A Comparison of Methods for Studying the Tumor Microenvironment's Spatial Heterogeneity in Digital Pathology Specimens

Ines Panicou Nearchou[1], Daniel Alexander Soutar[2], Hideki Ueno[3], David James Harrison[1,4], Ognjen Arandjelovic[2], Peter David Caie[1]

[1]School of Medicine, University of St Andrews, St Andrews, Scotland, UK, [2]School of Computer Science, University of St Andrews, St Andrews, UK, [3]Department of Surgery, National Defense Medical College, Tokorozawa, Saitama, Japan, [4]Lothian University Hospitals, Little France Crescent, Edinburgh, Scotland, UK

Abstract

Background: The tumor microenvironment is highly heterogeneous, and it is understood to affect tumor progression and patient outcome. A number of studies have reported the prognostic significance of tumor-infiltrating lymphocytes and tumor budding in colorectal cancer (CRC). However, the significance of the intratumoral heterogeneity present in the spatial distribution of these features within the tumor immune microenvironment (TIME) has not been previously reported. Evaluating this intratumoral heterogeneity may aid the understanding of the TIME's effect on patient prognosis as well as identify novel aggressive phenotypes which can be further investigated as potential targets for new treatment. **Methods:** In this study, we propose and apply two spatial statistical methodologies for the evaluation of the intratumor heterogeneity present in the distribution of $CD3^+$ and $CD8^+$ lymphocytes and tumor buds (TB) in 232 Stage II CRC cases. Getis-Ord hotspot analysis was applied to quantify the cold and hotspots, defined as regions with a significantly low or high number of each feature of interest, respectively. A novel spatial heatmap methodology for the quantification of the cold and hotspots of each feature of interest, which took into account both the interpatient heterogeneity and the intratumor heterogeneity, was further developed. **Results:** Resultant data from each analysis, characterizing the spatial intratumor heterogeneity of lymphocytes and TBs were used for the development of two new highly prognostic risk models. **Conclusions:** Our results highlight the value of applying spatial statistics for the assessment of the intratumor heterogeneity. Both Getis-Ord hotspot and our proposed spatial heatmap analysis are broadly applicable across other tissue types as well as other features of interest. **Availability:** The code underpinning this publication can be accessed at https://doi.org/10.17630/c2306fe9-66e2-4442-ad89-f986220053e2.

Keywords: Getis-Ord analysis, heatmap analysis, lymphocytic infiltration, spatial heterogeneity, tumor prognosis

INTRODUCTION

Cancer is heterogenous, with genetic and functional differences lying not only within different tumor types (intertumor heterogeneity) but also within single tumors (intratumor heterogeneity).[1] Although all types of tumor heterogeneity may contribute to the complexity of treatment approaches taken, intratumor heterogeneity is particularly problematic as it can aid the development of various mechanisms of drug resistance.[2] Intratumor heterogeneity is also closely related with cancer progression and clinical outcome.[3]

Colorectal cancer (CRC) is one of the most common cancers worldwide[4] and is highly heterogenous.[5] Currently, CRC is staged according to the gold standard of the tumor, node, metastasis staging system to stratify patients into prognostic subgroups.[6] However, the variation in survival outcome within the same stage patients demonstrates the need for additional prognostic factors.[7] The presence of tumor buds (TB), defined as single cancer cells or cancer clusters of up to 4 cells, has been recognized as an important adverse prognostic factor in

Access this article online

Quick Response Code:

Website:
www.jpathinformatics.org

DOI:
10.4103/jpi.jpi_26_20

CRC.[8] Lymphocytic infiltration is also a prognostic feature in CRC.[9,10] Previously, it was shown that the spatial relationship between these two features, TB and lymphocytes, confers great prognostic significance in Stage II CRC when analyzed at the entire invasive margin (IM).[11]

Whereas quantification of TB or tumor-infiltrating lymphocytes (TILs) at large regions of interest may represent well the processes occurring within the whole invasive front; it has the potential to miss any specific aggressive areas of the tumor that could promote invasion and metastasis. A number of studies have shown the role of spatial intratumor heterogeneity in the distribution of immune cells in breast cancer.[12,13] This spatial intratumor heterogeneity can be evaluated by applying the Getis-Ord hotspot analysis.[14] This reports the areas where features of interest are spatially clustered and therefore can identify hotspots within the images of interest. Through the application of such methodology, Nawaz *et al.* have found that areas with both high numbers of TILs and cancer cells were significantly correlated with better disease-specific survival (DSS) in estrogen receptor-negative breast cancer.[12] Another way to identify areas where a higher density of objects of interest occurs is through the use of spatial heatmaps.[15] However, their significance in evaluating intratumor heterogeneity remains unclear.

The aims of our study were to (i) apply the Getis-Ord hotspot analysis using data derived from automated image analysis on multiplexed immunofluorescence sections, (ii) develop a new methodology for studying the intratumor heterogeneity, and (iii) to assess how features derived from these methodologies may have an effect on predicting the survival outcome of Stage II CRC patients.

METHODS

Patient samples

The study population included 232 Stage II CRC patients who had undergone surgical resection, with detailed follow-up information of up to 11.5 years. Of these patients, 170 were treated at hospitals in Edinburgh, UK, from 2002 to 2004, and 62 at the National Defense Medical College Hospital, Japan, from 2006 to 2011. Clinicopathological characteristics are summarized in Table 1. This study was approved by the East of Scotland Research Ethics Service (approval ref: 13/ES/0126) and by the Ethics Committee of the National Defense Medical College (approval ref: No. 2992).

Data acquisition

Tissue labeling and segmentation of all features of interest through the use of automated image analysis have previously been described in detail.[11]

In summary, multiplexed immunofluorescence was performed on 3 μm thick formalin-fixed paraffin-embedded tissues using a DAKO Link48 Autostainer (Dako, Glostrup, DK). Primary antibodies against pan-T (CD3), cytotoxic T (CD8), and epithelial (Pan-cytokeratin, PCK) cells were

Table 1: Univariate Cox regression analysis for clinicopathological data, Getis-Ord-based model, and heatmap-based prognostic model

Features	Frequency (%)	HR (95%CI)	*P*
Clinicopathological			
Age			
≤70	110 (47.4)	1.464 (0.999-2.147)	0.051
71-79	64 (27.6)		
≥80	58 (25.0)		
Gender			
Male	133 (57.3)	0.670 (0.344-1.304)	0.239
Female	99 (42.7)		
pT stage			
pT3	190 (81.9)	**3.268 (1.695-6.301)**	**<0.001**
pT4	42 (18.1)		
Tumor site			
Left	69 (29.7)	1.102 (0.729-1.665)	0.646
Right	86 (37.1)		
Rectal	77 (33.2)		
Differentiation			
Moderate	125 (53.9)	1.105 (0.783-1.558)	0.570
Poor	35 (15.1)		
Well	71 (30.6)		
N/A	1 (0.4)		
EMLVI			
Yes	21 (9.1)	0.812 (0.580-1.137)	0.226
No	130 (56.0)		
N/A	81 (34.9)		
Tumor type			
Adenocarcinoma	206 (88.8)	1.020 (0.490-2.124)	0.958
Mucinous	18 (7.8)		
Mixed	7 (3.0)		
N/A	1 (0.4)		
Prognostic models			
Getis-Ord based			
Low-risk	137 (59.1)	**2.918 (1.909-4.461)**	**<0.001**
Mid-risk	80 (34.5)		
High-risk	15 (6.5)		
Spatial heatmap based			
Low-risk	130 (56.0)	**4.922 (3.041-7.966)**	**<0.001**
Mid-risk	92 (39.7)		
High-risk	10 (4.3)		

Significant features (*P*<0.05) are shown in bold. CI: Confidence interval, HR: Hazard ratio, EMLVI: Extramural lymphovascular invasion

used. Visualization of antibodies was performed using TSA fluorescein, TSA Cy5, and Alexa Fluor 555, respectively. Cell nuclei were counterstained using Hoechst 33342. Whole slide fluorescence images were captured with a ×20 objective using an Axioscan. Z1 (Zeiss, Oberkochen, DE). The captured images were uploaded into HALO® image analysis software (version 2.1.1637.18) (Indica Labs, Corrales, NM). The High-Plex FL (version 2.0) algorithm was applied to automatically classify CD3$^+$ and CD8$^+$ cells across the whole slide images [Figure 1a]. The invasive front was identified using the flood tool. Two Random Forest classifiers (one for high PCK intensity and one for low PCK intensity images)

were trained to segment tumor from nontumor areas. The High-Plex FL (version 2.0) algorithm in combination with the specific classifier (either for the high PCK or the low PCK intensity images) was then used for the segmentation and quantification of TB. Specifically, TB were classed as tumor clusters containing up to 4 PCK + cells. The quantification of TB was performed within a region of 1000 μm from the invasive front.

Data analysis

All object data (CD3+, CD8+, and TB) were exported in .csv format from HALO® software. Object data and their spatial coordinates were inputted to Python[16] in the Jupyter Notebook environment,[17] for the Getis-Ord hotspot and our spatial heatmap analyses. The Cartesian points defining the invasive front of each sample were also imported. These data were manipulated throughout using NumPy.[18] PySAL library[19] was used for the Getis-Ord hotspot analysis. Visualizations were performed using the Matplotlib library.[20]

From the Cartesian points defining the invasive front, we created two regions of interest, being the IM and the tumor core (CT) [Figure 1a]. We classified objects as being within the IM if the distance between the object's centroid and the tumor invasive front was ≤500 μm. Any object within the classified and segmented tumor area, but not in the IM, was considered in the CT. The union of these two regions was also considered as a distinct region (IMCT). Objects that were not located within the IM or CT areas were discarded. We considered lymphocytes and TB in isolation, as well as some measures of their juxtaposition. For the latter case, we evaluated the lymphocyte ratio (CD3+/CD8+) as well as evaluated all grids of lymphocytes within some Euclidean distance d of any TB. We considered d at 50 μm and 100 μm.

This led to 25 features of interest which are shown in Supplementary Table 1. Spatial plots for each feature of interest were generated. The sets of segregated objects were divided into grids using a fixed tile length t of 886 μm, giving an area of 0.785 mm² per tile following recommendations from the International Tumor Budding Consensus Conference.[7] This was our setup for both the Getis-Ord hotspot and our heatmap analyses [Figure 1b].

Getis-Ord hotspot analysis

The Getis and Ord's local G_i^* statistic was used in a similar manner to previous work by Nawaz *et al*.[12] This statistic can be defined as follows:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{s \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}}$$

$$s = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X} \right)^2}$$

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

Where i is the tile index in the flattened grid, $w_{i,j}$ the weight between tiles i and j, n the number of tiles, and χ_j being the count of feature of interest for tile j. This statistic (a Z-score) is acquired by first setting binary weights defining the spatial relationships between the tiles of the grid. Given the size of the tiles, we restricted ourselves to first-order neighbors, meaning that for each tile t, only those tiles that share an edge or vertex with the current tile are considered neighbors [Figure 1c]. Each tile was compared along with its neighbors against the entire image to establish whether the current tile is in an area with a high concentration of extreme values and correspondingly with a substantially higher or lower value than the alternative null hypothesis. The statistic then returns Z-scores and *P* values per tile to indicate whether the tile is a statistically significant cold or hotspot [Figure 1d]. We used a Z-score threshold of ±1.96 to indicate a significantly high/low score and a corresponding *P* value threshold of 0.05 for determining statistical significance as per Getis and Ord.[14] The cold and hotspots of each feature of interest were then quantified for each patient. For each feature of interest, 2 candidate prognostic factors were formed (e.g., feature of interest: "CD3+ cells in the CT," candidate prognostic factors: (1) number of CD3+ coldspots in CT and (2) number of CD3+ hotspots in CT). This resulted in the generation of 50 candidate prognostic factors, 25 being the number of cold spots and 25 being the number of hotspots for each feature of interest.

Spatial heatmap analysis

We first constructed multiple spatial heatmaps of the different features of interest [lymphocyte densities and ratios, TB, and their spatial interrelationships; Figure 1e]. Each tile within each heatmap was automatically assigned a heat value based on the number of objects of interest per tile (e.g., tiles with higher CD3+ density were given a higher heat value than the tiles with lower CD3+ density). Average heat values from each heatmap per patient were exported. The optimal cutoff point for which a tile would be considered "hot" or "cold" (i.e., a tile with large or small numbers of the feature of interest respectively) was found using the maximally selected rank statistic from the R[21] surv_miner package[22] on the data of 114 Edinburgh cases from 2002 to 3. The cutoff point was applied to the rest of the patients from Edinburgh (2004) and Japan. All tiles within each heatmap were then assigned to be either cold or hotspots [Figure 1f]. Similarly to the Getis-Ord analysis, the number of cold and hotspots present within each heatmap of the feature of interest was quantified. This led to the generation of 50 candidate prognostic factors.
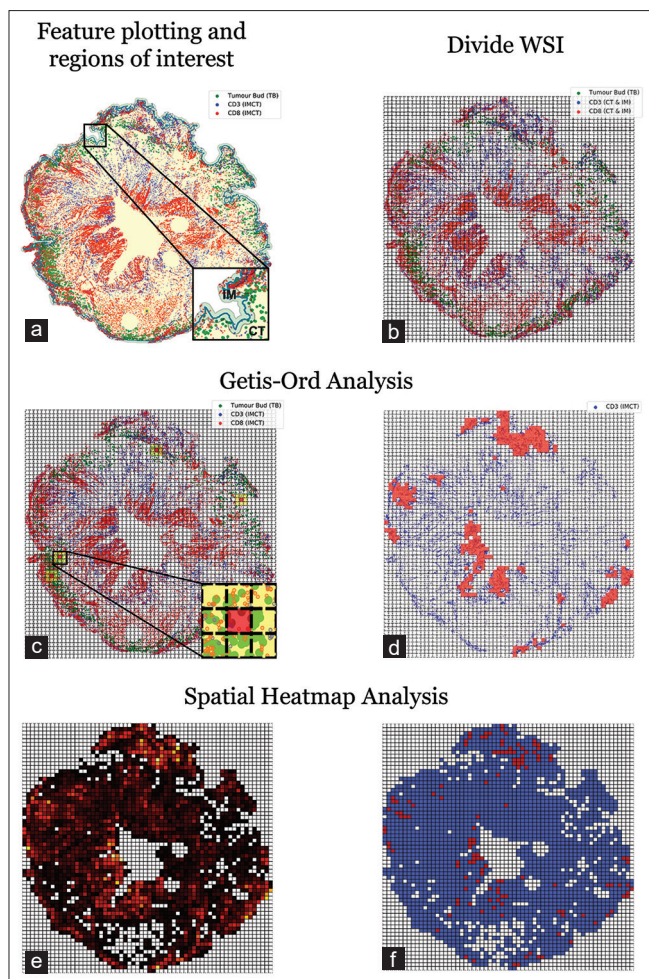
Figure 1: Methods for the assessment of intratumor heterogeneity. (a) Feature plot, tumor bud; green, CD3+; blue, CD8+; red, tumor core; yellow and invasive margin; teal. (b) Tiled Whole Slide Image (WSI). (c) Getis-Ord analysis: Tiles of interest overlaid in red and neighboring tiles overlaid in yellow. (d) Getis-Ord analysis: Detection of CD3 + cell hotspots in the invasive margin and tumor core, CD3+; blue and hotspot tile; red. (e) Heatmap analysis: Heatmap generation for CD3 + cells in the invasive margin and tumor core shows a spectrum through yellow (high CD3+) to black (low CD3+). (f) Heatmap analysis: Categorization of cold and hotspots within the WSI, hot tiles in red and cold tiles in blue

For both Getis-Ord and the spatial heatmap analyses, we excluded tiles devoid of TB or lymphocytes with a binary mask that classifies each tile based on the presence of the object of interest.

Survival analysis

Candidate prognostic factors derived from each type of analysis (Getis-Ord hotspot or spatial heatmap) were studied separately. However, the methodology for the survival analysis was kept consistent for both datasets. Image candidate prognostic factors together with the corresponding clinicopathological data from the original pathology report were imported into R studio 1.1.419[23] running R 3.4.3.[21] The least absolute shrinkage and selection operator (LASSO) penalized Cox proportional

hazard regression with 10 fold cross-validation was used to identify significant prognostic features using the glmnet package.[24] Random Forest decision tree model from the random Forest package[25] using out-of-bag validation was then employed to rank the significant features according to the corresponding decrease in the Gini coefficient. The least significant features were then removed in an iterative process, and the most prognostically significant model for each dataset was selected. Univariate Cox regression using the bootstrap resampling technique as well as Kaplan–Meier (KM) curve analysis was applied to evaluate the prognostic significance of the models. Using the Benjamini–Hochberg procedure,[26] P values from the KM analyses were corrected for false discovery rate. Univariate Cox regression was also applied to the data from the original clinicopathological report [Table 1].

RESULTS

Intratumor heterogeneity assessed using Getis-Ord Hotspot analysis

Using the spatial coordinates of each CD3 + or CD8 + T cell as well as those of the TB, we evaluated the intratumor heterogeneity using two different methods. In the first approach, we applied the Getis-Ord hotspot analysis to identify any statistically significant cold or hotspots of our objects of interest, in this case, the CD3+, CD8 + T cells, TB and their spatial inter-relationships within different tumor regions (IM, CT, IMCT). Each tile of area 0.785 mm^2 was compared to their first-order neighboring tiles and a Z-score with a P value was computed for each tile. We then quantified the number of statistically significant ($P < 0.05$) cold or hotspots [Figure 1d].

Intratumor heterogeneity assessed using spatial heatmap analysis

In the second approach, the spatial heatmaps of each object of interest were constructed. The average heat per heatmap of each patient was exported and the optimal cutoff point for which a tile was considered "hot" or "cold" was calculated [Supplementary Table 1]. The cold or hotspots within each patient for each object of interest were quantified.

Survival analysis

Cox regression with LASSO regularization was performed on each dataset in order to identify the most significant prognostic factors. Eight and 12 features were found to be significantly prognostic from the Getis-Ord and the spatial heatmap analysis, respectively [Table 2]. The significant features were further analyzed using the Random Forest decision tree model in order to rank the prognostic features according to their mean decrease Gini coefficient [Table 2]. The most prognostically significant model for each analysis was then selected. The Getis-Ord based prognostic model (GOPM) combined the number of TB hotspots and the number of the CD3 + coldspots within the IMCT. The spatial heatmap based prognostic model (SHPM) included the

Table 2: Least absolute shrinkage and selection operator penalized cox regression and random forest Gini coefficients for the significant features of each analysis

Features	Coefficients	
	LASSO	**Mean decrease Gini**
Getis-Ord		
Number of TB hotspots	**0.0048**	**12.6588**
Number of CD3+/CD8+ hotspots in CT	0.0025	11.3644
Number of CD3+ coldspots in IMCT	**−0.0017**	**10.8637**
Number of CD8+ coldspots in IMCT	−0.0003	7.8905
Number of TB coldspots	0.0230	6.8170
Number of CD3+ and CD8+ within 100 μm of TB coldspots	−0.0022	6.6115
pT	0.2185	3.7512
Age	0.3702	3.6080
Spatial heatmap		
Number of TB hotspots	**0.0080**	**9.2594**
Number of CD3+ within 100 μm of TB hotspots	**−0.0009**	**7.1527**
Number of CD3+ and CD8+ coldspots in IMCT	0.0003	6.9191
Number of CD3+ and CD8+ coldspots in CT	0.0003	6.8273
Number of CD3+ and CD8+ within 50 μm of TB coldspots	−0.0043	6.1868
Number of CD8+ within 100 μm of TB hotspots	−0.0033	6.0353
Number of CD3+/CD8+ hotspots in IMCT	0.0009	5.9058
Number of CD3+ within 50 μm of TB coldspots	−0.0009	5.6958
Number of CD3+ hotspots in CT	−0.0003	5.5083
pT	0.6508	2.1479
Site of tumor	0.0903	1.8549
Age	0.6082	1.3227

Features included in the final prognostic models are shown in bold. LASSO: Least absolute shrinkage and selection operator, TB: Tumor buds, CT: Tumor core, IMCT: Invasive margin and tumor core

number of TB hotspots and the number of CD3+ cells within 100 μm proximity of TB hotspots. Cutoff point values for the 4 selected features are shown in Supplementary Table 2. Concerning the GOPM, patients with low number of TB hotspots and high number of CD3+ coldspots were grouped into the low-risk category; patients with both features either low or high were grouped into the mid-risk category, and patients with high number of TB hotspots and low number of CD3+ coldspots were grouped into the high-risk category. In regard to the SHPM, patients with low number of TB hotspots and high number of CD3+ cells within 100 μm proximity of TB hotspots were grouped into the low-risk category, whereas patients with both features either low or high were grouped into the mid-risk category. Finally, patients with high number of TB hotspots and low number of CD3+ cells within 100 μm proximity of TB hotspots were grouped into the high-risk group. Univariate Cox regression showed both GOPM and SHPM to be highly significant in predicting DSS ($P < 0.001$, hazard ratio [HR] = 2.92; 95% confidence interval [CI]: 1.91–4.46, Bootstrap $P = 0.001$ and $P < 0.001$, HR = 4.92; 95% CI: 3.04–7.97, Bootstrap $P = 0.001$, respectively). Results from KM survival analysis further confirmed the prognostic significance of these models [Figure 2].

DISCUSSION

In this study, we assessed the spatial intratumor heterogeneity of TB and lymphocytes as well as its prognostic significance on 232 Stage II CRC specimens labeled with multiplexed immunofluorescence. By plotting the spatial coordinates of lymphocytes and TB identified by automated image analysis, we first applied the Getis-Ord hotspot method to quantify the tiles with statistically significant low or high numbers of lymphocytes, TB, or their associations (cold or hotspots, respectively). Second, we developed a new methodology for the classification and quantification of these cold or hotspots. Our results show that the evaluation and quantification of these features improve prognostic accuracy in Stage II CRC.

A number of studies have shown the prognostic significance of lymphocytic cell infiltration within distinct regions of the tumor microenvironment (TME) such as the IM and the CT in CRC.[27,28] In a meta-analysis by Zhao *et al.*, CD3+ cell infiltration was associated with disease-free survival only when assessed at the IM, whereas CD8+ cell infiltration was associated with disease-free survival when quantified at the CT.[28] Similarly, the prognostic significance of tumor budding has mainly been reported when evaluated at the invasive front,[29-31] although intratumoral budding has been suggested as a potential prognostic marker.[32] Quantification of such features across large regions of interest such as the IM or the CT may represent better the phenotypes found across the whole slide image. However, through averaging out the object densities within these large areas, it omits any specific areas where cells
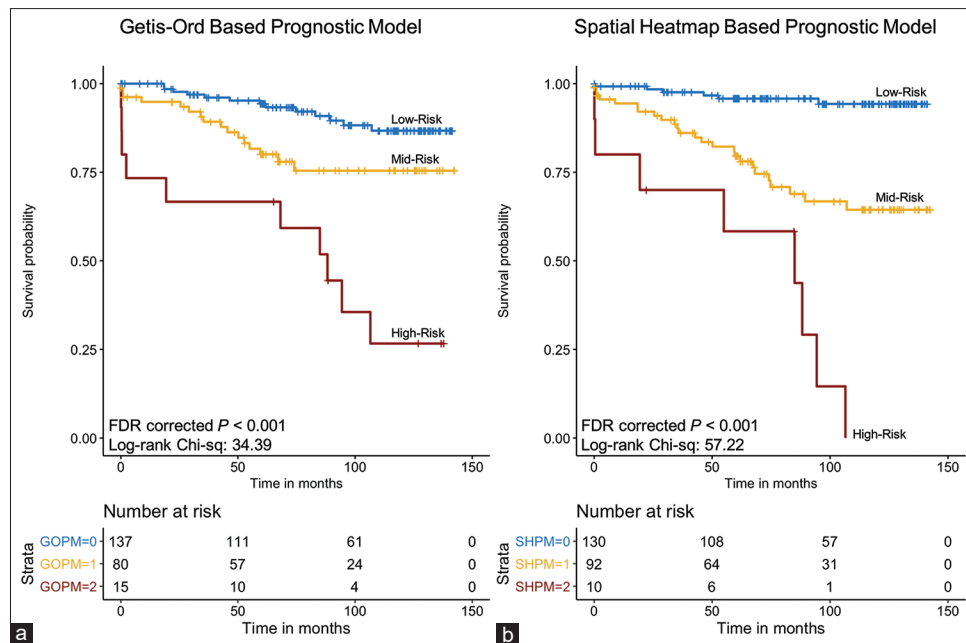
Figure 2: Kaplan–Meier survival analysis for Getis-Ord and spatial heatmap-based prognostic models. (a) Getis-Ord-based prognostic model. Blue line represents the "low-risk" patients, yellow line represents the "mid-risk" patients, and the red line represents the "high-risk" patients. (b) Spatial heatmap-based prognostic model. Blue line represents the "low-risk" patients, yellow line represents the "mid-risk" patients, and the red line represents the "high-risk" patients

are highly clustered which in turn might be sufficient to lead to metastasis and poor outcome.

In order to overcome this, we performed two spatial analysis methods, the Getis-Ord hotspot and the spatial heatmap analysis. Getis-Ord hotspot analysis has mainly been used in ecological studies, however in a study by Nawaz *et al.*, this type of analysis was shown to be promising when assessing the intratumoral heterogeneity in pathological images from breast cancer.[12] The Getis-Ord hotspot analysis offers the opportunity not only to identify any dense or sparse tiles within an image but also to identify statistically significant cold and hotspots. In addition, it has the strength of comparing the statistical significance of the different tiles in order to identify specific regions with cell clustering which can be further explored.[12] Our newly developed spatial heatmap analysis differs in methodology from the Getis-Ord hotspot analysis, whereas in Getis-Ord hotspot analysis, each tile of interest is compared to its first-order neighbors, the spatial heatmap analysis involves calculating the mean heat value across the whole tissue slide for each feature of interest. The cutoff point splitting the hot and cold heat values for each feature is calculated based on the data from all patients of interest therefore taking into account both intra- and interpatient heterogeneity. This cutoff point is then applied across the whole patient cohort, therefore ensuring consistency when considering the regions as cold or hotspots. Although the two methodologies for the establishment of the cold and hotspots within the images differ, the local regions within the image where clustering occurs were found to be similar, which strengthens our confidence in using such methodologies interchangeably.

While we present this method using CD3+, CD8+ cells, and TB as features of interest in CRC, such methodology could be applied to other types of tumor and other TME compartments including macrophage infiltration, vascular invasion, and immune checkpoint expression. Indeed, the distribution pattern of tumor-associated macrophages within the gastric cancer microenvironment was previously shown to be an independent prognostic factor.[33] Moreover, the ratio of CD68+/CD163+, specifically in the CT, has previously been shown to be highly prognostic in Stage II CRC.[34] Finally, a recent study by Tsakiroglou *et al.* has shown how evaluation of the spatial interaction of T+ and PD-L1+ cells can serve as a prognostic biomarker in oropharyngeal squamous cell carcinoma.[35] Therefore, by applying spatial analysis methods, such as the Getis-Ord hotspot or the spatial heatmap, may lead to the discovery of new prognostic factors as well as to the identification of specific regions of interest within the TME, which can be further analyzed for the potential discovery of targets for treatment.

When assessing the prognostic significance of the features derived from each spatial analysis type, many of them were shown to be promising prognostic factors. However, by applying a Cox regression with LASSO regularization and a Random Forest decision tree model, the most prognostically significant features were identified and combined into two prognostic models. The prognostic model derived from the results of the Getis-Ord hotspot analysis included the number of TB hotspots and the number of the CD3 + coldspots within the IMCT. The prognostic model derived from the results of the spatial heatmap analysis included the number of TB hotspots

and the number of CD3 $^+$ cells within 100 µm proximity of TB hotspots. These results concur well with previous findings,[11,34] where CD3 $^+$ density in the IMCT, the number of TB, and the spatial association between CD3 $^+$ and CD8 $^+$ cells with TB have been found to be highly prognostic. Unlike these studies,[11,34] although features which included CD8 $^+$ cells appeared to be significant, they were not selected to be included in the final prognostic model. These different results might be attributed to the methodology used in this study which differs significantly from the other two studies. Previously, the average numbers and densities of the lymphocytes, TB, and their spatial associations across large regions of interest such as the IM or the CT were quantified.[11,34] In this study, these large regions, namely IM and CT, are subdivided into smaller tiles which are then compared to each other in order to identify any specific aggressive areas with feature clustering that can potentially have an effect on tumor progression and patient prognosis. The two prognostic models developed in this study were shown to have high prognostic significance and were able to significantly stratify the Stage II CRC patients into low-, medium- and high risk of disease-specific death. Application of such models could aid clinical decision-making on adjuvant therapy and follow-up, and hence improve patient outcomes by identifying the high-risk patients who may benefit from further treatment, the low-risk Stage II subpopulation who may not need any further toxic treatment, and the mid-risk patients who may benefit from a more detailed follow-up.

CONCLUSIONS

This research highlights the importance of spatial statistics in investigating the heterogeneous TME. We first applied a known method for the investigation of the intratumor heterogeneity in Stage II CRC patient samples. We, secondly, proposed an alternative methodology which takes into account interpatient heterogeneity in addition to intratumor heterogeneity. Moreover, we have shown that quantification of tumor and lymphocyte hotspots within the TME using both methods was significantly associated with patient outcome and we hope that such studies will serve as a base for further research in the prognostic impact of the intra-tumor heterogeneity.

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Brafford P, Herlyn M. Gene expression profiling of melanoma cells-searching the haystack. J Transl Med 2005;3:2.
2. Rybinski B, Yun K. Addressing intra-tumoral heterogeneity and therapy resistance. Oncotarget 2016;7:72322-42.
3. Stanta G, Bonin S. Overview on clinical relevance of intra-tumor heterogeneity. Front Med (Lausanne) 2018;5:85.
4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
5. Sagaert X, Vanstapel A, Verbeek S. Tumor heterogeneity in colorectal cancer: What do we know so far? Pathobiology 2018;85:72-84.
6. Fotheringham S, Mozolowski GA, Murray EM, Kerr DJ. Challenges and solutions in patient treatment strategies for stage II colon cancer. Gastroenterol Rep (Oxf) 2019;7:151-61.
7. Lugli A, Kirsch R, Ajioka Y, Bosman F, Cathomas G, Dawson H, *et al*. Recommendations for reporting tumor budding in colorectal cancer based on the international tumor budding consensus conference (ITBCC) 2016. Mod Pathol 2017;30:1299-311.
8. Zlobec I, Lugli A. Tumour budding in colorectal cancer: Molecular rationale for clinical translation. Nat Rev Cancer 2018;18:203-4.
9. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, *et al*. International validation of the consensus immunoscore for the classification of colon cancer: A prognostic and accuracy study. Lancet 2018;391:2128-39.
10. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, *et al*. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. J Pathol 2014;232:199-209.
11. Nearchou IP, Lillard K, Gavriel CG, Ueno H, Harrison DJ, Caie PD. Automated analysis of lymphocytic infiltration, tumor budding, and their spatial relationship improves prognostic accuracy in colorectal cancer. Cancer Immunol Res 2019;7:609-20.
12. Nawaz S, Heindl A, Koelble K, Yuan Y. Beyond immune density: Critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. Mod Pathol 2015;28:766-77.
13. Yuan Y. Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer. Journal of The Royal Society Interface. 2015;12(103):20141153.
14. Getis A, Ord JK. The analysis of spatial association. Geogr Anal 1992;24:189-206.
15. Schönmeyer R., Brieu N., Schaadt N., Feuerhake F., Schmidt G., Binnig G. (2015) Automated Whole Slide Analysis of Differently Stained and Co-Registered Tissue Sections. In: Handels H., Deserno T., Meinzer HP., Tolxdorff T. (eds) Bildverarbeitung für die Medizin 2015. Informatik aktuell. Springer Vieweg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-46224-9_70
16. van Rossum G. Python Tutorial. Tech Rep CS-R9526, Cent Voor Wiskd en Inform; 1995.
17. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, *et al*. Jupyter Notebooks – a publishing format for reproducible computational workflows. Loizides, Fernando and Scmidt, Birgit (eds.) In Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press. 2016. p. 87-90 . (doi:10.3233/978-1-61499-649-1-87).
18. Oliphant TE. Guide to NumPy; 2006.
19. Rey SJ, Anselin L. PySAL: A python library of spatial analytical methods. Rev Reg Stud 2007;37:5-27.
20. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng 2007;9:90-5.
21. R Core Team. R: A Language and Environment for Statistical Computing; 2018. Available from: https://www.r-project.org/. [Last accessed on 2019 Oct 09].
22. Kassambara A, Kosinski M, Biecek P, Fabian S. Package "Survminer" Type Package Title Drawing Survival Curves using "ggplot2; 2019.
23. RStudio Team. RStudio Server: Integrated Development for R. Boston, MA: RStudio Inc.; 2016. Available from: http://www.rstudio.com/. [Last accessed on 2019 Oct 09].
24. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1-22.
25. Liaw A, Wiener M. Classification and Regression by random Forest. R News 2002;2:18-22.

26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Royal Statist Soc Series B (Methodological) 1995;57:289-300.
27. Guo G, Wang Y, Zhou Y, Quan Q, Zhang Y, Wang H, *et al*. Immune cell concentrations among the primary tumor microenvironment in colorectal cancer patients predicted by clinicopathologic characteristics and blood indexes. J Immunother Cancer 2019;7:179.
28. Zhao Y, Ge X, He J, Cheng Y, Wang Z, Wang J, *et al*. The prognostic value of tumor-infiltrating lymphocytes in colorectal cancer differs by anatomical subsite: A systematic review and meta-analysis. World J Surg Oncol 2019;17:85.
29. Ueno H, Murphy J, Jass JR, Mochizuki H, Talbot IC. Tumour 'budding' as an index to estimate the potential of aggressiveness in rectal cancer. Histopathology 2002;40:127-32.
30. Tanaka M, Hashiguchi Y, Ueno H, Hase K, Mochizuki H. Tumor budding at the invasive margin can predict patients at high risk of recurrence after curative surgery for stage II, T3 colon cancer. Dis Colon Rectum 2003;46:1054-9.
31. Nearchou IP, Kajiwara Y, Mochizuki S, Harrison DJ, Caie PD, Ueno H. Novel internationally verified method reports desmoplastic reaction as the most significant prognostic feature for disease-specific survival in stage II colorectal cancer. Am J Surg Pathol 2019;43:1239-48.
32. Lugli A, Vlajnic T, Giger O, Karamitopoulou E, Patsouris ES, Peros G, *et al*. Intratumoral budding as a potential parameter of tumor progression in mismatch repair-proficient and mismatch repair-deficient colorectal cancer patients. Hum Pathol 2011;42:1833-40.
33. Liu JY, Peng CW, Yang GF, Hu WQ, Yang XJ, Huang CQ, *et al*. Distribution pattern of tumor associated macrophages predicts the prognosis of gastric cancer. Oncotarget 2017;8:92757-69.
34. Nearchou IP, Gwyther BM, Georgiakakis EC, Gavriel CG, Lillard K, Kajiwara Y, *et al*. Spatial immune profiling of the colorectal tumor microenvironment predicts good outcome in stage II patients. NPJ Digit Med 2020;3:71.
35. Tsakiroglou AM, Fergie M, Oguejiofor K, Linton K, Thomson D, Stern PL, *et al*. Spatial proximity between T and PD-L1 expressing cells as a prognostic biomarker for oropharyngeal squamous cell carcinoma. Br J Cancer 2020;122:539-44.

Supplementary Table 1: Cutoff values for each feature of the spatial heatmap analysis used to categorize the tiles into cold or hotspots

Feature	Cutoff value
CD3$^+$cells in the IM	21.7013333
CD8$^+$ cells in the IM	13.7399103
CD3$^+$ and CD8$^+$cells in the IM	30.2813067
CD3$^+$cells in the CT	30.0809727
CD8$^+$cells in the CT	10.6572454
CD3$^+$ and CD8$^+$ cells in the CT	47.1046484
CD3$^+$ cells in the IMCT	26.4211475
CD8$^+$ cells in the IMCT	15.6260472
CD3$^+$ and CD8$^+$ cells in the IMCT	39.5460152
TB density	3.15805471
CD3$^+$/CD8$^+$ in the IM	43.8087855
CD3$^+$/CD8$^+$ in the CT	17.4055556
CD3$^+$/CD8$^+$ in the IMCT	17.1737661
TB number within 50 μm of CD3$^+$ cells	2.02352941
TB number within 100 μm of CD3$^+$ cells	4.00617284
TB number within 50 μm of CD8$^+$ cells	2.32098765
TB number within 100 μm of CD8$^+$ cells	2.64423077
TB number within 50 μm of CD3$^+$ and CD8$^+$ cells	2.71794872
TB number within 100 μm of CD3$^+$ and CD8$^+$ cells	6.08
Mean TB number within 50 μm of CD3$^+$ cells	0.2323185
Mean TB number within 100 μm of CD3$^+$ cells	1.54760868
Mean TB number within 50 μm of CD8$^+$ cells	0.16349674
Mean TB number within 100 μm of CD8$^+$ cells	0.57304754
Mean TB number within 50 μm of CD3$^+$ and CD8vcells	0.28206629
Mean TB number within 100 μm of CD3$^+$ and CD8$^+$ cells	1.14422595

IM: Invasive margin, CT: Tumor core, IMCT: Invasive margin and tumor core, TB: Tumor buds

Supplementary Table 2: Cutoff values for the Getis-Ord and the spatial heatmap prognostic model components

Feature	Cutoff value
Getis-Ord prognostic model	
Number of TB hotspots	35
Number of CD3$^+$ coldspots in IMCT	20
Spatial heatmap prognostic model	
Number of TB hotspots	120
Number of CD3$^+$ within 100μm of TB hotspots	56

TB: Tumor bud, IMCT: Invasive margin and tumor core