npj Breast Cancer

https://doi.org/10.1038/s41523-026-00897-1

Article in Press

Spatial gene expression analysis reveals drivers of extremely early lymph node metastasis in breast cancer

Received: 14 August 2025

Accepted: 14 January 2026

Cite this article as: Nagasawa, S., Kajiya, K., Ishikawa, E. *et al.* Spatial gene expression analysis reveals drivers of extremely early lymph node metastasis in breast cancer. *npj Breast Cancer* (2026). https://doi.org/10.1038/s41523-026-00897-1

Satoi Nagasawa, Keiko Kajiya, Erina Ishikawa, Akinori Kanai, Ayako Suzuki, Ai Motoyoshi, Tsuguo Iwatani, Manabu Kubota, Masaru Nakamura, Tatsuya Onishi, Akiyoshi Hoshino, Ichiro Maeda, Akihiko Morozumi, Kenji Takatsuka, Junki Koike, Masahide Seki, Koichiro Tsugawa & Yutaka Suzuki

Title

Spatial Gene Expression Analysis Reveals Drivers of Extremely Early Lymph Node Metastasis in Breast Cancer

Authors

Satoi Nagasawa,[1,2,4]†* Keiko Kajiya,[1]† Erina Ishikawa,[1] Akinori Kanai,[1] Ayako Suzuki,[1] Ai Motoyoshi,[2] Tsuguo Iwatani,[2] Manabu Kubota,[3] Masaru Nakamura,[3] Tatsuya Onishi,[4] Akiyoshi Hoshino,[5] Ichiro Maeda,[5] Akihiko Morozumi,[1,6] Kenji Takatsuka,[7] Junki Koike,[3] Masahide Seki,[1] Koichiro Tsugawa,[2] Yutaka Suzuki[1]*

Affiliations

[1] Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan.

[2] Division of Breast and Endocrine Surgery, Department of Surgery, St. Marianna University School of Medicine, 2-16-1, Sugao, Miyamae-ku, Kawasaki 216-8511, Japan.

[3]Department of Pathology, St. Marianna University School of Medicine, 2-16-1, Sugao, Miyamae-ku, Kawasaki 216-8511, Japan.

[4]Department of Breast Surgery, National Cancer Center Hospital East, 6-5-1, Kashiwanoha, Kashiwa, Chiba 277-8577, Japan.

[5] Department of Diagnostic Pathology, Kitasato University Kitasato Institute Hospital, 5-9-1, Shirokane, Minato-ku, Tokyo 108-8642, Japan.

[6] Designing and Development Department, Technology Solutions Sector, Healthcare Business Unit, Nikon Corporation, 1-5-20, Nishioi, Shinagawa-ku, Tokyo 140-8601, Japan.

[7] Investment Planning Department, Corporate Strategy, Nikon Corporation, 1-5-20, Nishioi, Shinagawa-ku, Tokyo 140-8601, Japan.

†These authors contributed equally to this work.

*Corresponding authors:

Satoi Nagasawa and Yutaka Suzuki

Laboratory of Functional Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba 277-8562, Japan

E-mail: s3nagasawa@edu.k.u-tokyo.ac.jp and ysuzuki@edu.k.u-tokyo.ac.jp

Tel./Fax: +81 4 7136 4076

Abstract

Lymph node metastasis correlate with breast cancer prognosis; however, the cellular mechanisms underlying the earliest metastatic events remain unclear. In spatial transcriptomic analysis of a patient with breast cancer at single cell resolution, we identified 30 tumor cells representing the initial metastatic seeding in a lymph node. These cells originated from multiple epithelial–mesenchymal (EM) transition status and included six distinct subpopulations with biological significance. Only cells exhibiting metabolic shift toward fatty acid metabolism successfully established lymph node colonies, implicating this shift in metastatic fitness. The tumor microenvironment surrounding these cells showed immunosuppressive and tumor-promoting features, supporting metastasis establishment. Cross-referencing these expression profiles with public datasets revealed that poor prognosis correlated not with fully mesenchymal or metastatic

populations, but with hybrid EM cells exhibiting epithelial and mesenchymal traits. These findings highlight the metabolic and phenotypic plasticity of metastatic cells and serve as translational bridges between the spatial evolution of tumor cells in the extremely early stages of lymph nodes metastasis and clinical prognosis in breast cancer.

Introduction

Axillary lymph nodes are considered the "first stop" in breast cancer metastasis [1-2]. The presence of metastatic tumor cells in these nodes correlate with poor prognosis [3-7]. Tumor cell profiles in metastases have been studied extensively. For example, Lee et al. used a mouse model and found that a shift toward fatty acid oxidation (FAO), driven by the selective activation of the transcription coactivator YAP in lymph node metastases, is essential for their establishment [8]. Liu et al. used single-cell RNA sequencing (RNA-seq) and spatial transcriptomics to compare the microenvironments of primary and lymph node metastatic tumors, reporting that in lymph node metastases, compared with primary tumors, T cell activation, cytotoxicity, and proliferation are markedly suppressed while dendritic cells show reduced ability to prime and activate T cells [9].

Although the requirements for lymph node metastasis and the niche environment are becoming clearer, how these molecular characteristics relate to cellular plasticity, possibly involving EMT and MET, remains uncertain. Notably, almost no studies have reported the first metastatic event in the lymph nodes. This is because it is difficult to track the trajectory of metastasis in humans, and it is extremely difficult to observe transient and plastic EMT states in vivo.

Substantial evidence supports the role of epithelial-to-mesenchymal transition (EMT) and mesenchymal-to-epithelial transition (MET) in metastasis [10-11]. EMT is a cellular process in which epithelial cells acquire mesenchymal cell characteristics, causing dramatic changes in tissue structure and function. Cells take on a fibroblast-like morphology and increase their motility (ability to move) and invasiveness (ability to invade surrounding tissues). MET is a process that reverses the changes in cellular phenotype induced by EMT. In cancer cells, suppression of EMT-Transcription factors like *TWIST1* is necessary to promote MET, which is required for proliferation after metastasis [12].Despite extensive debate, it remains unclear how EMT or MET contributes to metastatic potential and whether it is a necessary condition for metastasis [10,13-16]. A major challenge in addressing this question lies in EMT's broad and evolving definition. EMT was once considered a binary switch between epithelial and mesenchymal states. However, several *in vitro* studies have shown that EMT progresses gradually, forming hybrid subpopulations that occupy intermediate states between epithelial and mesenchymal identities [17-18]. Interestingly, tumor cell subpopulations nearest to the mesenchymal state do not necessarily exhibit the highest metastatic potential [19]. Indeed, high metastatic potential often lies in the hybrid state. Conversely, evidence suggests that metastatic potential and the reverse process, MET, are not correlated, indicating that EMT and MET may need to be considered separately [17]. To understand the molecular events driving metastasis, it is essential to examine how EMT progresses in the primary tumor by dissecting its intermediate states and how, and to what extent, metastasized cells restore epithelial traits in the lymph node [20].

In the present study, we used the Xenium in situ hybridization platform, which enables spatial gene expression profiling at the single-cell resolution, to track tumour cells in vivo during the

metastasis process. The purpose of the study was to clarify the spatial evolution of tumor cells in the extremely early stages of breast cancer lymph nodes metastasis.

Results

Multi-omics analysis in each breast cancer region

Before analyzing spatial patterns at the primary site and metastatic lymph node, we first examined the molecular features of the primary cancer via multiregional bulk analysis. The patient, an 80-year-old female (BRC-26), underwent surgery without neoadjuvant therapy. Her clinical diagnosis was stage II invasive breast carcinoma of the HER2 type, i.e., ER-negative, PgR-negative, and HER2-positive (for full clinicopathologic details, refer to Supplementary Table 1). For high-quality molecular analysis, especially RNA-seq, we freshly froze part of the tumor tissue harvested during surgery. We roughly dissected samples from histologically noncancerous regions (NC regions; NC-1 and NC-2) and cancerous regions (Ca regions; Ca-3 and Ca-4) (Fig. 1a and Supplementary Fig. 1a). Histologically, Ca-4 appeared sparser and less adhesive compared with Ca-3 (Supplementary Fig. 1b). We subjected these tissues to RNA-seq, enzymatic methylation sequencing (EM-seq), and whole-genome sequencing (WGS) (refer to Supplementary Table 2 for statistical analyses).

We first performed transcriptome analysis using RNA-seq to compare gene expression between the NC and Ca regions. Of 4,584 differentially expressed genes (DEGs), 2,248 and 2,336 were significantly upregulated and downregulated in the cancer region (Supplementary Fig. 1c). Gene pathway enrichment analysis of these DEGs showed that “cell adhesion” pathways were enriched in NC regions, whereas pathways for “carboxylic acid metabolism,” and “cell cycle” were enriched in Ca regions (Fig. 1b), suggesting progressive malignant transformation from NC

to Ca regions [21]. DNA methylation analysis revealed more extensive genome-wide hypomethylation in Ca regions compared with NC regions (Fig. 1c). Furthermore, promoters, intergenic regions, and gene bodies all showed reduced methylation (Supplementary Fig. 1d). Hypomethylation of intergenic and intron regions, especially repetitive sequences, can promote chromosomal instability and mutations, potentially increasing cancer risk [22-24]. These results confirm that breast cells undergo malignant transformation as they progress from NC to Ca regions.

Within Ca regions, we observed significant changes in key metastasis-related genes (Fig. 1d). Expression of EMT markers, such as Vimentin [25], and transcription factors Snail and ZEB2, considered, known EMT regulators [26-27], increased from Ca-3 to Ca-4. Correspondingly, DNA methylation levels declined from Ca-3 to Ca-4 (Supplementary Fig. 1e), consistent with previous reports that EMT-related genes are epigenetically regulated [28-30]. Similarly, methylation of the *THY1* gene, which was highly expressed in Ca-4, decreased from Ca-3 to Ca-4 (Fig. 1e, f). THY1, a GPI-anchored protein involved cell adhesion, migration, and polarity, can suppress lung metastasis when its integrin signaling is inhibited [31-32]. *THY1*, which regulates cancer cell migration and invasion, also appears epigenetically controlled. These molecular profiles indicate that malignant transformation from NC to Ca regions, along with DNA methylation differences within Ca regions, may underlie varying EMT phenotypes in tumor cells [33-34].

WGS results further supported cancer progression from NC to Ca regions. We detected mutations in KMT2C (Y987H), the most frequently mutated histone methyltransferase in breast cancer [35], across all regions (Fig. 1g). PIK3CA (H1047R), a known cancer driver, was mutated in all regions except NC-1. Mutations in transcription factor FOXA1 (R262P) and histone demethylase KDM6A (P281R) were specific to Ca regions (Fig. 1h). Copy number variation

(CNV) analysis revealed chromosomal abnormalities in Ca regions (Fig.1i). These findings suggest that tumor development occurred against a background of clonal mammary gland expansion [36]. Overall, multiregional analyses indicated that breast cancer follows a well-characterized molecular trajectory from NC to Ca regions. Importantly, these molecular features aligned closely with histological observations, especially regarding gene regulation.

Spatial analysis of the breast cancer at single-cell resolution

To investigate the molecular mechanisms of metastasis in detail, we performed spatial gene expression analysis at single-cell resolution on primary tumors and paired axillary metastatic lymph nodes from the patient shown in Fig. 1. In addition to the primary tumor adjacent to the bulk analysis region and lymph node metastases from the same case, we included a tumor-draining lymph node (TdL) as a comparison specimen (Fig. 2a). The TdL was a clinically diagnosed as metastasis-negative by rapid intraoperative evaluation.

We applied Xenium spatial gene expression profiling using a custom 380-gene breast cancer panel (refer to Supplementary Table 3 for the gene list). Across the primary tumor, metastatic lymph node and TdL, Xenium detected 1,005,436 cells (Fig. 2b). Using gene expression signatures [37-39] and histological location information, we annotated clusters by major cell types. The luminal cell cluster was subdivided into two clusters, (i) and (ii), with (i) further branching into four subclusters: L1: luminal 1; L2: luminal 2; B: basal; and M: mesenchymal like (Fig. 2c). We identified genes significantly upregulated within each cluster. *KRT8* and *KRT18*, markers of mammary ductal epithelial cells, were enriched in L1 and L2, whereas *KRT5* and *KRT14*, myoepithelial markers, were enriched in cluster B (Supplementary Fig. 2). Consistent with expression information, spatial localization of L1, L2, and B clusters matched the lobular

architecture of the mammary gland (Fig. 2d). High *ANKRD30A* expression in L2 and elevated *KIT* and *KRT23* expression in L1 aligned with two previously reported mature luminal cell types [40]: mature luminal cell 2 (*KRT18*+/*ANKRD30A*+) and mature luminal cell 1 (*KRT18*+/*KRT23*+/*KIT*+). L1, L2, and B clustered on the lobules, whereas M cells were distributed sporadically throughout the tumor, suggesting reduced cell adhesion (Fig. 2e). Consistent with this, the M cluster showed increased expression of EMT markers, including *THY1*, an EMT-related gene shown in Fig. 1e. Collectively, these findings support a role for *THY1* expression in EMT initiation from normal lobule cells (L1, L2, and B), potentially regulated through DNA methylation.

Although cluster (i) encompassed lobular and early EMT-stage cells, cluster (ii) localized pathologically to tumor regions and comprised definitive cancer cells. Strong *ANKRD30A* expression in cluster (ii) (Fig. 2f) suggested that L2 may be the tumor's cell of origin [40]. Uniform manifold approximation and projection (UMAP) showed that metastatic lymph node tumor cells formed distinct clusters. Based on high *ANKRD30A* expression, we defined L2 as the tumor-originating cell and applied pseudotemporal trajectory analysis using Monocle 3 [41]. The inferred trajectory showed branching from L2 into (ii)-1, progressing to (ii)-2, and finally differentiating into metastatic tumor cells in lymph nodes (Fig. 2g). Taken together, these findings suggest a progression pathway in which cancer cells originate from L2, acquire EMT potential in the M cluster, and ultimately establish metastases within cluster (ii) in lymph nodes.

Spatial analysis of the lymph nodes reveals six transcriptionally distinct EMT states

When inspecting spatial gene expression data to further investigate the mechanisms underlying metastasis, we unexpectedly identified a small cluster of only 30 cells within the TdL, a site

clinically diagnosed as metastasis-free; these cells spanned a region approximately 200 μm in diameter (Fig. 3a). Gene expression analysis revealed clear *KRT19* expression, an epithelial marker absent in normal lymph node cells. Additional hematoxylin and eosin (HE) staining confirmed their morphology was consistent with cancer cells. The identification of these cells in the TdL provided a rare opportunity to examine molecular events at an extremely early stage of metastasis. This form of early dissemination is clinically referred to as isolated tumor cells (ITCs), a term we adopt hereafter.

When we visualized ITCs on the UMAP plane, we observed transcriptional heterogeneity (Fig. 3b, left panel), indicating that ITCs comprise several distinct cell types. To better characterize the biology of ITCs and the roles of different cell states in metastasis, we applied nonhierarchical clustering to subclassify tumor cells into six groups (C1–C6) and assigned each ITC to one of these clusters (Fig. 3b, right panel). Owing to the small number of ITCs detected, subsequent analyses focused on the clusters to which each ITC mapped. For each subpopulation associated with an ITC, we identified distinct signaling pathways linking histology, spatial location, and gene expression (Fig. 3c–e and Supplementary Table 4).

C6 occupied a position on the UMAP closest to EMT initiation (corresponding to the M cluster in Fig. 2c). Its scattered distribution within the primary tumor suggested low intercellular adhesion. C6 cells expressed high levels of ETM-related markers, such as *VIM* and *CDH3*, as well as stemness markers characteristic of breast cancer, including $CD24^{low}$ and $CD44^{high}$ (Fig. 3d). In contrast, epithelial marker expression was low, indicating a sustained mesenchymal phenotype. Gene enrichment analysis showed activation of EMT, stem cell signaling, TGF-β signaling, WNT signaling, immune checkpoint pathways, and antianoikis signaling (Fig. 3e). Signals from the microenvironment, such as TGF-β and Wnt ligands, induce various EMT-

related transcription factors [42-43]. EMT also promotes *PD-L1* expression, activating immune checkpoint pathways [44-45]. These findings suggest that C6 represents tumor cells in an extreme mesenchymal phenotype, capable of evading anoikis and immune clearance. In contrast, C4 most strongly retained epithelial features and was the most predominant cell population within lymph node metastases, forming the tumor mass (Fig. 3c). C4 cells showed enrichment of cholesterol homeostasis and fatty acid metabolism pathways, with upregulation of fatty acid metabolism–related molecules in metastatic lymph nodes (Fig. 3e, f). A previous mouse-based lymph node metastasis study demonstrated that a metabolic shift toward FAO in tumor cells is required for lymph node colonization by cancer cells [8]. Thus, C4 represents a cancer cell population that has acquired an extreme epithelial phenotype and metabolic reprogramming suited for metastatic growth.

Clusters C1, C2, C3, and C5 represented hybrid epithelial–mesenchymal (EM) cell populations that coexpressed epithelial and mesenchymal markers (Fig. 3d). The degree of coexpression differs among clusters, each cluster exhibited distinct features. C5 showed the strongest enrichment for the G2M checkpoint pathway (Fig. 3g). Prior studies have indicated that breast cancers with G2M activity display higher proliferative activity, increased MYC pathway activation, earlier metastasis, and worse survival [46], suggesting that C5 may represent a more aggressive hybrid EMT subpopulation. C3 was enriched for VEGF signaling and glycolysis, likely reflecting a hypoxic environment in the tumor core that promotes angiogenesis and glycolysis [47]. C2, localized at the tumor periphery, i.e., the invasive front, displayed activation of matrix metalloproteinases (MMPs) and canonical Wnt signaling. Cells at tumor margins exhibit EM plasticity and migratory behavior [48], with MMP signaling [42] and the Wnt/β catenin pathway [49] known to promote invasion. Canonical Wnt signaling was also enriched in C2 and hybrid

EMT subpopulations (C3 and C6), consistent with reports that Wnt7A/B maintains the hybrid EM state [12,48].

C1 showed activation of retinoid metabolism and lipid transport pathways. Retinoic acid activates retinoic acid receptors and retinoid X receptors, driving expression of fatty acid metabolism genes [50]. Considering that C1 gene expression aligns closely with C4 (the most significantly proliferating cluster in lymph node metastases, involving elevated fatty acid metabolism), we speculate that C1 may act as a precursor population to C4.

To validate these results, we added spatial transcriptome and protein expression analysis of lymph node metastases in different patients. As a result, expression of the FASN which was the molecule shown in Figure 3f was confirmed in metastatic lymph nodes in both transcriptome and protein expression(Supplementary figure 3). We also conducted multiplex fluorescent immunostaining of EMT-associated molecules in tumour cells was performed. As a result, we identified tumour cells co-expressing both epithelial markers and mesenchymal markers simultaneously(Supplementary Figure 4). This observation in patient of breast cancer demonstrates that a genuine EMT continuum does indeed exist.

Collectively, these findings suggest that hybrid EM cells undergo functional adaptation within primary tumor, with C6 as a possible EMT origin, C2 as an intermediate, C3 and C5 representing divergent aggressive states, and C1 transitioning toward the epithelialized, proliferative C4. The important thing is that each hybrid EMT subtype has different functional capabilities and that these cells are present in primary tumour from the very early stages of metastasis (Fig. 3h).

Analysis of the tumor microenvironment in metastasized lymph node

The major bottleneck for cancer cells infiltrating a lymph node is the new tumor microenvironment (TME) [51]. To address this, we compared interactions between tumor cells and their surrounding TME in the TdL and the metastatic lymph node (Fig. 4a). Notably, we detected distinct interaction patterns specific to each TME (Fig. 4b). One interaction detected uniquely in the TdL was the CD45–MRC1 ligand–receptor interaction, occurring between CD45 on T cells and MRC1 on macrophages (Fig. 4c). MRC1, an endocytosis receptor belonging to the C-type lectin family, is expressed on cells like dendritic cells, macrophages, and endothelial cells [52]. It has been shown to inhibit CD45 activity on T cells via direct interaction, leading to upregulation of the immune checkpoint protein CTLA4 and induction of antigen-specific T cell tolerance [53-55]. Therefore, immune escape may begin through this mechanism in the TdL.

Metastatic lymph nodes uniquely exhibited coinhibitory ligand–receptor interactions, such as VEGFA–KDR and CD86–CTLA4, consistent with the promotion of angiogenesis and the establishment of a T cell–exhausted environment already formed at this stage (Fig. 4d). Among interactions involving collagen, cancer-associated fibroblasts (CAFs) exhibited the highest number of interactions (Supplementary Fig. 5). Representative interactions in metastatic lymph nodes included COL1A1–CD44 and COL4A1–CD44, both known to promote tumor proliferation and progression [56-57]. In the chemokine signaling category, the CXCL16–CXCR6 interaction was particularly prominent in metastatic lymph nodes (Supplementary Fig. 5). High CXCL16 expression is associated with histological malignancy and is implicated in tumor progression and metastasis via activation of the CXCL16–CXCR6 axis [58]. These TME-derived signals appear to constitute necessary conditions for successful metastasis.

Transcriptional profiling of hybrid EMT and its association with patient outcomes

Prompted by the preceding analyses, we sought to determine whether the transcriptional profiles we identified are associated with patient prognosis in a broader clinical context. To this end, we reanalyzed a previously reported scRNA-seq dataset from metastatic lymph nodes of patients with breast cancer [4] (Fig. 5a). Further clustering of cancer cells from the dataset revealed five clusters (sc8,sc4,sc2,sc15 and sc10) (Fig. 5b). According to publicly available information, sc2,sc4 and sc8 were metastatic lymph node-derived samples, and sc15, sc10 was derived from the primary tumor. A metabolic shift toward fatty acids were observed in metastatic lymph node-derived samples(sc2,sc4 and sc8)（Fig. 5c). These observations are consistent with our results shown in Figure 3 and support our findings that a metabolic shift toward fatty acid metabolism is necessary for the formation of metastatic colonies in lymph nodes. To investigate whether molecular subtype signatures derived from our Xenium data are associated with clinical outcomes, we evaluated their prognostic relevance using the METABRIC dataset [59].

Interestingly, the subtypes associated with poor prognosis were not clusters that formed metastatic colonies in lymph node (e.g., C4) or those showing strong mesenchymal features (C6) but rather clusters C5 and C3, which exhibited signatures of G2M cell cycle progression and increased glycolytic activity (Supplementary Fig. 6).

To validate this finding, we mapped the hybrid EM clusters identified in our Xenium analysis (C1–C6) to the corresponding cancer cell subclusters (sc2, sc4, sc8, sc10, and sc15) in the public scRNA-seq dataset. This revealed a one-to-one correspondence between C5 and sc8 and between C6 and sc15 (Fig. 5d). Consistent with our findings, only sc8 (corresponding to the C5 signature in our Xenium data) was significantly associated with poor prognosis [log-rank P = 2e-09, hazard ratio (HR) = 1.527, 95% confidence interval (CI) = 1.329–1.755] (Fig. 5e). In contrast, sc15, which displayed the highest mesenchymal programming and corresponded to the C6

signature in our Xenium data, tended to more favorable prognosis (log-rank P = 0.01, HR = 0.836, 95% CI = 0.727–0.961, Supplementary Fig. 7).

Discussion

In this study, we tracked the entire metastatic cascade of breast cancer at single-cell resolution and identified the relationship between spatial evolution of tumour cells and dynamic heterogeneity within the EMT spectrum. Our capture of early-stage metastatic events, including exceptionally early dissemination to lymph nodes, provides a rare in vivo snapshot of EMT dynamics during initial colonization. Thus, we were able to uncover distinct subpopulations with specific biological roles and their potential clinical relevance.

Through detailed analysis of pioneering metastatic tumour cells within metastatic cascade, we identified six distinct types of hybrid epithelial–mesenchymal (EM) cells that play key roles in the early stages of metastasis. The subpopulation that underwent a lipid metabolic shift successfully formed colonies in lymph nodes. This is consistent with previous studies indicating that FAO supports lymph node metastasis in breast cancer [8] . Interestingly, the subpopulation responsible for forming lymph node colonies that underwent a lipid metabolic shift, as well as the subpopulation with the most mesenchymal features, was not associated with poor clinical prognosis. This supports previous evidence that mesenchymal-like phenotypes may correlate with favorable outcomes in patients with breast cancer [60]. Instead, poor prognosis was linked to transcriptional programs associated with MYC and E2F signaling, along with increased aerobic glycolysis and G2/M cell cycle activity. In particular, MYC reprograms metabolism toward enhanced glycolytic flux [61], and breast cancers with enriched G2M signaling have been associated with elevated MYC pathway activity, increased proliferation, early distant metastasis,

and worse survival [46]. Collectively, these findings suggest that distinct EMT subpopulations drive different aspects of tumor progression: some mediate niche colonization, whereas others may drive aggressive metastasis and recurrence, implying the need for tailored therapeutic strategies [11]. A key question concerns how hybrid EM cells disseminate in lymph nodes: do metastases arise via the sequential model, where fully EMT-induced M cells colonize distant sites by MET, or via the cooperative model, which mesenchymal cells support Epithelial cells, which serve as metastasis-initiating cells [62]? Both models are considered plausible based on prior studies. In the present study, the observation of multiple EMT states in ITCs at an extremely early metastatic stage (Fig. 3a), along with similar mixtures of hybrid EM cells in the primary tumor (Fig. 3h), supports the cooperative metastasis model, where different cell states collaborate to initiate metastasis.

We also examined the early interactions between metastatic tumor cells and the lymph node microenvironment. Comparisons between the TdL and metastatic lymph nodes showed that a coordinated, immunosuppressive environment forms around tumor cells in newly colonized niches. These environments are likely shaped, in part, by the metabolic phenotypes of the tumor cells themselves. For example, lactic acid, a byproduct of glycolysis, has been shown to promote alternative splicing in T cells and enhance CTLA-4 expression in a Foxp3-dependent manner [63]. The plasticity of EMT observed in this study may contribute to the ecosystem that supports metastasis outgrowth, including immunomodulatory effects [64].

Despite the novel insights in this study, it has several limitations. First, our in-depth analysis was conducted on tissue from a single patient, limiting the generalizability of our conclusions. Second, owing to the limited number of cells available, we were unable to perform molecular or biochemical validations. However, by integrating our findings with prior knowledge and publicly

available single-cell RNA-seq datasets, we were able to infer the likely roles of key subpopulations and validate their clinical relevance. On the other hand, it is important to note that prognostic validation is indirect. In this analysis, we attempted to correlate the observed gene expression modules with the patient outcomes.

We have to carefully evaluate the confounding factor when the detected prognostic signatures of the tumor cells at the single cell level should be applied to the signature of the bulk tumor tissues. Neoplastic heterogeneity as well as the non-malignant cell types are represented in the latter profiles. Nevertheless, we believe this analysis has an important meaning, since the single cell analysis at this scale is not easy. Finally, we could not definitively distinguish whether the observed metastasis occurred via hematogenous or lymphatic dissemination. Future studies should also aim to elucidate whether distinct molecular features exist between these metastatic pathways.

Methods

Clinical sample

The analysis of tumor samples was performed in accordance with relevant national laws and recognized ethical guidelines (Declaration of Helsinki) for the protection of people participating in biomedical research. This study was approved by the Clinical Ethics Committee of St. Marianna University School of Medicine (approval number: 2297-i103). Informed consent was obtained from the patient. The surgical specimen was from an 80-year-old female patient, who received surgery for breast cancer (BRC-26). This study is based on samples taken from surgical residues that were available after histopathological analyses and were not required for diagnosis.

There is no interference with clinical practice. The patient had not received neoadjuvant or adjuvant chemotherapy (detailed clinicopathologic findings are provided in Supplementary Table 1). The same specimen was used in our previous study [65]. For multiregional, multi-omics analyses, fresh frozen samples of the primary tumor and adjacent normal tissue were used. Two nontumor regions (NC1 and NC2) and two tumor regions (Ca3, a solid tumor region, and Ca4, a sparse tumor region), as well as normal breast tissue from the same specimen (used as the control), were microdissected using the AVENIO MilliSect system (Roche, Pleasanton, CA) and subjected to multi-omics analysis, including RNA-seq, WGS, and EM-seq. For spatial analyses via Xenium, formalin-fixed paraffin-embedded (FFPE) samples of the primary tumor, lymph node metastases, and TdL were used.

RNA-seq

RNA-seq was conducted as previously described [65]. Briefly, total RNA was extracted from frozen normal breast tissue using the RNeasy Micro Kit (Qiagen). RNA-seq libraries were prepared using the SMART-Seq Stranded Kit (Takara Bio). Paired-end 150-bp sequencing was performed on the NovaSeq 6000 system (Illumina).

RNA-seq data analysis

RNA-seq data from the normal tissue obtained in this study and four tumor regions from our previous study [65] were analyzed. Adapter trimming was performed using fastp (v0.23.2) [66]. Reads mapping to rRNA were removed using Bowtie 2 (v2.3.4.3) [67]. The retained reads were aligned to the human reference genome GRCh38.p12 using STAR (v2.7.5c) [68]. Gene-level read counts were obtained using featureCounts (v2.0.2) [69], and RPKM values were calculated. DEGs

between histological classes were identified using DESeq2 (v1.42.1) [70] applying the Wald test with an adjusted P-value cutoff of <0.1. Gene enrichment analysis was performed using Metascape [71] .

Preparation of EM-seq and WGS libraries

EM-seq and WGS library preparations were conducted as described previously [72]. Briefly, gDNA was extracted from normal tissue using NucleoSpin Tissue XS (MACHEREY-NAGEL), and 100 ng of this gDNA was fragmented using the M220 Focused-ultrasonicator (Covaris). Adapter ligation was conducted using the NEBNext Enzymatic Methyl-seq Kit (New England BioLabs). Half of the adapter-ligated DNA was used for WGS library preparation (five cycles PCR amplification), and the other half was employed for EM-seq library preparation, which included TET oxidation, APOBEC conversion, and PCR amplification (six cycles). Libraries were sequenced as paired-end 150-bp reads on the NovaSeq 6000 system (Illumina).

EM-seq data analysis

EM-seq data from the normal tissue obtained in this study and four tumor regions from our previous study were analyzed following established protocols [65]. Briefly, adapter trimming was conducted using Trim Galore (v0.6.4_dev; https://github.com/FelixKrueger/TrimGalore). Trimmed reads were aligned to the human reference genome using Bismark (v0.22.1) [73]. Duplicate reads were removed using deduplicate_bismark, and methylation information in a CpG context was extracted via bismark_methylation_extractor. Genome-wide methylation profiles, including patterns in CpG islands, CpG shores, and promoter regions, were visualized as described previously [74]. Differentially methylated regions (DMRs) among NC-1, NC-2, Ca-3,

and Ca-4 and the broader NC and Ca regions were identified using metilene v0.2-8 [75], as per our previous study [65]. Promotor methylation was analyzed among NC-1, NC-2, Ca-3, and Ca-4 spanning 1 kb upstream to 500 bp downstream of transcription start sites, focusing on DMRs with methylation differences >10%. Visualisation was performed using IGV [76].

Point mutation detection

WGS reads were aligned to the human reference genome using BWA-MEM (v0.7.17) [77] with default settings. Duplicate reads were marked using Picard MarkDuplicates (v2.23.8) (https://broadinstitute.github.io/picard/). Somatic mutations were called using GATK Mutect2 and filtered with FilterMutectCalls (v4.1.3.0) [78]. Variants were annotated via ANNOVAR [79].

CNV detection

CNVs in NC-1, NC-2, Ca-3, Ca-4, and normal samples were detected using FACETS (v0.6.2), as described previously [74]. Copy number gains and losses were defined as ≥4 copies and ≤1 copy, respectively.

Spatial heatmaps of gene expression levels and average methylation rates

Spatial heatmaps of gene expression levels and average methylation rates were generated using a custom workflow implemented on Nikon's pilot analysis platform. Specifically, microscopic images (Olympus BX53, 2× objective) with substantial overlap were aligned using similarity transformation based on scale-invariant feature transform–detected feature point pairs. This alignment produced a composite image of the entire specimen with a common coordinate system. Manually dissected region contours were aligned to this coordinate system. Regions of

interest were color-coded by gene expression levels or average methylation rates. The average methylation rate was calculated as the mean frequency of methylation at CpG sites covered by at least five reads within the predefined promoter regions.

In situ gene expression analysis via Xenium

Spatial subclonal analysis at single-cell resolution was performed using the Xenium Slides & Sample Prep Reagents (10× Genomics), a predesigned human breast panel, and a custom panel, as previously described [80]. Briefly, FFPE tissue sections (5 μm thickness) of the primary tumor, lymph node metastases, and TdL were mounted onto Xenium slides. Following deparaffinization and decrosslinking, probe hybridization, probe circularization, and rolling circle amplification were performed. Detection of amplified probes was conducted using the Xenium Analyzer (10× Genomics). In total, 380 target genes, comprising 280 from the predesigned panel and 100 from the custom panel, are listed in Supplementary Table 1.

Computational processing of Xenium *in situ* expression data and analysis

Raw output files from the Xenium Analyzer were processed using Seurat (v5.0.2) [81]. Data normalization was performed using the SCTransform method. Clustering analysis and UMAP visualization were conducted using the first 30 principal components, with the former performed via the FindNeighbors and FindClustersing functions. The same parameters were applied to subclustering analyses (Figs. 2c and 3). For each cluster relevant to downstream analysis, DEGs were identified using the FindMarkers() function in Seurat with the default parameters (only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25). Cluster annotation was based on previously published gene expression signatures [37-39] and spatial histological context.

Trajectory analysis

Trajectory and pseudotime analysis of luminal cells was performed using Monocle 3 (v1.3.4) [41]. Preprocessing was performed using the preprocess_cds function, with the number of dimensions set to 100. Dimensionality reduction and clustering were implemented using the reduce_dimension and cluster_cells functions, respectively. Principal graph learning was conducted with the learn_graph function. Based on spatial and Xenium data, mature luminal cell 2 was designated as the trajectory root. Pseudotime ordering was visualized using the plot_cells function.

Identifying enriched gene signatures in Xenium clusters

To investigate the functional characteristics of Xenium clusters (C1–C6 in Fig. 3e), pathway and gene set enrichment analyses were performed using Metascape [71] based on the top 50 DEGs per cluster (Supplementary Table 5). Enrichment was assessed across GO categories, including Biological Process, Cellular Component, and Molecular Function, as well as other biological pathways.

Analysis of intercellular communication networks

Intercellular communication within metastatic lymph nodes and the TdL was analyzed using the CellChat package (v1.6.1) [82], which quantitatively identifies cell–cell interactions and communications. Following the official pipeline (https://github.com/jinworks/CellChat), analyses were performed using default settings. Statistically significant ligand–receptor

interactions ($P < 0.05$) were extracted. Visualization of cell–cell interactions was achieved using circle plots, bar plots, heatmaps, and violin plots.

Multiplexed immunostaining by PhenoCycler

Multiplexed immunostaining was performed using the PhenoCycler system (Akoya Biosciences) according to the manufacturer's instruction. Briefly, FFPE tissue sections were prepared at 5-µm thickness adjacent to those for Visium analysis and mounted onto the coverslip. The section was deparaffinized, and antigen activation was performed using pressure cooker for 20 min. Then, tissue sections were stained with 11 antibodies for 3 h (Supplementary Table 8). The section was washed, and the antibodies were fixed. Imaging analysis was conducted using the PhenoCycler instrument (Akoya Biosciences) and BZ-X810 fluorescence microscope (Keyence).

Computational processing of PhenoCycler multiplexed immunostaining data

Data processing was performed using the CODEX Processor (version 1.8). Visualization was performed using the obtained QPTIFF file by QuPath (version 0.3.2) [83]. Cell segmentation was performed using StarDist (QuPath StarDist extension, version 0.3.2)[84] on the QuPath software. Pixels in the QPTIFF images encoded as 8-bit integers (0–255) were used as expression levels of each protein.

Published dataset processing

We downloaded a publicly available single-cell RNA-seq dataset (GSE180286 from Guan et al. [4]). Data integration and clustering were performed in Seurat (v5.0.2) [81], focusing on HER2-positive breast tumors and two lymph node samples (GSM5457205, GSM5457206, and

GSM5457207). Louvain clustering at a resolution of 0.1 yielded nine clusters: one T cell cluster, three B cell clusters, one NK cell cluster, one CAF1 cluster, and one macrophage cluster. Clusters of the same cell type were merged (Fig. 5a). Subclustering was performed on the cancer cell populations (Fig. 5b).

Clinical validation using METABRIC data

Clinical validation was conducted using the METABRIC dataset [59]. METABRIC gene transcriptome data, as well as clinical and sample level metadata, were downloaded from cBioPortal (https://www.cbioportal.org). Analysis was performed on 1,980 breast cancer cases with available survival outcomes. For each cancer cluster defined in Fig. 5c, the top 100 DEGs (Supplementary Table 6) were used to score individual METABRIC samples. ROC curve analysis was used to define optimal cutoffs for stratifying patients into high and low score groups. Disease-free survival was analyzed using the survival (v3.5-7) and survminer (v0.4.9) R packages. HRs and 95% CIs were calculated using Cox proportional hazards models. Statistical significance was evaluated via the log-rank test.

Statistical Analysis

Statistical methods and tests are detailed in the figure legends. All analyses were performed using R (v4.3.2) or Python.

Data availability

The Xenium and bulk multi-omics data supporting the findings of this study are available at the DDBJ Japanese Genotype-phenotype Archive (https://gr-sharingdbs.dbcls.jp) under accession

number JGAD000946. The single-cell RNA-seq dataset from Guan et al. [4] was downloaded from Gene Expression Omnibus (GSE180286). METABRIC transcriptome and clinical data were downloaded from cBioPortal (https://www.cbioportal.org). Additional data are available from the corresponding authors upon reasonable request.

Acknowledgements: We thank K. Imamura, K. Abe, M. Satake, J. Zenkou, E. Sekimori, R. Fujinaga, and A. Gouda for their technical assistance. The authors would like to thank Enago (www.enago.jp) for the English language review.

Funding: This work was supported by the Japan Agency for Medical Research and Development (AMED) P-PROMOTE, grant number JP21ck0106700 (to S.N.), the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (KAKENHI Grants) Number JP24K11738 (Platform for Advanced Genome Science), as well as by Nikon Corporation. The supercomputing resource was provided by Human Genome Center, the University of Tokyo (http://sc.hgc.jp/shirokane.html). Computational resources were also supplied by Kashiwa-no-ha Omics Gate (https://www.kog.or.jp/en/server.html)

Author contributions: S.N. and K.K collected the human samples, performed experiments and computational analyses, and generated all figures and tables.

A.M., T.I., T.O., and K.T. aided in human sample collection.

T.O. generated Fig. 2a.

E.I. performed computational analyses.

M.K. and M.N. assisted in the Xenium experiments.

A.H., I.M. and J.K. conducted the pathological review.

A.M. and K.T. developed viewer software.

A.K. and A.S. offered advice and reviewed the manuscript.

S.N., M.S., and Y.S. designed the project and wrote the manuscript.

All authors read and approved the final manuscript.

Competing interests: A.M. and K.T. are employees of Nikon Corporation. The other authors do not have a competing interest.

References

1. Veronesi, U., et al. A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer. *N Engl J Med*. **349**, 546-553 (2003).
2. Turner, R. R., Ollila, D. W., Krasne, D. L. & Giuliano, A. E. Histopathologic validation of the sentinel lymph node hypothesis for breast carcinoma. *Ann Surg*. **226**, 271-276; discussion 276-278 (1997).
3. Nagasawa, S., et al. Genomic profiling reveals heterogeneous populations of ductal carcinoma in situ of the breast. *Commun Biol*. **4**, 438 (2021).
4. Xu, K., et al. Microenvironment components and spatially resolved single-cell transcriptome atlas of breast cancer metastatic axillary lymph nodes. *Acta Biochim Biophys Sin (Shanghai)*. **54**, 1336-1348 (2022).
5. Pesce, C. & Morrow, M. The need for lymph node dissection in nonmetastatic breast cancer. *Annu Rev Med*. **64**, 119-129 (2013).
6. de Boer, M., van Dijck, J. A., Bult, P., Borm, G. F. & Tjan-Heijnen, V. C. Breast cancer prognosis and occult lymph node metastases, isolated tumor cells, and micrometastases. *J Natl Cancer Inst*. **102**, 410-425 (2010).
7. Hinshaw, D. C. & Shevde, L. A. The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res*. **79**, 4557-4566 (2019).
8. Lee, C. K., et al. Tumor metastasis to lymph nodes requires YAP-dependent metabolic adaptation. *Science*. **363**, 644-649 (2019).
9. Cohen, M. R. Check concentrations of medications ordered for children. *Nursing*. **16**, 8 (1986).

10. Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. *Cell*. **166**, 21-45 (2016).
11. Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol*. **16**, 488-494 (2014).
12. Rubtsova, S. N., Zhitnyak, I. Y. & Gloushankova, N. A. Dual role of E-cadherin in cancer cells. *Tissue Barriers*. **10**, 2005420 (2022).
13. Fischer, K. R., et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature*. **527**, 472-476 (2015).
14. Ye, X., et al. Upholding a role for EMT in breast cancer metastasis. *Nature*. **547**, E1-E3 (2017).
15. Diepenbruck, M. & Christofori, G. Epithelial-mesenchymal transition (EMT) and metastasis: yes, no, maybe? *Curr Opin Cell Biol*. **43**, 7-13 (2016).
16. Aiello, N. M., et al. Upholding a role for EMT in pancreatic cancer metastasis. *Nature*. **547**, E7-E8 (2017).
17. Pastushenko, I., et al. Identification of the tumour transition states occurring during EMT. *Nature*. **556**, 463-468 (2018).
18. Kroger, C., et al. Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proc Natl Acad Sci U S A*. **116**, 7353-7362 (2019).
19. Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell*. **22**, 725-736 (2012).
20. Haerinck, J., Goossens, S. & Berx, G. The epithelial-mesenchymal plasticity landscape: principles of design and mechanisms of regulation. *Nat Rev Genet*. **24**, 590-609 (2023).
21. Bian, X., et al. Carboxylic acid metabolism in cancer: Mechanisms, microenvironment interactions, and therapeutic opportunities. *Metabolism*. **171**, 156334 (2025).
22. Vymetalkova, V., Vodicka, P., Vodenkova, S., Alonso, S. & Schneider-Stock, R. DNA methylation and chromatin modifiers in colorectal cancer. *Mol Aspects Med*. **69**, 73-92 (2019).
23. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics*. **1**, 239-259 (2009).
24. Rodriguez, J., et al. Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer Res*. **66**, 8462-9468 (2006).
25. Usman, S., et al. Vimentin Is at the Heart of Epithelial Mesenchymal Transition (EMT) Mediated Metastasis. *Cancers (Basel)*. **13** (2021).
26. Wang, Y., Shi, J., Chai, K., Ying, X. & Zhou, B. P. The Role of Snail in EMT and Tumorigenesis. *Curr Cancer Drug Targets*. **13**, 963-972 (2013).

27. DaSilva-Arnold, S. C., et al. ZEB2, a master regulator of the epithelial-mesenchymal transition, mediates trophoblast differentiation. *Mol Hum Reprod*. **25**, 61-75 (2019).
28. Sekimata, M., Murakami-Sekimata, A. & Homma, Y. CpG methylation prevents YY1-mediated transcriptional activation of the vimentin promoter. *Biochem Biophys Res Commun*. **414**, 767-772 (2011).
29. Chen, Y., Wang, K., Qian, C. N. & Leach, R. DNA methylation is associated with transcription of Snail and Slug genes. *Biochem Biophys Res Commun*. **430**, 1083-1090 (2013).
30. Xie, H., et al. Significance of ZEB2 in the immune microenvironment of colon cancer. *Front Genet*. **13**, 995333 (2022).
31. Valdivia, A., Avalos, A. M. & Leyton, L. Thy-1 (CD90)-regulated cell adhesion and migration of mesenchymal cells: insights into adhesomes, mechanical forces, and signaling pathways. *Front Cell Dev Biol*. **11**, 1221306 (2023).
32. Brenet, M., et al. Thy-1 (CD90)-Induced Metastatic Cancer Cell Migration and Invasion Are beta3 Integrin-Dependent and Involve a Ca(2+)/P2X7 Receptor Signaling Axis. *Front Cell Dev Biol*. **8**, 592442 (2020).
33. van Zijl, F., et al. Hepatic tumor-stroma crosstalk guides epithelial to mesenchymal transition at the tumor edge. *Oncogene*. **28**, 4022-4033 (2009).
34. Sethi, S., Macoska, J., Chen, W. & Sarkar, F. H. Molecular signature of epithelial-mesenchymal transition (EMT) in human prostate cancer bone metastasis. *Am J Transl Res*. **3**, 90-99 (2010).
35. Tinsley, E., Bredin, P., Toomey, S., Hennessy, B. T. & Furney, S. J. KMT2C and KMT2D aberrations in breast cancer. *Trends Cancer*. **10**, 519-530 (2024).
36. Nishimura, T., et al. Evolutionary histories of breast cancer and related clones. *Nature*. **620**, 607-614 (2023).
37. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. **18**, 220 (2017).
38. Lim, E., et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med*. **15**, 907-913 (2009).
39. Pliner, H. A., Shendure, J. & Trapnell, C. Supervised classification enables rapid annotation of cell atlases. *Nat Methods*. **16**, 983-986 (2019).
40. Nguyen, Q. H., et al. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat Commun*. **9**, 2028 (2018).
41. Qiu, X., et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. **14**, 979-982 (2017).

42. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. **15**, 178-196 (2014).
43. Zeng, Y. A. & Nusse, R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell*. **6**, 568-577 (2010).
44. Zhang, Y. C., Zhang, Y. T., Wang, Y., Zhao, Y. & He, L. J. What role does PDL1 play in EMT changes in tumors and fibrosis? *Front Immunol*. **14**, 1226038 (2023).
45. Imodoye, S. O. & Adedokun, K. A. EMT-induced immune evasion: connecting the dots from mechanisms to therapy. *Clin Exp Med*. **23**, 4265-4287 (2023).
46. Oshi, M., et al. G2M Cell Cycle Pathway Score as a Prognostic Biomarker of Metastasis in Estrogen Receptor (ER)-Positive Breast Cancer. *Int J Mol Sci*. **21** (2020).
47. Graziano, F., et al. Glycolysis gene expression analysis and selective metabolic advantage in the clinical progression of colorectal cancer. *Pharmacogenomics J*. **17**, 258-264 (2017).
48. Sacco, J. L. & Gomez, E. W. Epithelial-Mesenchymal Plasticity and Epigenetic Heterogeneity in Cancer. *Cancers (Basel)*. **16** (2024).
49. Teeuwssen, M. & Fodde, R. Cell Heterogeneity and Phenotypic Plasticity in Metastasis Formation: The Case of Colon Cancer. *Cancers (Basel)*. **11** (2019).
50. Keller, H., et al. Fatty acids and retinoids control lipid metabolism through activation of peroxisome proliferator-activated receptor-retinoid X receptor heterodimers. *Proc Natl Acad Sci U S A*. **90**, 2160-2164 (1993).
51. Galimberti, V., et al. Axillary dissection versus no axillary dissection in patients with sentinel-node micrometastases (IBCSG 23-01): a phase 3 randomised controlled trial. *Lancet Oncol*. **14**, 297-305 (2013).
52. Martinez-Pomares, L. The mannose receptor. *J Leukoc Biol*. **92**, 1177-1186 (2012).
53. Schuette, V., et al. Mannose receptor induces T-cell tolerance via inhibition of CD45 and up-regulation of CTLA-4. *Proc Natl Acad Sci U S A*. **113**, 10649-10654 (2016).
54. Luca, M. E., et al. Mannosylated PLP(139-151) induces peptide-specific tolerance to experimental autoimmune encephalomyelitis. *J Neuroimmunol*. **160**, 178-187 (2005).
55. Kel, J., et al. Soluble mannosylated myelin peptide inhibits the encephalitogenicity of autoreactive T cells during experimental autoimmune encephalomyelitis. *Am J Pathol*. **170**, 272-280 (2007).
56. Choi, J. H., et al. Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. *Nat Commun*. **14**, 1055 (2023).
57. Wang, T., et al. COL4A1 promotes the growth and metastasis of hepatocellular carcinoma cells by activating FAK-Src signaling. *J Exp Clin Cancer Res*. **39**, 148 (2020).

58. Han, J., et al. CXCL16 Promotes Gastric Cancer Tumorigenesis via ADAM10-Dependent CXCL16/CXCR6 Axis and Activates Akt and MAPK Signaling Pathways. *Int J Biol Sci*. **17**, 2841-2852 (2021).
59. Curtis, C., et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. **486**, 346-352 (2012).
60. Grosse-Wilde, A., et al. Loss of inter-cellular cooperation by complete epithelial-mesenchymal transition supports favorable outcomes in basal breast cancer patients. *Oncotarget*. **9**, 20018-20033 (2018).
61. Dong, Y., Tu, R., Liu, H. & Qing, G. Regulation of cancer cell metabolism: oncogenic MYC in the driver's seat. *Signal Transduct Target Ther*. **5**, 124 (2020).
62. Tsuji, T., Ibaragi, S. & Hu, G. F. Epithelial-mesenchymal transition and cell cooperativity in metastasis. *Cancer Res*. **69**, 7135-7139 (2009).
63. Ding, R., et al. Lactate modulates RNA splicing to promote CTLA-4 expression in tumor-infiltrating regulatory T cells. *Immunity*. **57**, 528-540 e526 (2024).
64. Reticker-Flynn, N. E., et al. Lymph node colonization induces tumor-immune tolerance to promote distant metastasis. *Cell*. **185**, 1924-1942 e1923 (2022).
65. Keisuke Kunigo, et al. Deep long-read methylation analysis using targeted nanopore sequencing and application to local region analysis of clinical specimen, 2025).
66. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. **34**, i884-i890 (2018).
67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. **9**, 357-359 (2012).
68. Dobin, A., et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15-21 (2013).
69. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923-930 (2014).
70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550 (2014).
71. Zhou, Y., et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. **10**, 1523 (2019).
72. Sakamoto, Y., et al. Long-read whole-genome methylation patterning using enzymatic base conversion and nanopore sequencing. *Nucleic Acids Res*. **49**, e81 (2021).
73. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. **27**, 1571-1572 (2011).
74. Haga, Y., et al. Whole-genome sequencing reveals the molecular implications of the stepwise progression of lung adenocarcinoma. *Nat Commun*. **14**, 8375 (2023).

75. Juhling, F., et al. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res*. **26**, 256-262 (2016).
76. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. **14**, 178-192 (2013).
77. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. **26**, 589-595 (2010).
78. McKenna, A., et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **20**, 1297-1303 (2010).
79. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. **38**, e164 (2010).
80. Takano, Y., et al. Spatially resolved gene expression profiling of tumor microenvironment reveals key steps of lung adenocarcinoma development. *Nat Commun*. **15**, 10637 (2024).
81. Hao, Y., et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. **42**, 293-304 (2024).
82. Jin, S., et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. **12**, 1088 (2021).
83. Bankhead, P., et al. QuPath: Open source software for digital pathology image analysis. *Sci*. **7**, 1-7 (2017).
84. Schmidt, U., et al. Cell detection with star-convex polygons. *Lect. Notes Comput. Sci*. **11071 LNCS**, 265–273 (2018).

Figure Legends

Fig. 1 | Multi-omics analysis across regions of breast cancer tissue.
a)HE images of breast cancer tissue with regions sampled for multi-omics analysis. Two noncancerous regions (NC-1 and NC-2; pink lines) and two cancerous regions (Ca-3, a solid tumor region; Ca-4, a sparse tumor region; light blue lines) were dissected (also refer to Supplementary Fig. 1). Scale bar: 1 mm.
b) Gene enrichment analysis using DEGs from noncancerous (upper panel) and cancerous regions (lower panel). Bar graphs show the top 10 enriched clusters; color scale indicates −log10 adjusted P-values.

c) Violin plots showing genome-wide (50 kb window) methylation rates measured by nanoEM across each region. Red line indicate mean.P-values were calculated using the two-sided Wilcoxon rank-sum test with Bonferroni correction. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

d) Heatmaps showing Z-score–normalized gene expression levels (from TPM values) and promoter CpG methylation levels of selected genes involved in cancer metastasis in Ca-3 and Ca-4. Promoter regions were defined as 1 kb upstream to 500 bp downstream of the transcription start site. Genes with the most variable promoter methylation across NC-1, NC-2, Ca-3, Ca-4, and normal breast tissue were selected.

e) Spatial visualization of THY1 gene expression (left) and promotor methylation (right) using our custom viewer. Color scale represents Z-scores (normalized from TPM) for gene expression and average methylation rate for methylation.

f) Methylation pattern of THY1 detected by EM-seq visualized using the IGV; the middle panel shows a heatmap of methylation rates across regions.

g) Mutation status of breast cancer-related genes across regions. Categories include MAPK-PI3K signaling, cell cycle regulation, DNA damage and apoptosis, chromatin remodeling, cell adhesion, transcription regulation, and clinically relevant genes.

h) Representative mutations in KMT2C (Y987H) and KDM6A (P281R) across regions.

i) Genome-wide copy number profiles by chromosome for each region, with cancer-related genes showing gains (red) or losses (blue) indicated in the inset.

Fig. 2 | Spatial analysis of the breast cancer at single-cell resolution.

a)Workflow of the study. Primary tumor (P), metastatic lymph node (ML), tumor-draining lymph node (TdL) samples were collected and processed for Xenium analysis. Middle panel: HE staining; right panel: annotated major cell types from b).

b) UMAP of 1,005,436 cells analyzed via Xenium Explore, annotated by cell type using canonical markers and histological location. Right panel shows subpopulation of luminal cells (n = 245,385 cells).

c) UMAP visualization of luminal cells ; right panel shows subclustered labeled as mature luminal 1 (L1), mature luminal 2 (L2), basal (B), and mesenchymal like (M) cells.

d) Histological location and gene expression in clusters representing normal lobules. Violin plot shows log-normalized expression levels of markers for mature luminal (ANKRD30A and KIT), basal (KRT23) cells. Scale bar: 100 μm.

e) Histological location of M clusters. Violin plot shows log-normalized expression of DEGs for M clusters; bubble plot shows EMT-related gene expression. Dot size represents the percentage of cells expressing. Color indicates mean expression relative to other subsets. Scale bar: 500 μm.

f) Histological location in cancer cluster(ii)v and expression of L2 marker ANKRD30A.

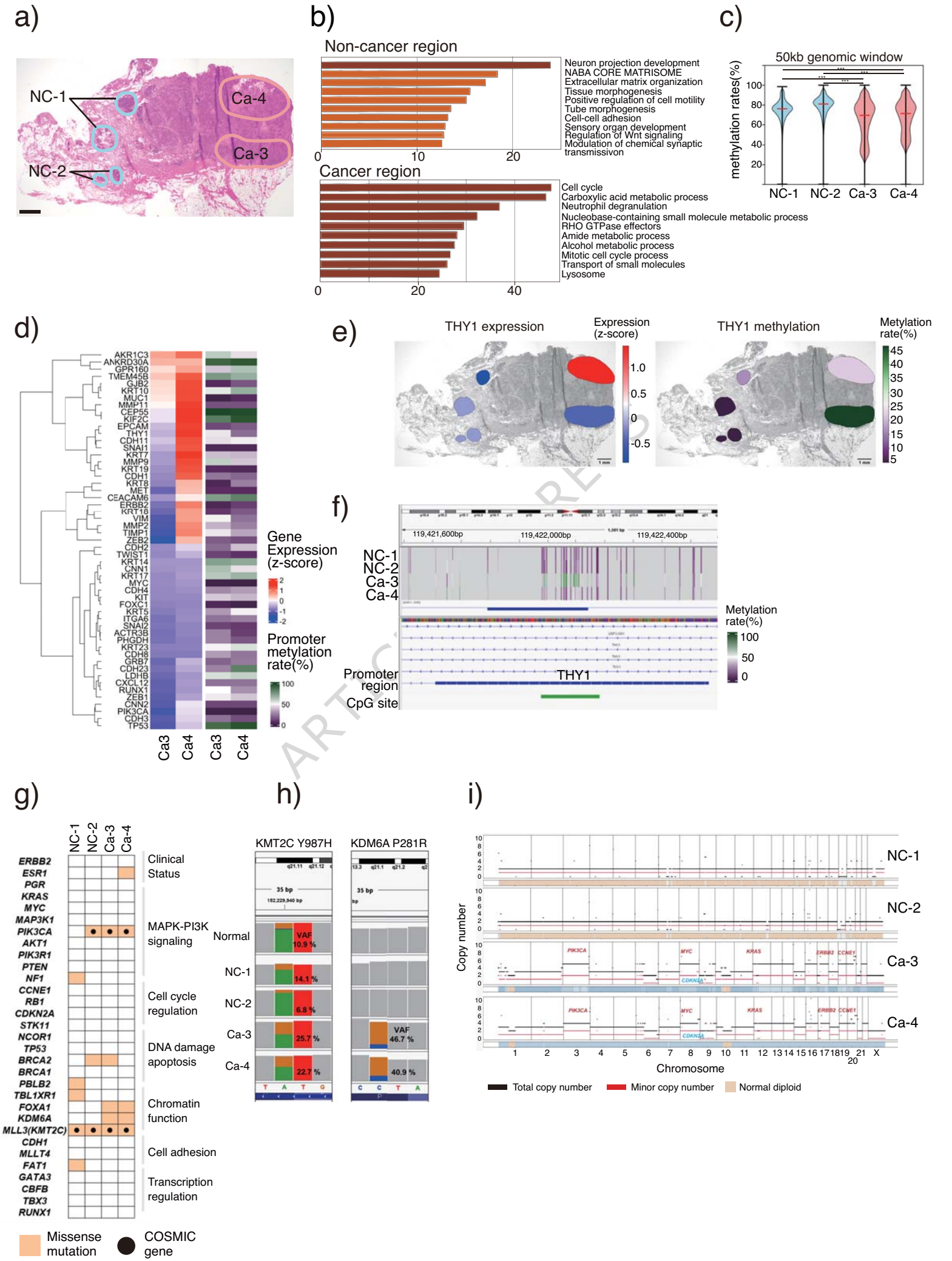
g) Monocle 3 trajectory analysis of luminal cell differentiation, revealing two main trajectories. The UMAP axes correspond to b). vCells are color-coded by pseudotime. Red winding lines in the objects indicate putative developmental trajectories.

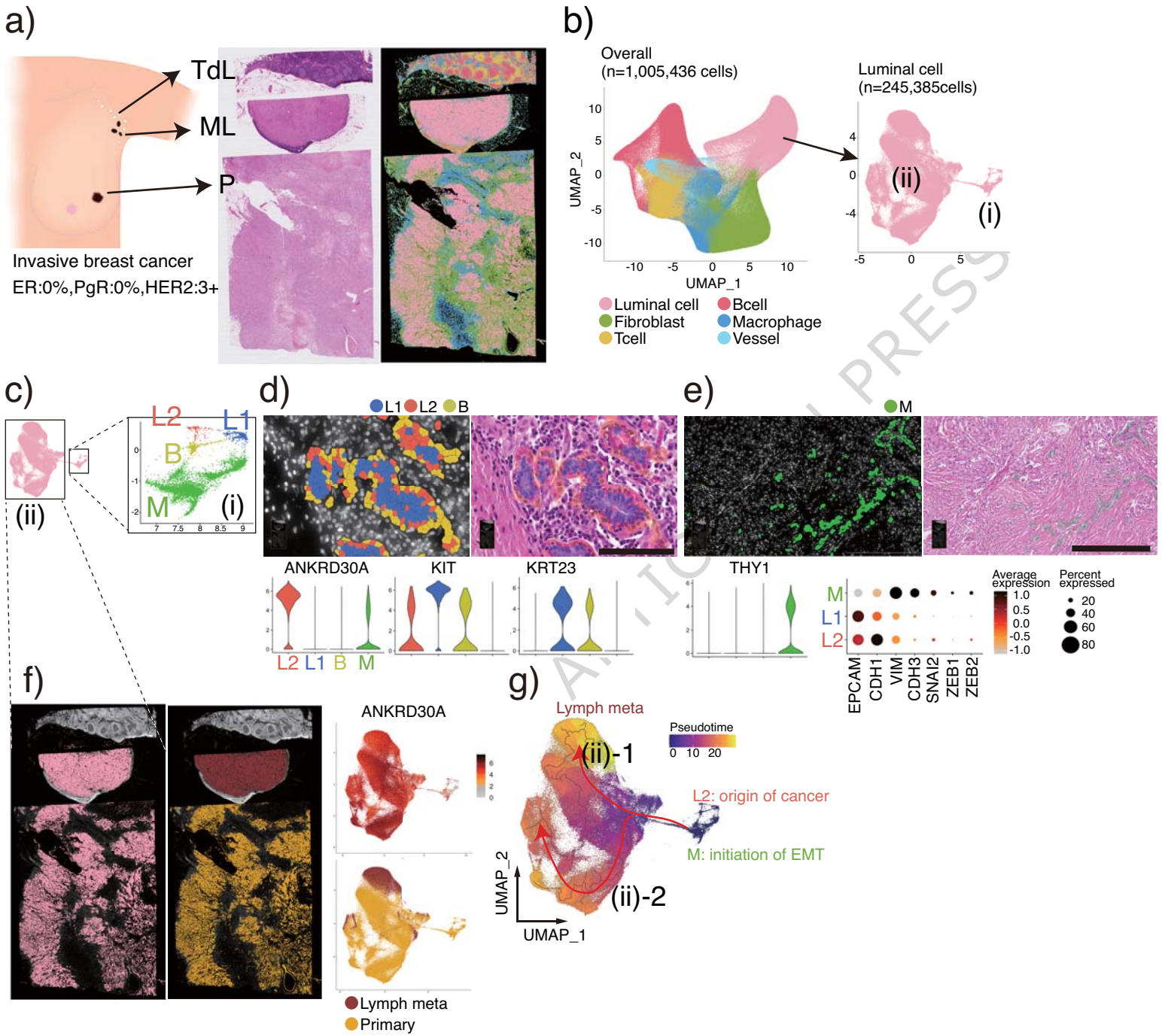
Fig. 3 | Spatial analysis of lymph nodes reveals six transcriptionally distinct EMT states.
a) Isolated tumor cells (ITCs) within the TdL. Enlarged view (upper panel), KRT19 expression (middle panel), and HE staining (lower panel). Scale bar: 100 μm.
b) UMAP placement of ITCs and reclustering of tumor cells, cell numbers per cluster shown in footnotes.
c) Histological localization of cancer cell clusters.
d) Bubble plot showing expression of related to EMT transcription factors, adhesion molecules, and stemness markers across clusters; violin plot shows representative marker expression.
e) Pathway enrichment using top 50 DEGs (Supplementary Table 5) in clusters C1–C6; heatmap shows enrichment scores (gray: not significant).
f) Xenium spatial plot of FASN and SCD transcripts. Color shows log-normalized expression levels.
g) Violin plot of G2M signature scores across cancer clusters using Seurat; P-values were determined via the Wilcoxon rank-sum test. *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.
h) Representation of cancer clusters within the primary tumor; cell counts shown in footnotes.

Fig. 4 | Analysis of the TME in metastatic lymph nodes.
a)ROI analyzed for cell–cell communication. Pie chart shows proportions of TME components. Scale bar: 100 μm.
b) Overview of intercellular interactions. Arrows and edge show directionality; color indicates cell type; circle size is proportional to cell count; edge thickness indicates the number of interactions between populations. Bar chart ranks signaling pathways by differential information flow between metastatic lymph nodes and the TdL (dark red shows enrichment in metastatic lymph nodes; dark green shows enrichment in the TdL).
c) Heatmaps showing CD45 signaling interactions in metastatic lymph nodes; color scale represents communication probabilities; empty squares show zero probability. Violin plot shows gene expression levels.
d) VEGF and CD86 signaling interactions in the TdL, shown as heatmaps and violin plots, as in c).

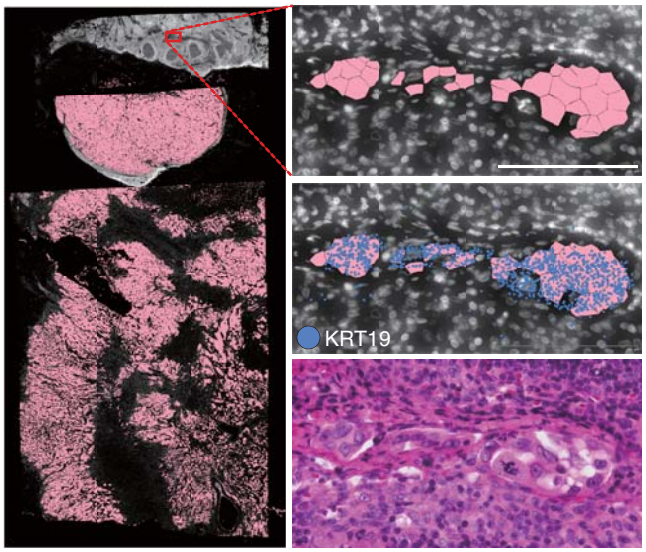
Fig. 5 | Transcriptional profiling of hybrid EMT and its association with patient outcomes.

a) UMAP annotated by dataset (left) and major cell types (right). Datasets: primary tumor (GSM5457205); lymph nodes (GSM5457206 and GSM5457207).
b) UMAP of reclustered tumor cells.
c) Dot plot compares metabolic pathway scores from MSigDB C2 gene sets across subclusters (sc2–sc15); color scale shows −log10-adjusted P-value.
d) Correlation of public tumor clusters (sc2, sc4, sc8, sc10, and sc15) with Xenium clusters (C1–C6). The top 50 DEGs (Supplementary Table 5) from Xenium clusters were used to generate signature scores. Violin and feature plots show Xenium signature scores. The UMAP axes correspond to b). C5 and C6 correspond to sc8 and sc15, respectively.
e) Survival analysis using the METABRIC cohort. Signature scores based on top 100 DEGs (Supplementary Table 6) for sc8 and sc15 were computed using UCell. Kaplan–Meier survival plots show disease-free survival stratified by high vs. low signature scores. P-values from log-rank test; adjusted HRs from univariate Cox proportional hazards model, with the low-scores as the reference.

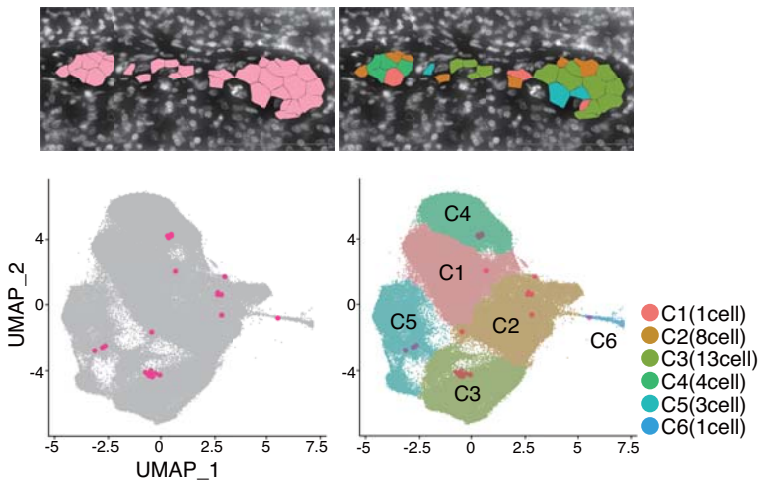

a)
NC-1
NC-2
Ca-4
Ca-3
b)
Non-cancer region
Neuron projection development
NABA CORE MATRISOME
Extracellular matrix organization
Tissue morphogenesis
Positive regulation of cell motility
Tube morphogenesis
Cell-cell adhesion
Sensory organ development
Regulation of Wnt signaling
Modulation of chemical synaptic transmissivon
Cancer region
Cell cycle
Carboxylic acid metabolic process
Neutrophil degranulation
Nucleobase-containing small molecule metabolic process
RHO GTPase effectors
Amide metabolic process
Alcohol metabolic process
Mitotic cell cycle process
Transport of small molecules
Lysosome
c)
50kb genomic window
methylation rates(%)
NC-1
NC-2
Ca-3
Ca-4
d)
Gene Expression (z-score)
Promoter metylation rate(%)
Ca3
Ca4
Ca3
Ca4
e)
THY1 expression
Expression (z-score)
THY1 methylation
Metylation rate(%)
f)
119,421,600bp
119,422,000bp
119,422,400bp
NC-1
NC-2
Ca-3
Ca-4
Metylation rate(%)
Promoter region
CpG site
THY1
g)
NC-1
NC-2
Ca-3
Ca-4
Clinical Status
MAPK-PI3K signaling
Cell cycle regulation
DNA damage apoptosis
Chromatin function
Cell adhesion
Transcription regulation
Missense mutation
COSMIC gene
h)
KMT2C Y987H
KDM6A P281R
Normal
NC-1
NC-2
Ca-3
Ca-4
VAF 10.9 %
14.1 %
6.8 %
25.7 %
22.7 %
VAF 46.7 %
40.9 %
i)
Copy number
Chromosome
Total copy number
Minor copy number
Normal diploid


a)
TdL
ML
P
Invasive breast cancer
ER:0%,PgR:0%,HER2:3+
b)
Overall
(n=1,005,436 cells)
UMAP_2
UMAP_1
Luminal cell
(n=245,385cells)
(ii)
(i)
Luminal cell
Bcell
Fibroblast
Macrophage
Tcell
Vessel
c)
L2
L1
B
M
(ii)
(i)
d)
L1 L2 B
ANKRD30A
KIT
KRT23
L2 L1 B M
e)
M
THY1
M
L1
L2
EPCAM
CDH1
VIM
CDH3
SNAI2
ZEB1
ZEB2
Average expression
Percent expressed
f)
ANKRD30A
Lymph meta
Primary
g)
Lymph meta
Pseudotime
(ii)-1
L2: origin of cancer
M: initiation of EMT
(ii)-2
UMAP_2
UMAP_1

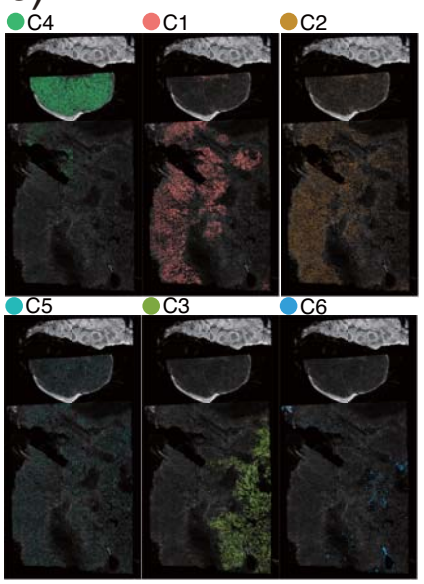
a)


KRT19

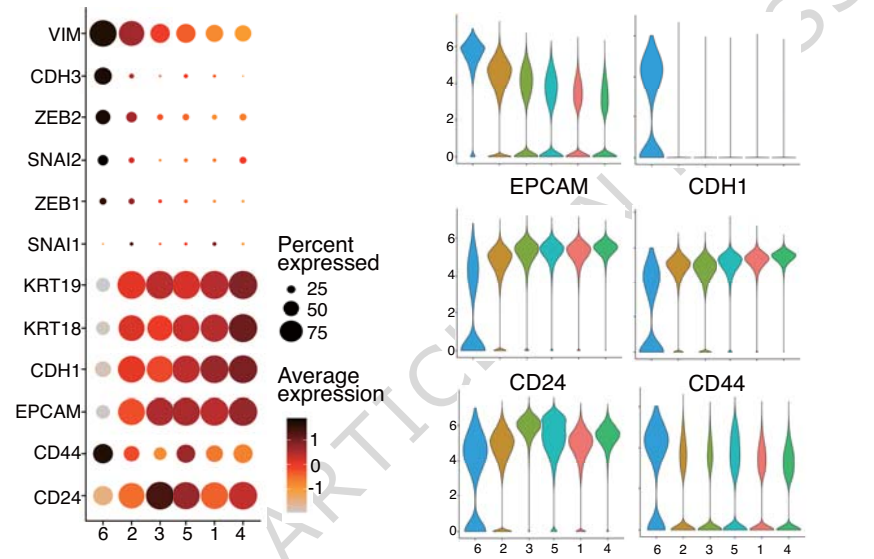
b)


UMAP_2
UMAP_1
C4
C1
C5
C2
C6
C3
C1(1cell)
C2(8cell)
C3(13cell)
C4(4cell)
C5(3cell)
C6(1cell)

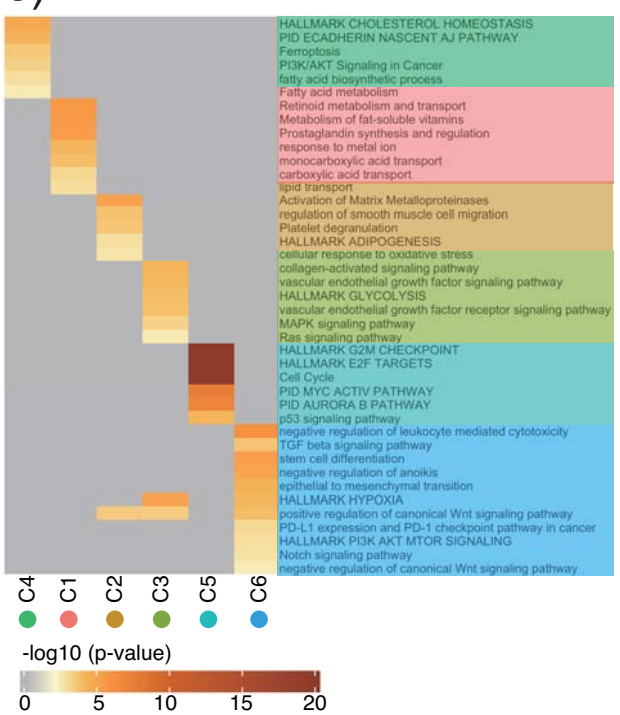
c)


C4
C1
C2
C5
C3
C6

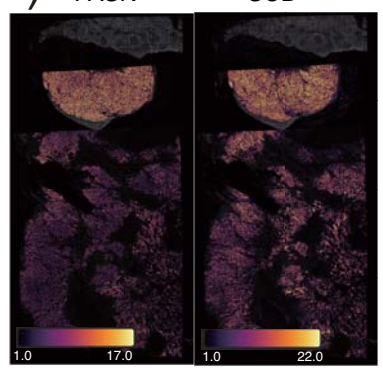
d)


VIM
CDH3
ZEB2
SNAI2
ZEB1
SNAI1
KRT19
KRT18
CDH1
EPCAM
CD44
CD24
6 2 3 5 1 4
Percent expressed
25
50
75
Average expression
1
0
-1
VIM
CDH3
EPCAM
CDH1
CD24
CD44

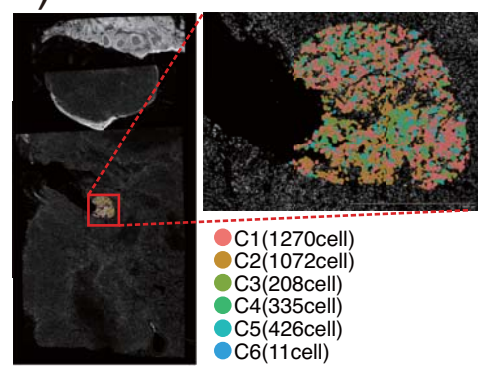
e)


HALLMARK CHOLESTEROL HOMEOSTASIS
PID ECADHERIN NASCENT AJ PATHWAY
Ferroptosis
PI3K/AKT Signaling in Cancer
fatty acid biosynthetic process
Fatty acid metabolism
Retinoid metabolism and transport
Metabolism of fat-soluble vitamins
Prostaglandin synthesis and regulation
response to metal ion
monocarboxylic acid transport
carboxylic acid transport
lipid transport
Activation of Matrix Metalloproteinases
regulation of smooth muscle cell migration
Platelet degranulation
HALLMARK ADIPOGENESIS
cellular response to oxidative stress
collagen-activated signaling pathway
vascular endothelial growth factor signaling pathway
HALLMARK GLYCOLYSIS
vascular endothelial growth factor receptor signaling pathway
MAPK signaling pathway
Ras signaling pathway
HALLMARK G2M CHECKPOINT
HALLMARK E2F TARGETS
Cell Cycle
PID MYC ACTIV PATHWAY
PID AURORA B PATHWAY
p53 signaling pathway
negative regulation of leukocyte mediated cytotoxicity
TGF beta signaling pathway
stem cell differentiation
negative regulation of anoikis
epithelial to mesenchymal transition
HALLMARK HYPOXIA
positive regulation of canonical Wnt signaling pathway
PD-L1 expression and PD-1 checkpoint pathway in cancer
HALLMARK PI3K AKT MTOR SIGNALING
Notch signaling pathway
negative regulation of canonical Wnt signaling pathway
C4 C1 C2 C3 C5 C6
-log10 (p-value)
0 5 10 15 20

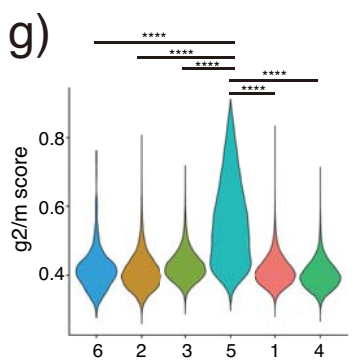
f)


FASN
SCD
1.0
17.0
1.0
22.0

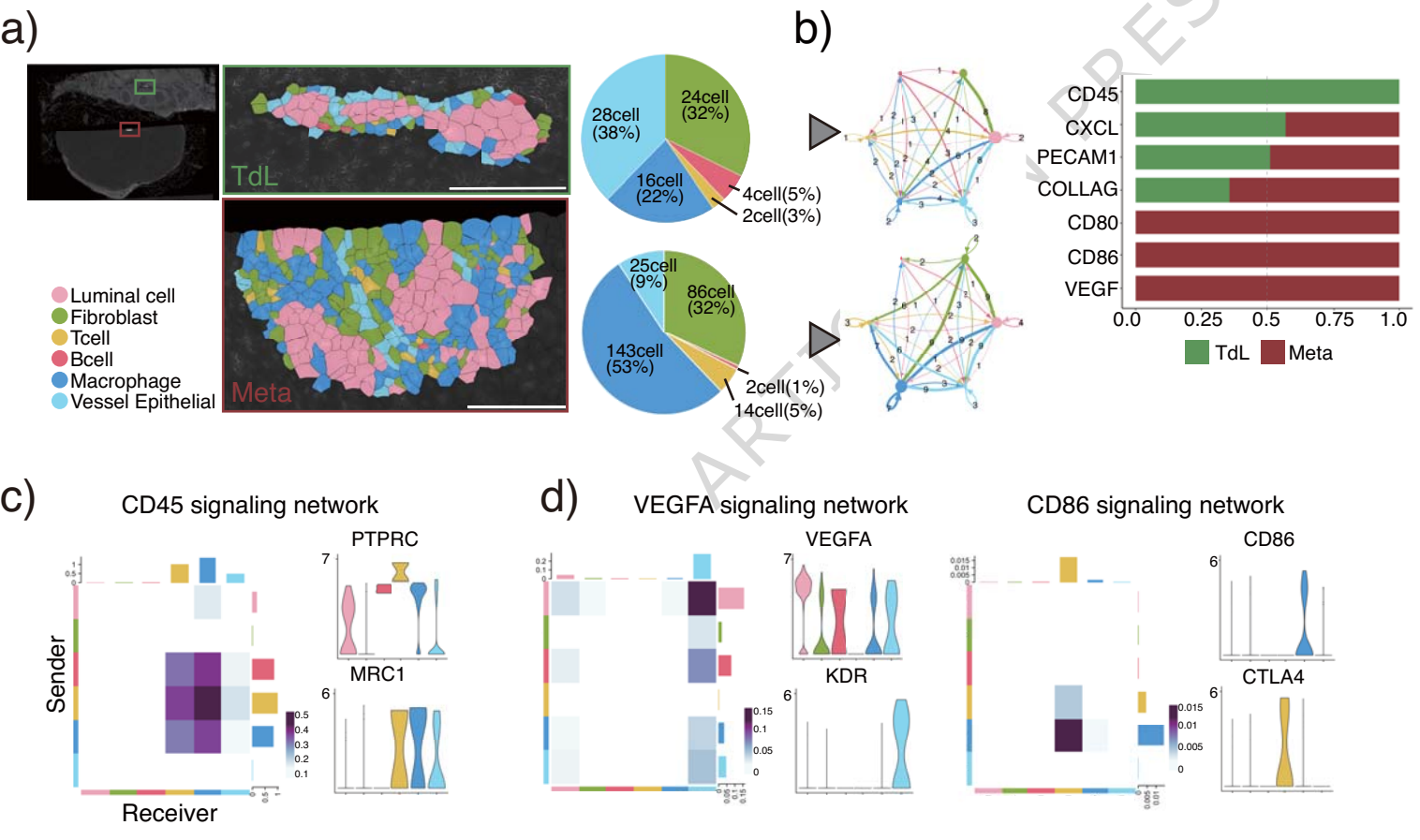
h)

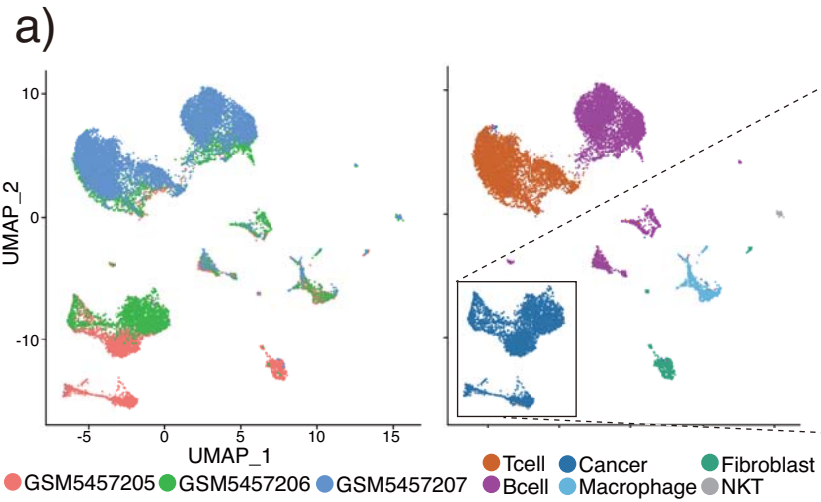

C1(1270cell)
C2(1072cell)
C3(208cell)
C4(335cell)
C5(426cell)
C6(11cell)

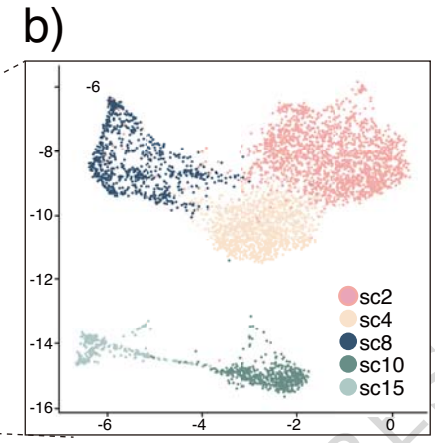
g)

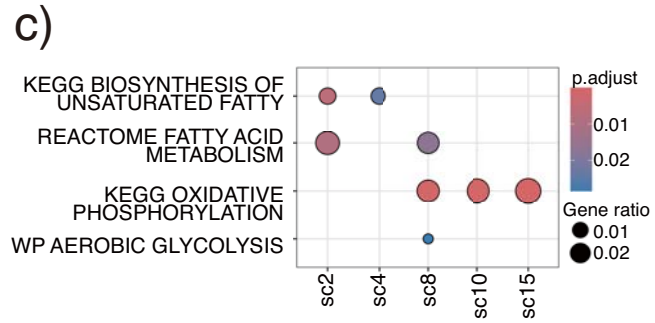

****
****
****
****
****
g2/m score
0.8
0.6
0.4
6 2 3 5 1 4

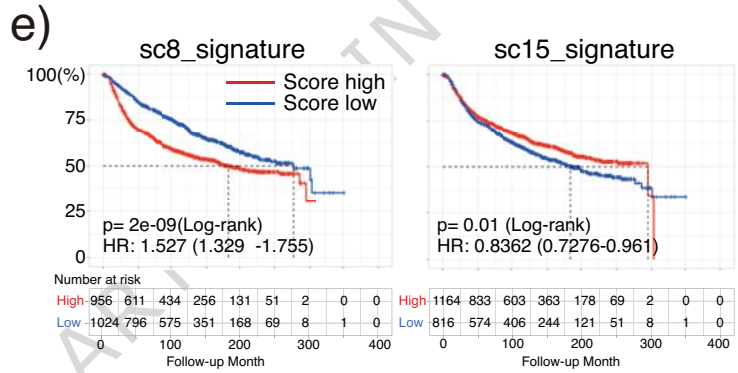
v

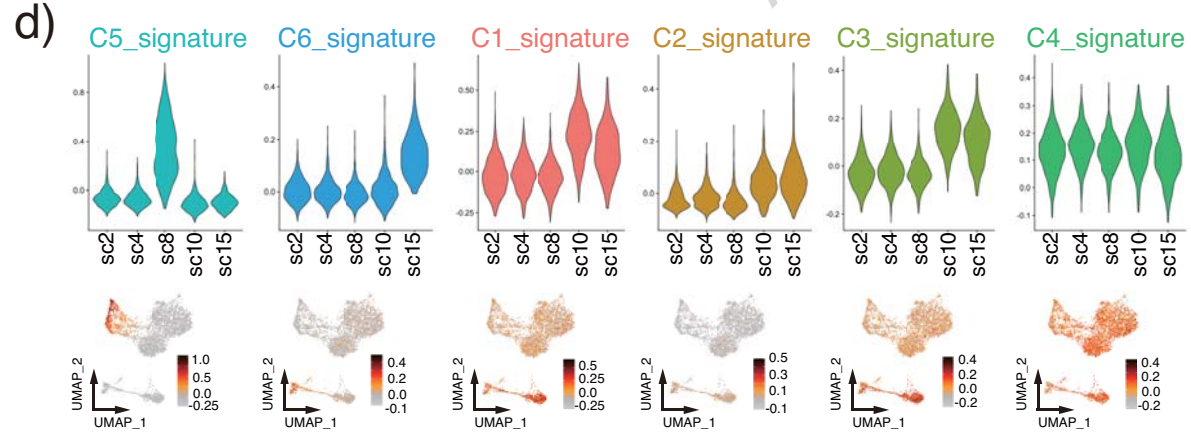



a)
UMAP_2
UMAP_1
GSM5457205
GSM5457206
GSM5457207
Tcell
Bcell
Cancer
Macrophage
Fibroblast
NKT


b)
sc2
sc4
sc8
sc10
sc15


c)
KEGG BIOSYNTHESIS OF UNSATURATED FATTY
REACTOME FATTY ACID METABOLISM
KEGG OXIDATIVE PHOSPHORYLATION
WP AEROBIC GLYCOLYSIS
p.adjust
Gene ratio
sc2
sc4
sc8
sc10
sc15


e)
sc8_signature
sc15_signature
Score high
Score low
p= 2e-09(Log-rank)
HR: 1.527 (1.329 -1.755)
p= 0.01 (Log-rank)
HR: 0.8362 (0.7276-0.961)
Number at risk
Follow-up Month


d)
C5_signature
C6_signature
C1_signature
C2_signature
C3_signature
C4_signature
UMAP_1
UMAP_2