SPARTYN Supplementary material

Nate Osher

Contents

Chapter 1

About

The purpose of this document is to serve as a "living" supplement to the first paper on the SPARTYN pipeline. While the paper itself will be static upon publication, this document will serve as both an ongoing supplement as well as a central location to give updated, detailed information on the pipeline and methods.

If you have a question that is not adequately addressed in this supplement, please email oshern (at) umich.edu.

Chapter 2

Introduction

While the staining and examination of tissue samples has been a ubiquitous medical practice for decades, it has only been relatively recently that high definition images of such stainings have begun to be explored through the lens of machine learning and statistical modeling. A considerable amount of work has thus far been put into building and training models capable of accurate classifications of overall tissue samples as well as subsections of tissue samples (Saltz et al. (2018); Amgad et al. (2019); Lu et al. (2020); Negahbani et al. (2021); Bian et al. (2021)). This work, often broadly referred to as Histopathological Image Analysis or Digital Pathology, encompasses a wide range of methods and goals ranging from developing models to assist in the scoring and staging of cancer biopsies to the classification of cells in tumor biopsies (Komura and Ishikawa (2018)). This work has important implications for biomedical research. Because of the precision with which such algorithms can assess biopsies at the cellular level, it is possible to uncover structures and associations not immediately apparent to human pathologists. This degree of precision allows for the rigorous modeling and examination of intratumoral heterogeneity with respect to arbitrary factors and patient characteristics. These methods also allow for the assessment of specific features of pathology images in a volume previously only available via the time and energy of a substantial number of trained pathologists.

While valuable in its own right such modeling has also created additional downstream opportunities for the application of more traditional modeling techniques. Specifically, the ability to quickly and accurately process high definition images of biopsies into cell level classification and location data has allowed for the development and application of more traditional spatial statistical methods to the resulting data. Despite this, there has been limited work to build off of the cell classification abilities of modern machine learning methods. To the author's knowledge, the most notable attempts to do so have been Li {*et al.*}'s work in 2019, which utilized data classified by the ConvPath pipeline (Wang

et al. (2019)).

There is ample prior reason to suspect that such applications may be valuable. Pagès et al. (2010) proposed immune reaction as the seventh hallmark of cancer, and laid out different associations between various types of immune cells and outcomes of interest. Of the different types of immune cells, the most well-studied are lymphocytes. Lymphocyte infiltration is both a meaningful prognostic indicator that can help inform treatment and predict survival across different types of cancer, including in colorectal cancer (Idos et al. (2020)), Breast cancer (Dieci et al. (2018); Denkert et al. (2015)), and melanoma (Fu et al. (2019)). This has led to increased interest in Computational TILs Assessment (CTA), a sub-field of digital pathology devoted to developing computational methods to assess lymphocyte infiltration in biopsy images. Thus far, results that have emerged from CTA research have provided additional evidence regarding the importance of tumor infiltrating lymphocytes and their spatial features in the assessment of pathology images. Saltz et al. (2018) specifically examined the presence of lymphocytes across biopsies in several types of cancer, and found that certain summarization metrics of lymphocyte clusters were significantly associated with survival in certain types of cancer. Lu et al. (2020) also found that various spatial statistical features of TILs present in a sample of breast cancer biopsies were significantly associated with survival both marginally and after adjusting for other factors. Further, they found that these associations actually differed by tumor subtype. In addition to this, they found that certain spatial features of TILs were also significantly associated with gene expression data.

While the presence of lymphocyte infiltration is consistently associated with positive prognosis and survival outcomes, the same cannot be said for all types of immune cells. It is well accepted that the presence of Tumor-Associated Macrophages (TAMs) generally have a negative impact on survival due to their association with inflammation and their inhibition of adaptive immune response (Sica et al. (2008), Tsutsui et al. (2005)). While TAMs are less studied in the domain of digital pathology, there has been recent work on the development of pipelines for the classification of TAMs. This work combined with the established effects of TAMs on cancer prognosis suggest both a possible avenue of investigation as well as a method by which to do so.

While this is only a sample of the different types of immune cells that may be present in the Tumor Microenvironment (TME) for a given patient, this broadly suggests that investigating the spatial association between general immune infiltration and the tumor itself may prove fruitful in better understanding the prognostic implications of tumor immunogenicity. Traditionally, immunogenic features of cancer have been assessed (and therfore investigated) at the biopsy level by pathologists. While this macro-assessment allows for assessment of inter-patient heterogeneity, by definition it does not allow for the investigation of \textit{intra-patient heterogeneity. Heterogeneity in the TME, including spatial heterogeneity, has proven to be an important component of understanding

cancers of various kinds (Heindl et al. (2015); Yuan (2016); Bareche et al. (2020); Boxberg et al. (2019); Yan et al. (2019); Hunter et al. (2021)). It is therefore plausible that developing methods to rigorously model the spatial variation in immune activation will yield insights into both the patient level clinical implications of such heterogeneity as well as possible genomic associations.

Spatial point processes have long been used in the domain of ecology to rigorously investigate spatial relationships between various organisms (Högmander and Särkkä (1999); King et al. (2012)). More recently, methods from this paradigm have been successfully applied within the domain of Biostatistics (Kang et al. (2011); Ray et al. (2015)). Given the relatively simple and granular nature of cell-level spatial imaging data, marked point processes are a natural way to simultaneously model the randomness in both the quantities and locations of the different cell types along with their relative spatial associations.

In this paper, we introduce the SPARTIN (SPatial Analysis of paRtitioned Tumor-Immune imagiNg) pipeline. SPARTIN is unique in the spatial pathological imaging analysis literature in that it uses statistical models to assess the association between tumor cells and immune cells across an entire partitioned biopsy. This allows for rigorous quantification of uncertainty in the style of Li et al. (2019) while still allowing for the assessment of full images in the style of Saltz et al. (2018). We accomplish this by partitioning the cell-level imaging data of each biopsy into non-overlapping sub-regions, which can then be modeled to capture the local infiltration patterns using standard techniques from point process theory. In section 2, we discuss the theory behind the models we chose to apply and the details of how the SPARTIN pipeline works. We also define Cell Type Infiltration Probability (CTIP), a measure of local immune cell infiltration. In section 3, we present results from a simulation study as well as an application to an actual data set consisting of SKCM biopsies. Finally, in section 4, we discuss the implications of this pipeline as well as potential future directions for the work.

The results of our simulations suggest that CTIP reliably identifies local positive spatial associations across a variety of different combinations of cell abundances and strengths of associations. We also found that CTIP was significantly associated with biopsy-level gene expression among genes related to immune cells. Additionally, we found that CTIP was significantly associated with various outcomes as assessed by Akbani et al. (2015), including transcriptomic class and pathologist assessment. Finally, at the patient outcome level, we found that immune cell infiltration as measured by CTIP was significantly associated with overall survival.

We focus our analysis on Skin Cutaneous Melanoma (SKCM). SKCM is an appealing target for the investigation of immune cell infiltration for several reasons. SKCM has been shown to be particularly responsive to Immunotherapy in some cases (Franklin et al. (2017); Achkar and Tarhini (2017)). It is possible that the ability to quantify immune infiltration at a large scale may allow for more detailed investigation into the scenarios in which this treatment may be most

effectively deployed. It is also well established that SKCM has an unusually high mutational load amongst the various cancer types (Berger et al. (2012)). The ability to quantify infiltration may allow for further investigation into not only genomic associations of this
occurrence but associations with mutations as well.

Chapter 3

Materials and Methods

3.1 Overview

Figure **??**b) illustrates the analysis pipeline for the processed data. For each biopsy, the smallest window that fit all cells in the biopsy is constructed by thresholding the image intensity (Figure 1, step A to step B). This window is then divided into tiles that fully partition it, while containing similar numbers of tumor cells (step B to step C). A Bayesian spatial point process model is then fit on each of these tiles separately, yielding a posterior distribution of the local interaction parameter (step C to step D). For each tile, this local distribution is compared to a tile-specific null distribution to compute the local CTIP value (step E), before being combined to summarize the overall level of interaction at the biopsy level (steps F and G).

3.2 Spatial Point Process Models

3.2.0.0.1 Data Structure Let $B_1 ... B_n$ be a set of independent regions with $B_i \in \mathbb{R}^2 \times \mathbb{R}^2$, such that each B_i has been partitioned into n_i non-overlapping sub-regions $b_{i1}..b_{in_i}$ each with well-defined boundaries. Denote the set of points observed within sub-region j of region i by $\mathbf{x}_{ij}$, and the marks of those points as $\mathbf{m}_{ij}$ where each mark $m_{ijk} \in \{K_1, ..., K_\ell\}$. Within each sub-region, this data can be naturally viewed as the realization of a marked point process and modeled as a Hierarchical Multitype Strauss Process.

3.2.0.0.2 Model The usage of a Strauss model to study spatial interaction in a marked point process is natural, since Strauss models were originally conceived for this purpose (Strauss (1975)) [While this paper introduced Strauss

models, the original model actually couldn't model positive interaction, which was pointed out a year later in a different paper (hence our usage of the hierarchical variant). Is this worth mentioning?]. As for the specific decision to use the hierarchical variant over the more standard multitype model, there were two major considerations that influenced our decision. First, in our application treating the locations of the immune cells as conditional upon the locations of the tumor cells is a priori biologically plausible. Moreover, it is quite plausible that while immune cells are responsive to the positioning of tumor cells, tumor cells may not be similarly responsive to the positioning of immune cells. The hierarchical Strauss models reflects the plausibly unidirectional nature of this spatial relationship in its modeling assumptions. Second, the hierarchical variant of the Strauss model allows for the proper modeling of positive interaction between points of different types, while the standard multitype model does not. Because we are interested in modeling not only negative interaction between different cell types but positive interaction as well, the hierarchical variant is the clear choice. It bears mentioning that this model can be extended to arbitrary numbers of marks. For our current application, we have limited ourselves to two, but it would be entirely possible to extend the chosen model to as many cell types as were available for study in the future.

The density function of the Hierarchical Multitype Strauss Process when there are $\ell = 2$ qualitative marks is defined by

$$f(\mathbf{p}_1, \mathbf{p}_2) \propto \beta_1^{n_1} \beta_2^{n_2} \gamma_{11}^{S_{R_{11}}(\mathbf{p}_1)} \gamma_{22}^{S_{R_2}(\mathbf{p}_2)} \gamma_{12}^{S_{R_{12}}(\mathbf{p}_1, \mathbf{p}_2)} \tag{1}$$

where $\mathbf{p}_t$ is a vector of points of type t, n_t is the number of points of type t, β_t is the first order intensity of points of type t, $S_{R_{tl}}(\cdot)$ counts the number of pairs of points of types t and l within R_{tl} of one another where R_{tl} is selected a priori based on subject specific knowledge, and γ_{tl} captures the tendency of points of type l to be near points of type t.

3.2.0.0.3 Interpretation As previously mentioned, γ_{tl} can be thought of as the degree to which points of type l tend to be close to or far away from points of type t. This allows us to distinguish between the mere relative abundance of different types of points (which is captured by the β_1 and β_2 parameters) and the actual spatial associations between the different types of points. This distinction is important in situations where there are substantial numbers of points of both types but no positive spatial association (and possibly a negative one). For examples, see section 3 for results from the simulation study. When $t = 1$ and $l = 2$, this association is interpreted as conditional upon the locations of the type 1 points. Under the Hierarchical Strauss Model, $\gamma_{12} \in [0, \infty)$, with $\gamma_{12} \in [0, 1)$ implying negative interaction, $\gamma_{12} \in (1, \infty)$ implying negative interaction, and $\gamma_{12} = 1$ implying no interaction at all. As one might intuitively expect, larger values of γ_{12} correspond to more positive interaction, and smaller values correspond to a more negative interaction. Due to mathematical constraints on

the density, we must have $\gamma_{11}, \gamma_{22} \in [0, 1]$. This constrains interaction between points of the same type to be modeled as negative. Since this constraint is unlikely to be satisfied in the context of a tumor biopsy, we set $\gamma_{11} = \gamma_{12} = 1$, in effect assuming no interaction between cells of the same type. Based on prior biological knowledge, we set $R_{12} = 30\ \mu m$.

In practice for various reasons these parameters are modeled on the log-scale. While this alters interpretation, because the $log(\cdot)$ function is a monotonically increasing function, the basic intuition still holds in that larger values on the log-scale are still indicative of stronger spatial interaction. An additional advantage of modeling on the log scale is the relative ease of interpretability. Whereas on the normal scale values of γ_{12} between 0 and 1 are indicative of negative interaction, values between 1 and ∞ are indicative of positive interaction, and 1 is indicative of no interaction, on the log scale negative values of $\log(\gamma_{12})$ indicate negative interaction, positive values indicate positive interaction, and a value of 0 indicates no interaction.

3.3 Pseudolikelihood Function

Except in the trivial case where $\gamma_{12} = 1$, the normalizing constant of this distribution is computationally intractable. This motivates the usage of the pseudolikelihood as outlined in Baddeley and Turner (2000). For the density $f(\cdot)$ of the simplified Hierarchical Strauss model outlined above, define the conditional intensity function at a point u given $\theta = \{\beta_1, \beta_2, \gamma_{12}\}$ and points $\mathbf{x}$ observed in window A by

$$\lambda(u|\theta, \mathbf{x}) = \begin{cases} \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})} & u \notin \mathbf{x} \\ \frac{f(\mathbf{x})}{f(\mathbf{x} - \{u\})} & u \in \mathbf{x} \end{cases} \tag{2}$$

Given this, the pseudolikelihood is defined by

$$PL(\theta|\mathbf{x}) = \prod_{x_i \in \mathbf{x}} \lambda(x_i|\theta, \mathbf{x}) \exp(-\int_A \lambda(u|\theta, \mathbf{x}) du) \tag{3}$$

This provides a computationally tractable alternative to the likelihood function that can be used in inference. Specifically, the integral in the pseudolikelihood function can be easily approximated by summing over a weighted quadrature on A. Thus, given a quadrature $\mathbf{u}$ on A with corresponding weights $\mathbf{w}$, Equation (3) can be approximated by

$$PL(\theta|\mathbf{x}) \approx \prod_{x_i \in \mathbf{x}} \lambda(x_i|\theta, \mathbf{x}) \exp(-\sum_{u_j \in \mathbf{u}} \lambda(u_j|\theta, \mathbf{x}) w_j) \tag{4}$$

The density of the selected quadrature $\mathbf{u}$ on A is chosen to balance the accuracy of the approximation against the computational load that larger quadratures impose.

Using this approximation of the pseudolikelihood function in place of the more standard likelihood fnction, Bayesian analysis can proceed in the style of King et al. (2012) by simply assigning priors to the parameters of interest and using techniques to sample from non-closed form posterior likelihoods. We assigned non-informative normal priors with mean 0 and variance 10^6 to $\log(\gamma_{12})$, $\log(\beta_1)$, and $\log(\beta_2)$. In all analysis presented below the quadrature and weights used to estimate the integral in the pseudolikelihood function was generated by the spatstat package (Baddeley and Turner (2005)). Samples from the posterior were taken using JAGS via the R2jags package (Su et al. (2015)).

3.4 Intensity Thresholding and Partition

3.4.0.0.1 Motivation for partition Each biopsy was partitioned into non-overlapping sub-regions, each of which was modeled separately as a point process. The primary motivation for structuring the analysis in this manner came from prior biological knowledge about tumor composition. Tumors are heterogeneous entities in many respects, and immune infiltration is no exception. Thus, partitioning the biopsy into non-overlapping sub-regions and fitting models on each sub-region is a natural way to capture this heterogeneity. An additional benefit of this method is that it allows for the parallelization of model fitting across the resulting sub-regions within a single biopsy. Much like selecting a quadrature when approximating an integral, calibrating the fineness of the partition entails a tradeoff between the precision with which one can assess the spatial heterogeneity within the biopsy and the computational load that a finer partition entails.

3.4.0.0.2 Partition pipeline In order to partition a given biopsy into non-overlapping sub-regions, we began with the smallest bounding rectangular window that contained cells $1...c_i$. We then applied an intensity thresholding algorithm in order to find the smallest possible window that still contained all c_i cells; see section [S] of the supplementary material for details. Next, we applied a voronoi tesselation to the T_i tumor cells within the intensity thresholded window, partitioning it into tumor cell specific sub-windows $1...T_i$ corresponding to tumor cells $1...T_i$. We then applied a modified version of k-means to the tumor cells, such that each of the K_i resulting clusters was constrained to be between a pre-defined range of cell counts. Finally, each tumor cell specific sub-window within a given cluster was combined into a tile, corresponding to each of the K_i clusters from the k-means clustering; see Figure 1 for an illustration of the process. This results in K_i tiles that fully partition the intensity thresholded window. This partition uniquely defines the membership of each of the c_i total

cells (tumor or immune) into one of the resulting tiles. Because each resulting tile has a well-defined boundary and each cell in the biopsy belongs to exactly one tile, a Hierarchical Strauss model can be fit on each tile to compute a tile specific value of each parameter in the model. Most notably, this allows for the computation of a tile-specific value of γ_{12}, which captures the local degree of tumor cell-immune cell interaction. By modeling this parameter locally to each tile, we are able to capture not only the overall level of infiltration in the biopsy, but also the potentially spatially heterogeneous nature of the infiltration.

It is worth emphasizing at this point the modular nature of our pipeline. The clustering method does not depend on the details of the model used, and the model in turn does not depend on details of the clustering method. Either can be exchanged for a different algorithm or model without disrupting the rest of the pipeline, so long as the output of the clustering algorithm is a spatial partition of the biopsy. Further, while the value ultimately selected to summarize the local infiltration at the tile level will obviously be informed by the exact model selected, there is considerable latitude in the choice of this quantity as well. We ultimately decided to summarize local infiltration using CTIP, defined in section 2.5.

3.5 Cell Type Interaction Probability

[Right now I am putting this section last, because the notation depends on the notion of the "tiles"/partition. If we re-organize this section so that the data structure explanation goes last, this could come right after the section on the pseudolikelihood function, which I think would flow best.]

3.5.0.0.1 Motivation For ease of notation denote the γ_{12} parameter corresponding to tile k of biopsy i by $\gamma_{i_{(k)}}$. While $\gamma_{i_{(k)}}$ allows us to distinguish between the relative abundance of cells of different types and their tendency to be spatially near one another, as well as the uncertainty around this tendency, this fundamentally does not capture the difference between the observed spatial association and what one would expect to observe by chance given a particular configuration of tumor cells. In order to properly assess this, the observed distribution of $\log(\gamma_{i_{(k)}})$ must be compared to a counterfactual distribution that captures the tendency when there is no interaction. This motivated the development of Cell Type Interaction Probability (CTIP).

3.5.0.0.2 Definition Let $f_{i_{(k)}}(\gamma)$ be the true posterior distribution of $\log(\gamma_{i_{(k)}})$, and $f_{i_{(k)},0}(\gamma_0)$ be the distribution of the log-interaction parameter absent interaction conditional upon the location of the tumor cells in tile k of biopsy i. Further, assume independence between the true distribution and the null distribution. Then we define the CTIP for tile k of biopsy i by:

$$\text{CTIP} \quad r_{i_{(k)}} = \int\int I(\gamma > \gamma_0) f_{i_{(k)}}(\gamma) f_{i_{(k)},0}(\gamma_0) d\gamma d\gamma_0 \tag{5}$$

3.5.0.0.3 Estimating the empirical null distribution The problem of estimating CTIP hinges on being able to estimate and sample from the null distribution of $\log(\gamma_{i_{(k)}})$, since estimation of the true posterior can proceed as described in section 2.3. In order to estimate the null distribution of $\log(\gamma_{i_{(k)}})$ conditional upon the location of the observed tumor cells, we used simulation to construct an empirical "null" distribution. Recall that when there is no interaction, by definition $\gamma_{i_{(k)}} = 1$. This means that the Hierarchical Strauss density reduces to two independent Poisson processes with intensities β_1, β_2, which are trivial to simulate.

[I could get even more rigorous with the notation and "algorithm" here, but this section is getting quite long as is- let me know if you think that would be a good idea.]

Given this, the estimation of the null distribution for a given tile proceeded as follows. First, the first order intensity of the lymphocytes was estimated using the standard estimator, $\widehat{\beta}_2 = \frac{I_{ik}}{A_{ik}}$ where I_{ik} is the number of immune cells observed in tile k of biopsy i and A_{ik} is the area in μm^2 of the tile. Next, S simulations of immune cells were generated from a poisson process with intensity $\widehat{\beta}_2$, S being selected a priori. Finally, each simulated set of immune cells was superimposed over the actual observed tumor cells, and samples were drawn from the resulting posterior distribution. These samples, combined across simulations $1...S$ served to estimate the tile specific null distribution of $\log(\gamma)$.

3.5.0.0.4 Computation of CTIP Having established the ability to sample from the true posterior distribution as well as the tile specific null distribution for $\log(\gamma)$, estimation of CTIP can be done via stochastic integration [Or should I call it "Monte Carlo integration?" I've heard it both ways]. Note that equation (5) is equivalent to $E_{\gamma,\gamma_0}[I(\gamma > \gamma_0)]$. Thus, given P posterior samples$\gamma_1, ..., \gamma_P$ from the true posterior f_γ and $\gamma_{01}, ..., \gamma_{0P}$ from the empirical null f_{γ_0}, (5) can be estimated by

$$\widehat{r_{i_{(k)}}} = \frac{1}{P} \sum_{j=1}^{P} I(\gamma_j > \gamma_{0j}) \tag{6}$$

3.5.0.0.5 Interpretation By definition, $r_{i_{(k)}} \in [0, 1]$. As previously mentioned, $r_{i_{(k)}}$ can be understood as the expected value of an indicator random variable with respect to the joint distribution of $\gamma_{i_{(k)}}$ and $\gamma_{i_{(k)},0}$, this naturally yields an interpretation of $r_{i_{(k)}}$ as a probability.

Specifically, it can be thought of as measuring the posterior probability that the observed value of $\gamma_{i_{(k)}}$ is larger than the value of $\gamma_{i_{(k)},0}$.
So $r_{i_{(k)}}$ captures immune infiltration for a particular tile, which separately capture the heterogeneity across a biopsy and can be aggregated across tiles to yield a measure of immune infiltration on a particular biopsy.

Chapter 4

Results

4.1 Simulation

4.1.0.0.1 Simulation overview To compare our model's detection of spatial association between different cell types, we ran a small-scale simulation study in which we tested the ability of this method to accurately classify positive interaction across a range of different simulated cell compositions and spatial associations. For each of four sets of simulations, the number of simulated tumor cells and immune cells were set a priori at T_s and L_s respectively. Further, interaction was controlled by a parameter $\phi \in [-1, 1]$, with -1 indicating the most negative possible interaction and 1 indicating the most positive possible interaction. Because the goal was classification as either positive interaction or non-positive (i.e. null or negative) interaction, we used AUC as our summary metric. This is a natural choice, since IP is interpreted as the posterior probability of positive interaction.

4.1.0.0.2 Negative simulations For a given combination of T_s and L_s, the positive and negative simulations proceeded differently. For the negative simulations ($\phi \in [-1, 0]$), T_s tumor cells and L_s immune cells were simulated as independent poisson processes in regions of the window that overlapped to varying degrees. The overlap was controlled by ϕ, such that the processes overlapped on $(100 \cdot (1+\phi))\%$ of the window in which the simulation occurred. Note that when $\phi = -1$ there was no overlap, and when $\phi = 0$ (complete overlap) the simulation reduced to simulating two independent Poisson Processes within the same window.

4.1.0.0.3 Positive simulations For the positive simulations ($\phi \in (0, 1]$), T_s tumor cells were simulated under a Poisson process. After their locations

were determined, L_s immune cells were simulated. For each immune cell l_i, a Bernoulli random variable $C_i \sim Bern(\phi)$ was drawn. If $C_i = 1$, l_i was simulated within 30 μm of a randomly selected tumor cell. Otherwise, l_i was simulated from a Poisson process. Thus, the level of interaction was again controlled by ϕ, with $\phi = 0$ again corresponding to two independent Poisson Processes and $\phi = 1$ corresponding to a situation in which all immune cells are within 30 μm of at least one tumor cell.

Across the different simulation settings, accurate classification was possible using IP, with the minimum AUC across simulations being 0.84. For detailed descriptions of simulation settings, simulated data, and results, see Figure 2.

4.2 Application

We applied the SPARTIN pipeline to a data set consisting of 335 high definition images of SKCM biopsies stained using hematoxylin and eosin (H&E). All images were taken from The Cancer Genome Atlas SKCM project. These images were processed using the cellular classification model of Rao and Krishnan [TODO: Figure out exact citation]. The result of this processing for each biopsy was cell type and location relative to the pathology slide. For this analysis, only data for tumor cells and immune cells were kept.

In order to compute IP, images were processed and models were fit in using the methods described in section 2. The same settings were used for all biopsies across the pipeline. For a sample of results, see Figure 3. The color of each tile indicates the value of CTIP estimated for that tile. More varied colors across a given biopsy are indicative of more spatial variation in CTIP, and thus infiltration. CTIP was summarized at the biopsy level by taking the empirical mean across all tiles for a given biopsy. Across biopsies, the median biopsy level CTIP was 0.69, and the interquartile range was 0.19. These results are consistent with the conventional wisdom that melanoma is a generally more immunogenic cancer.

4.2.1 Genomic Associations

4.2.1.0.1 Association with expression of significantly mutated genes In addition to associations with survival, we investigated the association between our measurement and gene expression. Gene expression data was acquired for all 335 patients in % A quick note about citations here: the Zhu one is the paper % I meant to cite- that paper lays out the TCGA assembler and % talks about the specific modules I used to get the data. % Re: Akbani, that's a little more complicated. See: % https://scholar.google.com/scholar?hl=en&q=genomic+classification+of+cutaneous+melanoma % This yields two different citations for (as far as I can tell % the same paper; the more utilized one is the one I

included. our sample using TCGA Assembler (Zhu et al. (2014)). Additionally, 42 significantly mutated genes (SMGs) of interest were identified using previous work investigating the genomic differences in SKCM (Akbani et al. (2015)).

Of the 335 patients in our sample, 330 had gene expression data for all genes of interest, while 95 were missing data for all genes of interest. We examined the marginal association between the normalized gene expression data for the 330 patients with complete data and average logit CTIP values. Marginal association was assessed via univariate simple linear regression, carried out separately for each gene. The Wald statistic of the coefficient corresponding to gene expression was used to produce a p-value. After correcting for multiple testing using the Benjamini -Hochberg procedure, we found that the expression of three genes were significantly associated with average logit IP: LRRC37A3, B2M, and TP53. LRRC37A3 and TP53 were positively associated with average logit CTIP ($\beta = 0.21$ and 0.16 respectively), while B2M was negatively associated ($\beta = -0.17$). See supplementary Table 2 for full details on significance of associations with each gene.

4.2.1.0.2 Association with expression of immune genes We also assessed the association between CTIP and genes that are associated with immune activity. Bhattacharya et al. (2018) have collected and classified a list of 1,793 unique genes associated with various aspects of human immune activity. Of these, gene expression data was available for 1,305 genes across the same 330 patients used in the previous gene expression analysis. We assessed the univariate association between biopsy level mean logit CTIP and these genes using Spearman Correlation. The advantage of Spearman correlation as opposed to the more standard Pearson correlation is that the former does not assume a linear relationship between the underlying variables of interest. Such assumptions can be problematic, particularly when there is no strong reason a priori to believe the relationship between the two variables is of a particular form. However, like Pearson correlation Spearman correlation is defined to lie in $[-1, 1]$, with each extreme indicating the same directionality and strength of association as Pearson Correlation. Finally, statistical significance of associations was calculated using the cor.test function of the R programming language (R Core Team (2021)). After applying a Bonferroni correction, we found that 28 genes were significantly associated with IP. For the complete list of genes, see supplementary table [N].

4.2.2 Association with Deconvolution Data

Using data from TIMER2.0 (Li et al. (2020)), we examined the association between the prevalence of different types of immune cells and biopsy level mean logit CTIP. We ultimately decided to use the MCP-counter algorithm (Becht et al. (2016)) based on the analysis of Sturm et al. (2019), since it was judged to be most effective in detecting the presence and prevalence of the most relevant

types of immune cells. We investigated the association of the score of each type of immune cell estimated by MCP-counter with biopsy level mean logit CTIP using Spearman correlation. Significance was assessed using the standard test of statistical significance of Spearman correlation as implemented by the pspearman package.

After applying a Bonferroni correction ($\alpha = 0.05$), we found that six different immune cell scores as computed by MCP-counter were significantly negatively associated with biopsy level mean logit CTIP: CD8+ T cells, B cells, Monocytes, Macrophages, Myeloid Dendritic Cells, and Natural Killer cells. No cell types were significantly positively associated with biopsy level mean logit CTIP after the Bonferroni correction, though the magnitude of the positive association with Cancer Associated Fibroblasts (CAFs) is notable, and while not statistically significant still highly consistent with a truly positive underlying association between biopsy level CTIP and prevalence of CAFs.

4.2.3 Association with Other Outcomes of Interest

4.2.3.0.1 Transcriptomic classes Akbani et al. (2015) also identified three transcriptomic classes by applying consensus hierarchical clustering techniques to gene expression data from 1,500 genes: the "immune" subclass, the "keratin" subclass, and the "MITF-low" subclass. Most notably for our current application, the immune subclass was characterized by overexpression of genes associated with T cells, B cells, and Natural Killer cells. Of the patients classified using these methods, 235 were present in our data set. Of these 235, 114 (49%) were in the immune subclass, 78 (33%) were in the keratin subclass, and 43 (18%) were in the MITF-low subclass. [Note: the proportions in the original paper are 51%, 31%, and 18% respectively, which are quite close- is this worth mentioning?] We examined mean differences in average logit-IP between each pair of classes using a standard two-sided t-test, and found that the mean logit-IP in the immune class was significantly lower than either the keratin or the MITF-low subclass ($p << 0.0001$ [$2.563e-06$ to be exact- is this how I should report it?] and $p = 0.011$, respectively). We found no significant difference between the mean logit-IP in the MITF-low subclass and the keratin subclass ($p = 0.21$).

4.2.3.0.2 Pathologist assessment In addition to classifying the biopsies into transcriptomic subclasses, Akbani et al. (2015) had pathologists assess biopsies for lymphocyte infiltration. Biopsies were scored from 0 to 3 on lymphocyte distribution, with 0 indicating no lymphocytes present in the tissue and 3 indicating that lymphocytes were present in over 50% of the tissue. They were also scored from 0 to 3 on lymphocyte density, with 0 indicating an absence of lymphocytes and 3 indicating a "severe" presence [this is the language used in the supplementary material]. These measures were added to create a Lymphocyte Score, a measure that ranged from 0 to 6 meant to summarize the general

presence and degree of lymphocyte infiltration in that biopsy. Of the 235 patients in our sample that were assessed by the pathologists, 29% had a score of zero, 23% had a score of two, 10% had a score of three, 13% had a score of four, 15% had a score of five, and 9% had a score of six. Note that by definition of the component scores, a score of one is not possible. We performed a linear regression of Lymphocyte Score on mean biopsy level logit-CTIP, treating Lymphocyte Score as continuous. We found that a one unit increase in mean biopsy level logit-CTIP was highly significantly associated with a 0.45 unit decrease in mean Lymphocyte Score ($p << 0.0001$).

4.2.4 Survival Analysis

In order to assess association between CTIP and survival, we fit a Cox Proportional Hazards model using clinical data as well as CTIP values. Clinical patient data were retrieved from the TCGA website and matched to biopsy images via TCGA identifier. Logit IP was computed at the tile level for each biopsy, and the resulting values were averaged to provide a summary of infiltration for each biopsy. In addition to adjusting for average logit IP, we adjusted for cancer stage, age in years, and sex. In both models, age in years was standardized by subtracting the mean value and dividing by the standard deviation, as was average logit IP.

Using this model, we found that after adjusting for other factors an increase in normalized logit-IP was significantly associated with increased hazard of death ($p = 0.02$). See [Figure 4] for hazard ratio estimates and associated confidence intervals. It also bears mentioning that the inclusion of normalized logit-IP modestly improved the C-statistic (Uno et al. (2011)) of the Cox model relative to the model that contained only demographic information. [Note: this is obviously a pretty negligible improvement, and MCON improved it slightly more, from 0.61 to 0.63, so I'm somewhat ambivalent about appealing to C-statistics in this section In addition to this, we found that normalized age was significantly positively associated with an increased hazard of death ($p < 0.05$), as was having stage 3 disease (relative to stage 1, $p < 0.05$). Having stage 4 disease was positively associated with increased hazard of death, though not significantly. This is most likely due to the lack of patients with stage 4 disease in the data set.

In order to assess the the performance of CTIP relative to a more traditional measure of spatial association the same model was fitted with average logit CTIP exchanged for the average normalized value of the estimated Mark Connection Function (MCON) evaluated at $r = 30$, which is the same as the radius of interaction used in our computation of IP. The Mark Connection Function is a natural point of comparison both because it is a commonly used tool for investigating spatial associations between points with discrete marks and also because it has been used elsewhere as a comparison point for similar survival modeling using spatial information (Li et al. (2019)). As with the CTIP model,

MCON was standardized by subtracting the mean and dividing by the standard deviation. While increased normalized MCON was also associated with an increased hazard of death, the association was not significant ($p = 0.057$).

Chapter 5

Discussion and Limitations

The results presented in this paper provide evidence of an immune phenotype in SKCM that corresponds to poor prognosis from a number of different perspectives. In terms of survival, we have demonstrated that adjusting for other relevant clinical factors, an increase in average logit-IP is associated with an increased hazard of death. In terms of gene expression data, average logit-IP is associated with decreased expression of genes related to immune cells that are generally associated with good overall prognosis. And finally, we found that average logit-IP is significantly lower in biopsies that have been classified as immune enriched through gene expression clustering, as well as biopsies that have been assessed by pathologists to have higher densities and spatial distributions of tumor infiltrating lymphocytes. These results fundamentally support the notion that CTIP is capturing an anti-immunogenic phenotype that is associated with poor prognosis, possibly due to an overall negative association with the presence of immune cells that are associated with improved prognosis such as tumor infiltrating lymphocytes and natural killer cells.

A fundamental limitation of this analysis is that immune cells were modeled as a single class rather than as separate cell types belonging to the same family. While the results paint a consistent picture from several different angles, the exact mechanism of these various associations remains fundamentally opaque because of this. It is worth emphasizing, however, that nothing about the structure of our pipeline depends on the presence of only one type of immune cell in addition to the cancer cells. Thus, while our present analysis raises questions about the impact of spatial immune activation on the progression of a tumor, it also offers the tools to further investigate such questions.

The SPARTIN pipeline represents a valuable contribution in and of itself to digital pathology through its ability to model and quantify immune infiltration across entire biopsies in a way that captures meaningful variation across patients. However, it also creates numerous opportunities for future work unrelated to the present investigation. As algorithms for cell-level image clas-

sification improve, the opportunities for more and more granular quantitative analysis of histopathological imaging data will become both more numerous and more fruitful. Specifically, as the ability to reliably distinguish between different types of immune cells will allow for more granular investigations into the prognoses associated with interaction between immune cells of different kinds and the tumor. Moreover, as spatial gene expression data becomes more and more common, so too will opportunities for synthesizing data on the relative spatial locations of different cell types along with local gene expression data through complex modeling. It would be quite trivial (particularly in a Bayesian framework) to model a function of one or all of the parameters of interest as a linear combination of the local gene expression data for genes that are known to be relevant to immune response. Alternatively, future methods in this area could be developed at the intersection of spatial data and high dimensional data in order to identify such genes.

The SPARTIN pipeline also creates numerous opportunities for further investigation not directly related to assessing immune infiltration. Partitioning each biopsy into non-overlapping sub-regions invites the application of other tools from the spatial statistical canon to arbitrary spatial features of the tumor microenvironment. In fact, the primary feature of interest need not even be spatial in nature; so long as a feature (such as average tumor cell size) can be quantified at the cellular level and may be expected to vary across the tumor microenvironment, SPARTIN provides a framework for mapping the variation the variation of that feature across the entire biopsy.

There is another limitation of our analysis that suggests an avenue for future methodological development. Note that the result of fitting models on the partitioned tumor microenvironment is a set of repeated observations with some underlying spatial covariance structure related to the spatial relationships between the different subspaces. For now, we have ignored this spatial covariance, and instead chosen to summarize the variation at the biopsy level. While there is nothing intrinsically incorrect about this approach, it is quite likely that incorporating the spatial structure across the biopsy will lead to gains in efficiency, and possibly insight into the nature of the spatial variation in immune activation in the tumor microenvironment. Thus, future work will likely include the development of methods to efficiently analyze this complex structured areal data.

Bibliography

Achkar, T. and Tarhini, A. A. (2017). The use of immunotherapy in the treatment of melanoma. *Journal of hematology & oncology*, 10(1):1–9.

Akbani, R., Akdemir, K. C., Aksoy, B. A., Albert, M., Ally, A., Amin, S. B., Arachchi, H., Arora, A., Auman, J. T., Ayala, B., et al. (2015). Genomic classification of cutaneous melanoma. *Cell*, 161(7):1681–1696.

Amgad, M., Sarkar, A., Srinivas, C., Redman, R., Ratra, S., Bechert, C. J., Calhoun, B. C., Mrazeck, K., Kurkure, U., Cooper, L. A., et al. (2019). Joint region and nucleus segmentation for characterization of tumor infiltrating lymphocytes in breast cancer. In *Medical Imaging 2019: Digital Pathology*, volume 10956, pages 129–136. SPIE.

Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns: (with discussion). *Australian & New Zealand Journal of Statistics*, 42(3):283–322.

Baddeley, A. and Turner, R. (2005). Spatstat: an r package for analyzing spatial point patterns. *Journal of statistical software*, 12:1–42.

Bareche, Y., Buisseret, L., Gruosso, T., Girard, E., Venet, D., Dupont, F., Desmedt, C., Larsimont, D., Park, M., Rothé, F., et al. (2020). Unraveling triple-negative breast cancer tumor microenvironment heterogeneity: towards an optimized treatment approach. *JNCI: Journal of the National Cancer Institute*, 112(7):708–719.

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W. H., and de Reyniès, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*, 17(1):218.

Berger, M. F., Hodis, E., Heffernan, T. P., Deribe, Y. L., Lawrence, M. S., Protopopov, A., Ivanova, E., Watson, I. R., Nickerson, E., Ghosh, P., et al. (2012). Melanoma genome sequencing reveals frequent prex2 mutations. *Nature*, 485(7399):502–506.

Bhattacharya, S., Dunn, P., Thomas, C. G., Smith, B., Schaefer, H., Chen, J., Hu, Z., Zalocusky, K. A., Shankar, R. D., Shen-Orr, S. S., et al. (2018). Immport, toward repurposing of open access immunological assay data for translational and clinical research. *Scientific data*, 5(1):1–9.

Bian, Y., Liu, Y. F., Jiang, H., Meng, Y., Liu, F., Cao, K., Zhang, H., Fang, X., Li, J., Yu, J., et al. (2021). Machine learning for mri radiomics: a study predicting tumor-infiltrating lymphocytes in patients with pancreatic ductal adenocarcinoma. *Abdominal Radiology*, 46(10):4800–4816.

Boxberg, M., Leising, L., Steiger, K., Jesinghaus, M., Alkhamas, A., Mielke, M., Pfarr, N., Götz, C., Wolff, K. D., Weichert, W., et al. (2019). Composition and clinical impact of the immunologic tumor microenvironment in oral squamous cell carcinoma. *The Journal of Immunology*, 202(1):278–291.

Denkert, C., Von Minckwitz, G., Brase, J. C., Sinn, B. V., Gade, S., Kronenwett, R., Pfitzner, B. M., Salat, C., Loi, S., Schmitt, W. D., et al. (2015). Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J Clin oncol*, 33(9):983–991.

Dieci, M. V., Radosevic-Robin, N., Fineberg, S., Van den Eynden, G., Ternes, N., Penault-Llorca, F., Pruneri, G., D'Alfonso, T. M., Demaria, S., Castaneda, C., et al. (2018). Update on tumor-infiltrating lymphocytes (tils) in breast cancer, including recommendations to assess tils in residual disease after neoadjuvant therapy and in carcinoma in situ: a report of the international immuno-oncology biomarker working group on breast cancer. In *Seminars in cancer biology*, volume 52, pages 16–25. Elsevier.

Franklin, C., Livingstone, E., Roesch, A., Schilling, B., and Schadendorf, D. (2017). Immunotherapy in melanoma: recent advances and future directions. *European Journal of Surgical Oncology (EJSO)*, 43(3):604–611.

Fu, Q., Chen, N., Ge, C., Li, R., Li, Z., Zeng, B., Li, C., Wang, Y., Xue, Y., Song, X., et al. (2019). Prognostic value of tumor-infiltrating lymphocytes in melanoma: a systematic review and meta-analysis. *Oncoimmunology*, 8(7):e1593806.

Heindl, A., Nawaz, S., and Yuan, Y. (2015). Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. *Laboratory investigation*, 95(4):377–384.

Högmander, H. and Särkkä, A. (1999). Multitype spatial point patterns with hierarchical interactions. *Biometrics*, 55(4):1051–1058.

Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I., and White, R. M. (2021). Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nature communications*, 12(1):1–16.

Idos, G. E., Kwok, J., Bonthala, N., Kysh, L., Gruber, S. B., and Qu, C. (2020). The prognostic implications of tumor infiltrating lymphocytes in colorectal cancer: a systematic review and meta-analysis. *Scientific reports*, 10(1):1–14.

Kang, J., Johnson, T. D., Nichols, T. E., and Wager, T. D. (2011). Meta analysis of functional neuroimaging data via bayesian spatial point processes. *Journal of the American Statistical Association*, 106(493):124–134.

King, R., Illian, J. B., King, S. E., Nightingale, G. F., and Hendrichsen, D. K. (2012). A bayesian approach to fitting gibbs processes with temporal random effects. *Journal of agricultural, biological, and environmental statistics*, 17(4):601–622.

Komura, D. and Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, 16:34–42.

Li, Q., Wang, X., Liang, F., and Xiao, G. (2019). A bayesian mark interaction model for analysis of tumor pathology images. *The annals of applied statistics*, 13(3):1708.

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., Li, B., and Liu, X. S. (2020). Timer2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Research*, 48(W1):W509–W514.

Lu, Z., Xu, S., Shao, W., Wu, Y., Zhang, J., Han, Z., Feng, Q., and Huang, K. (2020). Deep-learning–based characterization of tumor-infiltrating lymphocytes in breast cancers from histopathology images and multiomics data. *JCO clinical cancer informatics*, 4:480–490.

Negahbani, F., Sabzi, R., Pakniyat Jahromi, B., Firouzabadi, D., Movahedi, F., Kohandel Shirazi, M., Majidi, S., and Dehghanian, A. (2021). Pathonet introduced as a deep neural network backend for evaluation of ki-67 and tumor-infiltrating lymphocytes in breast cancer. *Scientific reports*, 11(1):1–13.

Pagès, F., Galon, J., Dieu-Nosjean, M.-C., Tartour, E., Sautès-Fridman, C., and Fridman, W.-H. (2010). Immune infiltration in human tumors: a prognostic factor that should not be ignored. *Oncogene*, 29(8):1093–1102.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ray, M., Kang, J., and Zhang, H. (2015). Identifying activation centers with spatial cox point processes using fmri data. *IEEE/ACM transactions on computational biology and bioinformatics*, 13(6):1130–1141.

Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K. R., Zhao, T., Batiste, R., et al. (2018). Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell reports*, 23(1):181–193.

Sica, A., Allavena, P., and Mantovani, A. (2008). Cancer related inflammation: The macrophage connection. *Cancer Letters*, 267(2):204–215.

Strauss, D. J. (1975). A model for clustering. *Biometrika*, 62(2):467–475.

Sturm, G., Finotello, F., Petitprez, F., Zhang, J. D., Baumbach, J., Fridman, W. H., List, M., and Aneichyk, T. (2019). Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*, 35(14):i436–i445.

Su, Y.-S., Yajima, M., Su, M. Y.-S., and SystemRequirements, J. (2015). Package 'r2jags'. *R package version 0.03-08, URL http://CRAN. R-project. org/package= R2jags*.

Tsutsui, S., Yasuda, K., Suzuki, K., Tahara, K., Higashi, H., and Era, S. (2005). Macrophage infiltration and its prognostic implications in breast cancer: The relationship with vegf expression and microvessel density. *Oncology Reports*.

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., and Wei, L.-J. (2011). On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*, 30(10):1105–1117.

Wang, S., Wang, T., Yang, L., Yang, D. M., Fujimoto, J., Yi, F., Luo, X., Yang, Y., Yao, B., Lin, S., et al. (2019). Convpath: A software tool for lung adenocarcinoma digital pathological image analysis aided by a convolutional neural network. *EBioMedicine*, 50:103–110.

Yan, Y., Leontovich, A. A., Gerdes, M. J., Desai, K., Dong, J., Sood, A., Santamaria-Pang, A., Mansfield, A. S., Chadwick, C., Zhang, R., et al. (2019). Understanding heterogeneous tumor microenvironment in metastatic melanoma. *PloS one*, 14(6):e0216485.

Yuan, Y. (2016). Spatial heterogeneity in the tumor microenvironment. *Cold Spring Harbor perspectives in medicine*, 6(8):a026583.

Zhu, Y., Qiu, P., and Ji, Y. (2014). Tcga-assembler: open-source software for retrieving and processing tcga data. *Nature methods*, 11(6):599–600.