

# Spatial Single-Cell Transcriptomic Analysis in Breast Cancer Reveals Potential Biomarkers for PD-1 Blockade Therapy

**Nan Wang**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Yan Song**

National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College

**Weifeng Hong**

Fudan University

**Hongnan Mo**

National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College

**Zhentao Song**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Wenshuang Dai**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Lianshui Wang**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Haiyang Zhang**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Yuyan Zhang**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Qi Zhang**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Hui Zhang**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Tao Zhang**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Yuyi Wang**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Yeyu Li**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Jiafei Ma**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Changchao Shao**

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

**Min Yu**

Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences, Southern Medical University)

**Haili Qian**

National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College

**Fei Ma**

National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College

**Zhiyong Ding**

[zhiyong.ding@fynnbio.com](mailto:zhiyong.ding@fynnbio.com)

Mills Institute for Personalized Cancer Care, Fynn Biotechnologies

---

**Research Article**

**Keywords:** Spatial single cell transcriptomics, Luminal type breast cancer, Triple-negative breast cancer, Cellular cross-talk, Ligand-receptor interaction, Checkpoint inhibitors, Gene expression inference, Cellular neighborhood

**Posted Date:** May 29th, 2024

**DOI:** <https://doi.org/10.21203/rs.3.rs-4376986/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Abstract

## Introduction:

Spatially defined cellular interaction and crosstalk are eminently important in deciphering key molecular messages driving oncogenesis and disease progression. To date, methods enabling high-plex true single-cell resolution profiling under spatial settings are gradually becoming available and those majorly include the expansion of spatial transcriptomics (ST) being utilized.

## Results:

Through in-depth spatial single-cell profiling on four breast cancer (BC) tissue samples bearing distinct biological characteristics, we evaluated the analytical performance benchmarked against conventional pathology and by selecting pre-defined region-of-interests (ROIs), we consolidated the technical robustness of this method in defining different molecular subtypes at the transcript level matching with canonical immunohistochemistry. Moreover, we demonstrated that high-dimensional ST data is capable of identifying a major cellular network inter-wired via macrophage and cytotoxic T cells interaction in tumor adjacent cellular neighborhood via PD-L1/CD80 and CD86/CTLA4 axis, a phenomenon reflecting an improved PD-1 mediated drug response observed clinically. By incorporating open-source computational methods (Tangram and SpaGE), we found compatible inference tools for in-situ expression imputation, an approach generalizable to enable deeper spatial profiling using Xenium in-situ or other parallel approaches.

## Discussion:

Our spatial single-cell ST sets as a technical and analytical prototype for those using similar approaches for high-dimensional in-situ profiling work.

## Materials:

We applied a newly developed spatial single-cell technology (Xenium in-situ) to interrogate the spatial single-cell architecture of the complex tumor microenvironment on a set of breast cancer patient tissues (luminal-type, HER2 2+/HR- and triple negative breast cancer, TNBC) and benchmarked against multiple clinicopathological features using bioinformatic tools.

# Introduction

Understanding the spatial organization and composition of tissues or organs is indispensable to unraveling elusive questions that remain unanswered by bulk or even single-cell profiling approaches (1). In recent years, spatial multi-omics technologies have made significant strides in addressing complex biological questions, especially within disease contexts (2–11). At the mRNA level, spatial transcriptomics (ST) has emerged as a powerful discovery tool that measures mRNA expression in a spatial-resolved manner, mainly by in-situ mRNA capturing, in-situ sequencing, cyclic probe hybridization,

or other approaches (9, 10, 12, 13). At the protein level, spatial proteomics (SP), with a limited number of detectable protein targets, has been applied to validate biomarkers at the single cell or cellular neighborhood levels, usually employing multiplex Immunohistochemistry or cyclic Immunofluorescence with labeled antibodies, or mass spectrometry-based methodologies (3, 11, 14, 15).

First-generation ST technologies face a trade-off between the target multiplexity and spatial resolution. The technologies generally acquire expression profiles from small cell-clusters using docking array spots or pre-defined region of interest (ROIs) (16–18), which cover almost the entire transcriptome but are not able to achieve single-cell resolution. Bioinformatics methods have been developed to bridge this gap by deconvoluting cellular phenotypes from small-bulk expression data and by incorporating single-cell RNA-seq (scRNA-seq) data, which led to the imputation of cellular composition at near single-cell resolution (19, 20). Cell-cell communication (CCC), often inferred from protein-protein interactions and ligand-receptor co-expression obtained via RNA-seq of small bulk samples, has also been used to map mRNA expression within specifically localized cells (21, 22). However, even with sophisticated bioinformatics techniques, accurately predicting complex cross-talk behaviors, such as autocrine, paracrine, and juxtacrine interactions, remains a significant challenge (21, 22). Therefore, spatial transcriptomics at true single cell resolution is urgently needed for more accurate dissection of the tumor microenvironment and many other diseases.

The recent advance of the second generation of ST can obtain spatial transcriptomic data at a single-cell resolution, which opens a new avenue especially in the cancer microenvironment and immunotherapy research field. Most of these technologies are imaging-based, using in-situ fluorescence cycling, decoding, and quantification in series, including MERSCOPE, Spatial Molecular Imager (SMI), and Xenium (23–25). Others use high-density index oligo matrix to capture mRNA via poly A tails at single cell or sub-cellular resolution such as Stereo-seq (26). Once emerged, these technologies have been quickly applied to address key questions in development biology, organogenesis, neuroscience, oncology and other diseases (25, 27–29).

Breast cancer (BC) has become a focal point for studies employing spatial transcriptomic and proteomic technologies (14) (30) (31, 32) (33). Spatial transcriptomics, in particular, has been instrumental in examining the breast cancer TME. This includes spatially-resolved profiling on mammalian breast tissues and the study of multicellular organization in HER2-positive breast cancer (34, 35). Preliminary studies have also been carried out on breast cancer to explore both intratumor and intertumor heterogeneity (36). However, despite these advances, the current transcriptomic data are still not comprehensive enough to elucidate the intricate interactions and crosstalk occurring at the single-cell level across various breast cancer types.

In this study, we characterized the complex TME across a diverse set of breast cancer tissues: two luminal types, one HER2 2+/HR-, and one triple-negative breast cancer (TNBC), each with unique clinical and pathological features, using the Xenium spatial single-cell transcriptomics technology (25). Our proprietary bioinformatics pipelines for the spatial single-cell transcriptomic data, have proven to be

robust and enable the accurate mapping of cellular phenotypes at the single-cell level. The results indicate that the samples with distinct clinicopathological features demonstrated varied immune profiles in the tumor-enriched regions, which were organized via different spatial cellular crosstalk or, to a higher order, their representative cellular neighborhood composition. Notably, the true single-cell spatial resolution enabled precise mapping of cell communication mediated through ligand-receptor interactions in the BC TME, suggesting potential avenues for the development of predictive or prognostic biomarkers.

## Materials and Methods

### Tissue Selection Criteria and Ethical Approval

To capture a range of tissue microenvironment-associated characteristics, four patients with diverse clinical and pathological profiles were chosen. The selection included two cases of luminal type breast cancer (LBC), one HER2 2+/HR- breast cancer, and one case of triple-negative breast cancer (TNBC), classifications based on immunohistochemistry (IHC). The samples were collected post-treatment, subsequent to neoadjuvant chemotherapy or PD-1 immunotherapy. Among these patients, two presented with lymph node metastatic carcinoma characterized by extensive tumor infiltration and lymphatic invasion - one TNBC and one HER2 2+/HR-. The remaining two samples were collected from a recurrent LBC that had metastasized to the chest wall and from an LBC with lung metastasis, respectively. All specimens were formalin-fixed and paraffin-embedded (FFPE) and archived before the experiments. The utilization of all clinical materials received approval from the Ethical Review Board of the Cancer Hospital Chinese Academy of Medical Sciences (National Cancer Center), under clinical registration number 22/074-3275.

### Clinical Charting and Histopathological Characterization of Breast Cancer Tissues

The clinical details for the samples included in this study are presented in Supp. Table 1. We assessed the universal histological markers typically evaluated in breast cancer tissues using immunohistochemistry (IHC), including the Estrogen receptor (ER), HER2, progesterone receptor (PR), and Ki-67. In addition, individual samples underwent further IHC analysis to evaluate androgen receptor (AR), GATA3, cytokeratin 5/6 (CK5/6), E-cadherin (E-cad), EGFR, P53, NapsinA, TTF-1, and PD-L1. The results of these histological assessments are shown in Supp. Table 1.

### Pathological Evaluation Using Immunohistochemistry (IHC) Staining

Immunohistochemistry (IHC) was conducted on a Ventana Benchmark Autostainer (Roche Diagnostics) or a Bond automation system (Leica, Bond Max) using protocols established under clinical pathological standards. Tissue sections, 5µm thick, were initially heated to 75°C, deparaffinized in EZ Prep (Roche), and then dehydrated in a ethanol series (80%, 90%, 95%, 95%, 100%) (Beijing Yili Fine Chemicals). Antigen retrieval was performed using Cell Conditioning 1 (CC1, Roche) before the application of primary antibodies. The specific primary antibodies, detailed with clone numbers, were incubated for 16–28

minutes at optimized concentrations (ER: SP1, PR: 1E2, HER2: NEU from Roche, and Ki-67: MIB-1 from Gene Technologies). Additional primary antibodies are listed in Supp. Table 2. A horseradish peroxidase (HRP)-conjugated secondary antibody (Roche) was applied subsequently, followed by colorimetric amplification with DAB chromogen and hematoxylin counterstaining. Experienced pathologists conducted all histological assessments, and representative images were captured at 100x and 200x magnification using a KFBio imaging station.

### Spatial Profiling Using Xenium In-Situ Technology and Subsequent Hematoxylin and Eosin (H&E) Staining

FFPE tissues of 5µm thickness were sectioned and arranged on Xenium slides (10X Genomics, PN-1000460) within specified sample areas (10.45mm x 22.45mm). The sections were deparaffinized in xylene, rehydrated through a graded ethanol series, and then treated with permeabilization enzyme to allow efficient RNA exposure (Perm Enzyme B, 10X Genomics 3000553). This was followed by overnight hybridization with Xenium DNA probes (Xenium Human Breast Gene Expression Probes 2000826, detailed in Supp. Table 3) at 50°C using a thermocycler (Bio-Rad C1000). After hybridization, the padlock DNA probes bound to target mRNA were ligated at 37°C (Xenium Ligation Enzymes A/B: 2000397/2000398) for 2 hours. Circularized DNA probes for each gene were PCR-amplified in situ by rolling circle amplification (RCA), generating multiple copies of target-specific DNA barcodes. Gene decoding was performed on the Xenium instrument (10X Genomics). Essentially, the workflow takes a cyclic approach and, in each cycle, fluorescence-labeled probes with complementary nucleotide sequences for individual targets react with RCA products, and signals are x, y, z registered. DAPI nucleus staining is used to allocate individual cells and for cell segmentation purposes using a built-in machine learning algorithm. As the fluorescence-guided decoding process takes place, recorded four-colored signal combinations are generated, counted and used to construct a gene-to-count matrix used for downstream processing.

Samples were immediately proceeded with H&E staining after the Xenium decoding process. In general, samples were washed in Mill-Q water for 2 mins and stained in hematoxylin (MHS16, Sigma Aldrich) for 20 mins, blued in bluing reagent (CS702, Dako) for 1 min, and then eosin (3801615, Leica) for 30 secs. Samples were finalized, dehydrated and covered with coverslips for digital imaging (Pannoramic MIDI, 3D Histech) at 40x magnification.

### Raw Data Handling, Quality Control, Preprocessing, and Dimension Reduction

Raw data from the Xenium instrument's output, including feature-cell matrices and cell boundary files, were integrated into the Seurat V5 pipeline. During quality control (QC), we aimed to retain a broad dataset, setting the inclusion criteria at a minimum of 5 gene features per cell and genes expressed in at least 5 cells across the tissue. The gene count upper limit per cell was determined based on the distribution of the number of features per cell. We assessed gene expression complexity by calculating the  $\log_{10}$  transformed gene feature count divided by the  $\log_{10}$  UMI count ( $\log_{10}\text{GenesPerUMI}$ ), setting the acceptable range from 0.6 to 0.985. The expression matrix underwent log normalization following the Seurat procedure. FindVariableFeatures from Seurat was utilized for the automated selection of

highly variable genes, identifying 252 such genes in our dataset. Data were then z-scored using the ScaleData function. For inter-sample integration, we applied the robust principal component analysis (RPCA) method via the IntegrateData function. Post-merging, the data underwent another z-score transformation to prepare for downstream processing. Dimension reduction was performed through t-Distributed Stochastic Neighbor Embedding (t-SNE) or Uniform Manifold Approximation and Projection (U-MAP), using the default Louvain clustering algorithm. To delineate clusters, the top 20 principal components were extracted, and 34 clusters were defined at a resolution of one, employing the FindNeighbors and FindClusters functions.

### Spatial Single-Cell Referencing, Phenotyping, and In-Situ Projection

Cell type annotation was performed according to a pre-defined marker gene list (detailed in Supp. Table 4). We initially annotated the main cell types: breast cancer epithelium using markers such as EPCAM, KRT5, KRT6B, ERBB2, GATA3, and EGFR; breast glandular cells using SERPINA3 and DSC2; immune cells using markers PTPRC, TRAC, and CD3E; fibroblasts using SPARC, PDGFRA, and CCDC80; smooth muscle cells using PDGFRB, MYH11, and ACTA2; and endothelial cells using PECAM1, VWF, and AQP1.

Subsequent immune cell re-clustering, based on the top 15 principal components at a resolution of 0.8, resulted in 23 identifiable clusters. These clusters were annotated as T cells (CD3E, CD3G), B cells (CD19, CD79A, MS4A1), macrophages (CD68, CD163), myeloid cells (ITGAM), plasma cells (TNFRSF17), monocytes (CD14, FCGR3A), natural killer cells (GNLY, NKG7, NCAM1), dendritic cells (CD83, FCER1A), neutrophils (CEACAM8), and mast cells (CPA3, CTSG, TPSAB1). To delineate T cell subsets further, we annotated them as CD4 + T cells (CD4), CD8 + T cells (CD8A, GZMB), regulatory T cells (FOXP3, IL2RA), naïve T cells (SELL, CCR7), effector T cells (PRF1, CCL5), effector-memory T cells (IL7R, EOMES), exhausted T cells (LAG3, TIGIT), and proliferative T cells (MKI67). For in-situ gene annotation, the feature-cell matrix and cell boundary files were utilized, and following the above described cross-sample data merging, genes of interest were plotted using ggplot2. The same approach was applied for cell-type annotation.

### Gene Regulatory Network Analysis Using Cytoscape

We analyzed genes co-regulated with ESR1, PGR, and ERBB2 using data imported from the STRING database, excluding any unconnected entities. The Cytoscape plugin CytoHubba was employed to identify key regulatory networks, with the degree of connectivity score serving as a metric to determine key nodes within these networks. The color intensity of the nodes corresponds to their connectivity scores, while the intensities of edges reflect the confidence scores between linked nodes.

### Cellular Neighborhood Analysis of Breast Cancer Samples

For the cellular neighborhood (CN) analysis, we defined a cell neighbor window as the 10 cells nearest to each anchor cell, which is a cell designated as central within its respective neighborhood (15, 37). Each

CN was quantified as a collection of cell types, with frequencies summed to 100% based on our in-situ cell annotations. We performed hierarchical clustering of CNs across all cells in a tissue using K-means clustering within R (version 4.2.0), setting the number of clusters (K) to 10. The prevalence of each CN type within individual samples was normalized so that the sum across the tissue equaled 100%, thereby providing an estimation of the proportional representation of each CN type across the entire tissue sample.

#### Region-of-Interest Selection Criteria Based on H&E and Xenium Data

Regions exhibiting neoplastic characteristics and adjacent stromal areas, serving as controls, were delineated in all samples using hematoxylin and eosin (H&E) staining by a qualified pathologist. For spatial analysis, 20 ROIs were chosen from tumor-enriched areas and 10 from immune-stromal areas in each sample, resulting in a total of 120 ROIs with sizes ranging from 14,751 to 313,038  $\mu\text{m}^2$ . We also referred to paired IHC data for Ki-67 and HER2, in addition to H&E staining, to accurately identify the tumor-immune interface for subsequent ROI selection. Spatial x-y indices of individual ROIs were exported from the cell boundaries files, generating an expression matrix for each ROI. Those data were compiled and processed in the same way as described above.

#### Spatial Cell-Cell Interaction, Distance-Based Cellular Distribution, and Spatial Enrichment Analysis

To assess cell-cell associations at the pan-tissue level, we calculated the average expression levels of individual genes across different cell types as annotated by Xenium, comparing gene-wise correlations among all pairs. For interactions within specific ROIs, we employed the CellTrek R package, utilizing its SColoc module to estimate spatial compositional differences between samples using Kullback-Leibler divergence (KL). SColoc creates a minimum spanning tree (MST) to approximate spatial cellular adjacencies, with graphical representations generated through its Graphical User Interface (GUI). In these graphs, edges represent the strength of inter-cell type associations (with a threshold set at 0.3), while colors and sizes denote cell type and frequency within each ROI, respectively.

For profiling based on spatial distance, we defined each anchor cell type according to Xenium annotations and mapped the distribution of other cell subsets within a 100 $\mu\text{m}$  radius (38, 39). To assess spatial enrichment of single-cell subsets, we first determined the proportions of individual cell types within specific ROIs (denoted as Props-in-ROI). We then calculated the mean proportion of these cell types across all ROIs (MeanProps-in-ROI) for defined phenotype groups, such as responders and non-responders.

We further calculated the proportion of the same cell subsets across the entire tissue sample, referred to as Prop-in-slide. We then calculated the enrichment score (ESscore) as the ratio of MeanProps-in-ROI to Prop-in-slide (40). Additionally, we computed the standard deviation (SD) of cell proportions across the ROIs (SDProp-in-ROI). To assess the consistency of cell type distribution across different ROIs, we calculated the ratio of MeanProps-in-ROI to SDProp-in-ROI, yielding an estimate of inter-ROI variation

(ESV). For visualization plotting, we also calculated the log-transformed proportional ratio ( $\text{Log}(\text{ESscore})$ ) as  $\text{Log}_{10}(\text{MeanProps-in-ROI}/\text{Prop-in-slide})$ .

### Spatial Ligand-Receptor Analysis Using CellPhoneDB

The CellPhoneDB R package was employed to interrogate Xenium expression profiles within predefined tumor-enriched ROIs (41). Available ligand-receptor pairs with normalized expression data were extracted, including PTPRC-MRC1, CXCL12-CXCR4, CD86-CTLA4, CD80-CTLA4, and CD274-CD80. A permutation test was conducted to assess the co-expression levels of selected ligand-receptor pairs and their co-occurrence probabilities relative to a null distribution generated from random pairs. The results were then plotted for individual cell types. Significance was established at a p-value threshold of 0.05, with results displayed as the negative logarithm to the base 10 of the p-value ( $-\log_{10}(\text{p-value})$ ).

### In-Situ Gene Imputation Test Using Xenium Data

To assess the efficacy of current in-situ expression inference methods, we analyzed various input data formats. Single-cell RNA-seq data, serving as the reference, were obtained from the 10X Genomics preview dataset for human breast, which can be accessed at 10X Genomics Xenium in-situ Preview Dataset for Human Breast. Additionally, in-house generated Xenium in-situ breast cancer data, derived from this study, were utilized. Both normalized and non-normalized (raw) data formats were tested; for the latter, outputs were further processed with or without normalization transformation. Two principal spatial gene expression prediction algorithms were employed: SpaGE, which relies on k-nearest-neighbor regression, and Tangram, which is predicated on a deep-learning approach (42, 43). The predictive capacity of these algorithms was tested by split the Xenium in-situ expression data into three random subsets, using two-thirds for model training and the remaining third to validate the predictions against actual Xenium expression data. The accuracy of these in-situ gene prediction methods was evaluated by comparing predicted and true expression values across selected ROIs, with the Pearson correlation coefficient serving as the measure of expression correlation.

### Statistical Analysis Methods

All statistical analyses were conducted in R, version 4.2.0, employing specific packages as needed. The primary basis for comparing sample groups was treatment efficacy, categorized as PD-1 responders versus non-responders. Analyses incorporated cell proportional ratios, cell distance metrics, and frequencies within cellular neighborhoods (CN). A non-parametric Wilcoxon rank-sum test was utilized to determine statistical significance, with the following thresholds: not significant (n.s.) for  $p > 0.05$ , \* for  $p < 0.05$ , \*\* for  $p < 0.01$ , \*\*\* for  $p < 0.001$ , and \*\*\*\* for  $p < 0.0001$ . Bar graphs illustrate the mean expression with standard deviation (SD), while box-and-whisker plots depict the 25th to 75th percentiles, with whiskers extending from the minimum to the maximum values, including all data points.

## Results

# Spatial Single-Cell Transcriptomics Elucidates Distinct Cell Composition of Different Breast Cancer Types

We employed the Xenium spatial single-cell transcriptomics technology to analyze four breast cancer samples with diverse pathological backgrounds: luminal carcinoma with chest wall metastasis (LBC-Chest-Met), luminal carcinoma with lung metastasis (LBC-Lung-Met), HER2(2+)/HR- carcinoma with lymph node metastasis (Her2+/HR-LN-Met), and triple-negative carcinoma at a lymph node metastatic site (TNBC-LN-Met) (Supp. Table 1 and Fig. 1A). Using a 280-gene probe panel (10X Genomics, CA), spatial single-cell transcriptomic data were generated for 1,224,277 cells in total, with counts ranging from 99,803 to 627,775 cells per sample (Supp. Figure 1B). The specificity of gene detection was validated by comparing probe signal-to-noise ratios against internal spike-in controls (Supp. Figure 1C). After filtering cells with fewer than five detected genes, we retained 95.3% of the dataset, with a median of 41.5 genes per cell (Supp. Figure 1D).

We constructed a knowledge-based marker gene chart for potential cell type annotation at three tiers (Supp. Table 4). At the first tier, five major cell categories including immune cells, endothelial, epithelial, fibroblasts, and smooth muscle cells were annotated (Fig. 1A top row). At the second tier, immune cells were further subdivided into T cells, B cells, dendritic cells (DCs), macrophages, monocytes, natural killer T cells (NKTs), plasma cells, neutrophils, and mast cells (Fig. 1A, middle row). The third tier focused on T cells, annotating them into naive CD8, effector CD8, regulatory CD4, naive CD4, and proliferative T cells (Fig. 1A, bottom row). In total, 15 cell subtypes were annotated from all samples (Fig. 1B). Due to rare presence of histologically-defined normal breast epithelium in the tissues, these cells were excluded from the cell clustering and annotation processes (Supp. Figure 1C).

Different types of the BC samples corresponded to diverse TME compositions (Fig. 1B, left and right panels). LBC-Chest-Met was notable for its abundant tumor content (66.3%), in contrast to the other samples, which contained a lower fraction of cancer cells (ranging from 21.2–27.6%), likely attributable to differences in sampling locations (Fig. 1B, right panel). Notably, Her2+/HR-LN-Met and TNBC-LN-Met, the two samples obtained from lymph node metastasis, exhibited similar cellular profiles. In contrast, LBC-Lung-Met differed significantly, with a high prevalence of endothelial cells (ECs) and smooth muscle cells (SMCs), reflecting the unique cellular milieu of the lung TME (44).

We subsequently created in-situ maps based on the annotated cell types, which were then compared with H&E-stained images of the corresponding samples. The spatial localization of individual cells annotated from Xenium data corresponded well with anatomically defined features, effectively highlighting the contours of tumor epithelium, fibrotic/connective tissue, and immune stromal regions (Fig. 1C). Furthermore, when comparing gene expression profiles among cell types, a lineage-dependent correlation was observed. Epithelial, endothelial/smooth muscle, and immune cells formed distinct hierarchical clusters (Fig. 1D). Within the immune lineages, myeloid and lymphoid cells formed two major clusters: the myeloid cluster comprises monocytes, mast cells, macrophages, and the lymphoid cluster consists of effector T cells, regulatory T cells, naïve T cells, and NKT cells (Fig. 1D).

# Validation of Spatial Single-Cells Transcriptomic Data by IHC

In evaluating the quality of Xenium spatial single-cell data, we compared the mRNA data with IHC quantitative protein data from corresponding samples. The expression of four key histological markers, including estrogen receptor (ESR1), progesterone receptor (PGR), Her2 (ERBB2), and Ki-67 (MKI67), were examined by IHC at the protein level, identifying ESR1 positivity in LBC-lung-Met and LBC-Chest-Met at 80% and 1%, respectively, and PGR positivity at 70% and 5%, respectively (Fig. 2A). Her2 showed a 2+ positivity in both LBC-lung-Met and Her2+/HR-LN-Met, and Ki-67 levels varied between 50–70% across the sample set (Fig. 2A). To compare Seurat-normalized single-cell mRNA levels with the IHC data, 20 tumor-enriched regions of interest (ROIs) were randomly selected as a representative set (Fig. 2B). In line with IHC findings, the spatial transcriptomic data accurately mirrored the protein expression profiles. Specifically, LBC-Lung-Met displayed the highest ER $\alpha$  protein expression (80%) also showed significantly higher mRNA levels than other samples ( $p < 0.001$ ) (Fig. 2B). Remarkably, LBC-Chest-Met, where ER $\alpha$  protein was only indicated as 1% positive by IHC, was distinguished from two samples (Her2+/HR-LN-Met and TNBC-LN-Met) that were negative for ER $\alpha$  protein by the Xenium mRNA data ( $p < 0.001$ ) (Fig. 2B; Supp. Figure 2).

PR mRNA was the highest in LBC-Lung-Met, consistent with high PR protein (70%). PR mRNA in LBC-Chest-Met did not recapitulate the 5% positivity in protein, potentially due to region selection bias or the discordance between mRNA and protein expression (Fig. 2B). Her2+/HR-LN-Met exhibited significantly higher Her2 mRNA compared to other samples ( $p < 0.001$ ), while LBC-Lung-Met showed a modest increase mRNA level, albeit not statistically significant. Additionally, the correlation between Ki-67 mRNA and protein levels across samples further confirmed the accuracy of the Xenium mRNA data for quantitative analysis (Fig. 2B).

After validating the quality data with IHC, we profiled the spatial single-cell mRNA of all panel-included genes, focusing on gene-wise correlations to identify regulatory networks associated with ESR1, PGR, and HER2. Hierarchical clustering confirmed a more robust co-regulation between ESR1 and PGR, which was anticipated given their joint roles as hormone receptors. Conversely, HER2 exhibited a distinct co-regulatory pattern (Fig. 2C, upper panel). CytoSCAPE for network computations revealed that ESR1 and PGR are not only co-regulated but also closely associated with steroid hormone receptor co-activator PPARG, and transcription regulators FOXA1 and RUNX1, indicative of their integrated functions with hormonal pathways (Fig. 2C lower panel) (45, 46). These hormone receptors also participate in immune regulation, interfacing with B cells and macrophages, which underscores their involvement in the TME (47, 48). In contrast, the receptor tyrosine kinase HER2 anchors a separate network with pivotal tumor-associated signaling molecules, including EGFR, keratins (KRT7/8/14), CD9, PECAM, and MKI67 (Fig. 2C, lower panel). Taken together, the findings of the network diversity underscore the value of the spatial single-cell transcriptomics data for extensive downstream investigation.

# Spatial Single-Cell Transcriptomics Reveals Distinct Cell Composition in Specific Regions of Interest

We subsequently profiled the regulatory networks in the specific regions in the TME. The 20 tumor ROIs from each sample described above, showing spatial mRNA expression nicely mirrored histological patterns, were used for in-depth analysis. We further chose 10 immune-stroma ROIs from each of three samples for comparison. LBC-lung-Met was excluded from this selection due to the absence of pathologically identifiable immune-stroma regions.

The tumor-enriched ROIs showed diverse TME architectures, characterized by variations in tumor, immune, fibrotic, and vascular components. LBC-Chest-Met, with a notable high tumor cell presence of 78.5%, contrasted with lymph node and lung metastatic samples, which contained a lower proportion of tumor cells (ranging from 37.2–63%) (Fig. 3A). These latter samples were also distinguished by significant infiltration of lymphocytes, including monocytes, dendritic cells (DCs), various T cell subsets, natural killer T (NKT) cells, and B cells/plasma cells ( $p < 0.05$ , Fig. 3B, Supp. Figure 3).

In contrast to the tumor-enriched ROIs, the immune-stroma ROIs revealed a wide array of non-tumoral cellular compositions. Notably, LBC-Chest-Met and TNBC-LN-Met presented similar cellular profiles, with smooth muscle cells, macrophages, and endothelial cells as the predominant types (Fig. 3A). However, the distribution of plasma cells, effector CD8 T cells, and endothelial cells varied significantly between these two samples (Fig. 3A,  $p < 0.05$ ). In contrast, Her2+/HR-LN-Met exhibited a unique cellular profile with a prominent presence of B cells and effector-memory CD8 T cells (Fig. 3B left).

All the samples exhibited an inflamed TME, despite varying levels of immune cell infiltration. We then characterized cellular co-localization in the tumor ROIs using neighborhood proximity measures. As anticipated, each biologically unique BC sample presented a distinct cellular co-localization profile. Specifically, LBC-Chest-Met featured a pronounced cellular network where smooth muscle cells co-localized with a variety of immune cells, while cancer cells engaged predominantly with B cells (Fig. 3C). LBC-lung-Met displayed an SMC-centric regulatory module including a variety of immune cells with abundant macrophages tightly connected with CD8 effector T cells (Fig. 3C). TNBC-LN-Met revealed a network interconnected via CD4 regulatory T cells, with macrophages and effector CD8 T cells in direct contact with cancer cells (Fig. 3C). Her2+/HR-LN-Met had the most distinct network profile with loosely connected immune cell clusters containing a few T cell subsets, DCs, and NKT cells, whereas the SMC-centric immune network had direct interaction with cancer cells. Macrophages in this sample were noteworthy for their links to cytotoxic T cells and, to a lesser degree, cancer cells (Fig. 3C). Taken together, our spatial transcriptomics analysis, centered on regions with high tumor content, elucidated the intricate TME architectures unique to each breast cancer subtype.

## Spatial Single-Cell Transcriptomics Reveals Localization of Immune Cells as Therapeutic Implications

With the spatial single-cell transcriptomic profile of the BC TME established, we sought to examine the cellular network associated with clinical responses to anti-PD-1 therapy. We first compared the tumor-enriched ROIs of the responders (R: TNBC-LN-Met and LBC-Lung-Met) with the non-responders (NR: LBC-Chest-Met and Her2+/HR-LN-Met). Despite the inherent biological differences among these samples, the non-responders typically exhibited higher proportions of cancer cells ( $p < 0.05$ ). In contrast, the responders displayed a pronounced increase of T cell subsets, including cytotoxic, proliferative, and regulatory T cells ( $p < 0.001$ ,  $p < 0.05$ , and  $p < 0.05$ , respectively) (Fig. 4A). Plasma cells, monocytes, and macrophages exhibited a higher proportion in the responders compared to the non-responders, implicating their potential anti-tumor roles ( $p < 0.05$ , Fig. 4A).

Recognizing that cell functions are markedly influenced by its neighbor cells in the TME, we characterized higher-order cellular units by grouping the 10 nearest cells of every anchoring cell across the entire tissue (15). By constantly identifying 10 cellular neighborhoods (CNs) across 4 samples, we generated higher-ordered spatial CN map of the samples (Fig. 4B). Those included a breast tumor dominating cluster (CN1), a multiple T cell dominated fibroblasts cluster (CN2), a tumor-SMCs interacting cluster (CN3), an endothelial-immune interacting cluster (CN4), a tumor-interacting macrophage/effector-CD8 T cells cluster (CN5), a smooth muscle-fibroblast cluster (CN6), a CD8 cytotoxic T cell-dominating immune cluster (CN7), a macrophage/monocytes-T cells interacting cluster (CN8), a B cells-enriched cluster (CN9) and an endothelium/SMCs/tumor interacting cluster (CN10) (Fig. 4B/C left). We then assessed the CN frequencies between the two anti-PD1 response groups and identified 7 out of 10 CNs exhibited statistically significant difference in proportion ( $p < 0.05$ , Fig. 4C right). Besides the tumor-dominated cluster (CN1), the effector T cell-enriched cellular unit (CN7) showed the most pronounced upshift in responders ( $p = 3.6e^{-10}$ ), underscoring its canonical role in mediating cytotoxic tumor cell killing (49). For the CN clusters in close proximity to tumors, CN5 involving a higher frequency of tumor-interacting macrophages and effector T cells was more prevalent in responders, which likely signifies their important anti-tumor function in the local TME (Fig. 4B/C,  $p = 0.004$ ).

To extend above observation, we calculated the distance of each cell from tumor beds within a 100  $\mu\text{m}$  radius. In consistent, cytotoxic T cells, which exhibited a tumor-recruiting pattern in CN5, were generally located closer to tumor in the responders than in the non-responders (Fig. 4D). Similarly, tumor-interacting macrophages, another cell type implicated in the response, exhibited closer localization with tumor cells in the responders. Remarkably, within CN5, a subset of proliferative T cells marked by Ki-67 expression—though small in proportion (1.3% of the total)—was significantly more prevalent in tumor-adjacent niches in responders as opposed to non-responders, underscoring their role in T-cell mediated tumor cell killing (Fig. 4D and 4E). We further examined the cellular levels of spatial enrichment score (ESscore) of all cell types in tumor-enriched areas compared to the entire tissues. In the responders, a pronounced spatial co-recruitment of effector T cells, proliferative T cells, and macrophages was observed (Supp. Figure 5: higher  $\text{Log}(\text{ESscore})$  indicates greater enrichment in tumor regions between groups), consistent with previous studies (40). This observation was also corroborated by Ki-67 IHC staining (Fig. 4F). Taken together, these spatial analyses validate a robust framework that integrates

clinical phenotypes with spatial single-cell transcriptomic data, a methodology that promises further insights with more samples.

## **Cellular Communication Reveals Spatially-Defined Ligand/Receptor Pairs as Potential Biomarkers**

Upon identifying cellular niches of tumor-interacting T cells and macrophages in the ICI responders, we applied CellphoneDB for a deeper exploration of the cell-cell communication in the TME. In the tumor-enriched TME, comprehensive cell-cell communication networks demonstrated variable interaction intensities. However, the overall crosstalk among macrophages, effector T cells, and tumor cells was not significantly different between the two response groups, which likely attributed to the distinct biological background of the samples (Fig. 5A). To further explore potential difference in cell-cell communication between the response groups, we interrogated ligand-receptor pairs detected by Xenium including PTPRC/MRC1, CXCL12/CXCR4, CD86/CTLA4, CD80/CTLA4, and CD274/CD80. Macrophage cell communication with T cells mediated by the CD86/CTLA4 interaction, which is inhibitory to T cells, was identified in all samples (Fig. 5B, upper panel). In contrast, Macrophage-expressed CD274 (PD-L1) exhibited pronounced interaction with CD80 on DCs, T cells (proliferative T cells and CD4 Treg cells), and macrophages themselves exclusively in the responders (Fig. 5B, upper panel). The CD274/CD80 interaction, either in cis or in trans, is known to be stimulatory to T cells (50) (51), suggesting a T cell-activating cellular communication in the responder TME. In addition to macrophage, T cell subsets exhibited the inhibitory CD86/CTLA4 interaction exclusively in the responders (Fig. 5B, lower panel), potentially resulted from extensive activation of CD8 T cells. Notably, the stimulatory CD274/CD80 interaction was also observed between CD8 T cells and other immune cells in one responder, TNBC-LN-Met (Fig. 5B, lower panel). The CD274/CD80 interaction was also observed between tumor cells and immune cells in the responders but not in non-responders (Supp. Figure 6). Consistently, the spatial cell annotation from Xenium data demonstrated that the infiltrated CTLA4-positive effector T cells and PD-L1-positive macrophages directly communicate with the cancer cells exclusively in the responders (Fig. 5C).

Given the established role of PD-L1 as a biomarker for the ICI therapy response, we examined mRNA level of CD274 and other genes in tumor cells of the samples. As expected, CD274 (PD-L1) expression was significantly higher in the responders than in the non-responders (Fig. 5D). Notably, mRNA expression of PDCD1LG2 (PD-L2, associated with T cell proliferation), NKG7 (indicative of CD8 + T cell degranulation), and CD80 (implicated in M1-like macrophage-mediated T cell activation) was also significantly higher in the responder tumor cells. The differential expression of these genes was even more significant than that of CD274, suggesting them as superior potential biomarkers (Fig. 5D and Supp. Figure 7). This underscores the validity of our findings and suggests the promise of biomarker discovery within spatially resolved tissue contexts.

## **Gene Imputation to Integrate scRNA-seq and Xenium Data**

To enhance the depth of spatial single-cell transcriptomic analysis, we applied gene imputation methodologies to integrate single-cell RNA sequencing (scRNA-seq) data with the Xenium data. We

applied two imputation algorithms compatible with Xenium data, SpaGE (42) and Tangram (43). Two-third of genes from the Xenium dataset was randomly selected as the training set for the algorithms, which then inferred the spatial expression profiles for all genes in the scRNA dataset. The accuracy of these predicted profiles was assessed by comparison with the actual Xenium-derived expression data for the rest one-third genes as a validation set. The resulting imputation by SpaGE and Tangram achieved median correlation coefficients of  $R^2 = 0.73$  and  $R^2 = 0.69$ , respectively (Fig. 6A, Supp. Figure 8). While genes with lower expression levels posed challenges for precision prediction, the imputation reliably mapped the genes with medium to high expression levels to spatially-resolved single cells (Fig. 6B). Therefore, the imputation algorithms expanded the utility of the spatial single-cell transcriptomic data, allowing us to map a broader range of gene expressions beyond the initial Xenium dataset.

## Discussion

In this study, we applied the newly-emerged Xenium spatial single-cell transcriptomics technology to characterize the TME of various types of breast cancer samples. In addition to confirming the quality of the spatial single-cell transcriptomics data and creating our bioinformatics pipelines for analysis, we identified distinct cellular interactions and crosstalk at the single-cell level in the TME between different response groups, which not only advances our understanding of the complex TME but also facilitates potential biomarker development.

Using a fixed BC probe panel targeting ~ 280 genes from 10x Genomics, together with our proprietary bioinformatics pipelines developed for spatial single-cell transcriptomic data analysis, we successfully annotated the spatially-resolved single cells from the samples to 15 distinct cell types. The analysis and the annotation not only align with but also expand the results from previous studies (36, 52). Furthermore, the consistency between single-cell annotations derived from our spatial data and the pathological features identified via IHC underscore the robustness and quality of our dataset, affirming the subsequent analysis of the BC TME.

Our data indicate that the samples, with distinct clinicopathological features, display distinct immune profiles in the tumor-enriched regions. The TME of these regions is organized via different cellular crosstalk or, to a higher order, cellular neighborhood clusters. The strength of the true single-cell spatial resolution enables precise mapping of cellular networks in specific regions of the tissues. We were able to identify the distinct cellular composition and communication in either tumor or stroma regions, which cannot be achieved by conventional scRNA-seq. Specifically, we identified that multiple types of immune cells, including effector T cells, proliferative T cells, and macrophages, are preferentially localized near tumor cells in the responders to the ICI therapy, correlating with the recognized tumor-killing functions of these cells. Therefore, our findings suggest that not only the quantity of infiltrated immune cells but also their spatial localization could serve as biomarkers for patient stratification and prognosis. Notably, we observed a marked increase in CD86/CTLA4 and PD-L1/CD80 ligand-receptor interactions in the responder tumor regions, compared to non-responders. The PD-L1/CD80 interaction, either in cis or in trans, has been demonstrated to play a critical role in activating T cell-mediated cytotoxicity. The

presence of this interaction in the responder likely contributes to the efficacy of the ICI therapy. Given these compelling findings, further investigation involving a larger cohort of samples is warranted to confirm these interactions and their therapeutic implications, which could significantly enhance precision cancer immunotherapy.

Spatial single-cell transcriptomic analysis has become increasingly important in dissecting the complex tissue microenvironment and advancing biomarker development in BC as well as other cancers, particularly in the realm of immunotherapy. Spatially-defined biomarkers, such as PD-L1 evaluated via quantitative tumor proportion score (TPS) and combined proportion score (CPS), highlight the importance of spatially-resolved data in predicting responses to the ICI therapies (53, 54). Additionally, the evaluation of tumor-infiltrating lymphocytes (TILs) and the “Immunescape”, which measures the distribution of CD3/CD8 T cells in tumor regions, also underscores the importance of spatially-resolved analysis in the TME (55). Therefore, our study on the BC TME using spatial single-cell transcriptomics has not only expanded the current understanding of the heterogeneous nature of the TME but also aided in the identification and development of new biomarkers, which is significant to the precision immunotherapy for BC and other cancers.

This study is apparently limited by the small sample size, which constrains us to draw solid clinical conclusions. Nonetheless, the detailed analysis of cell communication through specific ligand/receptor interactions, at the spatial single-cell levels, provides insightful observations on the differences between responders and non-responders to immunotherapy, even from limited number of samples. These findings highlight the potential for clinically relevant discoveries, justifying further investigation in larger patient cohorts.

Another limitation is the number of genes detected in the current study. Although only ~ 280 genes in the fixed probe panel were analyzed, our methodology allowed for the annotation of 15 distinct cell types and the imputation of approximately 20,000 genes from a public single-cell RNA-seq dataset, broadening the Xenium dataset's scope for more comprehensive analysis. As the technologies evolve rapidly, the number of genes assessable by spatial single-cell transcriptomics is expected to increase significantly. Future large-scale studies analyzing more genes in more clinically relevant samples, will validate our preliminary findings in this study and advance the development of predictive and prognostic biomarkers.

In summary, our spatial single-cell transcriptomics study on different types of breast cancer tissues validated the newly-emerged Xenium technology and our bioinformatics pipelines to deeply characterize cell-cell communications in the TME and identified the CD86/CTLA4 and CD274/CD80 interactions between cancer cells and immune cells or among different immune cells as potential biomarkers for future development.

## Abbreviations

TME: tumor microenvironment, ICI: immune checkpoint inhibitor, ICB: immune checkpoint blockade, IHC, immunohistochemistry, KNN: k nearest neighbor, BC: breast cancer, TNBC: triple-negative breast cancer,

LBC: luminal type breast cancer, SPT: proteotranscriptomics, OV: oncolytic virus, CAR-T: chimeric antigen receptor-T cells, TCR-T: T cell receptor-engineered T cells, ST: spatial transcriptomics, SP: spatial proteomics, ITH: intratumoral heterogeneity, FFPE: formalin-fixed and paraffin-embedded ER: estrogen receptor, PR: progesterone receptor, CK: cytokeratin, RCA: rolling circle amplification, QC: quality controls, UMI: unique molecular identifier, RPCA: robust principal component analysis, U-MAP: Uniform Manifold Approximation and Projection, t-SNE: T-distributed Stochastic Neighbor Embedding, CN: cellular neighborhood, ROI: region-of-interest, KL: Kullback-Leibler divergence, MST: minimum spanning tree, GUI: Graphical User Interface, ESV: estimate of inter-ROI variation, LN: lymph node, TILs: tumor-infiltrating lymphocytes, ICI: immune checkpoint inhibitor

## Declarations

Ethics approval and consent to participate

All human resources were used under the standard regulation approved by the Ethical Review Board of Cancer Hospital Chinese Academy of Medical Sciences (National Cancer Center) with clinical registration number 22/074-3275.

Consent for publication

All authors made consent for publication.

Availability of data and materials

Xenium raw data files and pre-processing codes are available in supplementary files or can be found in Github including relevant source codes for analysis. Other computational codes and IHC raw data are available on request.

Completing interests

All authors declare no competing interests.

Funding

This work was supported by funding from Fynn Biotechnologies R&D division and grants from The National Key Research and Development Program of China (2021YFF1201300, 2022YFE0103600), National Natural Science Foundation of China (82230058,82172875,82073094), CAMS Innovation Fund for Medical Sciences (2021-I2M-1-014, 2022-I2M-2-001, 2022-I2M-C&T-B-067), Beijing Hope Run Special Fund of Cancer Foundation of China (LC2021A08).

Author contributions

NW: experimental design, analysis design and writing manuscripts, YS: all pathological experiment, support and analysis. WFH: analysis, relevant guidance and manuscript revision, HNM: clinical

supervision, guidance and manuscript revision, ZTS: analysis and bioinformatics supervision, WSD: analysis, LSW: analysis, HYZ: analysis, YYZ: experiment and supervision, QZ, experiment, HZ: experiment, TZ: technical support, YYW: technical support, YYL: technical support, JFM: technical support, CCS: general guidance and experimental supervision, MY: manuscript revision and guidance, HLQ: funding, supervision, clinical guidance and manuscript revision, FM, clinical guideline, project designs, coordination and supervision, ZYD: project supervision, funding and manuscript revision.

This project was financially supported by funding from Fynn Biotechnologies and grants from The National Key Research and Development Program of China, National Natural Science Foundation of China, CAMS Innovation Fund for Medical Sciences and Beijing Hope Run Special Fund of Cancer Foundation of China.

## Acknowledgments

Not available.

## References

1. Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. An introduction to spatial transcriptomics for biomedical research. *Genome Med.* 2022;14(1):68.
2. Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, et al. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn.* 2012;14(1):22-9.
3. Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A, et al. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc Natl Acad Sci U S A.* 2013;110(29):11982-7.
4. Giesen C, Wang HA, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods.* 2014;11(4):417-22.
5. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, et al. Highly multiplexed subcellular RNA sequencing in situ. *Science.* 2014;343(6177):1360-3.
6. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science.* 2015;348(6233):aaa6090.
7. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc.* 2015;10(3):442-58.
8. Eng CL, Lawson M, Zhu Q, Dries R, Koulina N, Takei Y, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature.* 2019;568(7751):235-9.
9. Salmen F, Stahl PL, Mollbrink A, Navarro JF, Vickovic S, Frisen J, Lundeberg J. Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat Protoc.* 2018;13(11):2501-34.

10. Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol.* 2020;38(5):586-99.
11. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, et al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell.* 2018;174(4):968-81 e15.
12. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science.* 2016;353(6294):78-82.
13. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol.* 2008;26(3):317-25.
14. Keren L, Bosse M, Marquez D, Angoshtari R, Jain S, Varma S, et al. A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell.* 2018;174(6):1373-87 e19.
15. Schurch CM, Bhate SS, Barlow GL, Phillips DJ, Noti L, Zlobec I, et al. Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell.* 2020;182(5):1341-59 e19.
16. Asp M, Bergenstrahle J, Lundeberg J. Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *Bioessays.* 2020;42(10):e1900221.
17. Wang N, Li X, Wang R, Ding Z. Spatial transcriptomics and proteomics technologies for deconvoluting the tumor microenvironment. *Biotechnol J.* 2021;16(9):e2100041.
18. Wang WJ, Chu LX, He LY, Zhang MJ, Dang KT, Gao C, et al. Spatial transcriptomics: recent developments and insights in respiratory research. *Mil Med Res.* 2023;10(1):38.
19. Li B, Zhang W, Guo C, Xu H, Li L, Fang M, et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat Methods.* 2022;19(6):662-70.
20. Yan L, Sun X. Benchmarking and integration of methods for deconvoluting spatial transcriptomic data. *Bioinformatics.* 2023;39(1).
21. Andersson A, Larsson L, Stenbeck L, Salmen F, Ehinger A, Wu SZ, et al. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat Commun.* 2021;12(1):6012.
22. Dimitrov D, Turei D, Garrido-Rodriguez M, Burmedi PL, Nagai JS, Boys C, et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat Commun.* 2022;13(1):3224.
23. He S, Bhatt R, Brown C, Brown EA, Buhr DL, Chantranuvatana K, et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol.* 2022;40(12):1794-806.
24. Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature.* 2021;598(7879):137-43.

25. Janesick A, Shelansky R, Gottscho A, Wagner F, Rouault M, Beliakoff G, et al. High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. *bioRxiv*. 2022:2022.10.06.510405.
26. Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*. 2022;185(10):1777-92 e21.
27. Xu Y, Zhang T, Zhou Q, Hu M, Qi Y, Xue Y, et al. A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nat Cell Biol*. 2023;25(4):604-15.
28. Chu Y, Dai E, Li Y, Han G, Pei G, Ingram DR, et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. *Nat Med*. 2023;29(6):1550-62.
29. Garrido-Trigo A, Corraliza AM, Veny M, Dotti I, Melon-Ardanaz E, Rill A, et al. Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease. *Nat Commun*. 2023;14(1):4506.
30. Ali HR, Jackson HW, Zanutelli VRT, Danenberg E, Fischer JR, Bardwell H, et al. Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nat Cancer*. 2020;1(2):163-75.
31. Lan B, Li Q, Ma F, Xu B. The road toward breast cancer single-disease quality control in China. *Cancer Innov*. 2023;2(5):319-22.
32. Zhai J, Wu Y, Ma F, Kaklamani V, Xu B. Advances in medical treatment of breast cancer in 2022. *Cancer Innov*. 2023;2(1):1-17.
33. Wang XQ, Danenberg E, Huang CS, Egle D, Callari M, Bermejo B, et al. Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature*. 2023;621(7980):868-76.
34. Salmén F, Vickovic S, Larsson L, Stenbeck L, Vallon-Christersson J, Ehinger A, et al. Multidimensional transcriptomics provides detailed information about immune cell distribution and identity in HER2+ breast tumors. *bioRxiv*. 2018:358937.
35. Kumar T, Nee K, Wei R, He S, Nguyen QH, Bai S, et al. A spatially resolved single-cell genomic atlas of the adult human breast. *Nature*. 2023;620(7972):181-91.
36. Seo ES, Lee B, Hwang I, Kim J-Y, Park K, Park W-Y. Decoding spatial organization maps and context-specific landscapes of breast cancer and its microenvironment via high-resolution spatial transcriptomic analysis. *bioRxiv*. 2023:2023.10.25.563904.
37. Karimi E, Yu MW, Maritan SM, Perus LJM, Rezanejad M, Sorin M, et al. Single-cell spatial immune landscapes of primary and metastatic brain tumours. *Nature*. 2023;614(7948):555-63.
38. Yang X, Wang G, Song Y, Zhuang T, Li Y, Xie Y, et al. PD-1(+)/CD8(+) T Cells Proximal to PD-L1(+)/CD68(+) Macrophages Are Associated with Poor Prognosis in Pancreatic Ductal Adenocarcinoma Patients. *Cancers (Basel)*. 2023;15(5).
39. Shahidehpour RK, Nelson AS, Sanders LG, Embry CR, Nelson PT, Bachstetter AD. The localization of molecularly distinct microglia populations to Alzheimer's disease pathologies using QUIVER. *Acta Neuropathol Commun*. 2023;11(1):45.

40. Ruf B, Bruhns M, Babaei S, Kedei N, Ma L, Revsine M, et al. Tumor-associated macrophages trigger MAIT cell dysfunction at the HCC invasive margin. *Cell*. 2023;186(17):3686-705 e32.
41. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc*. 2020;15(4):1484-506.
42. Abdelaal T, Mourragui S, Mahfouz A, Reinders MJT. SpaGE: Spatial Gene Enhancement using scRNA-seq. *Nucleic Acids Res*. 2020;48(18):e107.
43. Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods*. 2021;18(11):1352-62.
44. Guo M, Morley MP, Jiang C, Wu Y, Li G, Du Y, et al. Guided construction of single cell reference for human and mouse lung. *Nat Commun*. 2023;14(1):4566.
45. Hong D, Fritz AJ, Gordon JA, Tye CE, Boyd JR, Tracy KM, et al. RUNX1-dependent mechanisms in biological control and dysregulation in cancer. *J Cell Physiol*. 2019;234(6):8597-609.
46. Robinson JL, Carroll JS. FoxA1 is a key mediator of hormonal response in breast and prostate cancer. *Front Endocrinol (Lausanne)*. 2012;3:68.
47. Jiang XF, Tang QL, Li HG, Shen XM, Luo X, Wang XY, Lin ZQ. Tumor-associated macrophages correlate with progesterone receptor loss in endometrial endometrioid adenocarcinoma. *J Obstet Gynaecol Res*. 2013;39(4):855-63.
48. Asaba J, Bandyopadhyay M, Kindy M, Dasgupta S. Estrogen receptor signal in regulation of B cell activation during diverse immune responses. *Int J Biochem Cell Biol*. 2015;68:42-7.
49. Koh CH, Lee S, Kwak M, Kim BS, Chung Y. CD8 T-cell subsets: heterogeneity, functions, and therapeutic potential. *Exp Mol Med*. 2023;55(11):2287-99.
50. Zhang Y, Song Q, Cassady K, Lee M, Tang H, Zheng M, et al. Blockade of trans PD-L1 interaction with CD80 augments antitumor immunity. *Proceedings of the National Academy of Sciences*. 2023;120(16):e2205085120.
51. Zhao Y, Lee CK, Lin CH, Gassen RB, Xu X, Huang Z, et al. PD-L1:CD80 Cis-Heterodimer Triggers the Co-stimulatory Receptor CD28 While Repressing the Inhibitory PD-1 and CTLA-4 Pathways. *Immunity*. 2019;51(6):1059-73.e9.
52. Janesick A, Shelansky R, Gottscho AD, Wagner F, Williams SR, Rouault M, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat Commun*. 2023;14(1):8353.
53. Doroshow DB, Bhalla S, Beasley MB, Sholl LM, Kerr KM, Gnjatic S, et al. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat Rev Clin Oncol*. 2021;18(6):345-62.
54. Xue L, Li Y, Jiang L, Liu C, Cheng N, Guo C, et al. Concordance between four PD-L1 immunohistochemical assays and 22C3 pharmDx assay in esophageal squamous cell carcinoma in a multicenter study. *Journal of the National Cancer Center*. 2023.

# Figures

Figure 1

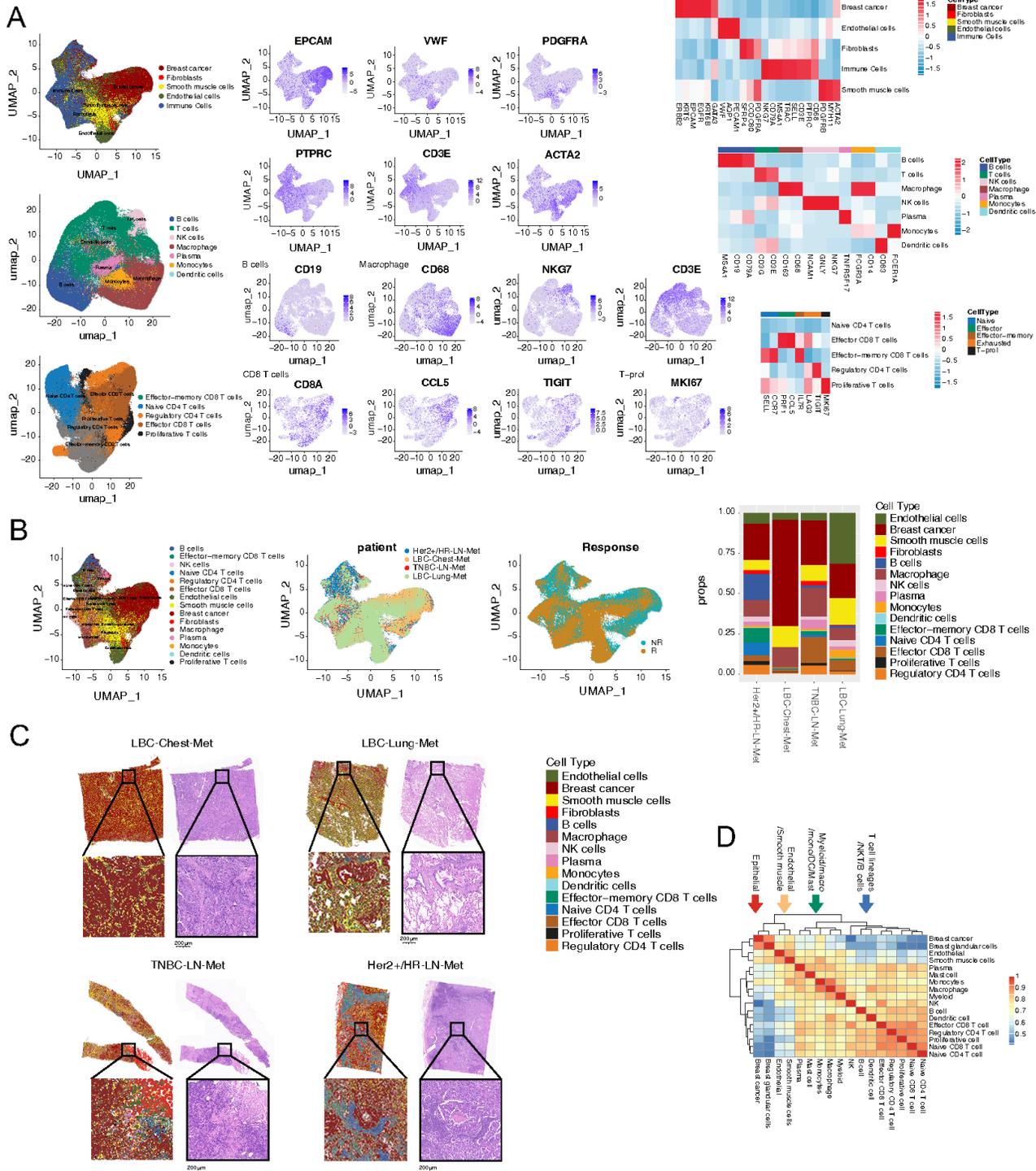


Figure 1

## In-situ Characterization of Cellular Phenotypes in Breast Cancer Tissues Using Spatial Single-Cell Transcriptomics

A. U-MAP Dimensional Reduction: Clustering based on gene annotations highlights large clusters and immune subsets with T cell subsets depicted separately (left panel). Middle panels display representative gene expressions projected onto U-MAPs. Right panels illustrate marker gene expression within each cluster with all color bars indicated. B. U-MAP Overviews: Plots display all annotated cell types, patient data, and drug response levels. The right chart quantifies cell proportions per sample, scaled to 100%, from the complete tissue dataset. C. In-Situ Cellular Distributions: Displays cell types for individual samples alongside parallel H&E staining post-Xenium experiment, with scale bars provided for reference. D. Cell Type Correlations: Correlations among all defined cells based on gene expression with lineage differentiation indicated.

Figure 2

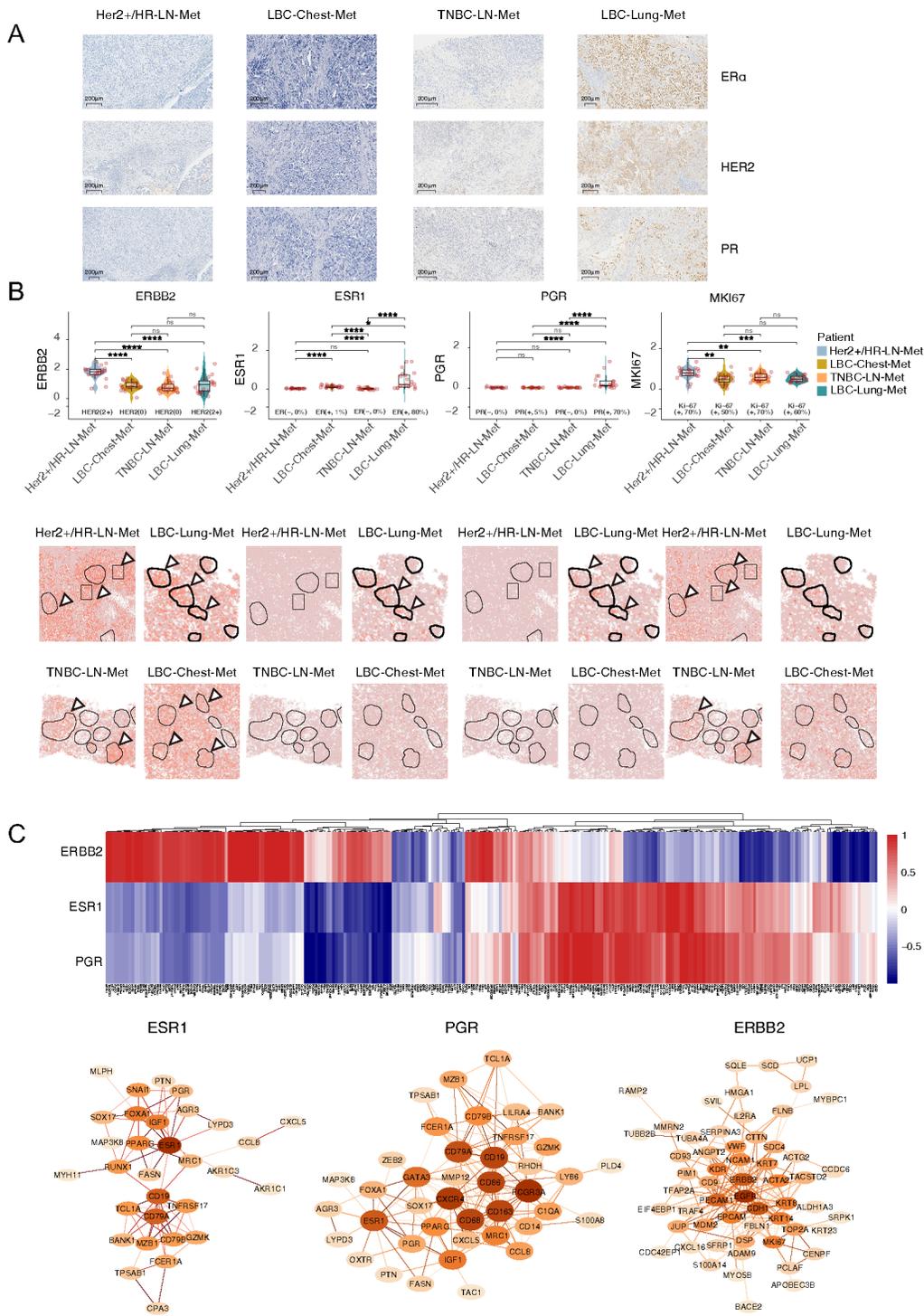


Figure 2

Comparative Analysis of IHC and Xenium-Derived Spatial Transcriptomic Data for Breast Cancer Marker Genes

A. Immunohistochemistry (IHC) Visualizations: ER $\alpha$ , Her2, and PR at 100X magnification for individual samples with scale bars provided. B. Spatial Transcriptomic Profiling: Upper Panel: Spatial

transcriptional profiles of ERBB2, ESR1, PGR, and MKI67 from selected tumor ROIs (20 per sample). Significant p-values are shown with asterisks. Lower Panel: Representative spatial gene expression maps for ERBB2, ESR1, PGR, and MKI67. Tumor ROIs are outlined, and color bars denote expression levels normalized across samples, with IHC expression levels indicated below each panel. C. Gene-Gene Correlation and Network Analysis: Left Panel: Unsupervised hierarchical clustering of gene-gene correlations across ESR1, PGR, and ERBB2, computed from the most to least co-expressed genes in four breast cancer samples. Correlation scales range from -1 (uncorrelated) to 1 (correlated). Right Panel: Hub genes and regulatory networks associated with ESR1, PGR, or ERBB2. Node colors represent connectivity degree, and edges indicate confidence scores (darker indicates higher confidence).

Figure 3

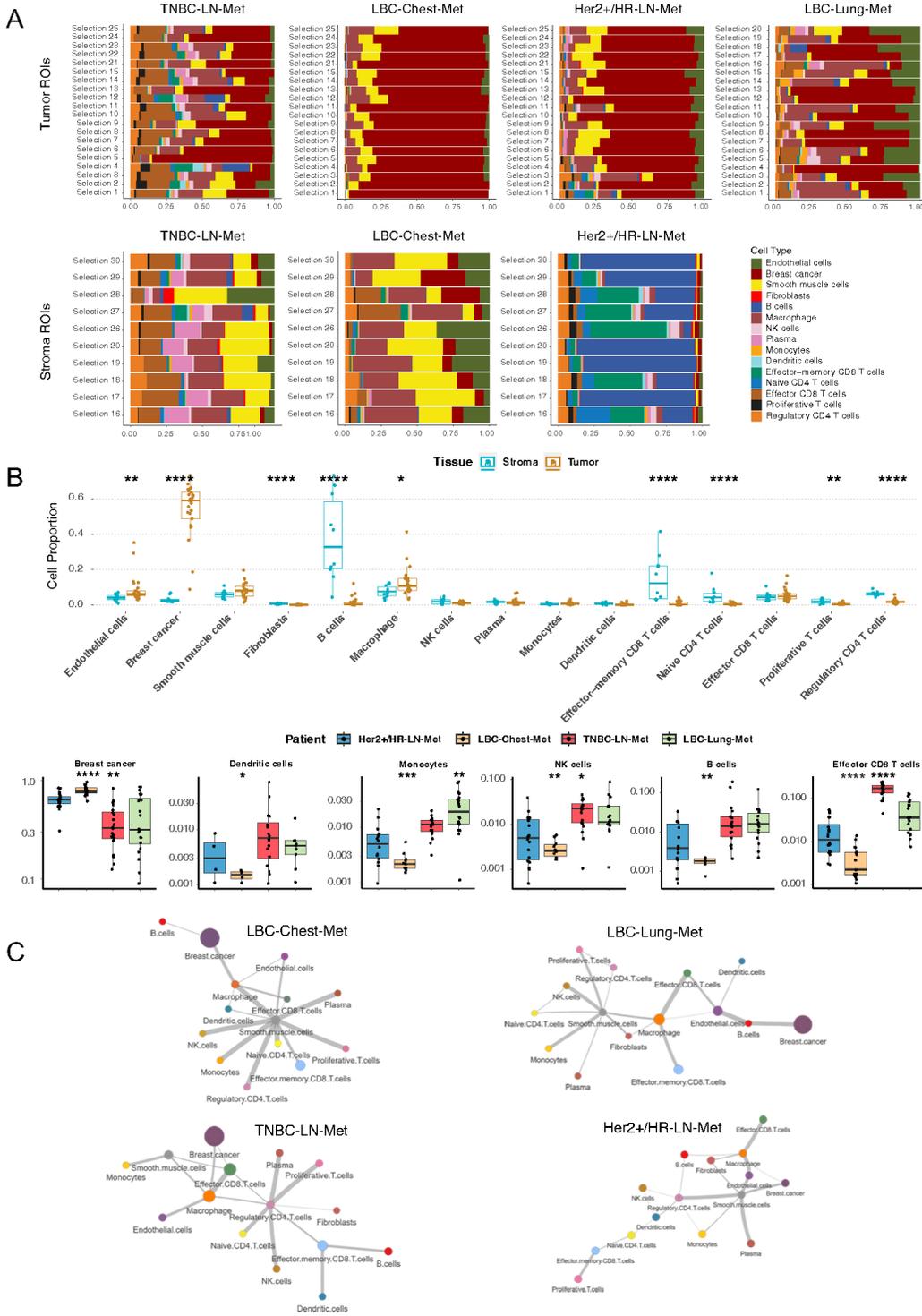


Figure 3

### Cell Composition and Communication in the Breast Cancer TME

A. Cellular Composition Mapping. ROI-level of the cellular composition of breast cancer samples in tumor regions (upper panel) or stroma regions (lower panel). B. Quantitative Analysis of Cellular Distribution. Boxplots depicting differential cell distribution between tumor and stromal ROIs of all

samples (upper panel) or individual cell alterations across the four samples. Significant p-values are shown in asterisks. The y-axis quantifies cell distribution as a percentage of the total. C. Cell-Cell Communication in the Tumor ROIs. Dot sizes indicate cell frequencies within ROIs, while edges represent the intensity of interaction between pairs of cell types. Pairs with strength above 0.3 are shown (range from 0.3 to 1).

Figure 4

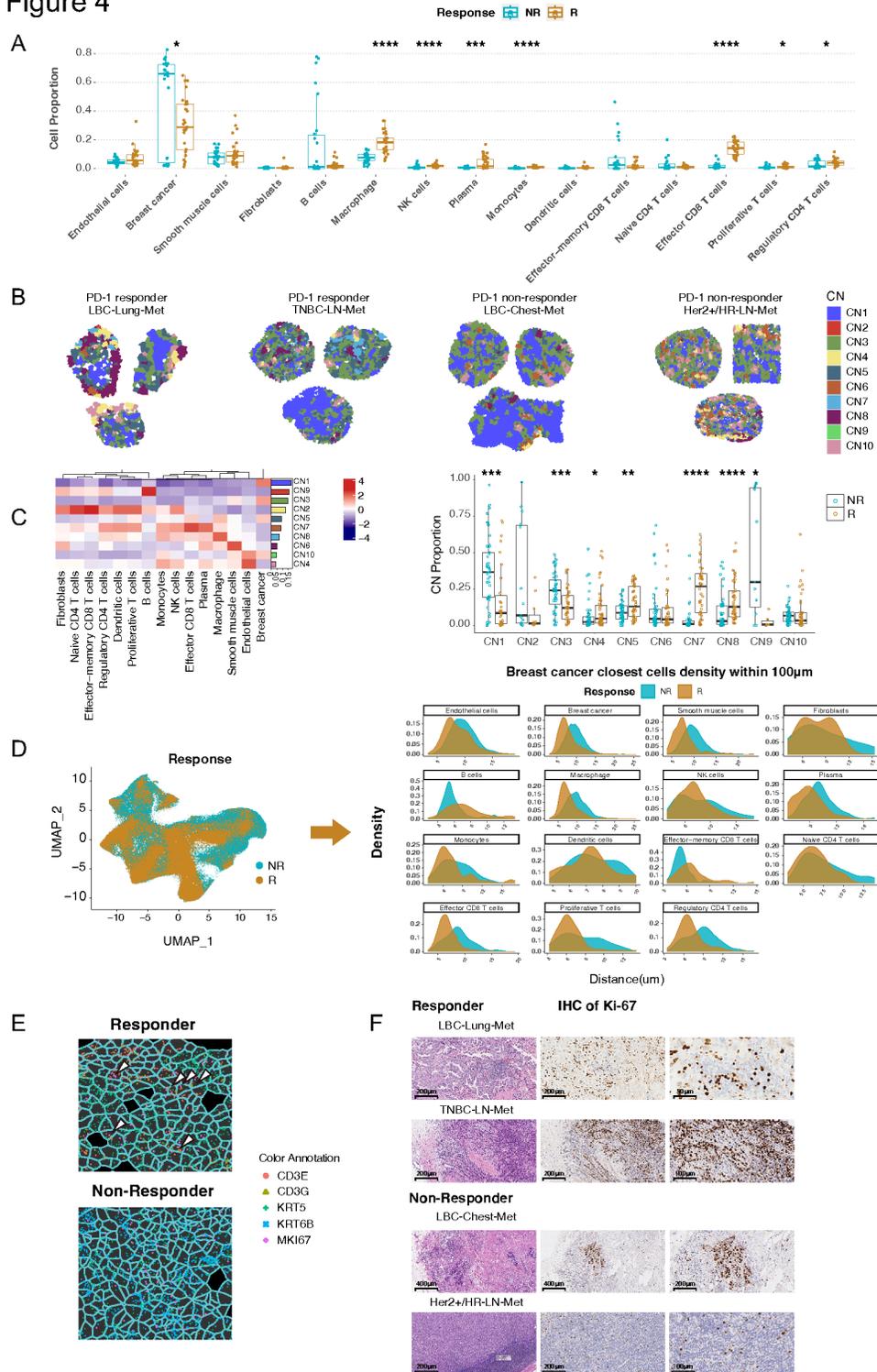


Figure 4

## Distinct Cellular Organization in Tumor ROIs at Cellular Neighborhood (CN) and Single-cell Levels

A. Differential cellular composition in tumor-enriched ROIs between ICI responders (R) and non-responders (NR). Statistic significances are marked by asterisks (\* for  $p < 0.05$ , \*\* for  $p < 0.01$ , \*\*\* for  $p < 0.001$ , and \*\*\*\* for  $p < 0.0001$ ). B. Recurrent CN detected (total of 10) in tumor-enriched ROIs across samples. Representative ROIs are shown. C. Semi-supervised hierarchical clustering using a KNN-based method (k-10). The 10 CNs are ranked by their distribution frequencies. Cellular compositions of each CN are shown at the bottom (left panel). The right panel shows the different CN composition in the responders compared with the non-responders (N vs. NR). Statistic significances are marked by asterisks as in A. D. Spatial distribution of individual cell types measured within 100  $\mu\text{m}$  of any anchoring breast cancer cells in the responders compared with the non-responders (N vs. NR). E. Graphical presentation of in-situ probe mapping for CD3E/CD3G (T cells), KRT5/KRT6B (tumor), and MKI67 (proliferation marker) comparing R and NR patients. Arrows denote proliferative T cells. Images are derived from Xenium data. Cell boundaries were defined by Xenium. F. Representative IHC staining for ki-67 in individual patients (R and NR), captured at 100X magnification. Scale bars provided.

Figure 5

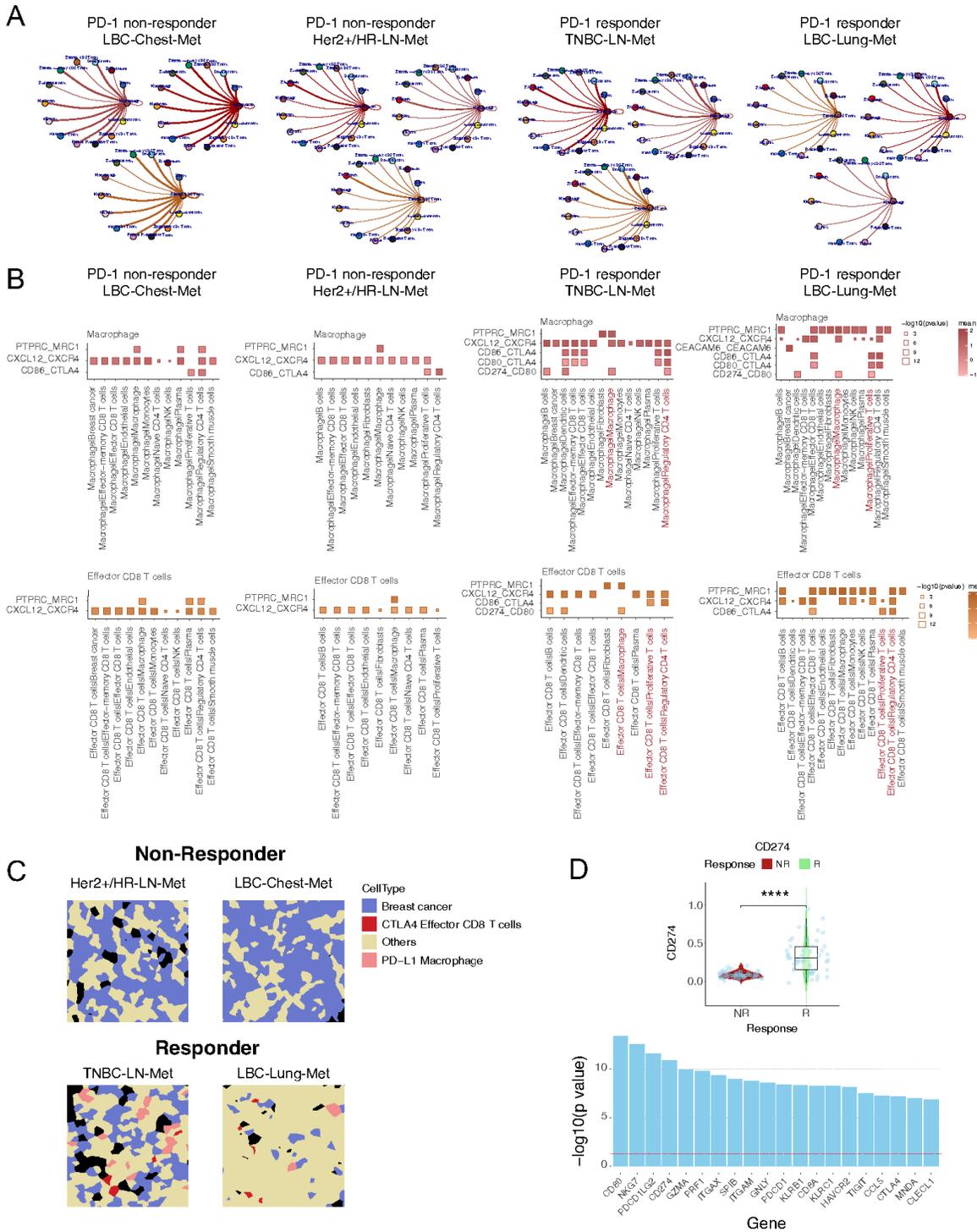


Figure 5

Cellular crosstalk analysis reveals ligand-receptor interactions among macrophages and T cells as potential biomarkers

A. Global cell-cell communications centering macrophages, effector T cells, and breast cancer cells, respectively. Crosstalk intensity is represented by the thickness of the connecting lines. The statistical

cut-off is set at 0.05. B. Ligand-receptor interaction analysis performed via available molecule pairs for individual patients. Different cell types are color-coded (Sienna for effector T cells, IndianRed for macrophages), and color intensity represents  $-\log_{10}(\text{p-value})$  measuring probabilistic co-presence of pairs. C. Distinct localization of specific immune cells in the responders compared with the non-responders. Infiltrated PD-L1+ macrophages (pink) and CTLA4+ effector T cells (red) are in direct contact with tumor cells in the responders. Tumor cells are marked in blue. D. Differential expression of CD274(PD-L1) in tumor cells from tumor-enriched ROIs between the responders and the non-responders (R and NR,  $\text{p-value} < 0.00001$ , Wilcoxon ranked sum test). The lower graph shows the top 20 differentially expressed genes in tumor ROIs between R and NR.

Figure 6

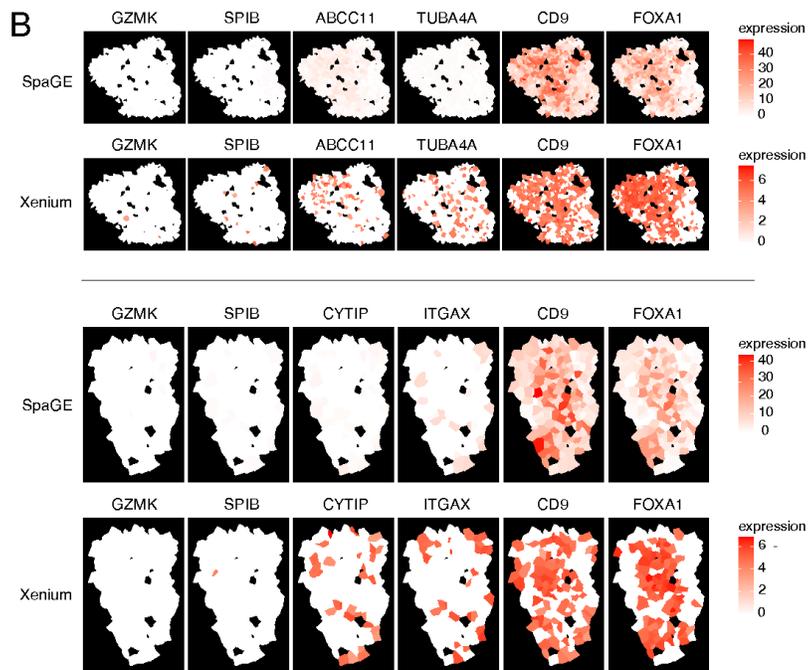
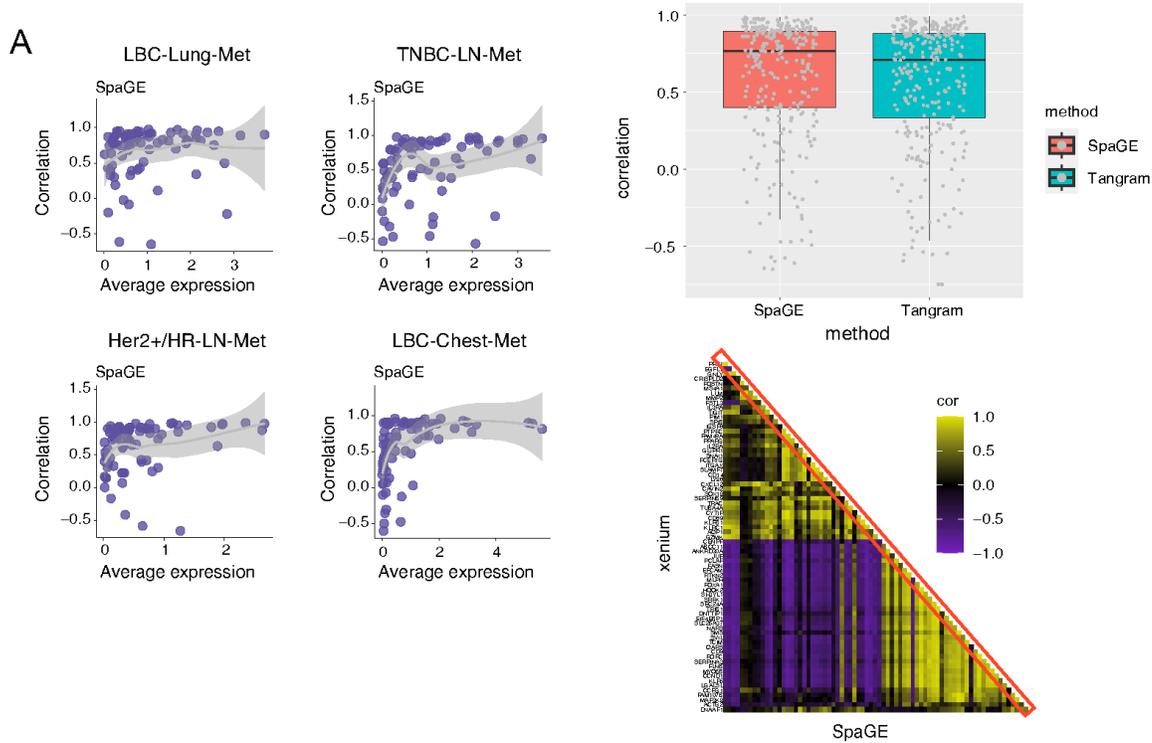


Figure 6

### Spatial Gene Imputation Using SpaGE and Tangram

A. Spatial gene expression inference using SpaGE and Tangram. The true average expressions of individual gene targets collected from Xenium are shown on the X-axis, and the predicted expression correlation on the y-axis ( $R^2$  range between -1 and 1). Box plots on the right depict the distribution of

overall prediction correlations, with lines indicating median value. A representative correlation matrix of true Xenium detected (ground truth) and SpaGE-imputed (predicted) expression is shown (lower right,  $R^2$  between -1 and 1). Red diagonal lines indicate each pair of genes (Xenium detected versus imputed). B. Two representative ROIs for selected example genes (expression levels from low, median to high from left to right) predicted by SpaGE (upper) and measured by Xenium (lower).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigureLegends.docx](#)
- [SuppFigure1.pdf](#)
- [SuppFigure2.pdf](#)
- [SuppFigure3.pdf](#)
- [SuppFigure4.pdf](#)
- [SuppFigure5.pdf](#)
- [SuppFigure6.pdf](#)
- [SuppFigure7.pdf](#)
- [SuppFigure8.pdf](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)