

# Highly multiplexed imaging reveals prognostic immune and stromal spatial biomarkers in breast cancer

Jennifer R. Eng<sup>1,2\*</sup>, Elmar Bucher<sup>2</sup>, Zhi Hu<sup>2</sup>, Cameron R. Walker<sup>5</sup>, Tyler Risom<sup>6</sup>, Michael Angelo<sup>5</sup>, Paula Gonzalez<sup>3</sup>, Melinda Sanders<sup>3,4</sup>, A. Bapsi Chakravarthy<sup>3</sup>, Jennifer A. Pietenpol<sup>3,7</sup>, Summer L. Gibbs<sup>2</sup>, Rosalie C. Sears<sup>1</sup>, Koei Chin<sup>2,8\*</sup>.

## AFFILIATIONS

1. Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR, USA.
2. Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR, USA.
3. Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA.
4. Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA.
5. Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA.
6. Department of Research Pathology, Genentech, South San Francisco, CA, USA.
7. Department of Biochemistry, Vanderbilt University, Nashville, TN, USA.
8. Cancer Early Detection Advanced Research Center, Oregon Health and Science University, Portland, OR, USA.

## Disclosures:

*The authors have declared that no conflict of interest exists.*

## \*Corresponding author:

Koei Chin, 3181 SW Sam Jackson Park Rd, Portland, OR 97239, USA. 503-494-1617, chinkoei@ohsu.edu.

Jennifer R. Eng, 3181 SW Sam Jackson Park Rd, Portland, OR 97239, USA. 503-346-1127, engje@ohsu.edu.

## ABSTRACT

Spatial profiling of tissues promises to elucidate tumor-microenvironment interactions and generate prognostic and predictive biomarkers. We analyzed single-cell, spatial data from three multiplex imaging technologies: cyclic immunofluorescence (CycIF) data we generated from 102 breast cancer patients with clinical follow-up, and publicly available imaging mass cytometry and multiplex ion-beam imaging datasets. Similar single-cell phenotyping results across imaging platforms enabled combined analysis of epithelial phenotypes to delineate prognostic subtypes among estrogen-receptor positive (ER+) patients. We utilized discovery and validation cohorts to identify biomarkers with prognostic value. Increased lymphocyte infiltration was independently associated with longer survival in triple-negative (TN) and high-proliferation ER+ breast tumors. An assessment of ten spatial analysis methods revealed robust spatial biomarkers. In ER+ disease, quiescent stromal cells close to tumor were abundant in good prognosis tumors, while tumor cell neighborhoods containing mixed fibroblast phenotypes were enriched in poor prognosis tumors. In TN disease, macrophage/tumor and B/T lymphocyte neighbors were enriched and lymphocytes were dispersed in good prognosis tumors, while tumor cell neighborhoods containing vimentin-positive fibroblasts were enriched in poor prognosis tumors. In conclusion, we generated comparable single-cell spatial proteomic data from several clinical cohorts to enable prognostic spatial biomarker identification and validation.

## INTRODUCTION

Recent advances in the breast cancer treatment landscape have motivated the characterization of the breast tumor microenvironment for deeper understanding of tumor-stroma interactions. For example, identifying biomarkers predicting breast cancer immunotherapy response is still an urgent clinical need (1). In metastatic TNBC, only a quarter of PD-L1 positive patients respond to single-agent immune checkpoint blockade (1). In contrast, in early-stage TNBC, response rates to neoadjuvant immune checkpoint plus chemotherapy were similar in PD-L1 positive and negative groups (2). This highlights the need for biomarker development for better patient stratification across disease stages and spatial characterization for improved understanding of tumor-microenvironment interactions in resistant patients.

Highly multiplex imaging methods enable quantification of dozens of biomarkers in a single tissue section at sub-cellular resolution while retaining spatial context (3–8). Tissue structures such as tertiary lymphoid structures, identified with multiplex imaging, are predictive biomarkers of immunotherapy response in melanoma (9,10). Spatial proximity between tumor, immune and stromal cell types is associated with response to neoadjuvant therapy in HER2+ breast cancer (11). In several breast cancer multiplex imaging studies, single-cell spatial context has prognostic relevance and shows correlations with transcriptomic and genomic features of tumors (12–15).

However, spatial biomarkers can be difficult to reproduce due to limited numbers of patients used to develop them and difficulties in comparing data from different imaging platforms. Furthermore, overfitting is an issue in biomedical imaging data due to the number of steps in the processing pipeline and the number of variables and parameters involved. Overfitting can be addressed through use of a discovery cohort to tune analytical methods, which are then fixed and subsequently applied to a validation cohort (16). In theory, validation cohorts can be readily obtained by incorporating publicly available data from disparate imaging platforms into biomarker studies. In practice, integrated analysis of such data remains a challenge. Furthermore, metadata documentation, analysis protocols and code are essential for data reuse and reproducibility of findings, preferably using open-source software tools (16). We developed an open-source python software, *mplexable* (17), for multiplex image processing and analysis, which we use herein to process and analyze three multiplex imaging breast cancer cohorts: a cyclic immunofluorescence (CycIF) dataset which we generated, and publicly available imaging mass cytometry (IMC) and multiplex ion-beam imaging (MIBI) datasets (12,14,18). This proof-of-concept study identifies prognostic single-cell spatial biomarkers common across imaging platforms. As imaging datasets become more widely available, our tools can facilitate biomarker discovery with high accuracy, reliability, and efficiency.

## RESULTS

### *Multiplex imaging datasets from different platforms produce similar single-cell phenotypes*

We generated CycIF data from two tissue microarrays (TMAs) containing surgical breast cancer specimens and obtained publicly available breast cancer multiplex images generated with antibody panels targeting similar antigens from IMC and MIBI platforms (Figure 1A-B). Image processing was standardized across all three datasets using our mplexable pipeline (Figure 1A). The CycIF data was comprised of 42 biomarkers imaged at a resolution of 0.325  $\mu\text{m}$  per pixel in entire cores with 1.2 to 1.4 mm diameters, with 1-2 full cores imaged per patient (Figure 1C top). The IMC data included 35 biomarkers imaged in the largest square area contained within the TMA cores (from 0.6 - 0.8 mm diameter), at a resolution of 1  $\mu\text{m}$  per pixel (Figure 1C middle) (14). The MIBI data included 36 biomarkers imaged in 0.8 x 0.8 mm square ROIs at a resolution of 0.5  $\mu\text{m}$  per pixel (Figure 1C bottom)(12).

Clinicopathological information was available for the IMC and CycIF datasets. Stage, tumor size and patient age at diagnosis were available for 364 patients in the CycIF and IMC datasets and grade was available for 275 patients in the IMC dataset (Supplemental Figure 1A). Similar fractions of early and advanced stage patients were analyzed by IMC and CycIF, apart from the CycIF TNBC group lacking stage 4 patients (Supplemental Figure 1B). Treatment information was available for the CycIF dataset; sixteen TNBC patients and one ER+ patient were treated with neoadjuvant chemotherapy (NAC). NAC-treated patients had shorter recurrence-free survival (RFS) with no difference in overall survival (OS, Supplemental Figure 1C). Multivariable CPH modeling revealed that tumor size but not NAC independently predicted RFS (Supplemental Figure 1C), and NAC treated patients had larger tumors ( $p < 0.001$ , Supplemental Figure 1D). Consistent with this, NAC-treated tissues had a higher fraction of Ki67+ proliferating cells within the epithelial compartment as well as a lower fraction of Vimentin+ fibroblasts in the non-epithelial stroma; but these were not significant after correcting for multiple testing (Supplemental Figure 1E). These results are consistent with NAC being given to downsize tumors before surgery in a subset of patients.

We used the same analytical methods for each platform to generate single-cell phenotypes via unsupervised Leiden clustering followed by visual inspection and annotation. We validated this method on serial breast cancer TMA slides approximately 20  $\mu\text{m}$  apart stained with similar CycIF panels on different dates (17). Across five cores, the average Pearson correlation R-value between cell type fractions was 0.99, SD=0.02, demonstrating excellent reproducibility (Supplemental Figure 1F-G).

In the CycIF dataset, we clustered cells using 23 markers plus nuclear area. We visualized the single cells using a UMAP embedding, confirming good separation of lineage specific markers: CD31, endothelial; E-cadherin (Ecad), epithelial; Collagen I (Coll), extracellular matrix; vimentin, mesenchymal cells including activated fibroblasts; and CD45, immune infiltrate (Supplemental Figure 2A). The UMAP visualization showed good mixing of cells from different TMA sources, indicating minimal batch effects, and separation of tumor cells from different breast cancer subtypes (Supplemental Figure 2B). Clustering resulted in 23 cell type clusters (Supplemental

Figure 2C). The mean expression of each biomarker in each cluster was used to annotate cell types and lineages (i.e., endothelial, epithelial, fibroblast, immune and other stromal, Supplemental Figure 2D). The most common cell types included luminal and luminal ER+ tumor, CD4 T cells, vimentin+ fibroblasts and quiescent stroma (Supplemental Figure 2D). To confirm our clustering-based cell typing, we performed manual gating on lineage specific markers (Supplemental Figure 2E). Gating and clustering-based cell lineages localized to similar areas of the UMAP and had 73% agreement on a single cell level, as calculated using metrics.accuracy\_score in scikit-learn (20) (Supplemental Figure 2E, F).

As in the CycIF dataset, we visualized lineage marker expression, TMA batch effects, breast cancer subtype and clustering results on a UMAP in the IMC dataset, confirming separation of cell lineages and mixing of cells from separate TMAs (Supplemental Figure 3A-C). We used 21 markers plus nuclear area for clustering, resulting in 24 cell types (Supplemental Figure 3D). Upon annotation, we found the most common cell types were similar to those in the CycIF samples, namely luminal, luminal ER+, and ER+ HER2+ tumor, vimentin or fibronectin (FN)+ fibroblasts, quiescent stroma, and T cells (Supplemental Figure 3D). Clustering and gating-based cell lineages localized similarly on a UMAP and had 77% agreement (Supplemental Figure 3E-F).

The MIBI panel included more immune-specific markers than the other panels and had only 15 markers shared with our CycIF panel (Figure 1B). To audit deeper immune contexture, we clustered using all 33 markers plus nuclear area and eccentricity. Again, we visualized lineage-specific markers and clustering with a UMAP to confirm separation of cell lineages and annotated 22 cell types of which luminal tumor, fibroblasts, T cells and quiescent stroma were the most common, similar to CycIF and IMC (Supplemental Figure 4A-C). Gating and clustering-based localized similarly on a UMAP and had 72% agreement (Supplemental Figure 4D-E).

### *Cell type fractions are similar across platforms*

While the most common cell types were similar across platforms, for further validation we compared the fractions of the five main cell lineages across platforms. In all three platforms, cell lineage identity and spatial distribution matched the underlying imaging data (Figure 1C-D). We calculated the fraction of cells in each lineage for each platform and breast cancer subtype. Both gating and clustering cell types showed high correlation (Pearson R=0.97 gating and 0.96 clustering) across platforms for ER+ (n=30 CycIF, 170 IMC) and TNBC (n=59 CycIF, 50 IMC, 41 MIBI) while HER2+, which had a smaller number of samples (n=8 CycIF, 22 IMC) had more variability between platforms (Figure 1E, Supplemental Figure 5A). We did note some platform-specific bias; for example, IMC showed a smaller fraction of immune cells defined by clustering in all three subtypes (Figure 1E). Therefore, when setting high/low cutoffs for cell abundances, we calculated high and low relative to each platform and subtype, as opposed to the whole dataset. Since different antibody clones, probes and imaging systems were used, resulting in different signal-to-background ratios between platforms, even for the same target, we believe this is necessary to account for technical variability.

As an additional validation, we obtained CycIF and MIBI data collected from sections of the same TMA containing normal tonsil, lymph node, liver, pancreas and placenta tissues. CycIF data was collected from the entire TMA core, while MIBI data was collected from 500 x 500  $\mu\text{m}$  ROIs in

each core (Supplemental Figure 5B). Despite the order of magnitude difference in the number of cells analyzed, the MIBI and CycIF data showed good concordance, with an overall Pearson correlation of 0.88 ( $p=0.008$ , Figure 1F). Cell types showed similar spatial localization between platforms and general consistency across the nine cores analyzed (mean Pearson's  $r=0.89$ ,  $SD=0.13$ , Supplemental Figure 5B, C).

Finally, we considered whether heterogeneity across the tumor would confound observations of cell type abundances from TMA cores. We collected CycIF data from full tissue sections of an ER+ and a TNBC tumor (21), and selected ROIs from within the tumor core, on the tumor/stroma border, and in the adjacent normal tissue (Supplemental Figure 5D-E). As expected, we found significant differences in epithelial and stromal cell phenotypes between tumor and adjacent normal after normalizing for epithelial content (Supplemental Figure 5F). In contrast, in both tumor sections we found no significant difference in cell type abundance in the tumor core and at the tumor/stroma border after normalizing for the number of epithelial cells (Supplemental Figure 5G-H). Without normalizing for epithelial content, border ROIs in TNBC had more CD8 T cells and quiescent stroma (False Discovery Rate,  $FDR<0.05$ ) and trended towards more B cells and macrophages ( $FDR<0.06$ ) while tumor ROIs had more luminal tumor (Supplemental Figure 5I). Intratumoral heterogeneity is a concern when analyzing TMAs, but if we normalize for epithelial content, variations in the location of the TMA core punch, i.e., closer to the tumor core or the tumor/stroma border, had no substantial impact on cell type abundance in this small sample.

#### *Unsupervised clustering defines prognostic tumor subtypes consistent across platforms*

First, we examined the prognostic value of epithelial cell type-based subtypes. To normalize for epithelial content, we calculated the fraction of each epithelial cell phenotype in the total epithelial cells in that sample. We then normalized across platforms by z-scoring cell fractions within each platform (Supplemental Figure 6A-D). For subtyping based on single-cell epithelial phenotypes, z-scored fractions of epithelial cell types (excluding rare cell types  $< 4\%$  of total) were used to Leiden cluster ER+ and TNBC patients from all platforms (Figure 2A, Supplemental Figure 6E). The resulting seven clusters included tissues enriched for luminal, basal, luminal ER+, myoepithelial, cytokeratin-low, and proliferating tumor cells, as well as a heterogeneous group not dominated by one phenotype (Figure 2A-B, Supplemental Figure 6F). Each subtype cluster contained a mixture of patients from multiple platforms, with no significant relationship between platform and subtype (Chi-squared  $p=0.2$  ER+,  $p=0.52$  TNBC, Figure 2C-D). The epithelial subtypes present in the ER+ patients were prognostic (log-rank OS  $p=0.04$ ,  $n=162$  patients, Figure 2E). Cox proportional hazards modelling (CPH) showed that the heterogeneous subtype 6 had a hazard ratio  $> 1$ , suggesting poor prognosis ( $p=0.079$ , Figure 2F). The CPH hazard ratios (HRs) were similarly ordered across the IMC and CycIF cohorts, with heterogeneous subtype 6 having a  $HR > 1$  in both platforms (Supplemental Figure 7A, B). Investigation of the poor prognosis subtype 6 tissues revealed expression of CD44 and EGFR in ER+ tumor cells (Supplemental Figure 7C). Quantification of epithelial marker expression showed increased CD44 and EGFR expression in subtype 6, but not more epithelial heterogeneity quantified by Shannon entropy ( $FDR<0.05$ , Figure 2G, Supplemental Figure 7D-E). We asked whether poor prognosis subtype 6 harbored any distinctive stromal cell types. Quantification showed significant enrichment of CD31+ endothelial cells in subtype 6 (Tukey HSD  $p<0.05$  Figure 2H). In TNBC patients the epithelial subtypes were not significantly associated with prognosis (Supplemental Figure 7F-H).

### *Microenvironment subtypes associate with clinical subtypes*

Next, we examined the prognostic value of stromal cell type-based subtypes. We clustered patients based on the stromal cell type fraction in each tissue in a similar manner to the epithelial subtyping above, selecting  $k=6$  for the number of clusters (Supplemental Figure 8A-C). Since the stromal phenotypes differed across platforms due to different markers, we clustered patients from each platform separately, using the fractions of stromal cell types representing greater than 2% of the stromal compartment. The stromal subtypes were not prognostic, except for the MIBI platform (log-rank=0.003, CPH  $p=0.056$ ,  $n=39$  patients, Supplemental Figure 8D-F). However, we observed significant correlation between stromal subtypes and clinical subtypes ER+ and TNBC. In the CycIF cohort, ER+ patients had significantly more of the vimentin+ (Vim) stromal subtype 0 and less T cell-rich stroma compared to TN (Chi-squared  $p=0.098$ , Bonferroni  $p\text{-adj}<0.05$  for subtype 0,  $n=89$  patients, Supplemental Figure 8G). Similarly, in the IMC cohort, ER+ patients had more Vim+ fibronectin+ (FN) fibroblast stromal subtype 0 and significantly less T cell stromal subtype 4 compared to TN (Chi squared  $p=0.002$ , Bonferroni  $p\text{-adj}<0.05$  for subtype 4,  $n=220$  patients, Supplemental Figure 8H). Our characterization of stromal subtypes supports the observation that ER+ breast cancer is immune-poor (22) and shows significant enrichment for Vim+ and FN+ fibroblasts relative to TNBC.

### *T cells are an independent prognostic factor in TNBC and high proliferation ER+ tumors*

To further investigate the prognostic value of multiplex imaging-defined cell types we utilized a discovery cohort to tune analytical methods, which were then fixed and subsequently applied to a validation cohort (16). The CycIF TMA1 dataset served as a discovery cohort to identify cell types whose fraction in the tissue were significantly associated with OS and determine cut-offs for prognostic high/low abundance. We tested the 0.33, 0.5 and 0.66 quantile to binarize tissues into low and high cell abundance (Supplemental Figure 9). For any cell type showing prognostic significance (log-rank  $p<0.05$ ), we selected the cut-off with the lowest  $p$ -value for validation in the other datasets, correcting for testing of multiple cell types using the Benjamini-Hochberg method. Using this methodology, we found that a high abundance of T and B cells were associated with longer OS in TNBC in both the discovery and validation cohorts (validation FDR $<0.05$ , Figure 3A-B).

Prognostic biomarkers significant as a single variable were combined with the clinical variables of stage, patient age and tumor size in a multivariate Cox-proportional hazards (CPH) model. In TNBC samples with clinical data, high CD3 T abundance remained significantly associated with longer OS in the multivariate model (CPH  $p=0.038$ ,  $n=88$ , Figure 3C). High CD20 B cell abundance trended towards longer OS in the multivariate model (CPH  $p=0.073$ ,  $n=89$ , Figure 3D).

CD3 T cells were not associated with OS in the ER+ breast cancer patients (FDR=0.12,  $n=162$ ), however, in tumors with proliferation above the median, increased CD3 T cells were associated with longer OS (FDR=0.0028,  $n=74$ ) (Figure 3E). Multivariate CPH modelling revealed that high CD3 T cells were independently prognostic for longer OS in high proliferation but not low proliferation ER+ tumors (CPH  $p=0.028$  high proliferation,  $p=0.99$  low proliferation, Figure 3F). Separation of tumors into high and low by median proliferation and median T cell abundance showed that in ER+ disease high proliferation, high T cell tumors had the best prognosis, while

high proliferation low T cell had the worse prognosis (log-rank  $p=0.0071$ , Figure 3G). In TNBC, both tumors with high proliferation and low proliferation (i.e., above and below the median, respectively), had similarly good survival if they had high-T cells (i.e., above the median), while low-T cell tumors had similarly poor survival regardless of proliferation status (log-rank  $p=0.011$  Figure 3G). Multivariate CPH modelling revealed that patients with high CD3 T cell abundance trended towards longer OS for high proliferation but not low proliferation TNBC tumors (CPH  $p=0.067$ ,  $p=0.31$ , Figure 3H). Survival analysis in each platform revealed outcome stratification by T cell abundance and proliferation in ER+ patients from the IMC cohort but not the CycIF cohort (log-rank  $p=0.028$  and  $0.4$ ), and in TNBC patients from the IMC and CycIF cohorts but not MIBI cohort (log-rank  $p=0.056$ ,  $0.012$  and  $0.49$ , Supplemental Figure 10A-B).

### *T cells in patients deriving a survival benefit from infiltration have distinct functional states*

To elucidate T cell functional states present in patients deriving a survival benefit from T cell infiltration, we compared T cell spatial localization and marker expression in T cell infiltrated groups with different survival outcomes. There was no significant difference in T cell abundance or T cell to macrophage or endothelial ratio in high versus low proliferation ER+ tumors (Supplemental Figure 10C). Interestingly, high-proliferation ER+ tumors in the IMC cohort, which gained survival benefit from CD3 T cells, showed more clustering of CD3 T cells than low-proliferation ER+ tumors, quantified by the mean number of T cell neighbors of each T cell (Tukey HSD  $p=0.01$ , Figure 3I). High proliferation ER+ tumors in the CycIF cohort, which did not gain a survival benefit from CD3 T cells, did not show increased T cell clustering (Kruskal-Wallis  $p=0.61$ , Supplemental Figure 10D). Similarly, CD3 T cells in high-proliferation ER+ tumors from IMC cohort had higher levels of the proliferation marker Ki67 and the memory/effector marker CD44 than in low proliferation ER+ tumors, indicating a more activated functional state (Tukey HSD  $p=0.04$  and  $0.001$ , Figure 3J-K). In the CycIF cohort, similar differences in Ki67 and CD44 expression were observed between ER+ and TNBC subtypes, consistent with an activated T cell state correlating with a survival benefit derived from increased T cell infiltration (Supplemental Figure 10E). High-proliferation TNBC, in which T cells independently predicted OS, showed increased levels of PD-1, FoxP3, IDO and Lag3 expression in T cells, consistent with upregulation of negative feedback checkpoints following immune activation (Supplemental Figure 10E-F). Epithelial cells in high-proliferation TNBC had increased expression of the antigen presentation molecule HLA-Class-1 and immune checkpoint PD-L1 (Supplemental Figure 10G).

### *Analysis of tumor-stromal proximity reveals robust spatial biomarkers*

Intrigued by the finding that T-cells in high-proliferation tumors had increased T cell neighbors and were associated with a survival benefit (Figure 3I), we leveraged our discovery and validation cohort analysis to systematically investigate cellular spatial relationships as prognostic biomarkers in breast cancer. We calculated the number of immune, stromal and tumor neighbors within proximity of each other to derive previously described biomarkers including mean neighbor counts (11,13), tumor-immune mixing score (12), immunoregulatory interactions (18), lymphocyte clustering and lymphocyte occupancy (15). We also used common statistical methods for quantification of spatial correlation (Ripley's L, Kcross and Gcross functions) (23). Analysis of proximity between cells of different lineages revealed that increased stromal (i.e., non-fibroblast, non-immune, non-endothelial) neighbors within 40  $\mu\text{m}$  of epithelial cells predicted longer



recurrence-free survival (RFS) in the discovery (log-rank  $p=0.018$ ) and validation ER+ cohorts (FDR=0.028), and independently predicted RFS in a multivariate CPH model with clinical co-variates (CPH  $p=0.02$ ), while increased immune neighbors within 40  $\mu\text{m}$  of immune cells associated with longer OS in both TNBC cohorts (validation FDR=0.019), but not in the multivariable model (CPH  $p=0.14$ , Figure 4A-D). Analysis of cell type neighbors showed increased macrophage neighbors within 40  $\mu\text{m}$  of tumor and increased B cell neighbors within 40  $\mu\text{m}$  of T cells trended towards longer RFS in TNBC in both cohorts (validation FDR=0.052) and remained significant in the multivariate model (CPH  $p=0.047$  Macrophage-tumor,  $p=0.017$  B cell-T cell, Figure 4E-H). Additional spatial metrics, including Ripley's L, Kcross and Gcross, did not yield any significant biomarkers in the validation cohort (Supplemental Figure 11A).

Previously, Ali *et al.* (13) showed that heterotypic neighbors of myofibroblasts, fibroblasts, cytokeratin low tumor cells, and vimentin+ Slug-macrophages were associated with poor outcome and homotypic neighbors of fibroblasts and myofibroblasts were associated with good outcomes in all breast cancer subtypes. We tested the prognostic value of heterotypic and homotypic neighbors of fibroblast subsets, CK low tumor and macrophages but did not find the significant association with survival (log-rank FDR>0.3, Supplemental Figure 11B-C). Previously, Keren *et al.* (12) showed that a high tumor-immune mixing score was associated with poor survival in TNBC. We were able to reproduce with our cell typing the prognostic value of the mixing score in the MIBI cohort, where it was developed (log-rank  $p=0.027$ , Figure 4I). However, in a validation cohort containing CycIF and IMC patients, the mixing score was not prognostic (log-rank  $p=0.26$ , Figure 4I), nor was it independently prognostic in samples with clinical outcome (CPH  $p=0.4$ , Supplemental Figure 11D).

Wortman *et al.* developed metrics for lymphocyte isolation and spatial dispersion which were linked to longer RFS in TNBC (15). Like Wortman *et al.*, we found that in our TNBC cores, the majority (57 percent) of lymphocytes near tumor cells (i.e. within 20  $\mu\text{m}$ ) were isolated, defined as fewer than 5 lymphocytes per 20  $\mu\text{m}$  radius, but unlike their findings, isolated lymphocytes near tumor in our data were not significantly associated with RFS (Supplemental Figure 11E). OS was significant, with greater numbers of isolated B cells near tumor cells associated with longer survival (log-rank FDR=0.014), which trended in a model with clinical co-variates ( $p=0.058$ , Supplemental Figure 11F). For lymphocytes near tumor cells, we calculated occupancy AUC, or area-under-the-curve of lymphocyte quadrant counts at different length scales, and fractal dimension (FD) difference, determined from slope of the log-log plot of the number of squares with at least one tumoral lymphocyte vs. the inverse box size, as described (15). Higher values of both metrics are indicative of spatial dispersion while lower values are associated with clustering. Similar to Wortman *et al.* (15), higher tumoral B and T cell occupancy AUC was associated with longer RFS in TNBC, which remained prognostic for B cells in the multivariable model (Figure 4J-K). While FD differences could only be calculated in tissues with lymphocytes present near tumor and intact tissue  $>200\mu\text{m}^2$  sampled, higher tumoral T and B cell FD differences were associated as single variables with longer OS (Figure 4L), and T cell FD trended in the multivariable model ( $p=0.063$ , Supplemental Figure 11G).

Formerly, Patwa *et al.* demonstrated in the MIBI cohort that increased spatial interaction between cells expressing immunoregulatory proteins PD-1, PD-L1, IDO and Lag3 were associated with longer RFS (18). We repeated this analysis using PD-1 in the CycIF TNBC dataset and found that

high PD-1 interactions were associated with longer OS, but not RFS (log-rank  $p=0.025$ , 0.2), and this trended in a multivariate model of OS with clinical covariates (CPH  $p=0.05$ , Figure 4M-N, Supplemental Figure 11H). Similar to Patwa *et al.*, we found no prognostic value of lineage marker interactions (Supplemental Figure 11I). We were not able to replicate the prognostic value of functional protein interactions (log-rank OS  $p=0.06$ ) and co-expression (log-rank OS  $p=0.38$ ) that Patwa *et al.*, previously reported (18) (Supplemental Figure 12A-B), although the CycIF panel's functional proteins only partially overlapped (7 of 18) with those analyzed in the MIBI dataset. Overall, our broad evaluation of spatial metrics across datasets revealed both examples of biomarkers with opposite survival associations in different platforms (Supplemental Figure 11A) and concordance with some previously identified biomarkers (15, 18), demonstrating the importance of a validation cohort (Supplemental Figure 12C).

To identify spatial metrics which provided additional information beyond abundance, we calculated the Pearson correlation between each spatial metric and cell type abundance within each patient's tissue (Supplemental Figure 13). Stromal neighbors of epithelial (good prognosis in ER+) correlated with quiescent stroma abundance and macrophage neighbors of tumor (good prognosis in TNBC) correlated with macrophage abundance (Supplemental Figure 13). CD20 B cell neighbors of T cells, isolated lymphocytes and lymphocyte occupancy AUC (good prognosis in TNBC) correlated with each other and T and B cell abundance (Supplemental Figure 13). The tumor-immune mixing score was positively correlated with tumor abundance and negatively correlated with immune abundance (Supplemental Figure 13). Many of the Kcross and Ripley's L function results correlated with each other and were not as strongly correlated with abundance (Supplemental Figure 13).

### *Neighborhood analysis reveals multicellular spatial biomarkers*

Finally, we analyzed multicellular spatial neighborhoods by considering stromal cells within 100  $\mu\text{m}$  radius of each tumor cell. We used spatial latent Dirichlet allocation (LDA) to model the neighborhood around each tumor cell as a combination of topics, utilizing a spatial parameter to increase the likelihood that adjacent cells share the same topics (24). LDA analysis can capture smoothly transitioning microenvironments (24) by assigning a probability for each topic to each neighborhood (Figure 5A). Each topic describes a microenvironment containing one or more cell type and each cell type can be in one or more topic; for example, in CycIF data, topic-0 in TNBC tissues was enriched in macrophages, vimentin+ fibroblasts and CD4 T cells (Figure 5B, cyan arrowhead), while CD4 T cells are found in topic-0, 4, 5 and 6 (Figure 5B, magenta arrowhead). After topic modelling, K-means clustering was run on the single-cell topic matrix to define "tumor neighborhood" clusters which contained one or more topics (Figure 5A-C). Clustering the topic matrix rather than the neighbor count matrix is believed to be less sensitive to noise (25). The spatial LDA neighborhood clusters were annotated based on their topics and examination of the images showed neighborhoods reflected the spatial distribution of the markers in the tissue (Figure 5A, Supplemental Figure 14A-D). As expected, we observed transitioning/mixed neighborhoods within both TNBC and ER+ neighborhood cluster results (Figure 5B-E). A TNBC tissue in our CycIF cohort, for example, showed tumor cell neighborhoods with more T cells (blue) on the tumor margin, with macrophage-rich neighborhoods in the tumor core (purple) (Figure 5A). These neighborhoods transition into a mixed neighborhood (brown), and finally a vimentin+ FB neighborhood (green) distant from the infiltrating T cells (Figure 5A). An ER+ tissue showed

quiescent stroma neighborhoods in the tumor core, vimentin+ FB neighborhoods on the tumor margin and T cell neighborhoods in tumor nests isolated in the stroma (Supplemental Figure 14A, D). Similar neighborhoods were identified in the IMC TNBC and ER+ cohorts (Supplemental Figure 14E-H) but the MIBI cohort was excluded from neighborhood analysis due to lower stromal cell type overlap (Figure 1B). For survival analysis, we used the CycIF TMA1 as a discovery cohort and CycIF TMA2 plus IMC patients as a validation cohort. Due to the small sample size, we used an FDR threshold of 0.1 rather than 0.05 to report findings. Increased Vimentin+ FB neighborhoods around tumor cells were associated with shorter OS in both TNBC cohorts (validation log-rank FDR=0.07, Figure 5F). Increased vimentin+ fibroblast neighborhoods were associated with shorter OS and RFS in the multivariate model (CPH  $p=0.049$  OS, 0.053 RFS Figure 5G). Interestingly, vimentin+ fibroblast abundance alone was not prognostic in TNBC (Supplemental Figure 14I). In ER+ tumors, increased mixed fibroblast neighborhoods containing vimentin-positive and -negative fibroblasts around tumor cells were associated with shorter OS (validation log-rank  $p=0.088$ ) and trended significant in the multivariable model (CPH  $p=0.087$  OS,  $p=0.046$  RFS, Figure 5H-I). Finally, similar to other groups (25), we found that directly clustering the neighborhood counts using Kmeans (rather than running LDA and clustering the topics) did not result in robust prediction of prognosis (Supplemental Figure 15).

#### *Tumor phenotypes correlate with stromal cell abundance and spatial neighborhoods*

We hypothesized that there would be significant correlation between tumor cell types and the spatial LDA neighborhoods, correlations that could shed light on biologically and clinically relevant tumor-stroma crosstalk. First, we visualized a matrix of pairwise correlation between epithelial and stromal cell fractions and spatial LDA neighborhoods across subtypes (Supplemental Figure 16). Epithelial cell types were inversely correlated with each other ( $p<0.05$  TNBC, except basal-like;  $p<0.001$  ER+ BC), indicating most tumors had just one main epithelial cell type (Supplemental Figure 16). The exception was luminal tumor, which correlated with cytokeratin low tumor in ER+ breast cancer, indicating mixing of these ER-negative phenotypes within the same tissues (Figure 5J,  $p<0.001$ ).

Immune cells exhibited distinct tissue-level correlations in the different subtypes. In ER+ breast cancer, T cells correlated with B cells (Figure 5J,  $p<0.001$ ), while proliferating tumor and macrophages correlated with endothelial cells ( $p<0.01$  and  $p<0.05$ , Supplemental Figure 16B) but not T cells (Figure 5J,  $p=0.35$ ). In TNBC, T cells correlated with proliferating tumor (Figure 5K,  $p=0.01$ ), and macrophages correlated with CD8 T cells (Figure 5K  $p=0.028$ ). Vimentin+, FN+, and ColI+ fibroblasts, as well as quiescent stroma were inversely correlated with immune cells ( $p<0.05$ , Supplemental Figure 16). In both subtypes, spatial LDA neighborhoods correlated strongly with the abundance of their respective stromal cell types; however, neighborhoods showed unique correlations to other cell types present. For example, proliferating tumor cell abundance did not correlate with T cell abundance in ER+ breast cancer, but it did correlate with the fraction of T cell neighborhoods (Figure 5L,  $p=0.015$ ). In TNBC, vimentin+ and fibronectin+ fibroblast abundances were not correlated (Supplemental Figure 16), but their respective neighborhoods were inversely correlated (Figure 5L,  $p=0.048$ ), suggesting exclusivity for a single fibroblast phenotype near tumor cells in each tissue. Therefore, although spatial neighborhoods tend to correlate with cell abundance, they can reveal unique features of tumor-stromal organization in tissues.

## DISCUSSION

Our approach of standardized processing and analysis across multiple imaging platforms shows the power of our methods for biomarker discovery. We incorporated analysis of two publicly available imaging datasets with our own CycIF data for efficient discovery of robust biomarkers.

We utilized our validated method for CycIF staining and image processing (17) to generate multiplex imaging data of 42 markers in single tissue sections from two TMAs with clinical follow-up. Our dataset alone represents a valuable new clinical cohort that provides improved plex, resolution, and ROI size compared to previously published datasets (12,14). We then developed an analysis pipeline ([https://github.com/engjen/cycIF\\_TMAs](https://github.com/engjen/cycIF_TMAs)) to generate single-cell phenotyping data from our CycIF dataset and two publicly available datasets (12,14). The advantage of using our pipeline for image processing is the development of smoothing algorithms so that pixelated IMC and MIBI data can be segmented with deep-learning models trained on higher resolution images and an algorithm to match nuclear and cell segmentation results from separate deep-learning segmentation models to extract features from subcellular compartments such as the nucleus and cytoplasm. Using our methods, we generated single-cell data that produced a high correlation between cell types across cohorts from the same breast cancer subtype and different platforms, as well as serial slides profiled with different platforms (Figure 1E-F).

Additionally, we identified similar epithelial phenotypes across platforms and clustered patient data from all platforms to separate seven epithelial subtypes without platform-specific bias (Figure 2). Our subtypes were consistent with the intrinsic breast cancer subtypes (26), including a luminal ER+ luminal A-like group with good prognosis and a cytokeratin low group previously shown to share features with luminal B tumors (13). Triple-negative breast cancers also fell into categories similar to those defined by gene expression profiling (27), including a highly proliferative, basal-like 1 (BL1)-like group, a luminal androgen receptor (LAR)-like group with luminal epithelial phenotypes, and a group with a basal/myoepithelial phenotype reminiscent of the basal-like 2 (BL2) group. We also identified a heterogeneous subtype with low cytokeratin and elevated CD44 expression that may represent tumors with mesenchymal features. Jackson *et al.* identified a similar single-cell pathology cluster of hormone-receptor-positive mixed tumors with poor prognosis (14). Our analysis showed that the ER+ tumors in the heterogeneous subtype had a poor prognosis and increased angiogenesis. An EMT program in breast cancer cells is linked to increased vascular endothelial growth factor A expression, increasing angiogenesis and the capacity for tumor initiation (28), a mechanism that could explain these correlated tumor and stromal phenotypes and their association with poor outcome.

Tumor-infiltrating lymphocytes have been linked to good prognosis in TNBC (29), and we confirmed that T and B cells are independently prognostic in TNBC in the multiplex imaging datasets analyzed herein. Previous gene expression profiling studies link productive anti-tumor immunity and tumor proliferation. Nagalla *et al.* found that immune signatures were prognostic solely in breast cancer patients with the highest proliferation gene expression (30). Subsequently, the same group showed that immune gene signatures were prognostic in highly proliferative basal-like, HER2-enriched and luminal B subtypes, but not those with low proliferation (31). Similarly, we have shown that CD3 T cells are independently prognostic specifically in high-proliferation ER+ and TNBC tumors.

Our analysis of immune functional states showed increased T cell proliferation, activation, checkpoint molecule expression and epithelial antigen presentation in high-proliferation tumors, consistent with IFN $\gamma$  pathway activation. Consistent with our analysis, gene network analysis previously showed activation of TNF $\alpha$ /IFN $\gamma$  signaling pathways in tumors with productive anti-tumor immunity and TGF- $\beta$ , an immunosuppressive cytokine, in tumors with unproductive anti-tumor immunity (31). TGF- $\beta$  also has anti-proliferative effects and is associated with good outcomes in ER+ breast cancer cohorts (32), suggesting that it could mechanistically link lower proliferation rates with immunosuppression and represent a rational drug combination with immune checkpoint targeting (33).

Thomas et al. recently showed that immune gene signatures were prognostic exclusively in tumor-mutation burden (TMB)-high breast cancer tumors (34). Thirty-seven percent of basal-like tumors had high TMB, while only 11.5% of luminal A tumors did (34), explaining the poor immunogenicity of the latter subtype. Together, these data point to a model of high TMB correlating with high proliferation and both linked to productive anti-tumor immunity. It had been hypothesized that oncogenes driving sustained proliferation also induce DNA replication stress, which generates genomic instability and presumably increase TMB (35). In summary, TMB provides a mechanistic link between proliferation and anti-tumor immunity and should be investigated in future studies. Furthermore, our analysis shows enrichment of potential immune checkpoint targets in high-proliferation breast cancer, including PD-1, Lag3, IDO and PD-L1 elevation.

One of the main goals of this study was to provide methods and a framework for robust identification of spatial biomarkers. Using external cohorts to validate biomarkers discovered in our CycIF data increases our confidence in biomarker identification. In ER+ breast cancer, we found that increased stromal neighbors of tumor correlated with better prognosis, similar to previous studies showing a survival benefit of high stroma in ER+ tumors (36). We found that macrophage proximity to tumor was associated with good prognosis in TNBC, which is unexpected given previous publications. Specifically, tumor-associated macrophages (TAMs) (37) were associated with shorter OS in a cohort of ER+ and ER- patients, but the prognostic value of macrophages specifically in TNBC was not investigated. Furthermore, Medrek *et al.* (38) found that CD68+ macrophages in close proximity to tumor cells were not associated with poor survival, but those out in the stroma were, suggesting that it may be difficult to compare our metric of macrophage-tumor neighbors in a 40  $\mu$ m radius with previous studies and further investigation is warranted. Finally, numerous immune-related spatial biomarkers, including immune-immune proximity, B cell-T cell proximity, immunoregulatory interactions, isolated lymphocyte abundance and lymphocyte occupancy, were associated with good prognosis in TNBC, supporting a model of productive anti-tumor immunity in the triple-negative subtype. Encouragingly, our results for the prognostic value of lymphocyte spatial metrics were similar to those of Wortman *et al.* (15) and Patwa *et al.* (18)

We utilized spatial LDA modelling to analyze multicellular neighborhoods of stromal cells surrounding tumor cells. We identified a neighborhood enriched for vimentin+ fibroblasts that was independently associated with shorter survival in TNBC. Given the high levels of vimentin and low levels of alpha-SMA, these cells may have an inflammatory phenotype similar to CAFs that

differentiate under  $\text{TNF}\alpha + \text{IL-1}\beta$  stimulation (39). Interestingly,  $\text{TNF}\alpha + \text{IL-1}\beta$  have been shown to stimulate pro-metastatic chemokine expression (CXCL8, CCL2 and CCL5) and aggressive characteristics in TNBC cell lines, mediated in part by direct CAF-tumor cell contact in co-cultures (40), consistent with proximity between putative poor-prognosis CAFs and tumor cells in spatial LDA neighborhoods.

The limitations of our study include different antibody probes and imaging systems resulting in different signal-to-background ratios for biomarkers across platforms. Therefore, our integrated analysis relied on matching annotated clusters across platforms. This introduces uncertainty since our annotations may not correspond to the same cell types in each platform. Some well-defined phenotypes, such as T cells and proliferating tumor, are relatively straightforward, and we found high correlation between normal cell types on adjacent slides analyzed on MIBI and CycIF platforms, respectively (Figure 1F, Supplemental Figure 5B-C). However, variable performance of antibodies, such as anti-ER, for example, could lead to variability in the classification of phenotypes such as luminal ER+ versus luminal tumor across platforms. To correct for platform-specific bias in cell types, we binarized patients into high/low expression within each subtype and platform for survival analysis. However, such binarization may not reflect underlying heterogeneity in quantitative biomarker abundance.

Another limitation of our study is the use of 1-2 TMA cores per patient for analysis. It has been shown that a limited number of TMA cores ( $\leq 3$ ) are needed to binarize patients into high and low TILs, although a larger number of cores ( $\geq 11$ ) are needed to accurately estimate the mean TIL abundance of a full tissue section (41). Our survival analysis relied on binarizing patients; therefore, the use of TMAs may be appropriate in this context. We undertook a limited analysis of large tissue sections from which we selected “virtual TMA” punches from the tumor core and border. We found that tumor and stromal cell type abundances were not markedly different in the core versus border after normalizing for epithelial content (Supplemental Figure 5D, G-I). Sampling larger tissue areas could potentially improve spatial biomarker performance, although increased heterogeneity in large sections could introduce noise, especially if they include adjacent normal, which we found to have significantly different epithelial and stromal phenotypes compared to the tumor (Supplemental Figure 5E-F). The optimal balance between the area analyzed in each tissue and number of patients included for estimation of prognostic tumor microenvironment composition and spatial architecture remains an open and important question in the field.

Overall, our spatial analysis supports the utility of spatial information in uncovering novel biomarkers of patient outcome in breast cancer. The tools developed in this study can be used to analyze additional cohorts further to characterize biomarkers in breast cancer and other tumor types.

## METHODS

### *Patient samples*

Two breast cancer tissue microarrays were provided by Dr. Jennifer Pietenpol (Vanderbilt University Medical Center [VUMC], Nashville, TN). All samples were collected at time of surgical resection (mastectomy or breast-conserving surgery) at VUMC with the same fixation protocol. JP-TMA1 had 131 cores of approximately 1.2 mm diameter, with duplicate cores from 19 TNBC, 8 HER2+ and 36 ER+ patients. Four of the TMA1 TNBC patients received neoadjuvant chemotherapy. JP-TMA2 contained a single, slightly larger (~1.4 mm diameter) core from 39 triple-negative tumors and 1 ER+/HER2+ core. Thirteen of the patients in TMA2 received neoadjuvant therapy. Clinical outcome and clinicopathological information were available for TMA1 and TMA2.

### *Sex as a biological variable*

Only female patients were included, as females account for more than 99% of breast cancer (42).

### *Imaging data generation and sources*

CycIF staining of tumor tissue was completed on JP-TMA1 and TMA2 using our protocol: [dx.doi.org/10.17504/protocols.io.23vggn6](https://doi.org/10.17504/protocols.io.23vggn6). Antibodies used for staining are available in the source data. The whole tissue core was imaged using fluorescence microscopy as described (17). MIBI imaging data was previously published by Keren *et al.* (12) and the images were downloaded from <https://mibi-share.ionpath.com/tracker/imageset> under the name “Keren et al., Triple Negative Breast Cancer.” Survival and recurrence data were obtained from a second publication by the same group (18), and were downloaded from <https://github.com/aalokpatwa/rasp-mibi>. IMC imaging data were previously published by Jackson *et al.* (14) and images and clinical data were downloaded from <https://doi.org/10.5281/zenodo.3518284>.

### *Image Processing*

CycIF tiff images were registered, segmented and single-cell intensity as well as nuclear size and shape features were extracted as described (17). Nuclear and cell segmentation were run using the Cellpose algorithm (43), which showed visually superior performance on CycIF data compared to a watershed algorithm (Supplemental Figure 17A). Nuclear and cell segmentation masks were matched using mplexable, enabling subtraction of nuclear mask from cell mask to obtain segmentation of the cytoplasm (Supplemental Figure 17B).

MIBI and IMC images were downloaded as multipage OMETiffs. Hot pixels (13) were detected by identifying pixels that were 10 standard deviations above a median filtered image with a 2x2 pixel kernel size. Hot pixels were set to the median filter values and resulting images were saved as tiffs for downstream feature extraction. For nuclear segmentation preprocessing, DNA images were rescaled between the 3rd and one and a half times the 99.999 quantile. The gamma value was

adjusted by 0.6 in MIBI data and 0.4 in IMC data to enhance dimly stained nuclei. A two-channel nuclear plus cytoplasm image was generated for cell segmentation. For MIBI cytoplasm segmentation preprocessing, the  $\beta$ -catenin, vimentin, CD45 and CD31 channels were combined into a maximum intensity projection image and the gamma value was adjusted by 0.6. For IMC cytoplasm segmentation preprocessing, E-cadherin, vimentin, CD44 and CD45 were combined into a maximum intensity projection image and gamma adjusted by 0.4. Chamboelle total variation denoising, implemented in scipy (44), was used to smooth out pixelated nuclear and cytoplasmic projection images (weight=0.1, except weight=0.05 for IMC cytoplasm). All parameters were selected by testing segmentation results at <https://www.deepcell.org/predict> and [cellpose.org](https://cellpose.org) (Supplemental Figure 17C-D). Skipping either nuclear or cytoplasmic Chamboelle total-variation de-noising resulted in failure of deep learning-based algorithms on the IMC data (Supplemental Figure 17E). Mesmer segmentation (45) performed better than Cellpose (43) in IMC data due to improved detection of dim nuclei, likely due to the incorporation of cytoplasmic staining in the nuclear segmentation model (Supplemental Figure 17D, yellow box). Cellpose was successful in CycIF images, which had brighter DNA staining. For IMC and MIBI data, nuclear and cellular segmentation were performed on preprocessed segmentation images using Mesmer (45). Matching of cell IDs in the nuclear and cell masks was done with mplexable (17), with cell masks relabeled to match the ID of the nucleus to which they had most overlap. Cytoplasm masks were calculated by subtracting the nuclear mask from the matching cell mask. Nuclear and cytoplasmic mean intensity, nuclear size and shape features, and nuclear centroid coordinates were extracted with mplexable (17).

Mesmer segmentation was compared to the watershed-based segmentation originally published by Jackson *et al.* (14). The cell counts across the two methods had a Pearson correlation of 0.98 (Supplemental Figure 18A). Visual examination of ROIs with discordant cell numbers revealed that Mesmer segmentation performed better in tissues with necrosis and high background noise in the DNA channel (Supplemental Figure 18B-D).

### *Image Quality Control*

In IMC data, artifacts include non-specific background staining, necrotic regions, and bright antibody aggregates. IMC data were collected from small ROIs ( $\sim 600^2 \mu\text{m}$ ) within TMA cores and some samples annotated as estrogen-receptor (ER)+ tissues did not show any ER+ staining in the ROI. Therefore, quality control (QC) was performed on ER-stained images, a marker noted to exhibit non-specific background staining on the IMC platform (46). QC images of ER staining were generated and sorted in a blinded fashion into negative and positive for nuclear-specific staining (Supplemental Figure 19A). Only ROIs from clinically annotated ER+ patients that were classified as ER+ during QC or ROIs that came from ER- patients and classified as ER- were used for analysis (Supplemental Figure 19B). Samples that passed ER QC did not have significantly different grade, PR status, TMA block, age of specimen, age of patient or tumor size compared to those that failed QC (Supplemental Figure 19C-D). There were no significant survival differences between QC passed versus failed tumors from ER+, TNBC or ER+HER2+ patients (Supplemental Figure 19E). Additional QC steps included: necrotic regions were manually circled using the napari (47) image viewer and excluded and bright aggregates in the CD3 channel were excluded by removing cells above a threshold set at the intensity of CD3+ cells showing an appropriate membranous staining pattern.



In the CycIF data, imaging artifacts included autofluorescence (AF), non-specific background, floating tissue and tissue loss. Background AF images were obtained halfway through CycIF data collection, and these images were scaled by exposure time in each round of staining and subtracted from the AF488, AF55 and AF647 channels using `mplexable` (17). Feature extraction was performed on AF subtracted images. Areas of floating tissue, air bubbles or necrotic regions were manually circled using the `napari` (47) image viewer and excluded. Non-specific background staining was removed by setting intensity thresholds for selected markers and subtracting those values from extracted data. The PD1 antibody had bright aggregates that were excluded with an upper threshold. Tissue loss was detected by cells that lacked DAPI staining in the last round of imaging, and these cells were excluded.

In all three platforms, additional artifacts caused by floating tissue or imaging problems (e.g., dark or bright bands across IMC and MIBI images, perhaps caused by problems with the rastering process) were detected through unsupervised clustering and visual inspection of clusters on the images. Clusters comprised of artifacts showed atypical, very bright or dim staining in many channels, formed distinct artifact clusters and were removed.

### *Single-cell phenotyping*

Cell types were defined in two ways: manual gating and unsupervised clustering. Unsupervised clustering was conducted using the `scanpy` (19) software. Single-cell mean intensity values were selected from either the nucleus or cytoplasm masks for each marker, depending on expected subcellular distribution. Since the CycIF and IMC platforms had more marker- and breast cancer subtype-overlap than the MIBI panel (Figure 1B, Supplemental Figure 20) twenty matching markers were selected for clustering in these datasets, plus selected markers for immune, epithelial and fibroblast subsets (collagen I [ColI], CD4, CD8 in CycIF, and fibronectin [FN], pan-cytokeratin [panCK] in IMC). For MIBI data, all available markers were used for clustering. Additionally, the nuclear area feature was used for clustering. Each marker was divided by its standard deviation, without zero-centering, and clipped above twenty standard deviations. A Uniform Manifold Approximation and Projection (UMAP) embedding was generated using 30 k-nearest neighbors and clustered using the Leiden community detection algorithm (48). The Leiden resolution parameter was selected that resulted in 20 - 25 clusters. Each cluster was annotated and categorized as epithelial, endothelial, fibroblast, immune or stromal. Some clusters were comprised of multiple expected cell types, and these were manually split, for example, the CD44+ cluster was split into CD44+ tumor and CD44+ stroma based on manual gating results (described below).

We then performed manual gating to verify our annotated-cluster cell type. A threshold was set for each gating marker based on the expected pattern of positive staining in images. Fibroblasts were defined as positive for one or more of vimentin, fibronectin (FN) or collagen I (ColI). Epithelial cells were defined as positive for one or more of Ecad, cytokeratins, or  $\beta$ -catenin. Endothelial cells were defined as CD31+. Immune cells were defined as CD45+. Stromal cells were defined as all non-fibroblast, non-endothelial, non-epithelial, non-immune segmented nuclei.

### *Patient Subtyping*

Epithelial and stromal subtypes were determined by unsupervised clustering of patients based on the fraction of epithelial or stromal cell types within each compartment, respectively. Cell types representing greater than 2-4% of the total cell population in the respective tissue compartment were used for clustering. This cutoff was chosen to ignore rare cell types that may represent method-specific artifacts. Cell fractions were normalized across platform using standard scaling. Unsupervised clustering of patients was performed using the Leiden algorithm implemented in scanpy (19). For epithelial subtypes, the resolution of clustering was selected to minimize differences between the platforms (Figure 2). For stromal subtypes, we selected the minimum number of clusters needed to separate T cells from other clusters ( $k=6$ ).

### *Survival Analysis*

The CycIF TMA1 dataset was used as a discovery dataset to determine the quantile separating high and low abundance of each cell type or spatial metric that was most predictive of survival. Three quantiles were tested: 0.33 (e.g., split patients into 1/3 low and 2/3 high), 0.5 and 0.66. The most prognostic cutoff value was selected for each cell type and for cell types having prognostic value ( $\alpha < 0.05$ ), these cutoffs were applied in the validation dataset, which included CycIF TMA2, MIBI and IMC samples. Since overall cell type fractions differed between platforms and subtypes (Figure 1), high and low values were determined relative to other samples from the same platform and subtype, using cutoffs from the discovery cohort. In the validation cohort, the log-rank test p-values were corrected for multiple testing using the Benjamini–Hochberg method. For biomarkers with  $FDR < 0.1$ , multivariate Cox proportional hazards (CPH) modelling was used to combine imaging biomarkers with patient age, tumor size and clinical stage to test if they were independently prognostic. Collectively, 89 TNBC and 160 ER+ patients had these additional clinical parameters (not available in the MIBI dataset).

### *Spatial Analysis*

Spatial distributions of cells were calculated as follows. For analysis of cellular neighbors (11) and homotypic/heterotypic interactions (13), each cell's neighbors within a 40  $\mu\text{m}$  radius were counted. For tumor-immune mixing score a 25  $\mu\text{m}$  radius was selected to replicate Keren *et al.* (12) For lymphocyte clusters, a 20  $\mu\text{m}$  radius was used and for lymphocyte occupancy a 10-300  $\mu\text{m}$  grid squares were used, at 10  $\mu\text{m}$  steps, both selected to replicate Wortman *et al.* (15) Voronoi tessellation was used to replicate spatial interactions as defined in Patwa *et al.* (18). Ripley's L (a density-normalized measure of clustering) and the multitype K function ( $K_{\text{cross}}$ ; a density-normalized measure of two cell types co-localization) and G function ( $G_{\text{cross}}$ , a measure of two cell types co-localization) were calculated using spatstat (23) with a radius of 50  $\mu\text{m}$ . Spatial latent Dirichlet allocation (LDA) analysis was done using spatialLDA (24), using the default radius of 100  $\mu\text{m}$  and 8 topics. Shorter distances of  $\sim 25$   $\mu\text{m}$  may be interpreted as cells nearly or directly touching, while 100  $\mu\text{m}$  represents a distance at which oxygen, nutrients and potentially other molecules diffuse in tissues (49). Survival analysis was done as described above, using the CycIF TMA1 dataset as a discovery cohort and the other datasets as the validation cohort. For previously published biomarkers, the validation cohort included patients not used in developing those biomarkers. Specifically, the tumor-immune mixing score, developed using the MIBI cohort, was

validated with the CycIF and IMC cohorts. Immunoregulatory interactions, also developed using the MIBI dataset, were validated in the CycIF cohort only, since the IMC panel lacked immunoregulatory markers. Lymphocyte clustering, lymphocyte occupancy and heterotypic neighbor biomarkers were initially developed in external cohorts, so all samples were included in the validation cohort (and no discovery cohort was used).

### *Statistics*

For comparison of continuous variables, two-sided Pearson correlation was used. For categorical correlations, two-sided Mann-Whitney U test (two groups) or Kruskal-Wallis H-test (three or more groups) were used, with Tukey HSD correction for pairwise comparisons between groups. Two-way Chi-squared analysis was used for categorical data, with Bonferroni correction for pairwise comparisons between groups. Survival curves were estimated with Kaplan-Meier analysis and the log-rank test was used to calculate significance. Cox proportional hazard modeling was used to estimate hazard ratios and p-values in multivariate survival models. All p-values < 0.05 were considered significant. When multiple variables were tested for survival, Benjamini-Hochberg correction was applied in the validation cohort. False-discover rate (FDR) <0.05 was considered significant, except with spatial LDA neighborhoods, where FDR <0.1 was considered significant.

### *Study approval*

Samples were obtained with written informed consent; sample collection complied with all relevant ethical regulations and was approved under Vanderbilt IRB protocol VICC BRE03103. Use of human samples for our study at Oregon Health and Science University was approved under IRB protocol STUDY00016712.

### *Data availability*

Spatial and survival analysis code; CycIF, MIBI and IMC processing pipelines, single-cell phenotyping pipelines and precomputed data are available here: [https://github.com/engjen/cycIF\\_TMAs](https://github.com/engjen/cycIF_TMAs). Supporting data values and antibody information is available in our Source\_Data.xlsx file. Images and data files larger than 10 MB are available here: <https://www.synapse.org/#!/Synapse:syn50134757/>

## AUTHOR CONTRIBUTIONS

J.R.E., S.L.G., and K.C. conceived of the project. M.S., B.C., P.G., and J.A.P. collected patient tissues, clinical outcome data, and constructed the TMAs. Z.H. performed staining and imaging experiments. J.R.E. and E.B. did single-cell image processing. C.R.W., T.R., and M.A., generated MIBI data on normal tissues. J.R.E. ran the analysis, drafted the manuscript, and prepared the figures. S.L.G., K.C., E.B., B.C., R.C.S., and J.A.P. edited the manuscript.

## ACKNOWLEDGEMENTS

We would like to thank Dr. Joe Gray for invaluable advice on analysis and feedback on the manuscript. We would also like to thank Drs. Sandra Rugonyi, Sara Courtneidge, Young Hwan Chang, and Pepper Schedin for feedback on earlier versions of this work. We would like to thank Damir Sudar for technical support related to image storage and Elliot Gray for sharing his neighbor counter script. We sincerely appreciate Drs. Clare Yu, Peter Lee and Julie Wortman for sharing their lymphocyte occupancy code. We appreciate the sample scanning assistance of both Drs. Stefanie Kaech Petrie and Crystal Chaw at the OHSU Advanced Light Microscopy Core. This work was supported by funding from NIH/NCI U54 CA209988, the Prospect Creek Foundation, and OHSU Foundation. We received the services from the Knight Cancer institute Histopathology Shared Resource and Advanced Light Microscopy Core, supported by the Cancer Center Support Grant (NIH/NCI P30CA69533). Additionally, J.A.P acknowledges support from NIH/NCI P50 CA098131 and Komen SAB210301. The normal tissue TMA was created with support from the Brenden Colson Center for Pancreatic Care.

## REFERENCES

1. Rizzo, A. & Ricci, A. D., Biomarkers for breast cancer immunotherapy: PD-L1, TILs, and beyond. *Expert Opin Investig Drugs* 1–7 (2021)
2. Schmid, P., *et al.* Pembrolizumab for Early Triple-Negative Breast Cancer. *New England Journal of Medicine* 382, 810–821 (2020).
3. Tsujikawa, T., *et al.* Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Rep* 19, 203–217 (2017).
4. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science (1979)* 361, (2018).
5. Lin, J. R., *et al.* Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat Commun* 6, 1–7 (2015).
6. Goltsev, Y., *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* 174, 968-981.e15 (2018).
7. Giesen, C., *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* 11, 417–422 (2014).
8. Eng, J., *et al.* Cyclic Multiplexed-Immunofluorescence (cmIF), a Highly Multiplexed Method for Single-Cell Analysis. in *Biomarkers for Immunotherapy of Cancer. Methods in Molecular Biology* (eds. Thurin, M., Cesano, A. & Marincola, F. M.) vol. 2055 521–562 (Humana, New York, NY, 2020).
9. Cabrita, R., *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* 577, 561–565 (2020).
10. Helmink, B. A., *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* 577, 549–555 (2020).
11. Janiszewska, M., *et al.* The impact of tumor epithelial and microenvironmental heterogeneity on treatment responses in HER2+ breast cancer. *JCI Insight* 6, (2021).
12. Keren, L., *et al.* A Structure Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam imaging. *Cell* 174, 1373-1387.e19 (2018).
13. Ali, H. R., *et al.* Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nat Cancer* 1, 163–175 (2020).
14. Jackson, H. W., *et al.* The single-cell pathology landscape of breast cancer. *Nature* (2019)
15. Wortman, J. C., *et al.* Spatial distribution of B cells and lymphocyte clusters as a predictor of triple-negative breast cancer outcome. *NPJ Breast Cancer* 7, (2021).
16. Parmar, C., *et al.* Data Analysis Strategies in Medical Imaging. *Clinical Cancer Research* 24, 3492–3499 (2018).
17. Eng, J., *et al.* A framework for multiplex imaging optimization and reproducible analysis. *Communications Biology* 2022 5:1 5, 1–11 (2022).
18. Patwa, A., *et al.* Multiplexed imaging analysis of the tumor-immune microenvironment reveals predictors of outcome in triple-negative breast cancer. *Commun Biol* 4, 1–14 (2021).
19. Wolf, A., Angerer, P. & Theis, F. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19 (2018).
20. Pedregosa, F., *et al.* Scikit-learn: Machine Learning in Python Fabian. *Journal of Machine Learning Research* 12, 2825–2830 (2011).

21. Trinh, A., *et al.* Genomic Alterations during the In Situ to Invasive Ductal Breast Carcinoma Transition Shaped by the Immune System. *Mol Cancer Res.* 19(4):623-635 (2021).
22. Goldberg, J., *et al.* The Immunology of Hormone Receptor Positive Breast Cancer. *Front Immunol* 12, 1–22 (2021).
23. Baddeley, A., *et al.* *Spatial Point Patterns: Methodology and Applications with R.* (Chapman and Hall/CRC Press, London, 2015).
24. Chen, Z., *et al.* Modeling Multiplexed Images with Spatial-LDA Reveals Novel Tissue Microenvironments. *Journal of Computational Biology* 27, 1204–1218 (2020).
25. Nirmal, A. J., *et al.* The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single-Cell Resolution. *Cancer Discov* 12, 1518–1541 (2022).
26. Perou, C. M., *et al.* Molecular portraits of human breast tumours. *Nature* 406, 747–752 (2000).
27. Lehmann, B. D., *et al.* Refinement of triple-negative breast cancer molecular subtypes: Implications for neoadjuvant chemotherapy selection. *PLoS One* 11, (2016).
28. Fantozzi, A., *et al.* VEGF-mediated angiogenesis links EMT-induced cancer stemness to tumor initiation. *Cancer Res* 74, 1566–1575 (2014).
29. Denkert, C., *et al.* Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol* 19, 40–50 (2018).
30. Nagalla, S., *et al.* Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome Biol* 14, (2013).
31. Miller, L. D., *et al.* Immunogenic Subtypes of Breast Cancer Delineated by Gene Classifiers of Immune Responsiveness. *Cancer Immunol Res* 4, 600–611 (2016).
32. Sato, M., *et al.* An integrated genomic approach identifies persistent tumor suppressive effects of transforming growth factor- $\beta$  in human breast cancer. *Breast Cancer Research* 16, 1–23 (2014).
33. Li, S., *et al.* Cancer immunotherapy via targeted TGF- $\beta$  signalling blockade in TH cells. *Nature* 587, 121–125 (2020).
34. Thomas, A. *et al.* Tumor mutational burden is a determinant of immune-mediated survival in breast cancer. *Oncoimmunology* 7, 1–12 (2018).
35. Macheret, M. & Halazonetis, T. D. DNA replication stress as a hallmark of cancer. *Annual Review of Pathology: Mechanisms of Disease* 10, 425–448 (2015).
36. Downey, C. L. *et al.* The prognostic significance of tumour-stroma ratio in oestrogen receptor-positive breast cancer. *Br J Cancer* 110, 1744–1747 (2014).
37. Leek, R. D. *et al.* Association of macrophage infiltration with angiogenesis and prognosis in invasive breast carcinoma. *Cancer Res* 56, 4625–4629 (1996).
38. Medrek, C., *et al.* The presence of tumor associated macrophages in tumor stroma as a prognostic marker for breast cancer patients. *BMC Cancer* 12, 1–9 (2012).
39. Rubinstein-Achiasaf, L., *et al.* Persistent Inflammatory Stimulation Drives the Conversion of MSCs to Inflammatory CAFs That Promote Pro-Metastatic Characteristics in Breast Cancer Cells. *Cancers* 2021, Vol. 13, Page 1472 13, 1472 (2021).
40. Liubomirski, Y., *et al.* Tumor-stroma-inflammation networks promote pro-metastatic chemokines and aggressiveness characteristics in triple-negative breast cancer. *Front Immunol* 10, 757 (2019).

41. Lee, A. T. J. *et al.* The adequacy of tissue microarrays in the assessment of inter- and intra-tumoural heterogeneity of infiltrating lymphocyte burden in leiomyosarcoma. *Sci Rep* 9, 14602 (2019).
42. Larønningen S., *et al.* NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 9.3 (02.10.2023). Association of the Nordic Cancer Registries. Cancer Registry of Norway. Available from: <https://nordcan.iarc.fr/> accessed on 11/14/2024.
43. Stringer, C., *et al.* Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* 18, 100–106 (2021).
44. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17, 261–272 (2020).
45. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol* 40, (2021).
46. Danenberg, E. *et al.* Breast tumor microenvironment structures are associated with genomic features and clinical outcome. *Nat Genet* 18, (2022).
47. napari contributors. napari: a multi-dimensional image viewer for python. Preprint at (2019).
48. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 9, 1–12 (2019).
49. Hanahan, D. and Weinberg, R. A. The Hallmarks of Cancer. *Cell* 107, 251–262 (2000).

## FIGURE LEGENDS

Figure 1. Concordant cell phenotypes in multiplex imaging data from different platforms.

**A.** Three multiplex imaging datasets from breast cancer tissue microarrays were processed through single cell segmentation and feature extraction using the mplexable pipeline. The single cell datasets were separately clustered using the unsupervised Leiden algorithm resulting in cell types which were annotated with similar names across platforms. We generated a suite of spatial statistics for each tissue and the resulting cellular and spatial features were used for discovery and validation of prognostic cell abundance and spatial biomarkers across datasets. CycIF, cyclic immunofluorescence; IMC, imaging mass cytometry; MIBI, multiplex ion beam imaging. **B.** Overlap of markers (left) and annotated cell types (right) in each multiplex imaging dataset. **C.** Representative images from the three multiplex imaging platforms showing epithelial (orange), immune (red) and fibroblast (green) markers. Scale bar = 100  $\mu$ m. Total of 413 patient tissues imaged. **D.** Cell lineage types showing cell location and lineages: epithelial (orange), immune (red), fibroblast (green), endothelial (blue) and other stromal (purple) CycIF, top; IMC, middle; MIBI, bottom. **E.** Correlation between platforms of fraction of each cell lineage per total cells per subtype, per platform, using unsupervised clustering and annotation to determine lineage. N number of patients for each subtype and platform shown in panel title. **F.** Correlation between cell types on adjacent sections of a TMA stained with MIBI and CycIF. N=9 tissues. **E-F.** Pearson's correlation  $r$  (two-sided) between platforms and p-value shown in panel title.

Figure 2. Prognostic ER+ breast cancer subtypes in multiplatform multiplex imaging data.

**A.** Hierarchical clustering of all ER+ and TNBC patients (n=350) based on the z-scored fraction in each patient's tissue of the six most common epithelial cell types. Heatmap annotation row colors show the Leiden clustering resulting in seven epithelial (Epithel.) subtype clusters (left), clinical subtype (center) and platform (right). **B.** Mean cell frequency of epithelial cell types per subtype cluster. **C.** UMAP embedding of patients by fraction of epithelial cell types in all tumor cells, colored by Leiden epithelial subtype cluster (top) and platform (bottom). N=350 patients. **D.** Two-sided Chi-squared analysis of epithelial subtypes versus platform, p-values in panel title. **E.** Kaplan-Meier curves (p-value from log-rank test) comparing overall survival (OS) in the seven epithelial subtypes present in ER+ tumors. **F.** Cox proportional hazard (CPH) model estimating hazard ratios for epithelial subtypes of ER+ tumors. The hazard ratio estimates marked by boxes and error bars are the 95% confidence intervals. **G.** CD44 intensity in epithelial cells from poor prognosis epithelial subtype 6 compared to other ER+ patients. FDR corrected for multiple cell markers given in panel title; p-values from Mann-Whitney U rank test. **H.** Fraction of endothelial cells in tissue stromal cells of patient tissues from each epithelial subtype cluster. Kruskal-Wallis H-test P-value given in panel title. Post-hoc Tukey HSD p-values for pairwise comparisons between groups. **D-H.** N number of patients given in panel title. **G-H.** Boxplots show the median and interquartile range (IQR) and whiskers 1.5x the IQR.

Figure 3. T cell infiltrate has prognostic value and distinct states in TN and high proliferation ER+ breast cancer.

**A.** Kaplan-Meier analysis of abundance of CD3 T cells versus overall survival (OS) in TNBC discovery (left) and validation cohort (right). **B.** Kaplan-Meier analysis of abundance of CD20 B cells versus OS in TNBC discovery (left) and validation cohort (right). **C.** Multivariate CPH modeling adding patient age, tumor size and stage to CD3 T cell high variable defined in (A). **D.** Multivariate CPH modeling adding patient age, tumor size and stage to CD20 B cell high variable defined in (B). **E.** Kaplan-Meier analysis of abundance of CD3 T cell versus OS in all ER+ patients (left) and ER+ patients with high (above the median) tumor proliferation (right). **F.** CPH modeling of CD3 T cell abundance plus clinical variables in high and low proliferation ER+ tumors. **G.** Kaplan-Meier analysis of all ER+ and TNBC patients stratified into four groups by median tumor proliferation and median T cell abundance. **H.** CPH modeling of CD3 T cell abundance plus clinical variables in high and low proliferation TNBC tumors. **I.** Mean number of T cell neighbors (within 25  $\mu$ m) of T cells in tissues from high and low proliferation ER+ or TNBC tumors in IMC cohort. **J.** Ki67 intensity indicating proliferation levels of T cells in tissues from high and low proliferation ER+ or TNBC tumors in IMC cohort. **K.** CD44 intensity in T cells, indicating memory/effector phenotypes in IMC tissues. **A-H.** All Kaplan-Meier p-values obtained from the log-rank test, validation cohort corrected for testing multiple cell types with Benjamini-Hochberg method. CPH modelling p-values for cell type variable given in panel titles; the hazard ratio estimates marked by boxes and error bars are the 95% confidence intervals. N number of patients given in panel titles. **I-K.** Kruskal-Wallis H-test P-value and N number of patients



given in panel title. Post-hoc Tukey HSD used for pairwise comparisons between groups. Boxplots show the median and interquartile range (IQR) and whiskers 1.5x the IQR.

#### Figure 4. Reproducible prognostic spatial metrics in breast cancer cohorts.

**A.** Spatial locations of cells in a TMA core with each cell colored by number of CD3 T cell neighbors in a 40  $\mu\text{m}$  radius. Tumor cells shown in orange. **B-C.** Kaplan-Meier (K-M) estimates of B) recurrence free survival (RFS) of ER+ patients with high versus low stromal neighbors of epithelial cells or C) overall survival (OS) of TNBC patients with high versus low immune neighbors of immune cells in a 40  $\mu\text{m}$  radius in the discovery (left) and validation (right) cohorts. **D.** Multivariate CPH modeling of the effect of patient age, tumor size and stage with (left) ER+ patient stromal neighbors of epithelial cells from (B) on RFS or (right) TNBC patient immune neighbors of immune cells (from C) on OS. **E & G.** K-M estimate of RFS of TNBC patients with high versus low macrophage neighbors of tumor (E) or B cell neighbors of T cells (G) in a 40  $\mu\text{m}$  radius in the discovery (left) and validation (right) cohorts. **F & H.** Multivariate CPH modeling of TNBC patient RFS versus macrophage neighbors of tumor (F) or CD20 B cell neighbors of CD3 T cells (H). **I.** K-M estimate of TNBC OS stratified by tumor immune mixing score developed by Keren et al. in the MIBI cohort (top) and validation cohort (i.e. CycIF and IMC; bottom). **J.** K-M estimate of TNBC OS stratified by occupancy AUC of lymphocytes in quadrants with length 10 – 300  $\mu\text{m}$  of T (left) or B lymphocytes (right) within 20  $\mu\text{m}$  of a tumor cell. **K.** Multivariable CPH modeling of T (left) and B lymphocyte (right) occupancy AUC with clinical co-variables. **L.** K-M estimate of OS in TNBC stratified by fractal dimension slope difference of large (<100  $\mu\text{m}$ ) minus small (10-40  $\mu\text{m}$ ) quadrants for T (top) or B lymphocytes (bottom) within 20  $\mu\text{m}$  of a tumor cell. **M.** Top: Representative image of CycIF core showing DAPI-stained nuclei (blue) and PD-1 staining (red), scale bar = 130  $\mu\text{m}$ . Bottom: Voronoi tessellation of tissue from top, with all cells (blue), PD1+ cells (red) and interactions (black line). **N.** Top: K-M estimate of high (>1 interaction in tissue) versus low PD1 interactions in CycIF TNBC cohort. Bottom: CPH modeling of PD-1 interaction metric (originally developed by Patwa et al. in the MIBI cohort). **A-N.** All Kaplan-Meier p-values from log-rank test; validation cohort FDR corrected for multiple metrics tested with the Benjamini-Hochberg method. CPH modelling p-values for spatial variable given in panel titles; the hazard ratio estimates marked by boxes and error bars are the 95% confidence intervals. N number of patients given in panel titles. **I-L.** Metrics developed by Wortman *et al.* and higher values indicate spatial dispersion.

#### Figure 5. Prognostic multicellular neighborhoods surrounding tumor cells modeled with spatial latent Dirichlet allocation.

**A.** Top: CycIF staining of TNBC tissue showing tumor (panCK), T cell (CD4 and CD8), macrophage (CD68), fibroblast (vimentin), and endothelial (CD31) markers. Middle: Tumor cells colored by topic weights of select topics. Bottom: Tumor cells colored by their spatial latent Dirichlet allocation (LDA) neighborhood cluster. Tumor cells colored by T cell- (blue), macrophage- (purple), mixed fibroblast- (brown) and vimentin+ fibroblast- neighborhoods (green). N=308 patients analyzed with spatial LDA. **B, D.** Heatmap of stromal cell enrichment in spatial LDA topics in a 100  $\mu\text{m}$  radius of tumor cells in CycIF B) TNBC tissues (n=59) and D) ER+ tissues (n=30). **C, E.** Heatmap of fraction of each topic in each neighborhood cluster resulting from K-means clustering (k=8) of spatial LDA topics from C) TNBS and E) ER+ tissues. **F.** Kaplan-Meier (K-M) estimate of overall survival (OS) for high and low vimentin+ fibroblast tumor neighborhoods in TNBC tissues in discovery (left) and validation cohorts (right). **G.** CPH modelling of OS and recurrence-free survival (RFS) with clinical variables plus spatial LDA neighborhood from (F). **H.** K-M analysis of OS for high and low mixed fibroblast tumor neighborhoods in ER+ tissues in the discovery (left) and validation cohorts (right). **I.** CPH modeling of OS and RFS for mixed fibroblast neighborhoods in ER+ tumors. **J-K.** Pearson correlation of cell types within J) ER+ and K) TNBC tissues from all cohorts, colored by cohort (legend in L). Pearson correlation of neighborhood/cell type abundances, subtype in panel title, colored by cohort. **F, H.** p-values obtained from the log-rank test. **G, I.** CPH modelling p-values for spatial variable in panel title; the hazard ratio estimates marked by boxes and error bars are the 95% confidence intervals. **J-L.** Cell types, cohort, and two-sided Pearson correlation (r) and p-values given in panel titles. **F-L.** N number of patients given in panel titles.

Figure 1

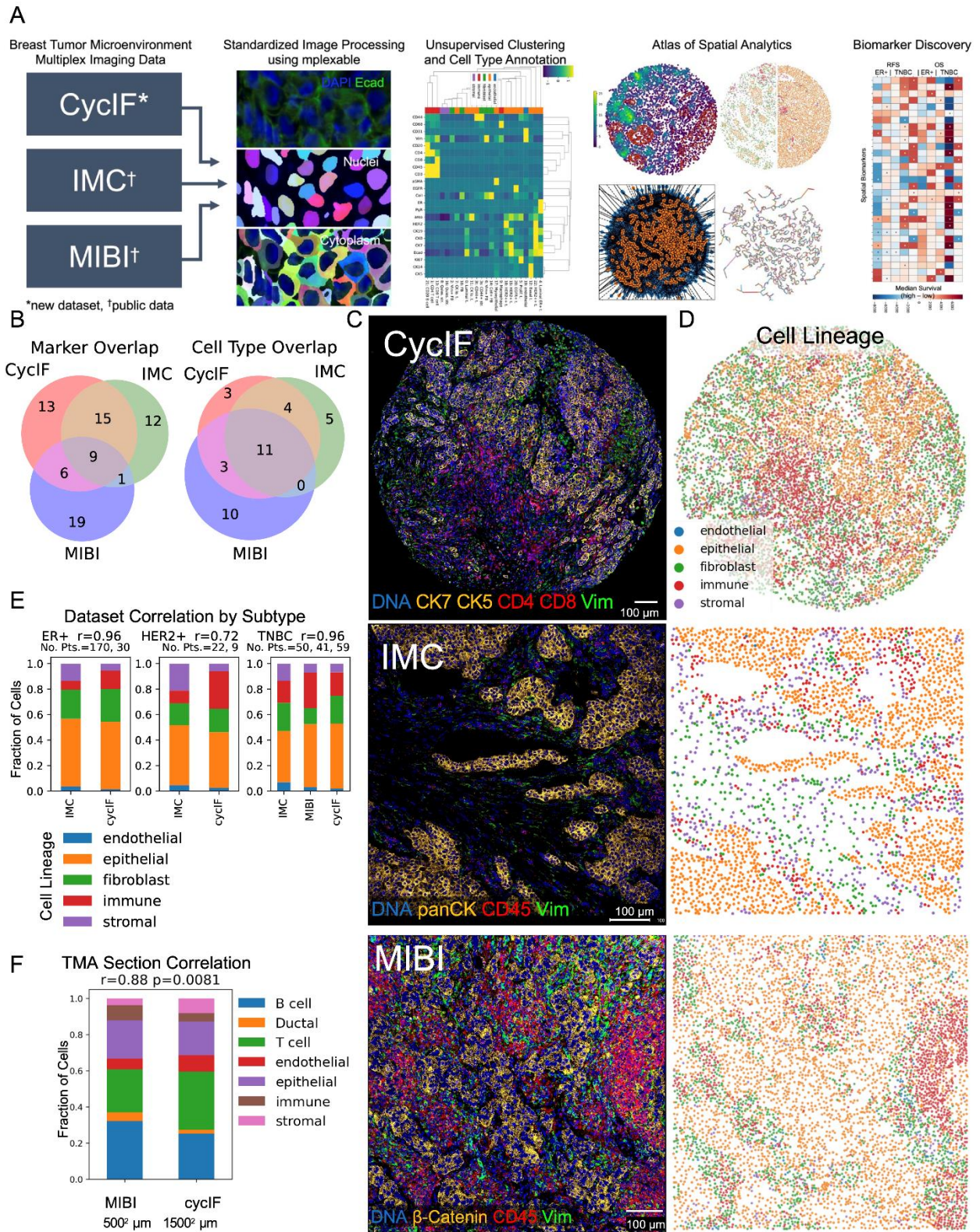


Figure 2

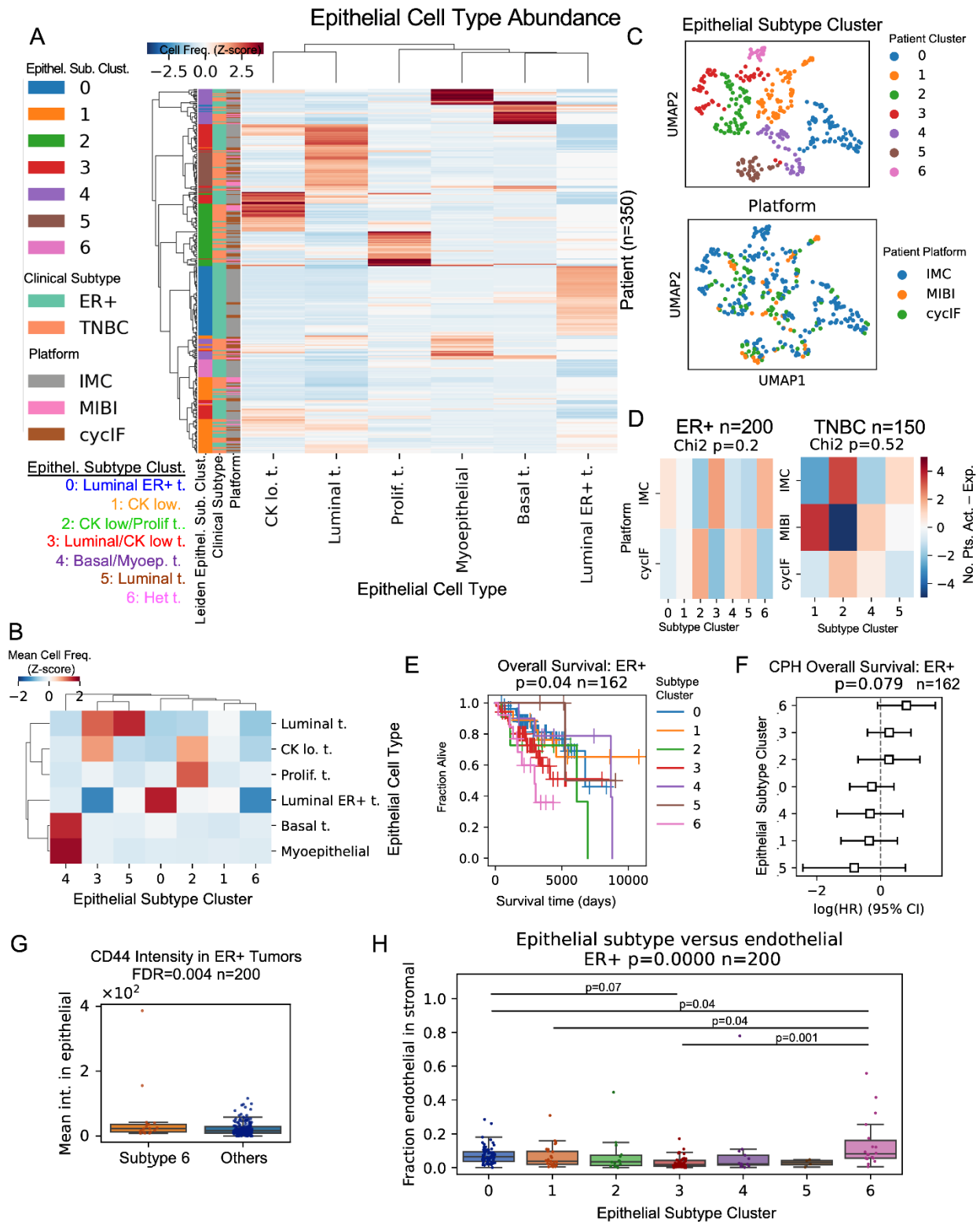




Figure 3

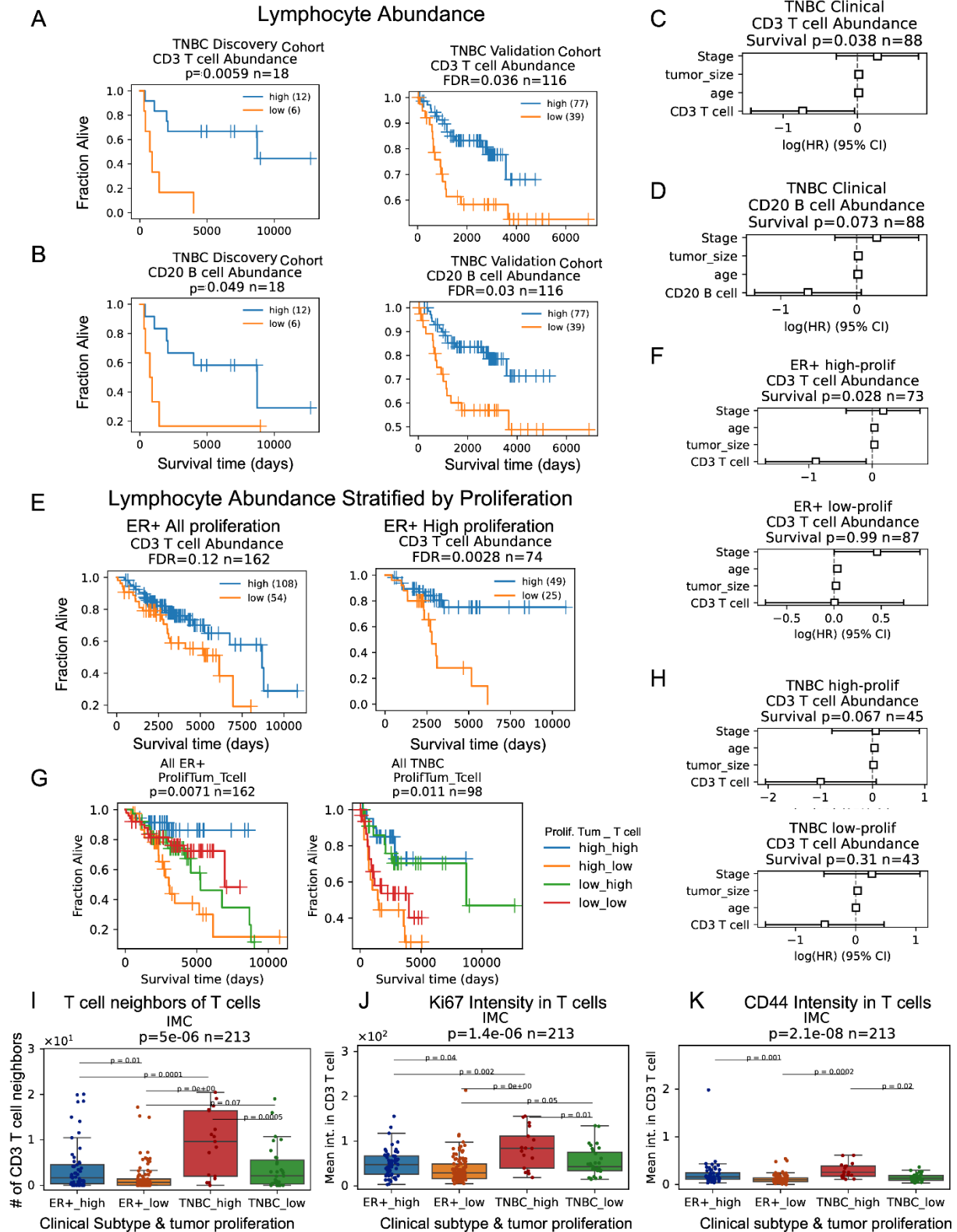


Figure 4

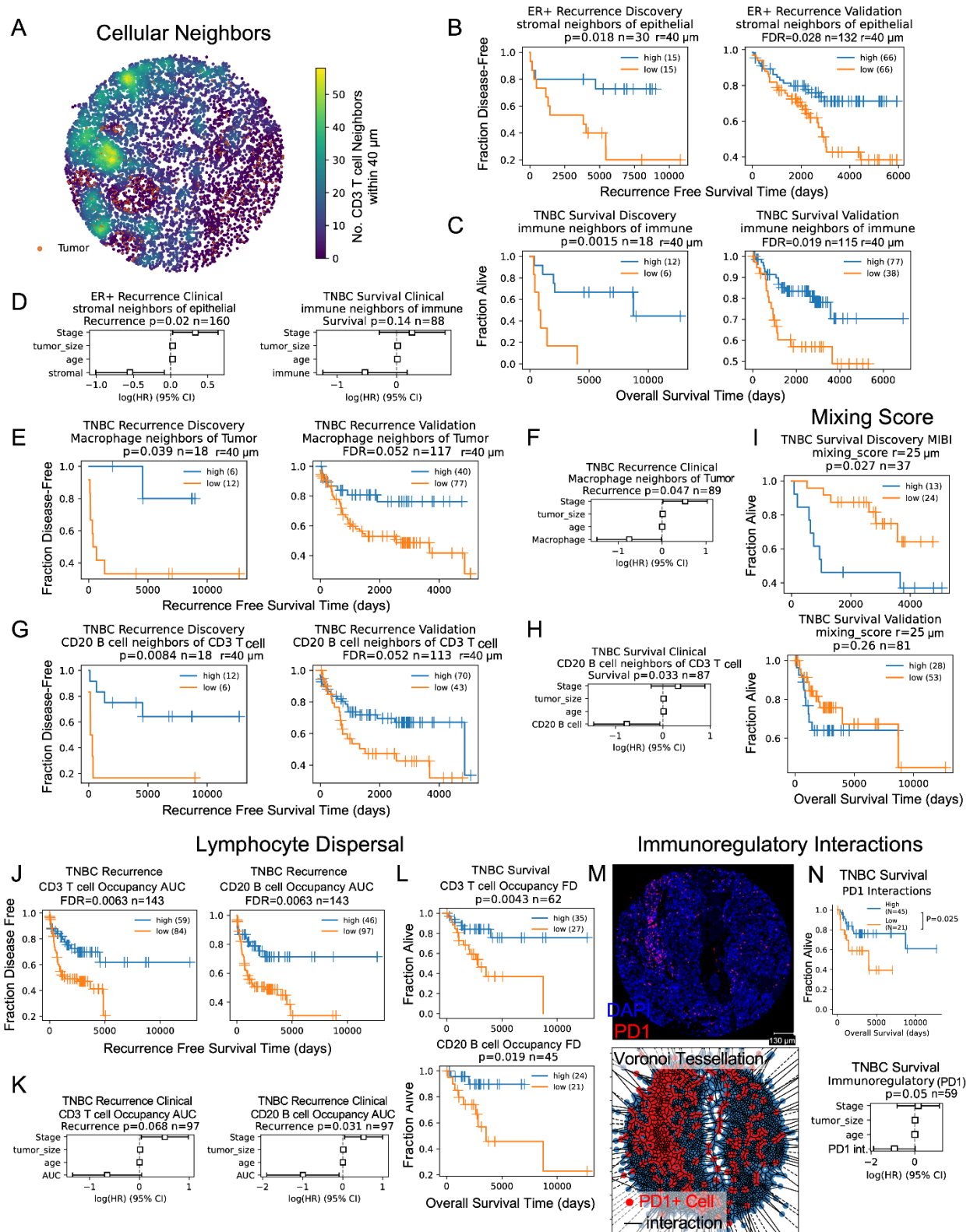


Figure 5

