Robust biomarker discovery through multiplatform multiplex image analysis of breast cancer clinical cohorts

Jennifer Eng[1,2,7], Elmar Bucher[2], Zhi Hu[2], Melinda Sanders[3,4], Bapsi Chakravarthy[3], Paula Gonzalez[4], Jennifer A. Pietenpol[3,5], Summer L. Gibbs[2,7], Koei Chin[2,6,7].

AFFILIATIONS

1. Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR, 97239, USA.
2. Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR, 97239, USA.
3. Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, 37232, USA, USA.
4. Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA.
5. Department of Biochemistry, Vanderbilt University, Nashville, TN, 37232, USA.
6. Center for Early Detection Advanced Research, Oregon Health and Science University, Portland, OR, 97239, USA.
7. Knight Cancer Institute, Oregon Health and Science University, Portland, OR 97239, USA

ABSTRACT

Spatial profiling of tissues promises to elucidate tumor-microenvironment interactions and enable development of spatial biomarkers to predict patient response to immunotherapy and other therapeutics. However, spatial biomarker discovery is often carried out on a single patient cohort or imaging technology, limiting statistical power and increasing the likelihood of technical artifacts. In order to analyze patient cohorts profiled on different platforms, we developed methods for comparative data analysis from three disparate multiplex imaging technologies: 1) cyclic immunofluorescence data we generated from 102 breast cancer patients, 63 with clinical follow-up, in addition to publicly available 2) imaging mass cytometry and 3) multiplex ion-beam imaging data. We demonstrate similar single-cell phenotyping results across breast cancer patient cohorts imaged with these three methods and identify cellular abundance and proximity-based biomarkers with prognostic value across platforms. In multiple platforms, we identified T cell infiltration as independently associated with longer survival in high-proliferation breast tumors. Functional marker analysis of T cells in these tissues showed enrichment for activated and spatially clustered T cells. A comparison of six spatial analysis methods revealed robust spatial biomarkers, including tumor-macrophage and tumor-fibroblast proximity associated with poor prognosis in estrogen receptor-positive and triple-negative tumors, respectively. Our methods enable assembly of larger clinical cohorts from diverse platforms to aid in prognostic spatial biomarker identification and validation.

STATEMENT OF SIGNIFICANCE

Our single-cell spatial analysis of multiple clinical cohorts uncovered novel biomarkers of patient outcome in breast cancer. Additionally, our data, software, and methods will help advance spatial characterization of the tumor microenvironment.

INTRODUCTION

Recent advances in the treatment landscape of breast cancer, including the approval of immune checkpoint inhibitors for early TNBC[1], have motivated the characterization of the breast tumor microenvironment for deeper understanding of immune-tumor interactions. Identification of biomarkers predicting breast cancer immunotherapy response is still an urgent clinical need[2]. In metastatic TNBC, only a quarter of PD-L1 positive patients respond to single-agent immune checkpoint blockade[2]. In contrast, in early stage TNBC, response rates to neoadjuvant immune checkpoint plus chemotherapy were similar in PD-L1 positive and negative groups[3]. This highlights the need for biomarker development for better patient stratification across disease stages. Furthermore, trials evaluating novel immune targeted therapies in breast cancer should be accompanied by biomarker development for maximum therapeutic efficacy[4].

Highly multiplex imaging methods enable quantification of dozens of biomarkers in a single tissue section with at sub-cellular resolution while retaining spatial context[5–10]. Tissue structures such as tertiary lymphoid structures, identified with multiplex imaging, are predictive biomarkers of immunotherapy response in melanoma[11,12]. Recently, spatial biomarkers involving multiple cell types have been identified that predict response to immune checkpoint inhibitors in cutaneous T cell lymphoma[13] and patient survival pancreatic ductal adenocarcinoma[14]. Employment of multiplex imaging methods in breast cancer has revealed that single-cell spatial context has prognostic relevance and shows correlations with transcriptomic and genomic features of tumors[15–17]. However, different imaging platforms can introduce method-dependent technical and computational idiosyncrasies. Identification of biomarkers in additional platforms can strengthen biological and clinical findings. Furthermore, due to the number of steps in the processing pipeline and the number of variables and parameters involved, overfitting is an issue in biomedical imaging data. Overfitting can be addressed through use of a training cohort to tune analytical methods, which are then fixed and subsequently applied to a validation cohort[18]. Such validation cohorts can be obtained through incorporation of data from additional imaging platforms into biomarker studies. Documentation of metadata, analysis protocols and code is essential for reuse of data and reproducibility of findings, preferably using open-source software tools[18]. We developed an open-source python software, mplexable[19], 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 generate, and publicly available imaging mass cytometry (IMC) and multiplex ion-beam imaging (MIBI) datasets [15,17,20]. In this proof-of-concept study, we identify prognostic single-cell spatial biomarkers common across the imaging platforms. As such imaging datasets become more widely available, tools such as ours can facilitate biomarker discovery with high accuracy, reliability, and efficiency.

METHODS

Patient samples

Breast cancer tissue microarrays were graciously provided by Dr. Jennifer Pietenpol (JP). All samples were collected at time of surgical resection (mastectomy or breast conserving surgery) at Vanderbilt University Medical Center with the same fixation protocol. JP-TMA1 had 131 cores of

approximately 1.2 µm diameter, with duplicate cores from 19 TNBC, 8 HER2+ and 36 ER+ patients. It also included one control core each of inflamed appendix, colon cancer, muscle, pancreas and normal breast. Four of the TMA1 TNBC patients received neoadjuvant chemotherapy. Clinical outcome and clinicopathological information were available for TMA1. JP-TMA2 contained a single, slightly larger (~1.4µm diameter) core from 39 triple-negative tumors and 1 ER+/HER2+ control core. Thirteen of the patients in TMA2 had received neoadjuvant therapy. Clinical outcome data were not yet available for TMA2.

Imaging data generation and sources

CyCIF staining of tumor tissue was completed on JP-TMA1 and TMA2. The whole tissue core was imaged, as described previously[19]. MIBI imaging data was previously published by Keren *et al*[15] 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[21], and were downloaded from https://github.com/aalokpatwa/rasp-mibi. IMC imaging data were previously published by Jackson *et al*.[17] 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 previously described[19]. The CyCIF pipeline is available at https://gitlab.com/engje/mplexable/-/tree/master/jupyter. Nuclear and cell segmentation were conducted using the Cellpose algorithm, which showed visually superior performance on CyCIF data compared to a watershed algorithm (S 1a). Nuclear and cell segmentation masks were matched using mplexable, enabling subtraction of nuclear mask from cell mask to obtain segmentation of the cytoplasm (S 1b).

MIBI and IMC processing pipelines are available as Jupyter notebooks at cycIF_TMAs (https://codeocean.com/capsule/7863588/tree). MIBI and IMC images were downloaded as multipage OMEtiffs. Hot pixels[16] 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. DNA channels were processed for nuclear segmentation as follows. 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. A mixture of markers was used to generate a composite image which stained the cytoplasm of the majority of cells. For MIBI data, the β-catenin, vimentin, CD45 and CD31 channels were combined into a maximum intensity projection cytoplasm image and the gamma value was adjusted by 0.6. For IMC data, E-cadherin, vimentin, CD44 and CD45 were combined into a maximum intensity projection cytoplasm image and gamma adjusted by 0.4. Chamboelle total variation denoising, implemented in scipy[22], 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 (S 1c-d). Skipping either nuclear or cytoplasmic

Chamboelle total-variation de-noising resulted in failure of deep learning-based algorithms on the IMC data (S 1e). Mesmer segmentation[23] performed better than Cellpose[24] in IMC data due to improved detection of dim nuclei, likely due to the incorporation of cytoplasmic staining in the nuclear segmentation model (S 1d). This issue with Cellpose was not detected in CyCIF images, which had brighter DNA staining. For IMC and MIBI data, nuclear and cellular segmentation were performed on processed nuclear and nuclear + cytoplasm images using Mesmer[23]. Matching of cell IDs in the nuclear and cell masks was done with mplexable[19], with cell masks relabeled to match the ID of the nucleus which they had most overlap with. 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 scikit-image[25].

Mesmer, a deep-learning-based approach for nuclear and whole-cell segmentation of tissue data, was run on IMC data and compared to the watershed-based segmentation originally published by Jackson *et al*[17]. The cell counts across the two methods had a Pearson correlation of 0.98 (S 2 a). 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 (S 2 b-d).

Image QC

In IMC data, artifacts include non-specific background staining, necrotic regions, and bright antibody aggregates. IMC data were collected from small ROIs (~600 μm) within TMA cores and some samples annotated as ER+ tissues did not show any ER+ staining in the ROI. Therefore, quality control (QC) was performed on ER stained images, a marker which has been noted to exhibit non-specific background staining on the IMC platform[26]. QC images of ER staining were generated and sorted in a blinded fashion into negative and positive for nuclear-specific-staining (S 3a). ROIs from clinically annotated ER+ patients that were classified as ER positive during QC or ROIs that came from ER negative patients and classified as ER negative were used for analysis. Samples that passed ER QC did not have significantly different grade, PR status, TMA block, age of specimen, age of patient or tumor size (S 3c-d). There were no significant survival differences between QC passed versus failed tumors from ER+, TNBC or ER+HER2+ patients (S 3e). In the IMC dataset, additional QC steps included: necrotic regions were manually circled using the napari[27] image viewer and excluded and bright aggregates in the CD3 channel were excluded by removing cells above a threshold set at the value 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 half-way through CyCIF data collection and these images were scaled by exposure time and subtracted from the AF488, AF55 and AF647 channels using mplexable[19]. Feature extraction was performed on AF subtracted images. Areas of floating tissue, air bubbles or necrotic regions were manually circled using the napari[27] image viewer and excluded. Non-specific background staining was removed by setting manual 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[28] software. Single-cell mean intensity values were selected from either the nucleus or cytoplasm masks for each marker, depending on expected intracellular distribution. β-catenin, which can localize on the membrane, in the cytoplasm or in the nucleus, was selected from both nuclear and cytoplasmic masks for separate analysis. Since the CyCIF and IMC platforms had more marker and breast cancer subtype overlap than the MIBI panel (**Figure 1**, **Figure 2**) 20 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 20 standard deviations. A Uniform Manifold Approximation and Projection (UMAP) embedding was generated using 30 k-nearest neighbors and the embedding was clustered using the Leiden community detection algorithm[29]. 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 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 positive pixel patterns in images. Fibroblasts were positive for one or more of vimentin, fibronectin (FN) or collagen I (ColI). Epithelial cells were positive for one or more of Ecad, cytokeratins, or β-catenin. Endothelial cells were defined as CD31+. Immune cells were CD45+. Stromal cells were all non-fibroblast, non-endothelial, non-epithelial, non-immune segmented nuclei. Jupyter notebooks recording single cell phenotyping pipelines are available here: cycIF_TMAs (https://codeocean.com/capsule/7863588/tree).

Statistical Analysis

The CyCIF dataset was used as a training dataset to determine the quantile separating high and low abundance of each cell type that was most predictive of survival. For analysis of single variable association with recurrence-free (RFS) or overall survival (OS), each cell type was tested (log-rank test) for significant association with RFS and OS in CyCIF TNBC, ER+ and HER2+ samples. For each cell type, fraction of cells of that type over all cells from combined ROIs/cores from each patient was calculated and binarized into high and low abundance based on tertile or median values within each subtype. 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 two additional

datasets. Since overall cell type fractions differed between platforms and subtypes (**Figure 3**), high and low values were determined relative to other samples from the same platform and subtype. High/low binarized abundances were combined for survival analysis on the whole cohort (n=98 TNBC and n=163 ER+). Kaplan-Meier analysis was used to estimate survival functions and the log-rank test was used to assess significant differences in survival. Multivariate Cox proportional hazards (CPH) modelling was used to combine cell abundance variables with patient age, tumor size and clinical stage. Collectively, 55 TNBC and 160 ER+ patients had these additional clinical parameters (additional clinical parameters were not included in the MIBI dataset).

Epithelial and stromal subtypes were determined by hierarchical 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, platform specific cell types that may represent method-specific artifacts. Unsupervised hierarchical clustering of patients was performed using the AgglomerativeClustering function in scikit-learn[30], using a Euclidean distance metric and complete linkage. The number of clusters/subtypes were selected as follows. Based on gene expression studies of breast cancer subtypes, we expected four triple-negative subtypes, two luminal clusters, a HER2 enriched and a normal-like cluster, therefore we selected eight as the number of epithelial clusters[31,32] (**Figure 4**). For stromal clusters, we selected the minimum number of clusters need to separate T cells from other clusters (k=6).

For analysis of tumor-stroma correlation, the fraction of cell type within the epithelial or non-epithelial stromal compartment (including fibroblast, immune, endothelial and other stromal) were calculated. Pearson correlation was used to assess significant correlations between cell frequencies within their respective tissue compartments.

Spatial distributions of cells were calculated in multiple ways. For analysis of cellular neighbors, each cell's neighbors within a 25 µm radius (~1 epithelial cell diameter) were counted. The 25 µm radius was selected to replicate the mixing score biomarker previously found to be prognostic in TNBC[15]; mean number of neighbors of specific cell types were also calculated in 25 µm neighborhoods. Ripley's L (a measure of clustering) and the multitype K function (Kcross; a measure of two cell types co-localization) functions were calculated using spatsat[33] with a maximum radius of 100 µm. Spatial aggregates were calculated with scimap, with the default radius of 30 µm and purity of 60%[34]. Spatial latent Dirichlet allocation (LDA) analysis was done using spatialLDA[35], using the default radius of 100 µm and 6, 8 or 10 topics for topic modelling. Shorter distances of ~25 µm may be interpreted as cells nearly or directly touching, while 100 µm represents a distance at which oxygen, nutrients and potentially other molecules diffuse in tissues[36]. Survival analysis was done as described above, using the CyCIF dataset to determine significant prognostic cell-cell interactions and cutoff values. These were then verified in the two additional datasets. Spatial and survival analysis code, and links to data are available here: cycIF_TMAs (https://codeocean.com/capsule/7863588/tree).

RESULTS

Multiplex Imaging Enables Single-Cell Phenotyping

We generated CyCIF data from two tissue microarrays (TMAs) containing surgical breast cancer specimens and analyzed publicly available breast cancer data from IMC and MIBI platforms generated with similar antigen targets (**Figure 1**). The CyCIF data was comprised of 47 biomarkers imaged at a resolution of 0.325 µm per pixel in entire cores with 1.2 to 1.4 mm diameters, with 1-2 full cores imaged per patient (**Figure 1** a-b, left). 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 µm per pixel (**Figure 1** a-b, center)[17]. The MIBI data included 36 biomarkers imaged in 0.8 x 0.8 mm square ROIs at a resolution of 0.5 µm per pixel (**Figure 1** a-b, right)[15].

We used the same methods to generate single-cell phenotypes via unsupervised clustering in each platform. In the CyCIF dataset, we generated a 30 k-nearest neighbors (k-NN) graph using a subset of 23 markers common between the CycIF and IMC datasets, plus nuclear area. Markers were scaled by the standard deviation (SD), clipped at 20 SDs, and nearest neighbors were calculated using scanpy[28]. We visualized the k-NN using UMAP embedding, confirming good separation of lineage specific markers, CD31, endothelial, E-cadherin (Ecad), epithelial, Collagen I (ColI), extracellular matrix, vimentin, mesenchymal cells including activated fibroblasts, and CD45, immune infiltrate (S 4a). The UMAP visualization showed good mixing of cells from different TMA sources and separation of tumor cells from different breast cancer subtypes (S 4b). We used the Leiden algorithm (implemented in scanpy[28]) to cluster the k-NN graph, resulting in 23 cell type clusters (S 4c). The mean expression of each biomarker in each cluster was used to annotate cell types (e.g., CD8 T cell, CD4 T cell, luminal ER+ tumor) and lineages (i.e., endothelial, epithelial, fibroblast, immune and other stromal, **Figure 2** a). The most common cell types included luminal and luminal ER+ tumor, CD4 T cells, vimentin+ fibroblasts and quiescent stroma (**Figure 2** a). To confirm our cell typing, we performed manual gating on lineage specific markers (S 4a). Gating and clustering-based cell lineages localized to similar areas of the UMAP and had 73% accuracy on a single cell level, as calculated using metrics.accuracy_score in scikit-learn[30] (S 4e, f).

In the IMC dataset, we used the same method as in the CyCIF dataset to visualize marker expression, TMA, and subtype and cluster cell types (S 4f-g). We used 21 markers plus nuclear area for clustering, resulting in 24 cell types (S 4h). Upon annotation, we found the most common cell types were luminal, luminal ER+, and ER+ HER2+ tumor, vimentin or fibronectin (FN)+ fibroblasts, quiescent stroma, and T cells (**Figure 2** b). Clustering and gating-based cell lineages localized to similar areas of the UMAP and had 77% accuracy (S 4i-j).

The MIBI panel included more immune-specific markers than the other panels and had 15 markers similar to our CyCIF panel. To generate a dataset in which to audit deeper immune contexture, we clustered on all 33 markers plus nuclear area and eccentricity. Again, we saw good separation of lineage specific markers and Leiden clustering generated 22 cell types of which luminal tumor, fibroblasts, T cells and quiescent stroma were the most common (S 5a-c). Gating-based cell types showed good overlap with clustering cell types with 72% accuracy (S 5d-e).

Cell type fractions are similar across platforms

As our most common cell types were similar across platforms, we were encouraged that we could successfully combine the data for cross-platform analysis. We visually validated individual clusters, as well as cell lineages by inspecting the images. In all three platforms, cell lineage identity and spatial distribution matched the underlying imaging data (**Figure 3** a, b).

For further validation, we sought to quantitatively compare the fractions of the five main cell lineages. 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 3** c, d, e). 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 3** d). 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 with were used, resulting in different signal-to-background ratios between platforms, even for the same target, we believe this is a necessary step to account for technical variability.

Hierarchical clustering defines prognostic tumor subtypes consistent across platforms

For subtyping based on single-cell phenotypes, we calculated the fraction of epithelial or non-epithelial stromal cells in the total epithelial or stromal cells in that sample, respectively. Fractions of epithelial cell types representing greater than 4% of the epithelial compartment were used to hierarchically cluster patients from all subtypes and platforms (**Figure 4**a). Given the five intrinsic breast cancer subtypes (luminal A and B, Erb-B2+, basal-like and normal-like)[31,37] and the sub-classification of four subtypes within TNBC/basal-like tumors[32,38], we set k=8 to capture the previously characterized biology of breast cancer epithelial cells. The resulting eight subtypes included tissues enriched for luminal, basal, luminal ER+, myoepithelial, HER2+ ER+, cytokeratin-low, and proliferating tumor cells, as well as a heterogeneous group not dominated by one phenotype (**Figure 4** a). Each subtype contained a mixture of patients from multiple platforms, with the exception of the HER2+ER+ subtype, which was expected given that all HER2+ER+ patients were in the IMC dataset (**Figure 4** a). Examination of images from tissues representing different subtypes showed the expected intensity differences in marker expression (**Figure 4** b). The seven epithelial subtypes present in the ER+ patients were prognostic (log-rank overall-survival (OS) p=0.00004, n=160 patients, **Figure 4** c). Cox proportional hazards modelling (CPH) showed that the heterogeneous subtype 6 had a log hazard ratio significantly > 1, indicating poor prognosis (**Figure 4** d). The CPH hazard ratios (HRs) were similarly ordered across the IMC and CyCIF platforms, with heterogeneous subtype 6 and cytokeratin-low subtype 1 having HRs > 1, or poor prognosis, and luminal ER+ subtype 2 and Luminal subtype 0 having HRs < 1, or good prognosis (S 6a). Investigation of the poor prognosis subtype 6 revealed expression of CD44 and EGFR in tumor (**Figure 4** e, S 6b). Quantification of stromal celltypes and epithelial expression showed significant enrichment of CD31+ endothelial cells and CD44 expression, respectively (**Figure 4** f-g). On the other hand, the luminal ER+ subtype was associated with better prognosis although not significantly, so we investigated luminal ER+ tumor cell fraction as a single variable,

which was found to predict longer OS (log-rank p=0.036, n=162 patients, S 6c). In a multivariate CPH model including stage, patient age and tumor size, the fraction of luminal ER+ tumor was no longer significantly associated with survival (p=0.18, S 6d). In TNBC patients, the epithelial subtypes had borderline prognostic significance (log-rank p=0.1; CPH p=0.051, n=98 patients, S 6e-f). However, the epithelial subtypes did not show a consistent order in hazard ratios obtained from CPH modelling across platforms in TNBC (S 6g-i).

Microenvironment subtypes associate with clinical subtype

We then 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 (S 7). 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, with the exception of the MIBI platform (log-rank=0.003, CPH p=0.056, n=39 patients S 7b-g). 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+ stromal subtype 0 and less T cell-rich stroma (Chi-squared p=0.098, Bonferroni p-adj<0.05 for subtype 0, n=89 patients [ER+ and TNBC, including patients without clinical outcome], S 7e). In the IMC cohort, ER+ patients had more Vim+/FN+ fibroblast stromal subtype 0 and significantly less T cell stromal subtype 4 (Chi squared p=0.002, Bonferroni p-adj<0.05 for subtype 4, n=220 patients [ER+ and TNBC, including patients without clinical outcome] S 7h). Our characterization of stromal subtypes supports the observation that ER+ breast cancer is immune-poor[39], and also shows significant enrichment for vimentin and fibronectin+ fibroblasts in ER+ relative to TNBC.

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

Next, we investigated the prognostic value of multiplex imaging-defined cell types within each clinical subtype. We used the CyCIF dataset as a training cohort to identify cell types whose fraction of all cells in the tissue significantly associated with overall survival and cut-offs for high or low abundance. We tested the 0.33, 0.5 and 0.66 quantile to binarize tissues into low and high cell abundance (S 8a-c). 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 cohorts (S 8d-e). High T and B cells were associated with longer OS in TNBC in multiple platforms (CyCIF p=0.006, n=18, IMC p=0.068, n=41, T cells; CyCIF p=0.049, n=18, MIBI p=0.0069, n=39 B cells, S 8 a, d, e). High macrophages (p=0.043), low endothelial cells (p=0.018) and high proliferating tumor (p=0.031) predicted longer OS in CyCIF TNBC samples (n=18) but were not significant in the other cohorts (S 8 a). In ER+ patients, high luminal ER+ tumor (IMC, p=0.034, n=132), low HER2+ Ki67+ (IMC p=0.058, n=132) and less cytokeratin low tumor (CyCIF p=0.067, n=30) were associated with longer survival, but only in a single cohort (S 8 b, e).

We then combined all samples from each subtype into a single cohort for determining the independent prognostic value of cell type fraction when combined with the clinical variables of stage, patient age and tumor size in a multivariate CPH model. CD3 T cells and CD20 B cells were identified in all cohorts, while only the CyCIF and MIBI panels could distinguish CD4 and CD8 T cells. The MIBI cohort lacked clinical variables of stage, age and tumor size. In combined TNBC

samples, high CD3, CD4 and CD8 T abundance were significantly associated with longer OS (log-rank, Holm-Šídák multiple testing correction, p=0.01, n=98; p=0.037 and 0.0025, n=57 respectively **Figure 5** a-b, S 9 a), and they remained significant in a multivariate model (CPH p=0.01, n=55, (CD3); p=0.049 (CD4) and 0.0071 (CD8), n=15, **Figure 5** c, S 9 b). High CD20 B cell abundance was associated with longer OS (log-rank, Holm-Šídák correction, p=0.011, n=98) but it did not remain significant in a multivariate model (CPH p=0.15, n=55, S 9 c). CD3 T cell abundance predicted longer RFS in the multivariate model (p=0.026, n=55, S 9 d). CD20 had borderline predictive value for RFS (multivariate CPH p=0.082, n=55 S 9 e), while CD8 and CD4 T cells were not independently predictive of RFS (n=15, S 9 f-g).

CD3 T cells did not predict survival in the combined ER+ breast cancer cohort (log-rank, Holm-Šídák multiple correction, p=0.83, n=162), however, in tumors with proliferation above the median, CD3 T cells did predict longer OS (log-rank, Holm-Šídák multiple correction, p=0.0013, n=74) (**Figure 5** d-e). Low proliferation ER+ did not derive a survival benefit from high CD3 T cells; in contrast, low proliferation TNBC did (log-rank, Holm-Šídák multiple correction, p=0.99 and p=0.023, S 9h). Multivariate CPH modelling revealed that CD3 T cells were independently prognostic for overall and recurrence-free survival in high proliferation but not low proliferation tumors from ER+ and TNBC subtypes (S 9 i-j). Stratification of tumors by proliferation and T cell abundance showed that high proliferation, high T cell tumors had the best prognosis in ER+ disease, while high proliferation low T cell had the worse prognosis (S 10 a). In contrast, in TNBC high- and low-proliferation, high-T cell tumors had similarly good survival, while high- and low-proliferation, low-T cell tumors had similarly poor survival (S 10 a). Independent analysis in each platform revealed significant 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 S 10 b), and in TNBC patients from the IMC and CyCIF cohorts but not MIBI cohort (log-rank p=0.056, 0.012 and 0.49 S 10 c).

Since the ER+ low-proliferation group was unique in not deriving survival benefit from CD3 T cells and had intermediate survival outcomes compared to other ER+ patients (S 10 b), we compared T cell abundance, spatial localization, and activation states to identify unique T cell characteristics in that group. There was no significant difference in T cell abundance or T cell to macrophage or endothelial ratio in low proliferation versus high proliferation ER+ tumors (S 10 e). 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 number of CD3 T cell neighbors of each CD3 T cell (Tukey HSD p=0.05, **Figure 5** f). High proliferation ER+ tumors in the CyCIF cohort, which did not gain survival benefit from CD3 T cells, did not show increased T cell clustering (Kruskal-Wallis p=0.61, S 10 d). 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=0.006 and 0.01, **Figure 5** g-h). 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 (S 10 f). High-proliferation TNBC, in which T cells independently predicted overall survival in the multivariate CPH model, also showed increased levels of PD-1, FoxP3, IDO and Lag3 expression in T cells, consistent with upregulation of negative feedback checkpoints following immune activation (S 10 f-g). Epithelial cells in high-

proliferation TNBC had increased expression of the antigen presentation molecule HLA-Class-1 and immune checkpoint PD-L1 (S 10 h).

Analysis of pairwise tumor-immune interactions reveals conserved 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 5** d, f), we leveraged cross-cohort analysis to systematically investigate spatial tumor-immune interactions as biomarkers in breast cancer. We counted the number of immune and tumor neighbors within close proximity (25μm) of each cell and derived previously published prognostic biomarkers including immune aggregates[34,40] and tumor-immune mixing score[15]. We also quantified spatial correlation of tumor and immune cells independent of density (Ripley's L and Kcross functions[33]). For each spatial metric, subtype and platform, we binarized patients by the median into high and low, and tested for association with overall and recurrence free survival. Only a minority of spatial metrics showed correlations with survival across multiple platforms (S 11). Of these, spatial aggregates of CD3 T cells predicted longer survival in the CyCIF and IMC cohorts, and independently predicted OS and RFS in a multivariate CPH model (**Figure 6** a-c, S 11 b). Similar to abundance, T cell spatial aggregates did not predict survival in the MIBI cohort (S 11 b). The Kcross function, which quantified spatial correlation of tumor cells and macrophages (50 μm radius) showed a negative association with RFS in both ER+ breast cancer cohorts, with a trend towards significance in the multivariate model (CPH p=0.083, **Figure 6** d-f). Macrophage neighbors of proliferating tumor predicted longer overall survival in TNBC in both the MIBI and CyCIF platforms, but were not significant in the multivariate model (**Figure 6** g-h). Epithelial neighbors of epithelial cells (within 25μm) were associated with longer RFS in ER+ tumors from both CyCIF and IMC platforms, but not in the multivariate CPH model (**Figure 6** i-j). Finally, mixing score, which had been developed in the MIBI cohort, was shown to predict shorter survival in that cohort, but not the CyCIF and IMC cohorts (**Figure 6** k), nor was it independently prognostic (S 11 a). Several variables, including the Kcross function of tumor to quiescent stroma, macrophage neighbors of T cells, and vimentin+ fibroblast neighbors of proliferating tumor showed opposite significant associations with survival in different platforms of the same subtype (S 11 c-f).

Although the spatial metric for CD3 T cell aggregates was independently prognostic, this was also true of CD3 T cell abundance, without the spatial information (**Figure 5**). Therefore, in order to identify spatial metrics which provided additional information beyond abundance, we calculated the Pearson correlation between all of our spatial metrics and cell type abundance within each tissue (S 12). As expected, CD3 T cell abundance significantly correlated with CD3 T cell spatial aggregates, as well as neighbor counts among T cells, B cells, macrophages, endothelial cells and tumor cells (S 12). CD3 T cell abundance negatively correlated with mixing score and abundance of quiescent stroma and fibronectin+ fibroblasts (S 12). Tissue abundance of fibroblasts and quiescent stroma correlated with high neighbor counts and spatial aggregates involving those cell types (S 12). True to their definitions as density independent, the Kcross and Ripley's L function did not correlate with cell type abundances (S 12) and are therefore most likely to provide unique insight not gained from abundance alone.

Neighborhood analysis reveals multicellular spatial biomarkers

Finally, we analyzed multicellular spatial neighborhoods by considering all stromal cells within 100 μm 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[35]. In LDA analysis, each topic can be enriched for multiple cell types and each cell type can be present in multiple topics. In our CyCIF data, for example, topic-0 in TNBC tissues was enriched in macrophages, vimentin+ fibroblasts and CD4 T cells, while CD4 T cells are found in topic-0, 4, 5 and 6 (**Figure 7** a). 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 7** b). Clustering the topic matrix rather than the neighbor count matrix is believed to be less sensitive to noise[34] and allows for smooth transitions between neighborhoods as opposed to arbitrary cutoffs[35]. We did observe transitioning/mixed neighborhoods within both TNBC and ER+ cluster results (**Figure 7** b-d, S 13). 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 7** c-d). A TNBC tissue in our CyCIF cohort, for example, showed tumor cell neighborhoods with more T cells (blue) on the tumor margin, with adjacent macrophage-rich tumor neighborhoods (purple) (**Figure 7** c-d). These neighborhoods transition into a mixed neighborhood (brown), and finally a vimentin+ FB neighborhood (green) distant from the infiltrating T cells (**Figure 7** c-d). Similar neighborhoods were identified in the IMC cohort (S 13a), and a combined survival analysis showed that T cell-rich neighborhoods were associated with longer overall survival (log-rank p=0.011) and vimentin+ FB neighborhoods were associated with shorter survival in TNBC (log-rank p=0.0039, **Figure 7** e). T cell neighborhoods trended towards significance in a multivariate CPH model of overall survival (p=0.056, **Figure 7** f) and recurrence-survival (p=0.072, S 14a). Vimentin+ fibroblast neighborhoods around tumor cells were significantly associated with shorter overall and recurrence-free survival in the multivariate model (p=0.03 OS, 0.023 RFS **Figure 7** f, S 14b). Interestingly, vimentin+ fibroblast abundance alone was not prognostic, although neighborhood abundance was correlated with fibroblast abundance (S 15). Finally, similar to other groups [34], we found that directly clustering the neighborhood counts using Kmeans (rather than running LDA and clustering the topics) did not result in robust neighborhood clustering, as judged by the inconsistency of prognostic value of neighborhoods defined in this manner (S 16).

In ER+ breast cancer, we identified similar neighborhoods within CyCIF and IMC cohorts that matched the spatial distribution of markers in the images (S 13b-e). Patients with more mixed fibroblast neighborhoods around tumor cells trended towards shorter overall survival in both cohorts and a multivariate CPH model (S 14c). Quiescent stroma neighborhoods showed association with longer recurrence-free survival in both cohorts, but this did not remain significant in the multivariate model (S 14d).

Tumor phenotypes correlate with stromal cell abundance and spatial neighborhoods

We hypothesized that there would be significant correlation between tumor cell types and the surrounding stromal cell 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 (S 15a, c). Epithelial cell types were inversely correlated with each other, indicating most tumors had

just one main epithelial cell type (S 15a, c). The exception was luminal tumor, which correlated with cytokeratin low tumor in ER+ breast cancer, indicating mixing of these phenotypes within the same tissues (S 15c). This conclusion is supported by our hierarchical clustering, which revealed luminal tumor as the main epithelial cell type likely to be found mixing with other phenotypes in the same tissue (**Figure 4** a).

Immune cells exhibited distinct tissue-level correlations in the different subtypes. In TNBC, T cells correlated with proliferating tumor, and macrophages correlated with CD4 T cells (S 15a). In ER+ breast cancer, T cells correlated with B cells while proliferating tumor, macrophages and endothelial cells were correlated with each other but not T cells (S 15c). Vimentin+, FN+, and ColI+ fibroblasts, as well as quiescent stroma were inversely correlated with immune cells (S 15a, c). In both subtypes, spatial LDA neighborhoods correlated strongly with the abundance of their respective enriched cell types; however, neighborhoods showed unique correlations to other cell types present. For example, proliferating and luminal ER+ tumor cell abundances did not correlate with total T cell abundance in ER+ breast cancer, but they did correlate with the fraction of T cell neighborhoods (S 15c). In TNBC, vimentin+ and fibronectin+ fibroblast abundances were not correlated, but their respective neighborhoods were inversely correlated, suggesting tissue exclusivity for a single fibroblast phenotype near tumor cells (S 15a, b). Therefore, although spatial neighborhoods tend to correlate with cell abundance, they do reveal unique features of tumor-stromal organization in tissues.

DISCUSSION

Our approach of combined analysis across multiple platforms shows the power of our methods for biomarker discovery. We were able to incorporate analysis of two publicly available imaging datasets with our own CyCIF data for efficient discovery of robust biomarkers.

We utilized our optimized method for CyCIF[19] to generate multiplex imaging data of 42 markers in a single tissue section 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[15,17]. Using our image processing and analysis pipeline, we generated single cell phenotyping data from our CyCIF dataset and two publicly available datasets[15,17]. The advantage of using mplexable for image processing and feature extraction is the ability to match nuclear and cell segmentation results from separate deep-learning segmentation models and extract features from subcellular compartments such as the nucleus and cytoplasm (S 1b). We found that cell type fractions across cohorts from the same breast cancer subtype were highly correlated across platforms.

Since we were able to identify similar epithelial phenotypes across platforms, we combined patient data from all platforms and identified eight epithelial subtypes through hierarchical clustering. Our subtypes were similar to the intrinsic breast cancer subtypes[37], including a luminal ER+ luminal A-like group with good prognosis, a cytokeratin low group previously shown to share features with luminal B tumors[16], and a HER2 enriched group. Triple-negative breast cancers also fell into categories similar to those defined by gene expression profiling[32], including a highly proliferative, basal-like 1 (BL1)-like group, a luminal androgen receptor (LAR)-like group with luminal

differentiation states, and subtypes of basal/myoepithelial phenotype reminiscent of the basal-like 2 (BL2) group. Our proliferative subtype was correlated with immune infiltrate, which was in turn correlated with better prognosis, supporting a link between our proliferative subtype and the good-prognosis proliferative BL1 subtype[38]. We also identified a heterogeneous subtype with low cytokeratin and elevated CD44 expression that may represent tumors with mesenchymal features. The ER+ tumors in this subtype had significantly worse prognosis, and subsequent analysis showed increased angiogenesis in these tissues. 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[41], 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[42], and we confirmed that T cells are an independent prognostic factor in TNBC in the multiplex imaging datasets analyzed herein. Furthermore, we showed that CD3 T cells were independently prognostic specifically in high-proliferation ER+ and TNBC.

Previous gene expression profiling studies linked anti-tumor immunity and proliferation. Nagalla *et al.* found that immune signatures were prognostic solely in breast cancer patients with the highest proliferation gene expression[43]. 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[44]. In luminal A and claudin-low subtypes, immune gene signatures were not prognostic, regardless of proliferation[44]. Similarly, in our analysis, CD3 T cell infiltration was not independently prognostic in low-proliferation ER+ and TNBC (S 9 i, j).

Our analysis of immune functional states showed increased T cell proliferation, activation and checkpoint molecules and epithelial antigen presentation in high-proliferation tumors, consistent with IFNγ pathway activation. Consistent with our analysis, gene network analysis previously showed activation of TNFα/IFNγ signaling pathways in tumors with productive anti-tumor immunity and TGF-β, an immunosuppressive cytokine, in tumors with unproductive anti-tumor immunity[44]. TGF-β also has anti-proliferative effects and is associated with good outcome in ER+ breast cancer cohorts[45], suggesting that it could mechanistically link lower proliferation rates with immunosuppression.

More recently, the group mentioned above showed that immune gene signatures were prognostic exclusively in tumor-mutation burden (TMB)-high breast cancer tumors[46]. Thirty-seven percent of basal-like tumors had high TMB, while only 11.5% of luminal A tumors did[46], 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 would increase TMB[47]. In summary, TMB would hypothetically provide a mechanistic link between proliferation and anti-tumor immunity and should be investigated in future studies.

Interestingly, high proliferation ER+ tumors showed no significant difference in CD3 T cell infiltration, or checkpoint marker intensity, but a significant increase in the number of T cell neighbors of each T cell, suggesting that spatial metrics could serve as improved biomarkers over

cell abundance alone. Herein, our investigation of spatial biomarkers yielded mixed results. For example, T cell aggregates predicted survival in TNBC but so did T cell abundance. Several biomarkers gave conflicting results in different cohorts (S 11). We also showed significant correlation between many spatial metrics and cell abundance, but the Ripley's L and Kcross functions did not suffer from that limitation (S 12).

Using the Kcross function we identified tissues with significant compartmentalization versus mixing of macrophages and tumor, with the latter phenotype associated with shorter RFS in ER+ breast cancer. Tumor associated macrophages (TAMs)[48] and TAMs specifically in the tumor core[49] have been associated with shorter overall survival independent of breast cancer subtype, but the prognostic value of macrophages specifically in ER+ patients was not established in these studies. Similarly, studies evaluating macrophages in the tumor stroma (TS) versus tumor nest (TN) found CD68[50] and CD163 positive macrophages[51] in the TS but not TN independently predicted shorter survival. This is the opposite of our findings which suggest tumor-nest macrophages (which would presumably have higher Kcross) are associated with survival; however, the prior mentioned studies did not specifically analyze ER+ breast cancer patients separately and did not quantify the degree of TN versus TS localization (i.e. what Kcross does) but rather analyzed them separately. Our finding should be further investigated in future cohorts, and the specific prognostic value of macrophage-tumor spatial correlation in ER+ breast cancer determined.

We utilized spatial LDA modelling to analyze multicellular neighborhoods of stromal cells surrounding tumor cells. In addition to T cell neighborhoods predicting good prognosis in TNBC as expected, we identified a neighborhood enriched for vimentin+ fibroblasts (VIM+ FB) that was independently associated with shorter survival in TNBC. Given the high levels of vimentin and low levels of alpha-SMA, these cells may be an inflammatory phenotype similar to CAFs that differentiate under TNFα + IL-1β stimulation[52]. Interestingly, TNFα + IL-1β 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[53], consistent with close proximity between CAFs and tumor in cellular 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, we relied on matching hand-annotated clusters across platforms for our integrated analysis. Since our annotations may not correspond to the same cell types in each platform, this introduces uncertainty. Some well-defined phenotypes, such as T cells and proliferating tumor, are relatively straightforward. 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. 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 (≤3) are needed to binarize patients in to high and low TILs, although a larger number of cores (≥11) are needed to accurately estimate the mean TIL abundance[54,55]. Our survival analysis relied on binarizing patients; therefore, the use of TMAs may be appropriate in this context. Lacking large tissue sections, we had no way to quantify sensitivity

of spatial metrics to sampling bias introduced by analyzing small ROIs, but our discovery that a number of spatial metrics showed opposite survival associations in different cohorts suggests caution when interpreting spatial biomarker results (S 11). Analysis of multiple cohorts should help identify robust spatial biomarkers. The optimal tissue area for various composition and spatial analytics 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 utilized to analyze additional cohorts for further characterization of biomarkers in breast cancer and other tumor types.

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 also like to thank Damir Sudar for technical support related to image storage and Elliot Gray for sharing his neighbor counter script. 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 P50 CA098131 and Komen SAB210301.

AUTHOR CONTRIBUTIONS

J.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.E. and E.B. did single-cell image processing. J.E. ran the analysis, drafted the manuscript, and prepared the figures. S.L.G., K.C., E.B., B.C. and J.A.P. edited the manuscript.

REFERENCES

1. FDA. FDA approves pembrolizumab for high-risk early-stage triple-negative breast cancer. 20–22 https://www.fda.gov/drugs/drug-approvals-and-databases/resources-information-approved-drugs (2021).

2. Rizzo, A. & Ricci, A. D. Biomarkers for breast cancer immunotherapy: PD-L1, TILs, and beyond. *Expert Opin Investig Drugs* 1–7 (2021) doi:10.1080/13543784.2022.2008354.

3. Schmid, P. *et al.* Pembrolizumab for Early Triple-Negative Breast Cancer. *New England Journal of Medicine* **382**, 810–821 (2020).

4. Gatti-Mays, M. E. *et al.* If we build it they will come: targeting the immune response to breast cancer. *NPJ Breast Cancer* **5**, (2019).

5. Tsujikawa, T. *et al.* Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Rep* **19**, 203–217 (2017).

6. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science (1979)* **361**, (2018).

7. Lin, J. R., Fallahi-Sichani, M. & Sorger, P. K. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat Commun* **6**, 1–7 (2015).

8. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).

9. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* **11**, 417–422 (2014).

10. 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, 2020).

11. Cabrita, R. *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577**, 561–565 (2020).

12. Helmink, B. A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* **577**, 549–555 (2020).

13. Phillips, D. *et al.* Immune cell topography predicts response to PD-1 blockade in cutaneous T cell lymphoma. *Nat Commun* **12**, (2021).

14. Mi, H. *et al.* Quantitative Spatial Profiling of Immune Populations in Pancreatic Ductal Adenocarcinoma Reveals Tumor Microenvironment Heterogeneity and Prognostic Biomarkers. *Cancer Res* **82**, 4359–4372 (2022).

15. 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).

16. Ali, H. R. *et al.* Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nat Cancer* **1**, 163–175 (2020).

17. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* (2019) doi:10.1038/s41586-019-1876-x.

18. Parmar, C., Barry, J. D., Hosny, A., Quackenbush, J. & Aerts, H. J. W. L. Data Analysis Strategies in Medical Imaging. *Clinical Cancer Research* **24**, 3492–3499 (2018).

19. Eng, J. *et al.* A framework for multiplex imaging optimization and reproducible analysis. *Communications Biology 2022 5:1* **5**, 1–11 (2022).

20. Eng, J. *et al. Cyclic Multiplexed-Immunofluorescence (cmIF), a Highly Multiplexed Method for Single-Cell Analysis*. *Methods in Molecular Biology* vol. 2055 (2020).

21. 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).

22. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).

23. 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).

24. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. 1–19 (2020) doi:10.1101/2020.02.02.931238.

25. van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

26. Danenberg, E. *et al.* Breast tumor microenvironment structures are associated with genomic features and clinical outcome. *Nat Genet* **18**, (2022).

27. napari contributors. napari: a multi-dimensional image viewer for python. Preprint at (2019).

28. Wolf, A., Angerer, P. & Theis, F. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, (2018).

29. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 1–12 (2019).

30. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python Fabian. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

31. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869–10874 (2001).

32. Lehmann, B. D. *et al.* Refinement of triple-negative breast cancer molecular subtypes: Implications for neoadjuvant chemotherapy selection. *PLoS One* **11**, (2016).

33. Baddeley, A. R. E. T. R. *Spatial Point Patterns: Methodology and Applications with R*. (Chapman and Hall/CRC Press, 2015).

34. 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).

35. Chen, Z., Soifer, I., Hilton, H., Keren, L. & Jojic, V. Modeling Multiplexed Images with Spatial-LDA Reveals Novel Tissue Microenvironments. *Journal of Computational Biology* **27**, 1204–1218 (2020).

36. Weinberg, D. H. and R. A. The Hallmarks of Cancer. *Cell* **107**, 251–262 (2000).

37. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).

38. Lehmann, B. D. B. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *Journal of Clinical Investigation* **121**, 2750–2767 (2011).

39. Goldberg, J. *et al.* The Immunology of Hormone Receptor Positive Breast Cancer. *Front Immunol* **12**, 1–22 (2021).

40. Nawaz, S., Heindl, A., Koelble, K. & Yuan, Y. Beyond immune density : critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Modern Pathology* **28**, 766–777 (2015).

41. Fantozzi, A. *et al.* VEGF-mediated angiogenesis links EMT-induced cancer stemness to tumor initiation. *Cancer Res* **74**, 1566–1575 (2014).

42. 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).

43. Nagalla, S. *et al.* Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome Biol* **14**, (2013).

44. Miller, L. D. *et al.* Immunogenic Subtypes of Breast Cancer Delineated by Gene Classifiers of Immune Responsiveness. *Cancer Immunol Res* **4**, 600–611 (2016).

45. Sato, M. *et al.* An integrated genomic approach identifies persistent tumor suppressive effects of transforming growth factor-β in human breast cancer. *Breast Cancer Research* **16**, 1–23 (2014).

46. Thomas, A. *et al.* Tumor mutational burden is a determinant of immune-mediated survival in breast cancer. *Oncoimmunology* **7**, 1–12 (2018).

47. Macheret, M. & Halazonetis, T. D. DNA replication stress as a hallmark of cancer. *Annual Review of Pathology: Mechanisms of Disease* **10**, 425–448 (2015).

48. Leek, R. D. *et al.* Association of macrophage infiltration with angiogenesis and prognosis in invasive breast carcinoma. *Cancer Res* **56**, 4625–4629 (1996).

49. Morita, Y. *et al.* Pathologic evaluation of tumor-associated macrophage density and vessel inflammation in invasive breast carcinomas. *Oncol Lett* **14**, 2111–2118 (2017).

50. Medrek, C., Pontén, F., Jirström, K. & Leandersson, K. *The presence of tumor associated macrophages in tumor stroma as a prognostic marker for breast cancer patients*. http://www.biomedcentral.com/1471-2407/12/306 (2012) doi:10.1186/1471-2407-12-306.

51. Esbona, K. *et al.* The Presence of Cyclooxygenase 2, Tumor-Associated Macrophages, and Collagen Alignment as Prognostic Markers for Invasive Breast Carcinoma Patients. *American Journal of Pathology* **188**, 559–573 (2018).

52. 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).

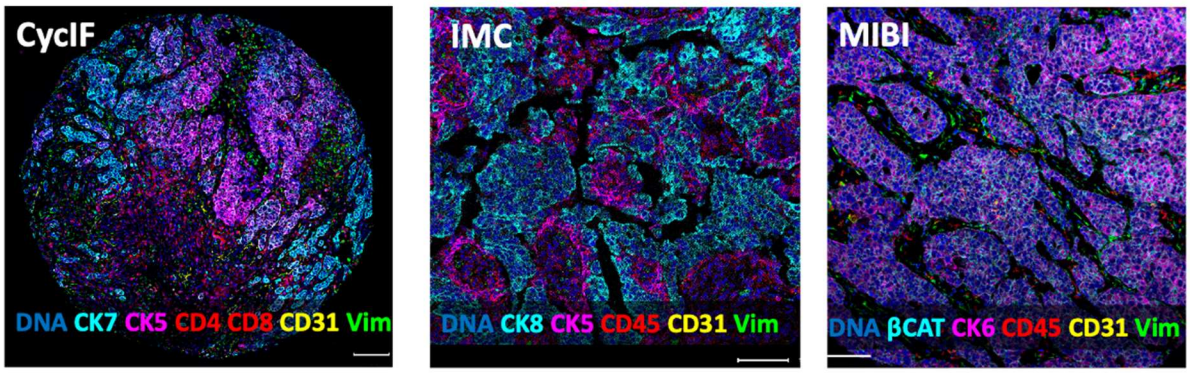
53. Liubomirski, Y. *et al.* Tumor-stroma-inflammation networks promote pro-metastatic chemokines and aggressiveness characteristics in triple-negative breast cancer. *Front Immunol* **10**, 757 (2019).

54. 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).

55. Liudahl, S. M. *et al.* Leukocyte heterogeneity in pancreatic ductal adenocarcinoma: Phenotypic and spatial features associated with clinical outcome. *Cancer Discov* **11**, 2014–2031 (2021).

FIGURES

a



b

CycIF

Antigen	Description	Antigen	Description
Cytokeratins		**Immune Context**	
CK5	Basal CK	CD45	pan-Immune
CK7	Luminal CK	CD3	T Cells
CK8	Luminal CK	CD20	B Cells
CK14	Basal CK	CD8	Cytotoxic T cell
CK17	Basal CK	CD4	Helper T cell
CK19	Luminal CK	CD68	Macrophage
Adhesion Molecules		PD1	T cell exhaustion
Ecadherin	Cell Adhesion	FoxP3	Regulatory T cell
CD44	Cell Adhesion	GRNZB	Cytotoxicity
Hormone Receptors		**Hypoxia and Metabolism**	
ER	Estrogen receptor	HIF1a	Hypoxia
PR	Progesterone Receptor	Glut1	Glucose Transport
AR	Androgen Receptor	CoxIV	Mitochndria
RTK Signaling		**Cell Growth and Division**	
HER2	RTK	Ki67	Proliferation
EGFR	RTK	PCNA	Proliferation
Endothelial		p-HH3	Mitosis
CD31	Endothelial	p-S6	Growth
PDPN	Lymphatic Endothelial	p-RB	Cell Division
Stromal Markers		**Nuclear**	
SMA	Myoepithelial cells	Lamin AC	Nuclear Membrane
Vimentin	Mesenchymal Cells	Lamin B1	Nuclear Membrane
Collagen IV	Basement Membrane	Lamin B2	Nuclear Membrane
Collagen I	Extracellular Matrix	**Cell Death**	
Epigenetic Mark		cleaved PARP	Apoptosis
H3K27	Gene Repression		
H3K4	Gene Accessibility		

IMC

Antigen	Description	Antigen	Description
Cytokeratins		**Immune Context**	
CK5	Basal CK	CD45	pan-Immune
CK7	Luminal CK	CD3	T Cells
CK8/18	Luminal CK	CD20	B Cells
CK14	Basal CK	CD68	Macrophage
CK19	Luminal CK	**Transcription Factors**	
AE1/AE3	Pan-CK	p53	Tumor Suppressor
Adhesion Molecules		cMyc	Proto-Oncogene
E/P-cadherin	Cell Adhesion	GATA3	Luminal TF
CD44	Cell Adhesion	Twist	EMT TF
Hormone Receptors		Slug	EMT TF
ER	Estrogen receptor	**Hypoxia and Metabolism**	
PR	Progesterone Receptor	HIF1a	Hypoxia
RTK Signaling		Glut1	Glucose Transport
HER2	RTK	CoxIV	Mitochondria
HER2	RTK	**Cell Growth and Division**	
Endothelial		Ki67	Proliferation
vWF	Endothelial	p-HH3	Mitosis
CD31	Endothelial	p-S6	Growth
Stromal Markers		P-MTOR	Growth
Fibronectin	Matrix Glycoprotein	**Nuclear**	
SMA	Myoepithelial cells	Histone H3	Chromatin
Vimentin	Mesenchymal Cells	**Cell Death**	
Epigenetic Mark		cleaved PARP	Apoptosis
H3K27	Gene Repression	cleaved Casp3	Apoptosis

MIBI

Antigen	Description	Antigen	Description
Cytokeratins		**Immune Context**	
CK6	Basal CK	CD45	pan-Immune
CK17	Basal CK	CD3	T Cells
pan-CK	Luminal & Basal	CD20	B Cells
Adhesion Molecules		CD8	Cytotoxic T cell
Beta catenin	Cell Adhesion	CD4	Helper T cell
Transcription Factors		CD68	Macrophage
p53	Tumor suppressor	PD1	Immune regulation
RTK Signaling		FoxP3	Regulatory T cell
EGFR	RTK	CD45RO	Memory T cells
Endothelial		CD56	NK cells
CD31	Endothelial	HLA Class 1	Antigen Presentation
Stromal Markers		HLA DR	Antigen Presentation
SMA	Myoepithelial cells	CD11c	Dendritic cells
Vimentin	Mesencymal Cells	CD11b	pan-Myeloid
Epigenetic Mark		CD138	Plasma cells
H3K27me3	Gene Repression	CD16	NK cells
H3K9ac	Gene Accessibility	CD209	Antigen Presentation
Cell Growth and Division		IDO	Immune regulation
Ki67	Proliferation	Lag3	Immune regulation
p-S6	Growth	MPO	Neutrophils
		PDL1	Immune regulation
		CD63	Monocytes

Figure 1. Breast cancer multiplex imaging datasets from different platforms.

a. We generated CyCIF data from two breast cancer TMAs and obtained publicly available IMC and MIBI data from three additional breast cancer TMAs. b. The antibody panels from the three datasets included markers for cytokeratins, adhesion molecules, hormone receptors, receptor tyrosine kinase (RTK) signaling, cell growth and division, endothelial, immune and stromal cells. CyCIF scale = 130 um, IMC scale = 100 um, MIBI scale = 96 um.

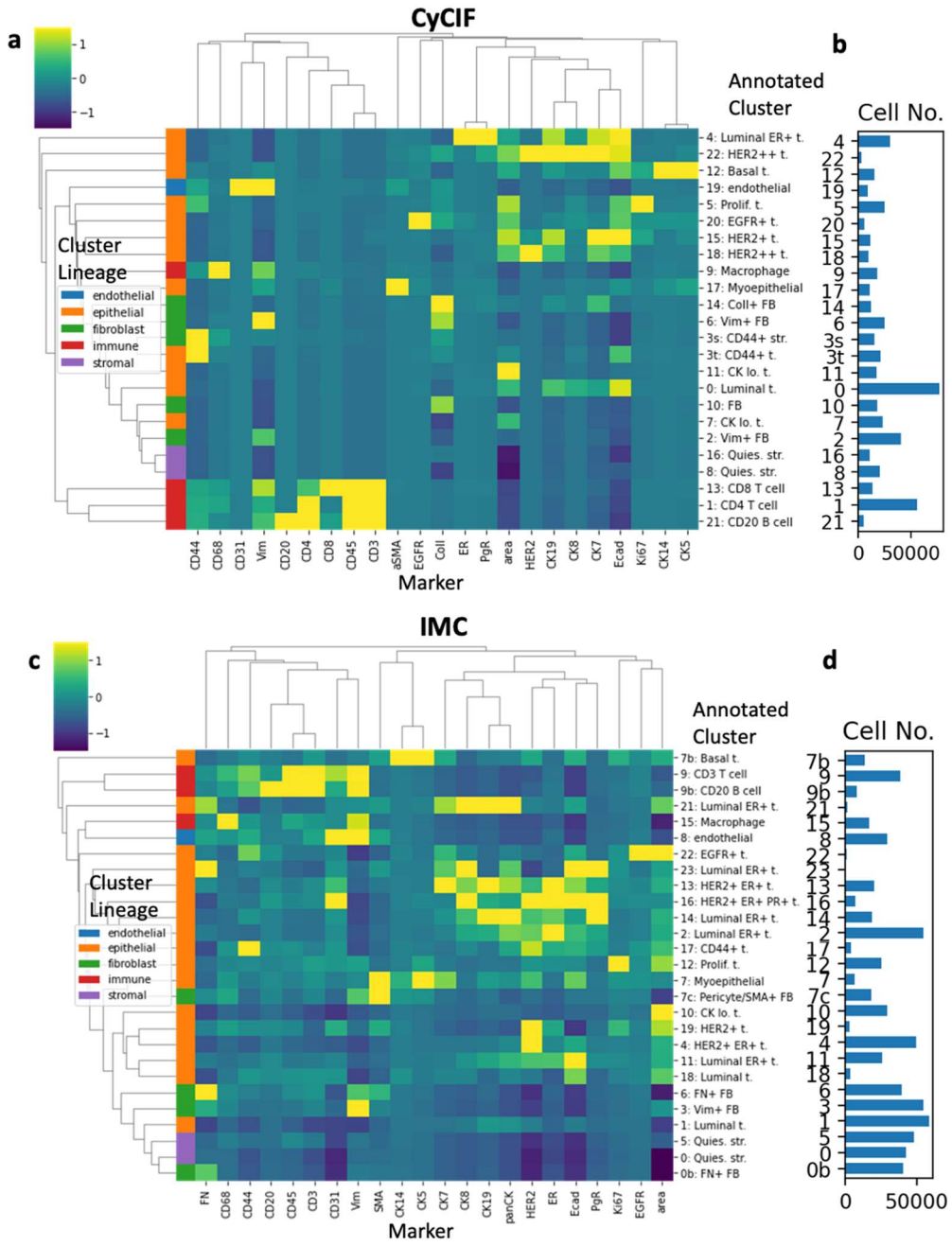


Figure 2. Unsupervised clustering defines cell types in CyCIF and IMC breast cancer datasets.

a. Heatmap of mean fluorescence intensity of each marker in CyCIF cell type clusters. Twenty-two markers and one morphology feature (nuclear area) were used for clustering. Clusters were annotated by lineage: endothelial, epithelial, fibroblast, immune or stromal (left color bar on heatmap) and named based on marker expression (right labels on heatmap). b. The number of cells of each cluster in the CyCIF dataset. c. Heatmap of mean fluorescence intensity of each marker in IMC cell type clusters with same annotation schema as (a). d. The number of cells of each cluster in the IMC dataset.

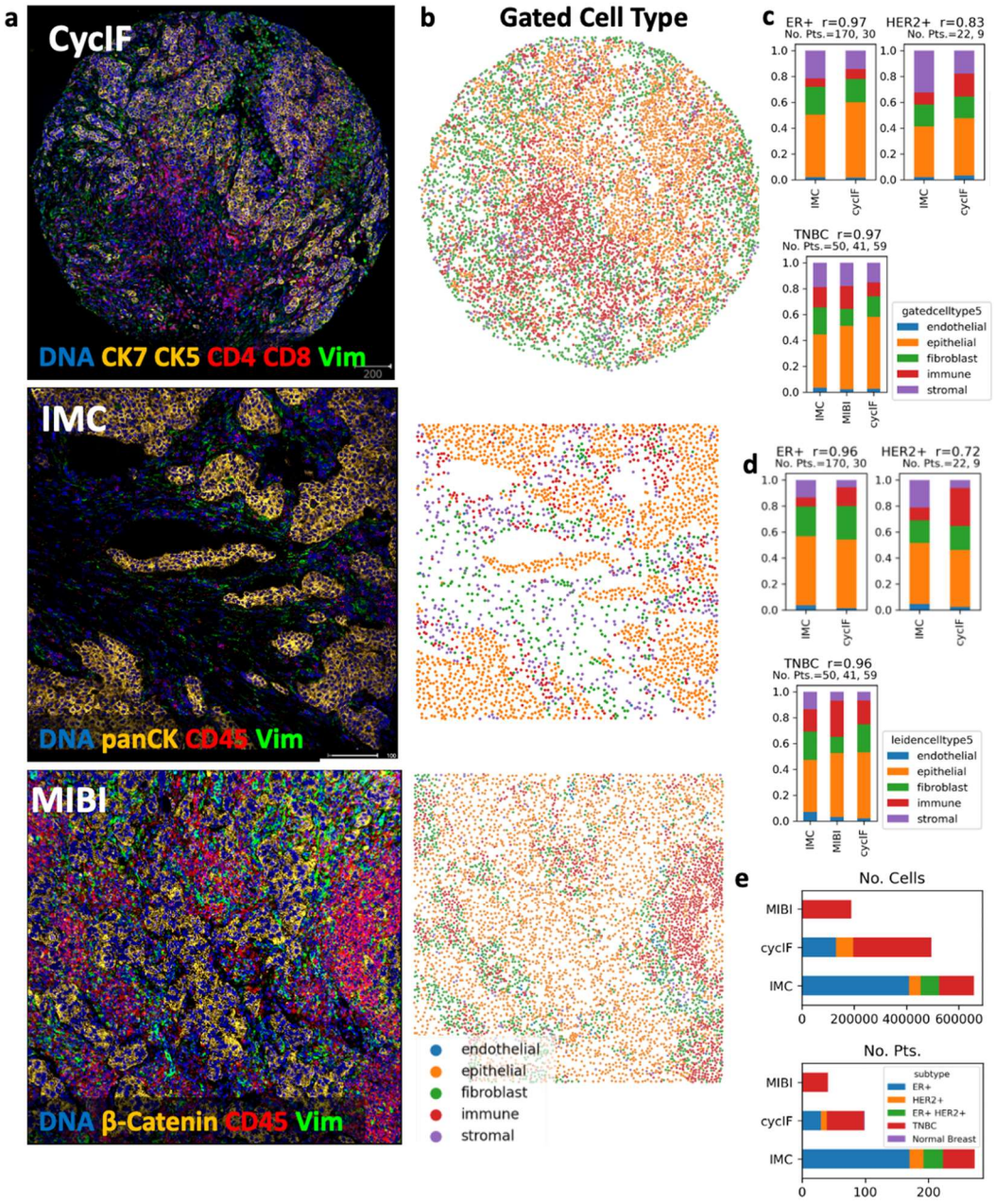


Figure 3. Concordance in abundance of cell types in breast cancer tissues across imaging platforms.

a. Representative images from the three multiplex imaging platforms showing epithelial (orange), immune (red) and fibroblast (green) markers. b. Gated cell types showing cell location and lineages: epithelial (orange), immune (red), fibroblast (green), endothelial (blue) and stromal (purple). c. Gated cell lineage totals per subtype (ER+, HER2+ and TNBC), per platform (IMC, CyCIF and MIBI). d. Cell type cluster lineage per subtype, per platform. c-d. Pearson's correlation between platforms shown in figure title (r=0.xx) e. Number of cells analyzed (top) and number of patients (bottom), colored by subtype, per platform.

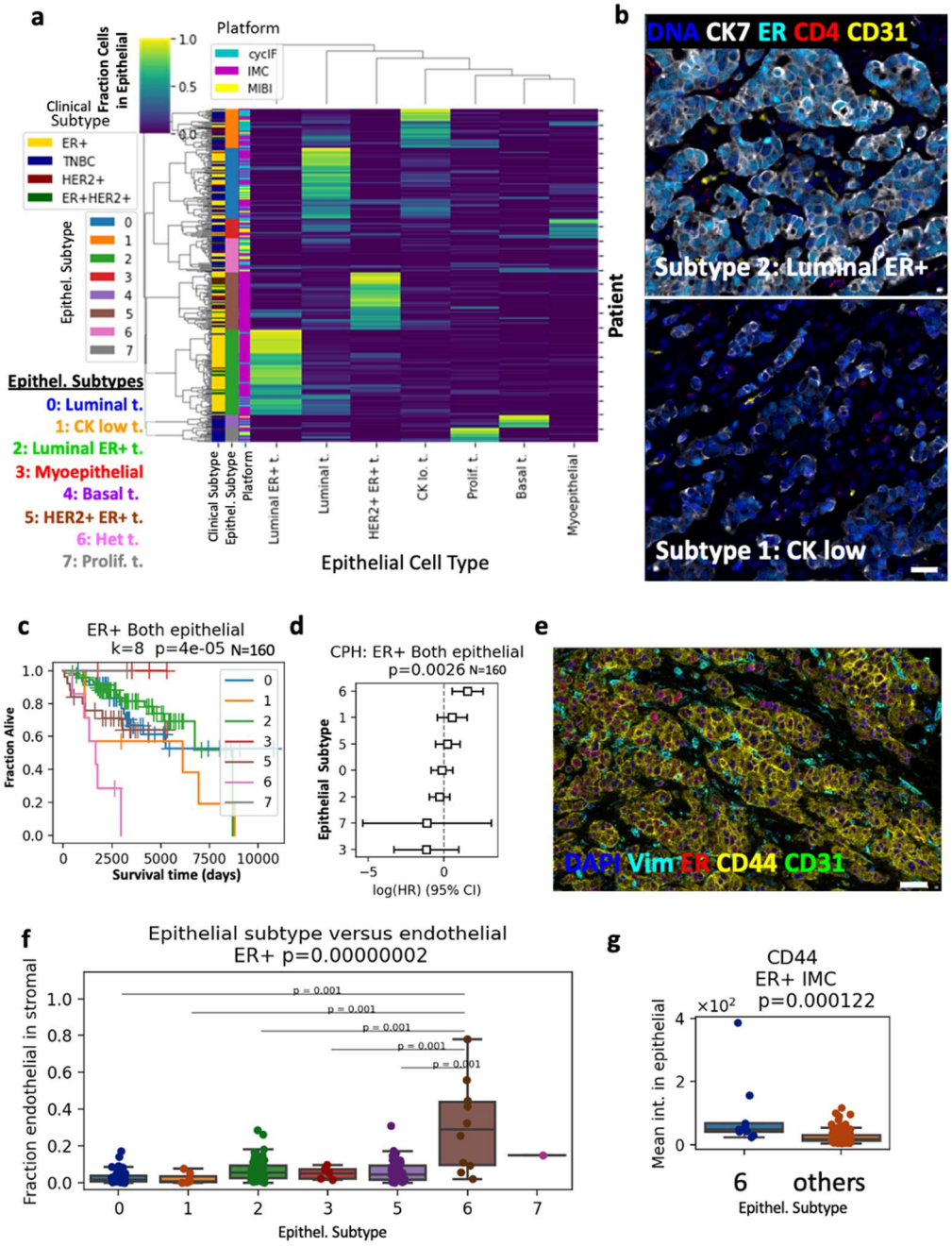


Figure 4. Prognostic breast cancer subtypes in integrated multiplex imaging data.

a. All ER+ and TNBC patients were hierarchically clustered based on the fraction in each patient's tissue of the seven most common epithelial cell types. This resulted in eight epithelial (Epithel.) subtypes. b. Representative CyCIF images of subtype 2 luminal ER+ tumor (top) and subtype 1 cytokeratin-low tumor (CK low, bottom). c. Kaplan-Meier curves (p-value from log-rank test) and Cox proportional hazard (CPH) models comparing overall survival (OS) in the seven epithelial subtypes present in ER+ tumors. c-d. N number of patients given in figure titles. E. Representative IMC image of poor prognosis ER+ tumor. Scale bar = 50 μm. f. Fraction of endothelial cells in tissue stromal cells of each epithelial subtype. Kruskal-Wallis H-test P-value given in figure title. Post-hoc Tukey HSD used for pairwise comparisons between groups. g. CD44 intensity in epithelial cells from poor prognosis epithelial subtype 6 compared to other ER+ patients in the IMC cohort. P-value from Mann-Whitney U rank test given in figure title.

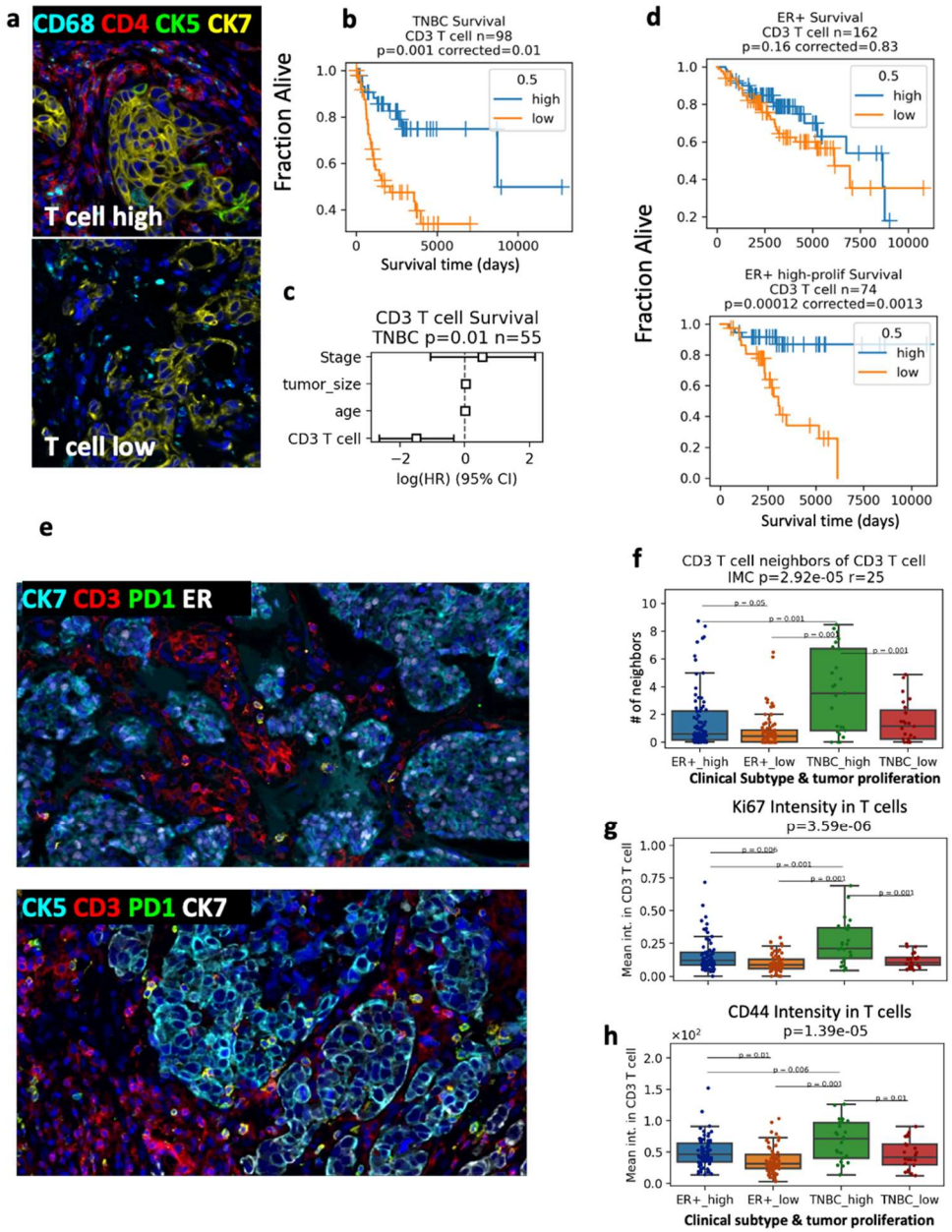


Figure 5. T cell infiltrate has prognostic value in TN and high proliferation ER+ breast cancer.

a. CyCIF images of T cell high and low TNBC tissues. b. Kaplan-Meier analysis of abundance of CD3 T cells versus OS in TNBC (high = above the median) c. Multivariate CPH modeling adding patient age, tumor size and stage to CD3 T cell high variable defined in (b). d. Kaplan-Meier analysis of abundance of CD3 T cell versus OS in all ER+ patients (top) and ER+ patients with high (above the median) tumor proliferation (bottom). b-d. All Kaplan-Meier p-values obtained from the log-rank test and multiple testing corrected with Holm-Šídák method; cutoff quantile for high/low indicated in figure, e.g. 0.5 indicates the median was used to binarize patients. CPH modelling p-values for cell type variable given in figure titles. n number of patients given in figure titles. e. Images of functional state of T cells in ER+ (top) and TNBC tumor (bottom). f. Mean number of T cell neighbors (within 25 μm) of T cells in tissues from high and low proliferation ER+ or TNBC tumors in IMC cohort. g. Ki67 intensity indicating proliferation levels of T cells in tissues from high and low proliferation ER+ or TNBC tumors in IMC cohort. h. CD44 intensity in T cells, indicating memory/effector phenotypes in IMC tissues. f-h. Kruskal-Wallis H-test P-value given in figure title. Post-hoc Tukey HSD used for pairwise comparisons between groups.

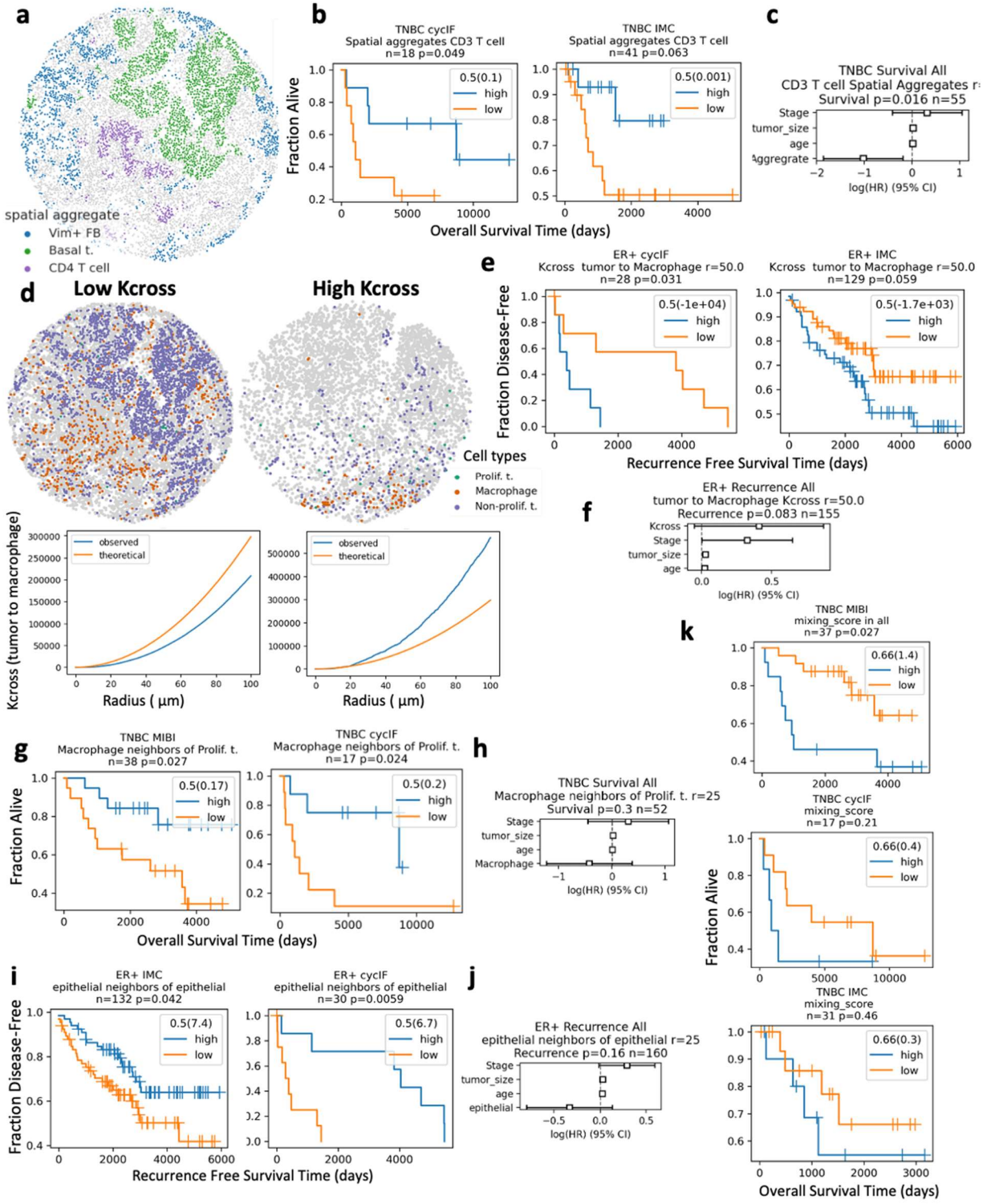


Figure 6. Prognostic tumor-immune spatial correlations in breast cancer cohorts

a. Spatial aggregates in CyCIF data on a TNBC tissue. Gray values indicate cells in areas of tissue with no aggregates. b. Overall survival Kaplan-Meier analysis of TNBC patients with high versus low CD3T cell spatial aggregates in CyCIF (left) and IMC (right) cohorts. c. Multivariate CPH modeling of TNBC patient survival adding patient age, tumor size and stage to CD3 T cell spatial aggregate variable defined in (b). d. Examples of tissues with low (left) and high (right) tumor-to-macrophage Kcross functions. Scatterplots show spatial locations of tumor (purple) and

macrophages (orange) in tissue, relative to all cells (gray). Line plots show theoretical Kcross function values at increasing distance (orange) versus observed values in blue. e. Recurrence-free survival Kaplan-Meier analysis of high versus low tumor-to-macrophage Kcross values at a 50 μm radius in ER+ patients for CycIF (left) and IMC cohorts (right). f. Multivariate CPH modeling of ER+ patient survival with clinical variables and Kcross variable defined in (e.). g-h. Overall survival Kaplan-Meier analysis and CPH modelling of number of macrophage neighbors in a 25 μm radius of proliferating tumor in TNBC cohorts. i-j. Recurrence-free survival Kaplan-Meier analysis and CPH modelling of number of epithelial neighbors in a 25 μm radius of other epithelial cells in ER+ cohorts. k. Overall survival Kaplan-Meier analysis of TNBC patients with high versus low tumor-immune mixing score in MIBI (top), CyCIF (middle) and IMC cohorts (bottom). b-k. All Kaplan-Meier p-values obtained from the log-rank test, n given in figure titles, cutoff quantile for high/low indicated in figure, e.g. 0.5 indicates the median was used to binarize patients. CPH modelling p-values for spatial variable given in figure titles, n number of patients given in figure titles.

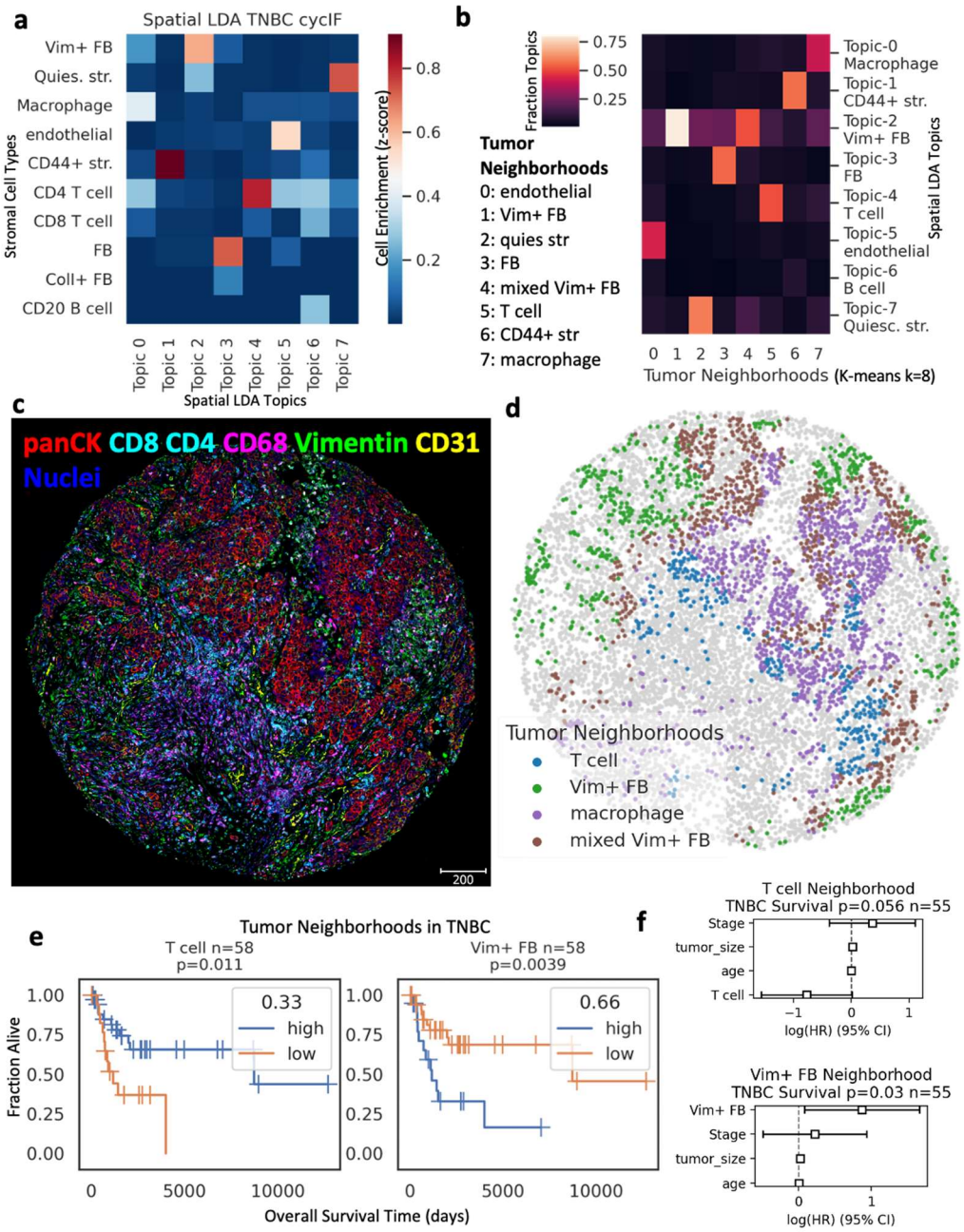


Figure 7. Multicellular neighborhoods surrounding tumor cells in TNBC.

a. Heatmap of stromal cell enrichment in spatial latent Dirichlet allocation (LDA) topic models of 100 μm tumor neighborhoods. b. Heatmap of fraction of each topic in each neighborhood cluster resulting from K-means clustering (k=8) of spatial LDA topics from (a). c. CyCIF staining of tissue showing tumor (panCK), T cell (CD4 and CD8) fibroblast (vimentin), macrophage (CD68) and endothelial (CD31) markers. d. Tissue from (c) with tumor cells colored by their spatial LDA neighborhood cluster from (b). Tumor cells colored by T cell- (blue), macrophage- (purple), mixed fibroblast- (brown) and vimentin+ fibroblast-neighborhoods (green). e. Kaplan-Meier analysis of overall survival versus high or low T cell (left) and vimentin+ fibroblast neighborhoods in TNBC tissues (right). p-values obtained from the log-rank test; n given in figure titles. Cutoff quantile for high/low indicated in figure, e.g., 0.33 indicates the lower tertile was used to binarize patients. f. CPH modelling of overall survival with clinical variables plus spatial LDA neighborhoods from (e). p-values and n number of patients given in figure title.