Development of computational tools for analysis of biological image data

Elliot Marshall Gray

Advisor: Dr. Y. H. Chang

Department of Biomedical Engineering
Oregon Health and Science University

This dissertation is submitted for the degree of
Master of Science

September 2021

I would like to dedicate this thesis to my parents.

Acknowledgements

Abstract

In order to improve treatment for a variety of forms of cancer, improved basic understanding of tumor microenvironments is needed. Multiplexed tissue imaging (MTI) is a family of novel diagnostic technologies that are capable of revealing critical molecular and architectural features of the tumor microenvironment, enabling the study of cancer cell processes in unprecedented detail. However, the data analysis process for MTI continues to present challenges for the research community. In this thesis, I review the state of the art MTI analysis methods (Chapter 1) and present two projects that address specific MTI analysis challenges. In the first project (Chapter 2), I present an algorithm specifically designed for automated quantification of a property of ductal carcinoma in situ measurable by MTI. The output of the algorithm is demonstrated to correlate with the state of the art results produced by an expert pathologist's manual estimation, and provides the benefits of reduced labor cost and no possibility of inter-user bias. In the second project (Chapter 3), I propose a novel metric of spatial proximity among immune and epithelial cell types, analyze a pancreatic ductal adenocarcinoma MTI data set, and test the hypothesis that the metric will improve the prognostic value of the MTI data using a machine learning model. In the context of the model, the proposed metric produces cross-validated predictions with prognostic value comparable to previously reported metrics, but the small sample size means that future validation experiments are required to confirm the findings. These two projects produced and tested new tools for MTI data analysis and may enhance the value and efficiency of MTI data analysis in future applications.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

AP   Average Precision

DCIS   Ductal Carcinoma In Situ

FFPE   Formalin-Fixed Parrafin-Embedded

HM   Harmonic Mean

MIHC Multiplexed Immunohistochemistry

MTI   Multiplexed Tissue Imaging

PDAC Pancreatic Ductal Adenocarcinoma

RFECV Recursive Feature Elimination with Cross-Validation

ROC AUC Receiver Operating Characteristic Area Under the Curve

SLIC   Simple Linear Iterative Clustering

TLS   Tertiary Lymphoid Structure

TME   Tumor Microenvironment

Chapter 1

Background of multiplexed tissue imaging (MTI) data analysis

1.1 Mapping the architecture of tumors is needed to improve basic knowledge that links tumor biology to patient outcomes

Cancer is the second leading cause of death in the United States and claimed nearly 600,000 lives in 2019[1]. In clinical management of cancer, tissue samples are routinely collected and submitted for pathological analysis, informing disease prognosis and treatment. Immunohistochemistry (IHC), which is an imaging technique based on the use of antibodies to specifically mark antigens within (frozen or formaldehyde-fixed) thin tissue sections, is a mainstay of diagnostic pathology, allowing the pathologist to identify antigens that are tumor specific or that have been upregulated by tumor-specific processes [23]. For example, the detection via IHC of upregulation of the receptor tyrosine kinase Her2 (erbb2) in estrogen receptor-negative breast cancer cells was demonstrated to predict improved benefit (meaning increased progression-free survival and overall survival times) from doxorubicin therapy [70]. While the analysis of IHC images by pathologists typically consists of visual assessment of the intensity (relative darkness or brightness, depending on whether the image contrast is generated by optical absorption or fluorescence by the imaging reagent) of stained cells, or counting of stained cells whose intensity surpasses a predefined threshold (commonly referred to as "positively stained") [76], IHC images also contain a wealth of information about the spatial arrangements of cells and the morphology of tissue structures, with the potential produce novel biomarkers and aid clinical cancer treatment.

Tumors are complex ecosystems with multiple interacting cell populations that collectively are referred to as the tumor microenvironment. Measurement of the spatial context of cells within the tumor microenvironment is useful for assessing patient prognosis in numerous ways. Assessment

---

[1]source at www.cdc.gov

of tumor grade, which is defined in part by the progressive loss of tissue structure such as regularly-shaped epithelial glands, is a cornerstone prognostic tool for solid tumors, e.g. pancreatic ductal adenocarcinoma (PDAC), where higher grade tumors are associated with reduced patient survival [91]. The invasion of tumor cells along nerves or blood vessels, which may be visible to the trained pathologist in histological images, is another important indicator of patient prognosis in PDAC [64], as is the presence of solitary pancreatic tumor cells that have lost the positive expression of the cellular adhesion molecule E-cadherin, a phenomenon that naturally requires IHC to observe [42]. Leukocyte spatial arrangements within tumor and adjacent tumor stroma, revealed by IHC, are particularly notable for prognostic value: in estrogen-receptor-positive breast cancer, for example, an increase in the degree of spatial clustering of lymphocytes was associated with improved prognosis, with a significance rivaling that of standard-of-care assays [39]. When IHC was used to invent the "Immunoscore", which delineates specific leukocyte populations such as $CD3^+$ T cells in their spatial context within colorectal tumors, prognostic value surpassing the gold-standard tumor stage and grade system was achieved in a large cohort of patients [69]. These examples reflect the biological and clinical significance of the spatial architecture of solid tumors, which we are still only beginning to understand [28].

Several advantages of tumor imaging, particularly IHC, combined with enhanced knowledge of tumor biology, have led to a proliferation of new and more sophisticated forms of IHC. While "omics" technologies, such as DNA microarrays [72] and single-cell RNA sequencing [89] to name only a few, have greatly expanded basic knowledge of tumor biology, including the discovery of extensive cellular heterogeneity within the neoplastic and stromal cells of individual tumors [89], these assays usually[2] require digestion and dissociation of the tumor tissue to form a cell suspension and so do not preserve the spatial context that those cells were taken from. Furthermore, IHC is considerably less expensive than genomic assays that utilize DNA or RNA sequencing, with a recent review placing costs for single-cell RNA sequencing experiments in the $0.30 to $10. per cell [50], and unlike certain advanced genomic technologies which can only be applied to fresh or frozen tissue samples, IHC can be applied to formalin-fixed parrafin-embedded tissue (FFPE) samples. While the sourcing and analysis of FFPE tissue samples comes with its own challenges due to preanalytical factors such as the conditions of formalin fixation, FFPE samples constitute a vast and valuable resource with great potential for retrospective studies [31]. These factors collectively made IHC an attractive technology for further development, including by increasing the number of antigens which could be simultaneously detected within single cells in what is now called Multiplexed Tissue Imaging (MTI).

MTI technologies use various methods of iterative staining and imaging to label dozens of protein biomarkers on single tissue sections [55, 90, 33, 47]. As an example, the method employed to generate the data analyzed in this thesis, multiplex immunohistochemistry (mIHC), starts with a standard IHC staining approach, but chromogen removal and antibody stripping steps are added, allowing re-staining and imaging of additional markers [90]. Using this approach, 12 to 29 markers can currently be imaged

---

[2]There exist new approaches to achieve spatially-resolved genomic measurements, such as Spatial Transcriptomics [61], which offer exciting new prospects for mapping tumor architecture.

in a single tissue section, allowing for comprehensive profiling of immune cell populations. MTI methods that currently exist have various advantages and disadvantages, including possessing different spatial resolution, field of view, image acquisition time, cost of implementation, dynamic range in the detection of antigen concentration, and number of antigens which can be simultaneously detected in a single sample. However, each MTI technology is similar in that they produce images with single-cell resolution, and possess a base of computational image processing tools that have been developed to quantify the spatial locations and extent of single cells, as well as to measure the magnitude (a.k.a. intensity) of multiple antigen concentrations within those same cells simultaneously. These MTI technologies hold the promise to greatly expand upon the diagnostic and prognostic potential of conventional IHC, but the technological development of the assays themselves and the computational tools required to effectively utilize the data they produce are still works in progress.

1.2 Analysis of rich 2D multiplexed tissue imaging data sets remains a challenge

While visual assessment of IHC images is still routinely performed by pathologists, the development of computational analysis tools to automate such tasks as the counting of positive cells or other quantitative evaluations, which require the valuable time of trained pathologists who do not always precisely agree on image quantification, will improve the utility of IHC and eventually MTI assays [76]. Focusing now on the general concept of 2D MTI data, assuming that images have sufficiently high resolution to detect single cells, the most common form of computational analysis consists of a set of canonical steps called a "pipeline". The steps in the typical pipeline are channel registration (or alignment), single-cell segmentation, cell-type classification, and pattern or tissue-structure analysis. The linking together of individual steps in a pipeline in a manner that is easily implemented by diverse users and is computationally scalable is also an obstacle, which has recently been addressed for example by the application MCMICRO [80], which allows users to run multiple algorithms for each step in the pipeline in parallel and visualize the results in a user friendly way. Finally, analysis of MTI image data using pipelines is critically supported by sophisticated open source image browsing and annotation tools such as Napari [62] and Minerva [43], which enable quality control and sanity checking of images and analysis results, respectively.

The individual steps of MTI analysis pipelines pose different degrees of challenge. For example, many good tools are available in ImageJ [81] for registration of channels, capable of aligning images as well as correcting rotational misalignment or deformation, with the caveat that the performance of individual algorithms generally depends on the qualities of the image data. That is, the images to be registered must either have adequate fiducial markers, or must possess a sufficient degree of similarity to each other in order to identify common landmarks. Registration of images that lack

adequate fiducial markers, contain little similarity, or possess strong distortion remains a challenge, however, improvements in the image acquisition process may ameliorate these challenges in the future.

Segmentation, whereby an algorithm assigns individual pixels belonging to either single cells (or their nuclei), or individual tissue structures, remains a difficult problem despite the availability of an abundance of tools for this purpose. In MTI images, segmentation of cell nuclei is usually achieved by a marker of DNA, such as DAPI or Hematoxylin, while segmentation of cell cytoplasm boundaries may be accomplished accurately only if molecular markers of cell membranes are available, which may not be the case for all cells in a tissue sample. While simple segmentation techniques based on the watershed algorithm [66] have seen success after they were popularized in CellProfiler [15], simple algorithms have historically suffered from difficulties of over-segmentation (that is, dividing one object into multiple parts), and from the challenge of identifying the boundary between touching objects. Also, compared to segmentation of *in vitro* cell populations, the segmentation of cells or objects in tissues is more challenging due to the greater diversity of cell types, shapes, and sizes (particularly in the case of cancer cells), as well as the presence of various kinds of noise and artifacts that are particularly common in FFPE tissues, arising from variable conditions of formalin fixation and tissue handling [80]. Recently developed segmentation algorithms based on deep learning such as Cellpose [88] and Deepcell [7] offer greatly improved performance in single-cell segmentation of MTI images, although care must be taken to assess the quality of each segmentation result as these models are trained using a limited variety of tissue types and imaging modalities.

Meanwhile, the segmentation of tissue structures is a niche subject, given the diversity of different normal and pathological tissue structures that exist in MTI images, and the relative scarcity of research studies aiming to quantify the morphological attributes of these structures. An example of a structure for which some segmentation algorithms exist is the epithelial glands of the prostate; Gleason grading, which expert pathologists assess by eye, is currently a gold-standard prognostic indicator for prostate cancer patients, and is based principally on the morphological appearance (i.e., shape) of prostate epithelium. Li *et. al.* developed a deep learning segmentation algorithm for segmentation of areas in prostate tissue exhibiting specific Gleason patterns, and demonstrated an automated Gleason grading system based on the segmentation results that correlated with the grading of expert pathologists [52]. Chapter 2 in this thesis includes (and motivates) an algorithm for the segmentation of ductal carcinoma in situ (DCIS) lesions and their surrounding myoepithelium, for which there had previously been no algorithm [35]. Computational segmentation and subsequent data mining of tissue structures in MTI images may pose many challenges and potentially yield great rewards in the future.

Given single-cell segmentation results paired with MTI image data, a critical next-step in analysis pipelines is to identify distinct cell types, either by examining intensity profiles of lineage markers (i.e., markers that are known to be expressed at high levels only by cells which have undergone a known differentiation process, such as the CD3 protein that is an important signal transduction molecule in the T-cell receptor complex and is specifically expressed by T cells [3]), or by using unsupervised

---

[3]source: http://www.proteinatlas.org/ENSG00000167286-CD3D

clustering to identify novel cell types [51, 78]. Unwanted variation in MTI image intensity caused by preanalytical variables such as formalin-fixation-time makes direct comparison of images of different tissue samples problematic, and as a result, a common approach to identification of cell types is for expert biologists to visually assess image intensity patterns and assign a threshold for each tissue and each marker, creating a binary positive/negative classification status for marker intensity in single cells, followed by definition of cell types using combinations of positive and negative for lineage markers; this process is commonly used in flow cytometry and is known as gating [4, 90]. When biological molecules do not have a well-defined lineage-specific expression pattern and the analyst wishes to use unsupervised clustering algorithms to identify cell types present in a collection of tissue samples, the intensity values for each image must be normalized to enable direct comparison of cells from different tissues; recent attempts to develop intensity normalization algorithms include RESTORE [19], which detects and takes advantage of the mutually-exclusive pattern of expression of some markers, for example immune lineage markers and markers expressed by epithelium, to identify and normalize the intensity of the background (negative cell population). Similar to image registration, the difficulty of the intensity normalization problem stems from data quality, in this case the unwanted variation in image intensity between tissues. In the future, tissue sample collections that are processed with standardized or automated tissue handling and imaging protocols may exhibit a reduced degree of unwanted variation due to preanalytical variables, possibly reducing the challenge of image intensity normalization; meanwhile, developing effective normalization is a crucial challenge.

Once cell types are determined, the final step in an MTI analysis pipeline is to assess the biological and clinical relevance of the frequency, density, and spatial distribution of those cell types. Evidence is emerging that measurement of the spatial distribution of cells within the tumor microenvironment, uniquely possible through the use of MTI or related imaging technologies among biological assays, contains prognostic relevance. For example, in breast cancer, immune hotspots [63, 39], tumor-immune mixing [46] and cytotoxic T cell localization [36], are all spatial features associated with survival. Similarly, proximity of cytotoxic T cells and tumor cells confers a better prognosis in PDAC [24, 16, 58]. These studies suggest that additional characterization of spatial features in tissue is needed to understand the biological processes underlying these associations. Indeed, in melanoma, tertiary lymphoid structures are prognostic and associated with response to immune checkpoint blockade, demonstrating the utility of spatial analysis for both prognosis and patient stratification [14].

There are several computational challenges associated with the development of a prognostic or therapy-response-predictive signature based on the spatial distribution of cells within tumor tissue. Taking, for example, the hypothesis that increased contact-dependent interaction between two cell types (for example, the lysis of tumor cells by cytotoxic T cells) would be a beneficial prognostic indicator, and making the assumption that increased spatial proximity between those two cell types in static images of fixed tumor tissue would be correlated with that contact-dependent interaction, the researcher must choose *a priori* from a collection of different spatial metrics [6], for example, the nearest-neighbor

G-function[4], and the reduced second moment function, or K-function[5], which quantify subtly different forms of spatial proximity; the G-function examines nearest-neighbor distances, while the K-function examines counts of nearby-neighbors within a given distance. Those metrics also require the *a priori* selection of a characteristic distance scale; indeed, the need to choose a spatial scale for analysis of proximity is common among proximity metrics, for instance the sliding-windows technique [87], which uses unsupervised clustering to identify small image regions, or "windows", that share a common cellular composition. The need to choose a distance scale is problematic because, while there exist *in vitro* studies [11] and a few *in vivo* studies of intact lymph nodes [9] that could be used to suggest a characteristic distance of the diffusion of chemokines or cytokines that might indicate enhanced probability of intercellular contact, it is not known whether these characteristic distance scales are the same in the more complex environment of an intact tumor or for arbitrary combinations of cell types. Furthermore, an exploratory analysis that attempts to analyze proximity between cell types at many different distance scales could suffer from the pitfalls of multiple hypothesis testing, resulting in a false-positive prognostic signature discovery. Several approaches aim to circumvent the choice of a distance scale by taking advantage of the Delaunay triangulation to identify "neighboring" cells [33, 61]. In the Delaunay triangulation [53], cells may have neighbors that are relatively far away, if the local cell density is very low, while cells in a region of high local density will tend to have neighbors that are close together; importantly, the Delaunay triangulation is unique, meaning that there is only one Delaunay triangulation for collection of cells from an MTI image. Because of the uniqueness and the lack of a specific distance parameter, the notion of "neighor" in a Delaunay triangulation is distance-agnostic. This, however, introduces the complication of changing the meaning of proximity to reflect the topology of the triangulation, rather than the absolute distances between cells. In sum, the choice of metrics or distance scales for proximity analysis may not be self-evident and requires careful justification, hindering the development of a general MTI pipeline.

Another complicating factor in the analysis of spatial proximity among cell types in MTI images is the identification of statistically significant proximity, which requires carefully justified *a priori* choice of a null spatial distribution of cells, often in the absence of real MTI data representing a true negative control tissue in which the only difference is the absence of the spatial proximity that is being tested for. Statistical significance can be assessed using an exact formula for some proximity metrics given a certain null hypothesis, or assumption of the meaning of "no proximity"; for example, the K-function [22] may be tested for significance using an exact formula for the p-value if the null hypothesis is that of complete spatial randomness, i.e., that within any arbitrarily shaped sub-region of tissue of area A, the expected number of cells N is equal to λA, where λ is the density of cells in the tissue. For alternative proximity metrics, a permutation test approach can be used. For example, Keren *et. al.* [46] identified cells within 39 μm as neighbors, and then performed a randomization procedure to identify cell-type-pairs that were significantly-spatially-proximal by choosing four tissue

---

[4]https://www.rdocumentation.org/packages/spatstat/versions/1.64-1/topics/Gcross
[5]https://www.rdocumentation.org/packages/spatstat/versions/1.64-1/topics/Kcross

compartments (Immune, Epithelial, Mesenchymal, and Endothelial), and then permuting the labels of cells which shared a tissue compartment 500 times to produce 500 null samples in which cell locations were randomized within their tissue compartment. This approach uses expert knowledge about the expected distribution of cell types in a no-significant-proximity scenario to test for statistical significance, but highlights the challenge inherent to the design of a simulated null distribution in the absence of data from a true null population.

1.3   Chapter Summary

MTI technologies hold great promise for expanding the capabilities of the modern pathologist and improving the clinical management of cancer. However, numerous technical challenges remain to be overcome, including the computational analysis of the data these assays can produce. In particular, the segmentation of single cells as well as tissue structures, the ability to overcome unwanted variation in staining intensity due to preanalytical variables and produce robust cell type classifications, and the rigorous development of prognostic spatial signatures with meaningful significance tests continue to challenge researchers in the field.

In this thesis, I present the results of two projects aimed at addressing several MTI analysis challenges. In Chapter 2, I describe a novel segmentation algorithm for the identification of DCIS lesions in MTI images. Building upon the results of this segmentation algorithm, I describe an algorithm to quantify a specific property called the "continuity" of a particular molecular marker (a.k.a. stain), which produced results that correlated with values estimated by a trained pathologist. These algorithms provide the means of estimating stain continuity in DCIS lesions without the expensive assistance of a pathologist, facilitating the discovery of new potential signatures of biological and clinical importance.

In Chapter 3, I perform a bioinformatics analysis of a PDAC MTI data set, with the goal of assessing the prognostic value of metrics quantifying the spatial proximity between immune and tumor cell populations. Because such metrics have to be chosen *a priori*, I analyze several proximity metrics and compare their prognostic value quantitatively. I also discuss the interpretations of each metric, and demonstrate the different correlations that various metrics of spatial proximity share with the simple cell density metric, which further elucidates the interpretation of these metrics. While this study does not provide a concrete answer about which proximity metric is the "best", it does provide additional evidence for the prognostic value of such metrics that supports the field of PDAC research, and its discussion of the comparison of those metrics could aid computational biologists seeking a proximity metric with a specific interpretation.

In the final chapter, I briefly conclude with a discussion of the impact of this work and my thoughts about future directions and lessons learned.

Chapter 2

Algorithm for automated quantification of continuity of calponin-positive myoepithelium in multiplexed immunohistochemistry images of ductal carcinoma in situ

Abstract

Ductal carcinoma in situ (DCIS) is tumor confined within mammary ducts, surrounded by an intact myoepithelial cell layer that prevents local invasion. A DCIS diagnosis confers an increased lifetime risk of developing invasive breast cancer (IBC) and results in surgical excision with radiation, and possibly chemotherapy. DCIS is known to be over treated, with associated co-morbidities. Biomarkers are needed that delineate patients at low risk of DCIS progression from patients requiring aggressive treatments. Research efforts to investigate the barrier function of the myoepithelium are anticipated to provide much needed insight into DCIS progression. Here, we develop a high throughput technique to assess the loss of myoepithelial differentiation markers. This work is published in the 2018 Proceedings of the International Symposium for Biomedical Imaging [35], and was presented as a poster at the conference held in April 2018 in Washington, DC. This method facilitates automated analysis of a clinically relevant histopathologic feature, as demonstrated by a high correlation with pathologist annotation (r = 0.959), and contributes analytical foundations to a multiplexed immunohistochemistry approach.

2.1   Introduction

Incidence of ductal carcinoma in situ (DCIS) increased from 3% in the early 1970's to 20% of all breast cancers due to widespread use of Screening Mammography [45]. In the U.S this year, $\sim$ 64,000 women will be diagnosed with DCIS. DCIS carries an increased lifetime risk of invasive breast cancer (IBC) with long term studies of untreated DCIS showing 30% progression to IBC [20]. Unfortunately, the clinical accuracy to predict which DCIS will remain indolent and which will progress to IBC remains low, leaving an urgent need to discover biomarkers of progression.

The myoepithelial cell is a 'gate keeper', exerting tumor suppressive effects by secreting tumor suppressive proteins and by forming a physical barrier that prevents intra-ductal tumor cell escape [3, 71, 73, 37]. Variable expression of myoepithelial differentiation markers such as Calponin and α-smooth muscle actin (αSMA) are reported in DCIS, and their loss is used clinically to define IBC [77]. Furthermore, a rodent model of DCIS revealed a progressive loss in myoepithelial differentiation markers Calponin and αSMA prior to transition to invasive carcinoma [77]. To date, none of the known myoepithelial cell proteins have been evaluated for their ability to distinguish between indolent and aggressive DCIS. In order to assess if loss of myoepithelial differentiation markers can serve as a biomarker for identifying gain of invasive tumor cell attributes, we need an unbiased, high throughput approach to quantify myoepithelial protein expression surrounding DCIS lesions.

2.2   Related Work

Histopathologic image analysis is a mainstay for quantitative evaluation of tissue structure and in situ protein expression [94]. The fields of histopathology and basic cancer research are increasingly turning to computational image processing techniques to aid in quantification of complex or subtle biological image features [60]. Quantified image features can then be used to predict relevant clinical attributes, such as disease diagnosis. The need for computational assistance is particularly true in the case of mIHC, where novel staining methods enable 12 or more antibodies to be imaged in one tissue section, substantially increasing pathologist burden [86, 90].

To this end, analysis methods such as Convolutional Neural Networks (CNN), which learn spatial features at multiple scales across an entire image, have been applied to discriminate DCIS from benign lesions in H&E images [74]. CNNs were also used to segment tumor-associated stroma and then to predict DCIS grade or cancer diagnosis [8]. Approaches which compute the morphologic features of cell nuclei, either focusing on epithelial cells, or encompassing all cells in a region of interest, have also found success in predicting DCIS grade or distinguishing between DCIS, benign neoplasia, and normal tissue [74, 93]. Yet, there remains a need for a method that can be applied to current mIHC data to quantify continuity of expression of myoepithelial cell proteins, as focal loss of myoepithelial differentiation may predict increased risk for progression.

Here, we demonstrate automated quantification of the continuity of protein positivity in the myoepithelial layer in DCIS, using a novel image processing pipeline and algorithm. Our automated method achieved high correlation with the manual ground truth, indicating its promise. Furthermore, the methods developed herein, including segmentation of tumor lesions and quantification of protein expression within different tissue regions, could readily be extended to other analysis tasks, particularly those involving analysis of the myoepithelium in DCIS.

2.3 Materials and Microscopy Methods

Formalin fixed paraffin embedded (FFPE) tissue of DCIS from women diagnosed with DCIS only (no evidence of IBC) were obtained through the OHSU Knight Biolibrary in full accordance of the OHSU IRB guidelines (OHSUIRB 4918). Tissue sections were stained with a mIHC method, using antibodies to calponin, alpha-smooth muscle actin (αSMA), and cytokeratin 18 (CK18) [90]. Hematoxylin staining was used to allow single cell image registration and color images were deconvolved with ImageJ as described previously [90].

In cross section of a normal mammary duct, the myoepithelium forms a closed, continuous boundary surrounding a single layer of luminal epithelium, whereas in DCIS, the myoepithelium encloses multiple layers of neoplastic cells. Here we detected tumor cells by CK18, delineated the myoepithelium with αSMA, and evaluated myoepithelial cell differentiation using calponin (Fig. 2.1a). DCIS-associated myoepithelium (n=6 DCIS cases, 27 DCIS lesions) were assessed for percent calponin expression (i.e. continuity), using visual assessment of the IHC signal by three independent investigators including a trained pathologist (S.J.). The annotated values were averaged and then compared to those produced by the computational method described here.

While assessing myoepithelial-layer integrity is routine clinical practice for pathologists to delineate DCIS from IBC, markers to distinguish indolent from progressing DCIS are lacking. We hypothesize that myoepithelial cell differentiation markers may predict risk of DCIS progression, and thus developed a method to assess continuity of the myoepithelial differentiation protein calponin as a test case. We utilized cases of DCIS where the myoepithelial layer was deemed intact. An image key was created by a pathologist (S.J.) to visually assess the percentage of the myoepithelium perimeter where calponin is expressed. For example, if 50% of DCIS myoepithelial cells are lacking calponin, then the lesion would be annotated with a continuity of 50%.

2.4 Results

To reduce reliance on visual pathologic review, we developed several ways to estimate calponin continuity using segmentation and image processing operations. Our segmentation procedure begins with thresholding by Otsu's method [68]. The result is three binary images comprised from αSMA,
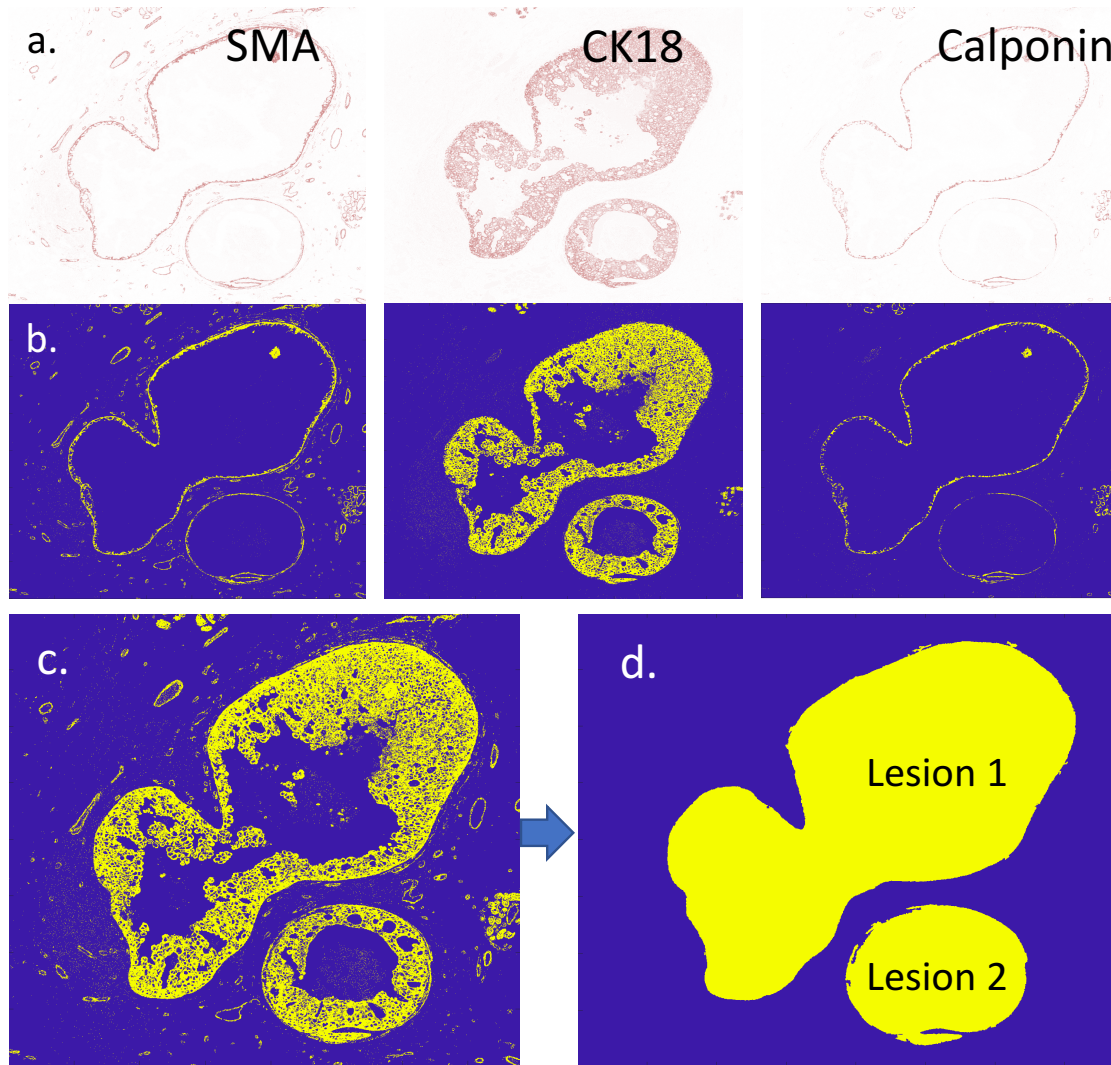
Fig. 2.1 Human DCIS lesion multiplex IHC stained to delineate myoepithelial cell layer (αSMA+) from tumor cells (CK18+), as well as assess myoepithelial cell differentiation (calponin). Otsu's method [68] is applied to color-deconvoluted, multiplexed IHC images (a) producing simple binary masks (b). The union of the masks is found (c) and processed with mathematical morphology to produce a segmentation of each DCIS lesion, including the αSMA- and calponin-positive myoepithelium (d).

CK18, and calponin IHC images (Fig. 2.1b). We refer to these individual IHC-derived images as "masks". These masks are then used to construct two "composite masks" described below, and which are used for geometric quantification.

The first composite mask, called the "DCIS mask" (Fig. 2.1c), is formed using the union of all three previously computed masks. DCIS lesions stained with our three markers tend to be significantly larger than other connected objects in such a binary image, such as noise, myofibroblasts, and blood vessels, so we used morphological operations defined in the MATLAB function "bwmorph ©", including

dilation and erosion, to refine the DCIS mask by smoothing edges and removing small objects. Each connected object (connected is defined here to mean pixelwise 8-connectivity) in the DCIS mask was assigned a unique ID, which was used to match the pathologist's annotation to our geometric calculations. Since the segmentation step was intermediate and prerequisite to the continuity estimation step, we evaluated the segmentation quality visually and only used the continuity estimation as the ground truth for our pipeline.

The second composite mask, called the "calponin mask" (Fig. 2.2a) was found by taking the intersection of the Otsu-thresholded calponin image with the DCIS mask. We refined it by setting pixels equal to zero that were further than 60 pixels (30 μm) inside the DCIS mask boundary - for method development, we chose 30 μm because it is slightly larger than the diameter of a typical myoepithelial cell in our experience. We set this constraint to omit calponin staining that can be found within the lesion center both specifically in papillary type DCIS and non-specifically in necrotic regions, which we found to complicate downstream analysis of the relevant specific staining of myoepithelial cells. Thus, each DCIS mask is associated with a corresponding calponin mask, which contains the region of positive calponin staining for a single myoepithelium.
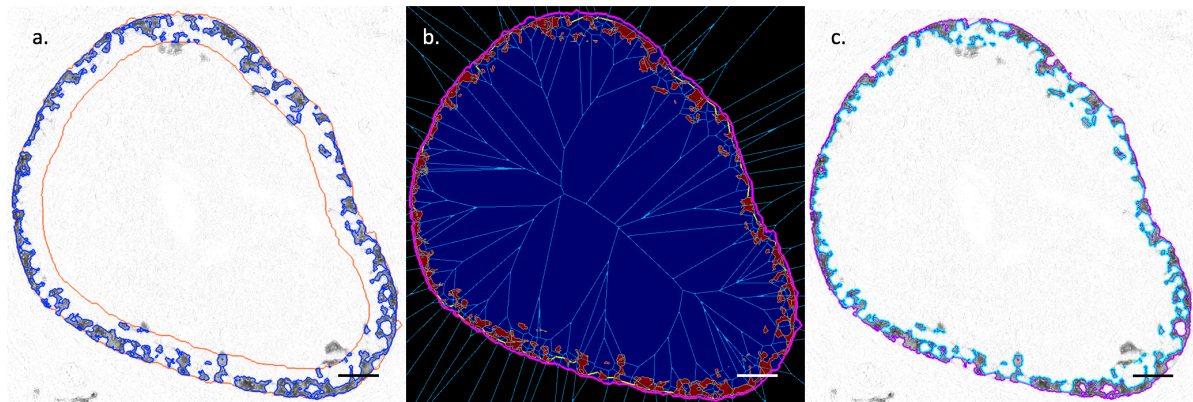

Fig. 2.2 Several intermediate steps of our image analysis method. (a) The grayscale image of calponin staining is shown outlined in blue by the result of "calponin mask" segmentation. The outer orange curve shows the boundary of the "DCIS mask" for this lesion. The inner orange boundary shows the morphological erosion of the DCIS mask. The naïve continuity is the area outlined in blue divided by the area between the orange curves. (b) calponin mask is shown in brown with SLIC superpixel partition. Voronoi cells computed for the superpixel centroids are drawn in light blue, and the DCIS mask is drawn in dark blue and outlined in pink. Connections made across calponin mask gaps are in yellow. (c) The grayscale image of calponin is shown again, with estimated outer (P^O, purple) and inner (P^I, teal) connected calponin mask perimeters shown. Scalebar 50 μm.

To enable continuity evaluation of calponin expression within the myoepithelial cell layer, we constructed line segments along the shortest path between disconnected objects in the calponin mask. In our data, DCIS lesions are frequently non-convex in shape (e.g. Fig. 2.1). Calponin masks also tend to have a high variability in the size, shape, and number of objects, which precludes the approach of

fully connecting all objects, or the use of convex hull or Matlab's Alpha Shape to connect only objects along the outside boundary. We also found that a calponin mask can have a single discontinuity, yet consist of a single fully connected object with one hole, that we could not distinguish mathematically from a perfectly continuous mask. We solved these problems in part by using the Simple Linear Iterative Clustering (SLIC) superpixel method to partition the calponin mask [2].

SLIC, as implemented in MATLAB-compatible C and available on the website of the authors of the SLIC paper [1]), has one parameter called compactness that governs the size and shape of each superpixel. The SLIC algorithm involves clustering of individual pixels using a distance metric that is the weighted sum of the distance between pixels in color space and the distance in image space (i.e., physical distance), and the compactness parameter governs the tradeoff between color and physical distance in computing this metric. Larger values of the compactness parameter produce smaller superpixels; we used a compactness of 100 based on empirical observation that the resulting superpixels were approximately the size of an epithelial cell, or about 10μm wide. By partitioning the calponin mask this way, we hoped to detect a single discontinuity in an otherwise topologically connected object, and also to provide a basis for constraining a search for discontinuities to neighboring regions of the mask.

To define local, or neighboring image regions, we first found the Voronoi tessellation of the set of centroids of all superpixels (Fig. 2.2b) [53]. This returns a one-to-one mapping of superpixels to Voronoi cells. We designated all superpixels whose Voronoi cells were touching as neighbors and as candidates for constructing a connecting line segment between them. It can also be stated that superpixels were called neighbors if they were connected by an edge in the Delaunay triangulation, which is the dual of the Voronoi tesselation [53].

Finally, we constructed the set of all "discontinuity" line segments by connecting a subset of neighboring superpixels along the shortest path between their respective perimeters. We connected two neighboring superpixels if they were not touching, and if their Voronoi cells were bridged by the DCIS mask boundary (pink line in Fig. 2.2b). We define the set L of superpixel pairs satisfying these properties as follows. Let $B(t)$ be an ordering of the pixels in the DCIS mask boundary, such that $B(t)$ is touching $B(t+1)$ for all $t = 1, \ldots, p-1$ where p is the number of pixels in the DCIS mask boundary, and $B(1)$ is touching $B(p)$. This is the precise output of the Matlab function "bwboundaries ©" on the DCIS mask. We denote the ID of the Voronoi region containing the pixel $B(t)$ as V_t, and write the list L of superpixel pairs as:

$$L = \{(V_t, V_{mod_p(t+1)})\}|V_t \neq V_{mod_p(t+1)}, 1 \leq t \leq p\} \tag{2.1}$$

We use the modulo function $mod_p(t+1)$ to denote the fact that pixels $B(1)$ and $B(p)$ are touching (8-connected). We performed a final refinement by discarding all of the line segments that do not

---

[1]https://www.epfl.ch/labs/ivrl/research/slic-superpixels/

intersect the outer perimeter of the resulting connected calponin mask (purple curve in Fig. 2.2c), and we denoted the refined set of "discontinuity" line segments $\{d_i\}$.

Since our continuity estimation is somewhat complex, we compared its performance to a simple method. This simple method, which we call the naïve continuity, is found simply by taking the ratio of calponin-positively-stained myoepithelium mask area to the total myoepithelium mask area, using a myoepithelial width of 30 μm. We defined

$$continuity\,(naive) = 100 * \frac{Area_{calponin}}{Area_{myoepithelium}} \tag{2.2}$$

where the numerator is the area of the calponin mask for the singular lesion in question, and the denominator is the area of the region between the DCIS mask boundary for the same lesion and the erosion of that boundary by 60 pixels (30 μm) (Fig. 2.2a). We found that the naïve estimate of continuity had a good correlation with the pathologist's annotations (Fig. 2.3 left), but that it tended to underestimate the continuity, especially for lesions with low continuity.

To demonstrate our improved method using the discontinuities using our superpixels and Voronoi tesselation method, we compared the total length of the discontinuities to an estimate of the perimeter of the myoepithelium. The line segments $\{d_i\}$, when added (i.e. set union) to the calponin mask, create an (8-)connected object with a hole for which inner and outer perimeters P^O and P^I are measurable (Fig. 2.2c). In the case of an object with more than one hole, we compute PI as the inner perimeter of the largest hole. We took the average of these inner and outer perimeters to approximate the total perimeter of the myoepithelium: $P \cong \left(P^I + P^O\right)/2$, and finally computed

$$continuity = 100 * \left(1 - \frac{2\sum_{i=1}^{n} d_i}{P^I + P^O}\right). \tag{2.3}$$

We found that the naïve method (Eq. 2.2) achieves a high correlation (R = 0.791) with the mean visually annotated continuity of Calponin. Abdallah, et. al. recently reported computational measurements of the sensitivity of Calponin, P63, CD10, and αSMA in demonstrating the myoepithelium of DCIS lesions, with results comparable to our "naïve method" [1].

The proposed method (Eq. 2.3) achieved a Pearson's-correlation of R = 0.959 with professional annotation. We also observed some discordance between pathologist and our method's assessments. One outlier (Fig. 2.3 right) was annotated at 10% continuity by pathologist assessment, but found to be 50% continuous by our computational method. This sample had faint staining and low signal-to-noise ratio, so it is likely that the discrepancy resulted from the choice of threshold by Otsu's method, which might be corrected using sample intensity normalization and flat threshold choice [48], or a more advanced segmentation method.
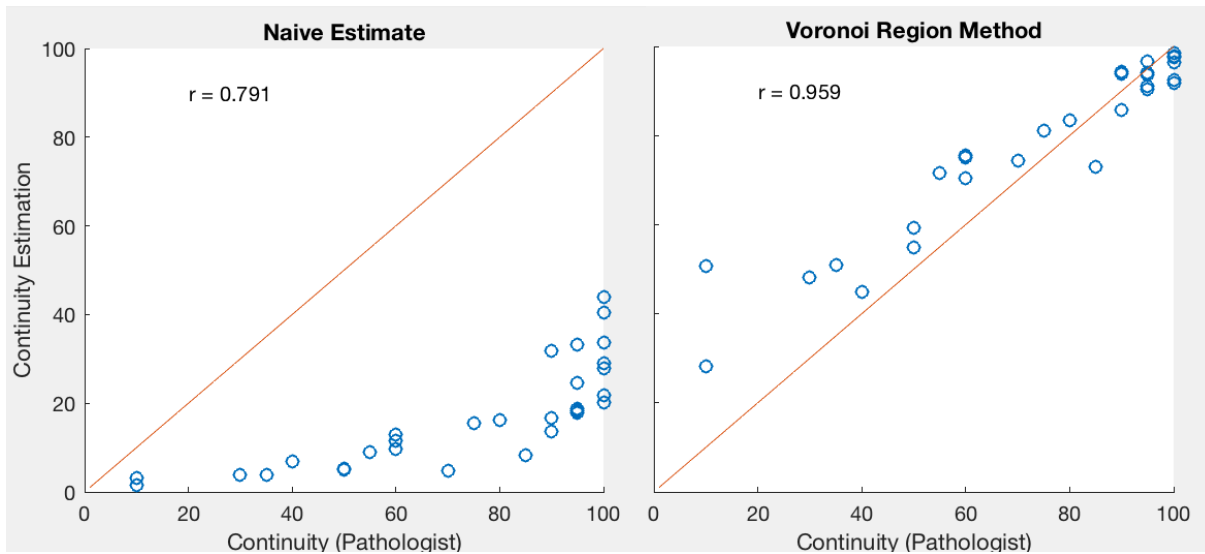
Fig. 2.3 Correlation between analytical calponin continuity estimation and pathologist's annotation for Naïve estimate and the proposed method.

2.5   Discussion

One drawback of our method is the reliance on segmentation of the DCIS mask boundary. Central necrosis or tissue-loss during the staining process were found to be the primary confounders of segmentation, although we note that cases like this would be difficult to analyze traditionally as well. Segmentation results must be either checked visually (as we did here), or filtered for artifacts. Future work to ensure the robustness of segmentation, for example by filtering out images with severe tissue loss during imaging, should be done to take full advantage of our method.

This study describes a computational method to assess continuity of calponin in mammary myoepithelium in cases of DCIS. This method, closely matches what can be detected by the trained human eye but in significantly reduced time and in the absence of observer bias. This spatial-structural analysis of continuity of myoepithelial differentiation proteins combined with multiplex IHC technique promises the capacity for integration of myoepithelial, immune cell and tumor intrinsic factors, which may help predict risk of DCIS progression to invasive disease.

Acknowledgements

and helped to analyze the data. Young Hwan Chang and Elliot Gray invented the algorithm, wrote the computational parts of paper, and edited the paper.

Funding

Chapter 3

A novel spatial proximity metric for multiplexed tissue imaging and its value for predicting survival or presence of tertiary lymphoid structures in pancreatic ductal adenocarcinoma

Abstract

Pancreatic cancer is a deadly disease for which new treatments are needed. The tumor microenvironment is known to contain tumor-specific T cells that could be harnessed for therapeutic benefit, but those T cells are spatially excluded and functionally repressed by other elements of the stroma. Tertiary lymphoid structures are functional aggregates of leukocytes that have been linked to immunotherapy response in melanoma and survival in pancreatic cancer, but whether they influence the spatial exclusion or functional activation of T cells in pancreatic cancer is unknown. Furthermore, while the spatial proximity of CD8$^+$ T cells to neoplastic epithelium has been linked to patient survival, it is unknown whether proximity among other leukocyte types is associated with survival, or dependent on the presence of tertiary lymphoid structures. In order to answer these questions, we studied a cohort of human pancreatic tumor resection specimens that were subjected to multiplexed immunohistochemistry for the purposes of auditing immune contexture in situ. We illustrate how existing metrics for spatial proximity are correlated with absolute cell densities, and present a new metric of relative proximity that does not have the same correlation, thus making it uniquely interpretable. Using a machine learning approach, we found that relative proximity and functional immune features worked together to predict the presence of tertiary lymphoid structures, and patient survival. Our approach improves

basic knowledge of the pancreatic tumor microenvironment and our new relative proximity metric can be used in future multiplexed imaging studies.

3.1   Introduction

Building on the foundation of successful image-based cancer biomarkers such as the "Immunoscore" [69], multiplexed tissue imaging (MTI) technologies are gaining traction as tools for cancer biomarker discovery [54, 90, 33, 47, 25]. Researchers can now take advantage of single-cell segmentation to quantify the spatial organization of multiple cell populations in a tissue, which has the potential to provide information about tissue organization that can independently predict patient survival [79, 33, 47, 13]. However, because MTI can resolve an unprecedented level of spatial and molecular detail, the complexity of data analysis poses new challenges. In particular, quantification and interpretation of spatial proximity (we use the word proximity generally to refer to a family of metrics) between multiple cell populations, which is thought to represent enhanced probability of contact-dependent interaction [27], is challenging. This challenge is one of interpretation, because of the diversity of epithelial, endothelial, stromal, and immunological structures that may exist simultaneously within a tumor, and the intrinsic dependence of distances between cells of different types on the cellular composition of the tissue.

The need to overcome this data analysis challenge pertains to the study of pancreatic ductal adenocarcinoma (PDAC). PDAC, while relatively rare compared to e.g. colon or breast cancer, is predicted to soon become the leading cause of cancer mortality in the US due to its tendency to be diagnosed at an advanced stage and the lack of effective treatments for advanced disease [49, 10]. MTI is a valuable tool to study PDAC because it enables profiling of the tumor in its microenvironmental context, which is histologically complex and is known to influence the efficacy of both FDA-approved as well as experimental therapies through the activities of stromal fibroblasts and infiltrating immune cells, which remain incompletely understood [64].

Spatial proximity, at the scale of 20-100μm, between tumor-infiltrating CD8$^+$ T cells and PDAC neoplastic epithelial cells, has been previously linked to patient prognosis using computational analysis [24, 16, 58]. Additionally, dense aggregates of CD20$^+$ B cells, versus diffuse intratumoral B cell infiltrates, were linked with improved PDAC patient survival [17]. We hypothesized that a metric of spatial proximity among different leukocyte types and PDAC neoplastic epithelial cells that was decoupled from cell densities would distinguish patients with long term survival while deepening our understanding of the prognostic value of these spatial arrangements by way of improved interpretability.

Members of the Coussens lab previously developed multiplexed immunohistochemistry (mIHC) [90], a novel form of MTI, and recently used it to generate a tissue atlas of the tumor immune microenvironment of PDAC [56]. Here, we leverage this PDAC data resource to demonstrate the correlation between density of certain leukocyte or neoplastic epithelial cell types in PDAC and several

standard metrics of spatial proximity among pairs of them. We then introduce a new metric of relative spatial proximity, and show that it is decorrelated from cell density. We investigate this new metric, along with cell density measures and measures of the functional activity of T cells and B cells, in a machine learning analysis that links the single-cell spatial organization of PDAC tumors to long-term survival of patients. Further, we also apply our machine learning model to the classification of the same tumors according to the presence of tertiary lymphoid structures (TLS), which are immunological structures that are thought to represent enhanced antitumor immunity and formation of immunological memory [29]. Our study elucidates the prognostic value of T cells and B cells in the PDAC tumor microenvironment, and contributes a new metric that can be applied in future tissue imaging studies.

3.2   Related work in spatial proximity analysis

In order to address our biological hypotheses, we needed a clearly interpretable metric for spatial proximity between cell types. However, as we will illustrate in Fig. 3.1, the interpretability of some metrics of proximity between two cell types is complicated by an association with the number of cells per unit area, i.e. their density. In Fig. 3.1, we simulate two different cell types (reference and target), whose locations are drawn independently uniformly at random within a box, so there is by definition no spatial dependence between them. In Fig. 3.1B and 3.1C, there are two examples that differ in the density of target cells, but not of reference cells. Because of the varying density of target cells, reference cells in Fig. 3.1C tend to be closer to target cells than those in Fig. 3.1B (quantified in Fig. 3.1D). This simple example illustrates our motivation to use a more interpretable metric of spatial proximity, in the sense that it should not be clearly dependent on cell density.
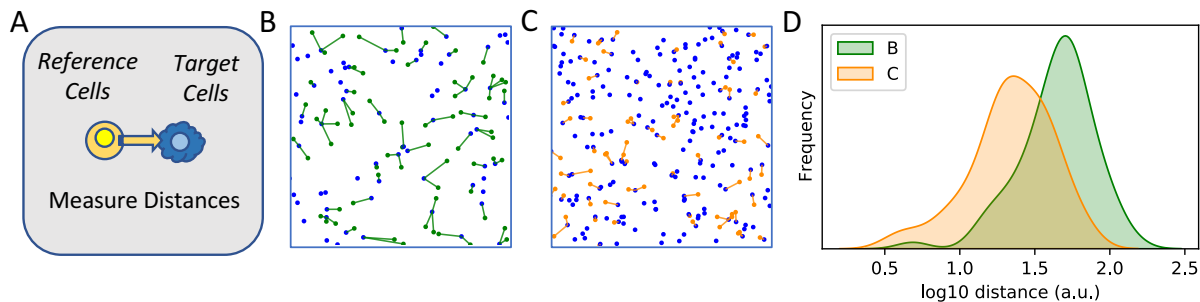
Fig. 3.1 Example of the association between cell density (number per unit area) and proximity (e.g. nearest-neighbor distances). **A)** Measuring proximity between two cell types (reference and target) fundamentally depends on the computation of distances, in this case from reference cells to their nearest target neighbors. **B)** Simulated reference and target cells, sampled completely at random spatially, with lines connecting reference cells to their nearest target neighbors. **C)** A similar sample, but with three times as many target cells as (B). **D)** The nearest neighbor distances in (C), from reference to target, are shorter than those in (B) due to the increased density of target cells in (C).

A number of spatial proximity metrics exist that have been used or developed for analyzing cells in tissue images, however, we found unique challenges associated with each of them that motivated us to develop a new metric. In general, the relevant methods had a common feature: they compare observed nearest-neighbor distances or near-neighbor counts in the observed data to those generated in randomly simulated data. They differ in the way they define the random data, illustrating the central importance of this concept.

Schapiro *et. al.* developed HistoCAT [79], in which the observed proximity among cell types is compared to a series of simulated random control tissues by permuting the labels (i.e., identities) of cells, and a proximity metric is defined as the z-score of neighboring cell counts at a certain spatial distance. The authors used this approach to distinguish breast tumors according to their grade in an unsupervised clustering analysis, showing that the metric could effectively represent the spatial disorganization of poorly differentiated epithelial tumors. However, we argue that interpreting this metric is challenging because the simulated random control tissues do not necessarily resemble actual tissue. The fact that the cell labels are shuffled at random to create the "negative control" data means that important architectural properties of the tissue could be lost, for example, the amount of intercellular adhesion of epithelial cells in PDAC (possibly represented by the formation of ducts or glands), which can be lost by some tumor cells (marked by solitary cells that do not express E-cadherin), is negatively associated with patient survival [59].

Helmuth *et. al.* [40, 84] developed a statistical framework that simulates structured random data by "smoothing" the observed spatial locations of cells with a Gaussian convolution operation, which returns a non-uniform probability distribution based on the data. This distribution represents an estimate for the local cell "density", parameterized by the smoothness parameter used in the convolution. Random samples of spatial locations for the cells can then be drawn from this distribution, and the observed data can be statistically compared to these samples using a proximity metric to determine whether it has significantly more or less proximity than the random samples. While similar in nature to the cell label permutation used in HistoCAT, this procedure requires the user to control the level of spatial smoothness in distribution of the random control tissues by the choice of a characteristic distance for the Gaussian kernel width. While this approach can be used to maintain the realism of the simulated point pattern, it is not clear how best to choose the important kernel parameter.

Keren *et. al.* [46] use a permutation test approach named "context-dependent spatial enrichment", which permutes target cell type labels only among other cell types that belong to a similar tissue compartment, defined as epithelial, mesenchymal, endothelial, and immune. Despite the limitation of needing to choose the tissue compartments, this approach effectively uses prior knowledge to overcome the direct dependence on target cell density while preserving overall tissue architecture in the simulated random control data. However, this approach does not directly take into account the effect that extremely low cell densities could have on the interpretation of the metric. For example, in the (extreme) hypothetical case that a user-defined tissue compartment consists of cells of only one type, the random permutation will have no effect and the control tissues will all be extremely similar

to the real tissue. Practically speaking, this approach could be improved by taking into account the absolute cell counts when performing spatial analysis, to elucidate the interpretation of z-scores.

Here, we define a new metric by directly comparing proximity values, between different reference cell types to a common target cell type. Using a statistical power analysis, we explicitly account for samples containing cell densities too low for quantification of significant spatial proximity, providing a guarantee that an absence of cells is not interpreted as an absence of relative spatial proximity. We demonstrate that this metric is not correlated with target cell density in PDAC mIHC data, unlike other proximity metrics. Finally, we use machine learning models to test our hypotheses (illustrated in Fig. 3.2) about how the formation of tertiary lymphoid structures in PDAC, and the improved prognosis of patients, are related to spatial and functional features of intratumoral leukocytes. The Python code we use to compute those features from processed single-cell data, as well as the functions used for machine learning, are available online at https://github.com/elliotmgray/Gray2021_code.
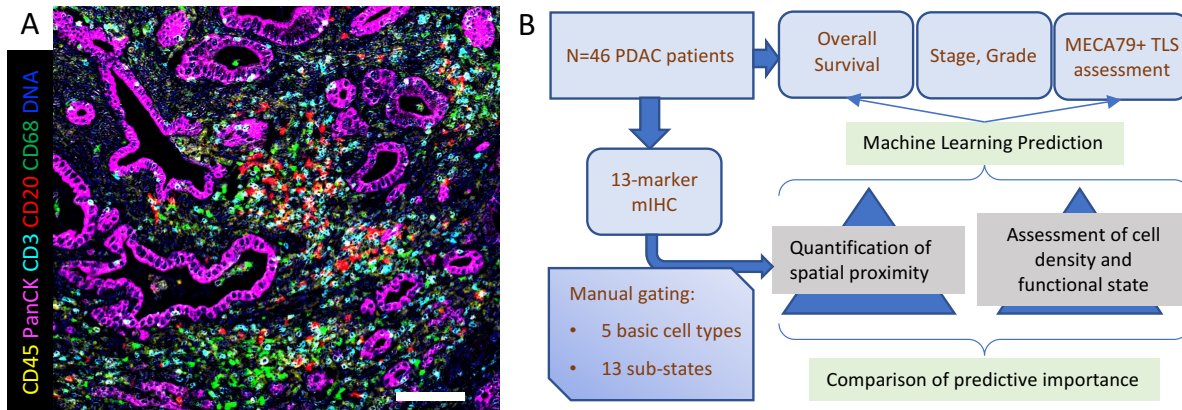
Fig. 3.2 **A)** Image of tumor core from a PDAC specimen stained with mIHC, displaying an agglomeration of lymphocytes in the tumor stroma, surrounded by infiltrating $CD68^+$ myeloid cells (scale bar 200μm). **B)** Analysis workflow. We utilized an mIHC data set to study the relative importance of spatial proximity features and density or functional state features with respect to patient survival and the presence of tertiary lymphoid structures in their tumor.

3.3 Materials and Methods

3.3.1 Patient cohort and multiplexed immunohistochemistry (mIHC)

PDAC surgical resection samples (N=46) were collected at OHSU from patients with no prior history of chemotherapy. Tumor samples were acquired with informed consent and OHSU Institutional Review Board (IRB) approval under IRB protocol #3609. Formalin-fixed, parrafin-embedded (FFPE) tissue sections were subject to mIHC staining, and areas containing invasive carcinoma and immune aggregates were digitally annotated by a pathologist, as described in [56]. mIHC images were processed

and cell type identification was performed by image cytometry in FCS Express 6 Image Cytometry RUO software using manual gating as described in [56]. Lymphoid aggregates were identified with pathologist assistance and verified quantitatively for MECA79 staining for classification as TLS. The image shown in Fig. 3.2A is for illustrative purposes and was manually rescaled and contrast adjusted independently of image quantification.

3.3.2   Calculation of frequency of positivity of leukocyte functional markers

The frequency of leukocyte and epithelial cell types, including states defined as positive or negative for markers of cellular function (e.g., Ki67) were defined according to Table 3.1. When computing the frequency of a cell type, the number of cells were divided by a chosen "base cell type", indicated in the right column of the table. The denominator for each cell type was chosen based on the cell type definition hierarchy: e.g., the frequency of $Ki67^+$ $CD8^+$ T cells was expressed as a fraction of the count of $CD8^+$ T cells. Notably, putative helper T cells that are throughout this paper referred to as "$CD4^+$ T cells", were defined with the use of a $PanCK^-$ $CD45^+$ $CD3^+$ $CD8^-$ marker combination, and not with our CD4 antibody, due to the quality of the CD4 antibody staining that was assessed. Finally, note that cells that resided in TLS were not counted when quantifying tumor-infiltrating leukocytes.

Cell Type	Protein Marker	Base Cell Type
Neoplastic Epithelium	$PanCK^+$	n/a
Leukocyte	$PanCK^-$ $CD45^+$	n/a
$CD4^+$ T cell	$CD3^+CD8^-$	Leukocytes
$CD8^+$ T cell	$CD3^+CD8^+$	Leukocytes
B cell	$CD3^-CD20^+$	Leukocytes
Monocyte or Macrophage	$CD68^+$	Leukocytes

Function	Protein Marker	Base Cell Type
Proliferation	$Ki67^+$	$CD4^+$ or $CD8^+$ T cells
Cytotoxicity	Granzyme B (Gzmb)$^+$	$CD4^+$ or $CD8^+$ T cells
Antigen Inexperienced	$CD27^-$ IgD^+	$CD20^+$ B cells
Innate-like Memory	$CD27^+$ IgD^+	$CD20^+$ B cells
Class-Switched Memory	$CD27^+$ IgD^-	$CD20^+$ B cells
$CD8^+$ T cell activation	$Eomes^+$	$CD8^+$ T cells
Early or Late Effector	$PD1^+$	$Eomes^+$ or $Eomes^-$ $CD8^+$ T
Presumptive helper T	$CD4^+T$, $PD1^+$	$CD4^+$ T cells
Immunoregulatory	$ICOS^+$	$PD1^+CD4^+$ T cells

Table 3.1 Classification & quantification scheme for each cell type.

3.3.3   Statistical power for spatial proximity analysis

Prior to analysis, patients and cell types were selected based on a statistical power analysis for the Mann-Whitney U test. A threshold was selected for the total cell count in order to ensure that sufficient cells were present to identify statistically significant spatial proximity under the assumption of existence of an underlying process driving that proximity with a known effect magnitude [75]. The power analysis was conducted in G* Power [26], assuming equal numbers of reference cells of each type (note that for this test, statistical power is an increasing function of the number of reference cells of each type, so assuming equal numbers gives us the worst-case scenario), a medium effect size (that is, Cohen's d) of 0.5, a power of 85%, and $\alpha = 0.15$. This resulted in a threshold of approximately 50 cells of each type. In 39 of the 46 patients, $CD4^+$ and $CD8^+$ T cells, B cells, and $PanCK^+$ epithelial cells were present in sufficient numbers (Fig. 3.3); all other cell types identified in our data set were too rare for statistically significant comparisons, and the 7 patients with insufficient cell of the selected types were not included in machine learning studies.

3.3.4   Calculation of spatial proximity metrics

In this section, cell types are named as either "reference-type" or "target-type" (illustrated in Fig. 3.1A), meaning that proximity was computed *from* reference cells *to* nearby target cells. Distances from single cells of a reference type to their k nearest neighbors of a target type (later referred to as "nearest neighbor distances" or just "neighbor distances") were computed. For each reference cell, the k neighbor distances were averaged. The counts of target cells within a radius r of each reference cell were computed (later referred to as "neighbor counts"). Neighbor counts at a distance r were chosen to represent interactions at a known distance scale, while (k) neighbor distances were chosen as an unbiased quantification of cell type proximity. Neighbor distances and counts were summarized for each patient, using the median for neighbor distances to reduce the potential impact of outliers, and using the mean for neighbor counts due to the discrete nature of counts.

The number k of nearest neighbors and the neighbor radius r were set to five and 40μm, respectively. A 40μm radius was chosen to represent an enhanced probability for cell-cell contact [24, 16, 47, 34]. A value of $k = 5$ was chosen empirically, instead of e.g. $k = 1$, to reduce the impact that solitary target cells would have on k-nearest-neighbor distances; in practice, choices of k between 1 and 5 produced similar proximity measurements for the cell types selected for spatial analysis. In our downstream analysis, there were no patients who had fewer than 5 cells of a type used in distance calculations.

A score of spatial mixing between pairs of cell types at predefined distance scales, called the "likelihood ratio" (*LR*) of counts of pairs of neighboring cells [33, 34], was defined as follows: for each image, and for a given radius r, a graph was constructed in which each pair of cells was connected by an edge if their spatial locations were closer than r in distance. Then, defining N_A as the number
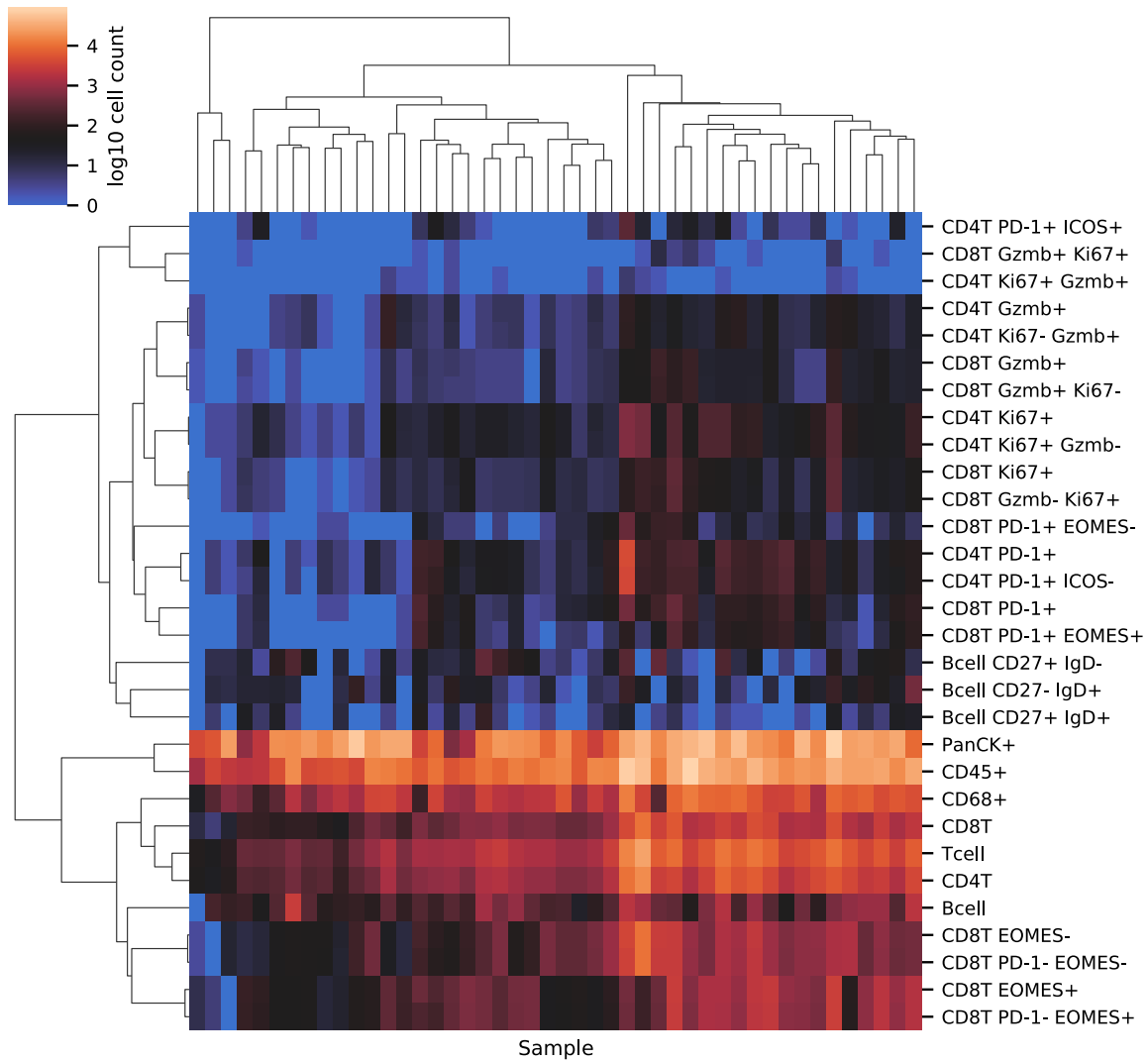
Fig. 3.3 Heatmap shows the absolute number of cells of each type that are present in each patient's tumor sample. Absolute cell counts were used in the statistical power analysis prior to selecting patients and cell types to analyze; note that the counts reflect the actual tissue size so comparison of patients this way not biologically meaningful.

of edges incident to nodes of type A, and N_{AB} as the number of edges connecting nodes of type A to nodes of type B, the likelihood ratio is given by:

$$LR_{AB}(r) = \frac{(N_A(r) + N_B(r)) \cdot N_{AB}(r)}{2 \cdot N_A(r) \cdot N_B(r)} \tag{3.1}$$

In previous works this metric was log-transformed [33, 34]; here, the log-transform was not used. In the downstream analysis that is presented, a log-transform would be beneficial to machine learning performance if the variables had a highly right-skewed distribution, since the model that was chosen (Extremely Randomized Trees) samples thresholds uniformly at random within the min/max range of the feature [32]. However, while we did not observe highly skewed distributions of LR (an example can be seen in Fig. 3.4B), the log transform did tend to produce outliers and so we omitted that step in the context of our machine learning analysis.

Defining also the number of nodes of types A and B, respectively, are n_A and n_B, and the area of the image or study region is $area$, we give Ripley's K function (neglecting edge-effect correction) [22] for purposes of illustration; Ripley's K is not used in our study but is very similar in interpretation to LR and has been used previously to study PDAC tumors stained with MTI [16].

$$K_{AB}(r) = \frac{area \cdot N_{AB}(r)}{n_A \cdot n_B} \tag{3.2}$$

"Fold change" is a metric used to quantify the relative difference in a measurement between two random samples. In order to de-correlate spatial proximity measurements from the underlying cell densities, we proposed a new relative proximity metric based on fold changes. Our new metric, the fold change of median neighbor distance (Eq. 3.3), or mean neighbor count (Eq. 3.4) between two reference cell types, was defined:

$$FoldChange_{distance}(Ref1, Ref2, Targ, k) = \frac{MedianDist._k(Ref1 \dashrightarrow Targ)}{MedianDist._k(Ref2 \dashrightarrow Targ)} \tag{3.3}$$

$$FoldChange_{count}(Ref1, Ref2, Targ, r) = \frac{MeanCount_r(Ref1 \dashrightarrow Targ)}{MeanCount_r(Ref2 \dashrightarrow Targ)} \tag{3.4}$$

where $Ref1$, $Ref2$, and $Targ$ indicate reference or target cell types, k indicates the number of nearest neighbors whose distances were averaged, and r indicates the distance below which two cells were counted as neighbors. When one reference cell type was a subset of the other, reference cells were not double-counted (e.g., for CD8$^+$ T cells and CD45$^+$ leukocytes, only CD45$^+$CD8$^-$ leukocytes were compared to CD8$^+$ T cells). As an example of our metric, given that PanCK$^+$ cells are the target, CD8$^+$ T cells are the first reference, and all other CD45$^+$ cells are the second reference, it would be correct to say that a fold change value greater than one would indicate that CD8$^+$ T cells are closer to PanCK$^+$ cells relative to all other CD45+ leukocytes.

3.3.5   Machine learning approach

For machine learning classification of patients, five categories of features were used, outlined in Table 3.1 and supplementary tables 3.2, 3.3, and 3.4, quantifying cell frequencies, densities, proximity likelihood ratios, and fold changes (neighbor distance and neighbor count are separate categories), respectively. For rationale behind choices of reference and target cell types, see section 3.3.6.

Cell Type
Bcell
$CD4^+T$
$CD8^+T$
$CD68^+$
$PanCK^+$

Table 3.2 Cell types quantified using density (number of cells per mm^2).

Reference	Target
$CD68^+$	$CD4^+T$
$CD8^+T$	$CD4^+T$
$CD4^+T$	Bcell
$CD8^+T$	Bcell
$CD68^+$	Bcell
Bcell	Bcell
$CD4^+T$	$PanCK^+$
$CD8^+T$	$PanCK^+$
Bcell	$PanCK^+$

Table 3.3 Cell type pairs for which proximity was quantified using the likelihood ratio (Eq. 3.1), with $r = 40\mu m$.

For classification of patient tumors according to long or short survival, patients were dichotomized according to median overall survival time, which was 20.4 months among those selected for survival analysis. For classification by presence of TLS, patients with at least one identified immune aggregate were considered as belonging to the positive group. The model for patient classification was chosen to be Extremely Randomized Trees (implemented in scikit-learn), based on its demonstrated superior accuracy and reduced tendency to overfit small data sets compared to similar machine learning algorithms, including Random Forests [32, 12]. Model performance was evaluated with leave-one-out cross-validation due to the limited data set size, and model hyperparameter tuning was performed using nested cross-validation and automated parameter search with the python package bayesian-optimization [67]. Classification performance was evaluated by the harmonic mean of the area under

Ref. 1	Ref. 2	Target
$CD8^+T$	$CD45^+$	$CD4^+T$
$CD68^+$	$CD45^+$	$CD4^+T$
$CD4^+T$	$CD45^+$	Bcell
$CD8^+T$	$CD45^+$	Bcell
$CD68^+$	$CD45^+$	Bcell
Bcell	$CD45^+$	Bcell
$CD4^+T$	$CD45^+$	$PanCK^+$
$CD8^+T$	$CD45^+$	$PanCK^+$
Bcell	$CD45^+$	$PanCK^+$

Table 3.4 Cell type triplets for which relative proximity between reference types with a common target type was quantified using either neighbor distance or neighbor count fold changes.

the receiver-operating-characteristic curve (ROC AUC) and the average precision (AP), again using functions defined in the sklearn python library. Recursive feature elimination with cross-validation (RFECV) was used to rank individual features and identify subsets of features with improved predictive value.

3.3.6   Selection of reference and target cell types based on prior knowledge

Given the results of our power analysis, we chose combinations of either Reference and Target (for likelihood ratios), or Reference 1, Reference 2, and Target (for fold changes), from among $CD20^+$ B cells, $CD8^+$ T, $CD4^+$ T, $CD68^+$ myeloid, and $PanCK^+$. In order to normalize proximity among leukocytes and epithelial cells to the overall distribution of the immune infiltrate, we used $CD45^+$ leukocytes as Reference 2 in all fold change metrics. We selected the combinations of cell types in Tables 3.3 and 3.4 based our previous knowledge and hypotheses about the known or potential interactions these cell types might have, in line with our previous work [34]. We used the same combinations of cell types in both long-term survival prediction and TLS presence prediction.

B cells have context dependent function and could be potentially either positive or negative regulators of tumor cell killing. They may act as professional antigen presenting cells, so their increased proximity to T cells could be associated with improved patient prognosis. Diffuse intraepithelial localization of B cells has been linked to poor outcomes in PDAC [17], so proximity between $PanCK^+$ cells and B cells is of hypothetical interest. Meanwhile, B regulatory cells may act as potent immuno-suppressive regulators, and can support tumor growth and improved prognosis in PDAC tumors via interaction with tumor associated macrophages [38]. Therefore, we hypothesize that proximity among B cells and $CD68^+$ myeloid cells should be associated with poor prognosis.

Both $CD4^+$ and $CD8^+$ T cells are capable of tumor cell killing, and their density within tumors and direct proximity to $PanCK^+$ neoplastic epithelial cells have previously been linked to PDAC

patient prognosis [58, 16]. $CD4^+$ helper T cells provide signals to $CD8^+$ T cells that support the formation of memory phenotype [82], and so we might expect that proximity among these cell types would be a positive prognostic factor and possibly associated with the presence of TLS or with effector function of $CD8^+$ T cells. However, T-regulatory cells ($FOXP3^+$ $CD4^+$ T cells), which are prevalent in advanced PDAC lesions, can suppress antitumor activity of $CD8^+$ T cells through a contact-dependent mechanism and are associated with worse prognosis [42], so spatial proximity between $CD8^+$ and $CD4^+$ T cells within PDAC tumor areas could be either positively or negatively associated with patient survival.

Lastly, because previous studies have shown Th2-polarized $CD4^+$ T cells interact with macrophages to promote tumor survival and metastasis in a breast cancer mouse model [21, 83], we hypothesized that proximity between $CD68^+$ myeloid cells (including macrophages) and $CD4^+$ T cells would also be linked with patient overall survival.

3.4 Results

3.4.1 Distance fold change is an unbiased metric of relative proximity

In human PDAC tumors, several metrics of spatial proximity are correlated with the density of the cell types in question. Fig. 3.4 illustrates this for proximity between $CD8^+$ T cells and $PanCK^+$ cells: both the median nearest-neighbor distance (3.4A), and the likelihood ratio (3.4B), are correlated with $PanCK^+$ cell density. As we investigated the impact of spatial proximity between these and other cell types on patient prognosis, independent of the densities of those cells, we sought an alternative definition of proximity that would have a reduced correlation.

We hypothesized that the correlation between $PanCK^+$ cell density and the proximity (e.g., likelihood ratio of proximity to $CD8^+$ T cells) could be removed, on average across our patient cohort, by normalizing by the proximity of $PanCK^+$ cell to all other $CD45^+$ leukocytes. To test our hypothesis, we computed the median distance from both $CD8^+$ T cells and all other $CD45^+$ leukocytes to their nearest five $PanCK^+$ neighbors, and found that the ratio (fold change) of these values was uncorrelated with $PanCK^+$ cell density (Fig. 3.4C). Indeed, we found that increasing density of target cells is associated with higher neighbor counts (Fig. 3.5A), shorter neighbor distances (Fig. 3.6A), and also some higher likelihood ratios (Fig. 3.5B), for a variety of combinations of reference and target cell types. Meanwhile, correlation between cell density and fold changes of either neighbor counts (Fig. 3.5C) or neighbor distances (Fig. 3.6B) is in most cases insignificant. This suggests that fold change is a robust metric of proximity between cell types in the sense that it is not correlated with cell density.

Some of the correlations between cell densities and fold changes were significant; there was significant correlation between the densities of T cells (and to a lesser extent, also $CD20^+$ B cells and $CD68^+$ cells) and the fold change in proximity between $CD8^+$ and $CD4^+$ T cells. The data in Fig. 3.5C show that increasing density of T cells is associated with a greater magnitude of *relative* proximity
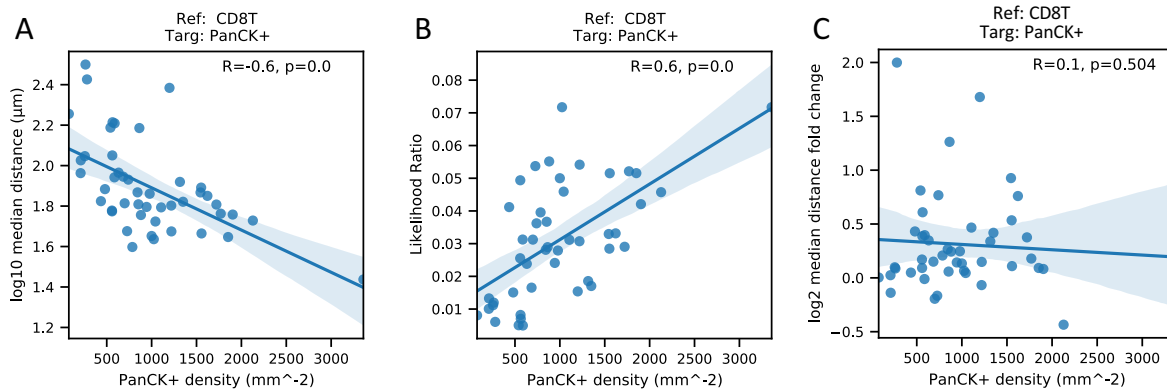
Fig. 3.4 **A)** The density of PanCK$^+$ cells is inversely correlated with the median distance from CD8$^+$ T cells to their nearest PanCK$^+$ neighbors, as expected. **B)** Somewhat surprisingly, the density of PanCK$^+$ cells is also positively correlated with the likelihood ratio (Eq. 3.1) of proximity between CD8$^+$ T cells and PanCK$^+$ cells at 40μm distance, despite this metric taking into account the number of cells of both types. **C)** Meanwhile, the log fold change in median distance to PanCK$^+$ cells, with CD8$^+$ T cell as reference 1 and CD45$^+$ leukocytes (not including CD8$^+$T cells) as reference 2, has no such correlation with PanCK$^+$ cell density. R is Spearman's rank correlation, and p-values for significant correlation are rounded to the thousandth decimal.
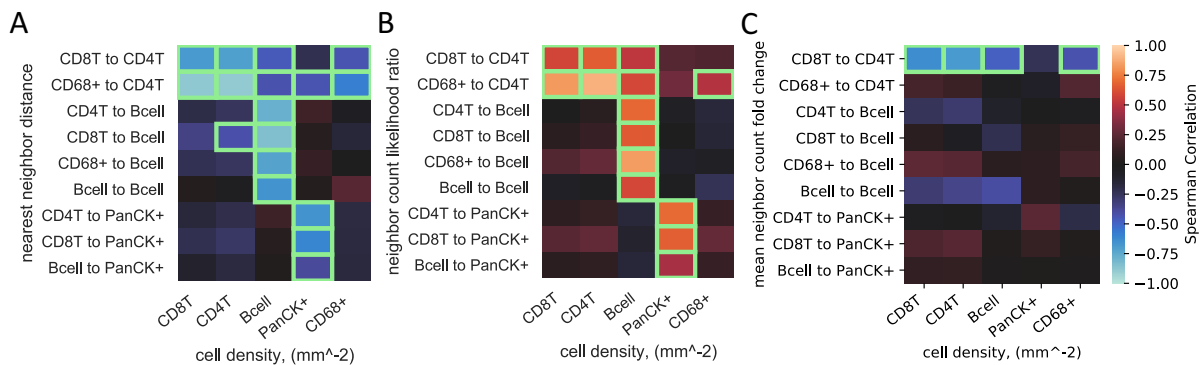

Fig. 3.5 Correlation among proximity and density features, with significant correlations highlighted in green (after Benjamini Hochberg-correction, $\alpha = 0.05$). **A)** The mean counts of neighboring cell types (rows) are correlated with the densities of those cell types (columns). CD4$^+$ T and CD8$^+$ T cell densities are interchangeable in this respect, indicating their tendency to be found together in clusters. **B)** Cell densities are positively correlated with proximity likelihood ratios, particularly among leukocytes. **C)** Fold change of mean neighbor counts is not significantly correlated with cell density of any type, except in the case of CD4$^+$ T and CD8$^+$ T cells. Reference1, reference 2, and target cell types are indicated in each row as (Ref1/Ref2→Targ).

among them, which is consistent with the formation of clusters or possibly lymphoid aggregates. This clustering behavior appears to explain the association of likelihood ratios with PanCK$^+$ cell density, and highlights the ability of the fold change metric to elucidate the spatial organization of tumor tissue.
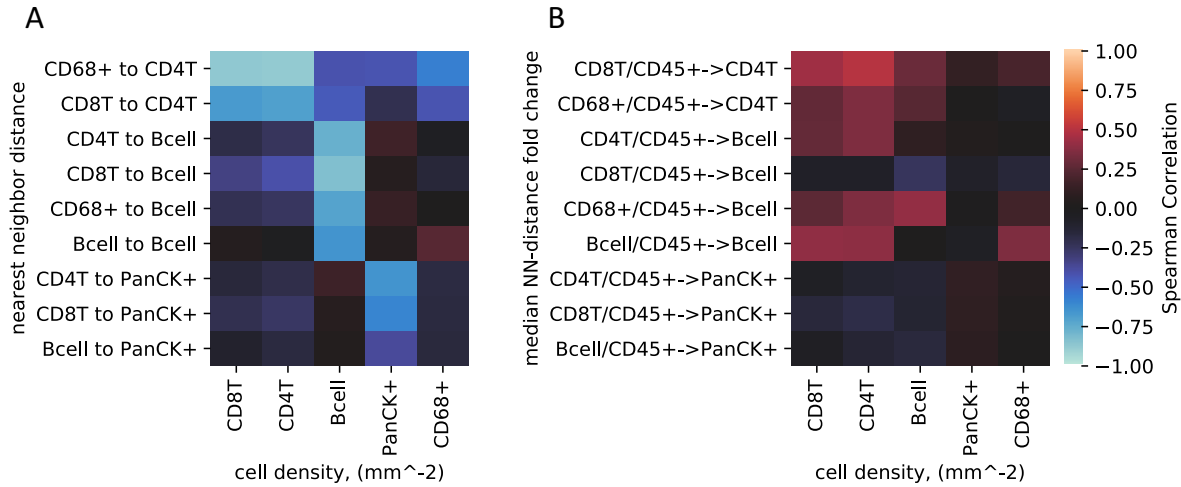
Fig. 3.6 Here we show an alternative way of computing the fold change metric - using k-nearest-neighbr distances instead of near-neighbor counts - and show that the same patterns of correlation with cell density exist that were seen with neighbor-count-based metrics. **(A)** Correlation between neighbor distances and cell densities. **(B)** correlation between neighbor count fold changes and cell densities.

3.4.2   Prognostic value of likelihood ratio and fold change proximity metrics

Next we evaluated how the likelihood ratio and fold change metrics compared in terms of prognostic value. We defined a survival classification task by grouping patients according to whether their overall survival was greater or less than the sample median. Before building models that combined the best elements of the different feature sets, we illustrate in Fig. 3.7A the impact that using fold change versus likelihood ratio versus density metrics has on prognostic performance. We previously studied the prognostic value of likelihood ratio features (using 40μm radius) [34], and found that CD8$^+$ T to PanCK$^+$ and CD4$^+$ T to CD20$^+$ B cell proximity contributed to the best model, however in that study we had focused only on patients with the most extreme survival times and did not filter out patients with insufficient leukocyte counts.

   Here, the comparison of likelihood ratio and fold change feature sets gave best-subset scores of 0.7 and 0.69, respectively, meaning that fold changes and likelihood ratios were equally successful at predicting long-term survival. Meanwhile, density features produced a best-subset score of 0.41 that is below the 0.5 level for better-than-chance performance. This contrasted with our expectations given previous studies showing prognostic value of CD8$^+$ and CD4$^+$ T cells, but could be explained by patient cohort size (smaller than ref. [58]) or the use of relatively large image regions-of-interest instead of a TMA (as used in ref. [16]). These data are consistent with the conclusion that fold changes, which represent spatial proximity not explained by cell density, are effective prognostic indicators. Interestingly, the top features for fold change and likelihood ratio categories differed despite their similar predictive value (Fig. 3.7B), but both included the proximity between CD68$^+$ and CD20$^+$ B
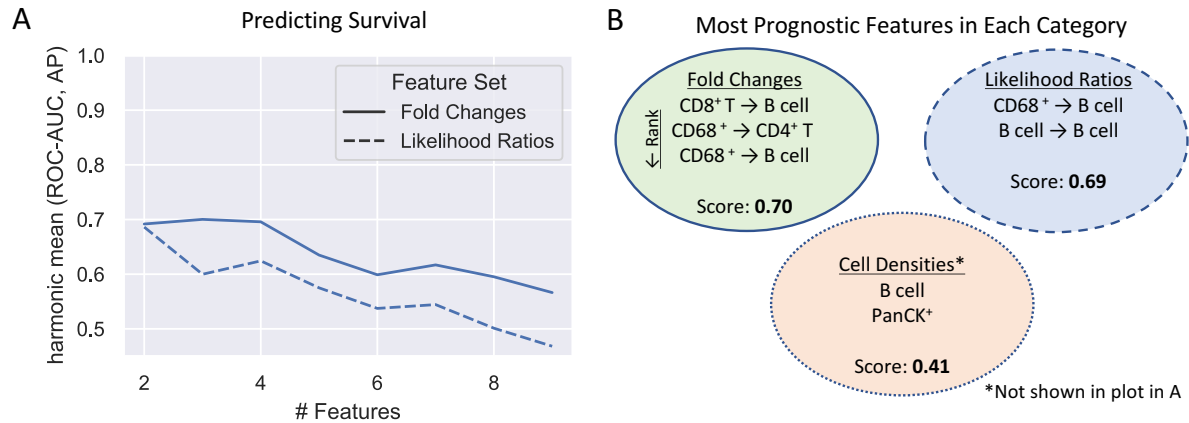
Fig. 3.7 Comparison of prognostic value of three categories of features. **A)** Recursive feature elimination with cross-validation shows that proximity fold change features have superior prognostic value to both density and likelihood ratio feature categories. The curves also indicate that the best scores in each group are produced with only 2 or 3 features. For fold changes, the reference 2 cell-type is not shown; it is always $CD45^+$. **B)** The best scoring feature subsets identified from each category. Models were trained with at least 2 features, and scores below 0.5 are not better than chance.

cells; this finding supports the preclinical work done that established the role of macrophage and B cell interactions via Bruton's tyrosine kinase as potent suppressor of $CD8^+$ T cells [38].

To test our hypothesis about the joint importance of density, proximity, and leukocyte functional features, we performed another machine learning classification experiment using a combination of these categories. In this experiment, we reduced the dimensionality of the combined data set by eliminating variables with high correlation (the variables we started with are listed in Tables 3.2-3.4). Using a correlation threshold set by the Spearman correlation test p-value and the Benjamini-Hochberg multiple testing correction, we identified features that were significantly correlated with multiple other features as those to remove. Although the prognostic feature ranking of likelihood ratios and fold changes was slightly different, given that they both included proximity between $CD20^+$ B cells and $CD68^+$ myeloid cells, and given that the proximity of B cells to other B cells was so strongly correlated with B cell density, we opted to exclude likelihood ratios from the feature set. We also excluded neighbor distance fold changes due to their similarity to neighbor count fold changes. We identified the following other correlated pairs: fold change proximity between $CD8^+$ and $CD4^+$ T cells and density of both $CD4^+$ and $CD8^+$ T cells (dropped fold change), $CD4^+$ and $CD8^+$ T cell densities (dropped $CD4^+$ densities out of preference for $CD8^+$ T cells given their prognostic value [24, 16, 58]); and finally we dropped the ratio of $CD8^+$ T to $CD68^+$ myeloid, the frequency of $CD4^+$ T, and the frequency of B cells, all due to strong correlations with densities. Since we re-used this feature set in another experiment, we called this feature set the "pruned feature set".

Following dimensionality reduction to create the pruned feature set, we compared our variables for their prognostic value by first ranking and scoring them within their categories using RFECV. Then, in

order to control for the size of the variable categories and create a final feature set that enabled direct comparison between variables, we selected the best feature subset from each category and pooled the features, followed by a final iteration of RFECV. For survival analysis, because the score produced by density features was below 0.5, we additionally removed density features from the pruned feature set.

Table 3.5 presents the ranking and cross-validation scores of models produced by this procedure. First we note that the frequency of Granzyme B positivity among $CD8^+$ T cells, and the frequency of PD-1 positivity among $EOMES^+$ $CD8^+$ T cells, were both included in the variables selected for ranking and were ranked as the top two features. Granzyme B secretion by $CD8^+$ T cells and subsequent uptake by target cells is a major mechanism of contact dependent cytotoxic lymphocyte-mediated cell lysis [57]. Meanwhile, concomitant upregulation of PD-1 and EOMES among $CD8^+$ T cells is associated with later stages of T cell exhaustion, and may indicate diminished cytotoxic potential among those cells [92], possibly enabling tumor survival. While the score (HM) for these features alone was only 0.63, when paired with the third ranked feature (fold change proximity from $CD68^+$ myeloid cells to $CD20^+$ B cells), they produced the best subset with a score of 0.71. Given the previously mentioned mechanism of $CD8^+$ T cell suppression by the interaction of B cells and macrophages through Bruton's tyrosine kinase [38], it is possible that the prognostic value of the fold-change proximity between $CD20^+$ B cells and $CD68^+$ myeloid cells reflects this mechanism of immunosuppression in human patients. Collectively, these data support the use of a multivariate prognostic model for PDAC that combines both spatial and non-spatial metrics and considers both immunosuppression and T cell exhaustion and cytotoxicity.

Rank	Variable	ROC AUC	AP	HM
1	Freq. CD8T Gzmb+	-	-	-
2	Freq. CD8T PD-1+ EOMES+	0.64	0.62	0.63
3	log2FC CD68+ →Bcell	0.74	0.68	0.71
4	log2FC CD68+ →CD4T	0.69	0.69	0.69
5	log2FC CD8T →Bcell	0.69	0.64	0.67
6	Freq. CD4T Gzmb+	0.70	0.68	0.69

Table 3.5 Performance of survival prediction models in RFECV, showing the marginal value of additional features, starting with the most important at the top. Scores of model with rank *n* correspond to use of features ranked 1 through *n*. ROC AUC: receiver operating characteristic area under curve. AP: average precision. HM: harmonic mean of ROC AUC and AP. Since we investigated multivariate models of survival, we did not train models with only one feature, and so the scores for feature rank 1 are marked "-".

3.4.3   Inferring tumor microenvironment features associated with tertiary lymphoid structures

Tertiary lymphoid structures (TLS) are ectopic cellular aggregates, primarily composed of T cells, B cells, and dendritic cells, recruited to the tumor through high endothelial venules that express the specific marker Meca79 [29]. TLS are thought to be local platforms for antigen presentation and costimulation necessary for the adaptive immune response to a tumor [29]. Evidence that TLS are independently associated with patient prognosis in PDAC has so far been mixed [41, 56], but TLS have been linked to PD-1 checkpoint inhibition response in melanoma [14], so studying their properties in PDAC could inform immunotherapy research. Currently, whether TLS are associated with changes in the spatial proximity between of other tumor infiltrating leukocytes in PDAC is not known, and this information could provide a link between spatial arrangements of infiltrating leukocytes, patient prognosis, and the presence of TLS in or near the periphery of PDAC tumors.

   We had two hypotheses about how the presence of TLS within or adjacent to PDAC tumor could affect the biology of the leukocytes (not themselves in TLS) within the tumor boundary. Given the aforementioned findings relating the spatial patterns of T cell and B cell infiltration in PDAC tumors to the prognosis of patients, we hypothesized that the presence of TLS could stimulate leukocytes to produce distinct spatial patterns of infiltration in tumors. That is, that the presence of TLS would change the localization pattern of leukocytes outside of defined TLS boundaries. Alternatively, we hypothesized that the presence of TLS would alter the function, e.g. rate of proliferation, frequency of positivity for the marker of cytotoxicity Granzyme B, or differentiation state of T cells or B cells. To test these hypotheses, we extended our machine learning approach to the task of distinguishing patient tumors with TLS from those without, using only the tumor regions not including TLS. Specifically, we used the same procedure of comparing feature categories by RFECV (Tables 3.2-3.4) and of performing RFECV on the final pruned feature set described in section 3.4.2. By training a machine learning model to predict the presence of TLS using features quantified in tumor regions outside of TLS boundaries, we aimed to deduce the influence that TLS had on other parts of the tumor microenvironment.

   In this context (Fig. 3.8), likelihood ratio, fold change, and density features produced best-subset scores of 0.83, 0.78, and 0.80 respectively, a substantial increase in performance over the models we trained to predict slong term survival. In TLS prediction, likelihood ratio features actually performed slightly better than fold change features, although both feature groups produced models with high scores compared to those in the survival prediction task. Fig. 3.8A shows how, for all three feature categories, scores near 0.8 were achieved with only the top two features. For cell density, performance dropped off steeply as more features were added. Meanwhile, for fold changes and likelihood ratios, the score curves were roughly level as more features were added. The steepness of each curve indicates the marginal value of additional features to the model; by comparing the curves, we see that, especially in the case of likelihood ratios, additional features beyond the first two were more beneficial (or less detrimental, as the case may be) to predictive performance than the corresponding cell density features.
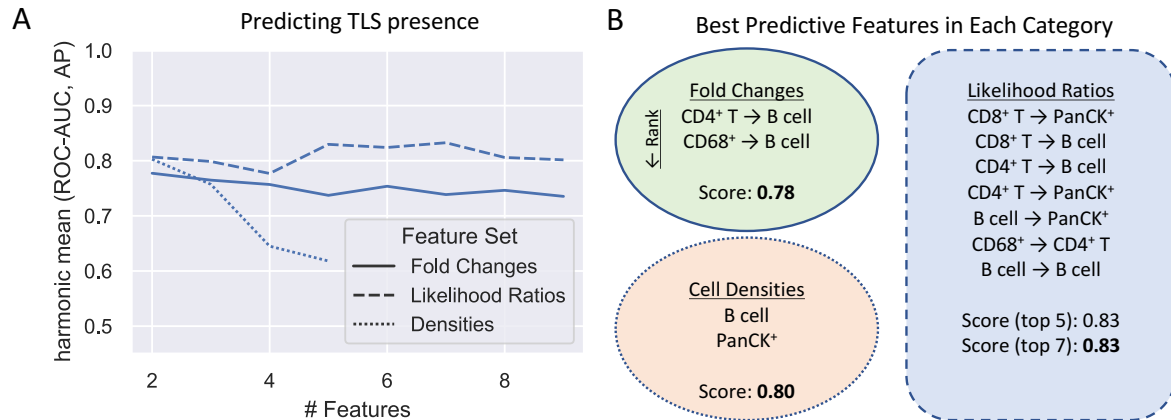
Fig. 3.8 Comparison of value in prediction of TLS presence. **A)** Proximity fold changes and likelihood ratios have similar performance curves, although the top score for likelihood ratios is better. Densities produce the best score overall, with only two features needed. **B)** The best scoring feature subsets identified from each category.

Fold change proximity between CD4$^+$ T cells and CD20$^+$ B cells was an important contributor to prediction of TLS presence; however, given the reported prevalence of T cells and B cells in TLS themselves, we speculate that clusters of these cell types in tumor regions could simply be parts of other TLS that were barely within the plane of the 5μm tissue section, and escaped quantification by the pathologist. Meanwhile, the fact that CD68$^+$ myeloid to CD20$^+$ B cell fold change proximity can predict TLS presence (Fig. 3.8B) and, to a lesser extent, overall survival (Fig. 3.7B), suggests potential influences of TLS presence on critical immunosuppressive and tumor promoting interactions between macrophages and B cells. As previously mentioned, Gunderson *et. al.* [38] showed that macrophages suppress CD8$^+$ T cells in a B cell dependent manner in a murine model of PDAC. Our findings here link this mechanistic understanding of the PDAC tumor microenvironment in that model system to the spatial proximity of CD20$^+$ B cells and CD68$^+$ myeloid cells in human tumors.

Finally, in order to directly compare the degree of association between of each of our feature categories (including those quantifying functional marker expression on T and B cells) and TLS presence, we applied the same nested RFECV procedure to the pruned feature set described in section 3.4.2. The results of this refined TLS classification in Table 3.6 show that, in addition to the proximity among CD68$^+$ myeloid, CD4$^+$ T, and CD20$^+$ B cells, the frequency of ICOS$^+$ cells among PD-1$^+$ CD4$^+$ T cells contributed to the best model, increasing the score from 0.80 to 0.87 upon inclusion. ICOS is a member of the CD28 family of receptors expressed on T cells and modulates the function of activated effector and regulatory CD4$^+$ T cells through promotion of survival, proliferation, and memory [85]. Although the downstream effect of ICOS signaling is dependent on the T cell subset in question [85], ICOS may play an important role in the stimulation of a robust antitumor response in PDAC. The contribution of the frequency of ICOS positivity among PD-1$^+$ CD4$^+$ T cells in our data

to the prediction of TLS supports a link between ICOS/ICOS-L pathway activation and the formation of TLS, which could be a potential therapeutic target.

Rank	Variable	ROC AUC	AP	HM
1	log2FC CD68+ $\rightarrow$Bcell	-	-	-
2	Freq. CD4T PD-1+ ICOS+	0.75	0.85	0.80
3	log2FC CD4T $\rightarrow$Bcell	0.86	0.88	0.87
4	PanCK+ density (mm^{-2})	0.88	0.89	0.88
5	Freq. Bcell CD27- IgD+	0.81	0.88	0.83
6	Freq. CD4T PD-1+	0.83	0.88	0.85
7	Bcell density (mm^{-2})	0.80	0.87	0.83
8	Freq. CD8T EOMES+	0.80	0.86	0.83

Table 3.6 Performance of TLS prediction models in RFECV. Format is the same as in table 3.5.

3.5   Discussion

With increasingly deep spatial molecular profiling of tumor tissues, image analysts are faced with untangling the complex network biological processes that make up a tissue. We demonstrated how the interpretation of spatial proximity between cells of a tissue can be complicated by the existence of multiple spatial processes: for instance, how apparent increases in (the likelihood ratio metric) of tumor infiltration by CD8$^+$ T cells could in fact be driven by a decrease in another spatial quantity, in this case agglomeration of T cells in the stroma. We contributed a new metric to separate distinct spatial processes using a reference cell population, and demonstrated its promise by linking it to patient survival and tertiary lymphoid structures in PDAC.

When we observed that spatial clustering of T cells was apparently intrinsically associated with their increasing numbers (Fig. 3.5), we wondered if our observation could be explained by increased molecular heterogeneity in the most dense leukocyte infiltrates. When we tested for correlation between the frequencies of positivity for leukocyte functional molecules (defined in Table 3.1) and the densities of the parent cell types, we found significant correlations (see Fig. 3.9, highlighted in green) between positivity among both CD4$^+$T and CD8$^+$T cells for PD-1 and ICOS and the density of both T cells and CD68$^+$ myeloid cells. The expression of PD-1 is generally associated with T cell exposure to antigen, so these associations may shed some light on the mechanism of T cell clustering in vivo, although expression of PD-1 on T cells has context dependent meaning [92]. We suggest that the characterization of such associations can inform future studies that aim to tell the difference between "random" spatial patterns of cells, and patterns that aren't explained by known biology.

In the task of survival analysis, we noted that the densities of CD8$^+$ T cells, CD20$^+$ B cells, PanCK$^+$ neoplastic epithelial cells, and CD68$^+$ myeloid cells did not produce a model that scored
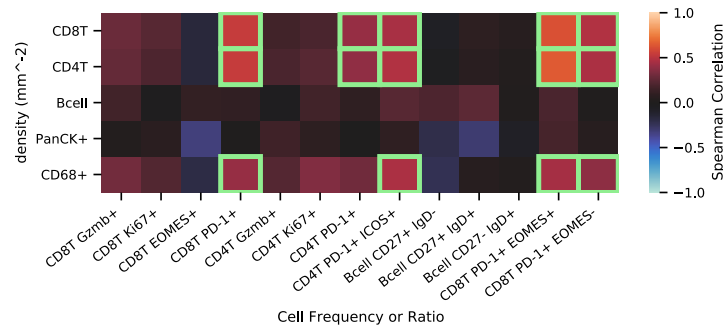
Fig. 3.9 Correlation between the frequencies of positivity of leukocyte functional markers and the density of the cell types expressing them, which could explain the relation between leukocyte density and spatial proximity. Significant correlations (after Benjamini-Hochberg correction with $\alpha = 0.05$) are marked with green boxes.

better than chance. This does run contrary to previously published studies linking the density of $CD8^+$ T cells to patient prognosis [16, 58]. There were important experimental differences between our studies and those cited above, including the use of mIHC instead of alternative antibody-based imaging methods, the amount of tissue profiled in each study, and the method for segmenting and classifying cells, which together may explain the discrepancy in our results. We also note that the papers cited in refs. [16, 58] also reported some discordance in prognostic findings. These differences highlight the importance of ongoing and future work to standardize data collection and quantification methods, and meta-analyses to clarify prognostic associations. The availability of our data and code will facilitate addressing the latter of these two challenges.

The advantages of the data set used for the study were the use of a cutting-edge imaging technology (and antibodies) that profiled leukocytes in significant detail, the relatively large tissue areas that were profiled (on average 20 mm^2 per patient), and the help of trained pathologists.

The study is limited by the small number of patients (N=46), further reduced to 37 by excluding the single stage 3 tumor and those with missing data or without at least 50 cells of each analyzed cell type, meaning that the associations we found will need validation in an independent cohort.

Lastly, a major challenge faced by our study and other antibody-based multiplexed imaging studies is the limited availability of effective antibodies for staining. It is well known now that the leukocytes profiled in this study are actually heterogeneous populations [29, 64], in particular monocytes, macrophages, B cells, and T cells. Also, functional processes we aimed to study such as T cell exhaustion have are under active research and more specific markers of cellular function are now available [5]. However, the multiplexing capacity of mIHC in particular is improving [90] along with expanding access to quality antibodies, and future studies will expand antibody panels to profile these cells in greater detail and link more up to date cell type definitions with the spatial tissue context.

Author Contributions

EG and YHC conceived of the study. EG performed all formal data analysis and wrote the paper. GT and YHC edited the paper. LC, JG, GT, and YHC supervised the study. CB, SL, and SS processed the imaging data. SL performed cell type classification. SL, DK, RS, and BS acquired patient samples. EG, YHC, GT, LC, JG, CB, SS, and SL contributed to manuscript revision, read, and approved the submitted version.

Funding

Chapter 4

Conclusion

In Chapter 1 of this thesis, I introduced Multiplexed Tissue Imaging and laid out some of the challenges that data analysts face in producing meaningful clinical or biological insights from these novel technologies. Although there are numerous challenges in the development and implementation of MTI analysis pipelines, I focused my research on two in particular. For the problem of detecting myoepithelial changes associated with DCIS progression, I created an automated pipeline to segment tumor lesions and quantify continuity of expression of differentiation markers within the myoepithelial border. Finally, to contribute new methods for spatial analysis of tissue, I evaluated proximity measures and proposed a novel metric that was not correlated to density and provided prognostic value in PDAC.

These projects represent first steps in tackling important problems in the field of biological image analysis. The DCIS project, for instance, contains an example of one tissue structure of biological and clinical relevance. Within breast cancer, other tissue structures such as vascular networks, extracellular matrix density and orientation, or necrotic regions are of clinical interest, and efforts are ongoing to build tissue databases using MTI technology [44]. The methods and code I developed to quantify the DCIS tissues, including segmenting of individual objects and calculation of continuity, could be readily adapted to other tissue structures. Ultimately, a toolbox of image processing routines and methods that could be applied to multiple tissue types and imaging modalities would be highly useful. I have provided my code in order to begin to build such a tool kit.

In Chapter 3, I focused deeply on improving spatial pattern analysis in tissues. Since existing metrics may be correlated by cell density or difficult to interpret due to randomization procedures, I developed a new metric for quantifying cell-cell proximity. I showed this metric to predict survival in PDAC, as well as to identify tertiary lymphoid structures. However, while this analysis and discussion could provide some insight into the structure of the PDAC tumor microenvironment and its relation to patient survival, the business of applying spatial proximity analysis to infer potential contact-dependent interactions in what was once a living tissue is still beset by difficulties. One essential problem is that observed proximity between cell types in a static image of fixed tissue is not necessarily the result of paracrine signaling between those cell types. Without some molecular marker of that signaling

(which of course might obviate the need for spatial proximity analysis), spatial proximity analysis can only suggest the possibility of such intercellular interaction. If researchers are to perform this kind of inference, whereby proximity is taken as a sign of ongoing intercellular communication, basic research that establishes the time and distance scales over which those processes take place *in living tumor tissues* must be conducted.

While mIHC and other MTI technologies are still in early stages of technological development, it is important to mention the limitation that a paucity of patient tissue samples imposes on statistical analyses such as those presented in this thesis. As discussed in Chapter 1 of this thesis, preanalytical variables (such as the amount of time a tissue is fixed in formalin) contribute to meaningful unwanted variation in MTI image intensity between patient tissues, and this variation can potentially cause systematic errors in analysis, warranting sophisticated normalization algorithms [19]. Given that the patient cohorts used in both of the projects presented here originated and were processed at a single institution, it is reasonable to expect that the results could be wildly different when algorithms were run on a new patient sample from another institution. Validation with a large, multi-institution cohorts of patients will be crucial for these and other algorithms developed for MTI data analysis.

When considering generalizability of automated methods to new data, such as those methods in this thesis, the complexity of the underlying model is another important factor. In Chapter 2, we demonstrated that a complex algorithm based on superpixels and Voronoi tesselation could outperform a naive algorithm; however, it is possible that we simply worked too hard to tailor our algorithm to the data we had, and that new data would show the simpler algorithm to be superior. A future study might take a closer look at the naive algorithm we presented; by varying its parameters slightly, I suspect that this simple algorithm might actually surpass the more complex one in terms of general performance in a larger cohort of patient tissue samples.

This consideration can be applied to the work in Chapter 3 as well, though in that case, it is just to reiterate the well-known phenomenon of high false-discovery rates with multiple hypothesis testing. The more variables that are given to a machine learning model, the greater the possibility (as long as the model has the capability) that the model will find a perfect fit in the training data that does not generalize to an unseen test data set [30]. This effect can even be observed when a model is cross-validated, if the cross-validation is performed many times and the best score is chosen (we did not precisely do this - RFECV produces a series of statistically dependent hypothesis tests, and their number is limited by the number of variables). In this case, the warning is simply that when patient tissue samples are limited, increasing the number of multiplexed channels or spatial proximity metrics that are simultaneously quantified and integrated into a prognostic signature has no guarantee of improving the generalizability of the resulting model; in fact, that could even be harmful [65]. Given the tissues we have now, the best course of action to remedy this issue is to rationally identify the patterns that most strongly reflect known biological processes; to prune the search space of variables to the ones that we know to be meaningful. Researchers studying genomics have already begun to identify molecular "pathways" representing cellular processes that comprise the activity of

many molecules [18], and so perhaps image analysts can begin to build a new kind of computational database to represent the relations between structure, i.e. of a tissue, and the molecular functions that it undergoes.

In sum, my work to advance image analysis methods towards a more automated, quantitative and reproducible approach has resulted in several tools that are available to the community. I have demonstrated the utility of these tools in mIHC datasets from breast and pancreatic cancer, automating previously manual tasks and creating a novel method for spatial analysis. Future endeavors will expand the use of these tools to quantify additional biology in diverse tissues and data types, and address the challenges of characterizing the interactions in human tissue from static 2D (and eventually 3D) MTI images.

References

[1] Abdallah D.M., M. E. D. N. (2017). Comparative Immunohistochemical Study of P63, SMA, CD10 and Calponin in Distinguishing In Situ from Invasive Breast Carcinoma. *Journal of Molecular Biomarkers & Diagnosis*, 8(4):342–346.

[2] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282.

[3] Adriance, M. C., Inman, J. L., Petersen, O. W., and Bissell, M. J. (2005). Myoepithelial cells: good fences make good neighbors. *Breast Cancer Research*, 7(5):1–8.

[4] Aghaeepour, N., Finak, G., Hoos, H., Mosmann, T. R., Brinkman, R., Gottardo, R., and Scheuermann, R. H. (2013). Critical assessment of automated flow cytometry data analysis techniques. *Nature methods*, 10(3):228–238.

[5] Alfei, F., Kanev, K., Hofmann, M., Wu, M., Ghoneim, H. E., Roelli, P., Utzschneider, D. T., von Hoesslin, M., Cullen, J. G., Fan, Y., Eisenberg, V., Wohlleber, D., Steiger, K., Merkler, D., Delorenzi, M., Knolle, P. A., Cohen, C. J., Thimme, R., Youngblood, B., and Zehn, D. (2019). TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature*, 9(June).

[6] Baddeley, A. and Turner, R. (2005). spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42.

[7] Bannon, D., Moen, E., Schwartz, M., Borba, E., Kudo, T., Greenwald, N., Vijayakumar, V., Chang, B., Pao, E., and Osterman, E. (2021). DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes. *Nature Methods*, 18(1):43–45.

[8] Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J. A. W. M., Hermsen, M., Manson, Q. F., and Balkenhol, M. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210.

[9] Beltman, J. B., Marée, A. F. M., and De Boer, R. J. (2007). Spatial modelling of brief and long interactions between T cells and dendritic cells. *Immunology and cell biology*, 85(4):306–314.

[10] Blackford, A. L., Canto, M. I., Klein, A. P., Hruban, R. H., and Goggins, M. (2020). Recent trends in the incidence and survival of Stage 1A Pancreatic Cancer: A Surveillance, Epidemiology, and End Results analysis. *JNCI: Journal of the National Cancer Institute*.

[11] Boquet-Pujadas, A., Olivo-Marin, J.-C., and Guillén, N. (2021). Bioimage analysis and cell motility. *Patterns*, 2(1):100170.

[12] Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., and Grobler, J. (2013). API design for machine learning software: experiences from the scikit-learn project. *arXiv preprint arXiv:1309.0238*.

[13] Bull, J. A., Macklin, P. S., Quaiser, T., Braun, F., Waters, S. L., Pugh, C. W., and Byrne, H. M. (2020). Combining multiple spatial statistics enhances the description of immune cell localisation within tumours. *Scientific reports*, 10(1):1–12.

[14] Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., Johansson, I., Phung, B., Harbst, K., Vallon-Christersson, J., van Schoiack, A., Lövgren, K., Warren, S., Jirström, K., Olsson, H., Pietras, K., Ingvar, C., Isaksson, K., Schadendorf, D., Schmidt, H., Bastholt, L., Carneiro, A., Wargo, J. A., Svane, I. M., and Jönsson, G. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature*.

[15] Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., and Moffat, J. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):1–11.

[16] Carstens, J. L., Correa de Sampaio, P., Yang, D., Barua, S., Wang, H., Rao, A., Allison, J. P., LeBleu, V. S., and Kalluri, R. (2017). Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nature Communications*, 8:15095.

[17] Castino, G. F., Cortese, N., Capretti, G., Serio, S., Di Caro, G., Mineri, R., Magrini, E., Grizzi, F., Cappello, P., Novelli, F., Spaggiari, P., Roncalli, M., Ridolfi, C., Gavazzi, F., Zerbi, A., Allavena, P., and Marchesi, F. (2016). Spatial distribution of B cells predicts prognosis in human pancreatic adenocarcinoma. *OncoImmunology*, 5(4).

[18] Cerami, E. G., Gross, B. E., Demir, E., Rodchenkov, I., Babur, Ö., Anwar, N., Schultz, N., Bader, G. D., and Sander, C. (2010). Pathway Commons, a web resource for biological pathway data. *Nucleic acids research*, 39(suppl_1):D685–D690.

[19] Chang, Y. H., Chin, K., Thibault, G., Eng, J., Burlingame, E., and Gray, J. W. (2020). RESTORE: Robust intEnSiTy nORmalization mEthod for multiplexed imaging. *Communications biology*, 3(1):1–9.

[20] Collins, L. C., Tamimi, R. M., Baer, H. J., Connolly, J. L., Colditz, G. A., and Schnitt, S. J. (2005). Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. *Cancer*, 103(9):1778–1784.

[21] DeNardo, D. G., Barreto, J. B., Andreu, P., Vasquez, L., Tawfik, D., Kolhatkar, N., and Coussens, L. M. (2009). CD4+ T Cells Regulate Pulmonary Metastasis of Mammary Carcinomas by Enhancing Protumor Properties of Macrophages. *Cancer Cell*, 16(2):91–102.

[22] Dixon, P. M. (2014). Ripley's K Function. *Wiley StatsRef: Statistics Reference Online*.

[23] Duraiyan, J., Govindarajan, R., Kaliyappan, K., and Palanisamy, M. (2012). Applications of immunohistochemistry. *Journal of pharmacy & bioallied sciences*, 4(Suppl 2):S307.

[24] Ene–Obong, A., Clear, A. J., Watt, J., Wang, J., Fatah, R., Riches, J. C., Marshall, J. F., Chin–Aleong, J., Chelala, C., and Gribben, J. G. (2013). Activated pancreatic stellate cells sequester CD8+ T cells to reduce their infiltration of the juxtatumoral compartment of pancreatic ductal adenocarcinoma. *Gastroenterology*, 145(5):1121–1132.

[25] Eng, J., Bucher, E., Gray, E., Campbell, L. G., Thibault, G., Heiser, L., Gibbs, S., Gray, J. W., Chin, K., and Chang, Y. H. (2019). cmIF: A Python Library for Scalable Multiplex Imaging Pipelines BT - Mathematical and Computational Oncology. pages 37–43, Cham. Springer International Publishing.

[26] Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4):1149–1160.

[27] Francis, K. and Palsson, B. O. (1997). Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proceedings of the National Academy of Sciences*, 94(23):12258–12262.

[28] Fridman, W. H., Pages, F., Sautes-Fridman, C., and Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. *Nature Reviews Cancer*, 12(4):298–306.

[29] Fridman, W. H., Zitvogel, L., Sautes–Fridman, C., and Kroemer, G. (2017). The immune contexture in cancer prognosis and treatment. *Nature Reviews Clinical Oncology*, 14(12):717–734.

[30] Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

[31] Gaffney, E. F., Riegman, P. H., Grizzle, W. E., and Watson, P. H. (2018). Factors that drive the increasing use of FFPE tissue in basic and translational cancer research. *Biotechnic & Histochemistry*, 93(5):373–386.

[32] Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1):3–42.

[33] Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S., and Nolan, G. P. (2018). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell*, 174(4):968–981.

[34] Gray, E., Liudahl, S., Sivagnanam, S., Betts, C., Link, J., Sheppard, B., Sears, R., Thibault, G., Gray, J. W., and Coussens, L. M. (2020). Activation vs. Organization: Prognostic Implications of T and B Cell Features of the PDAC Microenvironment. In *International Symposium on Mathematical and Computational Oncology*, pages 44–55. Springer.

[35] Gray, E., Mitchell, E., Jindal, S., Schedin, P., and Chang, Y. H. (2018). A method for quantification of calponin expression in myoepithelial cells in immunohistochemical images of ductal carcinoma in situ. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 796–799. IEEE.

[36] Gruosso, T., Gigoux, M., Manem, V. S. K., Bertos, N., Zuo, D., Perlitch, I., Saleh, S. M. I., Zhao, H., Souleimanova, M., and Johnson, R. M. (2019). Spatially distinct tumor immune microenvironments stratify triple-negative breast cancers. *The Journal of clinical investigation*, 129(4):1785–1800.

[37] Gudjonsson, T., Adriance, M. C., Sternlicht, M. D., Petersen, O. W., and Bissell, M. J. (2005). Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia. *Journal of mammary gland biology and neoplasia*, 10(3):261–272.

[38] Gunderson, A. J., Kaneda, M. M., Tsujikawa, T., Nguyen, A. V., Affara, N. I., Ruffell, B., Gorjestani, S., Liudahl, S. M., Truit, M., Olson, P., Kim, G., Hanahan, D., Tempero, M. A., Sheppard, B., Irving, B., Chang, B. Y., Varner, J. A., and Coussens, L. M. (2016). Bruton tyrosine kinase–Dependent immune cell cross-talk drives pancreas cancer. *Cancer Discovery*, 6(3):270–285.

[39] Heindl, A., Sestak, I., Naidoo, K., Cuzick, J., Dowsett, M., and Yuan, Y. (2018). Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence After Endocrine Therapy of ER+ Breast Cancer. *Journal of the National Cancer Institute*, 110(2).

[40] Helmuth, J. A., Paul, G., and Sbalzarini, I. F. (2010). Beyond co-localization: inferring spatial interactions between sub-cellular structures from microscopy images. *BMC bioinformatics*, 11(1):1–12.

[41] Hiraoka, N., Ino, Y., Yamazaki-Itoh, R., Kanai, Y., Kosuge, T., and Shimada, K. (2015). Intra-tumoral tertiary lymphoid organ is a favourable prognosticator in patients with pancreatic cancer. *British journal of cancer*, 112(11):1782.

[42] Hiraoka, N., Onozato, K., Kosuge, T., and Hirohashi, S. (2006). Prevalence of FOXP3+ regulatory T cells increases during the progression of pancreatic ductal adenocarcinoma and its premalignant lesions. *Clinical Cancer Research*, 12(18):5423–5434.

[43] Hoffer, J., Rashid, R., Muhlich, J. L., Chen, Y.-A., Russell, D. P. W., Ruokonen, J., Krueger, R., Pfister, H., Santagata, S., and Sorger, P. K. (2020). Minerva: a light-weight, narrative image browser for multiplexed tissue images. *Journal of open source software*, 5(54).

[44] Jackson, H. W., Fischer, J. R., Zanotelli, V. R. T., Ali, H. R., Weber, W. P., and Bodenmiller, B. (2019). The single-cell pathology landscape of breast cancer. *Nature*, (October 2018).

[45] Jansen, S. A. (2012). Biology of DCIS and Progression to Invasive Disease. In *Breast Cancer Biology for the Radiation Oncologist*, pages 35–48. Springer.

[46] Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S.-R., Kurian, A., Van Valen, D., and West, R. (2018). A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*, 174(6):1373–1387.

[47] Keren, L., Bosse, M., Thompson, S., Risom, T., Vijayaragavan, K., McCaffrey, E., Marquez, D., Angoshtari, R., Greenwald, N., Fienberg, H. G., Wang, J., Kambham, N., Kirkwood, D., Nolan, G. P., Montine, T. J., Galli, S., West, R., Bendall, S. C., and Angelo, M. (2019). MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. *Science Advances*, pages 1–17.

[48] Khan, A. M., Rajpoot, N., Treanor, D., and Magee, D. (2014). A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Transactions on Biomedical Engineering*, 61(6):1729–1738.

[49] Kleeff, J., Korc, M., Apte, M., La Vecchia, C., Johnson, C. D., Biankin, A. V., Neale, R. E., Tempero, M., Tuveson, D. A., Hruban, R. H., and Neoptolemos, J. P. (2016). Pancreatic cancer. *Nature Reviews Disease Primers*, 2(October 2017):1–23.

[50] Lafzi, A., Moutinho, C., Picelli, S., and Heyn, H. (2018). Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nature Protocols*, 13(12):2742–2757.

[51] Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., El-ad, D. A., Tadmor, M. D., Litvin, O., Fienberg, H. G., Jager, A., and Zunder, E. R. (2015). Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell*, 162(1):184–197.

[52] Li, Y., Huang, M., Zhang, Y., Chen, J., Xu, H., Wang, G., and Feng, W. (2020). Automated Gleason grading and Gleason pattern region segmentation based on deep learning for pathological images of prostate cancer. *IEEE Access*, 8:117714–117725.

[53] Liebling, T. M. and Pournin, L. (2012). Voronoi diagrams and Delaunay triangulations: Ubiquitous siamese twins. *Documenta Mathematics, ISMP*, pages 419–431.

[54] Lin, J.-R., Izar, B., Mei, S., Wang, S., Shah, P., and Sorger, P. (2017). A simple open-source method for highly multiplexed imaging of single cells in tissues and tumours. *eLife*, 7(e31657):1–46.

[55] Lin, J.-R., Izar, B., Wang, S., Yapp, C., Mei, S., Shah, P. M., Santagata, S., and Sorger, P. K. (2018). Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife*, 7:e31657.

[56] Liudahl, S. M., Betts, C. B., Sivagnanam, S., Morales-Oyarvide, V., da Silva, A., Yuan, C., Hwang, S., Grossblatt-Wait, A., Leis, K. R., Larson, W., Lavoie, M. B., Robinson, P., Dias Costa, A., Vayrynen, S. A., Clancy, T. E., Rubinson, D. A., Link, J., Keith, D., Horton, W., Tempero, M. A., Vonderheide, R. H., Jaffee, E. M., Sheppard, B., Goecks, J., Sears, R. C., Park, B. S., Mori, M., Nowak, J. A., Wolpin, B. M., and Coussens, L. M. (2021). Leukocyte Heterogeneity in Pancreatic Ductal Adenocarcinoma: Phenotypic and Spatial Features Associated with Clinical Outcome. *Cancer Discovery*, page candisc.0841.2020.

[57] Lord, S. J., Rajotte, R. V., Korbutt, G. S., and Bleackley, R. C. (2003). Granzyme B: a natural born killer. *Immunological reviews*, 193(1):31–38.

[58] Masugi, Y., Abe, T., Ueno, A., Fujii-Nishimura, Y., Ojima, H., Endo, Y., Fujita, Y., Kitago, M., Shinoda, M., Kitagawa, Y., and Sakamoto, M. (2019). Characterization of spatial distribution of tumor-infiltrating CD8+ T cells refines their prognostic utility for pancreatic cancer survival. *Modern Pathology*.

[59] Masugi, Y., Yamazaki, K., Hibi, T., Aiura, K., Kitagawa, Y., and Sakamoto, M. (2010). Solitary cell infiltration is a novel indicator of poor prognosis and epithelial-mesenchymal transition in pancreatic cancer. *Human Pathology*, 41(8):1061–1068.

[60] Menter, D. G., Hoque, A., Motiwala, N., Sahin, A. A., Sneige, N., Lieberman, R., and Lippman, S. M. (2001). Computerized image analysis of Ki-67 in ductal breast carcinoma in situ. *Analytical and quantitative cytology and histology*, 23(3):218–228.

[61] Miller, B. F., Bambah-Mukku, D., Dulac, C., Zhuang, X., and Fan, J. (2021). Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomics data with nonuniform cellular densities. *Genome Research*, pages gr–271288.

[62] Napari-Contributors (2019). napari: a multi-dimensional image viewer for python.

[63] Nawaz, S., Heindl, A., Koelble, K., and Yuan, Y. (2015). Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Modern Pathology*, 28(6):766–777.

[64] Neesse, A., Bauer, C. A., Öhlund, D., Lauth, M., Buchholz, M., Michl, P., Tuveson, D. A., and Gress, T. M. (2019). Stromal biology and therapy in pancreatic cancer: ready for clinical translation? *Gut*, 68(1):159–171.

[65] Ng, A. Y. (1997). Preventing" overfitting" of cross-validation data. In *ICML*, volume 97, pages 245–253. Citeseer.

[66] Ng, H. P., Ong, S. H., Foong, K. W. C., Goh, P.-S., and Nowinski, W. L. (2006). Medical image segmentation using k-means clustering and improved watershed algorithm. In *2006 IEEE southwest symposium on image analysis and interpretation*, pages 61–65. IEEE.

[67] Nogueira, F. (2014). Bayesian Optimization: Open source constrained global optimization tool for Python.

[68] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

[69] Pagès, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F.-S., Bifulco, C., Lugli, A., Zlobec, I., Rau, T. T., and Berger, M. D. (2018). International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *The Lancet*, 391(10135):2128–2139.

[70] Paik, S., Bryant, J., Park, C., Fisher, B., Tan-Chiu, E., Hyams, D., Fisher, E. R., Lippman, M. E., Wickerham, D. L., and Wolmark, N. (1998). erbB-2 and response to doxorubicin in patients with axillary lymph node-positive, hormone receptor-negative breast cancer. *JNCI: Journal of the National Cancer Institute*, 90(18):1361–1370.

[71] Pandey, P. R., Saidou, J., and Watabe, K. (2010). Role of myoepithelial cells in breast tumor progression. *Frontiers in bioscience: a journal and virtual library*, 15:226.

[72] Perou, C. M., Sørlie, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., and Akslen, L. A. (2000). Molecular portraits of human breast tumours. *nature*, 406(6797):747–752.

[73] Polyak, K. and Hu, M. (2005). Do myoepithelial cells hold the key for breast tumor progression? *Journal of mammary gland biology and neoplasia*, 10(3):231–247.

[74] Radiya-Dixit, E., Zhu, D., and Beck, A. H. (2017). Automated classification of benign and malignant proliferative breast lesions. *Scientific reports*, 7(1):1–8.

[75] Rajala, T., Olhede, S. C., and Murrell, D. J. (2019). When do we have the power to detect biological interactions in spatial point patterns? *Journal of Ecology*, 107(2):711–721.

[76] Rimm, D. L. (2006). What brown cannot do for you. *Nature biotechnology*, 24(8):914–916.

[77] Russell, T. D., Jindal, S., Agunbiade, S., Gao, D., Troxell, M., Borges, V. F., and Schedin, P. (2015). Myoepithelial cell differentiation markers in ductal carcinoma in situ progression. *The American journal of pathology*, 185(11):3076–3089.

[78] Samusik, N., Good, Z., Spitzer, M. H., Davis, K. L., and Nolan, G. P. (2016). Automated mapping of phenotype space with single-cell data. *Nature methods*, 13(6):493–496.

[79] Schapiro, D., Jackson, H. W., Raghuraman, S., Fischer, J. R., Zanotelli, V. R., Schulz, D., Giesen, C., Catena, R., Varga, Z., and Bodenmiller, B. (2017). HistoCAT: Analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature Methods*, 14(9):873–876.

[80] Schapiro, D., Sokolov, A., Yapp, C., Muhlich, J. L., Hess, J., Lin, J.-R., Chen, Y.-A., Nariya, M. K., Baker, G. J., and Ruokonen, J. (2021). MCMICRO: A scalable, modular image-processing pipeline for multiplexed tissue imaging. *bioRxiv*.

[81] Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9(7):671–675.

[82] Shedlock, D. J. and Shen, H. (2003). Requirement for CD4 T cell help in generating functional CD8 T cell memory. *Science*, 300(5617):337–339.

[83] Shiao, S. L., Ruffell, B., DeNardo, D. G., Faddegon, B. A., Park, C. C., and Coussens, L. M. (2015). TH2-polarized CD4+ T cells and macrophages limit efficacy of radiotherapy. *Cancer immunology research*, 3(5):518–525.

[84] Shivanandan, A., Radenovic, A., and Sbalzarini, I. F. (2013). MosaicIA: an ImageJ/Fiji plugin for spatial pattern and interaction analysis. *BMC bioinformatics*, 14(1):349.

[85] Simpson, T. R., Quezada, S. A., and Allison, J. P. (2010). Regulation of CD4 T cell activation and effector function by inducible costimulator (ICOS). *Current opinion in immunology*, 22(3):326–332.

[86] Stack, E. C., Wang, C., Roman, K. A., and Hoyt, C. C. (2014). Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods*, 70(1):46–58.

[87] Stoltzfus, C. R., Filipek, J., Gern, B. H., Olin, B. E., Leal, J. M., Wu, Y., Lyons-Cohen, M. R., Huang, J. Y., Paz-Stoltzfus, C. L., and Plumlee, C. R. (2020). CytoMAP: a spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues. *Cell reports*, 31(3):107523.

[88] Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106.

[89] Suvà, M. L. and Tirosh, I. (2019). Single-cell RNA sequencing in cancer: lessons learned and emerging challenges. *Molecular cell*, 75(1):7–12.

[90] Tsujikawa, T., Kumar, S., Borkar, R. N., Azimi, V., Thibault, G., Chang, Y. H., Balter, A., Kawashima, R., Choe, G., Sauer, D., El Rassi, E., Clayburgh, D. R., Kulesz-Martin, M. F., Lutz, E. R., Zheng, L., Jaffee, E. M., Leyshock, P., Margolin, A. A., Mori, M., Gray, J. W., Flint, P. W., and Coussens, L. M. (2017). Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Reports*, 19(1):203–217.

[91] Wasif, N., Ko, C. Y., Farrell, J., Wainberg, Z., Hines, O. J., Reber, H., and Tomlinson, J. S. (2010). Impact of tumor grade on prognosis in pancreatic cancer: should we include grade in AJCC staging? *Annals of surgical oncology*, 17(9):2312–2320.

[92] Wherry, E. J. and Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. *Nature Reviews Immunology*, 15(8):486–499.

[93] Yamamoto, Y., Saito, A., Tateishi, A., Shimojo, H., Kanno, H., Tsuchiya, S., Ito, K.-i., Cosatto, E., Graf, H. P., and Moraleda, R. R. (2017). Quantitative diagnosis of breast tumors by morphometric classification of microenvironmental myoepithelial cells using a machine learning approach. *Scientific reports*, 7(1):1–12.

[94] Zaha, D. C. (2014). Significance of immunohistochemistry in breast cancer. *World journal of clinical oncology*, 5(3):382.