Distinct tumor-immune ecologies in NSCLC patients predict progression and define a clinical biomarker of therapy response

Sandhya Prabhakaran[1], Chandler Gatenbee[1], Mark Robertson-Tessi[1], Amer A. Beg[2], Jhanelle Gray[2], Scott Antonia[4], Robert A. Gatenby[3] and Alexander R. A. Anderson[1]

[1] Department of Integrated Mathematical Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[2] Departments of Immunology and Thoracic Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[3] Department of Radiation Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[4] Department of Medicine, Duke University School of Medicine, Durham, NC, USA

Table of Contents

1. Abstract *2*

2. Introduction *2*

3. Results *4*

3.1 Non-spatial analyses indicate non-responders (PD) and responders (SD) have significantly different ecologies **4**

3.2 Non-responders have high PD1 and T effector cells (FoxP3) in both the tumor border and tumor regions **5**

3.3 Strong spatial association of T effector cells (FoxP3) with tumor in non-responders during treatment **7**

3.4 Spatial cellular neighbors confirm different neighborhoods are preserved across responders and non-responders. **8**

3.5 Fundamentally distinct architectures across PD and SD patients enable disease progression prediction and clinical biomarker identification **11**

3.5.1. BRT disease progression predictions using quadrats 11
3.5.2. SVM predictions using cell segments 11
3.5.3. Disease progression predictions combining cell segments and quadrats 14

3.6 Inferring underlying networks of SD and PD patients using prior information **16**

4. Discussion *17*

Validation of single cell and quadrat approaches on whole-slide NSCLC images 18
Comparison of single cell and quadrat approaches 20
Future work 21

5. Patient data *21*

6. Code availability *22*

7. Acknowledgments *22*

8. Supplementary *22*

S.1 Figures 22

S.2 Methods 27

9. References *28*

1. Abstract

Lung cancer is the leading cause of cancer death worldwide with 1.76 million people die as a result of the disease yearly. Analyzing high-throughput multiplexed images of tissues has recently emerged as a routine clinical procedure for early detection, clinical cancer diagnosis, treatment planning and prognosis for many cancer types including non-small cell lung cancer (NSCLC). Multiplexed images enable the precise interpretation of spatial distribution of cells and cellular states and the characterization of tumor-immune interactions in situ and at the single-cell level. Through two cornerstone image processing techniques - cell segmentation and image tiling generating quadrats - we perform an in-depth analysis of multiplexed images based on the frequency, phenotype, and spatial distribution of immune and tumor cells within the immune landscape of NSCLC. Multiplexed images were obtained from nine patients with advanced/metastatic NSCLC with progression who were treated with an oral HDAC inhibitor (vorinostat) combined with a PD-1 inhibitor (pembrolizumab). Images were collected from all patients both pre- and on-treatment (days 15-21). We profile ~100K cells and ~10K quadrats from tumor and adjacent tissues to characterize cellular composition and spatial neighborhoods and elucidate that distinct spatial cellular ecologies exist across progressive disease (PD) and stable disease (SD) patients where tumors of PD patients are characterized by a highly suppressed immune environment prior to treatment enabling higher chances of disease progression during treatment. These fundamentally distinct architectures across PD and SD patients enable disease progression prediction and clinical biomarker identification. Our results also provide a potential companion biomarker for PD-L1 in NSCLC. Validating the biomarker and its feasibility will require further in-depth investigation.

Keywords

Non-small cell lung cancer; NSCLC; Multiplexed images; cell segmentation; Species distribution modeling; cellular neighborhoods; quadrats; statistical modelling; prediction; Immune landscape; Immunotherapy.

2. Introduction

Lung cancer is the leading cause of cancer death worldwide with 1.76 million people die as a result of the disease yearly[1]. Non-small cell lung cancer (NSCLC) is the most prevalent type of lung cancer, accounting for about 85% of all cases[2]. Disease progression and treatment response in NSCLC vary widely among patients. Therefore, accurate diagnosis is crucial in treatment selection and planning for each NSCLC patient.

As targeted molecular therapies, immunotherapies, and combination therapies have become central in managing patients with NSCLC, the vital requirement for high throughput data analyses and clinical validation of biomarkers has become even more crucial. Multiplexed imaging of tissues, and their analysis, is an emerging and proficient approach aiding early detection, clinical cancer diagnosis, treatment planning and prognosis[3–6]. Multiplexed images

enable the precise interpretation of spatial distribution of cells and cellular states and the characterization of tumor-immune interactions in situ and at the single-cell level. Further, they allow simultaneous detection of various protein biomarkers on the same tissue sample permitting molecular and immune profiling of NSCLC, while preserving tumor tissue[4,7], and enable the prediction of response to a given treatment. However, image processing, and the subsequent interpretive and predictive tools for multiplexed image data are severely lacking, and thus is an active area of research.

In this study, we used images obtained from nine patients with advanced/metastatic NSCLC with progression who were treated with an oral HDAC inhibitor (vorinostat) combined with a PD-1 inhibitor (pembrolizumab). Images were collected from all patients both pre- and on-treatment (days 15-21). We developed a computational multiplexed-image analysis pipeline using two main state-of-the-art image analysis approaches – cell segmentation and quadrats - to analyze the spatial and temporal features of multiplexed NSCLC images, predict disease progression and identify clinical biomarkers (**Figure 1**). Using cell segments, we can learn relevant and distinct spatial neighborhoods, and interactions therein, at the cellular level. Using quadrats, we can infer spatial ecologies without the need of manual cell annotations using only patient-level labels. Combining both these approaches allows us to garner insights across multiple scales in the data, enables progression predictions based on the abundances and diversity of discovered marker phenotypes and neighborhoods, and provides clinicians with interpretable clinical decisions, since the predictions can be mapped back onto the original images generating '*risk maps'* (**Section 3.4** and **Figure 6**).

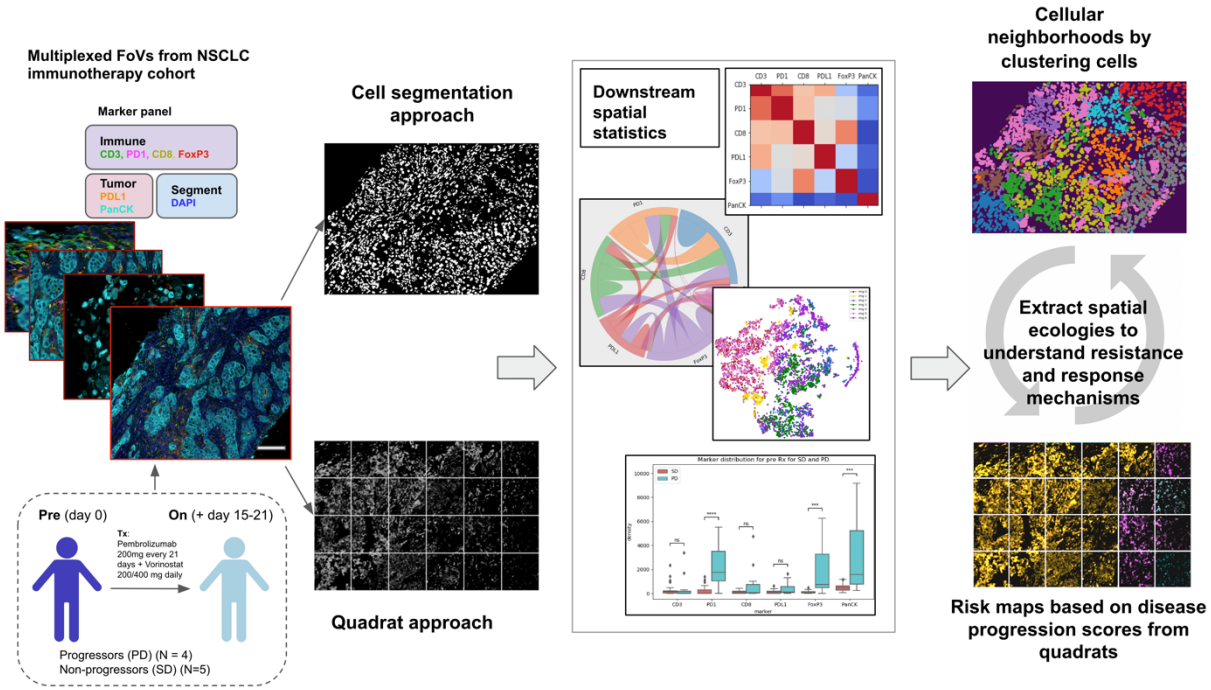


Figure 1: Graphical abstract showing the two computational approaches developed in this paper to study spatial and temporal interactions of tumor and immune cells using multiplexed images from Moffitt's NSCLC immunotherapy trial cohort. A longitudinal cohort of nine NSCLC patients were studied before (PRE) and during (ON) treatment. For the single cell approach, we extract the cell segments per Field of View (FoV) using a U-Net[8], build a count matrix with cells as rows and markers as columns, and cluster the count matrix to identify spatially heterogeneous cell types. For the quadrat approach, we view each patient sample as a collection of fixed-size grids, called quadrats to study the marker distribution and to detect global changes across the spatially diverse cell types. This approach is inspired by species distribution modeling (SDM), a popular tool in quantitative ecology and used by landscape ecologists[9,10]. Using an array of downstream analyses using spatial statistics, we identify key cellular neighborhoods distinguishing PD versus SD patients and can generate risk maps that aid interpretable clinical decisions. (Scale bar = 200 pixels/mm)

3. Results

3.1 Non-spatial analyses indicate non-responders (PD) and responders (SD) have significantly different ecologies

To understand the underlying patterns of tumor evolution in NSCLC amongst PD and SD patients and how this is affected by treatment, we start by analyzing the spatial density of markers across the FoVs using basic ecological analysis for example, average quadrat counts. In **Figure 2A**, an example 7-marker NSCLC Field of View (FoV) (with tumor regions in cyan) is shown (**i**) along with its corresponding cell segments (**ii**) and quadrats (**iii**). We plot the marker abundances (based on average positive pixel counts (PPC)) across all the quadrats of PD and SD patients, before and during treatment (**Figure 2Bi**). It is evident that PD and SD have significantly different marker abundances, with PD showing increased marker levels throughout (pre and during treatment). We further investigate the difference in PD and SD marker densities by looking at the CAP[11] (canonical analysis of principal components) projection of the average positive pixel counts (PPC) for each FoV (**Figure 2Bii**). The PERMAVOVA[12] (permutational analysis of variance) test performed on the ecological dissimilarity matrices (computed using PPCs of the quadrats) depicts that the centroids or spread of each group are different; this reinforces that PD and SD have significantly different ecologies (p-value = 0.001). This also indicates that treatment potentially has an effect and is successfully altering the tumor ecology, as shown by the arrows in **Figure 2Bii**, which indicate how each groups' centroid shifted while on treatment. Such significant differences in composition suggest that the pre-treatment ecology may be distinct enough between PD and SD to be predictive of response. Since these results were derived from basic non-spatial analyses, we further investigated the image cohort for differences pertaining to the spatial organization of the tumor with respect to its microenvironment.

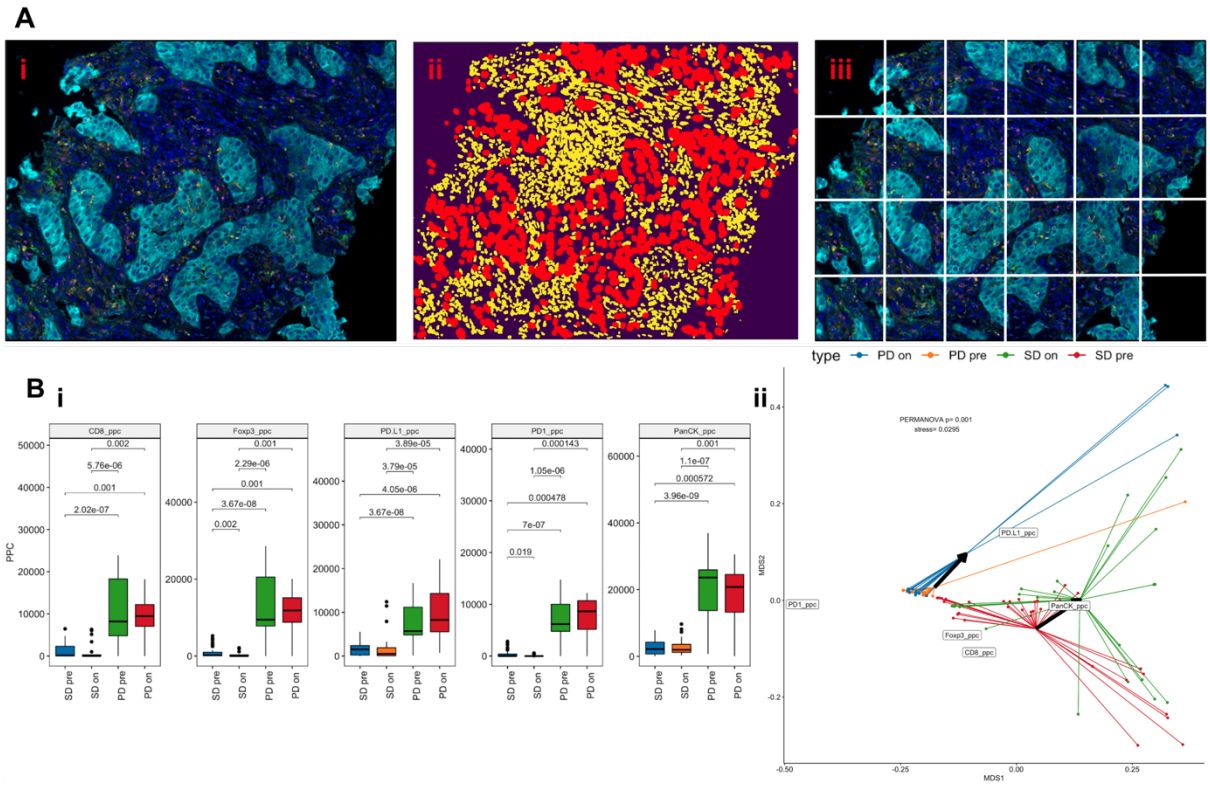


Figure 2: (A) (i) an example NSCLC FoV which is a 7-stain Vectra image with DAPI (blue; nuclei), CD3 (green; T cells), CD8 (yellow; T effector cells), FoXP3 (red; T regulatory cells), PD-1 (pink; inhibitory receptor on immune cells), PD-L1 (orange; immune checkpoint marker), and pan-cytokeratin (PanCK, cyan; epithelium). Corresponding

image showing **(ii)** cell segments (tumor cells shown in red) and **(iii)** image quadrats. **(B) (i)** Bar plots showing marker abundances (based on average positive pixel counts (PPC)) across SD and PD patients, before and during treatment, for each marker **(ii)** The CAP (canonical analysis of principal components) projection of the average positive pixel counts (PPC) across all FoVs. Shown also is the PERMAVOVA (permutational analysis of variance) test indicating that the centroids or spread of each group are different (note black arrows) reinforcing that PD and SD have significantly different ecologies.

3.2 Non-responders have high PD1 and T effector cells (FoxP3) in both the tumor border and tumor regions

Using our multiplexed imaging pipeline for single cells (**Figure 3i**), we extract the cell segments per FoV using a U-Net[8], build a count matrix with cells as rows and markers as columns, and cluster the count matrix to identify heterogeneous cell types. From these cell types, we automatically demarcate tumor-rich regions, across images, that are higher in PanCK expression (red regions in **Figure 3ii**, right panel). The tumor regions are approximated using multiple convex hulls to allow for spatial quantification of tumor and immune cells within and outside the tumor border. We note that for both pre and on treated patients, there is a higher presence of immune cells within the tumor (**Figure 3iii**) and a higher abundance of tumor and immune cells across the tumor border (**Figure 3iv**) in PD than SD patients signaling that there are distinct cellular compositions for PD and SD patients.

Next, to quantify the tumor-immune cell colocalization at the tumor border we cluster the tumor-immune cell counts at the tumor border approximated using convex hulls (see **Figure 3ii,** right panel). Clustering is performed using a Gaussian mixture model[13] (GMM). The input matrix to the GMM consists of cells (as rows) and their marker distribution as features (columns). The rows of the input matrix are ordered based on the cluster assignments and the z-score of the marker expression (columns) is averaged over vectors per cluster representing a cell type (rows). Those markers that have a higher z-score per cluster are identified as the differentially expressed markers for that cluster. The clusters are visualized using a 2D t-SNE plot where each point abstracts an image (**Supplementary Figure 1A, B**). We obtain distinct clusters showing higher colocalization of PanCK+PD-1+FoxP3+ for PD patients and higher colocalization of PanCK+PD-L1+ along with CD3+CD8+ immune cells for SD patients indicating underlying structural differences between PD and SD patient groups. Further, to better understand the relationship of the clusters to the actual images, we generate an image t-SNE (**Supplementary Figure 1C**) where we replaced each dot in the t-SNE of **Supplementary Figure 1A, B** with the corresponding multiplexed images. This image t-SNE was rendered using Mistic[14] and this arrangement of images as an image t-SNE clearly highlights the difference in immune cell abundance across PD and SD patient groups. The image t-SNE captures the immune cell gradient showing images higher in PanCK+PD-1+FoxP3+ colocalization (a higher abundance of green) to be generally observed across PD patients and images higher in CD3+CD8+PanCK+ colocalization (a higher abundance of yellow) to be generally observed across SD patients. The image t-SNE also helps in identifying groups of immunologically hot, warm or cold tumors across patients based on immune cell colocalization at the tumor margins and tumor regions.

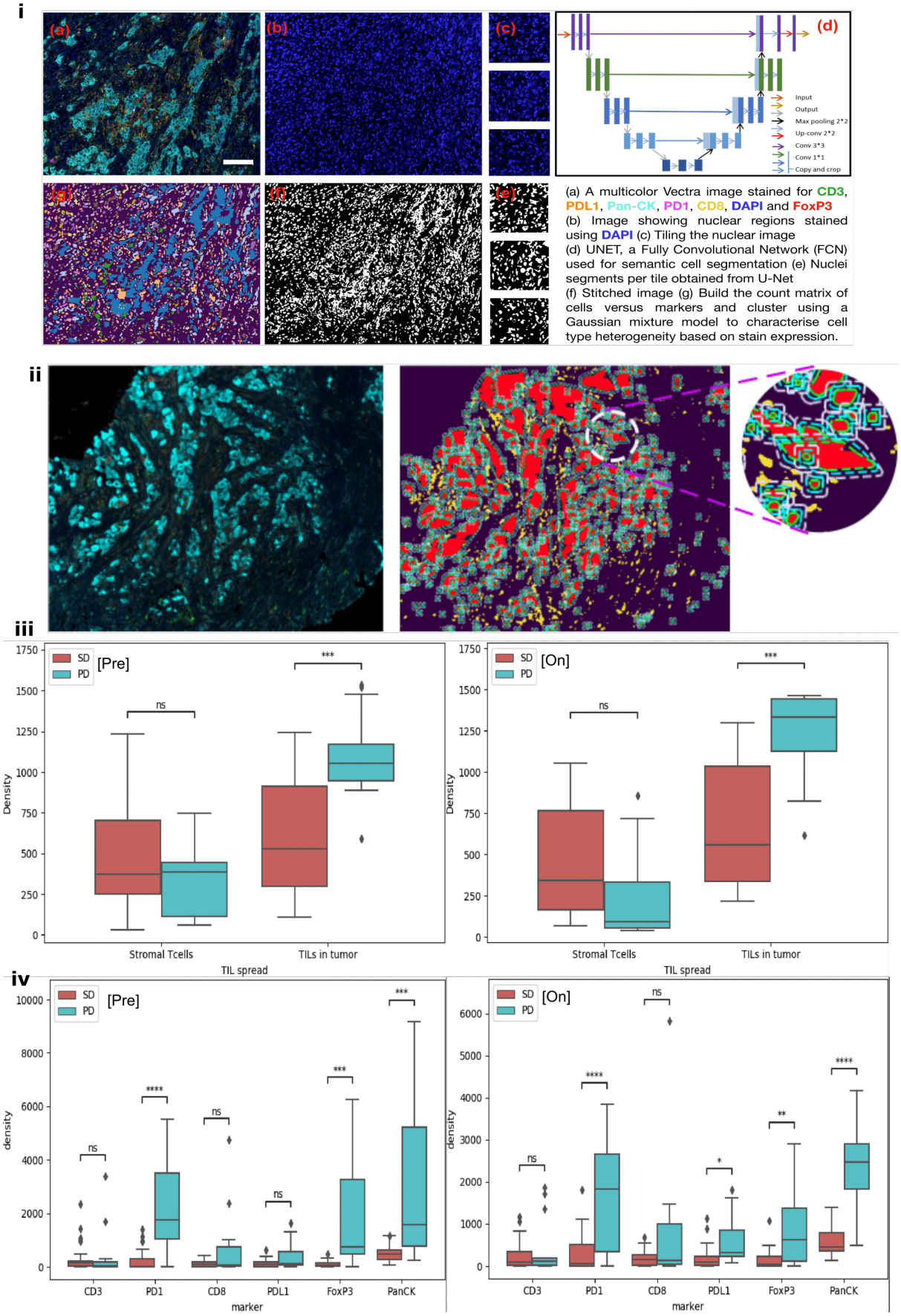


Figure 3: **(i)** Image processing pipeline: (a) the input is an original 7-stain Vectra image with DAPI (blue; nuclei), CD3 (green; T cells), CD8 (yellow; T effector cells), FoXP3 (red; T regulatory cells), PD-1 (pink; inhibitory receptor

on immune cells), PD-L1 (orange; immune checkpoint marker), and pan-cytokeratin (PanCK, cyan; epithelium) (b) the image with corresponding DAPI stains showing nuclei is extracted. The nuclear-stained image is tiled (c) and fed into a deep-learning architecture called U-Net[8] (d) to identify cell segments. (e) U-Net identified cell segments per tile which are then stitched to create the image in (f). A count matrix with cells (rows) and markers (columns) is computed from (f) and is used for clustering to identify heterogeneous cell types. In (g) cell segments are colored based on the cluster assignments obtained through a Gaussian mixture model. Scale bar = 200 pixels/mm.
Quantifying the tumor-immune cells at the tumor border using convex hull approximation: **(ii)** (Left) An example Original 7-color Vectra image multiplex FoV, (Middle) computationally identified tumor regions (in red) and their boundary approximations using convex hulls. Multiple convex hulls are drawn extending into and outside the tumor region to aid quantifying the immune-tumor cells colocalized at the tumor boundary. Inset: Zoomed in region showing tumor regions in red and three convex hull boundaries in dotted lines.
(iii) (Left) Box plots showing the presence of Stromal immune cells and immune cells within tumors in SD and PD patients in pre-treatment and (Right) on-treatment cases. For both pre and on treated cases there is a higher presence of immune cells in tumor in PD than SD patients. The Bonferroni-adjusted p-value with significance are: *: 1.00e-02 < p <= 5.00e-02, **: 1.00e-03 < p <= 1.00e-02, ***: 1.00e-04 < p <= 1.00e-03, ****: p <= 1.00e-04.
(iv) (Left) Box plots showing number of immune and tumor cells in the tumor border in SD and PD patients pre-treatment and (Right) on-treatment cases. For both pre and on treated patients there is a higher presence of immune cells (PD-1, FoxP3) with tumor (PanCK) in PD than SD patients. The p-values are Bonferroni-adjusted p-values with significance as in (iii).

3.3 Strong spatial association of T effector cells (FoxP3) with tumor in non-responders during treatment

In a parallel and complementary image-analysis approach, we look into species association networks[15–17] (SANs) that are a popular tool in quantitative ecology. SANs provide a modelling framework to extract direct spatial associations based on spatial co-localization, while removing indirect effects. The fundamental modeling steps in SANs are the data preparation, model fitting and assessment, and prediction. We view each patient sample as a collection of fixed-size grids, called quadrats, defined as partitions of an image into multicellular regions creating sequential tiles in order. Each quadrat is 100um x 100um in size and expresses the area normalized sum of marker intensity values of all cells present within that quadrat. The quadrats are combined and fit using a Gaussian copula graphical lasso (GCGL) using the *ecocopula* package for *R*[18,19] for the pre- and on-treatment groups (**Figure 4A**). Since the copula can directly model covariance and strength between markers, the GCGL creates a single association network with all direct links between markers and strength of association increases from red (negative/dispersed) to blue (positive/clustered). Associations were also clustered using the Leiden algorithm[20] so as to help identify spatial communities, as shown by the larger polygons under sets of markers. We note that the GCGL captures an increased influence of FoxP3 with tumor cells in PD patients during treatment and an increased influence of CD8 with tumor cells in SD patients during treatment.

We plot the marker abundance across all the FoV quadrats per patient response category (**Figure 4B**) and note that PD-1, FoxP3 and PanCK have a higher presence in PD (pre and on) than in SD (pre and on). Next, we cluster the quadrats using GMM and plot the t-SNE renderings (**Figure 4C i, ii**) and the individual marker expression spread (**Figure 4C iii).** It is again evident that there are distinct ecologies defining PD and SD patients such as PanCK+FoxP3+PD-1 (gray cluster) for PD patients and PanCK+CD3+CD8 (red cluster) for SD patients. By collectively examining each marker's importance, response, and interactions, a picture emerges wherein a distinguishing factor between PD and SD is that the tumors of the former are characterized by a more highly suppressed immune response prior to treatment. This in turn may help explain why these patients have a higher probability of progressing while on treatment.

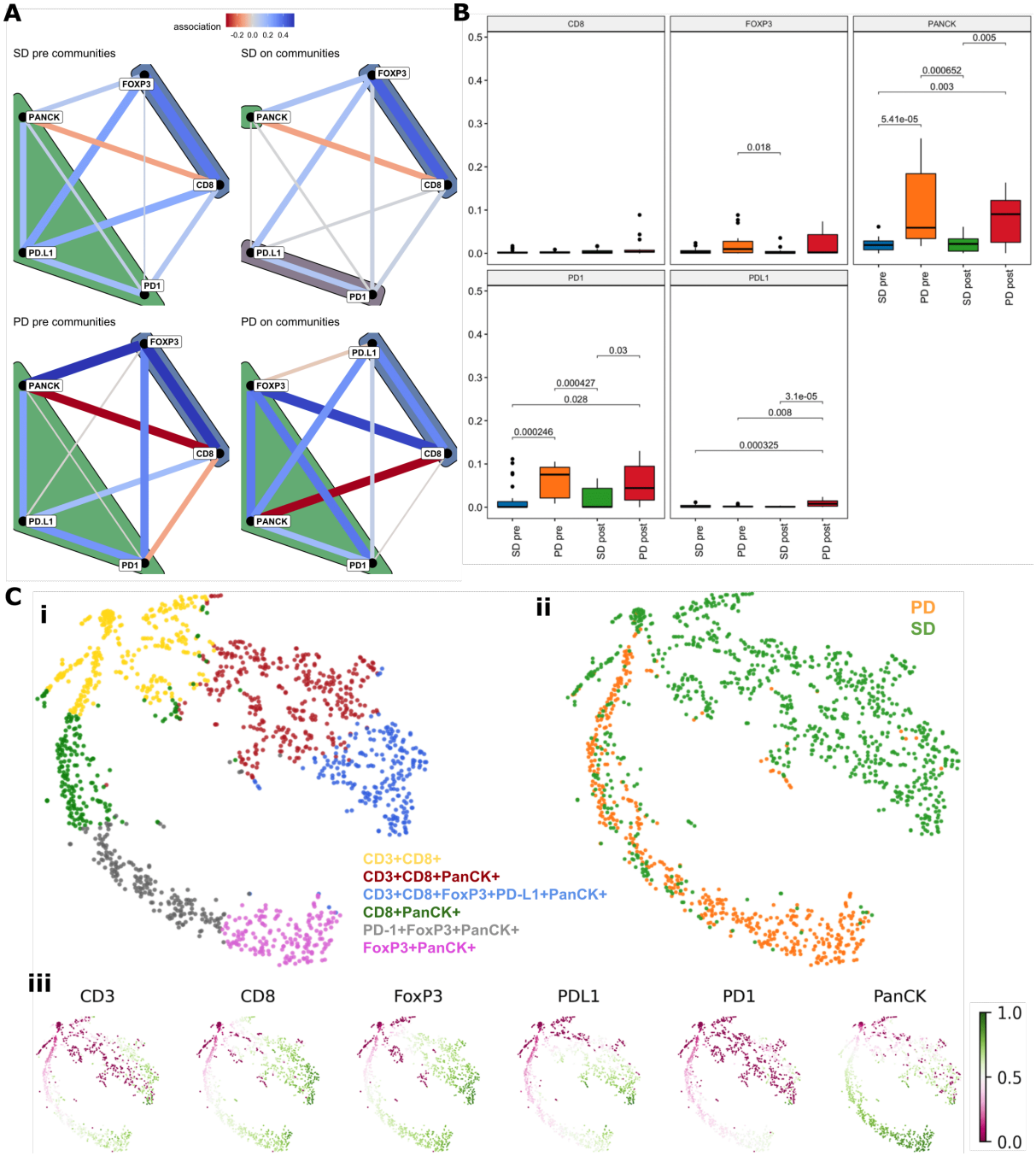


Figure 4: Spatial association of all markers across the four patient categories **A.** Optimal network of marker association obtained using a Gaussian copula graphical model (GCGM) shown per patient category. Spatial associations were clustered, as indicated by the underlying polygons **B.** Marker abundance across all the FoV quadrats shown per patient response category. PD-1, FoxP3 and PanCK have a higher presence in PD (pre and on) than in SD (pre and on). **C. (i)** 2d t-SNE rendering of all the quadrats (~5,000) colored by clusters (left panel), **(ii)** by patient category (right panel) and **iii**. Same t-SNE (as in **i**) depicting expression spread of the six markers.

3.4 Spatial cellular neighbors confirm different neighborhoods are preserved across responders and non-responders.

To further understand the spatial organization of tumor and immune cells across PD and SD patients, we investigate each cell in the context of its spatial neighbors. This is done by generating cellular neighborhoods[21] (CNs) where the marker expression of each cell is the

average of ten of its nearest spatial neighbors in Euclidean space. A cellular neighborhood can be defined as the minimal set of cell types that are both functionally and spatially similar. This approach is inspired by the cellular neighborhoods[21] with the major difference that we do not have a fixed set of cell types at the onset but work with cell types that are organically inferred through the clustering process.

By clustering cells from all FoVs along with their spatial neighbors, we obtain twelve distinct CNs (**Figure 5A**). An example of the spatial orientation of the CNs is shown in **Figure 5B** (lower panel). The t-SNE rendering of these CNs shows that there are characteristic cellular neighborhoods defining PD and SD patients (**Figure 5C i, ii**) such as PanCK+FoxP3+PD-1+ for PD patients and PanCK+CD3+CD8+ for SD patients (**Figure 5D**, associations are statistically significant using Bonferroni-corrected p-values). The individual marker expression spread is shown in **Figure 5C iii.** In **Supplementary Figure 2**, we show the marker co-expression covariances and corresponding t-SNE expressions captured by the CNs across patient categories. It is evident that prior to treatment PD patients show increased interactions between immune cells and tumor cells and tighter interactions between CD3, CD8, PD1, and PDL1 during treatment than SD patients.

It has been previously shown that infiltration of intra-tumoral CD3+ and CD8+ T cells in NSCLC is associated with better survival outcome[5,22,23]. It is also long been suggested that FoXP3 is a barrier to the efficacy of ICIs and therefore a potential target for therapy[24,25]. Further, the study of NSCLC patients[24] show that the spatial proximity of tumor and T regulatory cell predicts worse lung cancer survival. These previous studies corroborate our findings in that specific cellular compositions exist between PD and SD patients.

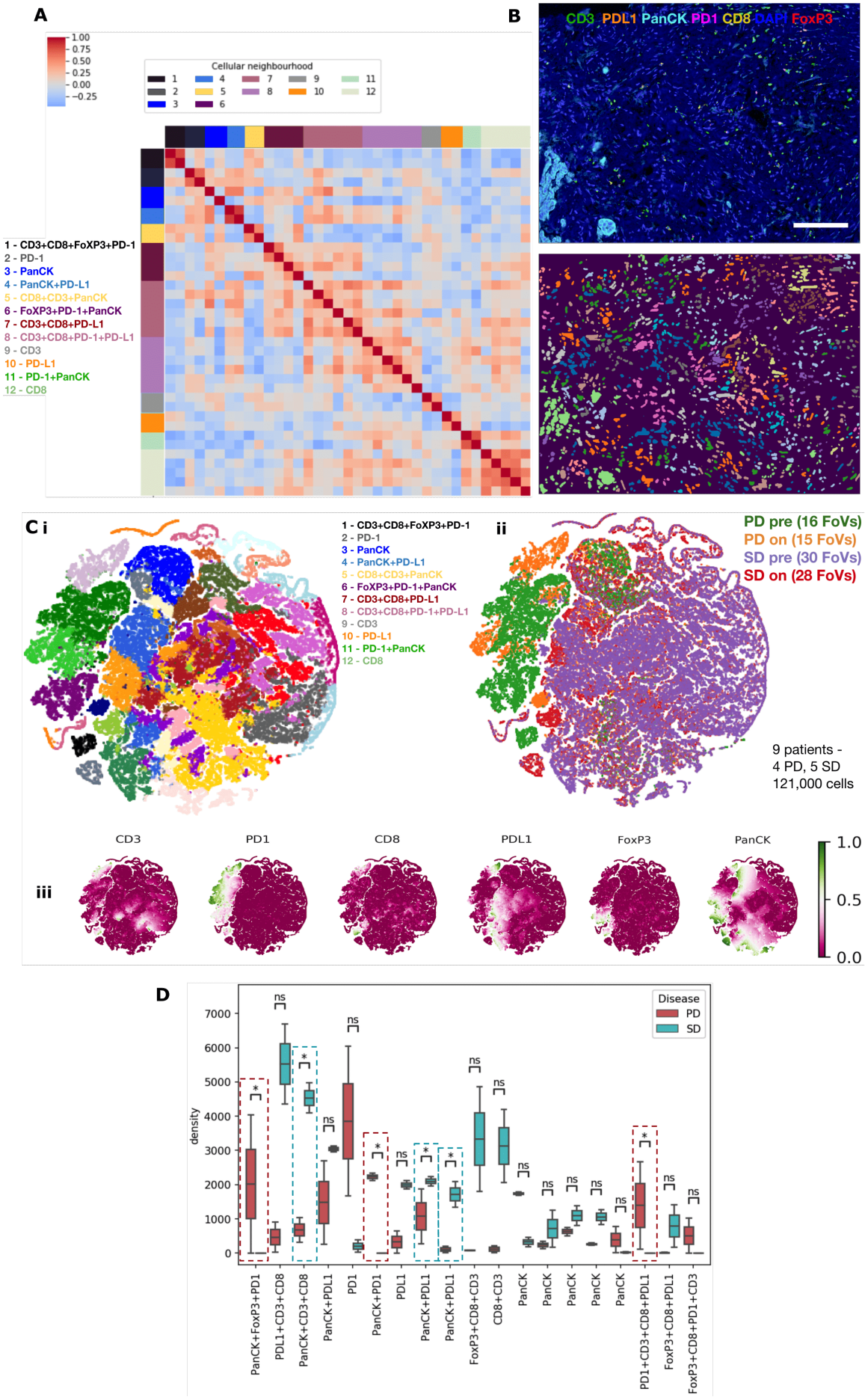

A
B
Cellular neighbourhood
1.00
0.75
0.50
0.25
0.00
−0.25
1 - CD3+CD8+FoXP3+PD-1
2 - PD-1
3 - PanCK
4 - PanCK+PD-L1
5 - CD8+CD3+PanCK
6 - FoXP3+PD-1+PanCK
7 - CD3+CD8+PD-L1
8 - CD3+CD8+PD-1+PD-L1
9 - CD3
10 - PD-L1
11 - PD-1+PanCK
12 - CD8
CD3 PDL1 PanCK PD1 CD8 DAPI FoxP3
C i
ii
PD pre (16 FoVs)
PD on (15 FoVs)
SD pre (30 FoVs)
SD on (28 FoVs)
9 patients -
4 PD, 5 SD
121,000 cells
iii
CD3
PD1
CD8
PDL1
FoxP3
PanCK
1.0
0.5
0.0
D
Disease
PD
SD
density
7000
6000
5000
4000
3000
2000
1000
0
ns
*
PanCK+FoxP3+PD1
PDL1+CD3+CD8
PanCK+CD3+CD8
PanCK+PDL1
PD1
PanCK+PD1
PDL1
PanCK+PDL1
PanCK+PDL1
FoxP3+CD8+CD3
CD8+CD3
PanCK
PanCK
PanCK
PanCK
PanCK
PD1+CD3+CD8+PDL1
FoxP3+CD8+PDL1
FoxP3+CD8+PD1+CD3

Figure 5: Diverse cellular neighborhoods captured across patient categories. **A.** Heatmap showing 12 distinct cellular neighborhoods (CNs) based on six markers and their respective z-scored frequencies within each CN (pooled data across both SD and PD groups and across pre and on treatment). **B.** (Top) Exemplar multiplex FoV in a seven-color overlay image with DAPI (blue; nuclei), CD3 (green; T cells), CD8 (yellow; T effector cells), FoXP3 (red; T regulatory cells), PD-1 (pink; inhibitory receptor on immune cells), PD-L1 (orange; immune checkpoint marker), and pan-cytokeratin (PanCK, cyan; epithelium). Scale bar = 200 pixels/mm. (Below) FoV showing the twelve clustered CNs mapped. **C.** 2d t-SNE rendering of all the cells (~121,000) colored (i) by CNs, (ii) by patient category and (iii) expression spread of the six markers. Note that the mapping between colors and the CNs are the same in A, B (below) and C(i) **D**. Box plot quantification of frequencies (in terms of number of cells) of CNs in PD and SD patients showing statistically significant CNs (Bonferroni corrected, p-value < 0.001).

3.5 Fundamentally distinct architectures across PD and SD patients enable disease progression prediction and clinical biomarker identification

We use images from the pre-treatment cases to train two separate ML classifiers - a binary support vector machine[26] (SVM) classifier trained using cell segments and a boosted regression tree[27,28] (BRT) using quadrat counts (100um x 100um) - to separate data into two classes: “PD/progression,” and “SD/non-progression/stable”.

3.5.1. BRT disease progression predictions using quadrats

Boosted regression trees are augmented regression trees that can associate a response variable (such as disease progression) with predictor variables (such as marker expressions) by recursively splitting and combining parsimonious trees to generate disease progression predictions. Using these quadrats with marker intensities, we trained the BRT to predict whether that quadrat came from a patient that progressed while on treatment, allowing us to create a ‘risk map’ that identifies areas of the tumor associated with high probabilities of progression while on treatment (**Figure 6i**). The generation of risk maps suggests that spatial measurements of the tumor ecology, and thus the interactions between cell types, can be used to make accurate predictions of how a patient will respond to therapy.

For each of the nine patients, the trained BRT can predict the probability of disease progression based on the image quadrats 91.4% of the time (**Figure 6ii A**). In addition to prediction response, the trained BRT provided numerous insights into what distinguishes responders from non-responders. By performing a feature importance analysis, it is possible to determine the degree to which each marker (each being a feature used as model input) affects the ability of the BRT to make accurate predictions. After performing this analysis, we found that PD-L1 is the most important variable in predicting response, closely followed by PanCK (tumor) and PD-1 (**Figure 6ii B**). By examining the response of each variable (marker) as shown in **Figure 6ii C**, we then see that high levels of PD-1 and PanCK are associated with higher probabilities of progression, while high levels of PD-L1 are associated with lower probabilities of progression. Finally, it is also possible to examine the interactions of the markers, based on their co-localization within each quadrat (**Figure 6ii D**), and we observe that the mix of PanCK and PD-L1, and the mix of PD-L1 and PD-1 have the strongest impact on the model’s predictions.

3.5.2. SVM predictions using cell segments

The SVM initially maps each data point into a 6-dimensional feature space where six is the number of markers used. It then identifies the hyperplane that separates the data into the two

classes while maximizing the marginal distance for both classes and minimizing the classification error[26]. The marginal distance for a class is the distance between the decision hyperplane and its nearest instance which is a member of that class. Due to this formulation, SVM yields better generalization performance.

We implemented a neural network based SVM in Keras[29,30] where we use the "Sequential model" option which is a linear stack of neural layers. Specifically, we use 64 dense 2D layers with a Rectified Linear Unit[31] (ReLU) as the activation function, and the final 2D dense layer has a sigmoid activation with a L2 kernel regularizer[13]. The model is trained for 20 epochs using a categorical hinge loss function and optimized using the Adadelta[32] stochastic gradient descent optimizer.

Our SVM classifier is trained on 47,000 pre-treatment cells where the cells are randomly split into training, validation and test set. The classifier is trained on the training set and predictions of disease progression are made using the test set. The response of the patients was known a priori. **Figure 6iii** depicts the UMAP projection of the pre-treatment data showing (a) patient data response and (b) SVM predictions. The predictions are averaged across 10 independent runs by randomly shuffling all the test data points for each run. Classification accuracy is the fraction of correct predictions per run. We obtain an overall accuracy of 94.25%, with class-specific accuracy of 95.4% for SD patients and 93.1% for PD patients.

<u>Both the BRT and SVM-based prediction experiments with high accuracy are suggestive that distinct ecologies are inherently present in the data and enable the predictions, and are not artefacts of the prediction algorithms themselves.</u>

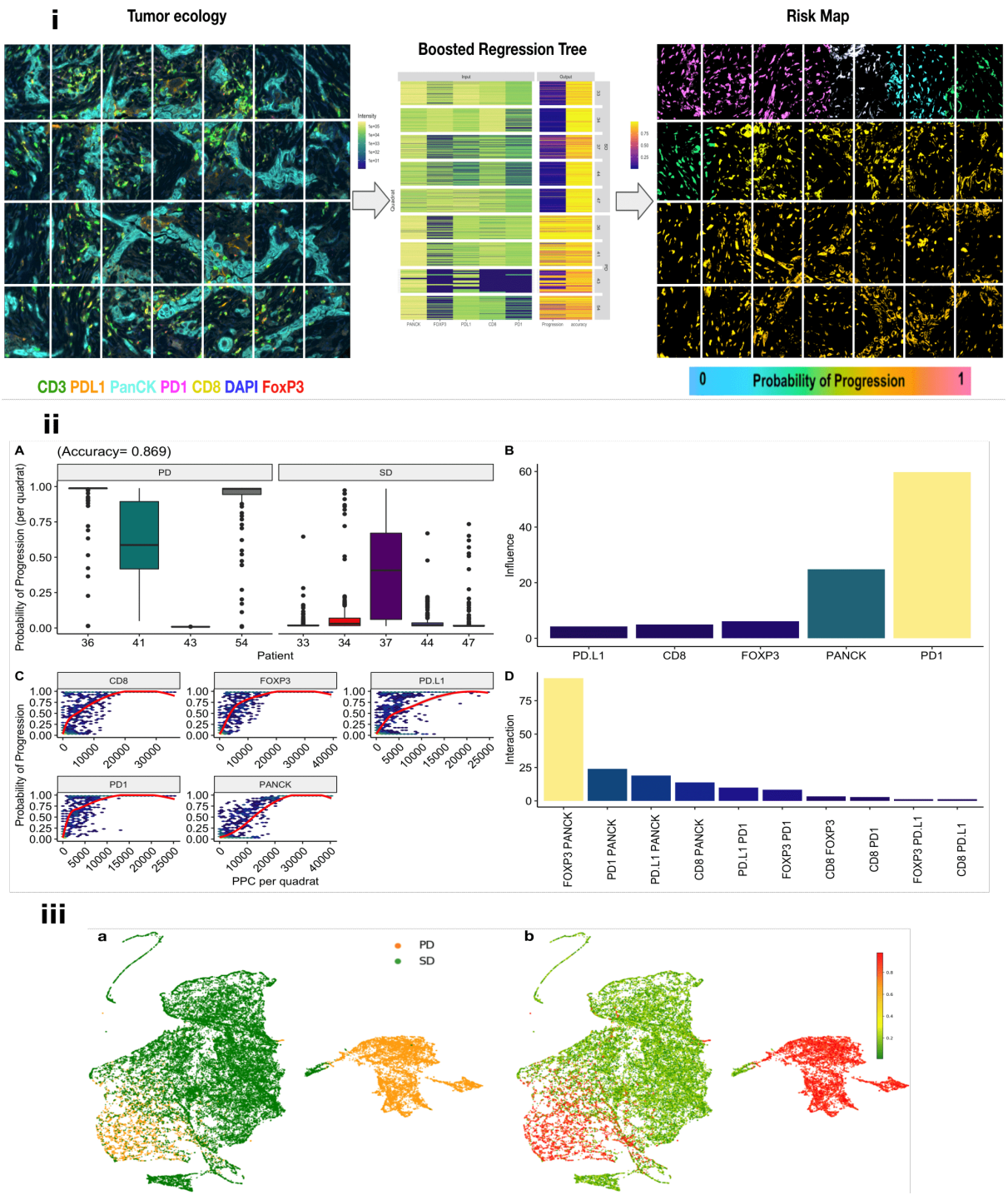


Figure 6: (i) The spatial co-localization of cells within each 100 µm x 100µm quadrat (outlined in white) from pre-treatment biopsies (left) where used to train a boosted regression tree (BRT) that predicts the probability a patient will progress while on treatment (middle). The BRT prediction probabilities are projected back on to the image, potentially identifying regions of disease progression (right). The images here are from one FoV from a patient that had progressive disease (PD). **(ii) (A)** The BRT was able to correctly predict whether a quadrat came from a patient with progressive disease (PD) 91.4% of the time. The distribution of quadrat probabilities could be used to assess risk of progression, as on average patients with stable disease (SD) had lower quadrat probabilities than PD patients. **(B)** A feature importance analysis revealed that PD-L1, PD-1, and PanCK had the greatest impact on the BRT's ability to accurately predict treatment response, suggesting these markers play the most important roles in determining whether a patient responds to treatment. **(C)** Variable response plots reveal that risk of progression increases with higher levels of tumor (PanCK) and PD-1 but decreases as levels of PD-L1 increase. **(D)** Analysis of interaction effects between variable (marker) values show that the mix of PanCK & PD-L1 and PD-L1 & PD-1 within each quadrat have strong effects on the BRT's prediction, suggesting that these pairs of markers strongly

influence whether or not a patient responds to therapy. (**iii**) Disease progression predictions using SVM trained on cells from pre-treatment images. UMAP projection of pre-treatment data showing **(a)** patient data response, and (**b**) SVM predictions with overall accuracy= 94.25%, accuracy for SD patients = 95.4% and for PD patients = 93.1%.

3.5.3. Disease progression predictions combining cell segments and quadrats

In this section, we describe cross over experiments between the cell segmentation and quadrat approaches to further reinforce the distinct architectures existing between PD and SD patients. For this, we use images from only the pre-treatment cases and redefine the quadrat to two different levels of granularity: a finer granularity where each quadrat is represented as a cell with its neighborhoods and a coarser granularity where each quadrat is a fixed-size grid with a collection of cells and their neighborhoods.

In the finer granularity case, we cluster the cells across all quadrats from the pre-treatment cases, and project the cells onto a 2d t-SNE, to obtain a clearer distinction between PD (orange) and SD (green) regions (**Figure 7Aa**). Next, to depict disease prediction using this data, we train BRTs as discussed in the previous section. We test a BRT in a cross-validation setting. For this, cells are randomly split into training, validation and test and the training set is used as input to the regression trees. BRT is used to make predictions of disease progression using the test set. This process of training and generating predictions is iterated, ensuring that all cells have partaken at least once in the validation and test sets. The disease progression prediction scores obtained from the BRT are averaged using Lowess smoothing[33,34] and are shown in **Figure 7Ab**. This overlaps perfectly with the PD and SD regions in **Figure 7Aa** showing high progression probability in orange/red corresponding to the PD region. The nine-patient spread is shown in **Figure 7Ac**. Note also in the marker spread (**Figure 7Ad**), the marker triplet (PD-1+PanCK+FoxP3) is differentially expressed in the PD region, indicating a highly suppressed immune environment, aiding disease progression.

In the coarser granularity setting where quadrats are sequential tiles of the images, we use them to predict disease progression using a BRT as discussed previously. The quadrat level probability of progression scores is used to color the quadrats for each FoV generating a risk map (refer **Figure 6i**). In **Figure 7B**, we depict these risk maps, arranged in the image t-SNE layout (as discussed in Section 3.2), where this particular layout is obtained by clustering tumor-immune cell counts at the tumor border. We observe that most of the PD images are colored in yellow correctly identifying these images, and therefore the patients, to have higher chances of progression. Similarly, SD images have pink/blue hues denoting lower chances of progression. In **Figure 7C**, the correlation matrix pair plots show relationships between both quadrat and cells, as well as for PD using the biomarker set of PanCK+FoxP3+PD-1+ and for SD for PanCK+CD3+CD8+. This allows us to see correlations across multiple variables. For example, the diagonal captures the similar distributions present between PD quadrats and cells as well as between SD quadrats and cells.

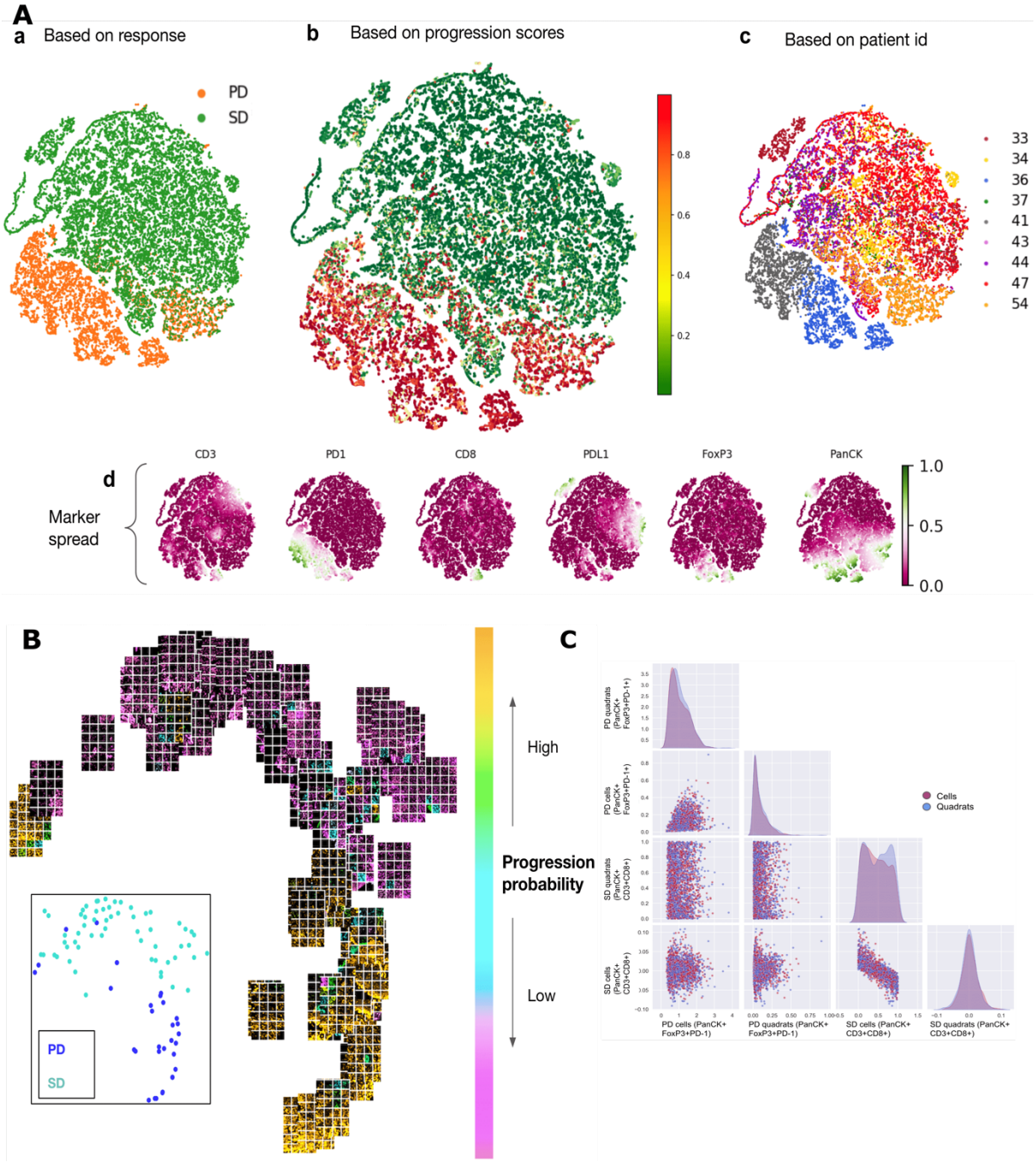


Figure 7: A. Combining cellular neighbors and quadrats using images from pre-treated patients where each quadrat is represented as a cell with its neighbors. Clustering the cellular neighborhoods and plotting the t-SNE showing **(a)** patient response, **(b)** progression scores predicted using boosted regression trees, **(c)** patient spread, and **(d)** the marker spread.

B. Combining cellular neighbors and quadrats using images from pre-treated patients where each quadrat is an image tile of 100um x 100um. Here shown is the image t-SNE with the risk maps per FoV where the quadrats for each FoV are colored using the probability of progression scores. Most of the PD images are colored in yellow correctly identifying these images, and therefore the patients, to have a higher chance to progress. Similarly, SD images have pink/blue hues denoting lower chances to progress. Inset: the t-SNE as in Supplementary **Figure 1B** showing PD and SD patient spread.

C. Correlation pair plots between pre-treated quadrats and cells across PD and SD patients for the biomarker sets of PanCK+FoxP3+PD-1+ and for SD for PanCK+CD3+CD8+, respectively.

3.6 Inferring underlying networks of SD and PD patients using prior information

In order to quantify the distinct architectures across PD and SD patients, we build a statistical model to infer the difference in marker interactions, in the form of a network, between these patient categories. This is done by utilizing the marker expression both at the pre-treatment and during-treatment states for each category and capturing the '*difference network*' that potentially drives the patients from the pre-treatment state to the on-treatment state. The inference problem can be stated as a least-squares minimization problem[13] and can be further regularized using prior biological knowledge of marker interactions.

<u>Problem formulation</u>

We denote the change in the ith marker expression in the pre-treatment case at time *t*, $m_i^{pre}(t)$ as:

Equation 1:

$$\frac{dm_i^{pre}(t)}{dt} = \sum_{j=1}^{n} W_{ij}\ m_j^{pre}(t) - m_i^{on}(t)$$

where *n* is the number of markers, *i, j* = 1, 2, …, *n*, and *W* is the weight matrix. Equation 1 can be written in matrix form as:

Equation 2:

$$\frac{dM^{pre}(t)}{dt} = WM^{\,pre}(t) - M^{on}(t)$$

where $M^*(t) = m_1^*(t), m_2^*(t), \ldots, m_n^*(t)$ and * denotes pre or on states. We can solve Equation 2 as a least-squares minimization problem given *n* markers and (M_i^{pre}, M_i^{on}) as steady states. This can be written as:

Equation 3:

$$W^{LS} = arg\min_{W} f(W) = arg\min_{W} \sum_{j=1}^{n} ||WM^{pre} - M^{on}||$$

Further, we can add prior knowledge of biological mechanisms between markers as additional constraints to the least-squares problem. These mechanisms can be captured in pairwise format as a matrix W^{prior} **(Figure 8A)** where a 1 indicates positive interaction, 0 indicates no interaction and -1 indicates negative interaction. Equation 3 can be written as follows to yield $W^{LSprior}$:

Equation 4:

$$W^{LSprior} = arg\min_{W} f(W) = arg\ \min_{W} \sum_{j=1}^{n} ||WM^{pre} - M^{on}|| + \lambda||W\ o\ W^{prior}||$$

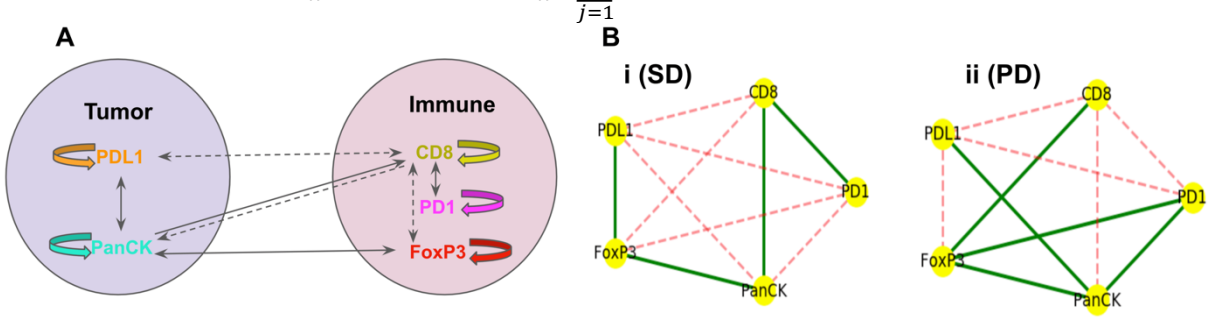


Figure 8: (A) W^{prior} matrix that captures prior biological interactions between markers (dotted links; negative interaction, solid links; positive interaction), **(B)** $W^{LSprior}$ depicted as a network for SD (i) and PD (ii) obtained by solving Eqn 4. $W^{LSprior}$ maps the markers from the 'pre' to the 'on' state while incorporating known mechanisms as prior information. Green links indicate inferred positive marker associations and red lines indicate inferred negative associations.

Implementation

We construct weighted count matrices for the four categories: SD (pre, on) and PD (pre, on), by combining the SANs obtained in **Figure 4A** and the marker expression data from the FoVs. Starting with the SAN for each category, a weighted adjacency matrix, A^{adj}, is created. A^{adj} is a square distance matrix representing connections between markers, wherein '1' indicates a marker-marker connection, and '0' otherwise, and each connection is multiplied by weights from the SAN where the weights are bound by -1 (red/negative association) and 1 (blue/positive association). For each A^{adj}, we construct a similarity matrix, S^{adj}, using $S^{adj} = -0.5(QA^{adj}Q^t)$ with $Q_{ij} = \delta_{ij} - 1/n$ where δ is the Kronecker delta function and n is the number of markers, based on the centering operation in kernel PCA[35,36]. The weighted count matrix, X^{adj} can be computed by the eigenvalue decomposition[37–39] of S^{adj}. Assuming that each marker follows a Gaussian distribution, we can model these weighted count matrices of cells by markers to follow a multivariate Gaussian distribution, without loss of generality. We compute the first and second order moments to describe each of the four categories. Next, for each category, we randomly draw 10,000 samples creating a random matrix of 10000 rows and 5 markers. This is repeated 100 times and the random matrices are averaged out to create one representative matrix denoting that category. This is done to ensure that we work with random matrices that a) are invariant to biases and shifts while generating them from distance and similarity matrices and b) have the same number of rows in the 'pre' and 'on' states that obey the observed/empirical distributional moments. Implementing Equation 4 by incorporating the biological prior, gives us the optimized $W^{LSprior}$ for SD and PD (**Figure 8B i and ii**) respectively. It becomes evident through $W^{LSprior}$ for SD that there is increased CD8 infiltration into the tumor with decreased PD-L1 expression. Note that the PanCK+CD8 interaction that differentiated SD from PD was previously also identified from the image t-SNE analysis (see **Section 3.2**).

The $W^{LSprior}$ in PD shows decreased CD8 infiltration, increased expression of PD-L1 and a positive interaction between PD-1+PanCK+FoXP3. This interaction was also previously observed from the image t-SNE analysis as well as the GCGL analysis using quadrats.

4. Discussion

We have computationally analyzed multiplexed images from nine patients with advanced/metastatic non–small cell lung cancer (NSCLC) with progression, who were treated with an oral HDAC inhibitor (vorinostat) combined with a PD-1 inhibitor (pembrolizumab). From both the single cell as well as quadrat analyses, it is evident that fundamental architectural differences exist, even pre-treatment, between PD and SD patients. In PD patients, the presence of tumor with exhausted T-cells (PD-1) and immune suppressed Tregs (FoxP3), along with decreased CD8 infiltration could potentially lead to evolution of treatment resistance whereas in SD patients, we observe increased CD8 infiltration in tumor regions that decelerate disease progression. We have quantified the number of immune cells colocalizing on tumor regions in **Supplementary Figure 3** which clearly illustrates the differences of tumor-immune cells across SD and PD patients. We further verify these differences using a statistical model that indicates ecological diversity to exist in pre-treated PD and SD patients.

Through this high-throughput data analysis study, we have developed an automated pipeline to analyze data in a quantitative and efficient manner, for both cell and quadrat-level analysis of immune and tumor cell profiling aiding the discovery of novel biomarkers for anticancer therapy and improving clinical management of patients with NSCLC.

Validation of single cell and quadrat approaches on whole-slide NSCLC images

We have shown that our ML classifiers (SVM and BRT) correctly predict disease progression on the nine-patient FoV data (**Figure 6** and **Section 3.5**). We plot the ratio of PD-L1+CD3+CD8+PanCK cells (indicating SD patients/green circle) versus PD-1+FoxP3+PanCK (indicating PD patients/orange square) for each patient and note that these two sets of patients can be separated by a line, indicating that the two sets of biomarkers can be used to distinguish responders versus non-responders in this cohort. In comparison, the PD-L1 status available for each of these nine patients do not clearly indicate PD or SD (**Figure 9A**). In **Figure 9B**, we replot the t-SNE rendering in **Figure 7Aa** but colored using the patient's PD-L1 status. While the PD patients are correctly predicted to progress using the BRT classifier (red region in **Figure 7Ab**), the patients' PD-L1 status maps to three different colors (pink/blue and orange for the PD region in **Figure 9B**) immediately suggesting the clinical challenge associated with using PD-L1 as a diagnostic biomarker in NSCLC.

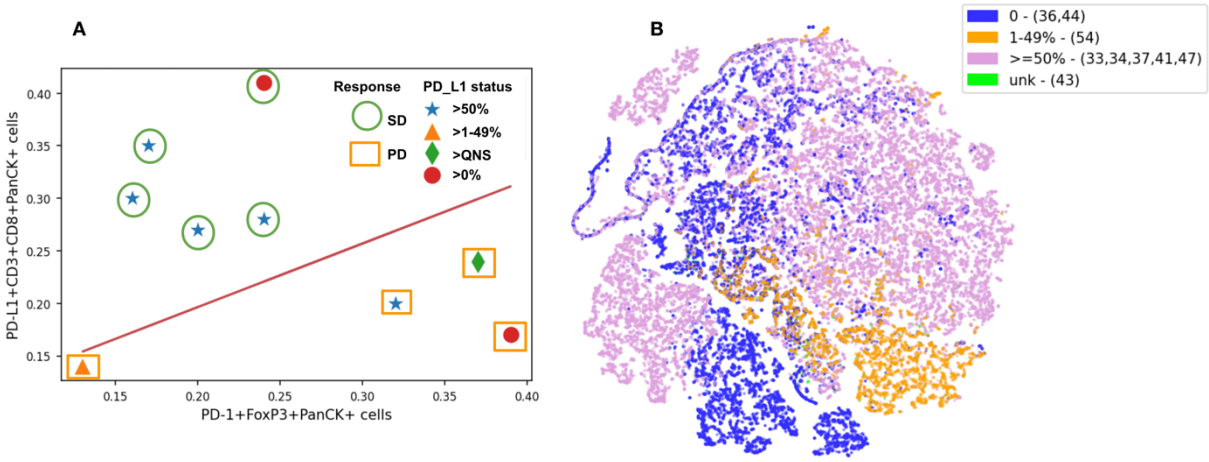


Figure 9: A. Ratio of PD-L1+CD3+CD8+PanCK cells versus PD-1+FoxP3+PanCK cells per patient. Each patient is marked according to their PD-L1 status and patient response. **B.** Combining cellular neighbors and quadrats using images from pre-treated patients where each quadrat is a cell with its neighbors. Clustering the cellular neighborhoods and plotting the t-SNE showing PD-L1 status per patient.

To further corroborate that our predictions are reliable and robust, and that the distinct cellular compositions can be used as a companion biomarker for PD-L1 status, we validate our predictions using patient data from a second NSCLC study (Lung2) using the same therapy, where there are three patients that are SD and three PD. The Lung2 cohort consists of whole-slide images (WSIs) that are orders larger than the nine-patient FoV data. We provide the training and test data summary in Table 1 (**Figure 10A**), single-cell and quadrats extracted from an example WSI (**Figure 10B**) and show the marker panels for both the datasets in **Supplementary Figure 4**.

SVM predictions on single-cell data from Lung2

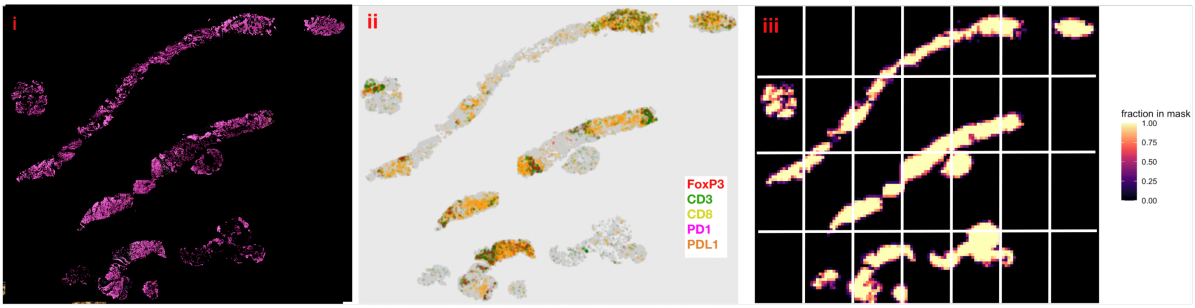
We retain the same SVM classifier trained on 47,000 pre-treatment cells (**Section 3.5.2**) to predict disease progression on 98,000 pre-treatment cells on Lung2 data, where the responses of the patients were known a priori. Figure **10C** depicts the t-SNE projection of

Lung2 pre-treatment data showing (**a**) patient data response (**b**) SVM predictions, and (**c**) the error in between training and test datasets. The predictions are averaged across 22 independent runs by randomly shuffling all the Lung2 data points for each run. Classification accuracy is the fraction of right predictions per run. We obtain an overall accuracy of 85.5%, with class-specific accuracy of 89% for SD patients and 82% for PD patients.

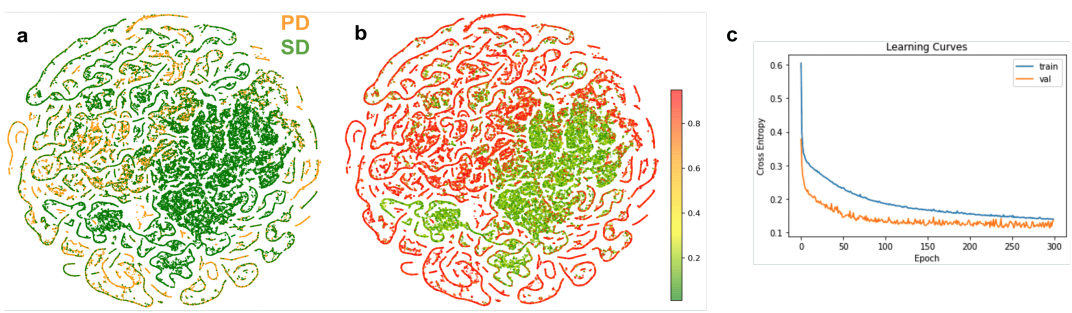
A

NSCLC multiplexed images	#Patients	#Images	#Response	#Images per response	#Cells	#Quadrats	Format	Dimension per image	Size per image
Training data	9	92 FoVs	5 SD (pre + on)	58	80K	1400	TIFF	1400 x 1000	10-50 MB
			4 PD (pre + on)	34	42K	1450			
Test data for predictions	6	6 WSIs (~1000 FoVs)	3 SD (pre)	3	56K	19K	Pyramid TIFF	14000 x 10000	300MB - 5GB
			3 PD (pre)	3	42K	15K			

B



C



D

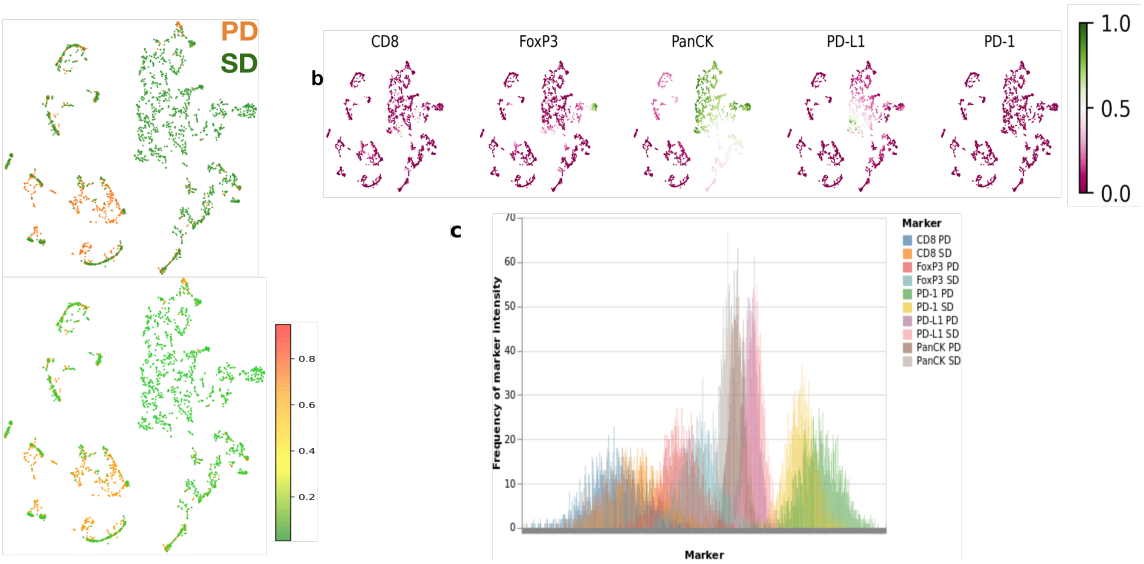


Figure 10: A: table comparing training and test datasets, in enabling the transfer learning-based predictions. **B**: (i) Example NSCLC WSI with corresponding (ii) cell segments colored per marker and (iii) quadrats with fraction of tissue retrieved using segmentation. **C:** Single cell: Between-study validation using SVM trained on FoV data to make predictions on Lung2 WSI data. t-SNE projection of Lung2 pre-treatment data showing **(a)** patient data response and **(b)** SVM predictions, with overall accuracy= 85.5%, accuracy for SD patients = 89% and for PD patients = 82%, and (**c**) error curves during training and test using a Keras-based SVM **D**. Quadrat: Between-study validation using MLE trained on FoV data to make predictions on Lung2 WSI data. t-SNE projection of Lung2 pre-treatment data showing **(a)** patient data response (top) and MLE predictions (bottom), with overall accuracy = 84.2%, accuracy for SD patients = 85.6% and for PD patients = 82.8%, **(b)** t-SNE showing marker expression spread, and (**c**) Lung2 pre-treatment quadrats: distribution of each marker across patient response.

MLE predictions on quadrats from Lung2

We assume that the marker expression in each quadrat in the nine-patient FoV dataset follows a multivariate Gaussian distribution. This enables us to view an image, in the form of a set of quadrats, as a mixture of quadrat-specific Gaussian distributions, and to use a Dirichlet process mixture model (DPMM) for clustering the quadrats into K different clusters. The quadrats for the PD (n= 500) and SD (n=1000) patients are clustered separately.
In order to assign a quadrat from Lung2, m, to one of the inferred clusters $k = 1 \ldots K$, we use the maximum likelihood estimation[13,40] (MLE) method. Every quadrat is tested for each cluster's parameters i.e., multivariate mean μ_k and covariance Σ_k to identify k_{ML} that maximizes the likelihood distribution. We assume the likelihood distribution to be that of a multivariate Gaussian. The analytic form of the ML method can be written as:

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\theta \in \Theta} p(m|\mu_k, \Sigma_k) \text{ and}$$

$$p(m|\mu_k, \Sigma_k) = \frac{\exp\left(-0.5\,(\mathrm{m} - \mu_{\mathrm{k}})^{\mathrm{T}} \Sigma_{\mathrm{k}}^{-1}\,(\mathrm{m} - \mu_{\mathrm{k}})\right)}{\sqrt{(2\pi)^{\mathrm{d}} - |\Sigma_{\mathrm{k}}|}}$$

where $\hat{\theta}_{\mathrm{ML}}$ is the maximum likelihood estimate of the Gaussian parameters $\theta = [\mu_k, \Sigma_k]$, Θ is the finite-dimensional parameter space for θ, d is the number of markers and $p(m|\theta)$ is the non-degenerate case of the multivariate Gaussian probability density function (pdf). The quadrat is assigned to the closest 'likely' k_{ML} described by $\hat{\theta}_{\mathrm{ML}}$. If k_{ML} belongs to a PD set of clusters, m is assigned to PD, and if k_{ML} belongs to a SD set of clusters, m is assigned to SD. Figure **10D** depicts the t-SNE projection of Lung2 pre-treatment data showing (**a**) (top) patient data response, (below) MLE predictions, and (**b**) the marker expression spread. In Figure **10D(c)**, we see the probability distribution of each marker in Lung2 for the SD and PD data and note that each marker has a distinct distribution but is slightly similar across response categories, which can pose inseparability issues in methods such as SVM, BRT but can be addressed at the distributional level. Using the MLE, we obtain an overall accuracy of 84.2%, with class-specific accuracy of 85.6% for SD patients and 82.8% for PD patients.

<u>Comparison of single cell and quadrat approaches</u>

In **Section 3.5**, we have shown that disease progression can be predicted using both single cell data as well as quadrats with an accuracy of 94.25% and 91.4%, respectively (**Figure 6**). In the transfer learning setting where training and predictions occur on separate NSCLC cohorts, both approaches can still predict with 85.5% and 84.2% accuracy, respectively. The decrease in prediction accuracy can be attributed to the sheer diversity between training and test datasets as shown in **Figure 10A**. Nevertheless, we can make predictions by exploiting the fact that prominent ecological differences exist between PD and SD prior to treatment.

Single cell data is a fundamental step in many biomedical studies, and it is regarded as a cornerstone of image-based cellular research. It allows us to measure the marker expression of individual cells in a FoV based on which we can identify cell morphologies, cell

subpopulations and neighborhoods, the stochasticity of marker expression, extract tumor and stromal regions as well as find the cells that colocalize in specific regions such as tumor, stroma or invasive margins. We can also perform dimensionality reduction to aid visualization of the high-dimensional single cell data. Single cell data still relies on a modest cell segmentation which can be cumbersome depending on the number of samples, and technical noise in the samples.

On the other hand, quadrat analysis, is an important process in image analysis where the high-dimensional image is sequentially tiled and summarized and classified. Quadrats can be used to identify aggregate marker expression and differential markers across the population of cells in a sample and is typically used to detect interesting regions across high-dimensional images and does not require any cell-specific information, making it a simple yet effective approach. In many high-throughput image analysis research, using tiles for pattern matching and classification is standard[41].

Further, we show in **Section 3.5.3** and **Figure 7**, that these methods can be synergized to predict disease progression, as well as visualize these predictions at the patient level through risk maps (**Figure 6i**, **Supplementary Figures 5, 6**).

Summary and Future work

There are technical and clinical challenges in using PD-L1 as a standard diagnostic biomarker and the need for having complementary diagnostic biomarker along with PD-L1 based on cancer types and treatment is highly recommended[42]. Herein, we have shown that the pre-treatment ecologies of NSCLC patients are distinct, so much so that their spatial organization (CN or quadrats) can be used to predict response to treatment more accurately than PD-L1 alone. While promising, the current study is limited to a handful of patients, but will be validated using WSIs from a larger cohort that are part of an on-going NSCLC clinical trial. While we have identified important biomarkers for treatment response, e.g., PD-1+PanCK+FoXP3, the biological underpinnings remain unclear. However, this will be investigated through the development of a mechanistic model to simulate response to treatment, which will aid in clarifying *why* the markers are predictive. We will also study if the signs of progression in tumor architecture, during-treatment, are brought about by the direct effects of combination therapy. In summary, we will continue to explore this promising new biomarker for treatment response in patients with NSCLC.

5. Patient data

We have 92 7-stain PerkinElmer Vectra images from nine patients with advanced/metastatic non–small cell lung cancer with progression[42]. They were treated with an oral HDAC inhibitor (vorinostat) combined with a PD-1 inhibitor (pembrolizumab). Tumor biopsies were collected from all patients both pre- and on-treatment (days 15-21). Of the 9 patients, five qualified as stable disease (SD) and four as progressive disease (PD). There are 58 images from the SD patients and 34 images from the PD patients. Figure 1b summarizes the number of images per treatment and response category.

Multiplex IF staining

The Opal 7 multiplexed assay was used to generate multiple staining slides. Staining was performed consecutively. The sequence of antibodies for multiplex staining was PanCK, PD-L1, CD8, CD3, PD-1, FoXP3 and DAPI. The slides were imaged using the Vectra 3.0 spectral imaging system (PerkinElmer).

6. Code availability

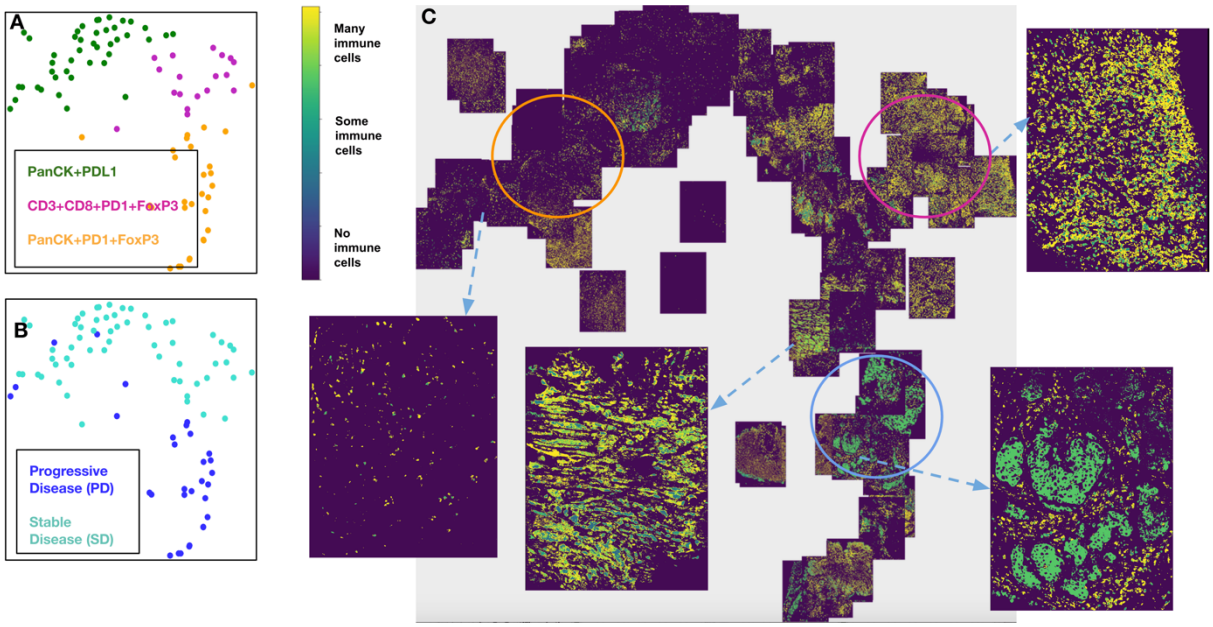
Code pipeline for the cell segmentation-based analysis and quadrat-based analysis will be made available upon user request. Mistic is downloadable at: https://github.com/MathOnco/Mistic.

7. Acknowledgments

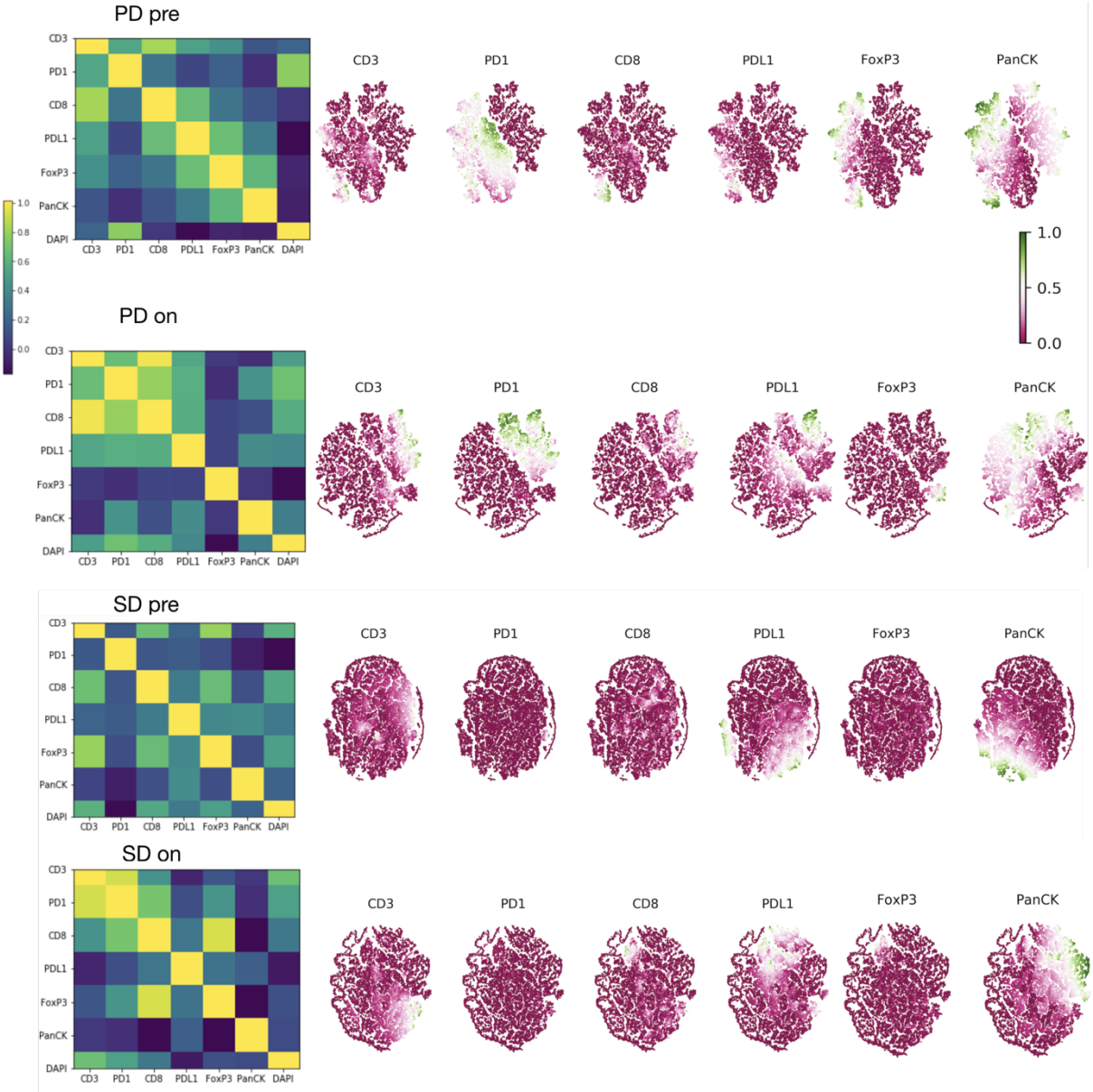
The authors gratefully acknowledge funding by the National Cancer Institute via the Cancer Systems Biology Consortium (CSBC) U01CA232382 (supporting S.P., C.G., M.R.-T., A.R.A.A), the Physical Sciences Oncology Network (PSON) U54CA193489 (supporting C.G., M.R.-T., A.R.A.A), and support from the Moffitt Center of Excellence for Evolutionary Therapy (supporting M.R.-T., and A.R.A.A).

8. Supplementary

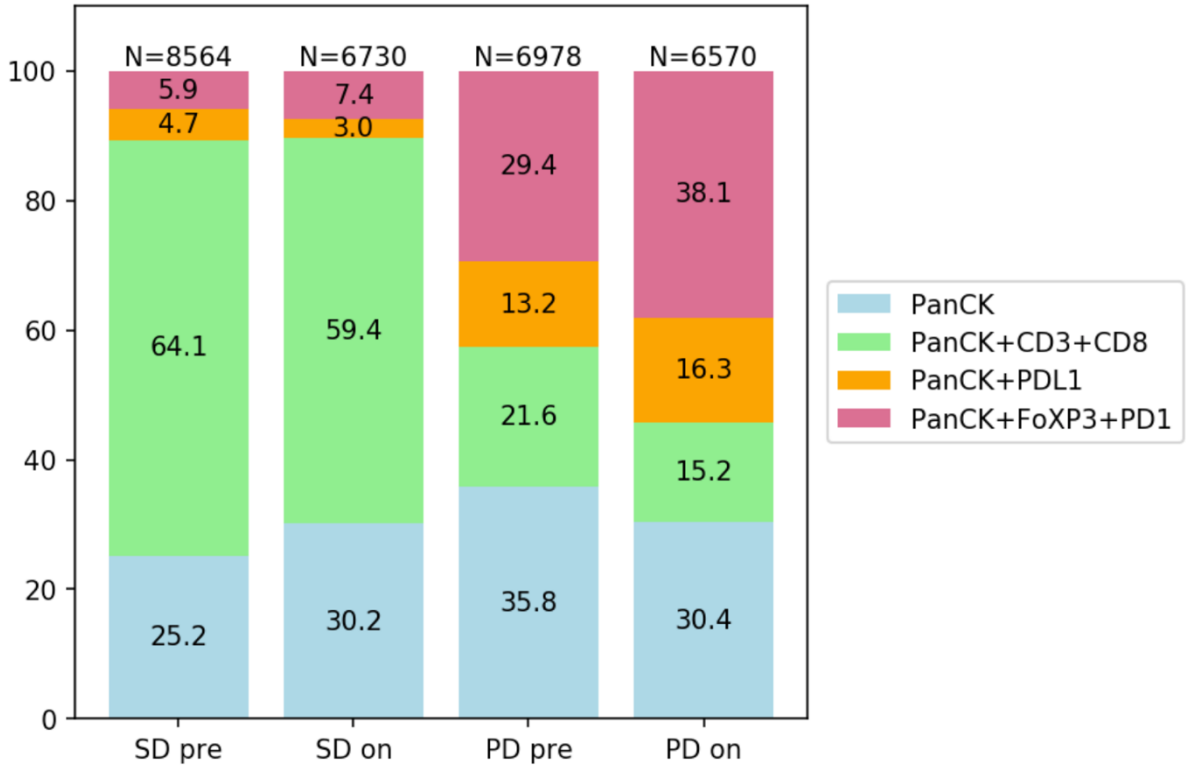
S.1 Figures



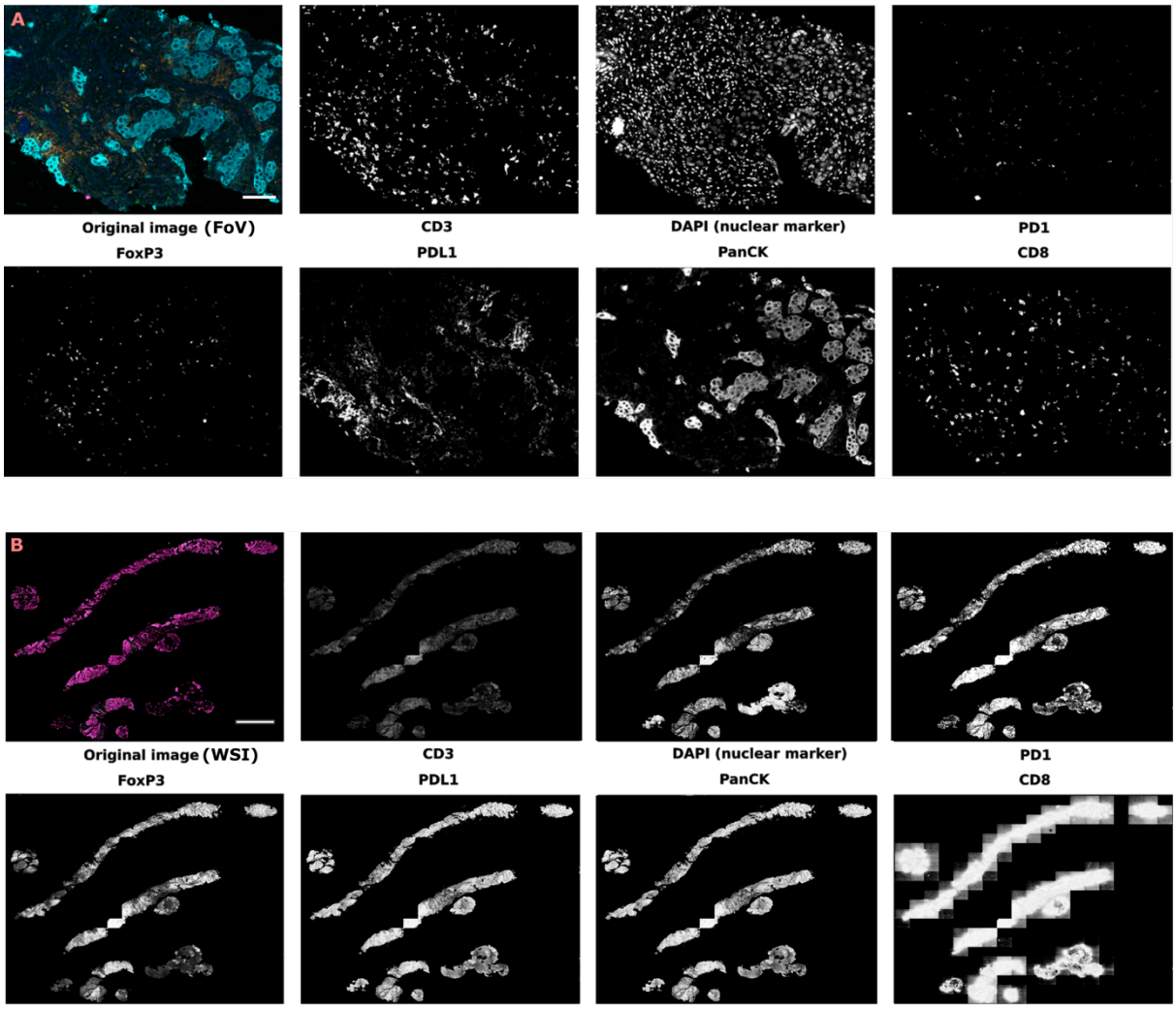
Supplementary Figure 1: PD and SD patients have significantly different cellular compositions. **A.** 2D t-SNE plot showing three clusters annotated with the differentially expressed markers per cluster. Clusters are obtained using the tumor-immune cell counts at the tumor border approximated by convex hulls (see **Figure 2**). **B.** Same t-SNE as in **(A)** depicting the disease response spread. **C.** Image t-SNE using same t-SNE coordinates as in **(A)** illustrating the gradient of immune cells (CD3, CD8, PD-1 and FoxP3) and tumor cells (PanCK) across images. A higher colocalization of PanCK+PD-1+FoxP3+ (shown in green) is seen for PD patients. Circles denote regions of immunologically cold (orange), immunologically warm (blue) and immunologically hot (pink) tumors.



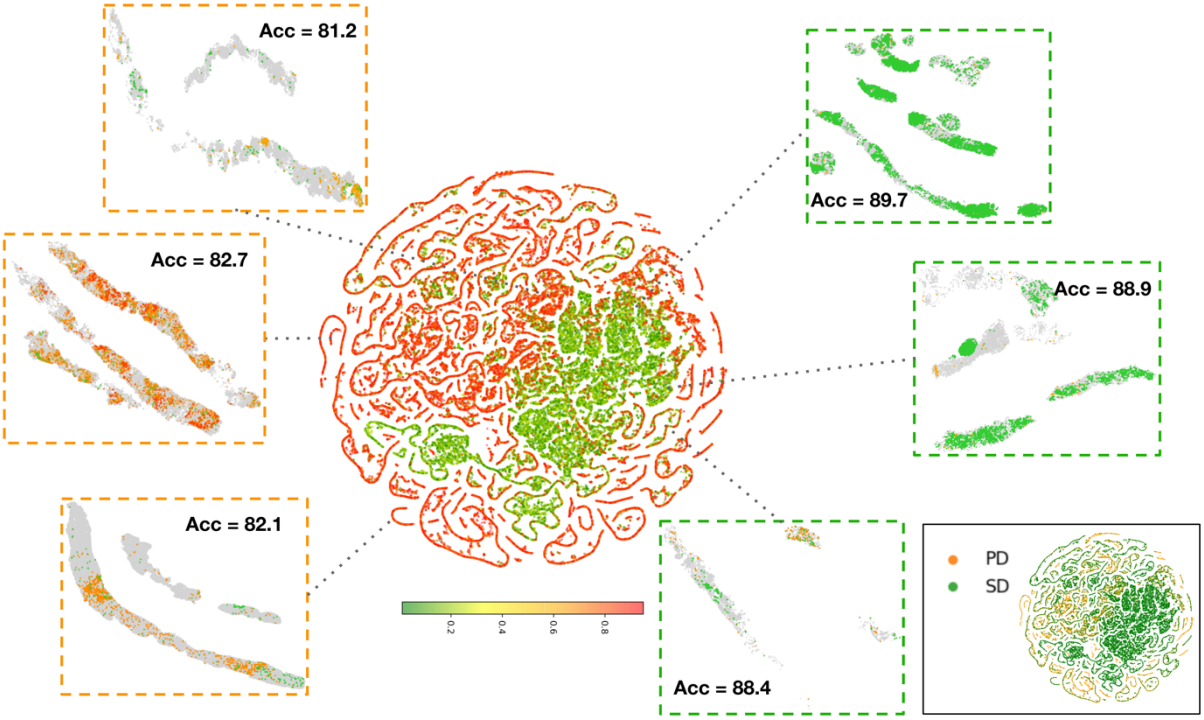
Supplementary Figure 2: Marker co-expression covariance captured by CNs across patient categories and t-SNE showing marker expressions for PD pre, PD on, SD pre and SD on (top to bottom panels). For marker co-expression covariances, blocks identify cells that both reside in spatially similar neighborhoods while expressing similar markers. Green-yellow shows high correlation (=cells are highly similar both spatially and functionally). For the marker t-SNEs, little to high expression is indicated by dark pink, pink to green. There are 21,000 cells in PD pre and PD on, 43,000 cells in SD pre and 36,000 cells in SD on.



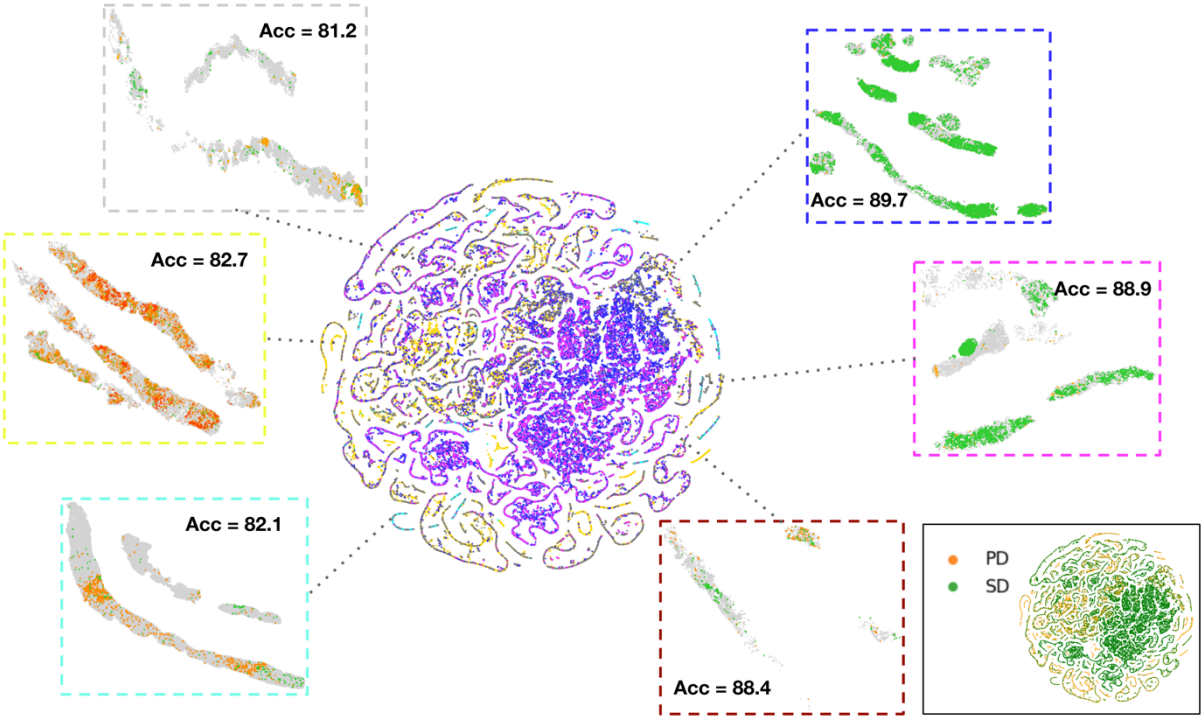
Supplementary Figure 3: Relative proportions of tumor cells (PanCK+) and tumor cells colocalized with one or more of the functional markers (CD3+ T cells, CD8+T cells, FoXP3+ (Treg cells), PD-L1, and PD-1) within each patient response category based on treatment timing. Pooled data from all patients are shown with cell numbers per category shown above the respective bar.



Supplementary Figure 4: Example marker panels for **A**. a FoV from the 9-patient NSCLC cohort. Scale bar = 200 pixels/mm. and **B**. a WSI from Lung2. Scale bar = 1236 pixels/mm.
The markers for the 7-stain Vectra image (FoV and WSI) are DAPI (nuclei), CD3 (T cells), CD8 (T effector cells), FoXP3 (T regulatory cells), PD-1 (inhibitory receptor on immune cells), PD-L1 (immune checkpoint marker), and pan-cytokeratin (PanCK, epithelium).



Supplementary Figure 5: Risk maps, based on SVM predictions, show the probability of disease progression for the six patients in the Lung2 data, and are oriented based on the patient response category given by the t-SNE (in the center) for SVM predictions. The probabilities of progression are well aligned to each patient's response where patient response is shown in the inset. PD patents are predicted to show higher progression, denoted by red-orange shades and SD patients show lower progression, denoted by yellow-green shades. Prediction accuracy for each patient is given within the risk map. Each risk map is outlined based on the patient response category i.e. three SD risk maps outlined in green and three PD in orange.



Supplementary Figure 6: Risk maps, based on SVM predictions, show the probability of disease progression for the six patients in the Lung2 data, and are oriented based on the patient spread given by the t-SNE (in the center). The probabilities of progression are well aligned to each patient's response where patient response is shown in the inset. PD patents are predicted to show higher progression, denoted by red-orange shades in the risk maps and SD patients show lower progression, denoted by yellow-green shades in the risk maps. Prediction accuracy for each patient is given within the risk map. Each risk map is outlined based on the patient.

S.2 Methods

Multiplexed image processing pipeline

Each 7-stain Vectra multiplexed image is available as a tiff stack with 7 sub tiffs: one tiff for each marker. We have built an image processing pipeline that consists of three parts to analyze each tiff stack (**Figure 3i**):

(1) the pre-processing unit that denoises and cleans each 1008-pixel x 1344-pixel sub tiff of the 7-stain Vectra Tiff stack (**Figure 3ia**). We start with generating the grayscale image of each sub tiff and choose the nuclear image stained with DAPI for cell segmentation (**Figure 3ib**). Since the images are ridden with technical noise, it is crucial to extract true stain signal from noise. For this we apply an Otsu thresholding to the images which involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels each side of the threshold, i.e., the pixels that either fall in foreground or background. Foreground pixels are regarded as true stain signals while background pixels are noise. Next, the images are tiled into 256-pixel x 256-pixel frames (**Figure 3ic**).

(2) The cell segmentation unit: The image tiles are used as input to U-Net[8], a Fully Convolutional Network, for cell segmentation (**Figure 3id**). We further trained UNET using nuclear segments available from Kaggle's Data Science Bowl[43].

(3) Downstream analysis unit: Once the cell segmentation predictions are made per tile (**Figure 3ie**), we stitch the tiles (**Figure 3if**) and perform clustering on the stitched image (**Figure 3ig**). For clustering, we use a Gaussian Mixture model[44–47] which identifies heterogeneous cell types based on stain expression and does not require the number of clusters to be known in advance.

Risk map generation

To enable patient-scale predictions, we generate risk maps for each patient using the SVM predictions. A risk map shows the probability of disease progression for any given patient along with the patient's overall chance to progress. In **Supplementary Figures 5** and **6**, we depict the risk maps for six patients, three of whom were identified as SD and three as PD. We note that the probabilities of progression are well aligned to each patient's response (patient response is shown in the inset (black box)).

9. References

1. World Health Organization. Lung Cancer: Estimated cancer incidence, mortality and prevalence worldwide. Published 2018. http://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf

2. American Cancer Society. What Is Non-Small Cell Lung Cancer. https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/what-is-non-small-cell-lung-cancer.html.

3. Tunali I, Gillies RJ, Schabath MB. Application of radiomics and artificial intelligence for lung cancer precision medicine. *Cold Spring Harb Perspect Med*. 2021;11(8):a039537.

4. Hofman P, Badoual C, Henderson F, et al. Multiplexed immunohistochemistry for molecular and immune profiling in lung cancer—just about ready for prime-time? *Cancers*. 2019;11(3):283.

5. Mezheyeuski A, Bergsland CH, Backman M, et al. Multispectral imaging for quantitative and compartment-specific immune infiltrates reveals distinct immune profiles that classify lung cancer patients. *J Pathol*. 2018;244(4):421-431.

6. Surace M, Rognoni L, Rodriguez-Canales J, Steele KE. Characterization of the immune microenvironment of NSCLC by multispectral analysis of multiplex immunofluorescence images. In: *Methods in Enzymology*. Vol 635. Elsevier; 2020:33-50.

7. Rashid R, Gaglia G, Chen YA, et al. Highly multiplexed immunofluorescence images and single-cell data of immune markers in tonsil and lung cancer. *Sci Data*. 2019;6(1):1-10.

8. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2015:234-241.

9. Guisan A, Thuiller W, Zimmermann NE. *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press; 2017.

10. Peterson AT, Soberón J, Pearson RG, et al. Ecological niches and geographic distributions (MPB-49). In: *Ecological Niches and Geographic Distributions (MPB-49)*. Princeton University Press; 2011.

11. Kettenring JR. Canonical analysis of several sets of variables. *Biometrika*. 1971;58(3):433-451.

12. Anderson MJ. Permutational multivariate analysis of variance (PERMANOVA). *Wiley Statsref Stat Ref Online*. Published online 2014:1-15.

13. Bishop CM, Nasrabadi NM. *Pattern Recognition and Machine Learning*. Vol 4. Springer; 2006.

14. Prabhakaran S, Gatenbee C, Robertson-Tessi M, et al. Mistic: an open-source multiplexed image t-SNE viewer. *Patterns*. Published online 2022:100523.

15. Carrington PJ, Scott J, Wasserman S. *Models and Methods in Social Network Analysis*. Vol 28. Cambridge university press; 2005.

16. Blüthgen N, Menzel F, Blüthgen N. Measuring specialization in species interaction networks. *BMC Ecol*. 2006;6(1):1-12.

17. Pocock MJ, Evans DM, Memmott J. The robustness and restoration of a network of ecological networks. *Science*. 2012;335(6071):973-977.

18. Team RC. R: A language and environment for statistical computing. Published online 2013.

19. Popovic GC, Warton DI, Thomson FJ, Hui FK, Moles AT. Untangling direct species associations from indirect mediator species effects with graphical models. *Methods Ecol Evol*. 2019;10(9):1571-1583.

20. Traag VA, Waltman L, Van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep*. 2019;9(1):1-12.

21. Schürch CM, Bhate SS, Barlow GL, et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell*. 2020;182(5):1341-1359.

22. Cruz AF, Parra ER, Jiang M, et al. Characterization of the immunologic intra-tumor heterogeneity in early stages of non-small cell lung carcinoma using multiplex immunofluorescence and image analysis approaches. *Cancer Res*. 2018;78(13_Supplement):1174-1174.

23. Shakya R, Nguyen TH, Waterhouse N, Khanna R. Immune contexture analysis in immuno-oncology: applications and challenges of multiplex fluorescent immunohistochemistry. *Clin Transl Immunol*. 2020;9(10):e1183.

24. Barua S, Fang P, Sharma A, et al. Spatial interaction of tumor cells and regulatory T cells correlates with survival in non-small cell lung cancer. *Lung Cancer*. 2018;117:73-79.

25. Principe DR, Chiec L, Mohindra NA, Munshi HG. Regulatory T-cells as an emerging barrier to immune checkpoint inhibition in lung cancer. *Front Oncol*. Published online 2021:2092.

26. Cortes C, Vapnik V. Support vector machine. *Mach Learn*. 1995;20(3):273-297.

27. Elith J, Leathwick J. Boosted Regression Trees for ecological modeling. *R Doc Available Online Httpscran R-Proj Orgwebpackagesdismovignettesbrt Pdf Accessed 12 June 2011*. Published online 2017.

28. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol*. 2008;77(4):802-813.

29. Gulli A, Pal S. *Deep Learning with Keras*. Packt Publishing Ltd; 2017.

30. Ketkar N. Introduction to keras. In: *Deep Learning with Python*. Springer; 2017:97-111.

31. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Icml*. ; 2010.

32. Zeiler MD. Adadelta: an adaptive learning rate method. *ArXiv Prepr ArXiv12125701*. Published online 2012.

33. Royston P. Lowess smoothing. *Stata Tech Bull*. 1992;1(3).

34. Hen I, Sakov A, Kafkafi N, Golani I, Benjamini Y. The dynamics of spatial behavior: how can robust smoothing techniques help? *J Neurosci Methods*. 2004;133(1-2):161-172.

35. Schölkopf B, Smola AJ, Bach F. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press; 2002.

36. Smola AJ, Schölkopf B. *Learning with Kernels*. Vol 4. Citeseer; 1998.

37. Vogt JE, Prabhakaran S, Fuchs TJ, Roth V. The translation-invariant Wishart-Dirichlet process for clustering distance data. In: *ICML*. ; 2010.

38. McCullagh P. Marginal likelihood for distance matrices. *Stat Sin*. Published online 2009:631-649.

39. McCullagh P, Yang J. How many clusters? *Bayesian Anal*. 2008;3(1):101-120.

40. Eliason SR. *Maximum Likelihood Estimation: Logic and Practice*. Sage; 1993.

41. Vu QD, Graham S, Kurc T, et al. Methods for segmentation and classification of digital microscopy tissue images. *Front Bioeng Biotechnol*. Published online 2019:53.

42. Gray JE, Saltos A, Tanvetyanon T, et al. Phase I/Ib study of pembrolizumab plus vorinostat in advanced/metastatic Non–small cell lung cancer. *Clin Cancer Res*. 2019;25(22):6623-6632.

43. Kaggle Data Science Bowl. Spot nuclei. Speed cures. Published 2018. https://www.kaggle.com/c/data-science-bowl-2018

44. Neal RM. Markov chain sampling methods for Dirichlet process mixture models. *J Comput Graph Stat*. 2000;9(2):249-265.

45. Teh YW. Dirichlet Process. *Encycl Mach Learn*. 2010;1063:280-287.

46. Teh YW, Jordan MI. Hierarchical Bayesian nonparametric models with applications. *Bayesian Nonparametrics*. 2010;1:158-207.

47. Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical dirichlet processes (technical report 653). *UC Berkeley Stat*. Published online 2004.