The Annals of Applied Statistics
2019, Vol. 13, No. 3, 1708–1732
https://doi.org/10.1214/19-AOAS1254

A BAYESIAN MARK INTERACTION MODEL FOR ANALYSIS OF TUMOR PATHOLOGY IMAGES[1]

BY QIWEI LI[*], XINLEI WANG[†], FAMING LIANG[‡] AND GUANGHUA XIAO[§,2]

University of Texas at Dallas[*], *Southern Methodist University*[†], *Purdue University*[‡] *and University of Texas Southwestern Medical Center*[§]

With the advance of imaging technology, digital pathology imaging of tumor tissue slides is becoming a routine clinical procedure for cancer diagnosis. This process produces massive imaging data that capture histological details in high resolution. Recent developments in deep-learning methods have enabled us to identify and classify individual cells from digital pathology images at large scale. Reliable statistical approaches to model the spatial pattern of cells can provide new insight into tumor progression and shed light on the biological mechanisms of cancer. We consider the problem of modeling spatial correlations among three commonly seen cells observed in tumor pathology images. A novel geostatistical marking model with interpretable underlying parameters is proposed in a Bayesian framework. We use auxiliary variable MCMC algorithms to sample from the posterior distribution with an intractable normalizing constant. We demonstrate how this model-based analysis can lead to sharper inferences than ordinary exploratory analyses, by means of application to three benchmark datasets and a case study on the pathology images of 188 lung cancer patients. The case study shows that the spatial correlation between tumor and stromal cells predicts patient prognosis. This statistical methodology not only presents a new model for characterizing spatial correlations in a multitype spatial point pattern conditioning on the locations of the points, but also provides a new perspective for understanding the role of cell–cell interactions in cancer progression.

1. Introduction. Cancer is a complex disease characterized by uncontrolled tumor cell growth. The pathological examination of hematoxylin and eosin (H&E)-stained tissue slides forms the basis of cancer diagnosis. It has been reported that cell growth patterns are associated with the survival outcome (Amin et al. (2002), Barletta, Yeap and Chirieac (2010), Borczuk et al. (2009), Gleason and Mellinger (2002)) and treatment response (Tsao et al. (2015)) of cancer patients. In addition, the interactions between tumor cells and other types of cells (e.g., immune cells) play important roles in the progression and metastasis of cancer

Received April 2018; revised March 2019.

[1] Supported in part by NIH Grants R01GM115473, R01GM117597, R01CA152301, R01CA172211, R15GM113157, R15GM131390, P50CA70907 and P30CA142543, and the Cancer Prevention and Research Institute of Texas Grant RP120732.

[2] Corresponding author.

Key words and phrases. Multitype point pattern, spatial correlation, Markov random field, double Metropolis–Hastings.

(Gillies, Verduzco and Gatenby (2012), Hanahan and Weinberg (2011), Junttila and de Sauvage (2013), Mantovani et al. (2002), Merlo et al. (2006), Orimo et al. (2005), Polyak, Haviv and Campbell (2009)). Spatial variations among cell types and their association with patient prognosis have been previously reported in breast cancer (Mattfeldt et al. (2009)). Pathological examination of tissue slides requires a pathologist to match the observed image slides with his/her memory for certain patterns and features (such as tumor content, nuclei counts and tumor boundary). This process is laborious, tedious and subject to errors. More importantly, due to the limitations of the human brain in interpreting highly complex pathology images, it is hard for pathologists to systematically explore those subtle but important patterns, such as tumor cell distribution and interaction with the surrounding micro-environment. Pathological examination by the human eyes is insufficient to decipher the large amount of complex and comprehensive information harbored in the high resolution pathology images.

With the advance of imaging technology, H&E-stained pathology imaging is becoming a routine clinical procedure, which produces massive digital pathology images on a daily basis. Digital pathology image analyses have been proven to be valuable in clinical diagnosis and prognosis of various malignancies, including precancerous lesions in the esophagus (Sabo et al. (2006)), prostate cancer (Tabesh et al. (2007)), neuroblastoma (Sertel et al. (2009a)), lymphoma (Sertel et al. (2009b)), breast cancer (Beck et al. (2011), Yuan et al. (2012)) and recently, lung cancer (Luo et al. (2016), Yu et al. (2016)). However, current studies of pathology image analysis mainly focus on the morphology features, such as tissue texture and granularity. For example, Tabesh et al. (2007) aggregate color, texture and morphometric cues at the global and histological object levels to predict the malignancy level of prostate cancer. Both Yu et al. (2016) and Luo et al. (2016) used a large number of objective morphological features (e.g., cell size, shape, distribution of pixel intensity in the cells and nuclei, texture of the cells and nuclei, etc.) extracted by CellProfiler (Carpenter et al. (2006), Kamentsky et al. (2011)) to predict lung cancer prognosis. Yuan et al. (2012) even integrate histopathology and genomics information, extending approaches that only use morphological features to predict breast cancer patient survival. However, these imaging data, which capture histological details in high resolution, still leave unexplored more undiscovered knowledge. Computer vision and machine learning algorithms have enabled us to automatically identify individual cells from digital pathology images at large scale (e.g. Yuan et al. (2012)). Recent developments in deep-learning methods have greatly facilitated this process. We have developed a convolutional neural network (CNN) to identify individual cells and classify their types into three categories: lymphocyte, stromal and tumor.

Consequently, a pathology image is abstracted into a spatial map of marked points, where each cell belongs to one of the three distinct types. The analysis of pathology images thus becomes an investigation of those marked point pattern, which will provide a new perspective for the role of cell–cell interactions in cancer progression. Currently, a patient cohort usually contains hundreds of patients, and each patient has one or more pathology images. These rich datasets provide a great opportunity to study the cell–cell interactions in cancer. Recently, Li et al. (2017) developed a modified Potts model to study the spatial patterns observed in tumor pathology images, by projecting irregularly distributed cells into a square lattice. However, this approximate method relies on selection of an *ad hoc* lattice. More importantly, it models the interaction among different regions (small squares defined by the lattice), but not those among individual cells.

The study of interactions between objects, which results in the spatial correlation of marks, has been a primary focus in spatial statistics. It is a key aspect in population forestry (Stoyan and Penttinen (2000)) and ecology (Dale (2000)) theory, but receives little attention in biology. Illian et al. (2008) discussed in detail a large variety of numerical, functional and second-order summary characteristics, which can be used to describe the spatial dependency between different types of points in a planar region. The most common approaches are based on generalizing the standard distance-dependent G-, K-, J- and L-functions to their "cross-type" versions (see, e.g., Besag (1977), Diggle and Cox (1981), Lotwick and Silverman (1982), Ripley (1977), Vincent and Jeulin (1989), van Lieshout and Baddeley (1996, 1999)). Mark connection functions (MCFs) are another well recognized tool for qualitative marks, which are more suitable for the detection of mark correlation in an exploratory analysis (Wiegand and Moloney (2004)). The *ad hoc* testing of hypotheses, such as spatial independences of the marks, based on some suitable summary characteristics (e.g., K-functions) has also been discussed in the literature (Grabarnik, Myllymäki and Stoyan (2011)). However, model-based analysis, which may sharpen inferences about the spatial pattern, is lagging. Marked point process models are usually constructed by familiar devices, such as Cox, cluster and thinned processes (Gelfand et al. (2010)). For example, Diggle and Milne (1983) defined the class of bivariate Cox processes and investigated the structure of correlated bivariate Cox models. Gibbsian point processes are more appropriate to account for interaction between points than Cox models, which tend to reflect underlying environmental variation. Ogata and Tanemura (1985) and Diggle, Eglen and Troy (2006) formulated pairwise interaction models based on Gibbs distributions for bivariate marked point patterns and argued that model-based inference is statistically more efficient. Inference for those mark-dependent/independent pairwise interactions are mainly based on frequentist approaches, which can only provide for point estimation. Bayesian inference of marked point processes has been notably underrepresented in the literature (Bognar (2008)). For more detailed examples of modeling approaches for marked point processes, see Baddeley and Turner (2006).

In this paper, motivated by the emerging needs of tumor pathology images analysis, we develop a novel geostatistical marking model, which aims to study the mark formulation at a finite known set of points through a Bayesian framework.

A local energy function of three groups of parameters, that is, first- and second-order intensities, and an exponential decay rate to the inter-point distance, is carefully defined, so is the related Gibbs distribution. The proposed model can serve as a novel model-based approach to characterize the spatial pattern/correlation among marks. We use the double Metropolis–Hastings (DMH) algorithm (Liang (2010)) to sample from the posterior distribution with an intractable normalizing constant in the Gibbs distribution. The model performs well in simulated studies and three benchmark datasets. We also conduct a case study on a large cohort of lung cancer pathology images. The result shows that the spatial correlation between tumor and stromal cells is associated with patient prognosis (P-value $= 0.007$). Although the morphological features of stroma in tumor regions have been discovered to be associated with patient survival, there is no strong statistical evidence to support this, due to a lack of rigorous statistical methodology. In the study, the proposed statistical methodology not only delivers a new perspective for understanding how marks (i.e., cell types in pathology images) formulate, given an independent unmarked point process, but also provides a refined statistical tool to characterize spatial interactions, which the existing approaches (e.g., MCF) may lack sufficient power to do so.

The remainder of the paper is organized as follows: Section 2 introduces the proposed modeling framework, including the local energy function and its related Gibbs distribution (i.e., the model likelihood), the choices of priors and the model interpretation. Section 3 describes the Markov chain Monte Carlo (MCMC) algorithm and discusses the resulting posterior inference. Section 4 assesses performance of the proposed model on simulated data. Section 5 analyzes three benchmark datasets and a large cohort of lung cancer pathology images from the National Lung Screening Trial (NLST). Section 6 concludes the paper with some remarks on future research directions.

2. Model. We describe a spatial map of cells in a Cartesian coordinate system, with n observed cells indexed by i. We use $(x_i, y_i) \in \mathbb{R}^2$ to denote the x- and y-coordinates and $z_i \in \{1, \dots, Q\}$, $Q \geq 2$ to denote the type of cell i. In spatial point pattern analysis, such data are considered as multitype point pattern data, where $(x_1, y_1), \dots, (x_n, y_n)$ are the point locations in a compact subset of the 2-dimensional Euclidean space $\mathbb{R}^2$ (note that the proposed model can be easily extend to a general case of $\mathbb{R}^k, k \geq 3$) and $z_1, \dots, z_n$ are their associated qualitative (i.e., categorical or discrete) univariate marks. The mark attached to each point indicates which type/class it is (e.g., on/off, case/control, species, etc.). Without loss of generality, we assume that the data points are restricted within the unit square $[0, 1]^2$. This can be done by rescaling each pair of coordinates (x_i, y_i) to (x_i', y_i'). Suppose all the points are within a known rectangle, with four vertices' coordinates denoted by $(v_x^{\text{lwr}}, v_y^{\text{lwr}})$, $(v_x^{\text{upp}}, v_y^{\text{lwr}})$, $(v_x^{\text{upp}}, v_y^{\text{upp}})$ and $(v_x^{\text{lwr}}, v_y^{\text{upp}})$, then $x_i' = (x_i - v_x^{\text{lwr}})/L$ and $y_i' = (y_i - v_y^{\text{lwr}})/L$, where $L = \max(v_x^{\text{upp}} - v_x^{\text{lwr}}, v_y^{\text{upp}} - v_y^{\text{lwr}})$.

Since our primary interest lies in the spatial correlations of different types of cells, we only treat the marks as random variables, conditional on their fixed locations within a bounded observation window. This removes the need to infer the unmarked point process, simplifying the modeling construction and eliminating error due to the estimation of point density. Baddeley (2010) argued that it is often appropriate to analyze a marked point pattern by conditioning on the locations and under some reasonable assumptions, the marks effectively constitute a random field.

2.1. *Energy functions.* In the analysis of tumor pathology images, cell distribution and cell–cell interaction may reveal important messages about the tumor cell growth and its micro-environment. Therefore, it is of great interest to study the arrangements of cell types associated with the observed cells, given their locations. In spatial point pattern analysis, such a problem is called *geostatistical marking* (Illian et al. (2008)), which is to study the formulation of the marks $\boldsymbol{z}$ in a pattern, given the points $(\boldsymbol{x}, \boldsymbol{y})$. In this subsection, we explore the formulation of energy functions, accounting for both of the first- and second-order properties of the point data. The energy function (also known as the potential function) originates in statistical physics. It can be interpreted as the energy required to obtain a stable arrangement of $\boldsymbol{z}$, which is contributed by each z_i as well as each pair of (z_i, z_i').

At the initial stage, we assume that each point interacts with all other points in the space. A complete undirected graph $G = (V, E)$ can be used to depict their relationships, with V denoting the set of points (i.e., the n observed cells) and E denoting the set of direct interactions (i.e., the $(n-1)n/2$ cell–cell pairs). We define G as the interaction network and define its potential energy as

$$V(\boldsymbol{z}|\boldsymbol{\omega}, \boldsymbol{\Theta}) = \sum_q \omega_q \sum_i I(z_i = q) + \sum_q \sum_{q'} \theta_{qq'} \sum_{(i \sim i') \in E} I(z_i = q, z_{i'} = q'), \tag{2.1}$$

where the notation $(i \sim i')$ denotes that points i and i' are the interacting pair in G (i.e., they are connected by an edge in G), and I denotes the indicator function. Note that $\theta_{qq'} = \theta_{q'q}$ as the edge between any pairs of points has no orientation. On the right-hand side of equation (2.1), the first term can be viewed as the weighted average of the numbers of points with different marks, while the second term can be viewed as the weighted average of the numbers of pairs connecting two points with the same or different marks. In the context of spatial point pattern analysis, the first and second terms are referred to the first- and second-order potentials/characteristics, respectively. Their corresponding parameters $\boldsymbol{\omega} = (\omega_1, \dots, \omega_Q)$ and $\boldsymbol{\Theta} = [\theta_{qq'}]_{Q \times Q}$ are defined as the first- and second-order intensities. These two groups of parameters control the enrichment of different marks and the spatial correlations among them simultaneously. A detailed interpretation of $\boldsymbol{\omega}$ and $\boldsymbol{\Theta}$ is discussed in Section 2.4.

In mathematical physics and statistical thermodynamics, the interaction energy between two points (i.e., particles and cells) is usually an exponential decay function with respect to the distance between the two points (see, e.g., Avalos and Bucci (2014), Chulaevsky (2014), Kashima (2010), Penrose and Lebowitz (1974), Rincón, Ganahl and Vidal (2015)). Similarly, exponential decay has also been observed in biological systems, such as cell–cell interactions (Hui and Bhatia (2007), Segal and Stephany (1984)) and gene-gene correlations (Xiao, Reilly and Khodursky (2009), Xiao, Wang and Khodursky (2011)). In this study, we assume the interaction energy between a pair of points decreases exponentially at a rate λ proportional to the distance,

$$\begin{aligned}V(\boldsymbol{z}|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda) = {} & \sum_q \omega_q \sum_i I(z_i = q) \\ & + \sum_q \sum_{q'} \theta_{qq'} \sum_{(i \sim i') \in E} e^{-\lambda d_{ii'}} I(z_i = q, z_{i'} = q'),\end{aligned} \tag{2.2}$$

where $d_{ii'} = \sqrt{(x_i - x_{i'})^2 + (y_i - y_{i'})^2}$ is the Euclidean distance between points i and i'. A larger value of the decay parameter λ makes the interaction energy vanish much more rapidly with the distance, while a smaller value leads to $e^{-\lambda d_{ii'}} \approx 1$ and equation (2.2) → equation (2.1). See Figure S1 in the Supplementary Material (Li et al. (2019)) for examples of exponential decay functions with different values of parameter λ. The decay function makes our approach similar in spirit to the other literature on spatial prediction, which are built from a fairly simple concept: spatial correlation suggests that one should give more weight to observations near the prediction location than to those far away (Waller (2005)). In addition to the exponential decay, we can also consider the other decay forms under different scenarios and for different applications, such as the step decay $I(d \le \lambda), 0 < \lambda < 1$, the power-law decay $d^{-\lambda}, \lambda > 0$ and the power-exponential decay $e^{-\lambda d^{\beta}}, \lambda > 0, \beta > 0$. However, note that different choices of decay functions may result in different estimations on the second-order intensities.

As shown in equation (2.2), it needs to sum over n data points and $(n-1)n/2$ pairs of data points to compute the potential energy, resulting in a tedious computation, especially when n is large. An alternative way is to obtain an approximate value of $V(\boldsymbol{z}|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)$ by neglecting those pairs with distance beyond a certain threshold $c, c \in (0, 1)$. It can be illustrated that a point (i.e., a cell) can only interact with its nearby points within a certain range c. Therefore, the complete network G reduces to a sparse network $G' = (V, E')$, with $E' \subseteq E$ denoting the set of edges joining pairs of points i and i' in G', if their distance $d_{ii'}$ is smaller than a threshold c. We write the potential energy of the interaction network G' as

$$\begin{aligned}V(\boldsymbol{z}|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda) = {} & \sum_q \omega_q \sum_i I(z_i = q) \\ & + \sum_q \sum_{q'} \theta_{qq'} \sum_{(i \sim i') \in E'} e^{-\lambda d_{ii'}} I(z_i = q, z_{i'} = q').\end{aligned} \tag{2.3}$$

Note that c is not a model parameter, but a user-defined value. We may determine its value from a mark connection function analysis (discussed in Section 4) or from the subjective assessment of an experienced expert in the related field. The choice of a large c causes a considerably complex network, while a too small value results in a sparse network that may neglect some important spatial information. See Figure S2 in the Supplementary Material (Li et al. (2019)) for an example of three-type point pattern data ($n = 100$) and its corresponding mark interaction networks G' under different choices of c. By introducing the sparse network G', we not only reduce the computational cost, but also define a local spatial structure.

2.2. *Data likelihood.* According to the fundamental Hammersley–Clifford theorem (Hammersley and Clifford (1971)), if we have a locally defined energy, such as equation (2.3), then a probability measure with a Markov property exists. This frequently seen measure in many problems of probability theory and statistical mechanics is called a *Gibbs measure*, which gives the probability of observing marks associated with their locations in a particular state,

$$p(z|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda) = \frac{\exp(-V(z|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda))}{\sum_{z'} \exp(-V(z'|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda))}. \tag{2.4}$$

The normalizing constant $C(\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda) = \sum_{z'} \exp(-V(z'|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda))$ is also called a partition function. An exact evaluation of $C(\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)$ needs to sum over the entire space of z, which consists of Q^n states. Thus, it is intractable even for a small size model. Take $Q = 2$ and $n = 100$, for example, it needs to sum over $2^{100} \approx 1.268 \times 10^{30}$ elements. To address this issue, we employ the double Metropolis–Hastings (DMH) algorithm (Liang (2010)) to make inference on the model parameters $\boldsymbol{\omega}$, $\boldsymbol{\Theta}$ and λ. DMH is an auxiliary variable MCMC algorithm, which can make the normalizing constant ratio canceled by augmenting appropriate auxiliary variables through a short run of the ordinary Metropolis–Hastings (MH) algorithm. More details are given in Section 3.1.

Equation (2.4) serves as the full data likelihood. Since the model satisfies the local Markov property, we can also write the probability of observing point i belonging to class q conditional on its neighborhood configuration(s),

$$\begin{aligned} &p(z_i = q|z_{-i}, \boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda) \\ &\quad\propto \exp\Bigl(-\omega_q - \sum_{q'} \theta_{qq'} \sum_{\{i':(i\sim i')\in E'\}} e^{-\lambda d_{ii'}} I(z_{i'} = q')\Bigr), \end{aligned} \tag{2.5}$$

where z_{-i} denotes the collection of all marks excluding the ith one. According to equation (2.5), the conditional probability depends on the first-order intensity ω_q, the second-order intensities $\theta_{qq'}, q' = 1, \dots, Q$, the decay parameter λ, and the neighborhood of the points defined by c. Equation (2.5) is essentially a multinomial logistic regression and hence the parameters ω_q and $\theta_{qq'}$ can be interpreted in terms of conditional odds ratios.

2.3. *Parameter priors.* The proposed model in the Bayesian framework requires the specification of prior distributions for the unknown parameters. In this subsection, we specify the priors for all three groups of parameters: $\boldsymbol{\omega}$, $\boldsymbol{\Theta}$ and λ. For the first- and second-order intensities $\boldsymbol{\omega}$ and $\boldsymbol{\Theta}$, we notice that an identifiability problem arises from equation (2.4) or (2.5). For example, adding a nonzero constant, say s, into $\omega_q, q = 1, \dots, Q$ does not change the probability of observing point i belonging to class q. Similarly, the settings of $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta} + s\mathbb{1}$ lead to the same conditional probability, where $\mathbb{1}$ is a Q-by-Q matrix of ones. Therefore, imposing an appropriate constraint is necessary. Without loss of generality, suppose the points with mark Q have the largest population and we set $\omega_Q = 1$ and $\theta_{QQ} = 1$. For the other parameters in $\boldsymbol{\omega}$ and $\boldsymbol{\Theta}$, we consider normal priors and set $\omega_q \sim \mathrm{N}(\mu_\omega, \sigma_\omega^2), q = 1, \dots, Q-1$ and $\theta_{qq'} \sim \mathrm{N}(\mu_\theta, \sigma_\theta^2), q = 1, \dots, Q-1, q' = q, \dots, Q$. We suggest users choose the standard normal distribution; that is, $\mu_\omega = \mu_\theta = 0$ and $\sigma_\omega = \sigma_\theta = 1$. For the decay parameter λ, we specify a gamma prior $\lambda \sim \mathrm{Ga}(a_\lambda, b_\lambda)$. One standard way of setting a weakly informative gamma prior is to choose small values for the two parameters, such as $a_\lambda = b_\lambda = 0.001$. The $\mathrm{Ga}(a, b)$ prior is an attempt at noninformativeness within the conditionally conjugate family, with both a and b set to a low value, such as 0.1, 0.01 and 0.001 (Gelman et al. (2014)).

2.4. *Interpretation.* In this subsection, we aim to interpret the meanings of the model parameters $\boldsymbol{\omega}$ and $\boldsymbol{\Theta}$. In order to understand the relationship between the estimated parameter values and the observed multitype point pattern.

Suppose there is only one point in the space. Then equation (2.5) reduces to $p(z_1 = q|\cdot) \propto \exp(-\omega_q)$, which implies the probability of observing a point with mark q in this single-point system is equal to $\pi_q = \exp(-\omega_q)/\sum_q \exp(-\omega_q)$. Note that the vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_Q)$ has a natural constraint; that is, $\sum_q \pi_q = 1$. Furthermore, suppose there are n points in the space and there are almost no mark interactions. This can be fulfilled by any one of the following conditions: (1) the distance between any pairs of two points is beyond the given value c, that is, $d_{ii'} > c, \forall (i \sim i') \in E$; (2) the second-order intensities are all equal, that is, $\boldsymbol{\Theta} = s\mathbb{1}, \exists s \in \mathbb{R}$; or (3) the decay parameter λ goes to infinity, that is, $\lambda \to \infty$. Then equation (2.5) converges to $p(z_i = q|\mathbf{z}_{-i}, \boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda) \propto \exp(-\omega_q) = \pi_q$, implying that the expected number of points with mark q is $n\pi_q$. Thus, after transforming the first-order intensities $\boldsymbol{\omega}$ to their probability measures $\boldsymbol{\pi}$, we find a clear path to describe the abundance of different marks in the above simplified situations.

Suppose there are only two points 1 and 2 in the space, with the type of the second point known; say $z_2 = q'$. For convenience, we further assume $\omega_1 = \cdots = \omega_Q$. We first consider the case of the two points being at the same location, that is, $d_{12} = 0$. Then equation (2.5) turns out to be $p(z_1 = q|z_2 = q', \cdot) \propto \exp(-\theta_{qq'})$, which implies the probability of observing the point with unknown mark belonging to type q, given the one with the known mark q' (at the same location), is

$\phi_{qq'} = \exp(-\theta_{qq'}) / \sum_q \exp(-\theta_{qq'})$. We use a Q-by-Q matrix $\boldsymbol{\phi}$ to denote the collection of $\phi_{qq'}, q = 1, \dots, Q, q' = 1, \dots, Q$. Note that each column in $\boldsymbol{\phi}$ should be summed to 1 and $\boldsymbol{\phi}$ is not necessary to be a symmetric matrix as $\boldsymbol{\Theta}$. In this duo-point system (and more complex cases therein), the larger the value of $\phi_{qq'}$, the more likely the points with mark q get attracted to the nearby points with mark q'. Thus, the spatial correlations among marks can be easily interpreted by the probability matrix $\boldsymbol{\phi}$.

In the aforementioned duo-point model with known parameters, if the assumption of equivalent first-order intensities is relaxed, then the probability of assigning mark q to point 1 conditional on the mark of point 2 is q' is a strictly monotonic function of their distance d,

$$\text{MIF}_{q|q'}(d) = \frac{\exp(-\omega_q - \theta_{qq'} e^{-\lambda d})}{\sum_{q''} \exp(-\omega_{q''} - \theta_{q''q'} e^{-\lambda d})}. \tag{2.6}$$

We call the above equation the *mark interaction function* (MIF) of mark q given mark q'. As the distance increases, its value ultimately converges to π_q. The plot of MIF is a more comprehensive way to describe the spatial correlation/interaction between marks.

In conclusion, $\boldsymbol{\pi}$, $\boldsymbol{\phi}$ and MIF directly characterize a single point behavior (i.e., the assignment of its mark) in a model with small size, such as $n = 1$ and 2. However, the observed multitype point pattern is a reflection of how each individual point reacts with its neighbors. Note that the mappings from $\boldsymbol{\omega}$ to $\boldsymbol{\pi}$ and from $\boldsymbol{\Theta}$ to $\boldsymbol{\phi}$ are one-to-one/unique, so we can implement this step after obtaining the estimates of $\boldsymbol{\omega}$ and $\boldsymbol{\Theta}$.

3. Model fitting.

In this section, we describe the MCMC algorithm for posterior inference. Our inferential strategy allows for simultaneously estimating (1) the first-order intensities $\boldsymbol{\omega}$, which reveal the abundance of different marks; (2) the second-order intensities $\boldsymbol{\Theta}$, which capture the spatial correlation among marks; and (3) the decay parameter λ. We first give the full details of our MCMC algorithm and then discuss the resulting posterior inference.

3.1. *MCMC algorithm.*

We are interested in estimating $\boldsymbol{\omega}$, $\boldsymbol{\Theta}$ and λ, which define the Gibbs measure based on the local energy function. However, the data likelihood, as shown in equation (2.4), includes an intractable normalizing constant $C(\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)$, making the Metropolis–Hastings (MH) algorithm infeasible in practice. To address this issue, we use the double Metropolis–Hastings algorithm (DMH) proposed by Liang (2010). The DMH is an asymptotic algorithm, which has been shown to produce accurate results by various spatial models. Unlike other auxiliary variable MCMC algorithms (Møller et al. (2006), Murray, Ghahramani and MacKay (2012)) that also aim to have the normalizing constant ratio canceled, the DMH sampler is more efficient because: (1) it removes the need for exact sampling while it only needs to generate auxiliary variables through a short run of the

MH algorithm initialized with the original observation; and (2) it does not require drawing the auxiliary variables from a perfect sampler, which can be very expensive or impossible for many models with intractable normalizing constants. Liang et al. (2016) also proposed an adaptive exchange algorithm, which generates auxiliary variables via an importance sampling procedure from a Markov chain running in parallel. However, this exact algorithm is computationally more intensive than the DMH. See the Appendix for details of our MCMC algorithm.

3.2. *Posterior estimation.* We obtain posterior inference by summarizing the MCMC samples after burn-in. Suppose that we obtain sequences of MCMC samples for the model parameters, that is, $\omega_q^{(1)}, \dots, \omega_q^{(U)}$, $1 \leq q \leq Q$, $\theta_{qq'}^{(1)}, \dots, \theta_{qq'}^{(U)}$, $1 \leq q, q' \leq Q$ and $\lambda^{(1)}, \dots, \lambda^{(U)}$, where U is the total number of iteration after burn-in, then an approximate Bayesian estimator of each parameter can be simply obtained by averaging over the samples, $\hat{\omega}_q = \sum_{u=1}^{U} \omega_q^{(u)}/U$, $\hat{\theta}_{qq'} = \sum_{u=1}^{U} \theta_{qq'}^{(u)}/U$ and $\hat{\lambda} = \sum_{u=1}^{U} \lambda^{(u)}/U$, where u indexes the iteration. We suggest to project the parameters $(\boldsymbol{\omega}, \boldsymbol{\Theta})$ to $(\boldsymbol{\pi}, \boldsymbol{\Phi})$, or plot the MIFs as given in equation (2.6).

4. Simulation. In this section, we use simulated data generated from the proposed model to assess performance of our strategy for posterior inference on the model parameters. In addition, we discuss how to choose the tunable parameter c based on the MCF plots and investigate the sensitivity of the model to the choice of c.

We considered to generate the points by using two different point processes: (1) a homogeneous Poisson process (HPP) with a constant intensity $\eta = 2000$ over the space $[0, 1]^2$; and (2) a log Gaussian Cox process (LGCP) with an inhomogeneous intensity $\eta(x, y) = \exp(6 + |x - 0.3| + |y - 0.3| + \mathcal{GP}(x, y))$, $x \in [0, 1]$, $y \in [0, 1]$ and $\mathcal{GP}$ denotes a zero-mean Gaussian process with variance equal to 1 and scale equal to 1. The mark of each point, $z_i \in \{1, 2\}$, was simulated by using a Gibbs sampler based on equation (2.5). We ran 100,000 iterations with a random starting configuration of $\boldsymbol{z}$. The true parameters were set as follows: (1) the decay parameter $\lambda = 60$ or $\lambda = 0$, and the threshold $c = 0.05$, which implies that any pair of points with distance large than 0.05 were not considered in the model construction; (2) the first-order intensities $\boldsymbol{\omega} = (1, 1)$, which correspond to $\boldsymbol{\pi} = (0.5, 0.5)$; and (3) the second-order intensities $\boldsymbol{\Theta}$ were set according to each of the five scenarios, as shown in Table S1 in the Supplementary Material (Li et al. (2019)). They represented the cases of high/low repulsion, complete randomness and high/low attraction. *Attraction* or inter-mark attraction is defined as the clustering of points with different marks, while *repulsion* (also known as inhibition or suppression) is defined vice versa. We repeated the above steps to generate 30 independent datasets for each point process and each setting of λ and $\boldsymbol{\Theta}$. See Figure S3 for examples of simulated data generated by the HPP under settings of $\boldsymbol{\Theta}$ and $\lambda = 60$, and their mark connection functions (MCFs) and multitype K-functions. MCF is used to describe the spatial correlations of marks, where its quantity $\mathrm{MCF}_{qq'}(d)$ is interpreted as the empirical probability that two points at distance d have marks q and q'. An upward trend in $\mathrm{MCF}_{qq}(d)$ with a downward trend in $\mathrm{MCF}_{qq'}(d)$ indicates attraction, while the opposite case suggests repulsion. K-function is a standard tool for exploratory analysis of spatial point pattern data and (Lotwick and Silverman (1982)) defined its "cross" variant that is suitable in the bivariate case. Let ρ_q denote the expected number of points with mark q per unit area, then $\rho_q K_{q',q}(d)$ represents the expected number of additional points with mark q within distance d of an arbitrary point with mark q'. In the bivariate case, a similarity between $K_{11}(d)$ and $K_{22}(d)$ suggests that different types of points are generated by the same underlying process, while $K_{11}(d) = K_{22}(d) = 0$ reveals an inhibitory effect within each of the component patterns. For the rigorous definitions and detailed explanations of these two summary statistics, please refer to Baddeley (2010).

For the prior on ω_1, we used a normal distribution $\mathrm{N}(1, 1)$, corresponding that $\pi_1 \in [0.125, 0.878]$ with 95% probability a priori. For the priors on θ_{11} and θ_{12}, we used a standard normal distribution $\mathrm{N}(0, 1)$. Note that we set the constraints $\omega_2 = 1$ and $\theta_{22} = 1$ to avoid the identifiability problem. We set the hyperparameters that control the gamma prior on the exponential decay to $a_\lambda = b_\lambda = 0.001$, which leads to a vague prior with variance equal to 1000. We chose the tunable parameter $c = 0.05$. Results we report below were obtained by running the MCMC chain with 50,000 iterations, discarding the first 50% sweeps as burn in. We started the chain from a model by randomly drawing ω_1, θ_{11}, θ_{12} and λ from their priors and assigning a random configuration of z. All experiments were implemented in `R` with `Rcpp` package to accelerate computations on a Mac PC with 2.60 GHz CPU and 16 GB memory. In our implementation, the MCMC algorithm ran about half an hour for each dataset.

Table 1 summarizes the results of posterior inference on the model parameters, for simulated datasets from the HPP and $\lambda = 60$. For the results for the other scenarios, please see Table S2–S4 in the Supplementary Material (Li et al. (2019)). Each estimate was obtained by averaging over 30 independent datasets. Overall, the tables indicate that our model fitting strategy based on the DMH algorithm works well. However, we notice that the decay parameter λ was greatly overestimated in the complete randomness scenarios. This is not surprising because all z_i's are completely irrelevant to each other (i.e., $p(z_i = q|\cdot) \propto \exp(-\omega_q)$) under this scenario. Therefore, λ was ill-defined when $\lambda = 0$. The observed large values of λ also indicate the weights, associated with the second-order intensities, decrease faster and thus explain why each mark was dominated by the first-order intensities. We also found that high repulsion scenarios had the worst performance on θ_{12}, which measures the interaction strength between different types of points. The reason is that we can only observe a small number of the interacting pairs between type 1 and 2 points. Take Figure S3(a) for example, such interacting pairs can be only seen near the border between the two clumps. Therefore, a biased estimation on θ_{12} is expected.

TABLE 1

Simulated datasets from the homogeneous Poisson process with $\lambda = 60$: *Results of posterior inference on the model parameters. Values are averaged over* 30 *simulated datasets for each scenario, with standard deviations indicated in parentheses*

	High repulsion	Low repulsion	Complete randomness	Low attraction	High attraction
ω_1	1.0	1.0	1.0	1.0	1.0
$\hat{\omega}_1$	1.30 (0.39)	1.05 (0.12)	1.04 (0.09)	1.08 (0.18)	1.05 (0.19)
θ_{11}	1.0	1.0	1.0	1.0	1.0
$\hat{\theta}_{11}$	0.71 (0.37)	0.97 (0.09)	0.84 (0.33)	0.94 (0.15)	0.95 (0.14)
θ_{12}	3.2	1.9	1.0	0.2	−1.2
$\hat{\theta}_{12}$	2.52 (0.26)	1.82 (0.17)	0.81 (0.30)	0.05 (0.19)	−1.14 (0.20)
λ	60	60	60	60	60
$\hat{\lambda}$	48.36 (6.79)	58.77 (7.49)	186.76 (118.09)	65.58 (11.82)	58.75 (4.81)

The proposed model contains a tunable parameter c, which defines the neighborhood for each point. A large value of c quadratically increases the computational cost, while a small value may cause biased estimates. We suggest users choose a value of 0.1 or less unless there is strong evidence in support of a larger value. Such evidence could be either subjective, such as an assessment from an experienced expert, or objective, such as MCF plots from the data. Take Figure S3 in the Supplementary Material (Li et al. (2019)), for example. For attraction scenarios, the MCF curve converged right after d passing over the true value of c. However, for repulsion scenarios, the curve tends to have a much bigger lag, especially for larger values of ϕ_{12} or ϕ_{21}. We also conducted a sensitivity analysis. We fit each of the 120 simulated datasets generated from the HPP (30 for each scenario, excluding the complete randomness) into the proposed model with $c = 0.03$, 0.05 and 0.1. Figure S4 show the boxplots of the three estimates $\hat{\omega}_1$, $\hat{\theta}_{11}$ and $\hat{\theta}_{12}$ under different values of c for each scenario. The model was quite robust to different choices of c.

5. Application. In this section, we first investigate the performance of our methodology using three benchmark datasets. Then we apply the model to a large cohort of lung cancer pathology images, and the result reveals novel potential imaging biomarkers for lung cancer prognosis.

5.1. `spatstat` *datasets.* `R` package `spatstat` is a major tool for spatial point pattern analyses. One of the basic data types offered by it is multitype point pattern data. We used two retinal cell datasets with marks on/off and one wood dataset with six species to quantify their attraction/repulsion characteristics using the proposed model.

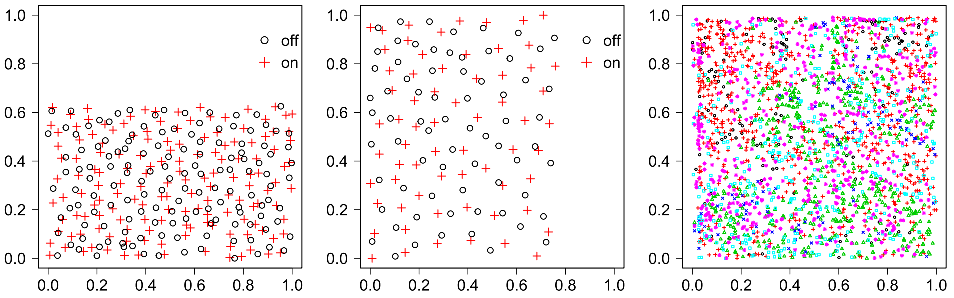

FIG. 1. *The plots of the rescaled marked points of* [*left*]: `amacrine` *with a unit standing for approximate* 1000 μm, *including* 142 *"light-off" cells* $(\circ)$ *and* 152 *"light-on" cells* $(+)$; [*middle*]: `betacells` *with a unit standing for approximate* 1000 μm, *including* 70 *"light-off" cells* $(\circ)$ *and* 65 *"light-on" cells* $(+)$; [*right*]: `lansing` *with a unit standing for approximate* 282 *m* ($\approx$ 924 *ft*), *including* 135 *black oaks* $(\circ)$, 703 *hickories* $(+)$, 514 *maples* $(\triangle)$, 105 *miscellaneous trees* $(\times)$, 346 *red oaks* $(\square)$ *and* 448 *white oaks* $(*)$.

Since 1970s, there has been considerable interest in studying the spatial pattern presented by particular types of mammalian retinal cell bodies (Wässle, Peichl and Boycott (1981), Wässle and Riemann (1978), Hughes (1981a, 1981b), Peichl and Wässle (1981), Rockhill, Euler and Masland (2000), Vaney, Peichl and Boycott (1981)). One of the two commonly used examples is the `amacrine` cells dataset (Diggle (1986)), consisting of two types of displaced amacrine cells within the retinal ganglion cell layer of a rabbit. The other is the `betacells` dataset (Wässle and Illing (1981)), composed of two types of beta cells that are associated with the resolution of fine details in the visual system of a cat. Figure 1 [left] and [middle] depict how the two different types of amacrine and beta cells distribute in restricted rectangular regions. Their MCF and multitype k-function plots are shown in Figure S5 in the supplemental material (Li et al. (2019)). Although the MCF plots clearly indicate strong attraction among cells with the different type and the interaction region radius around 0.1, no quantities can be accurately estimated further. We applied the proposed model with the same hyperparameter and algorithm settings as described in Section 4 and the choice of $c = 0.2$ for each dataset. We ran four independent MCMC chains and used the potential scale reduction factor (PSRF) (Gelman et al. (1992)) to evaluate convergence. PSRF is a statistic comparing the estimated between-chains and within-chain variances for each model parameter. Its value should be close to 1 if multiple chains have converged to the target posterior distribution. In this case, the PSRFs for all the model parameters were below 1.029, clearly suggesting that the MCMC chains converged. Then, for each dataset, we pooled together the outputs from the four chains and reported the estimated model parameters with their 95% credible interval in Figure S6 and S7. We plotted the estimated MIFs in Figure 2. Our method, as well as other methods (Diggle (1986), van Lieshout and Baddeley (1999)), suggest attraction between

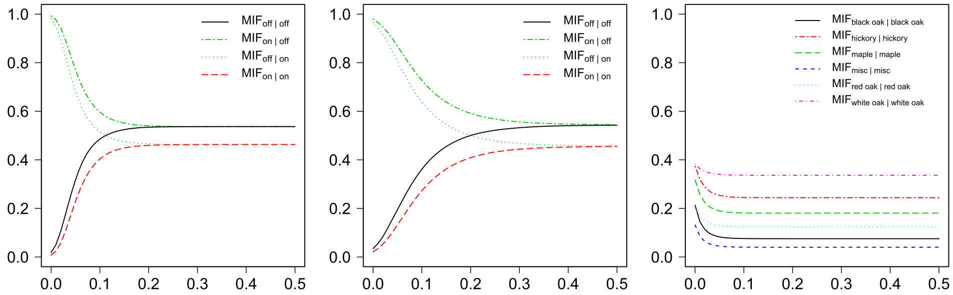

FIG. 2. *The MIF functions estimated from* [*left*]: `amacrine`; [*middle*]: `betacells`; [*right*]: `lansing` (*only the MIFs between the same mark are shown here*; *see Figure S*8 *for the numerical results of interactions between different marks*).

the cells (i.e., most cells have a nearest neighbor of the opposite type). The message about oppositely labeled pairs between neighbor cells would strengthen the assumption that there are two separate channels for brightness and darkness as postulated by Hering in 1874. Furthermore, our method is able to provide an accurate quantitative description, which can benefit the development and retinal sampling efficiency.

The third dataset in `spatstat` that was used is the `lansing` dataset. It contains the locations and botanical classification of trees in a 924 × 924 feet (19.6 acre) area of Lansing Woods, Clinton County, MI, USA. Figure 1 [right] shows the rescaled multitype point pattern that consists of $Q = 6$ types of trees. The MCF plots are shown in Figure S5 in the Supplementary Material (Li et al. (2019)), which indicate exhibition of clustering among the trees with the same type. With the same hyperparameter, algorithm and convergence diagnostic settings, we applied the proposed model with the choice of $c = 0.1$. The PSRFs were ranging from 1.003 to 1.022. The estimated model parameters are reported in Figure S8 and the MIFs are summarized in Figure 2 [right]. The pattern reveals that the first five types of trees exhibited clustering, especially for black oak and miscellaneous trees. This means if one species had a clump in an area, then no other species tended to form a clump there. We also found white oak had the least $\hat{\phi}_{qq}$ value, which suggests its spatial pattern was more likely random. Those findings were also reported in Cox and Lewis (1976) and Cox (1979). In addition, our MIF plots indicates that there was no interaction between the same type trees beyond 90 feet.

5.2. *Case study on lung cancer.* Lung cancer is the leading cause of death from cancer in both men and women. Non-small-cell lung cancer (NSCLC) accounts for about 85% of deaths from lung cancer. Current guidelines for diagnosing and treating NSCLC are largely based on pathological examination of H&E-stained tumor tissue section slides. We have developed a ConvPath pipeline (https://qbrc.swmed.edu/projects/cnn/) to determine the locations and types of

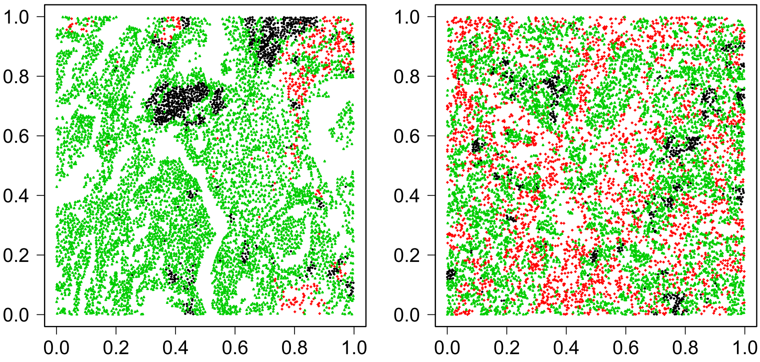

FIG. 3. *Lung cancer case study*: *Two examples of the rescaled marked point data from NLST dataset*, *where black*, *red and green points represent lymphocyte* (○), *stromal* (+) *and tumor* (△) *cells*. *For the data shown in the left*, $\hat{\lambda} = 172.102$, $\hat{\pi}_{\text{lym}} = 0.022$, $\hat{\pi}_{\text{str}} = 0.173$, $\hat{\pi}_{\text{tum}} = 0.805$ *and* $\hat{\phi}_{\text{tum,str}} = 0.012$; *For the data shown in the right*, $\hat{\lambda} = 169.268$, $\hat{\pi}_{\text{lym}} = 0.011$, $\hat{\pi}_{\text{str}} = 0.603$, $\hat{\pi}_{\text{tum}} = 0.386$ *and* $\hat{\phi}_{\text{tum,str}} = 0.162$.

cells observed in the processed tumor pathology images. Specifically, the classifier, based on a convolutional neural network (CNN), was trained using a large cohort of lung cancer pathology images manually labeled by pathologists, and it can classify each cell by its $Q = 3$ category: lymphocyte (a type of immune cell), stromal, or tumor cell. The overall classification accuracy is 92.9% and 90.1% in training and independent testing datasets, respectively (Wang et al. (2018)).

In this case study, we used the pathology images from 188 NSCLC patients in the National Lung Screening Trial (NLST). Each patient has one or more tissue slide(s) scanned at 40× magnification. The median size of the slides is 24,244 × 19,261 pixels. A lung cancer pathologist first determined and labeled the region of interest (ROI) within the tumor region(s) from each tissue slide using an annotation tool, ImageScope (Leica Biosystem). ROIs are regions of the slides containing the majority of the malignant tissues and are representative of the whole slide image. Then we randomly chose five square regions, each of which is in a 5000 × 5000 pixel window, per ROI as the sample images. The total number of sample images that we collected was 1585. For each sample image, the ConvPath (illustrated in Figure S9 in the Supplementary Material (Li et al. (2019))) software was used to identify cells from the sample images and classify each cell into one of three types, so that a corresponding spatial map of cells was generated and used as the input of our model. The number of cells in each sample image ranges from $n = 2876$ to 26,463. Figure 3 shows the examples of two sample images and Figure S10 displays their MCF and multitype K-function plots.

We applied the proposed model with the same hyperparameter and algorithm settings as described in Section 4 and different choices of $c = 0.02$, 0.05 and 0.1. We then computed the pairwise Pearson correlation coefficients between the estimated model parameters under different choices of c. These correlations indicated

TABLE 2

Lung cancer case study: The p-values of the transformed model parameters (in percentile %) by fitting a Cox proportional hazards model with survival time (defined as the number of days from diagnosis to death for participants who died or last contact for all other participants) and vital status (death or alive) as responses, and model parameters and clinical variables as predictors. The overall p-value (Wald test) is 0.003

Predictor	Coefficient	exp (Coef.)	SE	P-value
$\hat{\phi}_{\text{str,lym}}$	0.147	1.158	0.052	0.106
$\hat{\phi}_{\text{tum,lym}}$	0.030	1.030	0.025	0.546
$\hat{\phi}_{\text{lym,str}}$	−0.009	0.991	0.008	0.543
$\hat{\phi}_{\text{tum,str}}$	0.096	1.100	0.016	**0.002**
$\hat{\phi}_{\text{lym,tum}}$	−0.002	0.998	0.008	0.896
$\hat{\phi}_{\text{str,tum}}$	−0.059	0.943	0.019	0.128
$\hat{\pi}_{\text{lym}}$	0.034	1.035	0.015	0.219
$\hat{\pi}_{\text{str}}$	−0.032	0.969	0.007	**0.019**
$\hat{\lambda}$	−0.006	0.994	0.003	0.382
Age	0.038	1.039	0.009	0.176
Female/male	−0.138	0.871	0.091	0.631
Smoking/nonsmoking	−0.001	0.999	0.089	0.997

substantial agreement between any pair of settings, with values ranging from 0.967 to 0.997. The estimated parameters (with their summary statistics summarized in Table S5) that we used for the following three downstream analyses were obtained under the most conservative choice of $c = 0.1$.

5.2.1. *Association study.* With the estimated parameters in each sample image, we conducted a downstream analysis to investigate their associations with the other measurements of interest. Specifically, a Cox proportional hazards model (Cox (1992)) was fitted to evaluate the association between the transformed model parameters $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\Phi}}$ (in percentile %), and patient survival outcomes, after adjusting for other clinical information, such as age, gender and tobacco history. Multiple sample images from the same patient were modeled as correlated observations in the Cox proportional hazards model to compute a robust variance for each coefficient. The overall p-value for the Cox model was 0.002 (Wald test), and the p-value and coefficient for each individual variable are summarized in Table 2. The results imply that a low interaction between stromal and tumor cells is associated with good prognosis in NSCLC patients (p-value = 0.007). Interestingly, Beck et al. (2011) also discovered that the morphological features of the stroma in the tumor region are associated with patient survival in a systematic analysis of breast cancer. Besides, the abundance of the stromal cells itself (p-value = 0.017) is also a prognostic factor, while the underlying biological mechanism is currently unknown. The positive coefficient of the predictor $\phi_{\text{tum,str}}$ implies that a higher value may reveal a higher risk of death. Indeed, we obtained $\hat{\phi}_{\text{tum,str}} = 0.012$ for the data shown in Figure 3 [left] and it was from a patient who was still alive over 2615 days after the surgery, while the estimated value of $\phi_{\text{tum,str}} = 0.162$ for the data shown in Figure 3 [right] and it was from a patient who died on the 1246th day after the surgery. These two images have distinctive patterns, as the former clearly shows the same type cells tend to clump in the same area, while the latter displays a case where stromal and tumor cells are thoroughly mixed together, indicating the spread of stromal cells into the tumor region. Although the high/low interaction between stromal and tumor cells can be easily seen by eyes in these two images, the patterns are much more subtle for many other images. Therefore, the proposed model can be used to predict the survival time when human visualization does not work.

By contrast, we fitted a similar Cox proportional hazards model by using the MCF features as predictors. Specifically, we first used $\text{MCF}_{\text{lym,lym}}(d)$, $\text{MCF}_{\text{lym,str}}(d)$, $\text{MCF}_{\text{lym,tum}}(d)$, $\text{MCF}_{\text{str,str}}(d)$ and $\text{MCF}_{\text{str,tum}}(d)$, where $d = 0.1$ for each sample image as covariates. The results are summarized in Table S6 in the Supplementary Material (Li et al. (2019)). As we can see, there was no significant predictor and the overall p-value for the Cox model was 0.60 (Wald test). Next, we tried to vary d from 0 to 0.1. Figure S12(a) shows the p-values of those statistics against d. We repeated this analysis with the features from multitype K-functions and reported the results in Table S7 and Figure S12(b). Again, we were unable to find any association between cell–cell interactions and clinical outcomes. The comparison demonstrates the advantage of modeling the pathology images via the proposed model over the traditional explanatory analysis for characterizing spatial correlation.

5.2.2. *Predictive performance by cross-validation.* Lastly, we used leave-one-out cross-validation to evaluate the above Cox proportional hazards model. Specifically, we trained the model by using $(N - 1)$ sample images and then predicted the risk score of the left-out sample. After repeating this step for each of the N sample images, we calculated the average risk score for each patient from all the associated sample images. Based on the average risk score, we divided the patients into two equally sized groups (i.e., low and high-risk). Their corresponding Kaplan–Meier survival curves are shown in Figure 4. The log-rank test shows that there is a significant difference (p-value < 0.0001) between the two curves.

5.2.3. *Unsupervised clustering analysis.* Furthermore, we performed a model-based clustering analysis on the features extracted by the model. First, each of the eight parameters $\hat{\phi}_{\text{str,lym}}$, $\hat{\phi}_{\text{tum,lym}}$, $\hat{\phi}_{\text{lym,str}}$, $\hat{\phi}_{\text{tum,str}}$, $\hat{\phi}_{\text{lym,tum}}$, $\hat{\phi}_{\text{str,tum}}$, $\hat{\pi}_{\text{lym}}$, $\hat{\pi}_{\text{str}}$ of multiple sample images from the same patient were averaged. We then used the multivariate Gaussian mixture model (Fraley and Raftery (2002)) to cluster patients using those parameter. To estimate the number of clusters that best represents

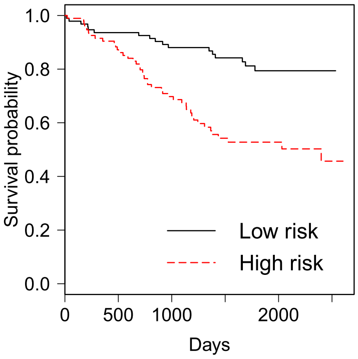

FIG. 4. *Lung cancer case study*: *The Kaplan–Meier plot for the low and high-risk groups obtained by leave-one-out cross-validation* (*log rank test p-value* < 0.0001).

the data as well as its covariance structure, we plotted the Bayesian information criterion (BIC) values against the number of clusters from 1 to 9, as shown in Figure 5 [left]. It shows that clustering patients into three groups achieves the best fit of the data measured by BIC, where the first, second and third groups have 79, 77 and 32 patients, respectively. Next, we visualized the means of these patient-level parameters for each group, shown as a radar chart in Figure 5 [middle], and plotted the Kaplan–Meier survival curve for each group in Figure 5 [right]. The patients from group 1 had higher survival probabilities, while the patients from the last group had the poor prognosis. The log-rank test shows that there are significant differences (p-value $= 0.015$) among the survival curves of the three groups. The analysis, again, demonstrates that the proposed mark interaction features can be used as a potential biomarker for patient prognosis.

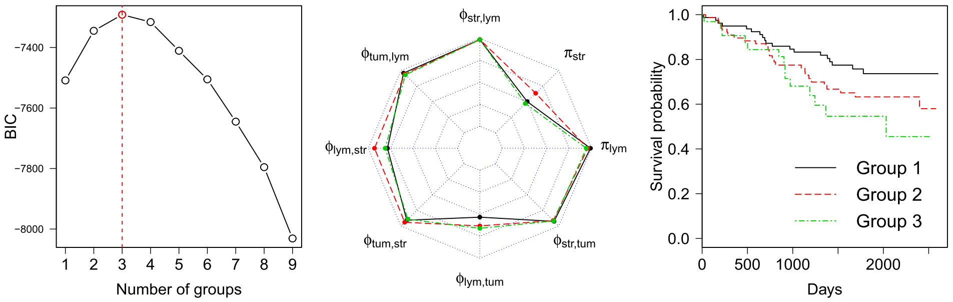

FIG. 5. *Lung cancer case study*: [*left*] *The BIC plot of the model-based clustering on the patient-level parameters*; [*middle*] *The radar chart of the averaged patient-level parameters of the three groups* (*shown in different colors*), *where the outer ring and the center have the values of* 0 *and* 1, *respectively*; [*right*] *The Kaplan–Meier plot for the three groups with patient survival* (*log rank test p-value* $= 0.015$).

6. Conclusion. The major cell types in a malignant tissue of lung are tumor cells, stromal cells and infiltrating lymphocytes. The distribution of different types of cells and their interactions play a key role in tumor progression and metastasis. For example, stromal cells are connective tissue cells, such as fibroblasts and pericytes, and their interaction with tumor cells is known to play a major role in cancer progression (Wiseman and Werb (2002)). Tumor-infiltrating lymphocytes have been associated with patient prognosis in multiple tumor types previously (Brambilla et al. (2016), Huh, Lee and Kim (2012)). Recent advances in deep learning methods have made possible the automatic identification and classification of cells at large scale. For example, the ConvPath pipeline could determine the location and cell type for thousands of cells. However, it is challenging to utilize the vast amount of information extracted digitally. In this study, we developed a Bayesian statistical method to model the spatial interaction among different types of cells in tumor regions. We focused on modeling the spatial correlation of marks in a spatial pattern that arose from a pathology image study. A Bayesian framework was proposed in order to model how the mark in a pattern might have been formed given the points. The proposed model can utilize the spatial information of thousands of points from any point processes. The output of the model is the parameters that characterize the spatial pattern. After a certain transformation, the parameters are identifiable and interpretable, and most importantly, transferable for conducting an association study with other measurements of interest. Furthermore, this statistical methodology provides new insights into the biological mechanisms of cancer.

For the lung cancer pathology imaging data, our study shows the interaction strength between stromal and tumor cells is associated with patient prognosis. This parameter can be easily measured using the proposed method and used as a potential biomarker for patient prognosis. This biomarker can be translated into real clinical tools at low cost because it is based only on tumor pathology slides, which are available in standard clinical care.

Several extensions of our model are worth investigating. First, the proposed model can be extended to finite mixture models for inhomogeneous mark interactions. Second, the correlation among first- and second-order intensity parameters could be taken into account by modeling them as a multivariate normal distribution. Third, in some scenarios, we may need to consider edge effects in that the marks associated with the points within the observation window may also interact with those marks of points outside the window. Therefore, some edge-correction methods, such as minus sampling, should be employed. Last but not least, the proposed model provides a good chance to investigate the performance of other approximate Bayesian computation methods, such as variational Bayes (Ren et al. (2011)). These could be future research directions.

APPENDIX: MCMC ALGORITHM

Update of $\boldsymbol{\omega}$: We update each of $\omega_q, q = 1, \ldots, Q - 1$ by using the DMH algorithm. We first propose a new ω_q^* from $\mathrm{N}(\omega_q, \tau_\omega^2)$. Next, according to equation (2.5), we implement the Gibbs sampler to simulate an auxiliary variable z^* starting from z based on the new $\boldsymbol{\omega}^*$, where all the elements are the same as $\boldsymbol{\omega}$ excluding the qth one. The proposed value ω_q^* is then accepted to replace the old value with probability $\min(1, r)$. The Hastings ratio r is given as

$$r = \frac{p(z^*|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)}{p(z|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)} \frac{p(z|\boldsymbol{\omega}^*, \boldsymbol{\Theta}, \lambda)}{p(z^*|\boldsymbol{\omega}^*, \boldsymbol{\Theta}, \lambda)} \frac{\mathrm{N}(\omega_q^*; \mu_\omega, \sigma_\omega^2)}{\mathrm{N}(\omega_q; \mu_\omega, \sigma_\omega^2)} \frac{J(\omega_q; \omega_q^*)}{J(\omega_q^*; \omega_q)},$$

where the form of $p(z|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)$ is given by equation (2.4). As a result, the normalizing constant in equation (2.4) can be canceled out. Note that the last fraction term, which is the proposal density ratio, equals 1 for this random walk Metropolis update on ω_q.

Update of $\boldsymbol{\Theta}$: We update each of $\theta_{qq'}, q = 1, \ldots, Q - 1, q' = q, \ldots, Q$ by using the DMH algorithm. We first propose a new $\theta_{qq'}^*$ from $\mathrm{N}(\theta_{qq'}, \tau_\theta^2)$ and set $\theta_{q'q}^* = \theta_{qq'}^*$ as the matrix is symmetric. Next, according to equation (2.5), an auxiliary variable z^* is simulated via the Gibbs sampler with z as the starting point. This simulation should be based on the new $\boldsymbol{\Theta}^*$, where all the elements are the same as $\boldsymbol{\Theta}$ except the two elements corresponding to $\theta_{qq'}$ and $\theta_{q'q}$. The proposed value $\theta_{qq'}^*$ as well as $\theta_{q'q}^*$ is then accepted to replace the old values with probability $\min(1, r)$. The Hastings ratio r is given as

$$r = \frac{p(z^*|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)}{p(z|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)} \frac{p(z|\boldsymbol{\omega}, \boldsymbol{\Theta}^*, \lambda)}{p(z^*|\boldsymbol{\omega}, \boldsymbol{\Theta}^*, \lambda)} \frac{\mathrm{N}(\theta_{qq'}^*; \mu_\theta, \sigma_\theta^2)}{\mathrm{N}(\theta_{qq'}; \mu_\theta, \sigma_\theta^2)} \frac{J(\theta_{qq'}; \theta_{qq'}^*)}{J(\theta_{qq'}^*; \theta_{qq'})},$$

where the form of $\Pr(z|\boldsymbol{\theta}, \lambda)$ is given by equation (2.4). As a result, the normalizing constant in equation (2.4) can be canceled out. Note that the last fraction term, which is the proposal density ratio, equals 1 for this random walk Metropolis update on $\theta_{qq'}$.

Update of λ: We update the decay parameter λ by using the DMH algorithm. We first propose a new λ^* from a gamma distribution $\mathrm{Ga}(\lambda^2/\tau_\lambda, \lambda/\tau_\lambda)$, where the mean is λ and the variance is τ_λ. Next, according to equation (2.5), we implement the Gibbs sampler to simulate an auxiliary variable z^* starting from z based on the new λ^*. The proposed value λ^* is then accepted to replace the old value with probability $\min(1, r)$. The Hastings ratio r is given as

$$r = \frac{p(z^*|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)}{p(z|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda)} \frac{p(z|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda^*)}{p(z^*|\boldsymbol{\omega}, \boldsymbol{\Theta}, \lambda^*)} \frac{\mathrm{Ga}(\lambda^*; a, b)}{\mathrm{Ga}(\lambda; a, b)} \frac{J(\lambda; \lambda^*)}{J(\lambda^*; \lambda)},$$

where the form of $\Pr(z|\boldsymbol{\theta}, \lambda)$ is given by equation (2.4). As a result, the normalizing constant in equation (2.4) can be canceled out. Note that the last fraction term, which is the proposal density ratio, equals 1 for this random walk Metropolis update on λ.

Acknowledgment. The authors would like to thank Jessie Norris for helping us in proofreading the manuscript.

SUPPLEMENTARY MATERIAL

Figures and tables (DOI: 10.1214/19-AOAS1254SUPPA; .pdf). We provide additional supporting plots and tables.

Code (DOI: 10.1214/19-AOAS1254SUPPB; .zip). We provide code in the form of R/C++ code. It can also be downloaded from GitHub (link: https://github.com/liqiwei2000/BayesMarkInteractionModel).

REFERENCES

AMIN, M. B., TAMBOLI, P., MERCHANT, S. H., ORDÓÑEZ, N. G., RO, J., AYALA, A. G. and RO, J. Y. (2002). Micropapillary component in lung adenocarcinoma: A distinctive histologic feature with possible prognostic significance. *Amer. J. Surg. Pathol.* **26** 358–364.

AVALOS, G. and BUCCI, F. (2014). Exponential decay properties of a mathematical model for a certain fluid-structure interaction. In *New Prospects in Direct, Inverse and Control Problems for Evolution Equations. Springer INdAM Ser.* **10** 49–78. Springer, Cham. MR3362986

BADDELEY, A. (2010). Multivariate and marked point processes. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. Diggle, P. Guttorp and M. Fuentes, eds.). *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 371–402. CRC Press, Boca Raton, FL. MR2730956

BADDELEY, A. and TURNER, R. (2006). Modelling spatial point patterns in *R*. In *Case Studies in Spatial Point Process Modeling. Lect. Notes Stat.* **185** 23–74. Springer, New York. MR2232122

BARLETTA, J. A., YEAP, B. Y. and CHIRIEAC, L. R. (2010). Prognostic significance of grading in lung adenocarcinoma. *Cancer* **116** 659–669.

BECK, A. H., SANGOI, A. R., LEUNG, S., MARINELLI, R. J., NIELSEN, T. O., VAN DE VIJVER, M. J., WEST, R. B., VAN DE RIJN, M. and KOLLER, D. (2011). Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Translational Med.* **3** 108–113.

BESAG, J. E. (1977). Comment on "Modelling spatial patterns." *J. Roy. Statist. Soc. Ser. B* **39** 193–195.

BOGNAR, M. A. (2008). Bayesian modeling of continuously marked spatial point patterns. *Comput. Statist.* **23** 361–379. MR2425167

BORCZUK, A. C., QIAN, F., KAZEROS, A., ELEAZAR, J., ASSAAD, A., SONETT, J. R., GINSBURG, M., GORENSTEIN, L. and POWELL, C. A. (2009). Invasive size is an independent predictor of survival in pulmonary adenocarcinoma. *Amer. J. Surg. Pathol.* **33** 462.

BRAMBILLA, E., LE TEUFF, G., MARGUET, S., LANTUEJOUL, S., DUNANT, A., GRAZIANO, S., PIRKER, R., DOUILLARD, J.-Y., LE CHEVALIER, T., FILIPITS, M. et al. (2016). Prognostic effect of tumor lymphocytic infiltration in resectable non-small-cell lung cancer. *J. Clin. Oncol.* **34** 1223–1230.

CARPENTER, A. E., JONES, T. R., LAMPRECHT, M. R., CLARKE, C., KANG, I. H., FRIMAN, O., GUERTIN, D. A., CHANG, J. H., LINDQUIST, R. A., MOFFAT, J. et al. (2006). CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7** R100.

CHULAEVSKY, V. (2014). Exponential decay of eigenfunctions in a continuous multi-particle Anderson model with sub-exponentially decaying interaction. Available at arXiv:1408.4646.

COX, T. F. (1979). A method for mapping the dense and sparse regions of a forest stand. *Appl. Statist.* **28** 14–19.

COX, D. R. (1992). Regression models and life-tables In *Breakthroughs in Statistics*, *Vol. II. Methodology and Distribution* (S. Kotz and N. L. Johnson, eds.) 527–541. Springer, Berlin.

COX, T. F. and LEWIS, T. (1976). A conditioned distance ratio method for analyzing spatial patterns. *Biometrika* **63** 483–491. MR0445713

DALE, M. R. (2000). *Spatial Pattern Analysis in Plant Ecology*. Cambridge Univ. Press, Cambridge.

DIGGLE, P. J. (1986). Displaced amacrine cells in the retina of a rabbit: Analysis of a bivariate spatial point pattern. *J. Neuroscience Methods* **18** 115–125.

DIGGLE, P. J. and COX, T. (1981). On sparse sampling methods and tests of independence for multivariate spatial point patterns. *Bulletin Internat. Statist. Inst.* **49** 213–229.

DIGGLE, P. J., EGLEN, S. J. and TROY, J. B. (2006). Modelling the bivariate spatial distribution of amacrine cells. In *Case Studies in Spatial Point Process Modeling*. *Lect. Notes Stat.* **185** 215–233. Springer, New York. MR2232131

DIGGLE, P. J. and MILNE, R. K. (1983). Bivariate Cox processes: Some models for bivariate spatial point patterns. *J. Roy. Statist. Soc. Ser. B* **45** 11–21. MR0701070

FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97** 611–631. MR1951635

GELFAND, A. E., DIGGLE, P., GUTTORP, P. and FUENTES, M. (2010). *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL.

GELMAN, A., RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3235677

GILLIES, R. J., VERDUZCO, D. and GATENBY, R. A. (2012). Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat. Rev. Cancer* **12** 487–493.

GLEASON, D. F. and MELLINGER, G. T. (2002). Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J. Urology* **167** 953–958.

GRABARNIK, P., MYLLYMÄKI, M. and STOYAN, D. (2011). Correct testing of mark independence for marked point patterns. *Ecol. Model.* **222** 3888–3894.

HAMMERSLEY, J. M. and CLIFFORD, P. (1971). Markov fields on finite graphs and lattices.

HANAHAN, D. and WEINBERG, R. A. (2011). Hallmarks of cancer: The next generation. *Cell* **144** 646–674.

HUGHES, A. (1981a). Cat retina and the sampling theorem: The relation of transient and sustained brisk-unit cut-off frequency to α and β-mode cell density. *Experimental Brain Res.* **42** 196–202.

HUGHES, A. (1981b). Population magnitudes and distribution of the major modal classes of cat retinal ganglion cell as estimated from HRP filling and a systematic survey of the soma diameter spectra for classical neurones. *J. Comparative Neurology* **197** 303–339.

HUH, J. W., LEE, J. H. and KIM, H. R. (2012). Prognostic significance of tumor-infiltrating lymphocytes for patients with colorectal cancer. *Archives of Surgery* **147** 366–372.

HUI, E. E. and BHATIA, S. N. (2007). Micromechanical control of cell–cell interactions. *Proc. Natn. Acad. Sci.* **104** 5722–5726.

ILLIAN, J., PENTTINEN, A., STOYAN, H. and STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. *Statistics in Practice*. Wiley, Chichester. MR2384630

JUNTTILA, M. R. and DE SAUVAGE, F. J. (2013). Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **501** 346–354.

KAMENTSKY, L., JONES, T. R., FRASER, A., BRAY, M.-A., LOGAN, D. J., MADDEN, K. L., LJOSA, V., RUEDEN, C., ELICEIRI, K. W. and CARPENTER, A. E. (2011). Improved structure, function and compatibility for CellProfiler: Modular high-throughput image analysis software. *Bioinformatics* **27** 1179–1180.

KASHIMA, Y. (2010). Exponential decay of correlation functions in many-electron systems. *J. Math. Phys.* **51** 063521, 40. MR2676498

LI, Q., YI, F., WANG, T., XIAO, G. and LIANG, F. (2017). Lung cancer pathological image analysis using a hidden Potts model. *Cancer Informatics* **16** 1176935117711910.

LI, Q., WANG, X., LIANG, F. and XIAO, G. (2019). Supplement to "A Bayesian mark interaction model for analysis of tumor pathology images." DOI:10.1214/19-AOAS1254SUPPA, DOI:10.1214/19-AOAS1254SUPPB.

LIANG, F. (2010). A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants. *J. Stat. Comput. Simul.* **80** 1007–1022. MR2742519

LIANG, F., JIN, I. H., SONG, Q. and LIU, J. S. (2016). An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants. *J. Amer. Statist. Assoc.* **111** 377–393. MR3494666

LOTWICK, H. W. and SILVERMAN, B. W. (1982). Methods for analysing spatial processes of several types of points. *J. Roy. Statist. Soc. Ser. B* **44** 406–413. MR0693241

LUO, X., ZANG, X., YANG, L., HUANG, J., LIANG, F., CANALES, J. R., WISTUBA, I. I., GAZDAR, A., XIE, Y. and XIAO, G. (2016). Comprehensive computational pathological image analysis predicts lung cancer prognosis. *J. Thorac. Oncol.* DOI:10.1016/j.jtho.2016.10.017.

MANTOVANI, A., SOZZANI, S., LOCATI, M., ALLAVENA, P. and SICA, A. (2002). Macrophage polarization: Tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends in Immunology* **23** 549–555.

MATTFELDT, T., ECKEL, S., FLEISCHER, F. and SCHMIDT, V. (2009). Statistical analysis of labelling patterns of mammary carcinoma cell nuclei on histological sections. *J. Microsc.* **235** 106–118. MR2731112

MERLO, L. M., PEPPER, J. W., REID, B. J. and MALEY, C. C. (2006). Cancer as an evolutionary and ecological process. *Nat. Rev. Cancer* **6** 924–935.

MØLLER, J., PETTITT, A. N., REEVES, R. and BERTHELSEN, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* **93** 451–458. MR2278096

MURRAY, I., GHAHRAMANI, Z. and MACKAY, D. (2012). MCMC for doubly-intractable distributions. Available at arXiv:1206.6848.

OGATA, Y. and TANEMURA, M. (1985). Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method. *Biometrics* **41** 421–433.

ORIMO, A., GUPTA, P. B., SGROI, D. C., ARENZANA-SEISDEDOS, F., DELAUNAY, T., NAEEM, R., CAREY, V. J., RICHARDSON, A. L. and WEINBERG, R. A. (2005). Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121** 335–348.

PEICHL, L. and WÄSSLE, H. (1981). Morphological identification of on-and off-centre brisk transient (Y) cells in the cat retina. *Proc. Roy. Soc. London Ser. B, Biolog. Sci.* **212** 139–153.

PENROSE, O. and LEBOWITZ, J. L. (1974). On the exponential decay of correlation functions. *Comm. Math. Phys.* **39** 165–184. MR0432092

POLYAK, K., HAVIV, I. and CAMPBELL, I. G. (2009). Co-evolution of tumor cells and their microenvironment. *Trends Genet.* **25** 30–38.

REN, Q., BANERJEE, S., FINLEY, A. O. and HODGES, J. S. (2011). Variational Bayesian methods for spatial data analysis. *Comput. Statist. Data Anal.* **55** 3197–3217. MR2825404

RINCÓN, J., GANAHL, M. and VIDAL, G. (2015). Lieb–Liniger model with exponentially decaying interactions: A continuous matrix product state study. *Phys. Rev. B* **92** 115107.

RIPLEY, B. D. (1977). Modelling spatial patterns. *J. Roy. Statist. Soc. Ser. B* **39** 172–212. MR0488279

ROCKHILL, R. L., EULER, T. and MASLAND, R. H. (2000). Spatial order within but not between types of retinal neurons. *Proc. Natn. Acad. Sci.* **97** 2303–2307.

SABO, E., BECK, A. H., MONTGOMERY, E. A., BHATTACHARYA, B., MEITNER, P., WANG, J. Y. and RESNICK, M. B. (2006). Computerized morphometry as an aid in determining the grade of dysplasia and progression to adenocarcinoma in Barrett's esophagus. *Lab. Investigation* **86** 1261.

SEGAL, D. M. and STEPHANY, D. A. (1984). The measurement of specific cell: Cell interactions by dual-parameter flow cytometry. *Cytometry A* **5** 169–181.

SERTEL, O., KONG, J., CATALYUREK, U. V., LOZANSKI, G., SALTZ, J. H. and GURCAN, M. N. (2009a). Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading. *J. Signal Processing Systems* **55** 169.

SERTEL, O., KONG, J., SHIMADA, H., CATALYUREK, U. V., SALTZ, J. H. and GURCAN, M. N. (2009b). Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. *Pattern Recognit.* **42** 1093–1103.

STOYAN, D. and PENTTINEN, A. (2000). Recent applications of point process methods in forestry statistics. *Statist. Sci.* **15** 61–78. MR1842237

TABESH, A., TEVEROVSKIY, M., PANG, H.-Y., KUMAR, V. P., VERBEL, D., KOTSIANTI, A. and SAIDI, O. (2007). Multifeature prostate cancer diagnosis and Gleason grading of histological images. *IEEE Trans. Med. Imag.* **26** 1366–1378.

TSAO, M.-S., MARGUET, S., LE TEUFF, G., LANTUEJOUL, S., SHEPHERD, F. A., SEYMOUR, L., KRATZKE, R., GRAZIANO, S. L., POPPER, H. H., ROSELL, R. et al. (2015). Subtype classification of lung adenocarcinoma predicts benefit from adjuvant chemotherapy in patients undergoing complete resection. *J. Clin. Oncol.* **33** 3439–3446.

VAN LIESHOUT, M. N. M. and BADDELEY, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Stat. Neerl.* **50** 344–361. MR1422574

VAN LIESHOUT, M. N. M. and BADDELEY, A. J. (1999). Indices of dependence between types in multivariate point patterns. *Scand. J. Stat.* **26** 511–532. MR1734259

VANEY, D. I., PEICHL, L. and BOYCOTT, B. (1981). Matching populations of amacrine cells in the inner nuclear and ganglion cell layers of the rabbit retina. *J. Comparative Neurology* **199** 373–391.

VINCENT, L. and JEULIN, D. (1989). Minimal paths and crack propagation simulations. *Acta Stereologica* **8** 487–494.

WALLER, L. A. (2005). Bayesian thinking in spatial statistics. In *Bayesian Thinking: Modeling and Computation. Handbook of Statist.* **25** 589–622. Elsevier/North-Holland, Amsterdam. MR2490540

WANG, S., WANG, T., YANG, L., YI, F., LUO, X., YANG, Y., GAZDAR, A., FUJIMOTO, J., WISTUBA, I. I., YAO, B. et al. (2018). ConvPath: A software tool for lung adenocarcinoma digital pathological image analysis aided by convolutional neural network. Available at arXiv:1809.10240.

WÄSSLE, H. and ILLING, R.-B. (1981). Morphology and mosaic of on-and off-beta cells in the cat retina and some functional considerations. *Proc. Roy. Soc. London Ser. B, Biolog. Sci.* **212** 177–195.

WÄSSLE, H., PEICHL, L. and BOYCOTT, B. B. (1981). Dendritic territories of cat retinal ganglion cells. *Nature* **292** 344–345.

WÄSSLE, H. and RIEMANN, H. (1978). The mosaic of nerve cells in the mammalian retina. *Proc. Roy. Soc. London Ser. B, Biolog. Sci.* **200** 441–461.

WIEGAND, T. and MOLONEY, K. A. (2004). Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* **104** 209–229.

WISEMAN, B. S. and WERB, Z. (2002). Stromal effects on mammary gland development and breast cancer. *Science* **296** 1046–1049.

XIAO, G., REILLY, C. and KHODURSKY, A. B. (2009). Improved detection of differentially expressed genes through incorporation of gene locations. *Biometrics* **65** 805–814. MR2649853

XIAO, G., WANG, X. and KHODURSKY, A. B. (2011). Modeling three-dimensional chromosome structures using gene expression data. *J. Amer. Statist. Assoc.* **106** 61–72. MR2816702

YU, K.-H., ZHANG, C., BERRY, G. J., ALTMAN, R. B., RÉ, C., RUBIN, D. L. and SNYDER, M. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **7**.

YUAN, Y., FAILMEZGER, H., RUEDA, O. M., ALI, H. R., GRÄF, S., CHIN, S.-F., SCHWARZ, R. F., CURTIS, C., DUNNING, M. J., BARDWELL, H. et al. (2012). Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. Translational Med.* **4** 157ra143.

Q. LI
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF TEXAS AT DALLAS
FOUNDERS BUILDING FO 2.604D
800 WEST CAMPBELL RD
RICHARDSON, TEXAS 75080
USA
E-MAIL: qiwei_li@utdallas.edu
URL: https://sites.google.com/site/liqiwei2000

X. WANG
DEPARTMENT OF STATISTICAL SCIENCE
SOUTHERN METHODIST UNIVERSITY
HEROY SCIENCE HALL 102
3225 DANIEL AVE
DALLAS, TEXAS 75275
USA
E-MAIL: swang@smu.edu
URL: http://faculty.smu.edu/swang/

F. LIANG
DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
MATH BUILDING 520
250 N. UNIVERSITY ST
WEST LAFAYETTE, INDIANA 47907
USA
E-MAIL: fmliang@purdue.edu
URL: http://www.stat.purdue.edu/~fmliang/

G. XIAO
DEPARTMENT OF POPULATION
AND DATA SCIENCES
UNIVERSITY OF TEXAS
SOUTHWESTERN MEDICAL CENTER
DANCIGER BUILDING H9.124
5323 HARRY HLINES BLVD
DALLAS, TEXAS 75390
USA
URL: https://sites.google.com/site/liqiwei2000
https://qbrc.swmed.edu/labs/xiaolab