Bachelor's Thesis Computing Science

Radboud University Nijmegen

Spatial Analysis of Lymphocyte Positions in Cancer Tissue

Author:
Zuzanna Pawlikowska
s1095105

Daily supervisor:
Evgenia Martynova

First assessor:
dr. Johannes Textor

Second assessor:
prof. dr. Marco Loog

June 18, 2025

Abstract

Understanding the immune response to tumors is essential for improving cancer treatment. The spatial distribution of immune cells likely reflects the characteristics of the ongoing immune response. However, a well-established methodology to quantify this spatial distribution while retaining sufficient information to capture the similarities and differences among multiple tumor samples is currently lacking. To address this gap, we applied spatial statistics to analyze the spatial distribution of B cells in 63 tumor samples from a publicly available cohort. We investigated whether distinct subgroups with similar B cell spatial organization could be discovered by quantifying their spatial distribution using the Local Correlation Function and clustering the results with an algorithm validated on synthetic data. The analysis did not reveal clearly separable clusters within the cohort; instead, a continuous landscape of spatial distribution was observed. This discovery challenges the conventional approach of categorizing tumor immune patterns into discrete subtypes. By fitting a generative model to the samples, it was found that the posterior estimates of the model parameters correspond well to the positioning of samples in the space of spatial distribution.

Contents

Chapter 1

Introduction

There are nearly 37 trillion cells in the human body, each carrying genes responsible for regulating cell growth and division. When a genetic change disrupts the normal splitting of a cell, it can multiply uncontrollably, thereby becoming cancerous. This leads to the dysfunction of the vital organs, eventually resulting in organ failure. However, your immune system has evolved to protect you. Macrophages and Natural Killer (NK) cells are the first line of defense; they attack and eliminate tumor cells. Dendritic and B cells help activate helper and cytotoxic T cells - the most efficient natural defense against cancer. Additionally, B cells produce antibodies that identify specific tumor antigens, marking them for destruction. There are mechanisms that regulate and suppress the immune response in order to protect healthy tissues from excessive immune activation. Regulatory T cells and immune checkpoints play this role; unfortunately, this means that the tumor can turn off the immune system by targeting inhibitory receptors.

The relationship between the immune system and cancer is complex, and the interactions between immune cells occurring during the immune response are largely unknown. The spatial organization of immune cells likely contains information about the ongoing immune response. This thesis focuses on an exploratory data analysis of the spatial organization of immune cells within cancer tissues. Such spatial analysis may, in the future, help understand how immune cells interact with tumor cells and predict treatment response and patient outcomes [13], but our focus here is on the spatial organization itself. To this end, we employ the methodology of spatial statistics that recognizes that the spatial arrangement of objects is scale-dependent and, therefore, provides metrics that can quantify it at a distance range. These metrics work with point patterns, data sets that contain the locations of objects or events in a region of space. In the context of the analysis of the tumor microenvironment, a tumor sample can be represented as a point pattern by noting the precise coordinates of cells of different types (like tumor

cells and lymphocytes) in the tissue (Figure 1.1). This creates a map of the cell positions in two-dimensional space.

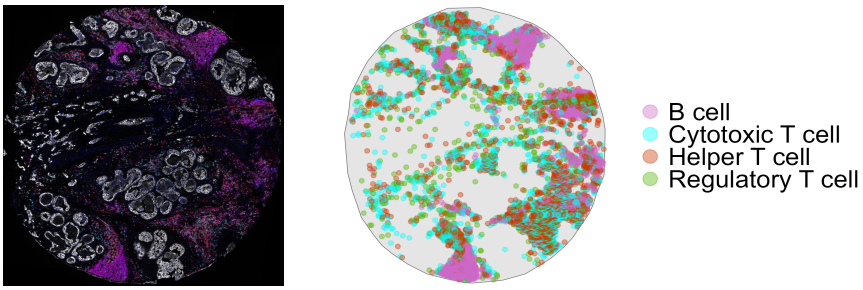


Figure 1.1: A tumor sample (left) presented as a point pattern (right).

There is growing interest in understanding how immune cells are distributed within tumors and how this influences cancer progression. Two tumors may have identical frequencies of different phenotypes of immune cells, but their distribution can vary. This variation can lead to serious clinical implications [24], which emphasizes the need for innovative spatial analysis methods.

While recent spatial profiling advances have improved our understanding of tissue structure, we still need better methods to analyze the interactions between immune and tumor cells at relevant spatial scales [12]. Building on this need, this thesis specifically focuses on the hypothesis that the spatial distribution of immune cells contains information about the ongoing immune response. Still, it is important to acknowledge that the dynamic interaction between tumor and immune cells influences these distributions. Therefore, we want to understand the broader immune-tumor relationship by analyzing specific immune cell types' arrangements.

To analyze point patterns, we selected the recently proposed Local Correlation Function (LCF) because its bounded and interpretable scale makes the results more accessible for clinicians. Our goal is to identify patients with similar spatial distributions of lymphocytes using LCF. We aim to achieve this by clustering the LCF curves using a clustering algorithm validated on synthetic data. If distinct patient subgroups are successfully identified, they can be used in survival analysis in future research [6].

To sum up, this research aims to fill the knowledge gap by answering the following question:

- Can spatial distribution patterns of lymphocytes, quantified using LCF, be used to identify patient subgroups with similar tumor spatial

structures?

To answer the main research question, we first focus on these three subquestions:

- Which clustering algorithm can recover the existing clusters of different types of point patterns based on their LCF curves?
- Does the identified clustering algorithm find different types of lymphocyte distribution in tumor samples of patients with different types of cancer?
- What are the characteristic spatial patterns of lymphocytes within identified tumor subgroups?

Chapter 2 covers the preliminary knowledge necessary to understand the thesis. Chapter 3 details the methods used, introduces the results, and examines these results along with any research limitations. Chapter 4 explores other studies investigating immune cells' spatial organization within the tumor microenvironment. Finally, Chapter 5 presents the conclusions.

Chapter 2

Preliminaries

The tumor microenvironment (TME) is a complex environment around the tumor that mainly consists of cancer cells, immune cells, blood vessels, signaling molecules, and extracellular matrix. TME plays a vital role in tumor growth and response to therapy. Before analyzing the spatial distribution of immune cells within the TME, we must introduce the basics and methods we will use in this thesis. We begin by examining different **immune cells** (Section 2.1) and presenting **spatial point patterns** (Section 2.2). Then, we continue with the **Local Correlation Function** (Section 2.2) that we use to quantify the spatial distribution of cells. **Hierarchical clustering** with **Mean Absolute Error** distance and **Ward's method** (Section 2.3) is applied to the obtained LCF curves. Then, we present ways to evaluate the results of clustering using the **Adjusted Rand Index**, **Cramér's V**, as well as the **Silhouette score** (Section 2.4), which helps us establish the optimal number of clusters. To help visualize the clustering results, we use **multidimensional scaling** (Section 2.4) to map the data into two dimensions. Finally, we implement **Approximate Bayesian Computation** (ABC) (Section 2.5). The intrinsic spatial structure within our samples can be described by using a point process along with its associated parameters. We use the ABC algorithm to estimate the parameters that represent patients' samples best. Additionally, we introduce the **Intraclass Correlation Coefficient** (Section 2.6) to estimate the agreement between measurements taken from two distinct tumor samples of the same patient.

2.1 Immune system overview

The immune system consists of two main branches: the **innate** and **adaptive** immune response. The innate immune system acts as the first line of defense. It responds quickly and non-specifically, aiming to stop the spread of infection or eliminate abnormal cells before adaptive immunity activates. The adaptive immune response is highly specific to each threat. The cells

involved in this response possess receptors that enable them to recognize threats based on their unique antigens - whether from bacteria, viruses, or cancer cells. The adaptive system can recognize nearly an infinite variety of specific antigens and mount tailored responses to each one (including tumor-associated antigens that distinguish cancer cells from healthy tissue). While the adaptive immune response is slower to kick in, it provides long-lasting immunity to the body. There are two types of adaptive responses: the **cell-mediated immune response** carried out by T cells and the **humoral immune response** controlled by activated B cells and antibodies.

The specialized cells that carry out these important functions are leukocytes, also called white blood cells. They are classified into two main groups: granulocytes and agranulocytes. The agranulocytes are further subdivided into monocytes and lymphocytes. The lymphocytes consist of B cells, T cells, and Natural Killer (NK) cells. B and T cells are part of the adaptive immune response, whereas NK cells belong to the innate immune system.

Dendritic cells, which develop from monocytes, serve as the primary envoy between the innate and adaptive immune systems. They are responsible for initiating most antigen-specific immune responses. Two distinct subsets of dendritic cells should be noted: **conventional** (cDCs) and **plasmacytoid** (pDCs). cDCs activate T cells by presenting antigens to them. pDCs have distinct functions - they produce type I interferons when fighting viral infections, and they also play important roles in immune tolerance and B cell activation.

B cells possess surface receptors that enable them to specifically attach to antigens with a distinct shape. In contrast to T cells, B cells can directly bind to antigens without needing them to be presented on an MHC (Major Histocompatibility Complex) molecule. After binding to an antigen, a B cell loads it onto an MHC II molecule and presents it to T cells. When a T cell becomes activated, it assists the B cell in maturing into a plasma cell. Plasma cells are capable of secreting large quantities of antibodies, which are Y-shaped proteins. The antibodies produced have the same specificity as the original B cells. These antibodies can identify specific antigens associated with tumors. They bind to cancer cells and mark them for destruction. Since antibodies are not attached to cells and circulate freely in the bloodstream, this process is known as humoral immunity.

T cells are antigen-specific, but they cannot secrete their antigen receptors. A naive T cell can be activated by any antigen-presenting cell, typically dendritic cells, enabling its transformation into a mature T cell. The two primary types of T cells are **helper T cells** and **cytotoxic T cells**. Helper T cells produce cytokines, which are signaling proteins that participate in-

directly in the destruction of infected cells. These proteins activate other immune cells, such as B cells and cytotoxic T cells. Helper T cells can only recognize antigens presented on MHC II molecules. In contrast, cytotoxic T cells are tasked with eliminating target cells that exhibit a particular antigen on MHC I molecules. We should also mention **regulatory T cells**, which help maintain immune balance. They suppress the immune response, so they act as a "self-check".

Natural killer (NK) cells work closely with T cells to eliminate infected and diseased cells, including cancer cells. More specifically, they kill cells that do not present MHC class I on their surface. NK cells kill these target cells by releasing cytotoxic granules, which release some molecules that get inside the cell, triggering apoptosis (programmed cell death).

2.2 Spatial statistics

The analysis of the spatial distribution of objects is the subject of a mature field of spatial statistics. A **spatial point pattern** is a data set used in spatial statistics that records the observed spatial locations of things or events within the bounded region in space, known as the **observation window**. As an example, we can give the positions of stars within a star cluster, the epicenters of earthquakes, trees in the forest (Figure 2.1), or, as in this research, the locations of immune cells within tumor tissue. Analysis of these point patterns can show underlying relationships and behaviors, such as the tendency of specific immune cell types to cluster or their spatial relation with tumor cells [1].

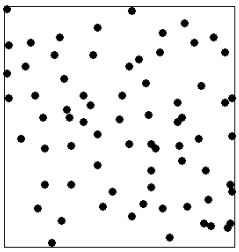


Figure 2.1: Swedish Pines dataset. It shows the locations of pine saplings (71 points) in a survey plot in a Swedish forest. Trees seem to be evenly dispersed throughout the plot, which could be caused by the competition for light, water, and nutrients.

A **point process** characterizes the underlying spatial distribution from which different point patterns can be drawn. Thus, a point pattern is a single realization of a point process - a relation similar to the one between an observed value of a random variable and its probability distribution. When you conduct a statistical experiment, you obtain a specific value of a random variable from its underlying probability distribution (with its specific parameters); however, if you repeat the experiment, you can get another value. This same principle applies in spatial statistics. We should also mention a key concept in spatial statistics - **Complete Spatial Randomness** (CSR), a point process where points are distributed within a given study region in a completely random and independent fashion (there is no interaction between points).

The spatial distribution of objects can generally be classified into three categories: random, dispersed, and clustered. However, it is important to note that spatial distribution is scale-dependent. For example, a pattern that appears clustered at a small scale may show dispersion at a larger scale. Since spatial distribution is scale-dependent, spatial statistics quantifies point patterns using summary functions, which take the distance between objects as an argument.

The most basic question about the spatial distribution of objects is whether they are clustered (dispersed) and to what extent. Most statistics used to quantify point patterns are based on **Ripley's K-function** [15], which counts the expected number of neighbors of a point at different distances and compares this to the expectation for randomly arranged points.

The K function is defined as:

$$K(r) = \frac{1}{\lambda}\mathbb{E}[N(r)] \tag{2.1}$$

where r is a distance of interest, λ is the intensity (the expected number of points per unit area), and E[N(r)] is the expected number of extra events within distance r of a randomly chosen event. Normalisation by intensity is necessary to avoid confusing high point density and clustering.

However, Ripley's K is unbounded, and its expected value depends on the area of the observation window. This makes it difficult to interpret the value of K as an effect size; in other words, it cannot answer the question "To what extent are objects clustered (dispersed)?". Other functions of spatial statistics share this limitation.

That is why a new method for spatial analysis was proposed: the **Local Correlation Function** (LCF). It is also based on Ripley's K but is de-

signed to be interpretable as a degree of clustering. To this end, LCF is limited to the [-1,1] range, with a value of 1 indicating maximal clustering, a -1 indicating maximal dispersion, and a 0 representing complete spatial randomness. Having these properties, LCF behaves similarly to a linear correlation coefficient, which makes interpreting LCF values and comparing patterns across different samples easier [10].

LCF formula:

$$LCF(r) = \begin{cases} 2\exp\left(-\frac{\ln 2}{2}\frac{rN'(r)}{N(r)}\right) - 1, & N(r) > 0 \\ -1, & N(r) = 0 \end{cases}$$

where N(r) denotes the expected number of neighbors of a random point within radius r, and N'(r) is its derivative. N'(r) originates from comparing the expected number of neighbors of a point at two close distances, r and hr, and taking the limit as h approaches 1.

To illustrate the behavior of the LCF and its capacity to differentiate among the fundamental types of spatial distribution, we created three types of point patterns: random, clustered, and dispersed. These patterns are shown in Figure 2.2, alongside their corresponding LCF curves. For the random point pattern, the LCF values stay around zero across the whole distance range (excluding very small values), which means no significant correlation, indeed expected for CSR. In contrast, the LCF for a clustered point pattern that contains approximately 20 clusters of 0.1 diameter on average is significantly higher than for CSR, demonstrating a positive spatial correlation between the positions of points. Lastly, for the dispersed point pattern where points are constrained to maintain a minimum distance of 0.05 from each other, the LCF shows values of -1 for distances r $\leq$ 0.05, which means a maximal dispersion within this range.

2.3 Hierarchical clustering

Clustering is the process of grouping similar objects into clusters. The similarity is determined by how close in space these objects are, which is based on a distance function. In contrast, dissimilar objects should be assigned different cluster labels.

Hierarchical clustering is a clustering method that groups objects into nested clusters, creating a hierarchical tree-like structure called a dendrogram. There are two types of hierarchical clustering: **agglomerative** (a bottom-up approach) and **divisive** (a top-down approach). The agglomerative approach starts with placing each object into a separate cluster and

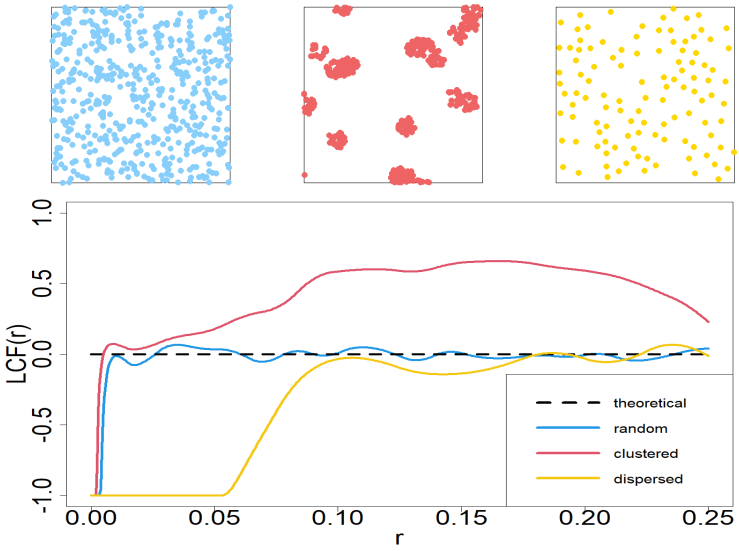


Figure 2.2: LCF curves analysis for characteristic point patterns. The LCF effectively differentiates between the three basic types of spatial point patterns. The top row shows random (left), clustered (middle), and dispersed (right) distributions.

subsequently merges the two most similar clusters at every step. This process continues until all groups are combined into a single cluster. In contrast, in the divisive approach, all objects start within a single group, which is then subdivided into smaller clusters. This splitting continues until each object is isolated in its cluster [3]. Regardless of the method selected (agglomerative or divisive), the fundamental requirement is to measure the distance between two clusters.

A significant drawback of hierarchical clustering is that it requires a **split point** (the distance used to cut the tree or the number of clusters); wrongly chosen values may lead to low-quality clusters.

2.3.1 Mean Absolute Error

We chose **Mean Absolute Error** (MAE) as a distance metric to measure the differences between two objects, represented as numerical vectors. It looks at the average size of the errors, giving a clear idea of the overall accuracy [4].

For two vectors $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |a_i - b_i| \tag{2.2}$$

2.3.2 Ward's method

Ward's method is a specific technique that decides which groups to merge at each step. When Ward's approach considers combining two groups, it looks at how much the variance within the new group would increase and selects the merger with the smallest increase. This method effectively tries to create clusters that are internally similar [11].

The error sum of squares is the sum of squared distances (d) from each observation i in cluster q to the cluster centroid c_q. Specifically:

$$\text{The error sum of squares is calculated as: } \sum_{i \in q} d^2(i, c_q). \tag{2.3}$$

Then the **variance** becomes:

$$\text{Variance: } 1/|q| \sum_{i \in q} d^2(i, c_q). \tag{2.4}$$

2.4 Clustering evaluation metrics

To evaluate the quality of clustering, we use **Adjusted Rand Index** and **Cramér's V**. These are supervised evaluation metrics, which means that they require knowledge of the true cluster labels to compare against the clustering results. Because our simulated point patterns have established ground truth labels, we can assess how effectively our clustering method retrieves the true cluster structure. This validation step provides us with greater confidence in our approach before applying it to the real dataset, where no ground truth labels exist.

2.4.1 Adjusted Rand Index

The **Adjusted Rand Index** (ARI) measures the similarity between two groupings. It is based on the Rand Index, which examines how pairs of objects are classified across two different partitions, U and V.

There are four possible scenarios for each pair of objects:

1. The objects are in the same class in both U and V.
2. The objects are in different classes in both U and V.
3. The objects are in different classes in U but in the same class in V.
4. The objects are in the same class in U but in different classes in V.

When a pair of items is categorized in either the same or different groups in both partitionings, we call this an agreement. To measure the overall agreement of two measures, the Rand Index counts all these agreements and divides them by the total number of possible pairs of items [5].

The Adjusted Rand Index addresses a limitation of the Rand Index - some matches might happen by chance. ARI takes into account the amount of agreement you would expect to see simply by chance alone. Most importantly, ARI works on a scale from 0 to 1. A value of 1 means a partition perfectly matches the intrinsic structure, and a value of 0 indicates a random partition.

2.4.2 Cramér's V

Cramér's V measures the association between two categorical variables, which, in the context of clustering, represent the cluster assignments for each object. It gives a value ranging from 0 (no association) to 1 (complete association).

The formula for calculating Cramér's V is as follows:

$$V = \sqrt{\frac{\chi^2}{n \cdot (k-1)}} \tag{2.5}$$

In this equation, χ^2 is a chi-squared test statistic, n is the total number of observations, and k is the smaller of the number of levels of the two variables (i.e., min(r,c), where r and c are the numbers of rows and columns in the contingency table) [14].

2.4.3 Silhouette score

When we apply the selected clustering algorithm to the real data, we do not know whether clusters exist and how many. Therefore, to aid our choice of the number of clusters, we need a measure of cluster quality that does not use labels.

The **silhouette score** is a metric used to measure the cluster quality, which takes into consideration two key distances: **the mean intra-cluster distance** (a) and **the mean nearest-cluster distance** (b). The former is the average distance of the data point to other points within the same cluster, and the latter is the average distance to all points in the nearest other cluster. It is calculated using the following formula:

$$\text{Silhouette score: } \frac{(b(i) - a(i))}{max(a(i), b(i))} \tag{2.6}$$

This formula calculates the silhouette score for just one point (i). Afterward, we use the mean of all Silhouette scores to evaluate clustering. The silhouette coefficient can range between -1 (which indicates that the point is matched with other points in its cluster poorly) and 1 (the data point is well matched). However, what is most important in this thesis is that it can serve as a suggestion for the number of clusters to choose. The highest silhouette score often corresponds to the optimal number of clusters for a given dataset [17].

2.4.4 Multidimensional scaling

To explore the structure of the space of LCF curves that describe the spatial distribution of immune cells, we use **multidimensional scaling** (MDS). MDS represents measurements of similarity or dissimilarity between pairs of objects as distances between points in a lower-dimensional space [2].

In this thesis, we use MDS to visualize the results of hierarchical clustering. This enables us to assess how effectively the hierarchical clustering solution captures the underlying structure of the data.

2.5 Approximate Bayesian Computation

To fit a point process to our data and find the parameters that describe the spatial arrangement in our samples best, we use **Approximate Bayesian Computation** (ABC). This method comes from **Bayesian statistics**.

ABC methods begin by sampling parameters from a prior distribution. A dataset is simulated for each sampled parameter, and summary statistics are calculated in the same way as those obtained from the observed data. Next, the simulated and observed summary statistics are compared using a chosen distance measure. If this distance is less than or equal to the predefined tolerance level (ϵ), then the sampled parameter value is accepted; otherwise, it is rejected.

After many iterations of this process, we obtain a set of accepted parameters that approximates a sample from the posterior distribution. An overview of this process is visible in Figure 2.3. It is vital to acknowledge that it is unlikely to find a perfect match between simulated and observed data, meaning the tolerance should be greater than zero [18].

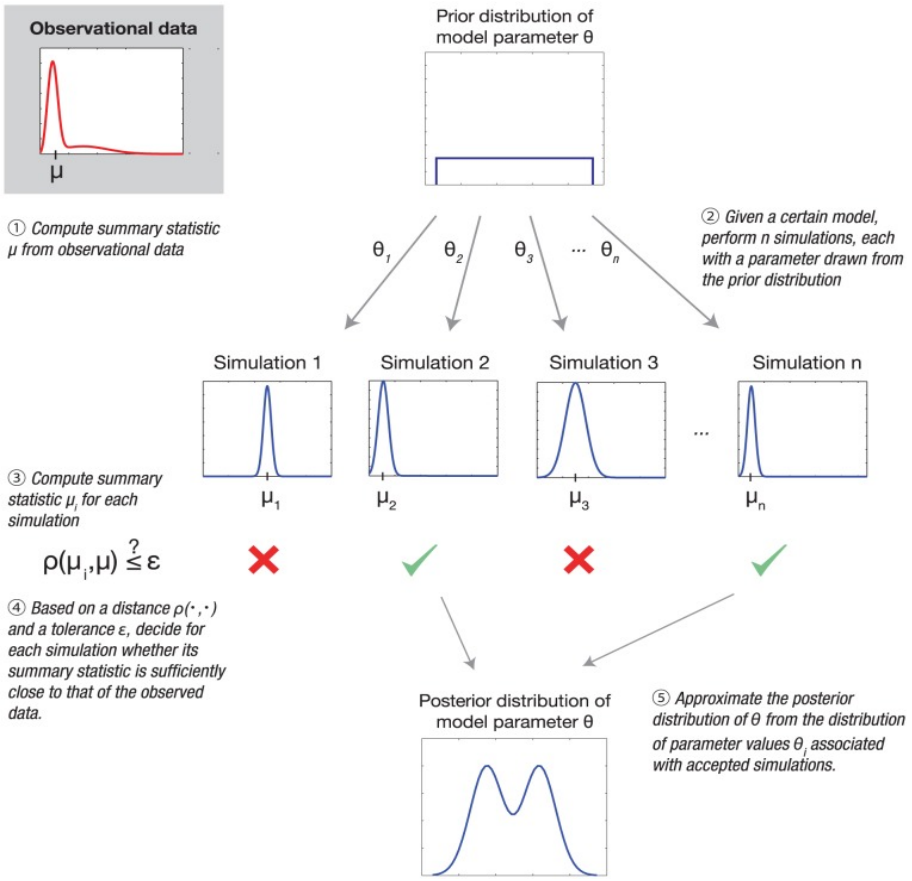


Figure 2.3: An overview of a parameter estimation by Approximate Bayesian Computation. Source: Sunnåker et al. (2013) [18]

2.6 Intraclass Correlation Coefficient

The intraclass correlation coefficient (ICC) is a key measure of reliability. We can define reliability as the consistency of measurements. Unlike typical correlations, which investigate correlations between two different classes of data, the ICC refers to correlations within a class of data. The ICC can help determine if two tissue samples from the same patient show similar spatial characteristics when compared to samples from different patients. Understanding the reliability of measurements is vital for determining possible applications of these measurements and the conclusions that can be drawn from them.

The ICC is calculated by examining the variance within groups and the variance between groups. We obtain a high ICC when measurements within the same group show low variance and measurements between different groups demonstrate high variance. A high ICC value indicates that the groups being studied are internally consistent.

However, calculating ICC is not straightforward because there are 10 different forms of ICC based on different models (one-way random effects, two-way random effects, or two-way mixed effects), types (single measurement or mean of multiple measurements), and definitions (consistency, where systematic differences between raters are acceptable, or absolute agreement, where raters must assign exactly the same scores). The choice of the appropriate version depends on the experimental situation.

When interpreting ICC results, it is important to recognize that no universally accepted standards exist for acceptable reliability. A low ICC might indicate not only a low degree of rater or measurement agreement but also, for example, the small number of subjects or the small number of raters being tested. Ideally, using at least 30 heterogeneous samples and engaging a minimum of three raters is recommended for robust results, but this may not always be possible.

The interpretation of ICC values can be summarized as follows: values below 0.5 suggest poor reliability, values ranging from 0.5 to 0.75 reflect moderate reliability, values between 0.75 and 0.9 indicate good reliability, and those exceeding 0.9 are considered indicative of excellent reliability [8].

Chapter 3

Research

The goal of our research is to determine whether we can identify patient subgroups with similar tumor spatial structures using LCF curves. If such subgroups are found, survival analysis can be conducted to assess whether the identified types of spatial distributions of lymphocytes influence patient prognosis. This chapter outlines the steps taken in this research, the used methods, the obtained results, and the limitations of this study.

3.1 Dataset

In our analysis, we utilized tissue microarray (TMA) data from van der Hoorn et al. [21], including samples from 15 different tumor types. For each tumor type, researchers selected tissue samples from two or three patients for the TMA. For many patients, two tissue cores were taken. All tissue samples were anonymized, and all patients gave permission to use their tissues for (clinical) research.

Each sample includes the tissue core's boundary, coordinates, and the types of immune cells detected. Additionally, each patient had a precomputed LCF. The samples were stained with the dendritic cell panel, which includes the markers BDCA2, XCR1, BDCA1, CD19, and CD14. These markers define the following cell populations: pDC (BDCA2+), cDC1 (XCR1+), cDC2 (BDCA1+CD19-), monocyte/macrophage (CD14+), B cell (CD19+BDCA1+/-), and double-positive cDC2 (BDCA1+CD19-CD14+) [21].

This thesis focuses specifically on B cells. We examine the subset characterized as CD19+BDCA1+/-, where CD19 is a broad marker of the B cell lineage, which also includes plasma cells that have distinct functions.

3.2 Methods

3.2.1 Exploring spatial patterns

In this study, our first step is to validate a selected clustering algorithm by checking whether it can recover existing clusters of different types of spatial distribution using the LCF curves of the corresponding point patterns. The LCF curves capture the spatial structure of point patterns, making them suitable inputs for clustering algorithms. Different types of spatial distributions can be obtained by selecting a few point processes; then, all realizations of a point process form a separate cluster. The following types of point processes embody random, dispersed, and clustered spatial distributions:

Poisson point process

The Poisson point process represents a completely random spatial distribution. It is uniquely defined by its intensity, λ, an expected number of points per unit area.

Hardcore point process

The Hardcore process represents a dispersed spatial distribution where no two points can be closer than a specified distance R. It is defined by two key parameters: the intensity β, and the hardcore distance R. This process better represents random object distribution in the real world, as objects have shapes.

Matérn point process

The Matérn cluster process represents clustered spatial distributions. It is defined by the cluster intensity, κ, which represents the expected number of clusters per unit area; the mean number of offspring, μ, which is the average number of points generated in each cluster; and the cluster radius, r, which indicates the maximum distance of offspring from the cluster center.

Thomas point process

The Thomas cluster process represents clustered spatial distributions with diffuse clustering. It is uniquely characterized by cluster intensity (κ), mean number of offspring per cluster (μ), and scale parameter that describes the spread of the clusters. Points are drawn from a 2D Gaussian distribution centered at each parent point, with a standard deviation equal to the scale parameter. Thomas's process is great for illustrating how immune cells behave because it captures how, from a spatial perspective, offspring cells generally cluster around parent cells, yet they tend to retain a certain degree of random movement.

The spatstat package in R allows the generation of realizations of these point processes using corresponding functions: rpoispp() for Poisson processes, rHardcore() for Hardcore processes, rMatClust() for Matérn cluster processes, and rThomas() for Thomas cluster processes (see Figure 3.1). Each function takes the respective parameters discussed above, along with a window within which to simulate the pattern.

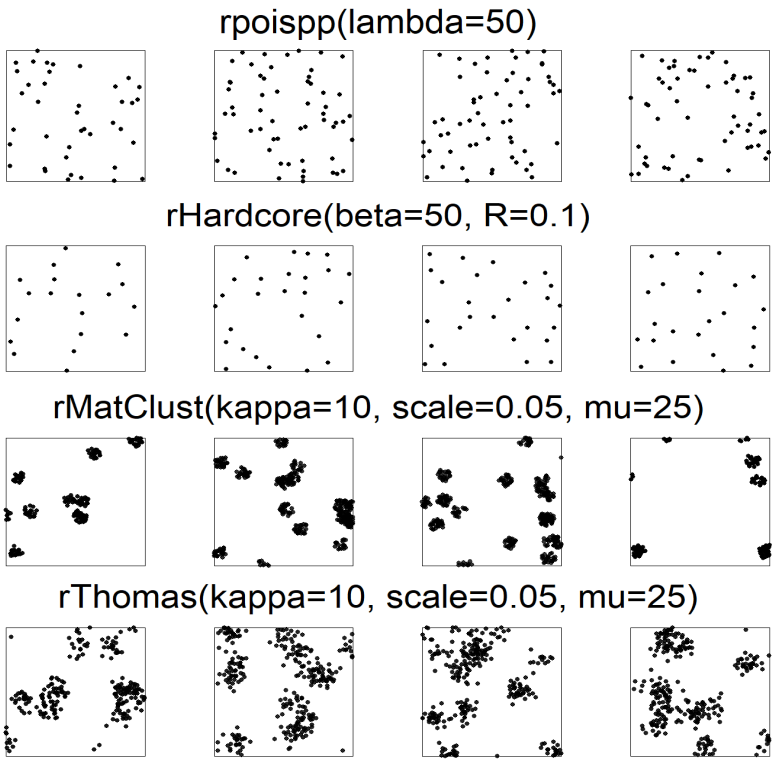


Figure 3.1: Realizations of spatial point processes.

3.2.2 Clustering algorithm for identifying point pattern groups using LCF curves

We chose hierarchical clustering as our primary clustering algorithm. This algorithm does not require specifying the number of clusters in advance, which is particularly advantageous when that number is unknown. It is easy to understand and provides very good visual clarity through dendrograms. We use it with the mean absolute error distance, which best captures the similarity of function curves, along with Ward's method to merge groups, which minimizes within-cluster variance.

Preprocessing of LCF Curves

Before applying hierarchical clustering, we preprocessed the LCF curves by filtering out values where r $\leq$ 50. Previous statistical analysis of the LCF has shown that the LCF's variance is high at small distances [10]. These small-distance values can disproportionately influence the clustering results due to their high variability, which stems from higher uncertainty in the LCF estimates rather than meaningful spatial differences.

3.2.3 Bayesian posterior estimation for tumor sample parameters

To investigate the intrinsic spatial characteristics of tumor samples and relate them to the clustering results, we decided to fit a Thomas cluster process to each sample. To obtain the posterior distribution of Thomas's process parameters, we employ a method called approximate Bayesian computation (ABC). By comparing simulated data to tissue sample data using summary statistics (LCF) and a distance metric (MAE), ABC enables us to efficiently approximate the posterior distributions. Our ABC function estimates two parameters (κ and scale) of a Thomas process; the μ parameter is calculated as $\frac{intensity}{\kappa}$, where intensity is derived from the observed data as the total number of B cells divided by the tissue sample area, representing the overall point density.

Algorithm Parameters

The algorithm takes the following parameters:

- summary statistics of a tissue sample: its LCF values.
- intensity of the point process.
- spatial boundary of the patient tissue for the simulation.
- population size, which determines the number of accepted parameter sets per iteration.
- bounds for the κ parameter.
- bounds for the scale parameter.
- target rejection rate for parameter acceptance.
- mutation threshold for perturbing previously accepted parameters.

Tolerance thresholds

We use three different tolerance thresholds: 0.3, 0.2, and 0.1, which we gradually decrease, allowing for more precise adjustments.

Workflow

Our ABC algorithm works by iteratively refining parameter estimates (κ and scale) through a comparison of observed data to simulated data. In the first iteration, parameters are sampled from the prior distributions (uniform distributions over the specified parameter ranges). In the subsequent iterations, new candidate parameters are obtained by perturbing the previously accepted parameters. The simulated point patterns are generated using the rThomas function, and their LCFs are calculated. They are then compared to the observed LCF using the MAE distance. We accept the parameters if this difference is less than our current tolerance level.

We calculate the rejection rate as the ratio of rejected simulations to total attempts and adjust the tolerance based on how this rejection rate compares to a target rejection rate. If the rejection rate is below the target, we reduce the tolerance to the next threshold. If the rejection rate is above the target, we increment a "stuck" counter, which terminates the process if it reaches a value of 2 (we give our algorithm a second chance to improve).

Finally, the algorithm returns the accepted parameters, their mean values, 95% credible intervals (range where we are 95% confident the true value lies, based on our data), and the rejection rates across iterations.

The five main steps of our algorithm are shown in Figure 3.2.

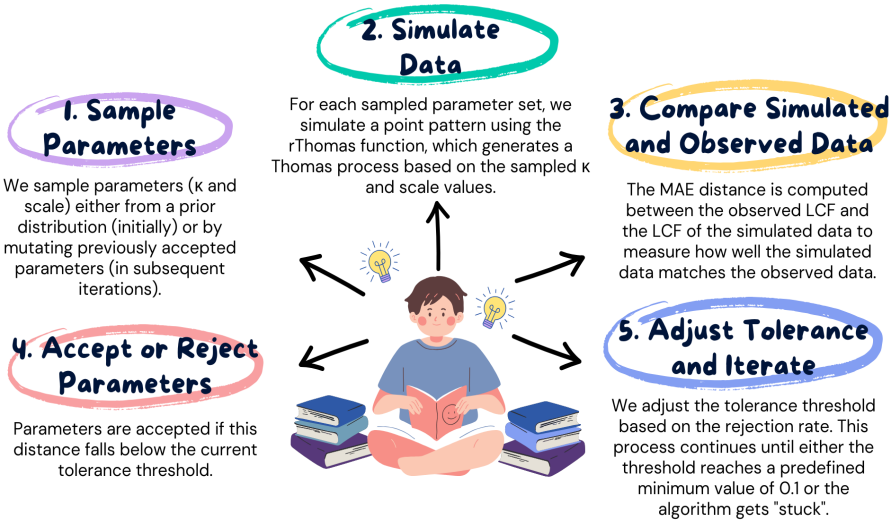


Figure 3.2: Five main steps of our ABC algorithm.

Configuration of the algorithm run

The ABC algorithm was configured with the following parameters:

- population size: 500
- parameter ranges:
  - κ: minimum = 1×10^{-8}, maximum = 1×10^{-5}.
  - scale: minimum = 1μm, maximum = 400μm.
- target rejection rate: 90%
- mutation thresholds: 5×10^{-7} for κ and 5 for scale.

3.2.4 Parameter reliability assessment across tissue samples

To evaluate the reliability of the inferred spatial parameters across different tissue samples from the same patient, we employed the intraclass correlation coefficient. For our analysis, we used a one-way random model with absolute agreement for single measurements. We chose a one-way model because we do not know how the two different tissue samples were taken from patients. It is quite likely that they were obtained at different times and by different specialists, which makes the sources of measurement variation random rather than systematically identifiable. We selected absolute agreement rather than consistency because we want to determine if the actual parameter values are similar between tissue samples from the same patient, not just if they are consistently ranked.

3.3 Results

We present our results in three parts: first, the validation of our clustering approach using synthetic data with known patterns (Section 3.3.1); second, the application of this method to patient tumor samples (Section 3.3.2); and third, the characterization of the underlying spatial processes via parameter estimation (Section 3.3.3).

3.3.1 Hierarchical clustering clearly shows distinct groups in synthetic data

We simulated 5 types of spatial point patterns (3 clustered, random, and dispersed) and computed their LCF as a summary statistic. Afterward, we used hierarchical clustering to group the patterns based on their LCF curves. Finally, we evaluated the clustering results using the Adjusted Rand Index and Cramér's V.

We worked with 5 types of spatial point patterns with 100 elements in each (for detailed parameters see Appendix A): random, dispersed, and 3 clustered patterns with varying cluster characteristics - a few large clusters (clustered1), many large clusters (clustered2), and small clusters (clustered3). We added noise to all clustered patterns using a Poisson process to make the problem less straightforward. A few randomly selected patterns from each type can be seen in Figure 3.3.

We also visualized different types of point patterns along with their corresponding LCF curves to evaluate whether the patterns could be clearly distinguished (Figure 3.4). Each LCF shows a distinct curve, indicating a unique spatial distribution.

The dendrogram obtained with hierarchical clustering reveals a clear structure in the synthetic data (Figure 3.5). Based on the distance between points in the dendrogram branches, 5 clusters seem to be a reasonable choice. Two clusters with the closest inter-cluster distance correspond to clustered1 and clustered2, which indeed have the most similar spatial structure. In Table 3.1 we include a confusion matrix, which shows slightly higher misclassification rates between these two clusters, indicating that it is harder to differentiate them than other clusters.

	Actual				
Calculated	clustered1	clustered2	clustered3	dispersed	random
clustered1	97	4	0	0	0
clustered2	3	96	0	0	1
clustered3	0	0	100	0	0
dispersed	0	0	0	100	0
random	0	0	0	0	99

Table 3.1: Confusion matrix with classification results. Overall, the classification is excellent: only 8 patterns are misclassified.

Using MDS visualization in which we show 5 identified clusters in different colors (Figure 3.6), we check whether the obtained clusters are well separated. The plot shows 3 separate clusters; however, it also reveals that 3 patterns—clustered1 (red), clustered2 (green), and random (cyan)—although not well-separated, occupy distinct regions in space obtained with MDS. This suggests that their LCF curves can be quite similar. This similarity may explain some of the classification mistakes shown in the confusion matrix (Table 3.1). On the other hand, clustered3 (blue) and dispersed (purple) patterns lie far from the other clusters. This is probably why the model clas-

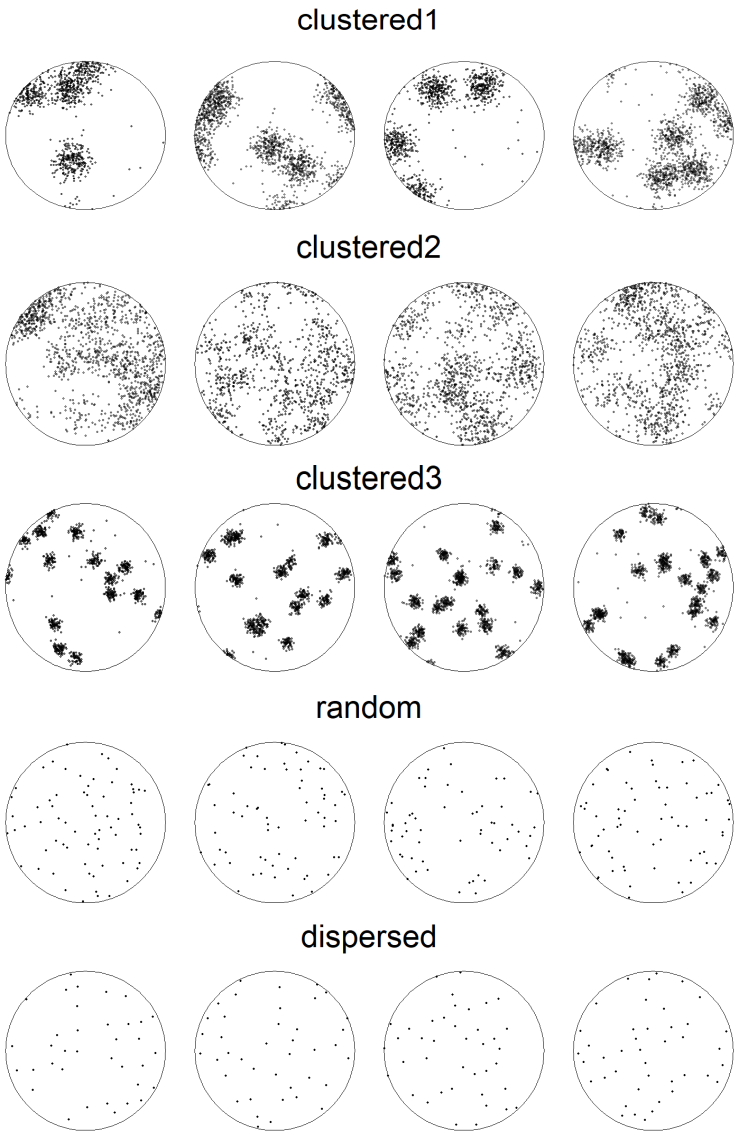


Figure 3.3: Four randomly selected realizations of each of the five point processes. In the first three rows, different clustering patterns created using the Thomas process, each using a different set of parameters, are shown. In the "random" row, the realizations of a Poisson process with an intensity of 2.228×10^{-5} can be seen. Lastly, the "dispersed" row illustrates a pattern drawn from the Hardcore process, where the points are kept at least 120 units apart.

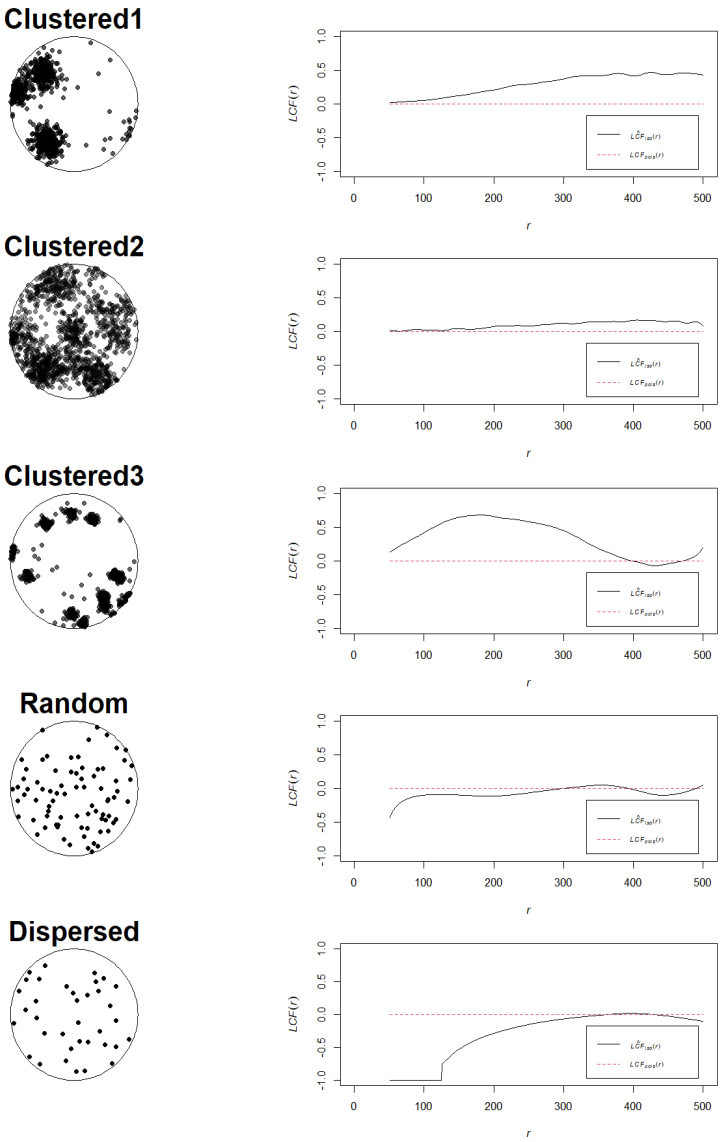


Figure 3.4: LCF curves revealing spatial organization. The first point pattern of each type is illustrated next to its corresponding LCF curve, represented by the solid line. The dashed line indicates CSR.

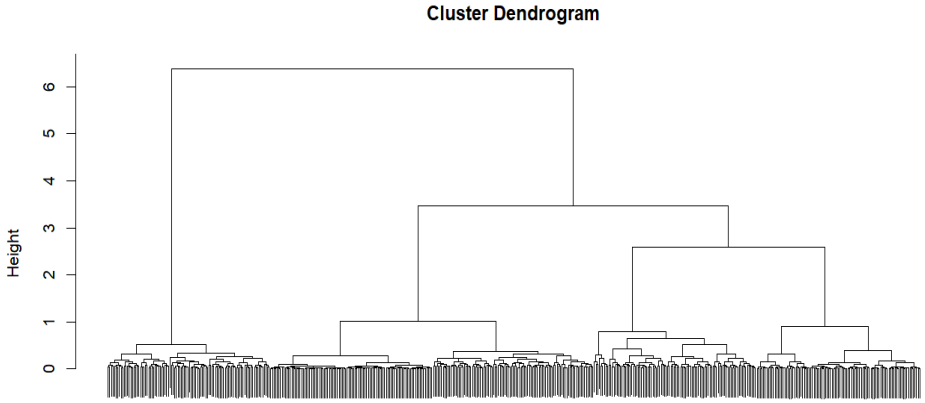


Figure 3.5: Hierarchical clustering dendrogram for synthetic data. The vertical axis (height) defines the distance between the clusters, where a lower value means greater similarity.

sified them correctly every time.

We can see an interesting pattern when we look at this MDS plot. Note how the data points are positioned. There is a smooth transition from clustered1 (red) to clustered2 (green), which then transitions into random patterns (cyan). Since clustered1 and clustered2 use the same clustering process and have the same scale parameter, their LCF curves look quite similar (see Figure 3.4). The main difference is that the LCF curves of clustered1 have higher peaks than clustered2. This happens because clustered1 has fewer parent points (κ parameter). Naturally, random patterns have LCF curves that stay close to zero. Since clustered2 has many clusters and high cluster size (scale), the overall distribution of points is quite close to random, and its LCF curves reach just slightly above zero. Therefore, clustered2 serves as a connection between clustered1 and random patterns.

Besides these smooth transitions, we can also observe how some clusters are tightly packed while others are more spread out in the MDS space. The spread within each cluster indicates the extent of variation possible for each pattern type. Clustered2 is grouped very closely because when there are many overlapping clusters spread throughout the observation window, there are limited ways this can actually appear - most realizations end up similar, generating nearly identical LCF curves. Random patterns show a slightly higher variation, primarily due to their lower point density compared to clustered patterns. Similarly, dispersed patterns demonstrate medium variability; the limitation here arises from the minimal distance restriction. Clustered1 and clustered3 are much more spread out, which is likely due to the fact that clusters within these patterns can be positioned and organized

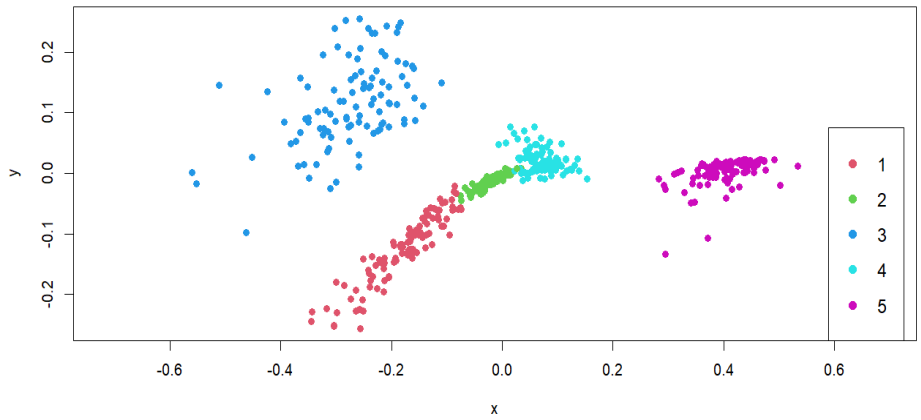


Figure 3.6: Multidimensional scaling of LCF curves of synthetic point patterns with the results of hierarchical clustering as coloring. It can be seen how our pattern types look when mapped onto a two-dimensional plane. Each dot represents one pattern, and colors show which group it belongs to (1-5). The spread of dots tells us how similar or different patterns are from each other.

in numerous ways, resulting in diverse LCF curves.

Performance evaluation of the clustering results is presented in Table 3.2. Both the ARI value of 0.961 and the Cramér's V value of 0.981 show an almost perfect agreement between the two clusterings. Overall, our clustering method did an outstanding job of finding the right groups, which shows that LCF is capable of capturing the distinct types of spatial organization.

Metric	Value
Adjusted Rand Index	0.961
Cramér's V	0.981

Table 3.2: Clustering accuracy metrics. This table shows two numbers that indicate the effectiveness of our clustering method in identifying the correct pattern groups.

3.3.2 Patient tumor samples show continuous spatial landscape

After ensuring that hierarchical clustering can identify distinct types of point patterns, we applied it to the patients' tumor samples. By clustering tissue samples based on the spatial distribution of cells quantified with LCF, we aim to identify distinct tumor subgroups that exhibit similar immune cell arrangements. We loaded each tissue sample's spatial data of lymphocyte positions, precomputed tumor boundaries, and observed LCF for B cells. The clustering methodology established using the synthetic patterns was applied. Since we do not know the actual number of clusters in real data, we used the silhouette score and the elbow method to aid in choosing the optimal number of clusters. The elbow method is based on calculating the sum of squared errors for different numbers of clusters. The "elbow" refers to the point where adding more clusters does not show significant improvement; it's the point where the curve starts to flatten out [20].

The resulting dendrogram is visualized in Figure 3.7.

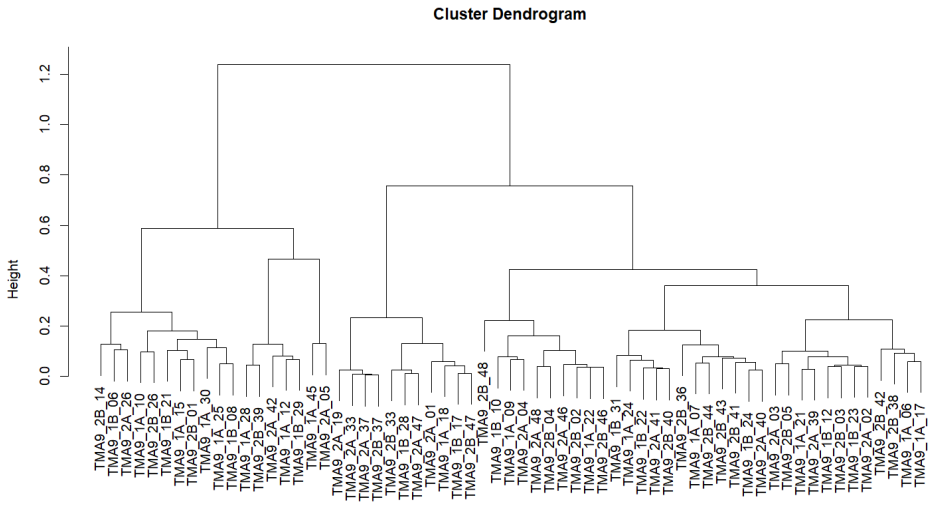


Figure 3.7: Hierarchical clustering dendrogram for tissue samples. This dendrogram organizes tumor samples into subgroups based on their B cell LCF. Each label at the bottom represents an individual tumor sample.

The silhouette score analysis revealed that the highest score was achieved with 2 clusters. However, this result was too broad for our analysis; we wanted to capture the finer distinction in cell distribution. The next best score was obtained with 5 clusters. The elbow method showed that at 5 clusters, the curve begins to flatten, suggesting that it is a suitable choice

(see Figure 3.8). Therefore, we opted to split the data into 5 clusters (see Appendix B for tissue sample assignments).

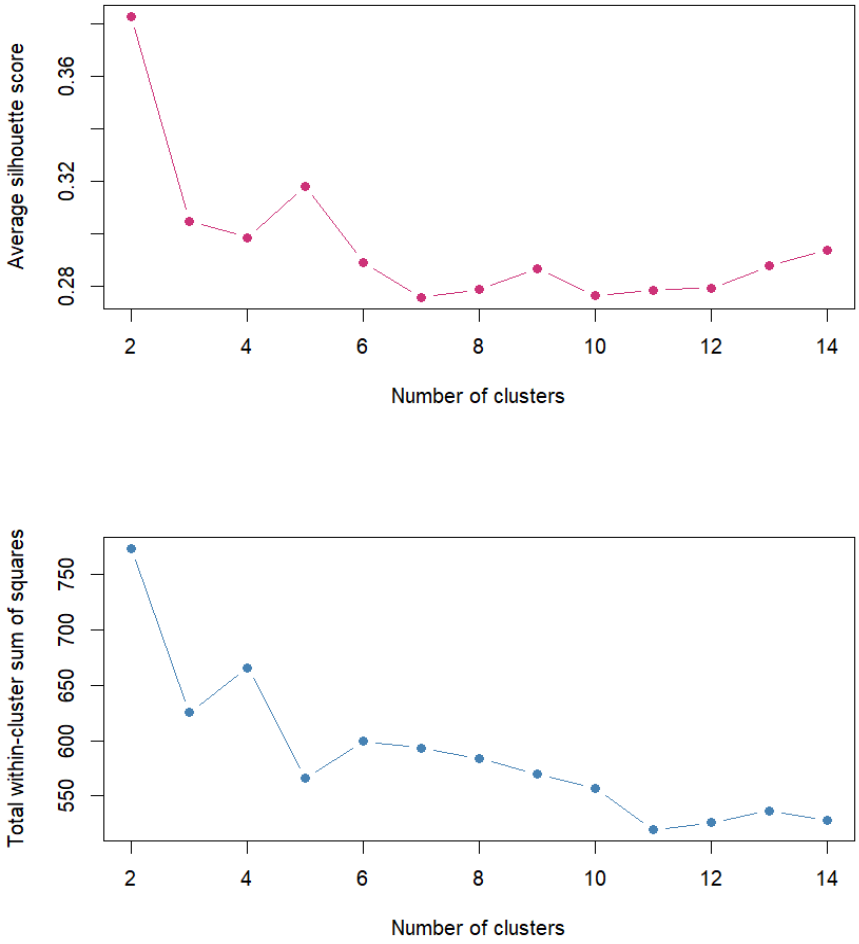


Figure 3.8: Silhouette score (top) and elbow method (bottom) for optimal cluster selection. Both methods show that 5 clusters are a reasonable choice.

We visualized these five clusters in two-dimensional space (Figure 3.9). The MDS plot reveals a pattern similar to what we observed with the synthetic clustered1, clustered2, and random patterns: although there is no clear separation between all the clusters, the assigned clusters occupy distinct regions within the MDS space. Notably, two samples from cluster 5 stand out because they are located far from all other samples. This suggests they have

drastically different immune cell patterns compared to the other samples.

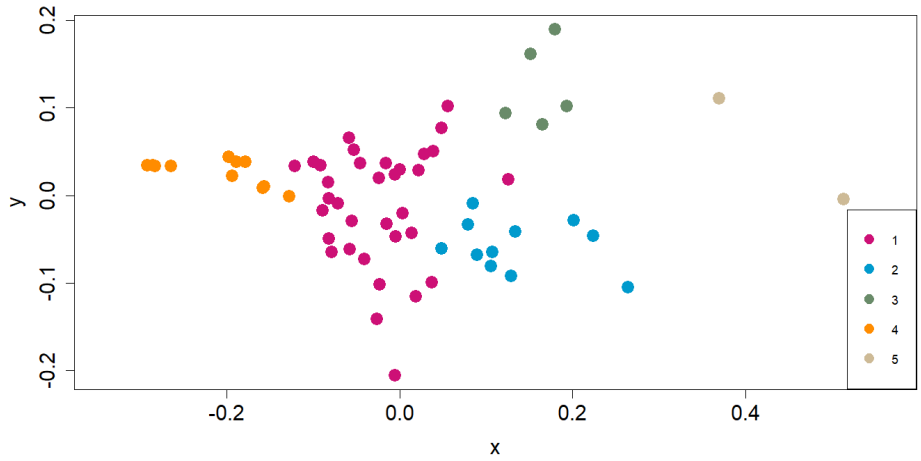


Figure 3.9: Multidimensional scaling of LCF curves of tissue samples with the results of hierarchical clustering as coloring. A simplified two-dimensional map of our tumor data can be seen, where each dot represents a tumor sample, with colors showing which of the five clusters it belongs to according to our analysis. Similar tumors are placed closer together based on their LCF curve for B cells.

3.3.3 Parameter estimates match observed spatial distributions

Although we did not find well-defined clusters of spatial distribution, samples from clusters selected through hierarchical clustering are located in distinct regions of the MDS space. We decided to investigate whether tumor samples that are closer together in a lower-dimensional space share similar underlying spatial characteristics by estimating Thomas process parameters for each sample.

To analyze these parameter estimates, we created a dataframe from the output of the ABC process, including the following variables: sample ID, κ along with its credible interval, and the scale with its credible interval. Additionally, we incorporated the μ parameter and the intensity of the process, as well as our identified clusters. All the plots presented in this subsection are based on this specific datasheet.

κ and scale distribution in different clusters

We examined whether the distributions of κ and scale differ across the chosen clusters (Figure 3.10). To create these plots, we found the mean of a parameter in each cluster, ordered the clusters by these means, and then plotted the parameters against the clusters.

We observe big differences in both κ and scale parameters between clusters. Upon examining the κ values, we find something interesting: most clusters display unique values, but cluster 1 appears to overlap with several others. The only clusters that stand out as different from it are 3 and 5. Cluster 1 shows the highest within-cluster spread (look at its wide box plot and dispersed data points). Cluster 2 also demonstrates higher within-cluster spread, while clusters 3, 4, and 5 have much tighter distributions with only slight differences in their κ values. Looking at the scale values, we see that there is more difference in scale parameters than κ across our clusters. We can again see that some clusters have more variation than others: clusters 2, 3, and 5 have very tight distributions with very little variability in their scale values, while clusters 1 and 4 demonstrate much higher within-cluster spread. In particular, cluster 4 stands out, showing the widest distribution.

In general, we can observe considerable overlap among the clusters we have identified. This supports our earlier observation that the differences between them are not very pronounced. This makes it difficult to truly categorize them as separate clusters. However, we would like to point out that some differences are still noticeable. These partially distinct groups suggest that the arrangement of tumors in space may be affected by a combination of clinical factors that we have not yet explored. It will be interesting to see if there is any correlation with clinical data.

To provide further context, we have included descriptive summary tables detailing the mean, standard deviation, minimum, and maximum values of κ and scale for each cluster (see Appendix C).

Relationship between κ and scale parameters

We continued our analysis by creating a scatterplot that shows how κ values relate to scale values, along with their 95% credible intervals, which help us understand the uncertainty in these estimates (Figure 3.11). Different colors represent clusters obtained with hierarchical clustering. It is remarkable that the inferred κ and scale parameters also form somewhat distinct regions when split by the selected clusters.

The scatterplot demonstrates a positive relation - as the scale increases,

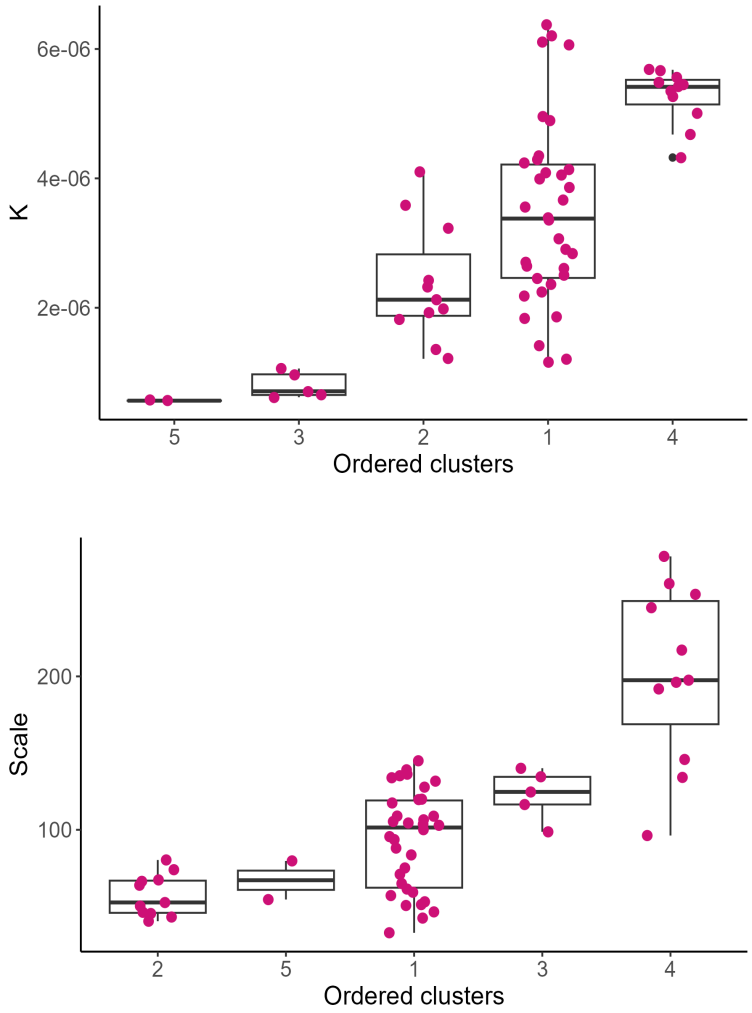


Figure 3.10: Distribution of κ (top) and scale (bottom) by cluster. Standard box plots show the distribution of values within each cluster. The individual dots scattered around show the actual tumor samples.

κ tends to grow. This indicates that tumors with B cells that are more tightly packed generally can be described by a process with lower cluster intensity, while tumors with dispersed B cells are typically characterized by a process with higher cluster intensity. However, this relationship is mainly driven by a few samples assigned to cluster 4.

In addition, note that as the values increase, so does the uncertainty. The points of cluster 4 (in orange), located in the upper right corner, have wide intervals for both dimensions, indicating that we have less confidence in the estimates of higher κ and scale values. The overlapping intervals between different clusters suggest that we cannot easily distinguish among them. As a result, some noted differences may lack statistical significance.

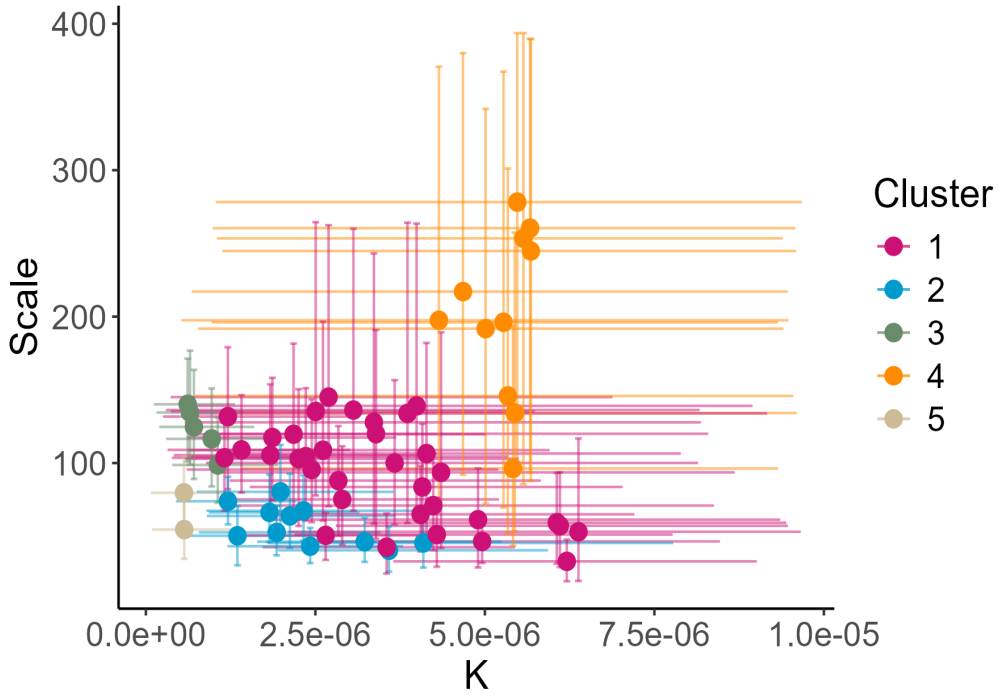


Figure 3.11: Scatterplot of κ vs scale with 95% credible intervals represented as lines extending from each point. This plot shows how uncertain we are about the parameter estimates, which gives us a clearer picture of the data's reliability.

When we examine Figures 3.11 (the scatterplot of κ vs. scale) and 3.9 (the MDS with hierarchical clustering), we observe an interesting pattern: the clusters are arranged in the same order in both figures. Specifically, cluster 4 is close to cluster 1 in both plots, cluster 1 is also located near clusters 2

and 3, and cluster 5 is near clusters 2 and 3. This similarity suggests that these identified patterns may reflect actual differences in the organization of the tumor microenvironment, even though the boundaries between clusters are not sharply defined.

ABC rejection rates across clusters

Figure 3.12 provides additional insights into the model fitting process across the different tumor groups, revealing an interesting pattern when compared to our earlier findings. Clusters 1, 2, 3, and 5 exhibit high rejection rates; it was similarly hard to fit these groups. In contrast, cluster 4 has much lower rejection rates. A possible interpretation is that samples in this cluster can be fitted with a wider range of scale and kappa parameters, which would explain both the lower rejection rates and the broad credible intervals shown in Figure 3.11.

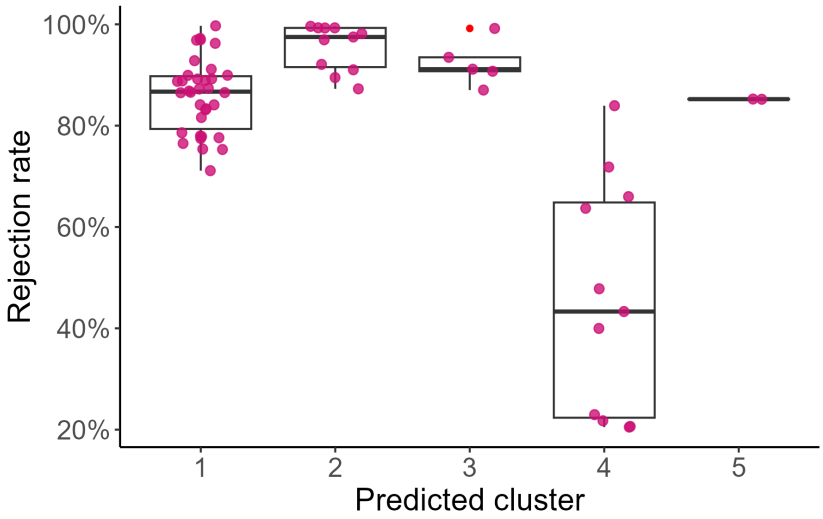


Figure 3.12: Distribution of rejection rates by cluster. This graph shows the distribution of final rejection rates for each cluster. Lower values indicate greater ease in finding the parameters that produce a similar spatial distribution.

It is noteworthy that the simulated patterns do not always match the real tumor data perfectly. Our model naturally creates circular clusters with some spread, which is insufficient to represent all types of spatial distribution that can occur in tissue samples. We analyzed tumor sample data from several patients and compared it to the simulated patterns (see Appendix E). Sometimes, the simulation matched the real data quite well, but at other

times, it failed to capture the complexity of the actual distribution of cells. This helps us explain certain rejection rates, as some patterns are more challenging to represent using the Thomas process.

Parameter reliability across tissue samples

Furthermore, we wanted to inspect the agreement between parameters inferred from two distinct tissues of the same patient. We calculated intraclass correlation coefficients, as demonstrated in Figure 3.13.

In the top plot of Figure 3.13, the points are scattered all over the graph, indicating a relatively poor agreement between the κ values from the two tissues. This analysis is supported by an ICC score of 0.49. On the contrary, for the scale parameter, points are clustered closer to the y = x line. This indicates a decent agreement, confirmed by an ICC value of 0.76. However, we believed that this agreement may be influenced by two outlier points with significantly higher scale values. Thus, we recalculated the ICC after excluding the outliers. This resulted in an ICC of 0.45, which confirmed that extreme values indeed drove the apparent superior reliability of scale measurements.

Additionally, we examined samples from the same patient and found that only about half (56.5%) of patients had their two tissue samples assigned to the same cluster (Table 3.3). This means that different parts of the same tumor can show different B cell patterns.

3.4 Discussion

In this thesis, we conducted a spatial analysis of B cell distribution in samples of several types of cancer. MDS output revealed a lack of distinct clusters in our data; instead, we found that the spatial distribution of B cells forms a continuous landscape. Nevertheless, we identified specific regions in the space of spatial distribution using the LCF curves. This section provides a reflection on the research process. Additionally, we include the limitations as well as the recommendations for future research.

3.4.1 Interpretation of the results

Reflecting on our hypothesis

The results of the analysis do not support the hypothesis that the spatial distribution patterns of lymphocytes, quantified using LCF, can help identify patient subgroups with similar tumor spatial structures. Although we did not find distinct clusters, our visualizations consistently revealed organized patterns in the tumors, which indicates differences in how the immune

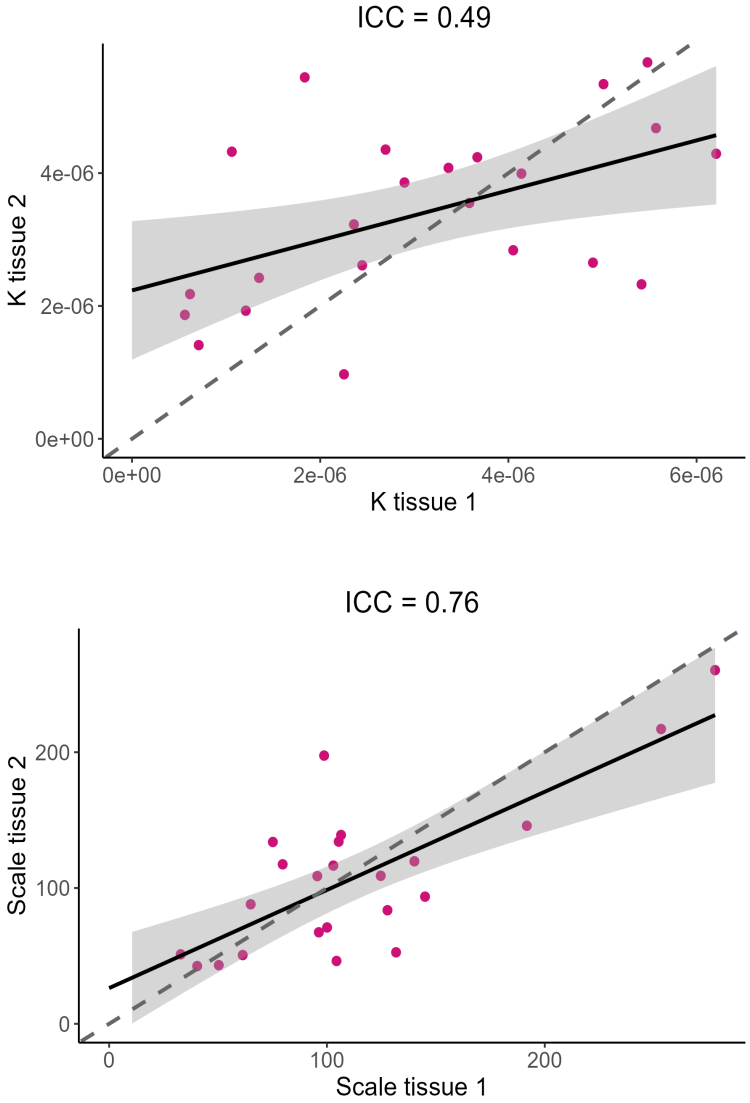


Figure 3.13: The relationship between κ (top) and scale (bottom) values measured in tissue samples 1 and 2. Each dot represents a patient, comparing their parameter value in tissue sample 1 (x-axis) to tissue sample 2 (y-axis). The dashed line represents perfect agreement (where both tissue samples would have identical parameter values). The solid black line shows the fitted regression line through the actual data points. The gray shaded area indicates the confidence interval.

Patient ID	$\kappa 1$ ($\times 10^{-6}$)	$\kappa 2$ ($\times 10^{-6}$)	Scale1	Scale2
1	5.41	2.33	96.3	67.4
2	3.67	4.24	100.00	71.00
3	2.45	2.61	95.60	109.00
4	6.21	4.29	32.90	51.20
5	0.56	1.87	79.70	117.00
6	1.21	1.93	132.00	52.60
10	3.58	3.55	40.40	42.50
12	0.62	2.18	140.00	120.00
17	1.84	5.44	105.00	134.00
21	2.36	3.23	104.00	46.30
22	2.90	3.86	75.20	134.00
24	2.69	4.35	145.00	93.60
26	1.35	2.42	50.30	43.20
28	1.06	4.32	98.70	198.00
33	5.57	4.68	253.00	217.00
37	5.48	5.67	278.00	260.00
39	2.25	0.97	103.00	117.00
40	4.14	3.99	106.00	139.00
41	3.36	4.08	128.00	83.70
42	0.71	1.41	125.00	109.00
46	4.05	2.84	65.00	88.00
47	5.01	5.34	192.00	146.00
48	4.90	2.65	61.30	50.60

Table 3.3: κ and scale measurements of patients with two tissue samples available. The tissue samples of patients in the highlighted rows were clustered together.

cells are arranged. This suggests that instead of distinct categories of spatial distribution of immune cells, there exists a continuous landscape of possible spatial arrangements. Therefore, utilizing a continuous representation of spatial distribution in cancer studies is more promising. For instance, PCA can be applied to the extracted LCF, and scores of principal components that explain a sufficient percentage of variance can be used. At this point, we cannot fully explain why these groups are so closely aligned. Are certain cancers alike in some ways? Are these patterns tied to the treatments patients received or their genetic makeup? Even with these uncertainties, it is pretty clear that we are seeing similar trends in different plots. This makes us believe that we are observing genuine biological differences in tumors.

Additional discovery

While analyzing our data, we made an interesting discovery: a highly dense point distribution can be best fit with a Thomas process having high scale and κ parameters. This occurs because when cell density is high, the cells occupy almost the entire sample area, causing their spatial distribution to approach a random pattern. Under these conditions, placing many parent points with large cluster sizes effectively approximates this random, high-density distribution. Additionally, our reliability analysis showed that the scale appeared to be much better than kappa at first (ICC = 0.76 vs. 0.49). However, this was influenced by two outlier points. When we removed them, both parameters showed similar and fairly low reliability (ICC = 0.49 vs. 0.45), indicating that both measurements vary significantly between tissue samples from the same patient. This suggests that neither parameter, when considered separately, is particularly reliable for characterizing individual patients. This presents an important limitation for clinical applications.

3.4.2 Limitations

Limited dataset size

A significant limitation in this data analysis is the data itself. In our study, we worked with a dataset of just 63 tissues, which is quite small. This may not adequately reflect the diversity of spatial patterns found in larger or more varied datasets. As a result, we may have missed some potential clusters.

Dataset

We worked with the dendritic cell panel that contains CD19 marker to characterize B cells (CD19+BDCA1+/-). The population defined this way differs from the conventional B lymphocytes in two important ways. First, CD19 is a broad marker of the B cell lineage, which also includes plasma

cells that have distinct functions compared to conventional B cells. Second, B cells expressing BDCA1 have distinct functions as well, which can affect the positions of these cells. Both factors could interfere with our results.

Computational limits

Some tissues contained a large number of cells, and without adjustments, processing them would have taken an excessive amount of time. To address this, we scaled down the intensity for these samples. Ideally, we should have analyzed whether downsampling significantly affects the computed LCF.

Small tissue sample size

We used tissue microarray cores with a diameter of just 2 mm, representing a small fraction of the overall tumor tissue. These small samples may fail to accurately represent the overall immune landscape of the tumor, which could affect our clustering results and the generalizability of our findings to the broader tumor microenvironment.

3.4.3 Future research

Looking only at the spatial distribution of B cells limits our ability to understand the immune landscape. To obtain its comprehensive picture, LCF curves of all cell types identified by the used panel should be analyzed jointly. Checking whether the way cells are organized is related to clinical information and genetic markers would be another valuable research direction. For instance, a spatial organization might be associated with the tumor stage, or there might be some genetic markers that affect a patient's immune response, which is visible through the spatial landscape of a tumor. Lastly, dimensionality reduction techniques, such as principal component analysis (PCA), can be promising to enhance our analysis. Essentially, PCA finds the strongest patterns in the data, and it can eliminate noise, allowing us to see clearer trends [19]. PCA can be applied to the extracted LCF and the scores of the resulting principal components that explain a sufficient percentage of variance can be used directly in survival analysis. This would demonstrate whether specific spatial patterns correlate with patient outcomes without the need to create separate groups. There is so much to discover!

Chapter 4

Related Work

Each year, we gain a deeper understanding of the spatial organization of immune cells within the tumor microenvironment. Numerous studies have investigated this spatial distribution, providing fresh insights and new methodologies, as shown in Table 4.1. In this chapter, we will describe the methods and results obtained from some of these studies and compare them to our research.

Paper	Immune cell investigated	Methodology employed	Research findings
A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging	Various immune cells (e.g., B cells, helper and cytotoxic T cells)	Multi-step pipeline to analyze Multiplexed Ion Beam Imaging (MIBI)	Three subtypes of tumor-immune interactions were identified: cold, mixed, and compartmentalized. A compartmentalized organization is associated with improved survival rates.
Physics approaches to the spatial distribution of immune cells in tumors	B cells, T cells	Maximum entropy method	Better outcomes are linked to dispersed B cells and higher cytotoxic T cell density, while poorer outcomes are associated with clustered B cells.
Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front	Various immune cells (e.g., T cells, macrophages)	CODEX multiplexed imaging, tensor decomposition	Identified nine cellular neighborhoods. The enrichment of PD-1 expressing helper T cells in granulocyte-rich areas correlated with improved survival.
Tumor immune cell clustering and its association with survival in African American women with ovarian cancer	Cytotoxic T-cells, regulatory T-cells	Permutation-based spatial analysis (Ripley's K function)	High abundance of tumor-infiltrating lymphocytes and low clustering are associated with improved survival. Additionally, the co-occurrence of cytotoxic and regulatory T cells also enhances survival.
Tumor-immune partitioning and clustering algorithm for identifying tumor-immune cell spatial interaction signatures within the tumor microenvironment	T cells, eosinophils, neutrophils	Multiplex Immunofluorescence, Tumor-Immune Partitioning and Clustering (TIPC) algorithm	Six tumor subtypes were identified based on spatial immune patterns. Three tumor subtypes are linked to better survival outcomes.

Table 4.1: Related research summary

4.1 A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging

In this paper, Keren et al. developed a multi-step pipeline to analyze multiplexed ion beam imaging by the time of flight data to investigate the spatial distribution of the immune microenvironment in triple-negative breast cancer. Researchers discovered that the presence of one cell type is frequently accompanied by the presence of others; for instance, all patients with B cells also had helper and cytotoxic T cells. This study also found three subtypes of tumor-immune interactions: cold (few immune cells), mixed (tumor and immune cells highly mixed), and compartmentalized (regions composed mostly of either immune or tumor cells). Patients with compartmentalized patterns showed better survival rates.

This paper explores a broader range of topics than our study, particularly the interactions between immune cells, although the focus remains the same—the spatial organization of the tumor-immune microenvironment. However, this work has an important limitation: a single distance was used to quantify interactions between cells. Our results present an interesting contrast: while Keren et al. identified distinct spatial subtypes, our analysis revealed a continuous landscape of patterns. Our results make us question the conclusions of this paper. However, our analysis is specifically related to the spatial distribution of B cells and cannot directly contradict their broader conclusions about multiple immune cell types. Interestingly, they themselves recognize that "the distinction between compartmentalized and mixed is not clear cut, as we observe a continuum of mixing scores between tumor and immune cells across patients" [7]. Yet despite this acknowledgment, they apply a cutoff of their "mixing score" to create discrete categories. This is a questionable approach. It seems to impose artificial limits on what they acknowledge is actually a continuous space.

4.2 Physics approaches to the spatial distribution of immune cells in tumors

This research presents a novel maximum entropy method to quantify the spatial distribution of discrete point-like objects. Entropy can be defined as the measurement of the degree of randomness. Yu et al.'s method divides the image into blocks of a chosen size, and further, each block is split into 3x3 squares. For each square, the binary question is asked: 'Is there at least one immune cell in the square?'. If this is the case, 1 is assigned to the square; otherwise, 0 is assigned. To measure the degree of randomness or

clustering of immune cells, entropy is calculated across all possible patterns of these 1s and 0s.

Yu et al. employ this new maximum entropy method to analyze the spatial distribution of B and T cells in tumor samples taken from triple-negative breast cancer patients and explore how the arrangement of these immune cells impacts cancer recurrence. This study closely resembles ours. Both studies employ mathematical methods that capture spatial organization at multiple scales to measure the distribution and clustering of immune cells in tumors. Our approach offers a few advantages over the entropy method. LCF provides better interpretability for medical specialists. Furthermore, Yu et al. use square tessellation to calculate their metric. This method is not ideal because the arrangement of the resulting matrix can affect the outcome, especially with larger matrix sizes. This is not an issue with our method since spatial statistics rely solely on the positions of the cells.

This study reveals significant differences in the distribution of immune cells between patients who did not experience cancer recurrence within at least five years and those who had a recurrence within three years. Patients with better outcomes showed a more dispersed spatial distribution of B cells. On the other hand, patients with poorer outcomes had a more clustered and irregular organization. Patients with better outcomes also had a higher density of cytotoxic T cells [23].

4.3 Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front

Schürch et al. use a technique known as co-detection by indexing (CODEX) to study the immune tumor microenvironment (iTME) of colorectal cancer (CRC). Researchers try to understand the spatial organization of the iTME of two groups: patients who exhibit de novo formation of numerous tertiary lymphoid structures (TLSs), known as a "Crohn's-like reaction" (CLR), and patients who lack TLSs and instead have a diffuse inflammatory infiltration (DII). These two groups have very different prognoses - patients with CLR survive much longer. The spatial organization of the iTME might be related to these differences in survival.

This study identified 28 distinct cell types (CT) within the iTME. The researchers identified 9 distinct cellular neighborhoods (CNs) that are crucial for antitumor immunity by clustering the cells according to the local density of various CTs. They define neighborhoods using distance-based methods. Specifically, they examine the ten nearest spatial neighbors around each cell

using a sliding window approach that moves across all cells in the tissue. However, their method uses only a single distance to obtain their cellular neighborhoods rather than examining multiple scales of a spatial organization, like our LCF does.

The methods they used are considerably more complex than ours. They employ non-negative Tucker tensor decomposition and canonical correlation analysis, which are difficult to interpret, even for data scientists. Our approach is more straightforward and interpretable.

Using these complex methods, the researchers found that the coupling of tumor and immune components is greater in DII patients than in CLR patients. This analysis indicates that in DII patients, a T cell-enriched CN is coupled to a macrophage-enriched CN, and these two CNs are also more spatially mixed. Additionally, Ki-67+ cytotoxic T cell density in the T cell-enriched CN was anti-correlated with the frequency of regulatory T cells in the macrophage-enriched CN in DII patients, suggesting that the interaction between T cells and macrophages may be immunosuppressive.

Researchers also found that the frequency of PD-1 expressing helper T cells within granulocyte CNs correlated positively with survival in DII patients. Interestingly, the overall frequency of PD-1 expressing helper T cells was not associated with survival, which suggests that the spatial localization of these cells is essential [16].

4.4 Tumor immune cell clustering and its association with survival in African American women with ovarian cancer

Wilson et al. developed a new technique for examining immune cell clustering within tumors and its effect on survival: a permutation-based spatial analysis framework that uses Ripley's K function. Researchers concentrated on epithelial ovarian cancer in African American women and studied tumor-infiltrating lymphocytes, cytotoxic T cells, and regulatory T cells.

This research is one of the closest to ours in terms of methodological approach. Wilson et al. use Ripley's K function, while we chose to use LCF, which is based on Ripley's K. However, Wilson et al. ultimately reduced their K-function results to a binary categorization (low/high degree of clustering) for use in Cox survival models (for clinically interpretable results), which discards most of the rich spatial information that the K-function provides. Since LCF's range is designed to be inherently interpretable, we avoid such reductive solutions. Additionally, our clustering approach to identify

tumor subtypes makes better use of the available information by assuming there might be more than just two types of spatial arrangements, rather than forcing a binary classification.

What they found is quite interesting: patients survived longer when they had many immune cells (tumor-infiltrating lymphocytes and T cell subsets) that were evenly distributed rather than tightly clustered. Additionally, the co-occurrence of cytotoxic T cells and regulatory T cells also turned out to have a positive influence on survival.

Notably, Wilson et al. found that models using spatial information predicted outcomes better than those using only immune cell counts [22].

4.5 Tumor-immune partitioning and clustering algorithm for identifying tumor-immune cell spatial interaction signatures within the tumor microenvironment

Lau et al. introduced an algorithm called Tumor-Immune Partitioning and Clustering (TIPC), designed to illustrate the organization of immune cells within the tumor microenvironment. It was created to overcome the limitations of other traditional methods; TIPC does not focus primarily on counting immune cells or examining their proximity. Instead, it measures immune cell partitioning between tumor epithelial and stromal areas and immune cell clustering versus dispersion.

Similarly to our work, TIPC focuses on unsupervised clustering of spatial patterns of immune cells within the tumor microenvironment. Our study uses the LCF curves to identify subgroups, while TIPC defines spatial subtypes based on how T lymphocytes are distributed between tumor and stroma. However, this work has several limitations. TIPC requires a subregion size to be selected. This can affect the outcome and make the applicability of this method more difficult. Additionally, while the researchers claim their method preserves more information than spatial statistics approaches, this is questionable. TIPC produces a six-element numerical vector that represents different spatial categories: tumor-only, I:T low, I:T high, stroma-only, I:S low, and I:S high; however, our approach uses the entire LCF curve to cluster patients, which preserves much more spatial information.

Six TIPC subtypes were discovered, with two "cold" and four "hot" subtypes: hot and disperse (include tumors with a dispersed distribution of T

cells across tumor and stromal regions), hot, tumor-centric clustering (T cells cluster in tumor intraepithelial regions), hot, stroma-centric clustering (T cells cluster in tumor stromal regions), hot and clustered (T cells cluster both in tumor intraepithelial and stromal regions), cold, tumor-rich (tumors with uniformly few T cells and predominant tumor regions), cold, stroma-rich (tumors with uniformly few T cells and predominant stromal regions). Importantly, three of the four hot subtypes showed significantly better survival rates than cold subtypes.

T cell densities across different TIPC subtypes did not affect survival outcomes. This means that immune cell distribution - not just quantity - impacts patient outcomes [9].

Chapter 5

Conclusions

In this thesis, we examined the arrangement of B cells in tumor tissue samples from patients with various types of cancer. Our study employed a new technique, the Local Correlation Function (LCF), to analyze the distribution of immune cells within tumors. Our primary goal was to check whether we could categorize patients based on the similarity of lymphocyte distribution in their samples using LCF. To this end, we clustered the LCF curves with a clustering algorithm validated on synthetic data. However, when applied to real tumor data, we found that tissue samples do not fall into distinct clusters. Instead, the spatial patterns of B cells reveal a continuous landscape of spatial organization. Interestingly, the clusters we chose occupied distinct regions within this continuous space and corresponded to different parameters of the Thomas process that we inferred for our samples. This suggests that the spatial landscape has a meaningful structure, even though the boundaries between tumor sample groups are not sharply defined.

Reliability analysis of the Thomas process parameters (κ and scale) revealed significant variability in both parameters across different tissue samples from individual patients. The ICC values ranged from 0.45 to 0.49 after we removed outliers. This suggests that individual parameters alone are not reliable enough for characterizing patient immune patterns, at least when obtained from these small tissue cores. It may be more reproducible if our tissue samples were larger.

Moving forward, to obtain a comprehensive picture of the immune landscape, it is essential to analyze the spatial organization of all immune cell types identified by the panel. Furthermore, using spatial information together with clinical data and genetic markers could also enhance our analysis. Rather than categorizing patients into distinct groups, we should consider employing methods like principal component analysis on LCF data. Since our space of spatial distribution is continuous, it is more effective to

use a few continuous variables to describe it (scores of principal components found with LCF). We could then apply these principal components in survival analyses to evaluate whether these patterns can assist in predicting patient outcomes, using that information to guide more informed treatment decisions.

Bibliography

[1] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R.* CRC Press, 2015.

[2] Ingwer Borg and Patrick J.F. Groenen. *Modern Multidimensional Scaling: Theory and Applications.* Springer, 2 edition, 2005.

[3] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques.* The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 3rd edition, 2011.

[4] Timothy O. Hodson. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14):5481–5487, July 2022.

[5] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, December 1985.

[6] Stephen P. Jenkins. *Survival Analysis.* 2005. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.

[7] Leeat Keren, Marc Bosse, Diana Marquez, Roshan Angoshtari, Samir Jain, Sushama Varma, Soo-Ryum Yang, Allison Kurian, David Van Valen, Robert West, Sean C. Bendall, and Michael Angelo. A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell*, 174(6):1373–1387.e19, September 2018.

[8] Terry K. Koo and Mae Y. Li. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2):155–163, June 2016.

[9] Mai Chan Lau, Jennifer Borowsky, Juha P. Väyrynen, Koichiro Haruki, Melissa Zhao, Andressa Dias Costa, Simeng Gu, Annacarolina Da Silva, Tomotaka Ugai, Kota Arima, Minh N. Nguyen, Yasutoshi Takashima, Joe Yeong, David Tai, Tsuyoshi Hamada, Jochen K. Lennerz, Charles S. Fuchs, Catherine J. Wu, Jeffrey A. Meyerhardt, Shuji Ogino, and

Jonathan A. Nowak. Tumor-immune partitioning and clustering algorithm for identifying tumor-immune cell spatial interaction signatures within the tumor microenvironment. *PLOS Computational Biology*, 21(2):e1012707, February 2025.

[10] Evgenia Martynova and Johannes Textor. A Uniformly Bounded Correlation Function for Spatial Point Patterns. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2177–2188. ACM, August 2024.

[11] Fionn Murtagh and Pierre Legendre. Ward's hierarchical clustering method: Clustering criterion and agglomerative algorithm. December 2011.

[12] Giovanni Palla, David S. Fischer, Aviv Regev, and Fabian J. Theis. Spatial components of molecular tissue biology. *Nature Biotechnology*, 40(3):308–318, March 2022.

[13] Edwin Roger Parra. Methods to Determine and Analyze the Cellular Spatial Distribution Extracted From Multiplex Immunofluorescence Data to Understand the Tumor Microenvironment. *Frontiers in Molecular Biosciences*, 8:668340, June 2021.

[14] R Project. Cramér's v. https://search.r-project.org/CRAN/refmans/confintr/html/cramersv.html.

[15] B. D. Ripley. Modelling Spatial Patterns. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 39(2):172–192, January 1977.

[16] Christian M. Schürch, Salil S. Bhate, Graham L. Barlow, Darci J. Phillips, Luca Noti, Inti Zlobec, Pauline Chu, Sarah Black, Janos Demeter, David R. McIlwain, Shigemi Kinoshita, Nikolay Samusik, Yury Goltsev, and Garry P. Nolan. Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell*, 182(5):1341–1359.e19, September 2020.

[17] Ketan Rajshekhar Shahapure and Charles Nicholas. Cluster Quality Analysis Using Silhouette Score. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 747–748, sydney, Australia, October 2020. IEEE.

[18] Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate Bayesian Computation. *PLoS Computational Biology*, 9(1):e1002803, January 2013.

[19] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. *Introduction to Data Mining*. Pearson, 2 edition, 2018.

[20] Edy Umargono, Jatmiko Endro Suseno, and S.K Vincensius Gunawan. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula:. In *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)*, Yogyakarta, Indonesia, 2020. Atlantis Press.

[21] Iris A.E. Van Der Hoorn, Evgenia Martynova, Beatriz Subtil, Jelena Meek, Kiek Verrijp, Johannes Textor, Georgina Flórez-Grau, Berber Piet, Michel M. Van Den Heuvel, I. Jolanda M. De Vries, and Mark A. J. Gorris. Detection of dendritic cell subsets in the tumor microenvironment by multiplex immunohistochemistry. *European Journal of Immunology*, 54(1):2350616, January 2024.

[22] Christopher Wilson, Alex C. Soupir, Ram Thapa, Jordan Creed, Jonathan Nguyen, Carlos Moran Segura, Travis Gerke, Joellen M. Schildkraut, Lauren C. Peres, and Brooke L. Fridley. Tumor immune cell clustering and its association with survival in African American women with ovarian cancer. *PLOS Computational Biology*, 18(3):e1009900, March 2022.

[23] Clare C Yu, Juliana C Wortman, Ting-Fang He, Shawn Solomon, Robert Z Zhang, Anthony Rosario, Roger Wang, Travis Y Tu, Daniel Schmolze, Yuan Yuan, Susan E Yost, Xuefei Li, Herbert Levine, Gurinder Atwal, and Peter P Lee. Physics approaches to the spatial distribution of immune cells in tumors. *Reports on Progress in Physics*, 84(2):022601, February 2021.

[24] Yinyin Yuan. Spatial Heterogeneity in the Tumor Microenvironment. *Cold Spring Harbor Perspectives in Medicine*, 6(8):a026583, August 2016.

Appendix A

Detailed parameters for spatial point pattern simulation

All spatial point patterns were simulated in a disk of radius 1000 units with 100 elements each. The specific parameters used for each pattern type are detailed below.

A.1 Clustered patterns

All clustered patterns were generated using the Thomas cluster process with the following specifications:

clustered1 (few large clusters):

- κ: 1.273×10^{-6}
- scale: 120
- μ: 250

clustered2 (many large clusters):

- κ: 6.366×10^{-6}
- scale: 120
- μ: 50

clustered3 (small clusters):

- κ: 4.775×10^{-6}
- scale: 40

- μ: 66.667

Noise addition: all clustered patterns had additional noise introduced using a Poisson process with λ of 5×10^{-6}, to make the problem less straightforward.

A.2 Random pattern

Generated using a homogeneous Poisson process with:

- λ: 2.228×10^{-5}

A.3 Dispersed pattern

Generated using a Hardcore process with:

- β: 2.228×10^{-5}
- minimum distance (R): 120

Appendix B

Tissue sample assignments

cluster 1: TMA9_1A_patient06, TMA9_1A_patient07, TMA9_1A_patient09, TMA9_1A_patient17, TMA9_1A_patient21, TMA9_1A_patient22, TMA9_1A_patient24, TMA9_1B_patient10, TMA9_1B_patient12, TMA9_1B_patient22, TMA9_1B_patient23, TMA9_1B_patient24, TMA9_1B_patient31, TMA9_2A_patient02, TMA9_2A_patient03, TMA9_2A_patient04, TMA9_2A_patient39, TMA9_2A_patient40, TMA9_2A_patient41, TMA9_2A_patient46, TMA9_2A_patient48, TMA9_2B_patient02, TMA9_2B_patient03, TMA9_2B_patient04, TMA9_2B_patient05, TMA9_2B_patient36, TMA9_2B_patient38, TMA9_2B_patient40, TMA9_2B_patient41, TMA9_2B_patient42, TMA9_2B_patient43, TMA9_2B_patient44, TMA9_2B_patient46, TMA9_2B_patient48

cluster 2: TMA9_1A_patient10, TMA9_1A_patient15, TMA9_1A_patient25, TMA9_1A_patient30, TMA9_1B_patient06, TMA9_1B_patient08, TMA9_1B_patient21, TMA9_2A_patient26, TMA9_2B_patient01, TMA9_2B_patient14, TMA9_2B_patient26

cluster 3: TMA9_1A_patient12, TMA9_1A_patient28, TMA9_1B_patient29, TMA9_2A_patient42, TMA9_2B_patient39

cluster 4: TMA9_1A_patient18, TMA9_1B_patient17, TMA9_1B_patient28, TMA9_2A_patient01, TMA9_2A_patient19, TMA9_2A_patient33, TMA9_2A_patient37, TMA9_2A_patient47, TMA9_2B_patient33, TMA9_2B_patient37, TMA9_2B_patient47

cluster 5: TMA9_1A_patient45, TMA9_2A_patient05

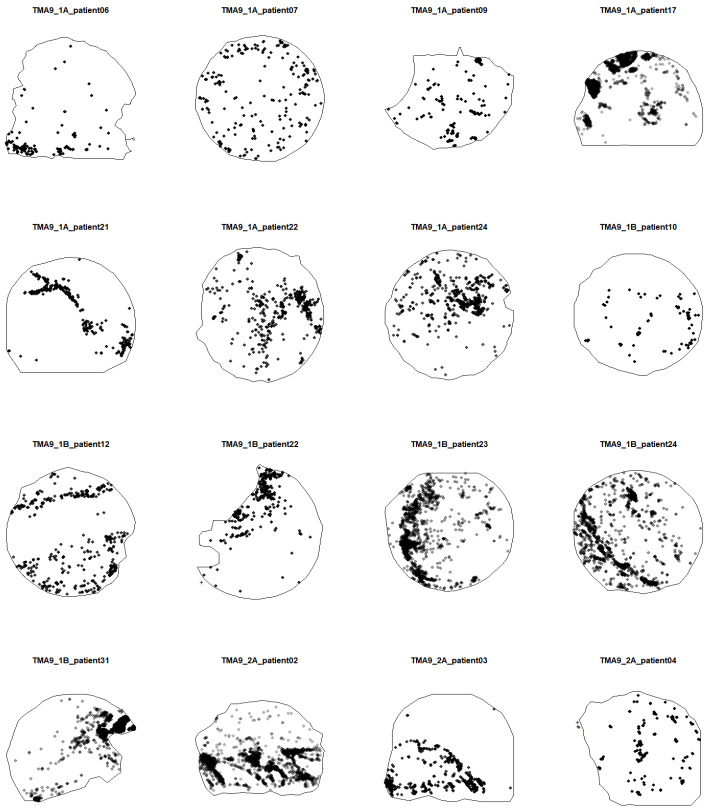
Appendix C

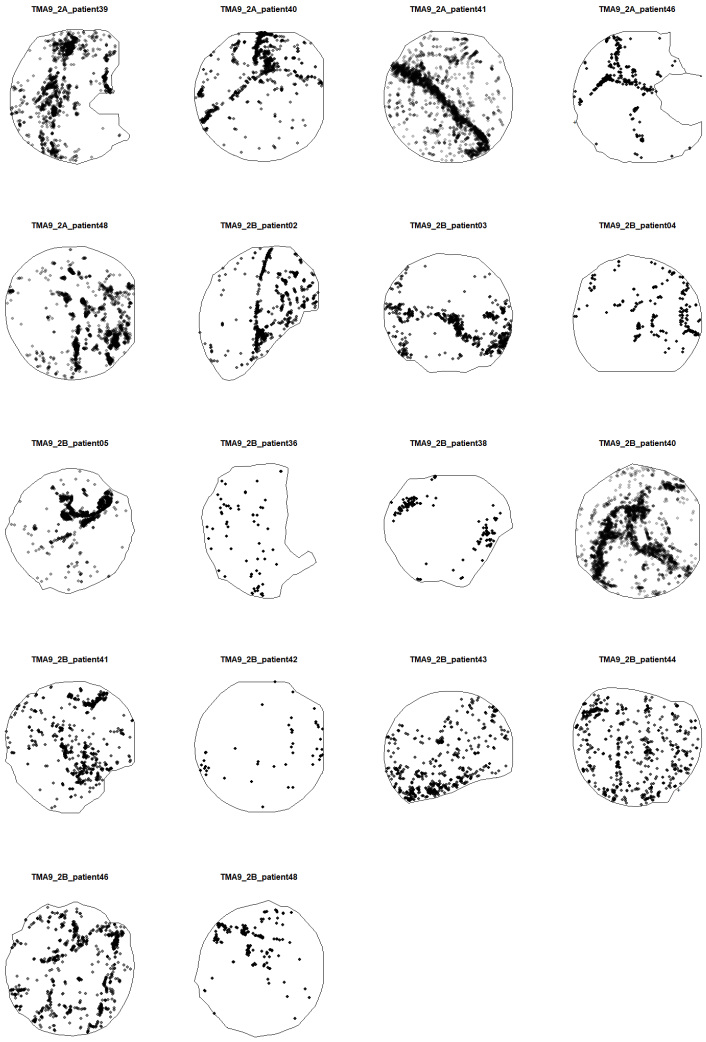
Statistics for κ and scale parameters by cluster

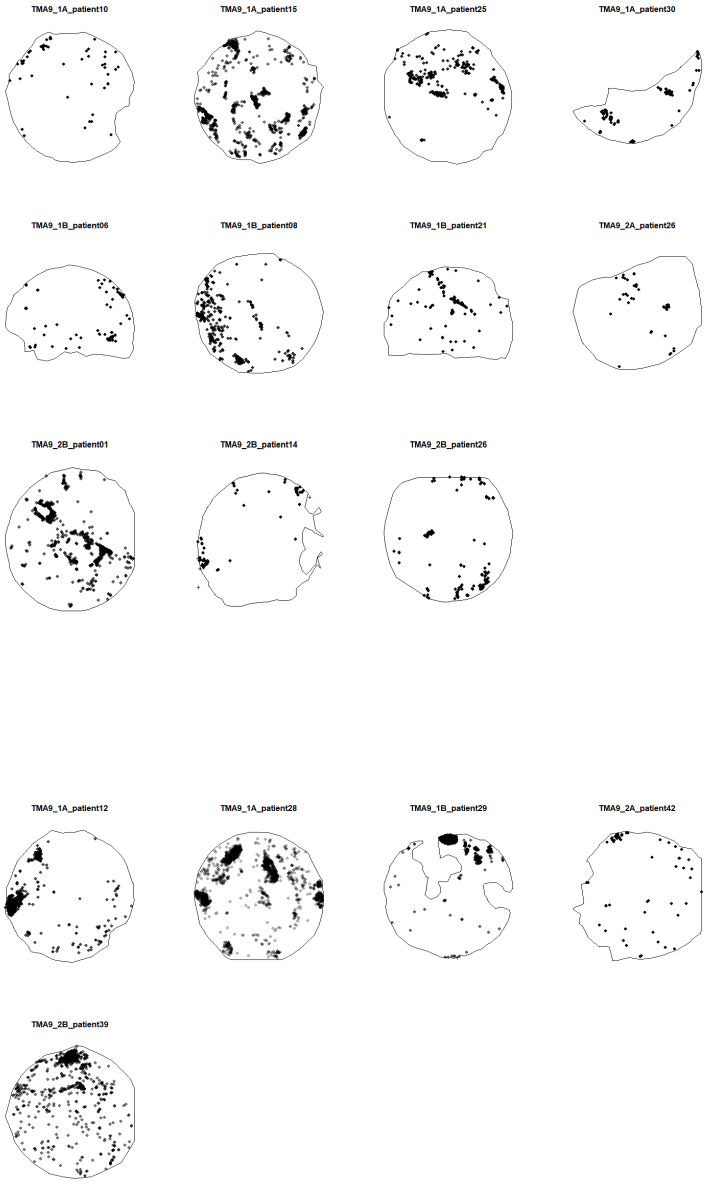
Cluster	n	κ ($\times 10^{-6}$)				Scale			
		Mean	SD	Min	Max	Mean	SD	Min	Max
1	34	3.46	1.42	1.16	6.38	93.30	32.40	32.90	145.00
2	11	2.37	0.91	1.21	4.09	57.30	13.60	40.40	80.30
3	5	0.80	0.20	0.62	1.06	123.00	16.30	98.70	140.00
4	11	5.26	0.43	4.32	5.68	201.00	57.50	96.30	278.00
5	2	0.56	0.00	0.56	0.57	67.10	17.80	54.60	79.70

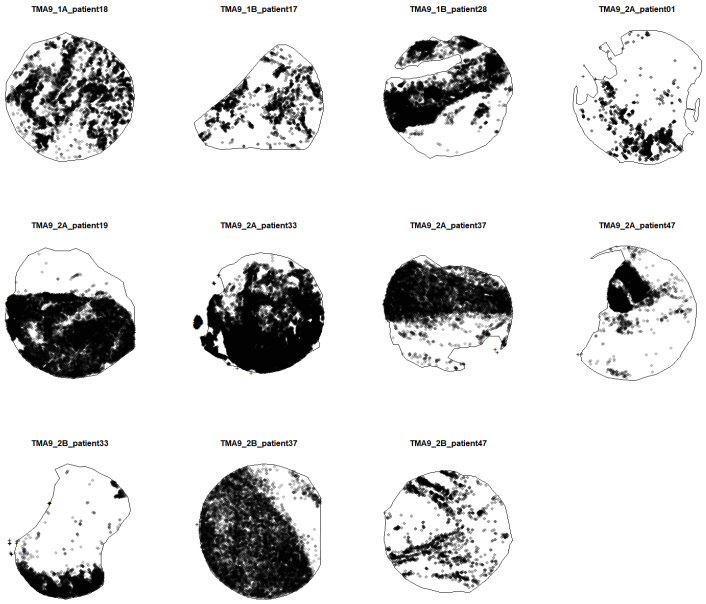
Appendix D

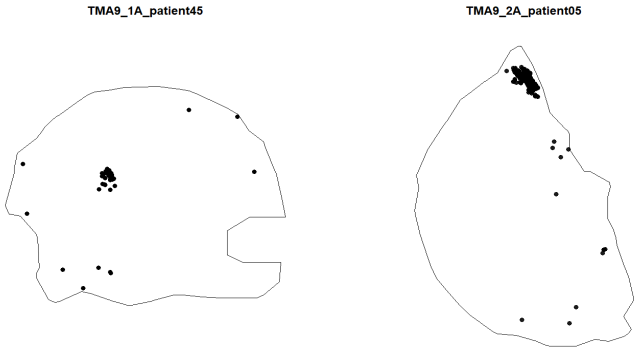
Data visualization




TMA9_2A_patient39
TMA9_2A_patient40
TMA9_2A_patient41
TMA9_2A_patient46
TMA9_2A_patient48
TMA9_2B_patient02
TMA9_2B_patient03
TMA9_2B_patient04
TMA9_2B_patient05
TMA9_2B_patient36
TMA9_2B_patient38
TMA9_2B_patient40
TMA9_2B_patient41
TMA9_2B_patient42
TMA9_2B_patient43
TMA9_2B_patient44
TMA9_2B_patient46
TMA9_2B_patient48

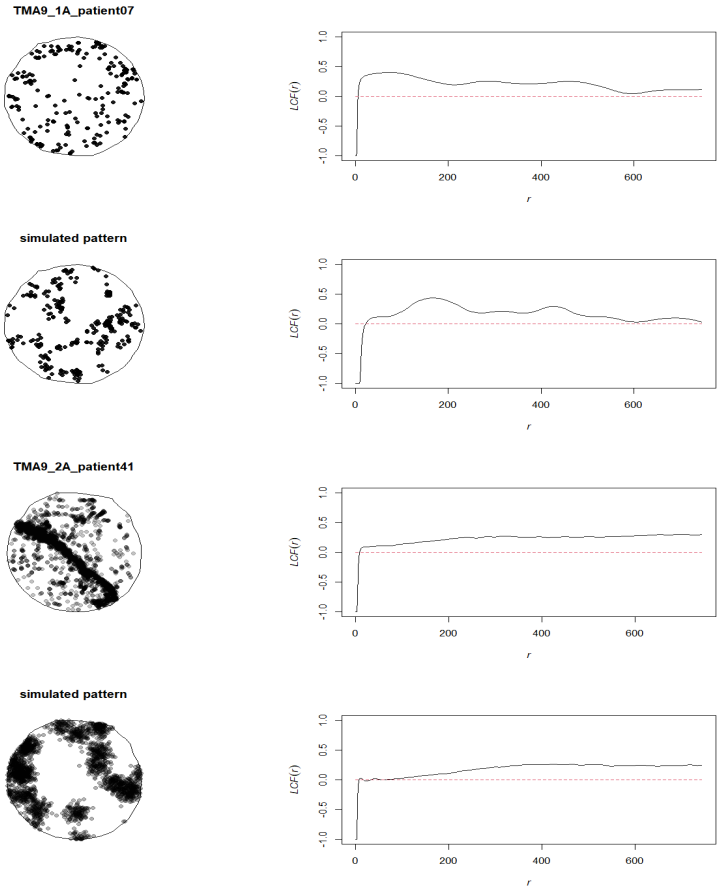

TMA9_1A_patient10
TMA9_1A_patient15
TMA9_1A_patient25
TMA9_1A_patient30
TMA9_1B_patient06
TMA9_1B_patient08
TMA9_1B_patient21
TMA9_2A_patient26
TMA9_2B_patient01
TMA9_2B_patient14
TMA9_2B_patient26
TMA9_1A_patient12
TMA9_1A_patient28
TMA9_1B_patient29
TMA9_2A_patient42
TMA9_2B_patient39


TMA9_1A_patient18
TMA9_1B_patient17
TMA9_1B_patient28
TMA9_2A_patient01
TMA9_2A_patient19
TMA9_2A_patient33
TMA9_2A_patient37
TMA9_2A_patient47
TMA9_2B_patient33
TMA9_2B_patient37
TMA9_2B_patient47

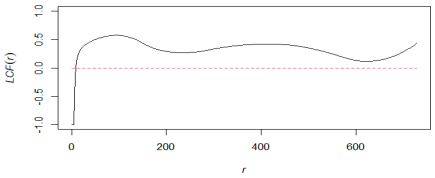

TMA9_1A_patient45
TMA9_2A_patient05

Appendix E

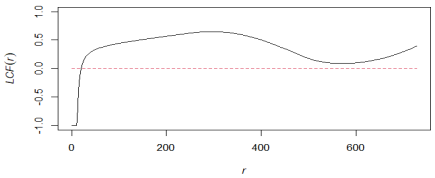
Observed vs. simulated patterns



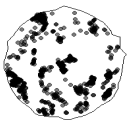
TMA9_1A_patient10

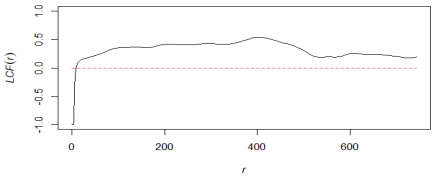


simulated pattern

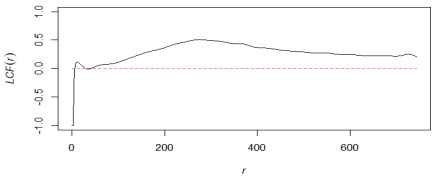


TMA9_1A_patient15

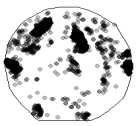


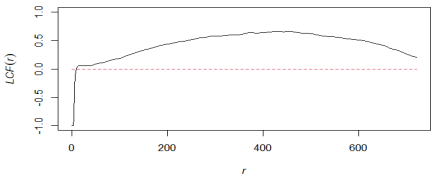


simulated pattern

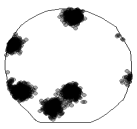


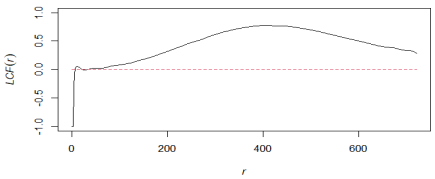
TMA9_1A_patient28



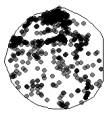


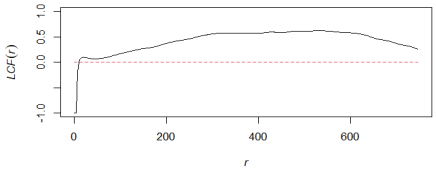
simulated pattern



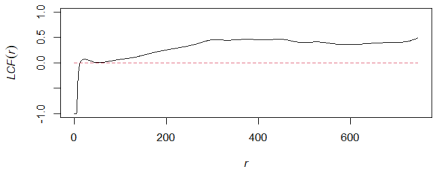


TMA9_2B_patient39

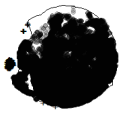


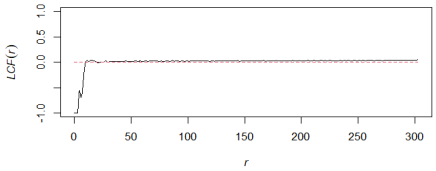


simulated pattern

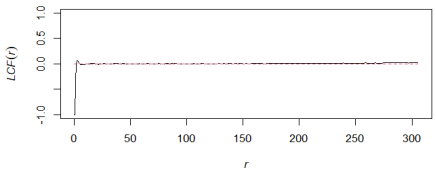


TMA9_2A_patient33

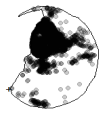


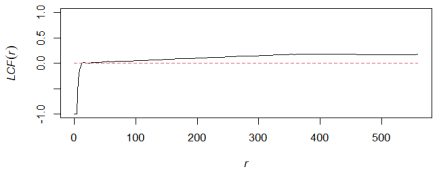


simulated pattern

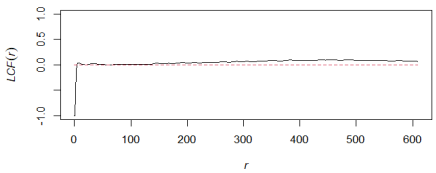


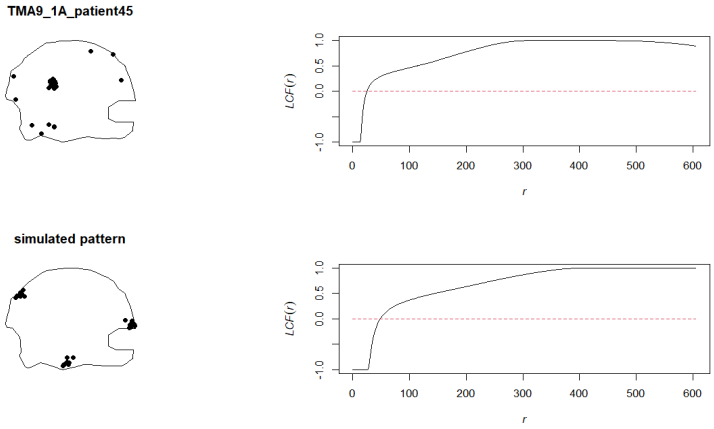
TMA9_2A_patient47





simulated pattern




TMA9_1A_patient45
simulated pattern
LCF(r)
r
0
100
200
300
400
500
600
-1.0
0.0
0.5
1.0