewness for the EPI signature, suggesting a more advanced stage of EMT transformation within these tumours, which was not an effect of the cancer subtype as these slides present different breast cancer pathologies.

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 that SpottedPy accurately measures hotspot distance using the shortest path approach, we simulated a hypoxia hotspot moving away from a mesenchymal hotspot of interest (**Supplementary Fig. 1e**). 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 tumour 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 (**Supplementary Fig. 2a**).

To further grasp the positioning of these EMT and EPI areas within the tumour, we used SpottedPy to determine the tumour perimeter (**Fig. 2d**) and calculated distances to it. We conducted a visual benchmark of our tumour perimeter calculation against the Cottrazm method(40), confirming that our algorithm accurately captures similar perimeter trends (**Supplementary Fig. 2b**). We find EMT hotspots closer to the tumour 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 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(45). 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(42). 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 farthest 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(46,47).

240 **EMT hotspots are immunosuppressed and shielded by myCAFs and macrophages**
241
242 Having confirmed that SpottedPy is able to recapitulate expected spatial hallmarks of EMT within the
243 breast cancer tissue, we next expanded our analysis to dissect the interplay between tumour cells
244 undergoing EMT and other immune and stromal cells in the microenvironment. Alongside EMT
245 hotspots, we calculated hotspots for 41 cell types in the TME, including different lymphocyte, myeloid
246 and fibroblast populations, as defined by Wu et al(12) (**Fig. 3a-c**). When visually inspecting these
247 hotspots, we observed myofibroblastic CAF (myCAF) hotspots and EMT hotspots tended to co-
248 localise (**Fig. 3a-c**). Quantifying hotspot distances using SpottedPy allowed us to confirm that tumour
249 EMT hotspots were indeed closer to myCAF hotspots (**Fig. 3d**). The relationship is particularly
250 highlighted when we look at the cellular niches that are significantly closer to EMT hotspots compared
251 to EPI hotspots, revealing a predominance of various CAF subtypes. This is well in line with existing
252 studies, as myCAFs have been shown to produce TGF-β, which is a well-known EMT trigger(48).
253 They have also been linked to ECM deposition and suppression of antitumor
254 immunity(49),(50),(51),(52).
255
256 Additionally, monocytes and particularly tumour-associated macrophages (TAMs) (LAM2+ APOE+
257 macrophages and SIGLEC1+ macrophages) were prominently closer to EMT hotspots as compared
258 to EPI hotspots. Monocytes, and TAMs derived from them, are known to modulate the environment
259 of tumour cells undergoing EMT, usually by promoting immune suppression in the TME, which
260 facilitates tumour progression and metastasis(53). Natural Killer (NK), NK T cells and CD8+ T cells,
261 the immune cells that can directly kill transformed cells, were amongst the least associated with EMT
262 hotspots, potentially reflecting a mechanism of immune evasion employed by tumour cells that have
263 undergone EMT (**Fig. 3f**). The T-cell sub-population closest to EMT hotspots when compared to EPI
264 were LAG3+ CD8+ T cells, an exhausted population suggestive of immune evasion capacity in these
265 EMT areas (54).
266
267 The cohort-level association between EMT hotspots and the myCAF s5 population was maintained
268 in individual tumours, suggesting that this is a universal pattern of EMT transformation in breast
269 cancer and not subtype-specific (**Supplementary Fig. 2a**). In contrast, levels of inter-patient
270 heterogeneity, often even within the same breast cancer subtype, were evident for a variety of cells
271 including macrophages, memory B-cells, naïve B-cells, iCAFs, NK cells, NKT cells, CD4+ T-cells and
272 CD8+ T-cells. However, distances to EMT hotspots were consistent across subgroups of cells
273 (**Supplementary Fig. 2c**), suggesting that, within individual patients, these cells share common
274 response patterns irrespective of the broader heterogeneity observed across the patient cohort.
275
276 Due to the close relationship with potential immunosuppressive factors, we next sought to test
277 whether EMT hotspots were indeed likely to be immunosuppressed. We found that EMT hotspots
278 displayed a significantly increased expression of immunosuppression and exhaustion markers (12,55)
279 compared to EPI hotspots (**Fig. 3f-g**), Highly expressed suppressive genes include *FAP*, which has
280 been shown to activate immune suppressive cells such as regulatory T cells (Tregs) and myeloid-
281 derived suppressor cells (MDSCs) (56,57), *INHBA*, which fosters a switch in macrophage polarisation
282 towards a tumour-promoting state (58), *VCAN*, which has been shown to inhibit T-cell proliferation(59)
283 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 the tumour slides (**Fig. 3g**). Indeed, EMT hotspots presented increased exhaustion **(Supplementary Fig. 2d-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 favourable 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 analysed the association of EMT hotspots with other immune and stromal areas within the same slide and across the different patient samples **(Fig. 4a).** Whilst the cells that had the strongest relationship with EMT hotspots when averaged over the slides, such as SIGLEC+ and LAM2+ APOE+ macrophages and CAFs, displayed the most consistent trends across the slides, it was evident that these relationships were still heterogenous. For example, in slide 4, whilst seven EMT hotspots were closer to LAM2+ APOE+ 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 LAM2+ APOE+ macrophage hotspots, enabled through SpottedPy functions **(Fig. 4b).** As expected, stronger associations with myCAFs and macrophage subtypes were pervasive in most tumours and hotspots **(Fig. 4a).** T-cells demonstrated noticeable heterogeneity across patients, but clustered together, 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 tumour 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).** Visualising these distributions highlights that whilst 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 neighbours for the Getis-Ord Gi* 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 tumour 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**, **Supplementary Fig. 3a**). 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 Gi* highlighted similar relationships (**Supplementary Fig. 3b**). 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 Methods). 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 (**Supplementary Fig. 3c**), the hotspots are significantly smaller than those that are biologically relevant (**Supplementary Fig. 3d**). 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).

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 **Supplementary Fig. 3e**, 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, neighbouring hotspot. Thus, the centroid approach may miss 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.

371 **Spatial EMT relationships in other cancer types**
372
373 We further investigated whether the relationships observed for EMT hotspots in breast cancer were
374 consistent across other cancer types. We assessed these relationships in publicly available datasets
375 from basal cell carcinomas (BCC)(67), pancreatic ductal adenocarcinomas (PDAC)(68) and colorectal
376 cancers (CRC)(69).
377
378 Within the BCC slides, angiogenetic and hypoxic hotspots were closer to EMT hotspots
379 **(Supplementary Fig. 4a-b).** Interestingly, proliferative hotspots were also closer to EMT hotspots,
380 suggesting an alternative relationship compared to breast cancer. POSTN+ fibroblasts were closer to
381 EMT hotspots, whilst there were no significant spatial relationships with T cells or NK cells, paralleling
382 the findings in breast cancer.

383 We next assessed the relationships within one available PDAC slide **(Supplementary Fig. 4c).** Both
384 angiogenesis and fibroblasts displayed a spatial relationship like that observed with EMT hotspots in
385 breast cancer. In contrast, immune cells were located nearer to EMT coldspots, and there was no
386 significant association between EMT hotspots and hypoxic environments, diverging from the patterns
387 observed in breast cancer.

388 In CRC, myofibroblasts, angiogenesis, and hypoxia showed comparable spatial relationships to those
389 seen in breast cancer **(Supplementary Fig. 4d-e).** Notably, regulatory T-cells, T-helper cells, and NK
390 cells were significantly closer to EMT hotspots, potentially indicating enhanced immune recognition
391 around EMT hotspots compared to other cancer types.
392
393 While limited in breadth and cell type resolution, these analyses suggest that the interplay between
394 tumour cells undergoing EMT and other immune and stromal cells within the TME is likely to be tissue
395 specific, and future work should explore this in more detail.
396
397 **Neighbourhood enrichment analysis**
398
399 The neighbourhood enrichment technique captures more localised, shorter-range relationships with
400 the TME. Additionally, it can assess spatial relationships of phenotypes that would be considered
401 scattered (states that do not occur spatially clustered and therefore might be overlooked by a hotspot-
402 based approach). We experimented with two approaches, ensuring a robust analysis that is less
403 sensitive to the MAUP (**Fig. 1bi-ii)**. We first assessed how the spatial relationships change by
404 correlating phenotypes across a central tumour spot and the direct neighbourhood surrounding it (a
405 ring encompassing six Visium spots). We then assessed how the phenotypes are linked within a spot
406 and then expanding what is considered a spatial spot**.** Varying the method and the number of rings in
407 both cases enables us to assess whether the observed hotspot relationships shift with the unit of
408 analysis and indicates how large of an influence the EMT regions have on surrounding spots.
409
410 Our analysis revealed that angiogenesis, myCAFs, macrophages and monocytes exhibited the most
411 significant correlation, in descending order, with cells undergoing EMT ($p<0.001$) across the 12 slides
412 (**Supplementary Fig. 5a**). 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 neighbourhood sizes (**Supplementary Fig 5b**).

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 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(70) (**Supplementary Fig. 6a**). The corresponding hotspots occupied distinct spatial locations within the tissue (**Supplementary Fig. 6b**). When visually inspecting these hotspots, we detected a progressive transformation in the tumour, 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-localised with the EPI state. The experimental study by Brown et al(41) 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 with the generic EMT hallmark signature employed previously for each tumour spot (**Supplementary Fig. 7a**). 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 the EMT hallmark 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. 7b**), 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. 7c**), 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. 7d**). 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”-type of analysis to spatial transcriptomic data(71),(72),(21), these methods do not offer a way to assess the confidence level in the identification of specific clusters/hotspots, 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 analyse the distances between hotspots. We build on these approaches to analyse 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 tumour cells or non-tumour 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 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(73),(44),(47).

We find a strong relationship between EMT and CAFs across all slides. CAFs have been linked to tumour 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 tumour settings(76). Here we have leveraged scRNA-seq for deconvolution, alongside detecting copy number aberrations to confirm the presence of tumour 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 tumour cells themselves by investigating signatures only within the tumour spots identified by STARCH and by using EMT reference datasets from pure tumour populations, further validation of tumour 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, LAM2+ APOE macrophages and EGR1+ macrophages, which are analogous to M2-like, tumour promoting, macrophages(12). Macrophages secrete TGF-β, TNF-α, IL-6 and IL-8, which are well-characterised 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). Whilst the relationships with CAFs and macrophages were displayed across the majority of tumour slides, within each slide there were EMT hotspots where this relationship was less clear. This indicates that other factors which we have not accounted 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(77),(52),(84). Indeed, our findings uncover notable associations between EMT hotspots

544 and immune suppression, alongside signatures indicative of a response to checkpoint therapy,
545 building upon evidence that EMT may offer crucial insights for existing strategies in
546 immunotherapy(85),(11).
547
548 Our analysis reveals that EMT occurs in discrete spatial locations distinct from proliferative signatures.
549 This finding is in line with a recent analysis of breast cancer by Jia et al[51] utilising a more focused
550 spatial transcriptomic dataset, previous research by Tsai et al(86) demonstrating that a departure
551 from a mesenchymal-like state is a prerequisite for tumour cell proliferation in mouse models, and a
552 recent study by Chen et al(87) investigating EMT states in scRNA-seq data. Such spatial
553 characterisations at various scales were largely unexplored.
554
555 Delving deeper into EMT, we observed that hybrid EMT states exhibit more heterogeneous and
556 weaker associations with tumour-promoting populations in the TME in comparison to the quasi-
557 mesenchymal M1 state. This disparity might be indicative of the inherent plasticity of hybrid EMT
558 states(88),(89), complicating our ability to delineate clear relationships, but might also suggest a
559 directed trajectory towards an M1 state. The M2 state demonstrates more similar distribution and TME
560 associations to the EPI state, which may be attributed to the activation of integrin β4 (a key epithelial
561 marker) when cultured, a limitation mentioned in the original study which potentially transformed the
562 state towards a more epithelial phenotype(41).
563
564 These results point to a highly dynamic and plastic nature of tumour cells in navigating the
565 complexities of their microenvironment. The interactions likely extend beyond a linear framework.
566 Hypoxia, a known catalyst for both angiogenesis and EMT(90), can set off a cascade of events that
567 not only amplify these processes but also create a conducive environment for the recruitment of
568 immunosuppressive cells such as macrophages(91),(92). These cells can in turn bolster
569 angiogenesis, thereby fuelling a self-perpetuating cycle that further complicates the tumour
570 landscape. These insights provide a further understanding of the cellular interactions and
571 environmental factors that underpin tumour progression and metastasis, and could in the future pave
572 the way for the development of targeted interventions aimed at disrupting these complex networks for
573 therapeutic benefit.
574
575 The consistency of the relationships we observed across different hotspot sizes and neighbourhood
576 scales further strengthens our confidence in the findings. The neighbourhood ring approach
577 predominantly captured TME cells that have infiltrated the tumour, offering insights into the immediate
578 cellular interplay at the tumour periphery. In contrast, the hotspot methodology provided a broader
579 view, encompassing interactions at more distal locations. By pinpointing statistically significant cellular
580 hotspots, we bolster the reliability of our observations, especially considering the inherent
581 inaccuracies that can arise from deconvolution algorithms applied to non-single cell transcriptomic
582 datasets such as those from the Visium platform.
583
584 To the best of our knowledge there are no direct comparisons available for our SpottedPy
585 methodology due to the unique nature of focusing on discrete spatial clusters of user-defined
586 continuous signatures at expanding scales, and performing differential spatial relationships compared
587 to a reference for downstream analysis.

Our insights into the spatial organisation of tumours during EMT progression are limited by the significant amount of uncertainty surrounding EMP programmes (31),(32) and their incomplete characterisation in different types of breast cancer and other neoplasms. In the future, integrating further hybrid states characterised 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 signalling information into the evaluation of spatial effects on cell populations will increase the confidence in the identified relationships.

Overall, our findings confirm expected spatial effects of EMT progression in tumours, demonstrating that SpottedPy can 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 primarily illustrate insights into breast cancer tissue organisation, 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.

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

The SpottedPy package

SpottedPy package is compatible with Python 3.9 and depends on scanpy, libpysal, 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 10x Genomics (slides 3-5)(34) 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(93). 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. Pre-processed BCC slides were obtained from Gania et al.(67), PDAC slides obtained from Ma et al.(68) and CRC slides 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 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 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 modelling to assign the cells to EPI and EMT clusters.

669 The scRNA regression model was trained with 500 epochs, and the spatial transcriptomic model
670 trained with 20,000 epochs using a GPU. To delineate the tumour cells within our spatial
671 transcriptomics dataset, we used the STARCH Python package designed to infer copy number
672 alterations (CNAs)(38). STARCH identifies tumour clones (setting *K*=2 clones) and non-tumour spots.
673 It confirms identification of normal spots by clustering the first principal component into two clusters
674 using K-means. Changing the value of *K* alters the number of identified tumour clones, but the number
675 of cells labelled as tumour cells remains the same. This approach is based on the principle that the
676 direction of maximum variance in the expression data typically reflects the division between non-
677 cancerous and cancerous spots. Only tumour cell spots were considered for EMT analysis. The EPI
678 and EMT spots identified using Cell2location were used to define the EPI and EMT hotspots in the
679 breast cancer downstream analysis.
680
681 **EMT state and hallmark signature scoring**
682
683 To identify more granular distinct EMT states, we employed data from Brown et al(41), consisting of
684 seven RNA-seq sequenced cell clones, derived from SUM149PT inflammatory breast cancer cell line
685 with 3 repeats spanning the EMT spectrum from epithelial-like (EPI), quasi-mesenchymal (M1), fully
686 mesenchymal (M2) and three distinct intermediates (EM1, EM2, EM3) . We used these data to derive
687 a weighted gene signature to represent the EMT states. We captured EMT gene patterns from this
688 data using non-negative matrix factorisation (NMF) by applying the CoGAPs workflow(70). We used
689 ProjectR's implementation of *lmfit* R function to map the captured EMT patterns onto the spatial
690 transcriptomic spots(94). This transfer learning approach assumes that if datasets share common
691 latent spaces, a feature mapping exists between them and can measure the extent of relationships
692 between the datasets. The final states were captured with 20 patterns and 10,000 training iterations.
693 The number of patterns were chosen based on capturing the discrete states with the highest accuracy.
694 The EM1 state was not distinguishable from the EPI state, so we merged the two states. Thus, overall
695 we obtained scores for one epithelial, two intermediate, a quasi-mesenchymal and a fully
696 mesenchymal state for each spot.
697
698 Hypoxia and angiogenesis were defined based on signatures deposited at MSigDB(95). The
699 proliferative signature was compiled from Nielsen at al(96). The immunosuppresion signature was
700 compiled from Wu et al(12) and Cui et al(55). The checkpoint blockade response signature was
701 compiled from Johnson et al(61) and Liu et al(62). The exhaustion signature comprised classical
702 exhaustion markers: *CTLA4, PDCD1,TIGIT, LAG3, HAVCR2, EOEMT, TBX21, BTLA, CD274,*
703 *PTGER4, CD244* and *CD160*(97). All these signatures were scored using *scanpy.tl.score_genes*
704 function. EMT hotspots and coldspots were identified in the BCC, CRC and PDAC slides using the
705 EMT hallmark signature(95).
706
707 **Graph construction**
708
709 The SquidPy(23) (Spatial Single-Cell Analysis in Python) package was used for graph construction
710 using *sq.gr.spatial_neighbors* and slide visualisation of the Visium spatial slides. NetworkX(98) was
711 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 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.

Two methods were developed to assess neighbourhood 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 neighbourhood of spots surrounding it (a ring encompassing six Visium spots), after filtering for tumour 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 neighbours and compared the change in 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 neighbourhood results.

Hotspot analysis

Hotspots were calculated using The Getis-Ord G* statistic as implemented using the PySAL package(99), using 10 as neighbourhood size parameter by default (number of spot neighbours surrounding the central Visium spot) 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. 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.

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. tumour cells. The *neighourhood_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 labelled 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 labelled using *.obs['batch']* within the *anndata* object. Additionally, the library ID in the *.uns* data slot should be labelled 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 tumour cells as labelled by STARCH, as we aimed to specifically capture these processes within the tumour cells themselves. EMT hotspots are the regions with a high proportion of mesenchymal tumour cells within the tumour-labelled spots. Therefore, they only include a subset of tumour spots. Similarly, cell proliferation hotspots are regions with high fractions of proliferating tumour 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 cells 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 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, 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 geneal 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 Generalised 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 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.

Tumour perimeter calculation

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

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

Where S denotes the set of all spots, T denotes the set of tumour spots $N(s)$ represents the neighbouring 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 neighbouring perimeter spots, thereby excluding isolated perimeter spots caused by a single non-tumour labelled spot within the tumour. 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 (called using *sp.calculate_tumour_perimeter*).

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

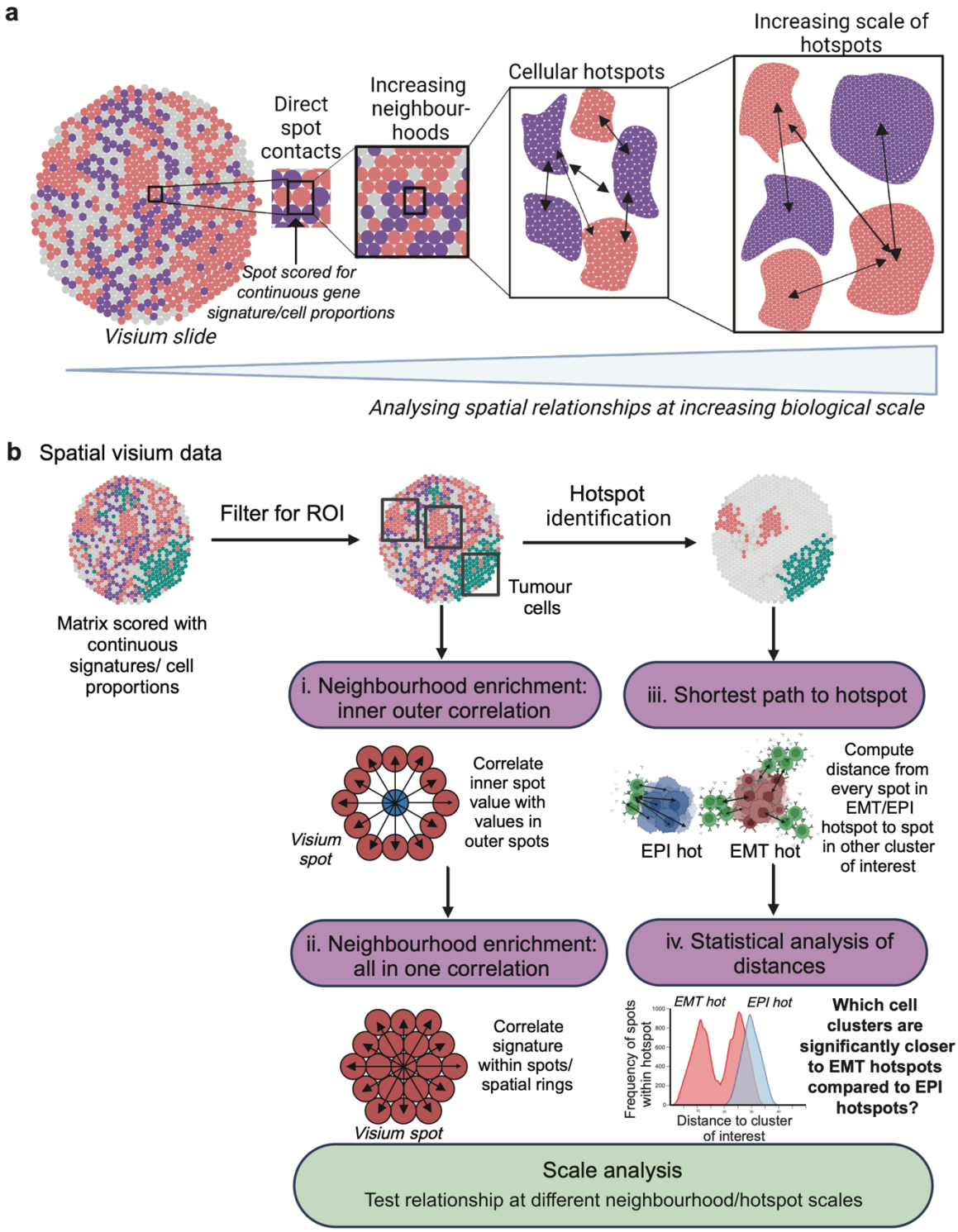
The sensitivity analysis to evaluate the impact of varying hotspot sizes on the spatial relationships was achieved by incrementally adjusting the neighbourhood parameter for the Getis-Ord statistic, which directly influenced the size of identified hotspots (*sp.sensitivity_calcs)*. As we expanded the neighbourhood 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, characterised 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.

875 **FIGURES**



876

877 **Figure 1: SpottedPy provides a multi-scale approach to analyse spatial transcriptomic**
878 **relationships. (a)** Overview of spatial scales captured in the SpottedPy workflow, from direct cellular
879 contacts to broader cellular hotspots. Figure created with BioRender.com. **(b)** SpottedPy workflow

880 overview. Visium spatial transcriptomic data is loaded as a pre-processed *AnnData* object where there
881 is the option to select the region of interest (ROI) within the slide e.g., *AnnData.obs* column labelled
882 with tumour cells. The default spatial analytics include: (i) Neighbourhood enrichment: inner outer
883 correlation, which correlates cell prevalence or signatures in individual spots with their immediate
884 neighbourhood, (ii) Neighbourhood enrichment: all in one correlation, which correlates cell prevalence
885 of signatures within a spot or spatial unit (iii) Shortest path to hotspot, which calculates the minimum
886 distance between each spot within a selected hotspot and the nearest spot in other hotspots, (iv)
887 Statistical analysis of distances, which compares distances from a reference hotspot to another
888 hotspot of interest, and assesses the statistical significance of the relationships. Scale analysis allows
889 us to compare relationships defined at different scales in both approaches, either by increasing the
890 number of rings included for neighbourhood enrichment or increasing the hotspot size. The outputs
891 for the different modules include various plots to highlight the relationships. Figure created with
892 BioRender.com.
893

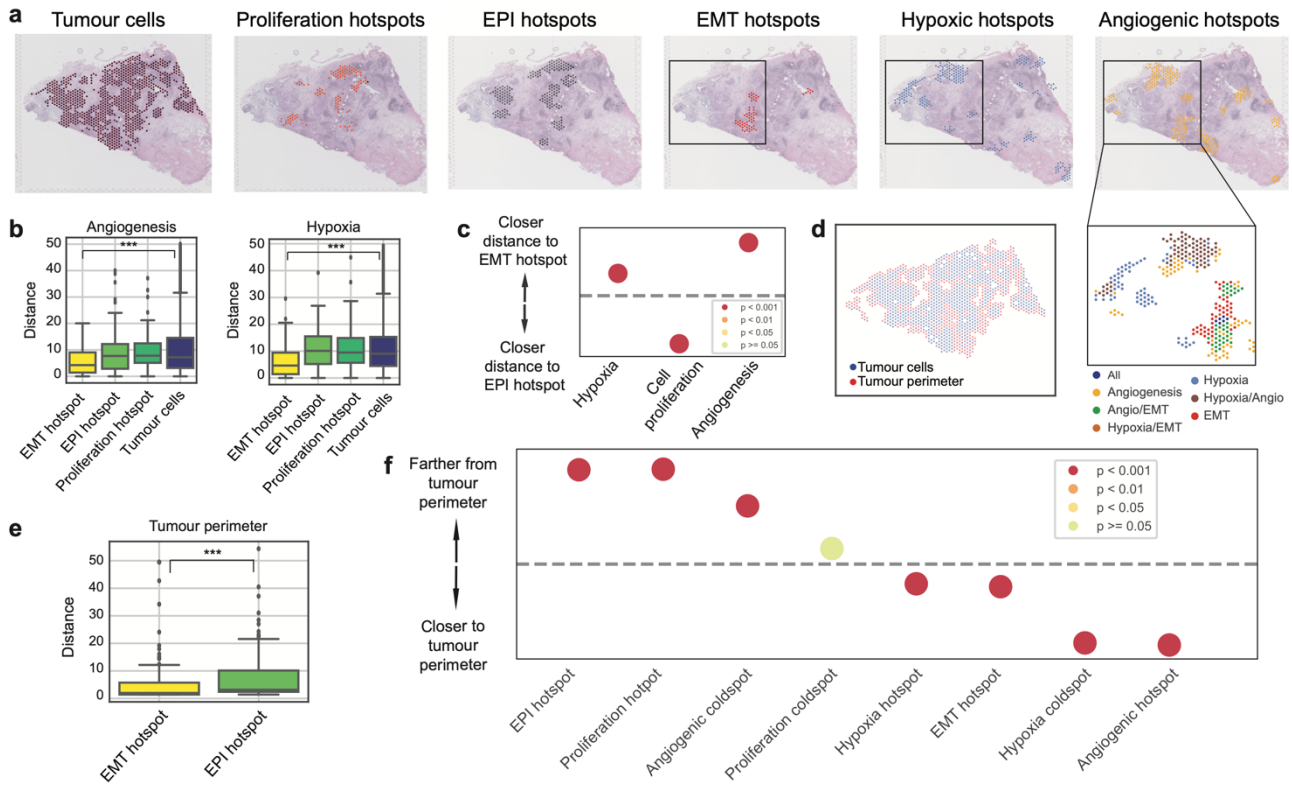


Figure 2: The spatial interplay between EMT progression and cancer hallmarks. (a) Spatial transcriptomics slide (Slide 0) highlighting from left to right: tumour spots, proliferation hotspots, EPI hotspots, EMT hotspots, hypoxic hotspots and angiogenic hotspots identified by SpottedPy. The black square 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 tumour 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, summarised across the 12 slides. The dashed line represents no difference in relative distance to EMT hotspots or EPI hotspots. 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 EPI hotspots, respectively (*** p<0.001). **(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.

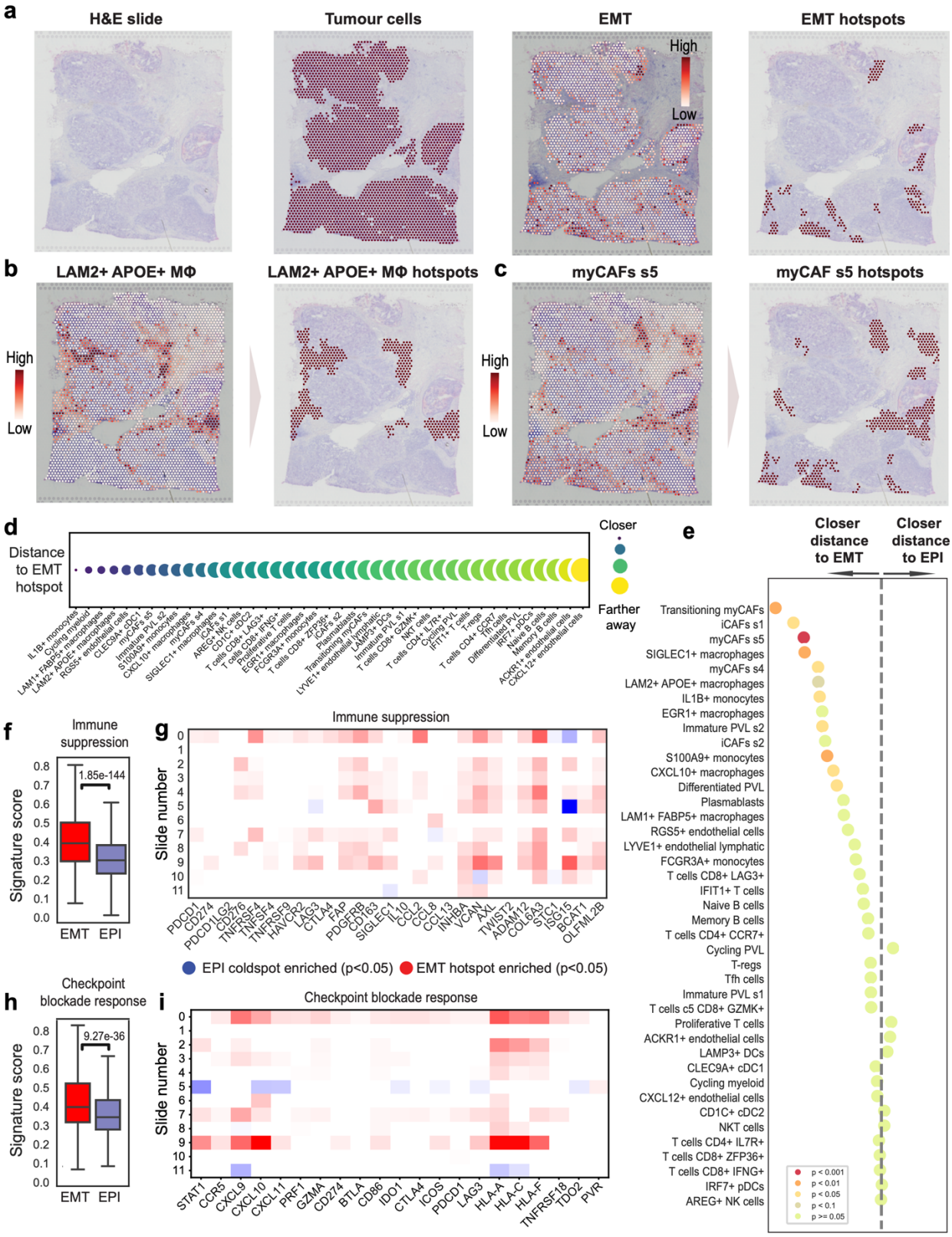


Figure 3: The spatial interplay between EMT progression and the TME. (a) Spatial transcriptomics plots highlighting tumour cell spots (left), the EMT gradient through these tumour

spots (middle) and EMT hotspots identified by SpottedPy (right) in Slide 5. **(b)** Spatial localisation of macrophage-enriched spots (left) and SpottedPy-defined LAM2+ APOE+ macrophage hotspots (right) in Slide 5. **(c)** Spatial localisation 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 colours 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.

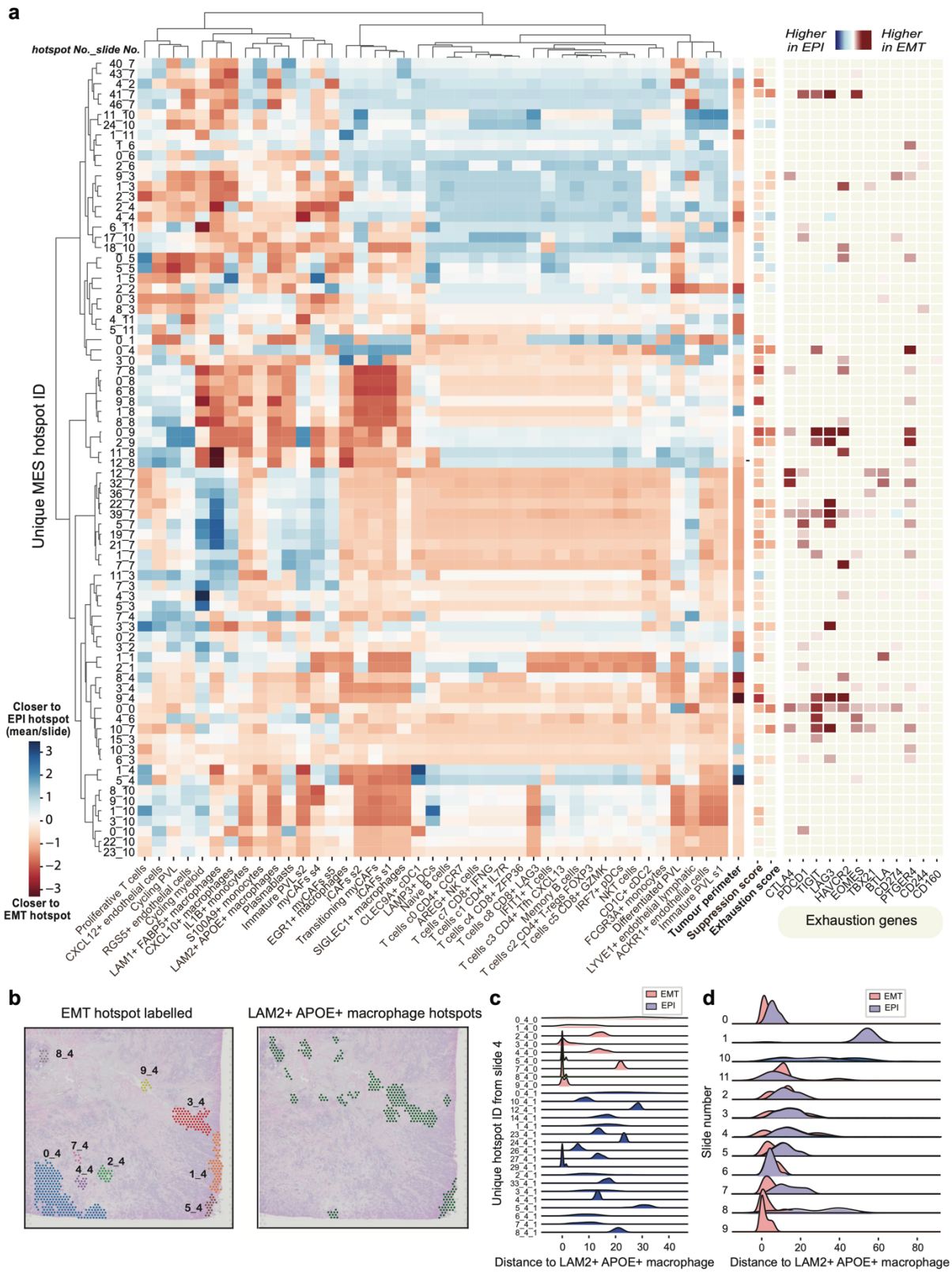


Figure 4: Inter- and intratumour heterogeneity of EMT hotspots. (a) Dendrogram highlighting the proximity of EMT and EPI hotspots to TME cell types. The dendrogram is clustered according to the

937 distances from EMT/EPI hotspots to regions enriched in immune/stromal cells. Red indicates that an
938 EMT hotspot is closer to a cell type, while blue suggests that the EPI hotspots in that slide are on
939 average closer. The x-axis displays individual EMT hotspots (label indicates hotspot number and slide
940 number). To the right of the dendrogram, distances to the tumour perimeter, suppression and
941 exhaustion signature scores are illustrated. Red indicates that the hotspot is significantly enriched in
942 these signatures compared to the average EPI hotspot in the slide ($p<0.05$), blue indicates EPI
943 hotspots are significantly enriched ($p<0.05$). Further to the right, individual genes associated with the
944 exhaustion signature are shown, with red indicating the gene expression is higher in EMT hotspots
945 ($p<0.05$) and blue indicating the gene expression is higher in EPI hotspots ($p<0.05$). (b) Slide 4 with
946 individual EMT hotspots labelled (left) and LAM2+ APOE+ macrophage hotspots highlighted (right).
947 (c) Distance distributions for each EMT hotspot in slide 4 to LAM2+ APOE+ macrophage hotspots.
948 (d) Distance distributions of EMT hotspots to LAM2+ APOE+ macrophages across all 12 slides in the
949 cohort.

950

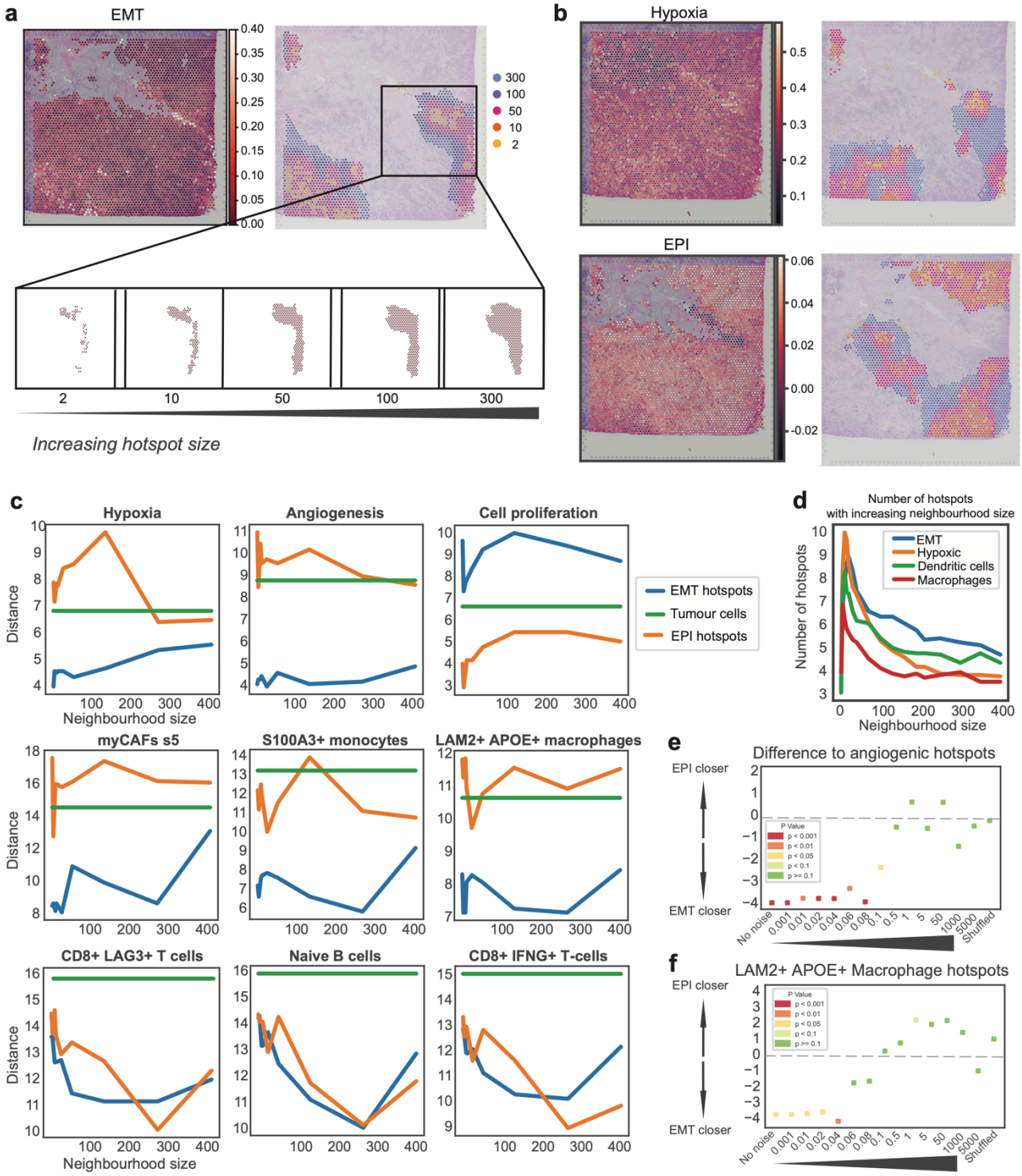


Figure 5: Sensitivity analysis of hotspot relationships. (a) EMT hotspot generation using a hotspot neighbourhood parameter of 2, 10, 50, 100 and 300, respectively. Increasingly larger neighbourhoods are highlighted in different colours as indicated in the legend. **(b)** Hypoxia and epithelial hotspot 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 EPI hotspots (orange) to regions enriched in various cancer hallmarks and TME components as the hotspot size increases.

958 The distances to the average of all tumour cells are used as a reference (green). Distances are
959 averaged over all 12 slides. **(d)** Number of distinct hotspots identified as the hotspot neighbourhood
960 parameter is increased, averaged over 12 slides. **(e)** Evaluating the impact of increased noise or spot
961 shuffling on the association between angiogenic hotspots and EMT/EPI hotspots. **(f)** Evaluating the
962 impact of increased noise or spot shuffling on the association between LAM2+ APOE+ macrophage
963 hotspots and EMT/EPI hotspots.
964

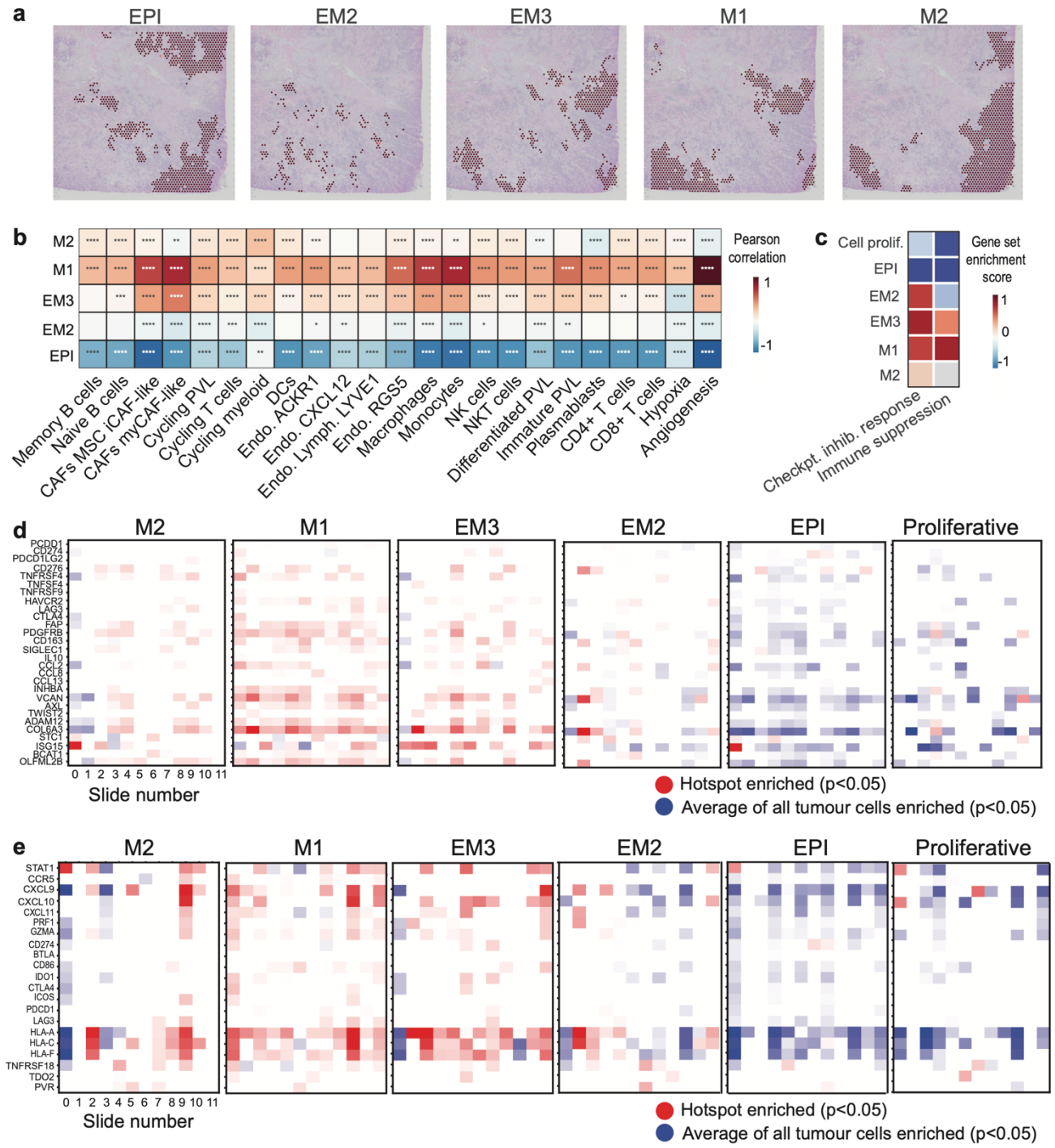


Figure 6: EMT state dynamics uncovered from the spatial exploration of breast cancer tissue. **(a)** Spatial plots depicting epithelial (EPI), intermediate (EM2, EM3), quasi-mesenchymal (M1) and fully mesenchymal (M2) hotspots in Slide 4. **(b)** Neighbourhood enrichment analysis depicting the association 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(65) 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 (column). 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. (28), Wu et al. (12) and the 10x Genomics website (34).

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/13284570.

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).

Author contributions

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

REFERENCES

1. Yuan Y. Spatial Heterogeneity in the Tumor Microenvironment. Cold Spring Harb Perspect Med. 2016 Aug;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 Jul 1;353(6294):78–82.
3. Asp M, Bergenstråhle J, Lundeberg J. Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration. BioEssays. 2020 Oct 1;42(10):1900221.
4. Moses L, Pachter L. Museum of spatial transcriptomics. Nat Methods. 2022 May;19(5):534–46.
5. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods. 2021 Jan;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 Apr 1;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 Oct 14;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 Aug;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 Nov 30;2022.06.18.496114.
10. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. Cell. 2020 Jul 23;182(2):497-514.e22.
11. He S, Jin Y, Nazaret A, Shi L, Chen X, Rampersaud S, et al. Starfysh reveals heterogeneous spatial dynamics in the breast tumor microenvironment [Internet]. bioRxiv; 2022 [cited 2023 Mar

1049 16]. p. 2022.11.21.517420. Available from:
1050 https://www.biorxiv.org/content/10.1101/2022.11.21.517420v1
1051 12. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and
1052 spatially resolved atlas of human breast cancers. Nat Genet. 2021 Sep;53(9):1334–47.
1053 13. Lengrand J, Pastushenko I, Vanuytven S, Song Y, Venet D, Sarate RM, et al. Pharmacological
1054 targeting of netrin-1 inhibits EMT in cancer. Nature. 2023 Aug;620(7973):402–8.
1055 14. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene expression,
1056 spatial location and histology to identify spatial domains and spatially variable genes by graph
1057 convolutional network. Nat Methods. 2021 Nov;18(11):1342–51.
1058 15. Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, et al. Spatial transcriptomics at
1059 subspot resolution with BayesSpace. Nat Biotechnol. 2021 Nov;39(11):1375–84.
1060 16. Walker BL, Nie Q. NeST: nested hierarchical structure identification in spatial transcriptomic data.
1061 Nat Commun. 2023 Oct 17;14(1):6554.
1062 17. Chitra U, Arnold BJ, Sarkar H, Ma C, Lopez-Darwin S, Sanno K, et al. Mapping the topography
1063 of spatial gene expression with interpretable deep learning. bioRxiv. 2023 Oct
1064 13;2023.10.10.561757.
1065 18. Palla G, Fischer DS, Regev A, Theis FJ. Spatial components of molecular tissue biology. Nat
1066 Biotechnol. 2022 Mar;40(3):308–18.
1067 19. Seferbekova Z, Lomakin A, Yates LR, Gerstung M. Spatial biology of cancer evolution. Nat Rev
1068 Genet. 2023 May;24(5):295–313.
1069 20. Varrone M, Tavernari D, Santamaria-Martínez A, Walsh LA, Ciriello G. CellCharter reveals spatial
1070 cell niches associated with tissue remodeling and cell plasticity. Nat Genet. 2023 Dec 8;1–11.
1071 21. Moses L, Einarsson PH, Jackson K, Luebbert L, Booeshaghi AS, Antonsson S, et al. Voyager:
1072 exploratory single-cell genomics data analysis with geospatial statistics [Internet]. bioRxiv; 2023
1073 [cited 2023 Oct 23]. p. 2023.07.20.549945. Available from:
1074 https://www.biorxiv.org/content/10.1101/2023.07.20.549945v2

1075 22. Bernstein MN, Scott D, Hession CC, Nieuwenhuis T, Gerritsen J, Tabrizi S, et al. Monkeybread:
1076 A Python toolkit for the analysis of cellular niches in single-cell resolution spatial transcriptomics
1077 data [Internet]. bioRxiv; 2023 [cited 2023 Oct 23]. p. 2023.09.14.557736. Available from:
1078 https://www.biorxiv.org/content/10.1101/2023.09.14.557736v1
1079 23. Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable
1080 framework for spatial omics analysis. Nat Methods. 2022 Feb;19(2):171–8.
1081 24. Fotheringham AS, Wong DWS. The Modifiable Areal Unit Problem in Multivariate Statistical
1082 Analysis. Environ Plan A. 1991 Jul;23(7):1025–44.
1083 25. Dries R, Zhu Q, Dong R, Eng CHL, Li H, Liu K, et al. Giotto: a toolbox for integrative analysis and
1084 visualization of spatial expression data. Genome Biology. 2021 Mar 8;22(1):78.
1085 26. Li Z, Zhou X. BASS: multi-scale and multi-sample analysis enables accurate cell type clustering
1086 and spatial domain detection in spatial transcriptomic studies. Genome Biol. 2022 Dec;23(1):168.
1087 27. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for
1088 research on epithelial–mesenchymal transition. Nat Rev Mol Cell Biol. 2020 Jun;21(6):341–52.
1089 28. Barkley D, Moncada R, Pour M, Liberman DA, Dryg I, Werba G, et al. Cancer cell states recur
1090 across tumor types and form specific interactions with the tumor microenvironment. Nat Genet.
1091 2022 Aug;54(8):1192–201.
1092 29. Aggarwal V, Montoya CA, Donnenberg VS, Sant S. Interplay between tumor microenvironment
1093 and partial EMT as the driver of tumor progression. iScience. 2021 Feb 19;24(2):102113.
1094 30. Pastushenko I, Blanpain C. EMT Transition States during Tumor Progression and Metastasis.
1095 Trends in Cell Biology. 2019 Mar;29(3):212–26.
1096 31. Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. Nat
1097 Commun. 2020 May 1;11(1):2142.
1098 32. Haerinck J, Goossens S, Berx G. The epithelial–mesenchymal plasticity landscape: principles of
1099 design and mechanisms of regulation. Nat Rev Genet. 2023 Sep;24(9):590–609.

1100 33. Malagoli Tagliazucchi G, Wiecek AJ, Withnell E, Secrier M. Genomic and microenvironmental
1101 heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer. Nat Commun. 2023 Feb
1102 11;14(1):1–20.
1103 34. Home Page - 10x Genomics [Internet]. [cited 2023 Mar 6]. Available from:
1104 https://www.10xgenomics.com/
1105 35. Kleshchevnikov V, Shmatko A, Dann E, Aivazidis A, King HW, Li T, et al. Cell2location maps fine-
1106 grained cell types in spatial transcriptomics. Nat Biotechnol. 2022 May;40(5):661–71.
1107 36. Sahoo S, Ramu S, Nair MG, Pillai M, Juan BPS, Milioli HZ, et al. Multi-modal transcriptomic
1108 analysis unravels enrichment of hybrid epithelial/mesenchymal state and enhanced phenotypic
1109 heterogeneity in basal breast cancer. bioRxiv. 2023 Oct 2;2023.09.30.558960.
1110 37. Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY, et al. Epithelial-mesenchymal transition
1111 spectrum quantification and its efficacy in deciphering survival and drug responses of cancer
1112 patients. EMBO Mol Med. 2014 Oct;6(10):1279–93.
1113 38. Elyanow R, Zeira R, Land M, Raphael BJ. STARCH: copy number and clone inference from
1114 spatial transcriptomics data. Phys Biol. 2021 Mar 9;18(3):035001.
1115 39. Pirrotta S, Masatti L, Corrà A, Pedrini F, Esposito G, Martini P, et al. signifinder enables the
1116 identification of tumor cell states and cancer expression signatures in bulk, single-cell and spatial
1117 transcriptomic data. bioRxiv. 2023 Mar 10;2023.03.07.530940.
1118 40. Xun Z, Ding X, Zhang Y, Zhang B, Lai S, Zou D, et al. Reconstruction of the tumor spatial
1119 microenvironment along the malignant-boundary-nonmalignant axis. Nat Commun. 2023 Feb
1120 20;14(1):1–16.
1121 41. Brown MS, Abdollahi B, Wilkins OM, Lu H, Chakraborty P, Ognjenovic NB, et al. Phenotypic
1122 heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and
1123 metastasis in breast cancer. Science Advances. 2022 Aug 3;8(31):eabj8002.
1124 42. Muz B, de la Puente P, Azab F, Azab AK. The role of hypoxia in cancer progression,
1125 angiogenesis, metastasis, and resistance to therapy. Hypoxia (Auckl). 2015 Dec 11;3:83–92.

1126 43. Weidemann A, Johnson RS. Biology of HIF-1α. Cell Death Differ. 2008 Apr;15(4):621–7.
1127 44. Tam SY, Wu VWC, Law HKW. Hypoxia-Induced Epithelial-Mesenchymal Transition in Cancers:
1128 HIF-1α and Beyond. Front Oncol. 2020 Apr 8;10:486.
1129 45. Nishida N, Yano H, Nishida T, Kamura T, Kojiro M. Angiogenesis in Cancer. Vasc Health Risk
1130 Manag. 2006 Sep;2(3):213–9.
1131 46. Almagro J, Messal HA, Elosegui-Artola A, Rheenen J van, Behrens A. Tissue architecture in
1132 tumor initiation and progression. Trends in Cancer. 2022 Jun 1;8(6):494–505.
1133 47. Saxena K, Jolly MK, Balamurugan K. Hypoxia, partial EMT and collective migration: Emerging
1134 culprits in metastasis. Translational Oncology. 2020 Nov 1;13(11):100845.
1135 48. Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, et al. IL-1-induced JAK/STAT signaling is
1136 antagonized by TGF-β to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. Cancer
1137 discovery. 2019 Feb;9(2):282.
1138 49. Ghahremanifard P, Chanda A, Bonni S, Bose P. TGF-β Mediated Immune Evasion in Cancer—
1139 Spotlight on Cancer-Associated Fibroblasts. Cancers. 2020 Dec 5;12(12):3650.
1140 50. Principe DR, Timbers KE, Atia LG, Koch RM, Rana A. TGFβ Signaling in the Pancreatic Tumor
1141 Microenvironment. Cancers (Basel). 2021 Oct 11;13(20):5086.
1142 51. Xu J, Lamouille S, Derynck R. TGF-β-induced epithelial to mesenchymal transition. Cell Res.
1143 2009 Feb;19(2):156–72.
1144 52. myCAFs are better than yours: targeting myofibroblasts potentiates immunotherapy. Trends in
1145 Cancer. 2023 Jan 1;9(1):1–2.
1146 53. Amer HT, Stein U, El Tayebi HM. The Monocyte, a Maestro in the Tumor Microenvironment (TME)
1147 of Breast Cancer. Cancers. 2022 Nov 7;14(21):5460.
1148 54. Grosso JF, Kelleher CC, Harris TJ, Maris CH, Hipkiss EL, Marzo AD, et al. LAG-3 regulates $CD8^+$
1149 T cell accumulation and effector function in murine self- and tumor-tolerance systems. J Clin
1150 Invest. 2007 Nov 1;117(11):3383–92.

1151 55. Cui C, Xu C, Yang W, Chi Z, Sheng X, Si L, et al. Ratio of the interferon-γ signature to the
1152 immunosuppression signature predicts anti-PD-1 therapy response in melanoma. NPJ Genom
1153 Med. 2021 Feb 4;6(1):7.
1154 56. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour
1155 microenvironment. Nat Rev Immunol. 2015 Nov;15(11):669–82.
1156 57. Yang X, Lin Y, Shi Y, Li B, Liu W, Yin W, et al. FAP Promotes Immunosuppression by Cancer-
1157 Associated Fibroblasts in the Tumor Microenvironment via STAT3–CCL2 Signaling. Cancer Res.
1158 2016 Jul 15;76(14):4124–35.
1159 58. Zhao K, Yi Y, Ma Z, Zhang W. INHBA is a Prognostic Biomarker and Correlated With Immune
1160 Cell Infiltration in Cervical Cancer. Front Genet. 2022 Jan 4;12:705512.
1161 59. Johnson KA, Emmerich P, Matkowskyj KA, Deming DA. Predicting CD8+ T-cell infiltration in
1162 colorectal cancer using versican proteolysis across molecular profiles. JCO. 2020
1163 Feb;38(4_suppl):189–189.
1164 60. Chen P, Cescon M, Bonaldo P. Collagen VI in cancer and its biological mechanisms. Trends Mol
1165 Med. 2013 Jul;19(7):410–7.
1166 61. Johnson AM, Bullock BL, Neuwelt AJ, Poczobutt JM, Kaspar RE, Li HY, et al. Cancer Cell-Intrinsic
1167 Expression of MHC Class II Regulates the Immune Microenvironment and Response to Anti-PD-
1168 1 Therapy in Lung Adenocarcinoma. J Immunol. 2020 Apr 15;204(8):2295–307.
1169 62. Liu J, Ling Y, Su N, Li Y, Tian S, Hou B, et al. A novel immune checkpoint-related gene signature
1170 for predicting overall survival and immune status in triple-negative breast cancer. Transl Cancer
1171 Res. 2022 Jan;11(1):181–92.
1172 63. Hazini A, Fisher K, Seymour L. Deregulation of HLA-I in cancer and its central importance for
1173 immunotherapy. J Immunother Cancer. 2021 Aug 1;9(8):e002899.
1174 64. Wuerfel FM, Huebner H, Häberle L, Gass P, Hein A, Jud SM, et al. HLA-G and HLA-F protein
1175 isoform expression in breast cancer patients receiving neoadjuvant treatment. Sci Rep. 2020 Sep
1176 25;10(1):15750.

1177 65. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-γ-related
1178 mRNA profile predicts clinical response to PD-1 blockade. J Clin Invest. 2017 Aug 1;127(8):2930–
1179 40.
1180 66. Zormpas E, Queen R, Comber A, Cockell SJ. Mapping the transcriptome: Realizing the full
1181 potential of spatial data analysis. Cell. 2023 Dec 21;186(26):5677–89.
1182 67. Ganier C, Mazin P, Herrera-Oropeza G, Du-Harpur X, Blakeley M, Gabriel J, et al. Multiscale
1183 spatial mapping of cell populations across anatomical sites in healthy human skin and basal cell
1184 carcinoma. Proceedings of the National Academy of Sciences. 2024 Jan 9;121(2):e2313326120.
1185 68. Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. Nat
1186 Biotechnol. 2022 Sep;40(9):1349–59.
1187 69. Valdeolivas A, Amberg B, Giroud N, Richardson M, Gálvez EJC, Badillo S, et al. Profiling the
1188 heterogeneity of colorectal cancer consensus molecular subtypes using spatial transcriptomics.
1189 npj Precis Onc. 2024 Jan 10;8(1):1–16.
1190 70. Sherman TD, Gao T, Fertig EJ. CoGAPS 3: Bayesian non-negative matrix factorization for single-
1191 cell analysis with asynchronous updates and sparse data structures. BMC Bioinformatics. 2020
1192 Oct 14;21(1):453.
1193 71. Spatial and single-nucleus transcriptomic analysis of genetic and sporadic forms of Alzheimer's
1194 Disease | bioRxiv [Internet]. [cited 2023 Nov 27]. Available from:
1195 https://www.biorxiv.org/content/10.1101/2023.07.24.550282v1
1196 72. Ospina OE, Wilson CM, Soupir AC, Berglund A, Smalley I, Tsai KY, et al. spatialGE: quantification
1197 and visualization of the tumor microenvironment heterogeneity using spatial transcriptomics.
1198 Bioinformatics. 2022 Apr 28;38(9):2645–7.
1199 73. Emami Nejad A, Najafgholian S, Rostami A, Sistani A, Shojaeifar S, Esparvarinha M, et al. The
1200 role of hypoxia in the tumor microenvironment and development of cancer stem cell: a novel
1201 approach to developing treatment. Cancer Cell Int. 2021 Jan 20;21(1):62.

1202 74. Wang X, Zhang W, Sun X, Lin Y, Chen W. Cancer-associated fibroblasts induce epithelial-
1203 mesenchymal transition through secreted cytokines in endometrial cancer cells. Oncol Lett. 2018
1204 Apr;15(4):5694–702.
1205 75. Jia C, Wang G, Wang T, Fu B, Zhang Y, Huang L, et al. Cancer-associated Fibroblasts induce
1206 epithelial-mesenchymal transition via the Transglutaminase 2-dependent IL-6/IL6R/STAT3 axis
1207 in Hepatocellular Carcinoma. Int J Biol Sci. 2020 Jul 19;16(14):2542–58.
1208 76. Tyler M, Tirosh I. Decoupling epithelial-mesenchymal transitions from stromal profiles by
1209 integrative expression analysis. Nat Commun. 2021 May 10;12(1):2592.
1210 77. Li X, Chen L, Peng X, Zhan X. Progress of tumor-associated macrophages in the epithelial-
1211 mesenchymal transition of tumor. Front Oncol. 2022 Jul 28;12:911410.
1212 78. Brabletz S, Schuhwerk H, Brabletz T, Stemmler MP. Dynamic EMT: a multi-tool for tumor
1213 progression. The EMBO Journal. 2021 Sep 15;40(18):e108647.
1214 79. Pang MF, Georgoudaki AM, Lambut L, Johansson J, Tabor V, Hagikura K, et al. TGF-β1-induced
1215 EMT promotes targeted migration of breast cancer cells through the lymphatic system by the
1216 activation of CCR7/CCL21-mediated chemotaxis. Oncogene. 2016 Feb;35(6):748–60.
1217 80. Kmieciak M, Knutson KL, Dumur CI, Manjili MH. HER-2/neu antigen loss and relapse of mammary
1218 carcinoma are actively induced by T cell-mediated anti-tumor immune responses. Eur J Immunol.
1219 2007 Mar;37(3):675–85.
1220 81. Santisteban M, Reiman JM, Asiedu MK, Behrens MD, Nassar A, Kalli KR, et al. Immune-induced
1221 epithelial to mesenchymal transition in vivo generates breast cancer stem cells. Cancer Res. 2009
1222 Apr 1;69(7):2887–95.
1223 82. Galsky MD, Wang L, Saci A, Szabo PM, Gong Y, Zhu J. 850PD - Epithelial-mesenchymal
1224 transition (EMT), T cell infiltration, and outcomes with nivolumab (nivo) in urothelial cancer (UC).
1225 Annals of Oncology. 2017 Sep 1;28:v297.

1226 83. Chockley PJ, Chen J, Chen G, Beer DG, Standiford TJ, Keshamouni VG. Epithelial-mesenchymal
1227 transition leads to NK cell–mediated metastasis-specific immunosurveillance in lung cancer. J
1228 Clin Invest. 128(4):1384–96.
1229 84. Datar I, Schalper KA. Epithelial–Mesenchymal Transition and Immune Evasion during Lung
1230 Cancer Progression: The Chicken or the Egg? Clin Cancer Res. 2016 Jul 15;22(14):3422–4.
1231 85. Terry S, Savagner P, Ortiz-Cuaran S, Mahjoubi L, Saintigny P, Thiery JP, et al. New insights into
1232 the role of EMT in tumor immune escape. Mol Oncol. 2017 Jul 1;11(7):824–46.
1233 86. Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J. Spatiotemporal regulation of epithelial-
1234 mesenchymal transition is essential for squamous cell carcinoma metastasis. Cancer Cell. 2012
1235 Dec 11;22(6):725–36.
1236 87. Cheng YC, Zhang Y, Tripathi S, Bv H, Jolly MK, Schiebinger G, et al. Reconstruction of single
1237 cell lineage trajectories and identification of diversity in fates during the epithelial-to-mesenchymal
1238 transition. bioRxiv. 2023 Sep 22;2023.09.19.558325.
1239 88. Sinha D, Saha P, Samanta A, Bishayee A. Emerging Concepts of Hybrid Epithelial-to-
1240 Mesenchymal Transition in Cancer Progression. Biomolecules. 2020 Nov 16;10(11):1561.
1241 89. Subhadarshini S, Markus J, Sahoo S, Jolly MK. Dynamics of Epithelial–Mesenchymal Plasticity:
1242 What Have Single-Cell Investigations Elucidated So Far? ACS Omega. 2023 Apr 4;8(13):11665–
1243 73.
1244 90. Zhang Y, Zhang H, Wang M, Schmid T, Xin Z, Kozhuharova L, et al. Hypoxia in Breast Cancer-
1245 Scientific Translation to Therapeutic and Diagnostic Clinical Applications. Front Oncol.
1246 2021;11:652266.
1247 91. Fu Z, Mowday AM, Smaill JB, Hermans IF, Patterson AV. Tumour Hypoxia-Mediated
1248 Immunosuppression: Mechanisms and Therapeutic Approaches to Improve Cancer
1249 Immunotherapy. Cells. 2021 Apr 24;10(5):1006.
1250 92. Lee CT, Mace T, Repasky EA. Hypoxia-Driven Immunosuppression: A new reason to use thermal
1251 therapy in the treatment of cancer? Int J Hyperthermia. 2010;26(3):232–46.

1252 93. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis.
1253 Genome Biology. 2018 Feb 6;19(1):15.
1254 94. Stein-O'Brien GL, Clark BS, Sherman T, Zibetti C, Hu Q, Sealfon R, et al. Decomposing Cell
1255 Identity for Transfer Learning across Cellular Measurements, Platforms, Tissues, and Species.
1256 Cell Systems. 2019 May 22;8(5):395-411.e8.
1257 95. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
1258 Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417–
1259 25.
1260 96. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A Comparison of PAM50
1261 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-
1262 Treated Estrogen Receptor–Positive Breast Cancer. Clinical Cancer Research. 2010 Oct
1263 31;16(21):5222–32.
1264 97. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nature reviews
1265 Immunology. 2015 Aug;15(8):486.
1266 98. Proceedings of the Python in Science Conference (SciPy): Exploring Network Structure,
1267 Dynamics, and Function using NetworkX [Internet]. [cited 2023 Mar 6]. Available from:
1268 https://conference.scipy.org/proceedings/SciPy2008/paper_2/
1269 99. Rey SJ, Anselin L. PySAL: A Python library of spatial analytical methods. The Review of regional
1270 studies. 2007 Dec 1;37(1):5–27.
1271