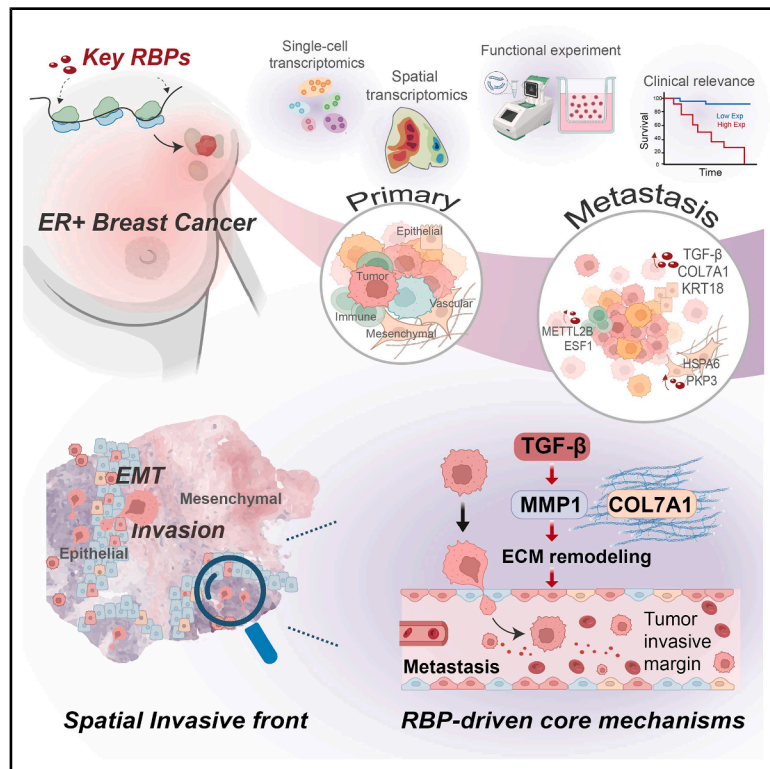


Single-cell transcriptomics profiling elucidates RBP-driven metastatic signaling pathways in ER⁺ breast cancer

Graphical abstract



Authors

Mingjie Dong, Xinyu Li, Songyu Yang, ..., Weixin Liang, Xia Li, Jing Bai

Correspondence

baijing@hrbmu.edu.cn

In brief

Biological sciences

Highlights

- A dynamic single-cell atlas delineating ER⁺ breast cancer metastatic progression
- Metastasis-enriched RBPs drive cytoskeletal remodeling and RNA metabolic rewiring
- RBP-mediated signaling gradients define invasion fronts in breast cancer
- TGF- β -centered signaling axes underlying metastatic progression



Article

Single-cell transcriptomics profiling elucidates RBP-driven metastatic signaling pathways in ER⁺ breast cancer

Mingjie Dong,^{1,3} Xinyu Li,^{1,3} Songyu Yang,^{1,3} Shujuan Li,¹ Zhengzheng Liu,¹ Chuo Peng,¹ Rongyao Li,¹ Weixin Liang,¹ Xia Li,^{1,2} and Jing Bai^{1,4,*}

¹College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China

²College of Biomedical Information and Engineering, Hainan Medical University, Haikou, Hainan, China

³These authors contributed equally

⁴Lead contact

*Correspondence: baijing@hrbmu.edu.cn

<https://doi.org/10.1016/j.isci.2026.116155>

SUMMARY

Estrogen receptor-positive breast cancer progression involves extensive cellular remodeling, yet the contribution of RNA-binding proteins to this process remains incompletely defined. Here, we constructed a single-cell transcriptomic atlas of 198,286 cells from 39 samples spanning normal tissue, primary tumors, and metastatic lesions. We identified stage-specific transcriptional and post-transcriptional reprogramming across different cell types. Several RBPs, including *KRT18*, *HSPA6*, and *PKP3*, were progressively upregulated during malignant progression and associated with cytoskeletal remodeling, extracellular matrix interactions, and stress adaptation. Cell-cell communication analyses revealed metastasis-associated signaling programs involving TIMP1-MMP1 and TGFB1-COL7A1/MMP1, which were spatially enriched at tumor margins and invasive fronts. Functional perturbation experiments in ER⁺ breast cancer cells showed that TGF- β stimulation enhanced invasive behavior, whereas knockdown of MMP1 or COL7A1 attenuated epithelial-mesenchymal transition (EMT) marker expression and invasion. Together, our study delineates RBP-associated regulatory programs linking cellular state transitions to invasive signaling in ER⁺ breast cancer.

INTRODUCTION

Estrogen receptor-positive (ER⁺) breast cancer represents the most common molecular subtype, accounting for approximately 70% of all breast cancer cases.^{1,2} Although endocrine therapies, such as selective estrogen receptor modulators, are initially effective, a significant proportion of patients eventually experience metastatic relapse, which remains the leading cause of mortality. Despite recent advances in deciphering the molecular underpinnings of breast cancer progression,^{3,4} the mechanistic transition from primary tumors to metastatic lesions, particularly the reprogramming of the tumor microenvironment (TME), is still not fully understood. RNA-binding proteins (RBPs), as central components of the post-transcriptional regulatory machinery, modulate mRNA splicing, stability, transport, and translation.⁵ Disruption of RBP function or expression is increasingly recognized as a driver of oncogenic transformation, therapeutic resistance, and tumor evolution.⁶ In metastatic settings, RBPs regulate the fate of metastasis-associated transcripts, facilitating epithelial plasticity, immune evasion, and cytoskeletal adaptation.^{7,8} However, the broader landscape of RBP-mediated regulatory networks, especially their roles in promoting metastatic phenotypes and modulating intercellular communication within the TME, remains largely uncharacterized.

Recent breakthroughs in single-cell RNA sequencing (scRNA-seq) have provided transformative insights into intra-tumor heterogeneity, cellular plasticity, and lineage hierarchies within the TME.⁹ Unlike bulk transcriptomic approaches, scRNA-seq enables deconvolution of gene expression programs at single-cell resolution, capturing rare cell populations, transitional states, and signaling networks. Ozmen et al.¹⁰ and Xu et al.¹¹ used single-cell and spatial transcriptomic approaches to define cellular states and MMP1⁺ malignant subsets in breast cancer, establishing transcriptional remodeling and extracellular matrix (ECM)-driven tumor-immune interactions as key features of disease progression, yet without elucidating the post-transcriptional regulatory mechanisms underlying these phenotypes. Notably, recent studies have begun to reveal cell-type-specific RBP-RNA interactions at single-cell resolution,¹² establishing RBPs as essential effectors of transcriptome fine-tuning and phenotypic heterogeneity. Given their pivotal role in coordinating post-transcriptional regulatory logic, RBPs are positioned to serve as signal integration hubs that regulate epithelial-mesenchymal transition (EMT), immune rewiring, angiogenic remodeling, and the formation of metastatic niches. Jiang et al.¹³ identified oncogenic RBPs in lung adenocarcinoma using a machine learning framework, but their bulk-based, single-cancer analysis could



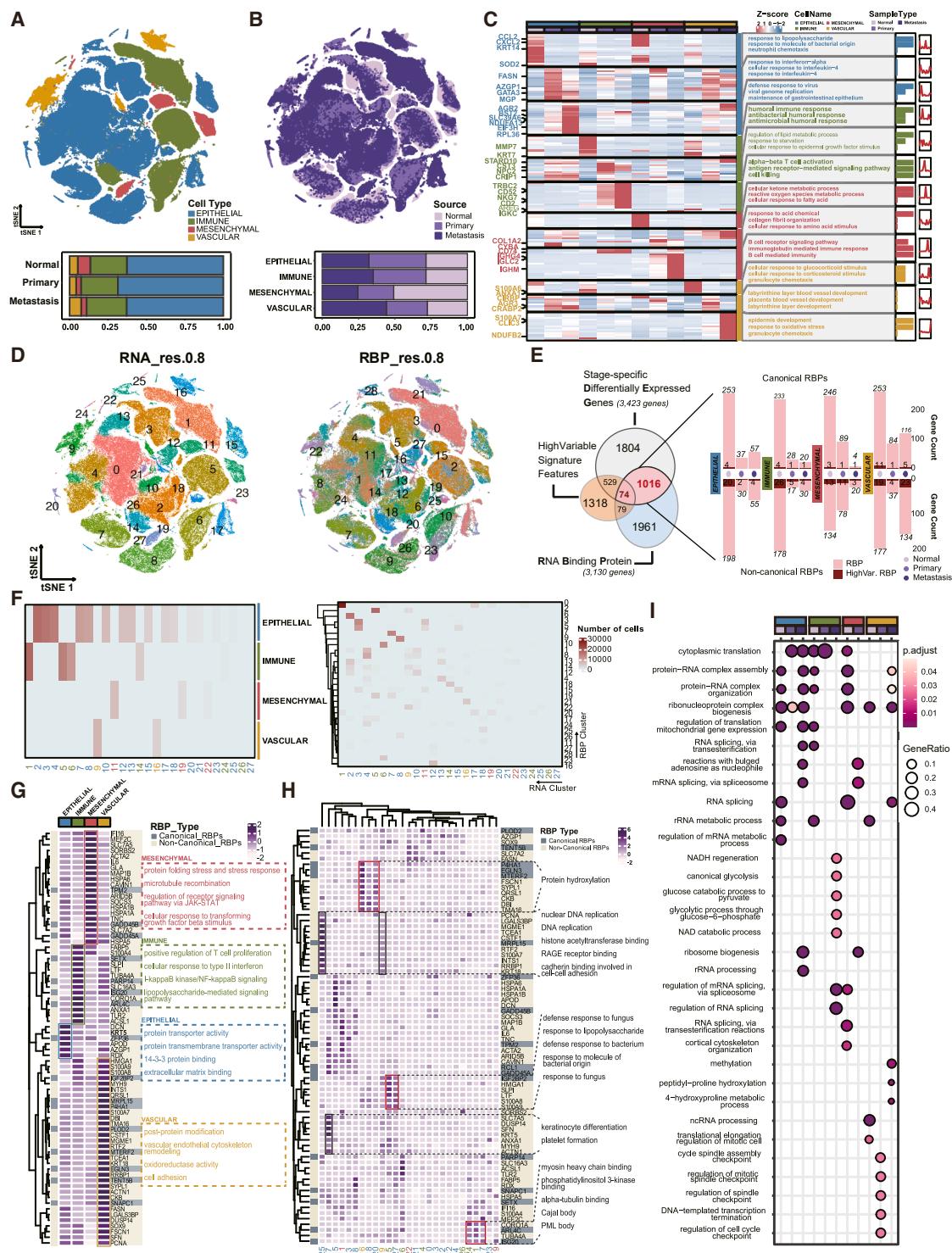


Figure 1. Single-cell atlas of ER⁺ breast cancer malignant progression

(A and B) t-SNE plots and corresponding cell proportions showing classification by cell type (A) and by sample stage (B).

(C) ClusterGV analysis across various progression stages identifies four cell types with distinct gene expression dynamics and associated biological processes.

(D) t-SNE plots showing cell clustering based on RNA expression modules (left) and RBP expression modules (right).

(legend continued on next page)

not resolve RBP dynamics across cell types, disease stages, or microenvironmental contexts.

In this study, we construct a high-resolution single-cell atlas of ER⁺ breast cancer progression, spanning normal, primary, and metastatic stages. By systematically profiling RBP expression and activity across epithelial, immune, mesenchymal, and endothelial lineages, we uncover stage-specific post-transcriptional regulatory networks that underpin malignant evolution. Notably, we identify key RBPs such as *KRT18* (keratin 18) and *HSPA6* (heat shock protein family A member 6) that orchestrate cytoskeletal remodeling and invasive adaptation. In parallel, we reveal metastasis-enriched intercellular signaling pathways, particularly the TIMP1-MMP1, TGFB1-MMP1, and TGFB1-COL7A1 axes (TIMP1, tissue inhibitor of metalloproteinases 1; MMP1, matrix metalloproteinase 1; TGFB1, transforming growth factor β 1; COL7A1, collagen type VII alpha 1), which are spatially and transcriptionally coupled to RBP-associated regulatory states. Specifically, RBP activity is linked to the establishment of pro-invasive signaling contexts, within which these ligand-receptor interactions become amplified during metastatic progression. Together, our findings illuminate a dynamic, RBP-driven regulatory architecture that links cell-intrinsic remodeling to intercellular communication during tumor progression. This work not only deepens mechanistic insight into the post-transcriptional regulatory logic of metastasis but also nominates RBPs as potential therapeutic targets for reprogramming the TME and disrupting pro-metastatic signaling networks.

RESULTS

Single-cell transcriptomic atlas of ER⁺ breast cancer

To comprehensively delineate the role of RBPs in the malignant progression of ER⁺ breast cancer, we constructed a high-resolution single-cell transcriptomic atlas encompassing 198,286 cells across 39 samples. These samples capture key pathological stages of tumor evolution, including 13 normal breast tissues, 12 primary ER⁺ breast tumors, and 14 lymph node metastases (Figures 1A and S1A). Using canonical cell type markers, we identified four major cell types including epithelial, immune, mesenchymal, and vascular endothelial cells. Systematic comparisons across the complete RBP gene set and the highly variable RBP subset (Figure S1B) consistently revealed that disease stage accounted for substantially stronger expression differences than any individual clinical variable, including age, tumor grade, ER expression level, or tumor size. These findings indicate that the predominant source of variation in RBP expression profiles arises from disease stage itself rather than from patient-specific clinical characteristics, thereby supporting the interpretation that the observed dynamic changes in RBPs primarily reflect tumor progression-associated molecular reprogramming rather than clinical covariate-driven effects. Comparative analysis across disease stages revealed a marked increase in epithe-

lial cell abundance in primary tumors (Figures 1B and S1C–S1E), indicating early epithelial expansion during tumorigenesis. Moreover, immune cells progressively increased with disease advancement, suggesting enhanced immune surveillance and cytotoxic activity in later stages.

Dynamic gene expression profiling (Figure 1C) demonstrated a stage-specific transcriptional shift in epithelial cell programs, progressing from structural maintenance and basal metabolism in normal tissues, to proliferative and adaptive responses in primary tumors, and finally to the activation of invasion and metastasis associated pathways in metastatic lesions, reflecting evolving molecular regulatory dependencies and regulatory adaptations. Immune cells displayed progressive enrichment of gene signatures associated with cytotoxicity, immunosurveillance, and metabolic reprogramming, reflecting their dynamic adaptation within the evolving tumor immune microenvironment. Mesenchymal cells transitioned from early involvement in ECM remodeling to promoting tumor migration and invasion in later stages. Endothelial cells exhibited a shift from homeostatic vascular functions to neo angiogenesis and vascular remodeling, supporting tumor expansion and dissemination. Collectively, coordinated molecular and cellular dynamics shape the TME and drive ER⁺ breast cancer progression.

Multifaceted regulatory patterns of RBPs in ER⁺ breast cancer progression

By integrating DEGs, highly variable genes, and RBPs list, we identified 1,090 RBP genes (581 canonical and 509 non-canonical) associated with ER⁺ breast cancer progression, among which 74 exhibited high intercellular variability (Figures 1D and 1E). In line with previous studies, canonical RBPs were predominantly enriched in essential biological processes, such as ribosome biogenesis, RNA splicing, and mRNA translation. In contrast, non-canonical RBPs were associated with protein folding, stability maintenance, and cell adhesion, which facilitate tumor adaptation to stress and support invasive progression (Figure S2A). Interestingly, an increased proportion of highly variable RBP genes in metastatic samples were identified as non-canonical and largely expressed in endothelial cells, implying that RBPs serve as critical mediators of tumor invasiveness within the vasculature.

Unsupervised clustering based on transcriptomic expression profiles of RBPs revealed distinct cell type-specific RBP signatures. For example, RNA cluster 1 (c1), representing a subset of immune cells, was predominantly associated with RBP cluster 0, whereas epithelial cells (c3) were characterized by RBP clusters 3 and 5 (Figure 1F). RBP cluster 21 was recurrently observed in both endothelial (c16) and epithelial (c20) compartments, highlighting the overlap and divergence of RBP-driven regulatory patterns across cell types. Pathway enrichment analysis further demonstrated that epithelial-specific RBPs were involved in protein transport and secretion (Figure 1G), supporting rapid growth

(E) Venn diagram (left) showing overlap between all RBP genes, stage-specific DEGs, and hypervariable genes; overlapping RBPs are highlighted in red. Bar chart (right) displays the number of classical (top) and non-classical (bottom) RBPs across cell types.

(F) Heatmap shows the correspondence between RNA cluster labels, cell types, and RBP cluster labels.

(G and H) Heatmaps showing significant differential expression of RBP and representative enrichment pathways in each cell type and clusters.

(I) Heatmap showing the enrichment corresponding biological pathways of stage-type specific RBPs.

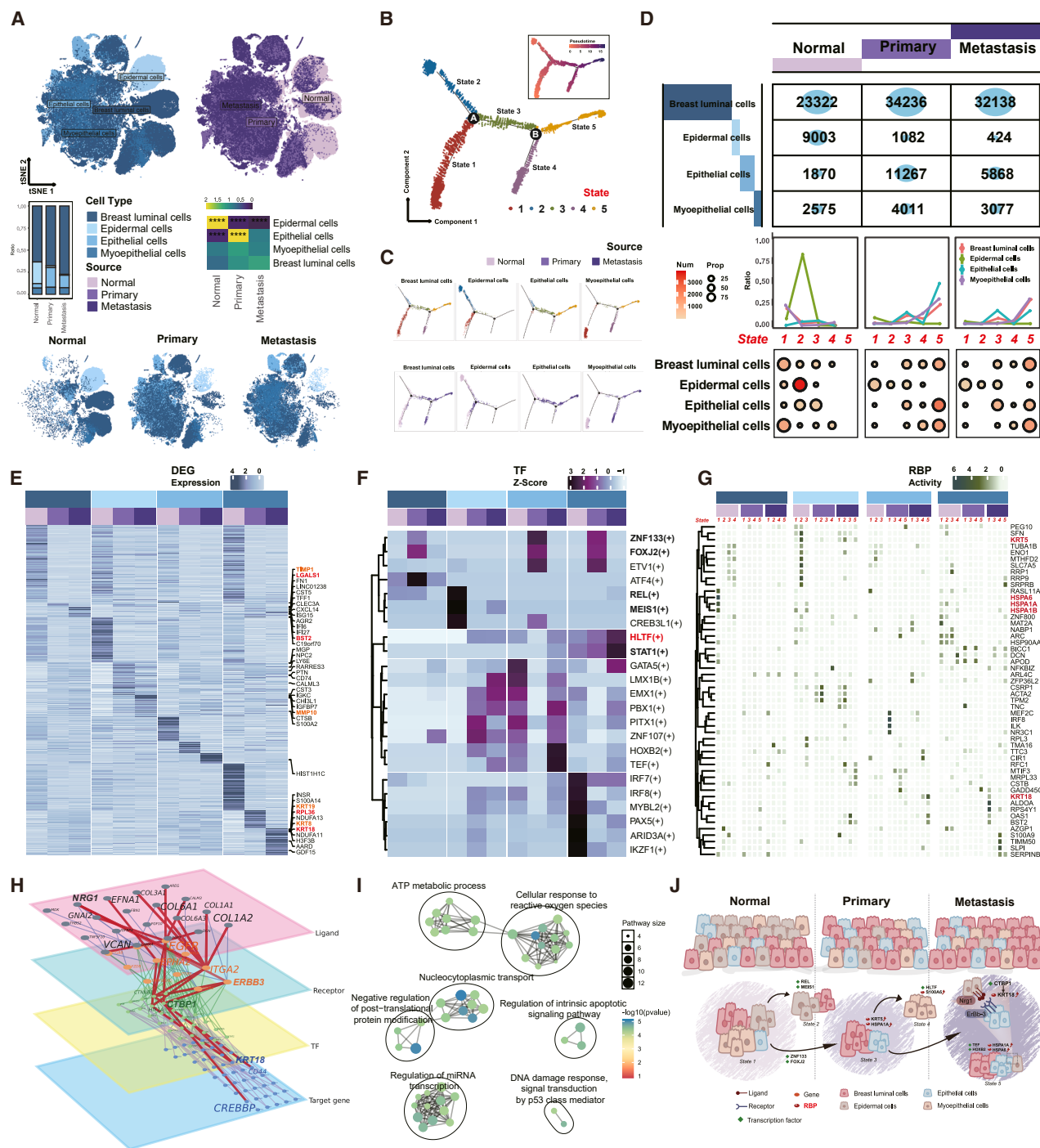


Figure 2. Transcriptional changes in epithelial cells with stage-specific RBP dynamics

(A) t-SNE plots and cell proportions of epithelial cells colored by subtype and sample stage. Heatmap shows the tissue preference of the cell subtype, Ro/e (B) and C) Pseudotime trajectory of epithelial cells in two-dimensional space. Pseudotime trajectory of epithelial subtypes colored by developmental state (top) and cell type (bottom).

(D) Bar plot, line plot, and bubble chart showing the distribution of epithelial subtypes across ER⁺ breast cancer stages.

(E–G) Heatmaps showing stage-specific expression of DEGs (E), transcription factors (F), and RBPs (G) in epithelial subtypes.

(H) Inter-cellular communication networks between epithelial cell subtypes.

(legend continued on next page)

and potential invasiveness. In immune cells, RBP signatures were enriched in T cell proliferation, type II interferon signaling, I-kappaB kinase/NF- κ B, and lipopolysaccharide-mediated pathways, highlighting their potential role in enhancing immune effector functions and modulating the tumor immune microenvironment. In mesenchymal cells, RBPs were associated with protein folding, cytoskeletal reorganization, and JAK-STAT signaling, indicating their involvement in TGF- β -mediated responses and tumor cell motility. Endothelial cell-specific RBPs were enriched in cytoskeletal remodeling, redox activity, and adhesion regulation, facilitating angiogenesis and microenvironmental reorganization (Figure 1G). Collectively, these findings underscore the cell-type-specific regulatory role of RBPs in coordinating malignant progression adaptation across ER⁺ breast cancer stages.

Of particular interest, we identified a subset of endothelial (c16) and a small number of epithelial (c18 and c20) that co-expressed RBPs involved in protein hydroxylation (Figure 1H), a process often linked to collagen biosynthesis and vascular homeostasis, which suggests a potential cooperative mechanism between endothelial and epithelial cells in regulating ECM maturation and vascular stabilization during metastasis.¹⁴ Additionally, immune (c5) and epithelial (c27) shared expression of RBPs, such as *IGF2BP2* (insulin like growth factor 2 mRNA-binding protein 2), *HMGA1* (high-mobility group AT-hook 1), *SLPI* (secretory leukocyte peptidase inhibitor), *LTF* (lactotransferrin), *S100A8* (S100 calcium-binding protein A8), and *S100A9* (S100 calcium-binding protein A9), all of which were associated with pathogenic response and immune defense pathways.¹⁵ *IGF2BP2* and *HMGA1* have been implicated in tumor cell proliferation and metastasis,^{16,17} *S100A8/A9* are known to modulate inflammation and immune responses, indicating dual roles in tumor progression and immune regulation. Similarly, immune (c1 and c24), along with epithelial (c27), co-expressed *CORO1A* (coronin 1A), *ARL4C* (ARF like GTPase 4C), *TUBA4A* (tubulin alpha 4a), and *ISG20* (interferon stimulated exonuclease gene 20), which enriched in pathways including myosin heavy chain binding, phosphatidylinositol 3-kinase binding, α -tubulin binding, and nucleolar processes. These processes are central to cytoskeletal plasticity, intracellular signaling, and post-transcriptional regulation, suggesting that multiple cell types coordinate through cytoskeletal and nuclear architecture remodeling during malignancy (Figure 1I).

In summary, we establish a high-resolution cellular atlas of ER⁺ breast cancer progression, revealing dynamic shifts in the composition and molecular programs of epithelial, immune, mesenchymal, and vascular compartments. The co-enrichment of hydroxylation-related RBPs in endothelial and epithelial cells points to shared regulation of ECM stability and vascular function. Moreover, cross-lineage RBP expression in immune and epithelial cells underscores their cooperative involvement in pathogenic defense and tumor adaptation. To refine the regula-

tory networks driven by RBPs, further investigation into the underlying molecular mechanisms is urgently needed.

Dynamic remodeling of RBPs defines epithelial transitional states

We focused on dissecting the transcriptional dynamics of RBPs within epithelial subpopulations during ER⁺ breast cancer progression. The epithelial compartment comprises distinct subtypes including luminal, epidermal, epithelial, and myoepithelial cells, each undergoing compositional and functional shifts during malignancy (Figure 2A). Breast luminal cells constituted the dominant epithelial population, increasing from 23,322 cells in normal tissue to 34,236 in primary tumors, before slightly declining to 32,138 in metastatic samples. This pattern suggests proliferative expansion during early tumorigenesis, followed by functional reprogramming at the metastatic stage. Epithelial cells displayed similar trends. In contrast, tissue preference analysis also showed epidermal cells were most abundant in normal tissue but declined sharply in both primary and metastatic samples, implying a limited role in late-stage progression. Myoepithelial cells exhibited moderate fluctuations, rising from 2,575 (normal) to 4,011 (primary), and then decreasing to 3,077 (metastasis), implying dynamic involvement in tissue remodeling and stress adaptation within the TME.

To systematically model the dynamic transitions of epithelial subpopulations, we performed pseudotime trajectory analysis and identified five distinct cellular states (state 1–5), with bifurcation points marking key developmental transitions from normal to malignant phenotypes (Figure 2B). State 1, representing the trajectory origin, was dominated by normal tissue-derived cells, particularly luminal and myoepithelial cells (Figures 2C and 2D). From bifurcation point A, two early branches emerged, state 2 and state 3, representing divergent early progression paths. State 2 retained homeostatic epithelial features and marked by transcription factors REL (REL proto-oncogene, NF- κ B subunit) and MEIS1 (Meis homeobox 1) activity associated with quiescence and tissue integrity.¹⁸

Conversely, state 3 initiating malignant transition with upregulation of proliferation associated factors FOXJ2 (forkhead box J2), ZNF133 (zinc finger protein 133), and RBPs *KRT5* (keratin 5) and *HSPA1A* (Figures 2E–2G), underscoring the importance of post-transcriptional regulation in early tumorigenesis.¹⁹ Bifurcation point (B) delineated the evolution toward more aggressive states. State 4, enriched in myoepithelial cells, was defined by activation of HLTF (helicase like transcription factor) and STAT1 (signal transducer and activator of transcription 1), transcription factors involved in ECM remodeling and TME reorganization.²⁰ In contrast, state 5 comprised highly invasive luminal and epithelial cells, exhibiting marked upregulation of pro-metastatic transcription factors TEF (thyrotroph embryonic factor), HOXB2 (homeobox B2), and non-canonical RBPs *HSPA6* and *KRT18* (Figures S2B and S2C). These molecular signatures

(I) Network plot of pathway enrichment for malignancy-associated RBPs in epithelial cells, generated using aPEAR. Node size reflects the number of genes per pathway, color indicates enrichment significance, and edges represent pathway similarity based on the Jaccard index.

(J) Schematic model summarizing the roles of key RBPs and ligand-receptor pairs in guiding epithelial subtype transitions during ER⁺ breast cancer progression. Figure created with BioRender.com.

reflect enhanced migratory capacity, adaptive stress responses, and metastatic potential, features associated with advanced disease.^{21,22} Intercellular communication analysis further revealed critical ligand-receptor interactions underpinning tumor progression (Figures 2H–2J and S3A). Specifically, in metastatic samples, epidermal cells secreted *NRG1* (neuregulin 1), which engaged *ERBB3* (Erb-b2 receptor tyrosine kinase 3) receptors on epithelial cells, leading to activation of *CTBP1* (C-terminal-binding protein 1) and transcriptional upregulation of *KRT18* in state 5 (Figure 2I). In parallel, alterations in ECM interactions, including downregulation of *CD44* and *EGFR* (epidermal growth factor receptor), were observed, likely facilitating cellular detachment and enhanced motility,^{23–25} thereby accelerating metastatic dissemination.

In brief, our results underscore the dynamic remodeling of RBPs as a defining feature of epithelial state transitions during ER⁺ breast cancer progression. Key RBPs such as *KRT18* exhibit stage-specific activity, shaping early invasive behavior, metastatic progression, and microenvironmental adaptation. Furthermore, the differential expression also showed consistent outcomes in additional validation cohorts (Figures S5 and S6A). These findings shed light on the epithelial plasticity by RBP-driven regulatory circuits and nominate promising candidates for biomarkers.

RBP-driven programs regulate stage-specific immune plasticity

For immune cells, we identified five major subtypes, proliferating T cells (*MKI67*⁺, marker of proliferation Ki-67), inflammatory T cells (*CXCL1*⁺, C-X-C motif chemokine ligand 1), neutrophils, plasma cells, and macrophages; we reconstructed a functional trajectory of immune responses during disease progression (Figure 3A). Each subset exhibited distinct transcriptional profiles and functional dynamics during disease advancement. In the normal tissue, *CXCL1*⁺ T cells represented the predominant immune population, followed by macrophages, with other immune subsets being comparatively scarce. Pseudotime analysis revealed that most immune cells clustered in state 1 (Figures 3B and 3C), characterized by low proliferative activity and minimal transcriptional change, indicative of immune homeostatic. In the primary tumor stage, neutrophils and macrophages increased substantially, reflecting activation of inflammatory and stress response pathways. These macrophages were predominantly located in state 2, reflecting early immune responses during tumor initiation. Of note, *MKI67*⁺ T cells were virtually absent in the normal stage but expanded dramatically to 245 cells in the primary tumors, highlighting their proliferative activation in the tumor immune microenvironment. As disease progressed to metastasis, further shifts in immune composition were observed. Macrophages remained dominant (14,989 cells), whereas neutrophils sharply declined (893 cells). At this stage, immune cells localized mainly to state 3 and displayed increased transcriptional heterogeneity and activation of metastasis-associated signaling (Figure 3D).

Differential expression analysis (Figure 3E) revealed stage-specific markers within immune subsets. In normal macrophages, *SAT1* (spermidine/spermine N1-acetyltransferase 1) and *AZGP1* (alpha-2-glycoprotein 1, zinc binding) were highly

expressed and associated with immune homeostasis. In the primary stage, pro-inflammatory mediators such as *S100P* (S100 calcium-binding protein P) were significantly upregulated in neutrophils, highlighting its critical role in initiating inflammatory responses. During metastasis, immune cells exhibited unique gene signatures including *IFI27* (interferon alpha inducible protein 27), *MUC1* (mucin 1), and *HBB* (hemoglobin subunit beta), implicating roles in immune evasion and tissue remodeling,^{26,27} similar upregulation trends of these genes were also observed in the independent validation cohorts (Figure S6B). Transcription factor prediction by SCENIC further illustrated dynamic regulatory shifts (Figure 3F). In the primary tumors, pro-inflammatory transcription factors (TFs) such as ATF3 (activating transcription factor 3) was strongly activated, promoting immune cell recruitment and activation. In metastasis, immunoregulatory TFs CREM (CAMP responsive element modulator) and TCF7 (transcription factor 7) became dominant, suggesting their role in immune suppression and functional plasticity. To understand the post-transcriptional landscape, we next assessed RBP activity within each subset (Figures 3G and 3H). In primary tumors, macrophages showed elevated expression of RBP genes *HSPB1* (heat shock protein family B) and *S100A4* (S100 calcium-binding protein A4) (Figure 3G), which are known to mediate stress and inflammatory responses. These expression patterns were consistently validated in the independent cohorts. In metastasis, RBPs such as *ATP1B1* (ATPase Na⁺/K⁺ transporting subunit beta 1) were upregulated, likely contributing to cytoskeletal remodeling and adaptation to the metastatic niche,²⁸ thereby supporting the plasticity of immune cells under tumor-induced pressures. Functional enrichment analysis highlighted immune-relevant biological processes including miRNA metabolism, nuclear transport, and apoptotic signaling (Figure 3I). These pathways exhibited stage-specific regulation, underscoring immune cells' dynamic response to tumor signals and microenvironmental remodeling.

A comprehensive analysis of RBPs and associated regulators (Figure 3J) showed that *MKI67*⁺ T cells in primary tumors expressed higher levels of RBPs and related genes compared to the normal and metastatic stages. These included *RPL27* (ribosomal protein L27, ribosomal and RNA-splicing factors) (Figure 3G), supporting the high proliferative potential of T cells.²⁹ *ATP1B1* and *AIFM1* (apoptosis inducing factor mitochondria associated 1, metabolic regulators), enhancing adaptation to metabolic stress. *TPM1* (tropomyosin 1) and *DPYSL2* (dihydropyrimidinase like 2, cytoskeletal modulators), reflecting enhanced motility and tumor interaction.^{30,31} Elevated expression of *CASP7* (caspase 7) and *DEDD2* (death effector domain containing 2) indicated regulation of apoptosis.³² Collectively, these factors underscore the functional flexibility of T cells in response to tumor-derived stress and immune demands. Plasma cells in metastasis showed elevated levels of *SPATS2L* (spermatogenesis associated serine rich 2 like), potentially aiding microenvironmental adaptation through post-transcriptional regulation. High expression of *ESF1* (ribosome biogenesis, ESF1 nucleolar pre-rRNA-processing protein) and *ZYX* (adhesion signaling, Zyxin) suggests involvement in translational upregulation and enhanced cell-ECM interactions.³³ Overall, ER⁺ breast cancer progression is marked by profound immune

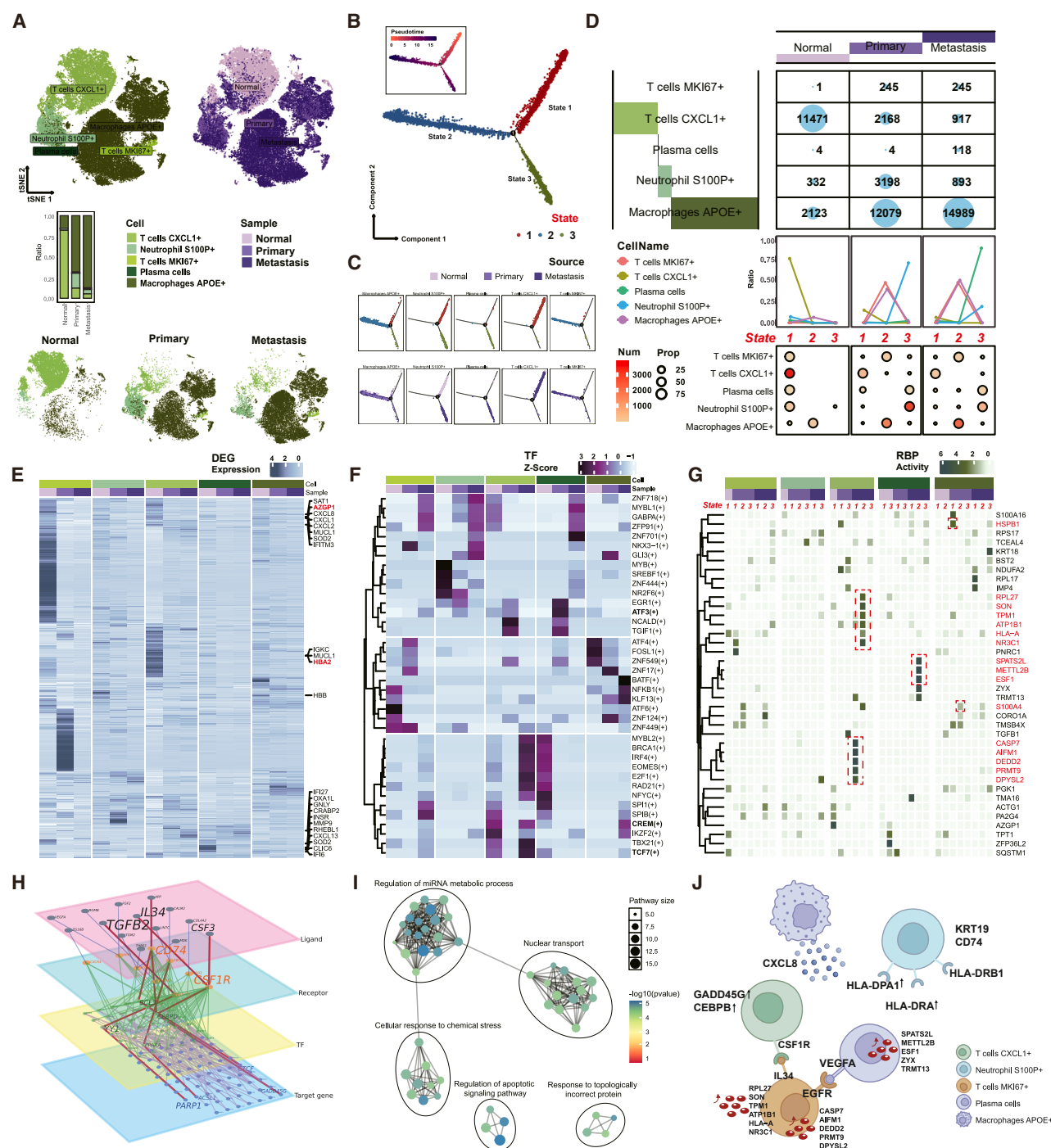


Figure 3. Immune cell composition and RBP expression patterns

(A) t-SNE plots (top and bottom) and cell proportions (middle) of immune cells colored by subtype and sample stage.
 (B and C) Pseudotime trajectory of immune cells in two-dimensional space, colored by developmental state (top) and cell type (bottom).
 (D) The distribution of immune subtypes across ER⁺ breast cancer stages.
 (E–G) Heatmaps showing stage-specific expression of DEGs (E), transcription factors (F), and RBPs (G) across immune subtypes.
 (H) Regulatory networks of ligand-receptor interactions in immune cell subtypes.
 (I) Network visualization of pathway enrichment for malignancy-associated RBPs in immune cells.
 (J) Schematic model summarizing the roles of key RBPs and ligand-receptor interactions in shaping immune subtype dynamics during ER⁺ breast cancer progression.

plasticity and RBP-driven regulatory circuits, which reveal novel targets for precision immunotherapy aimed at overcoming immune evasion and metastasis.

Post-transcriptional programs define the metastatic niche in mesenchymal cells

During ER⁺ breast cancer progression, mesenchymal populations, primarily fibroblasts and pericytes, exhibit marked changes in proportion and function. In normal tissue, fibroblasts accounted for 95.2% of mesenchymal cells and orchestrated ECM remodeling such as collagen deposition and fibronectin assembly.³⁴ In primary tumors, fibroblasts decreased to 84.6% while pericytes rose to 15.4%, suggesting enhanced vascular permeability via PDGF-BB and PDGFR β signaling in support of tumor invasion.³⁵ This trend intensified in metastases, with fibroblasts dropping to 23.4% and pericytes increasing to 76.6%. Fibroblasts also showed elevated tissue bias odds ratios, indicating a functional shift toward vascular support and immune modulation (Figures 4A and B).

A coordinated transcriptional and post-transcriptional program involving TFs and RBPs underpinned mesenchymal functional heterogeneity (Figures 4C–4E). We observed that HOXA7 and TBX15 were significantly enriched in fibroblasts at the metastatic stage (Figure 4D). Previous studies have reported that HOX family transcription factors can regulate multiple signaling pathways, including Wnt signaling.^{36,37} Consistent with this notion, the activity of the Wnt/ β -catenin signaling pathway exhibited a gradual increase during disease progression (Figure 4F). In parallel, ECM remodeling scores were significantly elevated during malignant progression (Figure 4G), accompanied by a progressive upregulation of *TIMP1* expression (Figure 4C). Notably, ECM remodeling scores showed strong positive correlations with the expression of TIMP family genes (Figure 4G). Together, these findings suggest that fibroblasts may progressively coordinate the balance between TIMPs and ECM-degrading enzymes (MMPs) during malignant progression,³⁸ thereby contributing to the formation of a pro-invasive TME. Co-activation of *GATA2* (GATA-binding protein 2) and *NR1H3* (nuclear receptor subfamily 1 group H member 3) drove lipid metabolic reprogramming, specifically fatty acid oxidation, promoting a cancer-associated fibroblast (CAF) phenotype marked by α -SMA (alpha smooth muscle actin) and FAP (fibroblast activation protein alpha). At the transcriptional level, *PKP3* (plakophilin 3) was significantly enriched in metastatic fibroblasts, probably enhancing translational output and ECM-remodeling capacity.³⁹

To further elucidate the functional implications of dynamic RBP changes, we performed pathway enrichment on RBPs upregulated in metastases. The results revealed significant enrichment in the miRNA metabolic process and regulation of cellular response to transforming growth factor β stimulus, suggesting that these RBPs are not only involved in mRNA stabilization and translation but may also contribute to metastatic adaptation by modulating non-coding RNA metabolism and the TGF- β signaling axis. Interestingly, we observed a series of stage-specific RBP expression patterns in fibroblasts paralleled cellular function shifts. In normal tissue, *ISG20* (interferon stimulated exonuclease gene 20), *PRSS23* (serine protease

23), and *BSPRY* (B-box and SPRY domain containing) were highly expressed and may play roles in RNA processing and cellular homeostasis. As the disease progressed, *SON* (SON DNA and RNA binding protein), *EDF1* (endothelial differentiation related factor 1), *IQGAP1* (IQ motif containing GTPase-activating protein 1), and *SF1* (splicing factor 1) were gradually upregulated, with *SON* and *SF1* likely participating in early post-transcriptional network remodeling as splicing-related regulators.⁴⁰

In pericytes, *SSR4* and *PRDX4* expression increased with malignancy, potentially supporting protein trafficking, redox homeostasis, and promoting vascular stability and immune evasion in advanced disease (Figure 4E). At the metastatic stage, RBPs, including *BSPRY*, *SEC61A1*, *PKP3*, *IL6* (interleukin 6), and *RNH1* (ribonuclease/angiogenesis inhibitor 1) became highly active (Figure 4E). *PKP3* likely promotes ECM remodeling by binding to the 3' UTR of ECM-related transcripts, thereby stabilizing their expression and enhancing translational output.³⁹ Notably, in addition to its classical cytokine function, *IL6* emerged as a functional RBP in metastatic fibroblasts. scMLnet (Single-cell Multi-layer Signaling Network) analysis revealed transcriptional regulation of *IL6* by CEBPD, REL, and DDIT3, indicating a coordinated TF-RBP axis driving a pro inflammatory, invasive fibroblast phenotype (Figures 4H and 4I).

Collectively, these results reveal that dynamic RBP programs, shaped by upstream transcriptional regulation, drive functional transitions in fibroblasts and pericytes during metastasis (Figure 4J). Targeting key RBPs such as *PKP3* and *PRDX4* may offer new strategies to disrupt the mesenchymal support for ER⁺ breast cancer progression.

RBP-mediated reprogramming of endothelial cells

Single-cell transcriptomic profiling of endothelial cells across normal, primary, and metastatic ER⁺ breast cancer stages revealed pronounced transcriptional heterogeneity and dynamic reprogramming. Distinct endothelial subclusters predominated at each stage, reflecting temporal shifts in vascular phenotypes (Figures 5A–5D). In normal tissue, clusters 3 and 4, corresponded to pseudotime state 1, expressed *AKAP12* (A-kinase anchoring protein 12) and *CXCL1*, with activation of transcription factors *PBX4* (PBX homeobox 4) and *ATF2* (activating transcription factor 2), suggesting involvement in vascular stability and early microenvironmental sensing.⁴¹ In primary tumors, endothelial cells mainly aggregated in cluster 0, whereas metastatic samples were enriched in clusters 1 and 2, aligned with pseudotime state 3, exhibiting more complex transcriptional signatures. These endothelial cells showed strong activation of TFs, such as *DDIT3* (DNA damage inducible transcript 3), *FOXC1* (forkhead box C1), *BCL11B* (BCL11 transcription factor B), and *ZNF* family members (Figures 5E and 5F), known to regulate stress responses, angiogenesis, and metastatic traits.⁴² Meanwhile, RBPs such as *MRPL15* (mitochondrial ribosomal protein L15) and *HSPA8* were significantly upregulated. *MRPL15* may regulate mitochondrial protein synthesis and metabolic adaptation, while *HSPA8* is involved in protein homeostasis under stress, suggesting that metastatic endothelial cells rely on metabolic and stress-regulatory programs to maintain functional plasticity (Figure 5G).

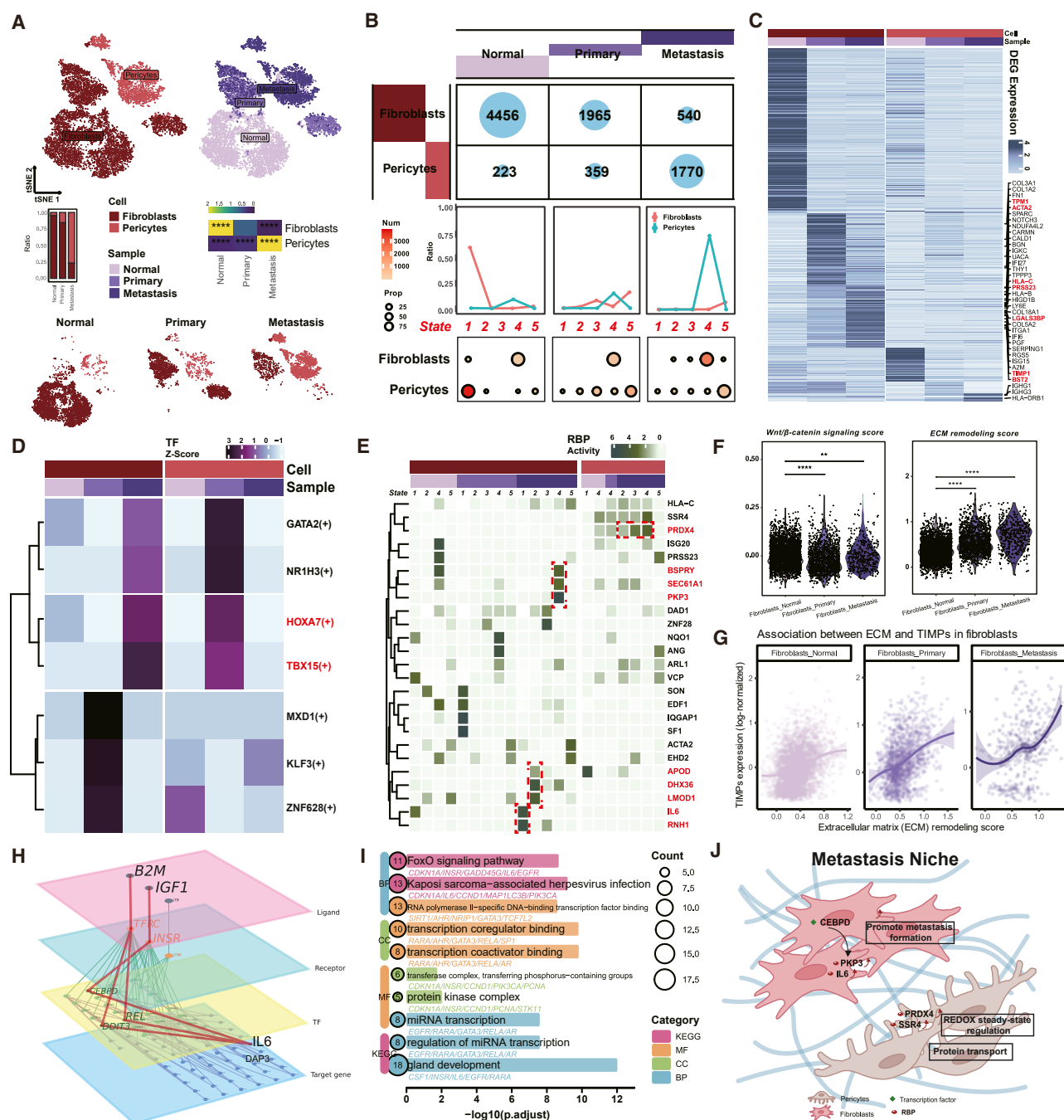


Figure 4. Mesenchymal cell differentiation and RBP modulation

(A) t-SNE plots (top and bottom) and cell proportions (middle) of mesenchymal cells colored by subtype and sample stage.

(B) The distribution of mesenchymal subtypes across ER⁺ breast cancer stages.

(C–E) Heatmaps showing stage-specific expression of DEGs (C), transcription factors (D), and RBPs (E) across mesenchymal subtypes.

(F) Heatmaps showing Wnt signaling pathway activity and ECM remodeling scores in fibroblasts across normal, primary, and metastatic states. Statistical differences between two groups were evaluated using the Wilcoxon rank-sum test, significance levels used are * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

(G) Fitted correlation curves between the expression of the TIMP gene family and ECM remodeling scores across different disease states.

(H) Regulatory network of ligand-receptor interactions between pericytes and fibroblasts.

(I) Visualization of significantly enriched GO terms (Biological Process, Cellular Component, and Molecular Function) and KEGG pathways, with bar length representing $-\log_{10}(\text{adjusted } p \text{ value})$ and circle size indicating the number of associated genes.

(J) Schematic model illustrating the roles of key RBPs and ligand-receptor pairs in guiding mesenchymal subtype transitions during ER⁺ breast cancer progression.

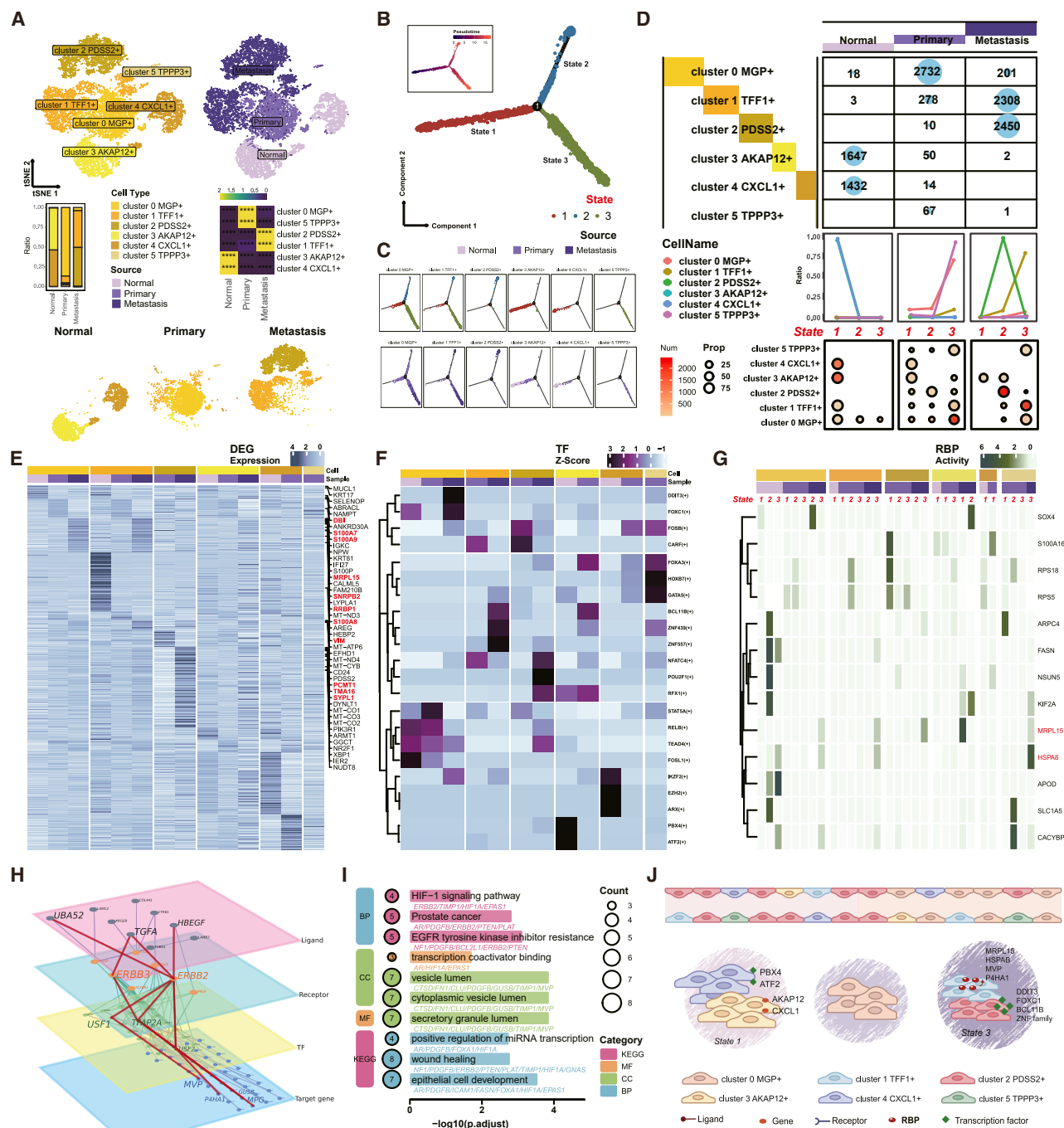


Figure 5. Dynamic RBP expression in vascular cells during tumor progression

(A) t-SNE plots (top and bottom) and cell proportions (middle) of vascular cells colored by subtype and sample stage.
 (B and C) Pseudotime trajectories of vascular cells in two-dimensional space, colored by developmental state (top) and cell type (bottom).
 (D) The distribution of vascular subtypes across ER⁺ breast cancer stages.
 (E–G) Heatmaps showing stage-specific expression of DEGs (E), transcription factors (F), and RBPs (G) across vascular subtypes.
 (H) Regulatory network of ligand-receptor interactions among vascular subpopulations.
 (I) Visualization of significantly enriched GO terms and KEGG pathways in vascular cells.
 (J) Schematic model illustrating how key RBPs and ligand-receptor pairs regulate vascular subtype transitions during ER⁺ breast cancer progression.

Functional network analysis (Figures 5H–5J) identified USF1 (upstream transcription factor 1), USF2⁴³ (upstream transcription factor 2), and TFAP2⁴⁴ (transcription factor AP-2 alpha) as key transcriptional regulators of an RBP cohort, including, *MPG* (N-methylpurine DNA glycosylase), *GUSB* (glucuronidase beta), and *MVP*⁴⁵ (major vault protein), enriched in ErbB signaling⁴⁶ (Figure 5H), suggesting their involvement in tumor-associated vascular remodeling. Functional enrichment of malignancy-associated RBPs further revealed their participation in biological processes such as regulation of response to endoplasmic reticulum stress, myeloid cell differentiation, and cellular response to interleukin-4. These findings suggest a potential role for RBPs in modulating ER stress, orchestrating immune differentiation, and mediating inflammatory responses, thereby contributing to vascular adaptation and immunoregulatory evolution within the TME. For instance, P4HA1 may facilitate ECM remodeling through the regulation of collagen synthesis,⁴⁷ indirectly influencing immune cell recruitment and activation.

In summary, endothelial cells undergo RBP-mediated reprogramming throughout ER⁺ breast cancer progression, enabling dynamic shifts in angiogenesis, metabolic adaptation, and intercellular communication. This reprogramming contributes to the restructuring of the tumor vasculature and the modulation of the immune microenvironment, highlighting the functional plasticity of endothelial cells. These findings underscore the potential of targeting endothelial RBP programs as a strategy for anti-angiogenic and immunomodulatory interventions in breast cancer and possibly other solid tumors.

RBP-signaling interactions drive intercellular crosstalk

Having delineated the lineage-specific regulatory roles of RBPs, we next sought to systematically characterize their cross-lineage coordination during malignant progression. Therefore, we further explored the dynamics of intercellular communication within the TME. In the normal stage, signaling axes were predominantly within the same cell type (Figure 6A). For instance, epidermal cells displayed interactions, such as TIMP1-MMP1 and IFITM1-CD81, while fibroblasts engaged in ECM-related signaling via *ITGA1*, which receives inputs from collagen ligands *Collagen family*, highlighting fibroblast-specific matrix regulation.⁴⁸ As malignancy progressed, these interactions intensified and novel inter-lineage signaling emerged.

We observed that compared to the normal stage, endothelial in the metastatic stage exhibited signal interactions involving the RBP gene *LGALS3* (galectin 3), which acted as a ligand interacting with receptors such as *THBS1* and *EFNA1*. *LGALS3* has been shown to regulate endothelial-to-mesenchymal transition via the PI3K/AKT signaling pathway in silica-induced pulmonary fibrosis,⁴⁹ supporting its role in modulating endothelial plasticity during tumor progression. Interestingly, in the malignant stage, a dense communication network formed among fibroblasts, endothelial, and epidermal cells (Figures 6B and S4). Fibroblasts exhibited strong outgoing signals, notably TIMP1, which activated the TIMP1-MMP1 axis in epidermal cells, thereby promoting ECM degradation and facilitating tumor cell migration. Concurrently, endothelial cells upregulated *TGFB1* in metastases, establishing a TGFB1-COL7A1 signaling axis via *COL7A1* recep-

tors on epidermal cells, which was associated with the induction of an invasive epithelial phenotype.

Through integrative analysis of cell-type-resolved transcriptomes and ligand-receptor networks, we systematically characterized the progressive remodeling of intercellular signaling during ER⁺ breast cancer metastasis. This remodeling is marked by a transition from predominantly intra-lineage communication toward coordinated cross-lineage interactions, accompanied by context-dependent involvement of RBPs. Notably, RBPs such as *LGALS3* emerge as metastasis-associated ligands in endothelial cells, engaging receptors including *THBS1* and *EFNA1*, and delineating RBP-linked communication routes across cellular compartments. The formation of RBP-associated signaling axes, including TIMP1-MMP1 and TGFB1-COL7A1, reflects a dynamic reprogramming of cell-cell interactions associated with ECM remodeling and invasive epithelial behavior (Figure 6B), similar transition patterns of these key signaling programs were observed in independent cohorts (Figures S6C and S6D).

The spatial relevance of these single-cell-inferred signaling programs was independently supported by spatial transcriptomics analyses (He et al., 2025; Wu et al., 2021; Table S1). By mapping single-cell-defined cell-type signatures onto spatial transcriptomic sections, we identified epithelial, mesenchymal, immune, and endothelial compartments within ER⁺ breast cancer tissues (Figures 6C and S7). Across independent spatial transcriptomics cohorts, key signaling axes enriched in single-cell communication analyses—including TIMP1-MMP1, TGFB1-COL7A1, and TGFB1-MMP1—displayed pronounced spatial enrichment at tumor epithelial margins and invasive fronts. These signals formed gradient-like distributions at the tumor-stroma interface, closely mirroring histological features associated with tumor invasion, indicating that these signaling interactions are spatially constrained to invasion-associated microenvironments.

We further identified key ligand-receptor pairs mediating intercellular communication in spatial sections of ER⁺ breast cancer. stLearn-based LR analysis was performed on four primary ER⁺ samples (P1A_ER, P1B_ER, CID4290, and CID4535). Collagen-integrin signaling was dominant in P1A_ER and P1B_ER (COL1A1-ITGB1 and COL1A1-ITGA5), mainly enriched at the tumor-stroma interface (Figure S8). In CID4290, the MDK-ITGB1 interaction was enriched, consistent with reported roles of mid-kine in tumor proliferation, immune modulation, and drug resistance,^{50,51} suggesting active paracrine signaling in this subtype. In CID4535, enrichment of the GNAI2-F2R (PAR-1) interaction was observed, aligning with the involvement of PAR-1 in coagulation- and inflammation-related signaling within the TME.⁵²

Functional perturbation experiments further supported the biological relevance of the invasion-associated signaling programs identified by single-cell and spatial analyses. TGF- β stimulation induced EMT-related molecular changes and enhanced invasive capacity in ER⁺ breast cancer cell lines, including ZR-75-1 and MCF7, with consistent trends observed across both models. In parallel, independent knockdown of MMP1 or COL7A1—two ECM-associated effectors enriched in metastatic states—partially reversed EMT-associated molecular programs, as reflected by increased E-cadherin and decreased vimentin expression at both the mRNA and protein levels (Figures 7A and 7B). Consistently, Transwell invasion assays showed that although TGF- β

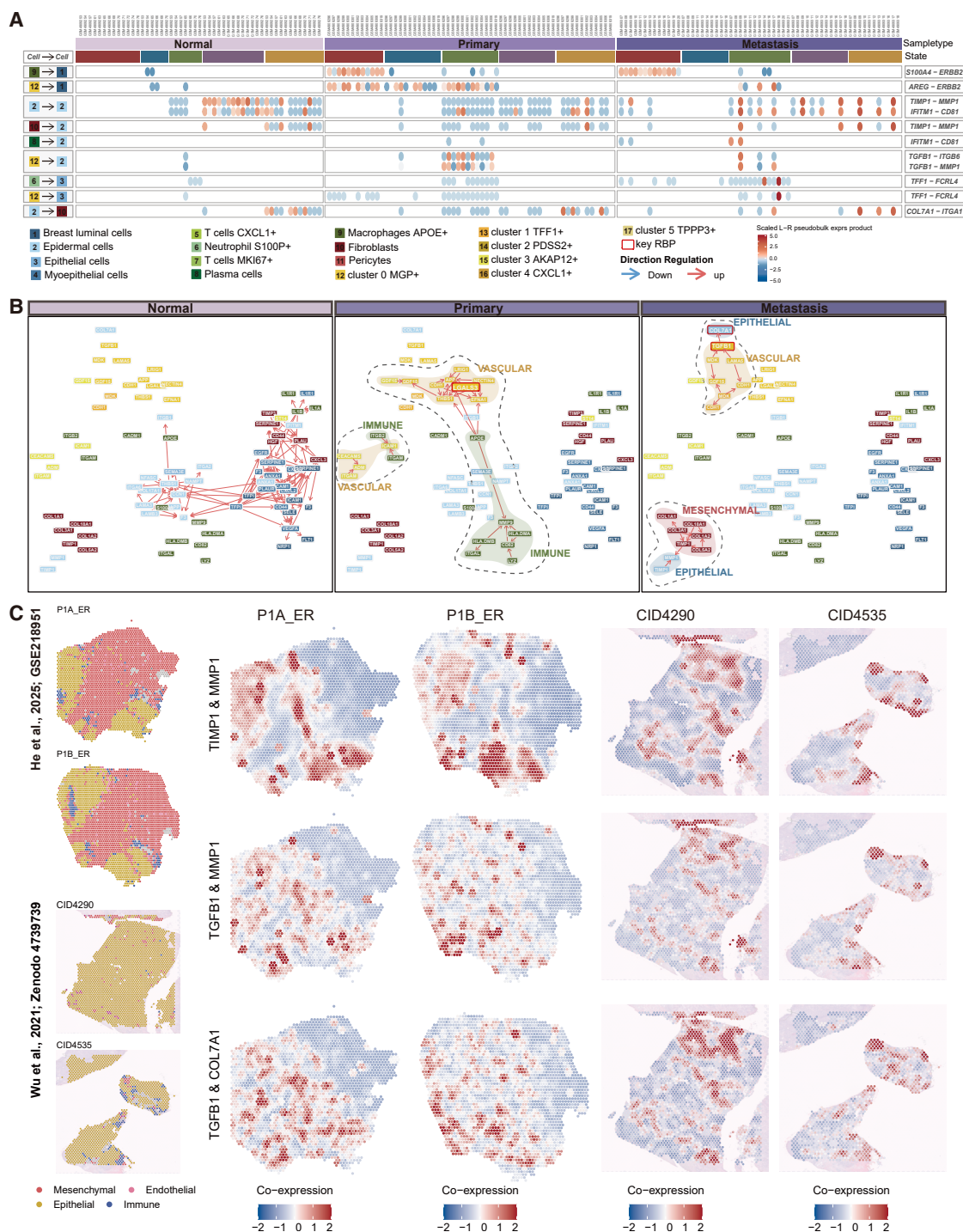


Figure 6. Inter-cellular communication network highlighting key interactions in ER⁺ breast cancer

(A) Correlation plot showing both ligand-receptor pair and predicted target gene expression were visualized with the predicted target ligand-target links in the metastatic tissue.

(B) Stage-specific diagram of cellular communication networks. Different colored nodes represent signals from different receptor ligands, and lines represent interaction relationships.

(C) Spatial cell type distribution and co-expression patterns of ER⁺ breast cancer (P1A_ER, P1B_ER, CID4290, and CID4535).

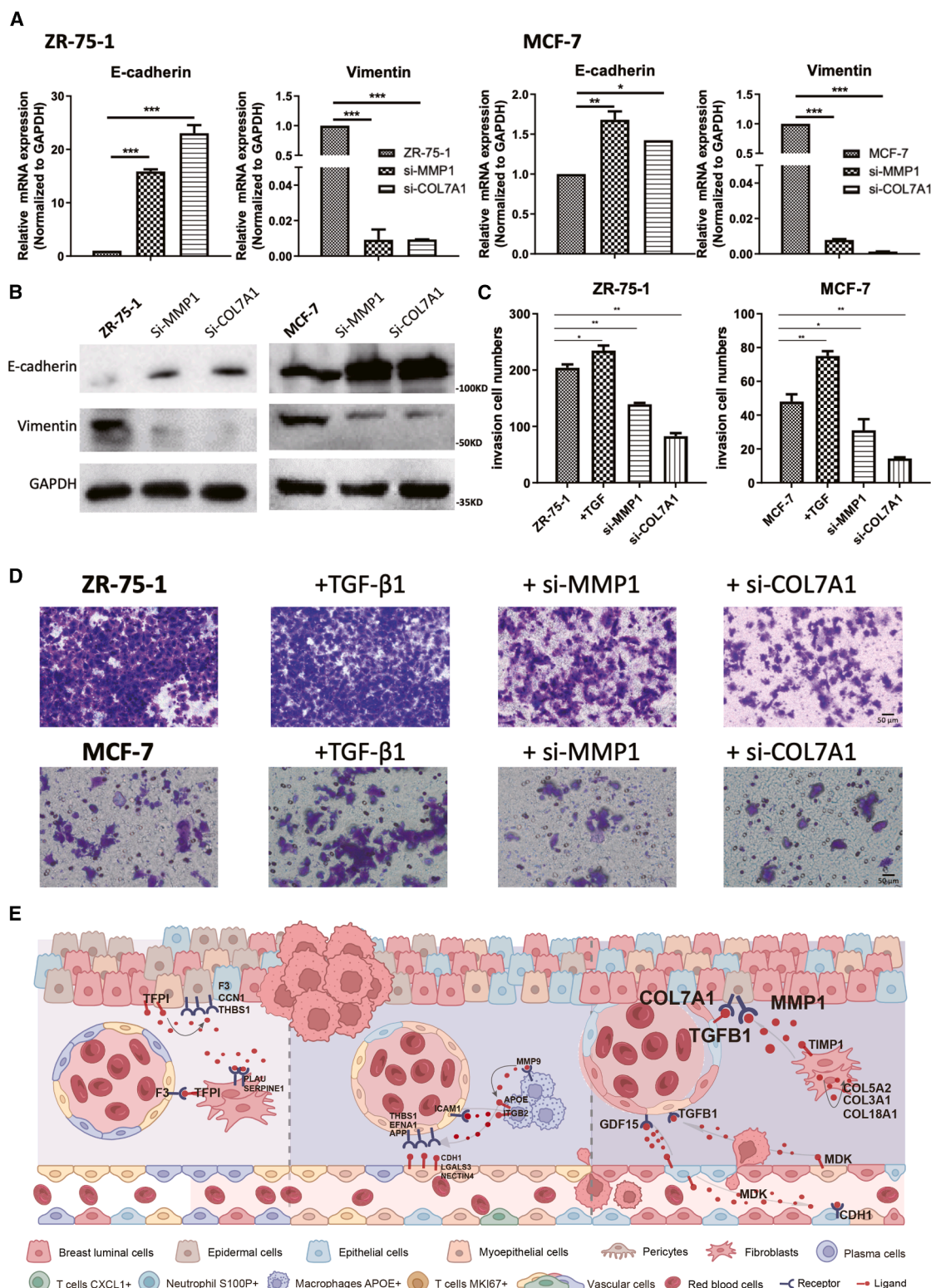


Figure 7. Functional validation of MMP1 and COL7A1 in EMT-associated phenotypes and invasive capacity of ER⁺ breast cancer cells

(A and B) Relative mRNA expression of E-cadherin and vimentin in ER⁺ breast cancer cells (ZR-75-1 and MCF7) under control, si-MMP1, and si-COL7A1 conditions, as measured by qPCR. Protein levels of E-cadherin and vimentin were assessed by western blot, with GAPDH used as a loading control.

(C) Quantification of invaded cells in Transwell invasion assays under control conditions, TGF- β treatment, si-MMP1, or si-COL7A1 conditions.

(legend continued on next page)

treatment promoted tumor cell invasion, knockdown of *MMP1* or *COL7A1* significantly attenuated invasive behavior (Figures 7C and 7D). As a downstream effector of *TGFB1* and other RBPs, *MMP1* was significantly upregulated in tumor tissues compared with normal tissues across multiple BRCA subtypes in the TCGA-BRCA cohort and was strongly associated with poor prognosis (Figures S9A–S9C). Among malignant phenotype signatures, *MMP1* expression showed the strongest correlations with invasion ($p = 0.23$, $p < 0.001$) and EMT ($p = 0.17$, $p < 0.001$), suggesting a potential role in promoting metastatic progression in breast cancer (Figures S9D and S9E).

Collectively, these results indicate that *TGFB1* signaling and *MMP1*/*COL7A1*-mediated ECM remodeling act as functionally convergent components of a metastasis-associated regulatory module (Figure 7E), jointly contributing to EMT-related and invasive phenotypes in ER⁺ breast cancer.

DISCUSSION

Our study presents a comprehensive single-cell transcriptional atlas of ER⁺ breast cancer progression, revealing cell type-specific regulatory programs that integrate RBPs activity with malignant evolution. By profiling epithelial, immune, mesenchymal, and endothelial compartments across normal, primary, and metastatic stages, we uncover dynamic post-transcriptional regulatory networks that guide lineage plasticity and tumor adaptation. A major highlight of this work is the identification of metastasis-enriched RBPs such as *KRT18*, *HSPA6*, and *LGALS3* that orchestrate cytoskeletal remodeling and RNA metabolism while amplifying ligand–receptor signaling pathways, notably *TIMP1*–*MMP1*, *TGFB1*–*COL7A1*, and *TGFB1*–*MMP1* axes. These pathways are transcriptionally activated and post-transcriptionally sustained via RBP-mediated mRNA stabilization, thereby linking intracellular RNA regulation with extracellular signaling and metastatic niche remodeling.

Previous single-cell studies have delineated the transcriptional landscape and microenvironmental remodeling of ER⁺ breast cancer. Ozmen et al.¹⁰ characterized transcriptional state transitions between primary and metastatic tumors, while Xu et al.¹¹ identified *MMP1*⁺ malignant subsets and established *MMP1* as a key effector of ECM remodeling and tumor–immune interactions. However, these studies largely focused on transcriptional programs and downstream effector functions, leaving the regulatory mechanisms underlying stage-specific activation unresolved. In contrast, our study explicitly interrogates post-transcriptional regulation by providing a systematic, stage-resolved characterization of both canonical and non-canonical RBPs across major cell types during ER⁺ breast cancer progression. We identify RBP-amplified signaling axes, including *TGFB1*–*MMP1* and *TGFB1*–*COL7A1*, and integrate trajectory inference, intercellular communication modeling, and RBP-centered regulatory analysis into a cohesive framework supported by independent scRNA-seq cohorts and spatial transcriptomics.

Importantly, beyond correlative single-cell associations, we performed targeted perturbation experiments to functionally validate key components of these signaling modules, demonstrating that interference with *MMP1* and *COL7A1* directly attenuates epithelial invasiveness and matrix-associated phenotypes. Together, these results link post-transcriptional regulation to metastatic behavior at a functional level, strengthening the causal interpretation of RBP-associated signaling programs and mitigating concerns regarding over-reliance on observational transcriptomic analyses.

Nonetheless, several limitations should be acknowledged. Functional validation relied primarily on gene knockdown, and the absence of overexpression experiments limits assessment of the sufficiency of the identified regulatory axes; moreover, the lack of *in vivo* models constrains evaluation within physiological contexts. scRNA-seq captures only static transcriptional states. To better describe the continuous regulatory changes during disease progression, we performed trajectory inference and regulatory network analyses to provide indirect evidence of dynamic processes, and this limitation was partially addressed by targeted perturbation experiments validating the functional roles of *MMP1* and *COL7A1*. In addition, to compensate for the incomplete characterization of rare cell subpopulations in large-scale analyses across disease stages, we incorporated independent cohorts and spatial transcriptomic data. This integration improves the robustness of our findings and adds spatial context, thereby refining the molecular features associated with dynamic changes during ER⁺ breast cancer progression. Future studies using time-resolved sampling and systematic perturbation-based approaches will be required to more directly define the dynamic structure of RBP-driven regulatory programs. Clinical variables were carefully controlled during cohort curation to ensure consistency in tumor subtype and receptor status, thereby defining a biologically homogeneous ER⁺ breast cancer cohort and minimizing confounding effects. Within this framework, clinical annotations and large-scale transcriptomic data were integrated to evaluate the prognostic and phenotypic relevance of key RBP-associated effectors, including *MMP1*. These analyses extend our findings beyond descriptive cellular heterogeneity and provide evidence linking RBP-driven signaling to clinically meaningful outcomes.

In summary, our study establishes RBPs as central organizers of post-transcriptional regulation and intercellular communication during ER⁺ breast cancer progression. By combining single-cell transcriptomics, functional perturbation, clinical annotation, and spatial validation, we provide a refined framework for understanding how RBP-driven signaling logic shapes metastatic evolution and for guiding the development of network-informed therapeutic strategies.

Limitations of the study

Despite the strengths of this study, several limitations should be acknowledged. First, scRNA-seq inherently captures static

(D) Representative microscopic images from the Transwell invasion assay, scale bars, 50 μ m.

(E) The evolving TME during all stages of ER⁺ breast cancer progression is depicted with key representative cell types and regulatory factors shown. **Data in (A) and (C) are presented as mean value $\pm$ SD of individual test performed in triplicate. Significant differences between 2 groups were assessed using unpaired *t* test. The significance levels used are * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

transcriptional snapshots, limiting direct inference of temporal dynamics. Although trajectory inference and regulatory network analyses, together with targeted perturbation of MMP1 and COL7A1, provide indirect support, they cannot fully resolve causal temporal relationships. Second, functional validation relied primarily on gene knockdown, the absence of overexpression experiments limits assessment of sufficiency, and the lack of *in vivo* models constrains physiological relevance. Future studies incorporating time-resolved sampling, multi-omics integration, and systematic gain-of-function and *in vivo* validation will help define the dynamic and causal architecture of RBP-driven regulatory programs and strengthen their clinical relevance.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Jing Bai (baijing@hrbmu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The scRNA-seq dataset analyzed in this study is publicly available from the NCBI Gene Expression Omnibus (GEO) under accession number GEO: GSE161529. In addition, three independent publicly available ER⁺ breast cancer single-cell transcriptomic datasets (GEO: GSE176078, GSE248288, and GSE306201), together with a normal breast tissue reference dataset (GEO: GSE174588), were incorporated as external validation cohorts to independently verify major cell-type transcriptional programs and stage-specific RBP dynamics. To validate key cell populations, signaling axes, and the spatial localization patterns of RBPs at the tissue level, spatial transcriptomics datasets GEO: GSE218951 and the spatial transcriptomics data reported by Wu et al. (Zenodo: <https://zenodo.org/records/4739739>) were further analyzed. Bulk RNA-seq expression profiles and corresponding clinical data for breast cancer were obtained from The Cancer Genome Atlas (TCGA-BRCA) through the Genomic Data Commons data portal (<https://portal.gdc.cancer.gov/>). Gene sets for EMT, Wnt signaling, and related functional programs were obtained from the CancerSEA database (<http://biocc.hrbmu.edu.cn/CancerSEA/>) and the MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb>). All downstream analysis scripts are publicly available at <https://github.com/MJayDong>.

ACKNOWLEDGMENTS

This work was supported by the ST12030-Major Projects (2021ZD0202400); National Natural Science Foundation of China (62102123, U23A20166); Hainan Provincial Natural Science Foundation, China (822RC698); Heilongjiang Chunyan Youth Science and Technology Talent team project, China (CYQN24009); Heilongjiang Touyan Innovation Team Program, China; Natural Science Foundation of Heilongjiang Province, China (LH2023C061).

AUTHOR CONTRIBUTIONS

Conceptualization, J.B., M.D., X.L., and S.Y.; data curation, M.D., X.L., and S.L.; formal analysis, M.D., S.L. and X.L.; funding acquisition, J.B.; investigation, M.D. and X.L.; methodology, M.D. and X.L.; project administration, J.B.; resources, J.B.; software, M.D. and S.L.; supervision, J.B.; validation, S.L., Z.L., C.P., and R.L.; visualization, M.D. and S.L.; writing – original draft, M.D., X.L., and S.L.; writing – review and editing, X.L. and J.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, we used AI-assisted tools to check for grammatical errors and to improve language clarity and readability. After using these tools, we carefully reviewed and edited the content as needed and take full responsibility for the final version of the manuscript.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Cell culture
- METHOD DETAILS
 - Single-cell RNA-seq data
 - Single-cell RNA-seq data quality control
 - Clinical covariate-aware within-stage analysis of RBP expression
 - Cell type identification
 - Definition and annotation of RBPs
 - Trajectory analysis
 - Enrichment analysis
 - Single-cell regulatory network inference and clustering (SCENIC) analysis
 - Construction of multilayer inter- and intra-cellular signaling networks
 - Spatial transcriptomic analysis
 - Cell type deconvolution
 - Spatial gene co-expression analysis
 - stLearn-based spatial cell-cell communication analysis
 - RNA interference
 - Real-time quantitative PCR (qPCR)
 - Western blot
 - Transwell invasion assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2026.116155>.

Received: August 25, 2025

Revised: December 23, 2025

Accepted: April 13, 2026

REFERENCES

1. DeSantis, C.E., Ma, J., Gaudet, M.M., Newman, L.A., Miller, K.D., Goding Sauer, A., Jemal, A., and Siegel, R.L. (2019). Breast cancer statistics, 2019. CA. A Cancer J. Clinicians 69, 438–451. <https://doi.org/10.3322/caac.21583>.
2. Waks, A.G., and Winer, E.P. (2019). Breast Cancer Treatment A Review. JAMA, J. Am. Med. Assoc. 321, 288–300. <https://doi.org/10.1001/jama.2018.19323>.
3. Iglesia, M.D., Jayasinghe, R.G., Chen, S., Terekhanova, N.V., Herndon, J.M., Storrs, E., Karpova, A., Zhou, D.C., Naser Al Deen, N., Shinkle, A.T., et al. (2024). Differential chromatin accessibility and transcriptional dynamics define breast cancer subtypes and their lineages. Nat. Cancer 5, 1713–1736. <https://doi.org/10.1038/s43018-024-00773-6>.
4. Kumar, T., Nee, K., Wei, R., He, S., Nguyen, Q.H., Bai, S., Blake, K., Pein, M., Gong, Y., Sei, E., et al. (2023). A spatially resolved single-cell genomic atlas of the adult human breast. Nature 620, 181–191. <https://doi.org/10.1038/s41586-023-06252-9>.

5. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583, 711–719. <https://doi.org/10.1038/s41586-020-2077-3>.
6. Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M., Valcárcel, J., and Eyraes, E. (2018). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks (vol 26, pg 732, 2016). *Genome Res.* 28, 1426. <https://doi.org/10.1101/gr.242214.118>.
7. Cornelius, V.A., Naderi-Meshkin, H., Kelaini, S., and Margariti, A. (2022). RNA-Binding Proteins: Emerging Therapeutics for Vascular Dysfunction. *Cells* 11, 2494. <https://doi.org/10.3390/cells11162494>.
8. Zhang, J., Xiang, F., Ding, Y., Hu, W., Wang, H., Zhang, X., Lei, Z., Li, T., Wang, P., and Kang, X. (2024). Identification and validation of RNA-binding protein SLC3A2 regulates melanocyte ferroptosis in vitiligo by integrated analysis of single-cell and bulk RNA-sequencing. *BMC Genom.* 25, 236. <https://doi.org/10.1186/s12864-024-10147-y>.
9. Shao, X., Lu, X., Liao, J., Chen, H., and Fan, X. (2020). New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. *Protein Cell* 11, 866–880. <https://doi.org/10.1007/s13238-020-00727-5>.
10. Ozmen, F., Ozmen, T.Y., Ors, A., Janghorban, M., Rames, M.J., Li, X., Doe, A.R., Behbod, F., Mills, G.B., and Mohammed, H. (2025). Single-cell RNA sequencing reveals different cellular states in malignant cells and the tumor microenvironment in primary and metastatic ER-positive breast cancer. *npj Breast Cancer* 11, 95. <https://doi.org/10.1038/s41523-025-00808-w>.
11. Xu, D.-M., Chen, L.-X., Xue, T., Zhuang, X.-Y., Wei, L.-C., Han, H., and Mo, M. (2025). Decoding the impact of MMP1+ malignant subsets on tumor-immune interactions: insights from single-cell and spatial transcriptomics. *Cell Death Discov.* 11, 244. <https://doi.org/10.1038/s41420-025-02503-y>.
12. Brannan, K.W., Chaim, I.A., Marina, R.J., Yee, B.A., Kofman, E.R., Lorenz, D.A., Jagannatha, P., Dong, K.D., Madrigal, A.A., Underwood, J.G., and Yeo, G.W. (2021). Robust single-cell discovery of RNA targets of RNA-binding proteins and ribosomes. *Nat. Methods* 18, 507–519. <https://doi.org/10.1038/s41592-021-01128-0>.
13. Jiang, H., Zheng, H., Zhao, X., Xiang, Y., Li, J., Wang, K., Wang, G., and Du, J. (2025). Identifying the crucial oncogenic mechanisms of DDX56 based on a machine learning-based integration model of RNA-binding proteins. *npj Precis. Onc.* 9, 240. <https://doi.org/10.1038/s41698-025-01039-9>.
14. Hodson, D.J., Screen, M., and Turner, M. (2019). RNA-binding proteins in hematopoiesis and hematological malignancy. *Blood* 133, 2365–2373. <https://doi.org/10.1182/blood-2018-10-839985>.
15. Li, M., Mao, J., and Zhu, Y. (2021). New Therapeutic Approaches Using Hydrogen Sulfide Donors in Inflammation and Immune Response. *Antioxidants Redox Signal.* 35, 341–356. <https://doi.org/10.1089/ars.2020.8249>.
16. Elcheva, I.A., Wood, T., Chiarolanza, K., Chim, B., Wong, M., Singh, V., Gowda, C.P., Lu, Q., Hafner, M., Dovat, S., et al. (2020). RNA-binding protein IGF2BP1 maintains leukemia stem cell properties by regulating HOXB4, MYB, and ALDH1A1. *Leukemia* 34, 1354–1363. <https://doi.org/10.1038/s41375-019-0656-9>.
17. Müller, S., Bley, N., Busch, B., Glaß, M., Lederer, M., Misiak, C., Fuchs, T., Wedler, A., Haase, J., Bertoldo, J.B., et al. (2020). The oncofetal RNA-binding protein IGF2BP1 is a druggable, post-transcriptional super-enhancer of E2F-driven gene expression in cancer. *Nucleic Acids Res.* 48, 8576–8590. <https://doi.org/10.1093/nar/gkaa653>.
18. Schulte, D., and Geerts, D. (2019). MEIS transcription factors in development and disease. *Development* 146, dev174706. <https://doi.org/10.1242/dev.174706>.
19. Bitaraf, A., Razmara, E., Bakhshinejad, B., Yousefi, H., Vatanmakanian, M., Garshasbi, M., Cho, W.C., and Babashah, S. (2021). The oncogenic and tumor suppressive roles of RNA-binding proteins in human cancers. *J. Cell. Physiol.* 236, 6200–6224. <https://doi.org/10.1002/jcp.30311>.
20. Addison, J.B., Voronkova, M.A., Fugett, J.H., Lin, C.C., Linville, N.C., Trinh, B., Livengood, R.H., Smolkin, M.B., Schaller, M.D., Ruppert, J.M., et al. (2021). Functional Hierarchy and Cooperation of EMT Master Transcription Factors in Breast Cancer Metastasis. *Mol. Cancer Res.* 19, 784–798. <https://doi.org/10.1158/1541-7786.MCR-20-0532>.
21. Boopathy, L., Jacob-Tomas, S., Alecki, C., and Vera, M. (2022). Mechanisms tailoring the expression of heat shock proteins to proteostasis challenges. *J. Biol. Chem.* 298, 101796. <https://doi.org/10.1016/j.jbc.2022.101796>.
22. Zhang, C., Smalley, I., Emmons, M.F., Sharma, R., Izumi, V., Messina, J., Koomen, J.M., Pasquale, E.B., Forsyth, P.A., and Smalley, K.S.M. (2021). Noncanonical EphA2 Signaling Is a Driver of Tumor-Endothelial Cell Interactions and Metastatic Dissemination in BRAF Inhibitor-Resistant Melanoma. *J. Invest. Dermatol.* 141, 840–851.e4. <https://doi.org/10.1016/j.jid.2020.08.012>.
23. Ji, Q., Zhou, L., Sui, H., Yang, L., Wu, X., Song, Q., Jia, R., Li, R., Sun, J., Wang, Z., et al. (2020). Primary tumors release ITGBL1-rich extracellular vesicles to promote distal metastatic tumor growth through fibroblast-niche formation. *Nat. Commun.* 11, 1211. <https://doi.org/10.1038/s41467-020-14869-x>.
24. Yang, G., Huang, L., Jia, H., Aikemu, B., Zhang, S., Shao, Y., Hong, H., Yesseyeva, G., Wang, C., Li, S., et al. (2021). NDRG1 enhances the sensitivity of cetuximab by modulating EGFR trafficking in colorectal cancer. *Oncogene* 40, 5993–6006. <https://doi.org/10.1038/s41388-021-01962-8>.
25. Menezes, S.V., Fouani, L., Huang, M.L.H., Geleta, B., Maleki, S., Richardson, A., Richardson, D.R., and Kovacevic, Z. (2019). The metastasis suppressor, NDRG1, attenuates oncogenic TGF- β and NF- κ B signaling to enhance membrane E-cadherin expression in pancreatic cancer cells. *Carcinogenesis* 40, 805–818. <https://doi.org/10.1093/carcin/bgy178>.
26. Laughney, A.M., Hu, J., Campbell, N.R., Bakhoun, S.F., Setty, M., Lavallée, V.P., Xie, Y., Masilionis, I., Carr, A.J., Kottapalli, S., et al. (2020). Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.* 26, 259–269. <https://doi.org/10.1038/s41591-019-0750-6>.
27. Nguyen, P.H.D., Wasser, M., Tan, C.T., Lim, C.J., Lai, H.L.H., Seow, J.J.W., DasGupta, R., Phua, C.Z.J., Ma, S., Yang, J., et al. (2022). Trajectory of immune evasion and cancer progression in hepatocellular carcinoma. *Nat. Commun.* 13, 1441. <https://doi.org/10.1038/s41467-022-29122-w>.
28. Zhang, J., Hu, S., and Li, Y. (2019). KRT18 is correlated with the malignant status and acts as an oncogene in colorectal cancer. *Biosci. Rep.* 39, BSR20190884. <https://doi.org/10.1042/BSR20190884>.
29. Park, S.Y., Seo, D., Jeon, E.H., Park, J., Jang, B.C., Kim, J., Im, S.S., Lee, J.H., Kim, S., Cho, C., and Lee, Y.H. (2023). RPL27 contributes to colorectal cancer proliferation and stemness via PLK1 signaling. *Int. J. Oncol.* 63, 93. <https://doi.org/10.3892/ijo.2023.5541>.
30. Molgora, M., Esaulova, E., Vermi, W., Hou, J., Chen, Y., Luo, J., Brioschi, S., Bugatti, M., Omodei, A.S., Ricci, B., et al. (2020). TREM2 Modulation Remodels the Tumor Myeloid Landscape Enhancing Anti-PD-1 Immunotherapy. *Cell* 182, 886–900.e17. <https://doi.org/10.1016/j.cell.2020.07.013>.
31. Abu Rmaileh, A., Solaimuthu, B., Khatib, A., Lavi, S., Tanna, M., Hayashi, A., Ben Yosef, M., Lichtenstein, M., Pillar, N., and Shaul, Y.D. (2022). DPYSL2 interacts with JAK1 to mediate breast cancer cell migration. *J. Cell Biol.* 221, e202106078. <https://doi.org/10.1083/jcb.202106078>.
32. Li, S., Qu, X., Qin, Z., Gao, J., Li, J., and Liu, J. (2023). Incfocs/miR-212-5p/CASP7 Axis-Regulated miR-212-5p Protects the Brain Against Ischemic Damage. *Mol. Neurobiol.* 60, 2767–2785. <https://doi.org/10.1007/s12035-023-03216-y>.
33. Prakash, V., Carson, B.B., Feenstra, J.M., Dass, R.A., Sekyrova, P., Hoshino, A., Petersen, J., Guo, Y., Parks, M.M., Kurylo, C.M., et al. (2019). Ribosome biogenesis during cell cycle arrest fuels EMT in development and disease. *Nat. Commun.* 10, 2110. <https://doi.org/10.1038/s41467-019-10100-8>.
34. DeLeon-Pennell, K.Y., Barker, T.H., and Lindsey, M.L. (2020). Fibroblasts: The arbiters of extracellular matrix remodeling. *Matrix Biol.* 91–92, 1–7. <https://doi.org/10.1016/j.matbio.2020.05.006>.

35. Molnár, K., Mészáros, Á., Fazakas, C., Kozma, M., Győri, F., Reisz, Z., Tiszlavicz, L., Farkas, A.E., Nyúl-Tóth, Á., Haskó, J., et al. (2020). Pericyte-secreted IGF2 promotes breast cancer brain metastasis formation. *Mol. Oncol.* **14**, 2040–2057. <https://doi.org/10.1002/1878-0261.12752>.
36. Yu, M., Zhan, J., and Zhang, H. (2020). HOX family transcription factors: Related signaling pathways and post-translational modifications in cancer. *Cell. Signal.* **66**, 109469. <https://doi.org/10.1016/j.cellsig.2019.109469>.
37. Wansleben, S., Peres, J., Hare, S., Goding, C.R., and Prince, S. (2014). T-box transcription factors in cancer biology. *Biochim. Biophys. Acta Rev. Canc* **1846**, 380–391. <https://doi.org/10.1016/j.bbcan.2014.08.004>.
38. Cabral-Pacheco, G.A., Garza-Veloz, I., Castruita-De La Rosa, C., Ramirez-Acuña, J.M., Perez-Romero, B.A., Guerrero-Rodriguez, J.F., Martinez-Avila, N., and Martinez-Fierro, M.L. (2020). The Roles of Matrix Metalloproteinases and Their Inhibitors in Human Diseases. *Int. J. Math. Stat.* **21**, 9739. <https://doi.org/10.3390/ijms21249739>.
39. Du, Y., Hou, S., Chen, Z., Li, W., Li, X., and Zhou, W. (2023). Comprehensive Analysis Identifies PKP3 Overexpression in Pancreatic Cancer Related to Unfavorable Prognosis. *Biomedicines* **11**, 2472. <https://doi.org/10.3390/biomedicines11092472>.
40. Kim, J.H., Jeong, K., Li, J., Murphy, J.M., Vukadin, L., Stone, J.K., Richard, A., Tran, J., Gillespie, G.Y., Flemington, E.K., et al. (2021). SON drives oncogenic RNA splicing in glioblastoma by regulating PTBP1/PTBP2 switching and RBFOX2 activity. *Nat. Commun.* **12**, 5551. <https://doi.org/10.1038/s41467-021-25892-x>.
41. Kwon, H.B., Choi, Y.K., Lim, J.J., Kwon, S.H., Her, S., Kim, H.J., Lim, K.J., Ahn, J.C., Kim, Y.M., Bae, M.K., et al. (2012). AKAP12 regulates vascular integrity in zebrafish. *Exp. Mol. Med.* **44**, 225–235. <https://doi.org/10.3858/emmm.2012.44.3.017>.
42. Manford, A.G., Mena, E.L., Shih, K.Y., Gee, C.L., McMinimy, R., Martínez-González, B., Sherriff, R., Lew, B., Zoltek, M., Rodríguez-Pérez, F., et al. (2021). Structural basis and regulation of the reductive stress response. *Cell* **184**, 5375–5390.e16. <https://doi.org/10.1016/j.cell.2021.09.002>.
43. Horbach, T., Götz, C., Kietzmann, T., and Dimova, E. (2015). Protein Kinases as switches for the function of upstream stimulatory factors: implications for tissue injury and cancer. *Front. Pharmacol.* **6**, 3. <https://doi.org/10.3389/fphar.2015.00003>.
44. Jin, C., Luo, Y., Liang, Z., Li, X., Kolat, D., Zhao, L., and Xiong, W. (2023). Crucial role of the transcription factors family activator protein 2 in cancer: current clue and views. *J. Transl. Med.* **21**, 371. <https://doi.org/10.1186/s12967-023-04189-1>.
45. Chen, D., Li, X., Zhao, G., Zhang, C., Gan, J., Li, T., Xu, N., Li, Q., Yi, Z., Xie, Y., et al. (2026). Transglutaminase 2 facilitates metastasis of basal-like bladder cancer by sustaining MVP-mediated MAPK/ERK1/2 signaling activation. *Cancer Lett.* **636**, 218146. <https://doi.org/10.1016/j.canlet.2025.218146>.
46. Wu, X., Zhou, Z., Xu, S., Liao, C., Chen, X., Li, B., Peng, J., Li, D., and Yang, L. (2020). Extracellular vesicle packaged LMP1-activated fibroblasts promote tumor progression via autophagy and stroma-tumor metabolism coupling. *Cancer Lett.* **478**, 93–106. <https://doi.org/10.1016/j.canlet.2020.03.004>.
47. Xu, Y., Xia, D., Huang, K., and Liang, M. (2024). Hypoxia-induced P4HA1 overexpression promotes post-ischemic angiogenesis by enhancing endothelial glycolysis through downregulating FBP1. *J. Transl. Med.* **22**, 74. <https://doi.org/10.1186/s12967-024-04872-x>.
48. Huang, H., Wright, S., Zhang, J., and Brekken, R.A. (2019). Getting a grip on adhesion: Cadherin switching and collagen signaling. *Biochim. Biophys. Acta Mol. Cell Res.* **1866**, 118472. <https://doi.org/10.1016/j.bbamcr.2019.04.002>.
49. Cheng, D., Lian, W., Jia, X., Wang, T., Sun, W., Liu, Y., and Ni, C. (2024). LGALS3 regulates endothelial-to-mesenchymal transition via PI3K/AKT signaling pathway in silica-induced pulmonary fibrosis. *Toxicology* **509**, 153962. <https://doi.org/10.1016/j.tox.2024.153962>.
50. Nair, H.B., Nair, A., Liu, Y.-G., Vijayan, D.K., Subramani, R., Lakshmanaswamy, R., Viswanadhapalli, S., Sareddy, G.R., Batra, S.K., and Vadlamudi, R.K. (2026). Midkine (MDK) as a central regulator of the tumor micro-environment: From developmental cytokine to therapeutic target. *Cancer Lett.* **641**, 218258. <https://doi.org/10.1016/j.canlet.2026.218258>.
51. Yıldırım, B., Kulak, K., and Bilir, A. (2025). Midkine (MDK) in cancer and drug resistance: from inflammation to therapy. *Discov. Onc.* **16**, 1062. <https://doi.org/10.1007/s12672-025-02941-1>.
52. Liu, X., Yu, J., Song, S., Yue, X., and Li, Q. (2017). Protease-activated receptor-1 (PAR-1): a promising molecular target for cancer. *Oncotarget* **8**, 107334–107345. <https://doi.org/10.18632/oncotarget.21015>.
53. Pal, B., Chen, Y., Vaillant, F., Capaldo, B.D., Joyce, R., Song, X., Bryant, V.L., Penington, J.S., Di Stefano, L., Tubau Ribera, N., et al. (2021). A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J.* **40**, e107333. <https://doi.org/10.15252/emboj.2020107333>.
54. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502. <https://doi.org/10.1038/nbt.3192>.
55. Hu, C., Li, T., Xu, Y., Zhang, X., Li, F., Bai, J., Chen, J., Jiang, W., Yang, K., Ou, Q., et al. (2023). CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res.* **51**, D870–D876. <https://doi.org/10.1093/nar/gkac947>.
56. Liao, J.-Y., Yang, B., Shi, C.-P., Deng, W.-X., Deng, J.-S., Cen, M.-F., Zheng, B.-Q., Zhan, Z.-L., Liang, Q.-L., Wang, J.-E., et al. (2025). RBPWorld for exploring functions and disease associations of RNA-binding proteins across species. *Nucleic Acids Res.* **53**, D220–D232. <https://doi.org/10.1093/nar/gkac1028>.
57. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386. <https://doi.org/10.1038/nbt.2859>.
58. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics-A Journal of Integrative Biology* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118>.
59. Aibar, S., González-Blas, C., Moerman, T., Van, A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083. <https://doi.org/10.1038/NMETH.4463>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
GAPDH antibody (rabbit polyclonal)	Proteintech	Cat# 5174; RRID: AB_10694233
E-cadherin antibody (rabbit polyclonal)	Proteintech	Cat# 874-1-AP; RRID: AB_11182677
Vimentin antibody (rabbit polyclonal)	Proteintech	Cat# 10366-1-AP; RRID: AB_2277027
HRP-conjugated anti-rabbit IgG	Absin	Cat# abs20040
Biological samples		
Human ER ⁺ breast cancer scRNA-seq dataset	NCBI Gene Expression Omnibus (GEO)	GSE161529;GSE176078;GSE248288;GSE306201
Human normal breast tissue scRNA-seq dataset	NCBI GEO	GSE174588
Human breast cancer spatial transcriptomics dataset	NCBI GEO	GSE218951
Human breast cancer spatial transcriptomics dataset	Zenodo (Wu et al.)	https://doi.org/10.5281/zenodo.4739739
Chemicals, peptides, and recombinant proteins		
TRIzol Reagent	Invitrogen	Cat# 15596026
Lipofectamine 2000 Transfection Reagent	Invitrogen	Cat# 11668019
PrimeScript TM RT Reagent Kit	Takara	Cat# RR037A
TB Green [®] Premix Ex Taq TM II	Takara	Cat# RR820A
Matrigel Basement Membrane Matrix	BD Biosciences	Cat# 354235
Critical commercial assays		
PrimeScript TM RT Reagent Kit	Takara	Cat# RR037A
TB Green [®] Premix Ex Taq TM II	Takara	Cat# RR820A
Deposited data		
ER ⁺ breast cancer scRNA-seq dataset	NCBI Gene Expression Omnibus (GEO)	GSE161529
ER ⁺ breast cancer scRNA-seq validation dataset	NCBI GEO	GSE176078;GSE248288;GSE306201
Normal breast tissue scRNA-seq dataset	NCBI GEO	GSE174588
Breast cancer spatial transcriptomics dataset	NCBI GEO	GSE218951
Breast cancer spatial transcriptomics dataset	Zenodo	https://doi.org/10.5281/zenodo.4739739
Analysis scripts repository	GitHub	https://github.com/MJayDong
Experimental models: Cell lines		
Human: ZR-75-1: ER-positive breast adenocarcinoma cell line	National Collection of Authenticated Cell Cultures (NCACC, Shanghai, China)	https://www.cellbank.org.cn
Human: MCF-7: ER-positive breast adenocarcinoma cell line	National Collection of Authenticated Cell Cultures (NCACC, Shanghai, China)	https://www.cellbank.org.cn
Oligonucleotides		
siRNA targeting COL7A1 (si-COL7A1-s)	China	Sequence: 5'-CGACCUCUGCU CCACGAGA[dT][dT]-3'
siRNA targeting MMP1 (si-MMP1-s)	China	Sequence: 5'-GGAGGUAUGAU GAAUAUAA[dT][dT]-3'

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
R (statistical computing environment)	R Foundation for Statistical Computing	Version 4.2
Seurat	Satija Lab (Butler et al., Nat Biotechnol 2018)	Version 5.1.0
Harmony	Korsunsky et al., Nat Methods 2019	R package
Monocle2	Trapnell Lab (Qiu et al., Nat Methods 2017)	Version 2.32.0
clusterProfiler	Yu et al., 2012	Version 4.7.1.3
ClusterGVis	CRAN R package	Version 0.1.1
SCENIC	Aibar et al., Nat Methods 2017	Python implementation
scMLnet	Wang et al., Bioinformatics 2021	Version 0.1.0
MultiNicheNet	Browaeys et al., Nat Methods 2020	Version 2.0.1
Python	Python Software Foundation	Version 3.9
Other		
RBPWorld database	RBPWorld (research.gzsys.org.cn)	http://research.gzsys.org.cn/eurpdb2/index.html
CellMarker 2.0 database	Zhang et al., Nucleic Acids Res 2019	http://bio-bigdata.hrbmu.edu.cn/CellMarker/
CancerSEA database	Yuan et al., Nucleic Acids Res 2019	http://biocc.hrbmu.edu.cn/CancerSEA/
Molecular Signatures Database (MSigDB)	Broad Institute	https://www.gsea-msigdb.org

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**Cell culture**

Human ER-positive breast cancer cell lines ZR-75-1 and MCF7 were used in this study. All cell lines were obtained from the National Collection of Authenticated Cell Cultures (NCACC, Shanghai, China) and cultured in RPMI-1640 supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin under standard conditions (37°C, 5% CO₂). Cell line identity was authenticated by NCACC. All cell lines were regularly tested for mycoplasma contamination using PCR-based methods and confirmed to be mycoplasma-free prior to use.

METHOD DETAILS**Single-cell RNA-seq data**

ER+ breast cancer scRNA-seq dataset GSE161529⁵³ was obtained from the Gene Expression Omnibus (GEO) database. All primary tumor samples correspond to invasive breast carcinoma, including invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). Focusing on ER+ disease progression, we analyzed samples from 13 normal individuals without breast cancer, 12 patients with ER+ invasive breast cancer, and 7 patients with ER+ breast cancer paired with lymph-node metastases, yielding a total of 198,286 high-quality cells (see also Table S1).

Single-cell RNA-seq data quality control

The single-cell RNA-seq data were imported into R using the read10X function from the Seurat (v 5.1.0)⁵⁴ package. Seurat was utilized for preprocessing and normalization, beginning with the creation of Seurat objects using the CreateSeuratObject function. Quality control measures were implemented to retain cells with unique molecular identifiers (UMIs) and gene counts within the range of 200 to 10,000, and mitochondrial gene expression below 20%. After filtering, each sample retained an average of 32,138,148 reads and more than 5,218 cells.

Normalization of the counts matrix for each sample was performed using the NormalizeData function in Seurat, followed by identification of the top 2,000 highly variable genes using the FindVariableFeatures function. Individual samples were integrated into a single dataset of 39 samples using the merge function. Subsequently, the integrated dataset was normalized using the ScaleData function, based on the 2,000 highly variable genes.

Principal component analysis (PCA) was conducted on the 2,000 highly variable genes using the RunPCA function. To address batch effects and enhance integration, Harmony, a PCA-based iterative method, was employed to embed the transcriptome expression profiles into a low-dimensional space. The first 10 principal components, along with a resolution parameter of 0.5, were utilized in the FindClusters function to identify cell clusters, which were subsequently visualized using the t-SNE method.

Clinical covariate-aware within-stage analysis of RBP expression

To evaluate whether RBP expression is potentially influenced by clinical variables, we constructed multidimensional clinical stratification schemes based on patient-level annotations (Table S1), including age groups (Patient_Age_group, defined in 10-year intervals), tumor grade (Grade 1/2/3), ER expression status (high/low), and tumor size categories (Tumor_Size_group: ≤ 20 mm, 20–50 mm, > 50 mm). For each RBP, expression values were first averaged at the patient level to minimize biases arising from unequal cell numbers across samples. Subsequently, within each disease stage, non-parametric Kruskal–Wallis tests were applied to assess differences in RBP expression across clinical subgroups. All P values were adjusted for multiple testing using the Benjamini–Hochberg procedure.

Cell type identification

Differentially expressed genes across clusters were identified using the FindAllMarkers function. Canonical cell markers for major cell types were annotated using the CellMarker 2.0⁵⁵ database. Iterative clustering and annotation were performed for each major cell type to identify sub-clusters and assign cell subtypes within each major category. Epithelial cells were defined by *EPCAM* and *AZGP1*, and further subdivided into breast luminal cells (*PIP*, *TFF1*), epidermal cells (*KRT14*, *KRT5*), epithelial cells (*KRT8*, *KRT18*), and myoepithelial cells (*DCN*, *APOD*). Immune cells were initially identified as *PTPRC* and subsequently stratified based on subpopulation-specific expression patterns, including proliferative/activated T cells (*CD3E*, *CXCL1*, *MKI67*), neutrophils (*S100P*), plasma cells (*IGHG1*, *XBP1*), and *APOE*⁺ macrophages. Mesenchymal cells and pericytes were distinguished by the expression of *RGS5* and *PDGFRB*, respectively. Endothelial cells were defined using canonical markers such as *EDN1*, with endothelial subpopulations further delineated according to significantly enriched marker genes (see also Table S2).

Definition and annotation of RBPs

All RBP genes analyzed in this study were curated from the RBPWorld⁵⁶ database (<http://research.gzsys.org.cn/eurpdb2/index.html>). RBPWorld currently catalogs 3,303 Homo sapiens RBPs and classifies them into canonical and non-canonical categories. Throughout all analyses, RBP classification and gene inclusion were strictly based on RBPWorld annotations without additional manual filtering or exclusion, thereby ensuring full traceability and reproducibility of the results (see also Table S3).

Trajectory analysis

We utilized Monocle2 (v 2.32.0)⁵⁷ package to investigate the single-cell trajectories and cell state transitions of different cell subtypes. The matrix of UMI counts was taken as input to Monocle and created an object with “newCellDataSet” function. To construct the trajectory, obtain the high-variable genes for each Monocle object using the “dispersionTable” function, filter with the mean_expression ≥ 0.1 & dispersion_empirical $\geq 1 \times$ dispersion_fit as the reference. All high variable genes of cell subtypes were selected to order cells using the “orderCells” function. Dimensionality reduction was performed using the “DDRTree” function, and the developmental trajectories were visualized using the “plot_cell_trajectory” under various situations.

Enrichment analysis

To explore the potential functions and mechanisms of cell subtypes, enrichment analysis was performed with KEGG pathway and GO term set using the clusterProfiler (v 4.7.1.3)⁵⁸ package in R software (v 4.2). The “compareCluster” function was employed to compare the enrichment results across different clusters. ClusterGVis (v 0.1.1) package was used to visualize the top genes and GO term set among cell subtypes.

Single-cell regulatory network inference and clustering (SCENIC) analysis

SCENIC analysis was performed on quality-controlled cells of each major cell type using Python (v3.9). The counts matrix files for each cell type were extracted and exported, then converted into a loom file format for SCENIC⁵⁹ analysis. “GENIE3” was used to detect the relationship between transcription factors and candidate target genes. The regulatory network between TFs and genes was constructed with a search region limited to a 10k distance from the transcription start sites (TSS) or 500bp upstream of the TSS; “RcisTarget” was employed to validate the regulatory network by performing motif-TF enrichment analysis, selecting upstream regulators, and identifying highly enriched motif modules while excluding indirect targets lacking motif support. The AUCell algorithm was used to score the regulon activity for each cell, producing a binary regulon activity matrix that determines which cell active regulons are.

Construction of multilayer inter- and intra-cellular signaling networks

scMLnet (v 0.1.0) package was used to construct multilayer inter- and intra-cellular signaling networks. Subsampled and annotated scRNA-seq data were formatted into input files containing cell barcodes and clusters. Using curated ligand-receptor, TF, and TF-target interaction libraries, we applied the RunMLnet function to infer signaling networks for all cell type pairs, with thresholds of pval = 0.05 and logFC = 0.15. The DrawMLnet function visualized these networks, highlighting stage-specific intercellular communication and regulatory dynamics. Python (v 3.13) was configured for graphical output, and all interactions were validated against curated databases.

MultiNicheNet (v 2.0.1) package was used to explore ligand, receptor and ligand information in intercellular and intracellular signaling pathways to determine interactions between cell clusters. Each condition was compared with the other two to determine the specific intercellular communication in the condition of interest. To obtain more reliable results and avoid cluster loss due to small numbers, we filtered genes using thresholds of `min_cells = 1`, `fraction_cutoff = 0.01`, and `min_sample_prop = 0.1`, retaining genes expressed in at least one cell, with a minimum fraction of 1%, and present in at least 10% of samples. Ligand activity was calculated, and ligand-target gene interactions were predicted using thresholds of `logFC_threshold = 0.50` and `p_val_threshold = 0.05`.

Spatial transcriptomic analysis

Spatial transcriptomic data were obtained from public databases and processed using the Seurat R package (v 5.1.0). Filtered feature barcode matrices were imported using ReadMtx, spatial metadata and histology images were added with Read10X_Image, and only tissue-covered spots were retained for downstream analyses.

Cell type deconvolution

To deconvolve the cell type composition in the spatial region, we used R `spacexr` encapsulation v2.0.1 RCTD. The integrated single-cell RNA-seq dataset was used as the reference for cell type annotation. For each ST slide, the gene expression matrix together with spatial coordinate information was extracted and used to generate the query object via the `SpatialRNA` function. Cell type proportion estimates for individual spots were subsequently inferred using the `create.RCTD` and `run.RCTD` functions.

Spatial gene co-expression analysis

Spatial gene co-expression was evaluated using normalized expression values from the Spatial assay in Seurat. For each predefined gene pair, Spearman correlation coefficients were calculated across all spatial spots and within annotated cell types. To visualize co-expression patterns, expression values were z-score standardized and combined into a composite co-expression score for each spot, which was displayed on tissue sections using spatial feature plots. Spatial colocalization was quantified by identifying high-expression spots based on upper quantile thresholds for each gene, followed by calculation of colocalization ratios and odds ratios to assess the extent of spatial overlap.

stLearn-based spatial cell-cell communication analysis

After quality control and normalization, spatial transcriptomic data were analyzed in Python (v 3.9) using stLearn for cell-cell interaction inference. Gene expression matrices were integrated with spatial coordinates to construct an AnnData object incorporating spatial adjacency. Ligand-receptor (LR) interactions were inferred using `st.tl.cci.run`, with `min_spots=20` to filter low-coverage LR pairs and `distance=None` to consider each spot and its immediate neighbors. Statistical significance was estimated by random permutation of LR pairs (`n_pairs=100`). P values were adjusted using the Benjamini-Hochberg method ($FDR < 0.05$). Significant LR pairs were used to construct intercellular communication networks and quantify interaction strength.

RNA interference

Cells were seeded and cultured until reaching approximately 60% confluence. The siRNA was transfected with Lipo2000 Transfection Reagent (Invitrogen, USA). The subsequent experiments were performed 48h later. The si-COL7A1-s sequence was CGACCU-CUGCUCCACGAGA/dT//dT/ and the si-MMP1-s sequence was GGAGGUAUGAUGAAUAUAA/dT//dT/.

Real-time quantitative PCR (qPCR)

The Trizol reagent (Invitrogen, USA) was utilized to extract total RNA from cells followed by cDNA synthesis using the PrimeScript™ RT reagent Kit purchased from Takara and using 1000 ng of total RNA. The experimental procedure involved the use of qPCR reactions utilizing the StepOne Plus Real-time equipment manufactured by Applied Biosystems. The TB Green Premix Ex Taq II (Takara, Japan) was used in experiments. The results were shown as linearized Ct values, and the normalization of target gene mRNA expression levels in relation to GAPDH protein was accomplished using the $2^{-\Delta\Delta Ct}$ approach. All experiments were conducted in triplicate.

Western blot

Protein samples were mixed with 4× SDS loading buffer. Samples were boiled for 10 min at 100°C, and proteins were separated by Biofuraw™ Precast Gel (Tanon) and transferred to PVDF membranes manufactured by Millipore (USA), and then, the following primary antibodies and dilutions were used: GAPDH (Proteintech #5174, 1:1000), E-cadherin (Proteintech #874-1-AP, 1:1000), Vimentin (Proteintech #10366-1-AP, 1:1000). Secondary antibodies were anti-rabbit HRP conjugated IgG ((1:5000; abs20040) purchased from Absin. Subsequently, the immunoreactive bands were observed via chemiluminescence.

Transwell invasion assay

The invasion experiment was conducted using Matrigel (BD, 354235) in Transwell chambers (Corning, 8.0 μm). Cells (ZR-75-1 and MCF7) were cultured into the top chamber of the Transwell where serum-free media (100 μL) was added. Simultaneously, the bottom chamber was filled with 1 ml of 1640 supplemented with 20 % FBS. Cells that invaded through the Matrigel-coated membrane after

24 h were fixed with paraformaldehyde and stained with crystal violet (0.1 %) for 15 min. Afterward, the upper chamber was rinsed with PBS. Four sections were chosen at random inside each chamber and then examined under a microscope and photographed using an Olympus IX-53 microscope.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using R software (v 4.2) and Python (v 3.9), unless otherwise specified. Detailed information regarding statistical tests, exact sample sizes (n), definitions of n, and summary statistics is provided in the corresponding figure legends, [results](#) section, and [STAR Methods](#) subsections.

Single-cell RNA-seq analyses were performed using Seurat (v 5.1). In single-cell analyses, n represents the number of cells unless otherwise stated. Differential gene expression between cell populations was assessed using the Wilcoxon rank-sum test implemented in Seurat, with p values adjusted for multiple testing using the Benjamini–Hochberg method. Gene set enrichment analysis (GSEA) was performed using clusterProfiler, and significance was assessed based on normalized enrichment scores and adjusted p values.

The GraphPad Prism software for Windows was used for experiment statistical analyses. The results are expressed as mean values $\pm$ SD of individual tests performed in triplicate. Significant differences between 2 groups were assessed using unpaired t-test. The significance levels used are * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.