SPATIAL INTERACTION ANALYSIS WITH GRAPH BASED MATHEMATICAL MORPHOLOGY FOR HISTOPATHOLGY

Bassem Ben Cheikh[1] *Nicolas Elie*[2] *Benoit Plancoulaine*[3] *Catherine Bor-Angelier*[4] *Daniel Racoceanu*[5]

[1] Sorbonne Universités, UPMC Paris 6, CNRS, INSERM, Biomedical Imaging Laboratory, Paris, France
[2]Normandie Univ, UNICAEN, SFR 4206 ICORE, CMABio3, 14000 Caen, France
[3]Normandie Univ, UNICAEN, INSERM, BioTICLA, 14 000 Caen, France
[4]Unicancer - Auvergne Rhône Alpes, Centre Jean Perrin - Service de Pathologie, France
[5]Pontifical Catholic University of Peru, Engineering Department, San Miguel, Peru

ABSTRACT

Exploring the spatial interactions between tumor and the inflammatory microenvironment using digital pathology image analysis can contribute to a better understanding of the immune function and tumor heterogeneity. We address this by providing tools able to reveal various metrics describing spatial relationships in the cancer ecosystem. The approach comprises nuclei segmentation and classification, using supervised learning algorithm, to detect lymphoid aggregates and tumor patterns, and spatial distribution quantification using sparse sets' mathematical morphology. Tumor patterns were classified into three groups: surrounded by lymphocytes, close to lymphoid aggregates or distant and might be protected from immune attack. The approach provides statistical assessment and comprehensive visual representation of the inflammatory tumor microenvironment.

Index Terms— Tumor-immune system interaction, spatial relation modeling, mathematical morphology, graph representation, digital histopathology

1. INTRODUCTION AND RELATED WORK

Tumor inhabits a cellular environment composed of parts of different types, such as immune cells, blood vessels, collagen, fat and many other cell types. The whole of these components is referred to as the tumor microenvironment (TME). The interactions between tumor and its TME are recognized as playing an important role in the progression of the disease. In particular, the inflammatory microenvironment (iTME) is known to have a great impact on the tumor behavior [1,2]. With the progress being made in the development of cancer immunotherapy, a considerable number of efforts have focused on the understanding of the immune system-cancer interactions. By understanding these interactions, tumor may be treated more effectively. Links have been shown to exist between clinical outcome and immune cell presence, relative abundance, as well as spatial proximity of immune cells to invasive cancer cells. In [3], metrics describing absolute cell number, relative cell type proportion (ratio) and density (minimum, median and maximum) based on the 50 nearest neighbors, were computed. The study was conducted among 768 H&E-stained surgical samples of breast tumors. From the set of parameters of this model, only median lymphocyte density was associated with one of the clinical variables; the predictor of pathological complete response (pCR), which is an important histological indicator of chemotherapy response. In [4], the cancer cell density was quantified using kernel estimator. Then, for every lymphocyte, its spatial proximity to cancer was quantified with the cancer density landscape at its location. A quantitative measure of tumor-lymphocyte ratio was found to be significantly associated with disease-specific survival based on statistical modelling conducted on 181 H&E-stained breast cancer tissues. In [5], a clinical outcome analysis revealed that the degree of clustering of dendritic cells in tumor-positive lymph nodes correlated with the duration of disease-free survival in breast cancer patients. The study was conducted on 59 tissue sections using different immunohistochemistry stains that highlight different immune cell types. To investigate the interactions between dendritic cells and T cells, the contact between two cells was defined as co-localization within a radius of 100 pixels. In [6], based on spatial statistics, spatial grouping patterns of T and B cells were found different between healthy and breast cancer lymph nodes. The cell density was estimated using the Gaussian kernel density estimation method. The study was conducted on 25 patients using different immunohistochemistry staining techniques. In another study [7], it was found that the amount of co-localized cancer and immune hotspots, weighted by tumor area, correlates with a better prognosis in univariate and multivariate analysis. The study was carried out on H&E-stained tumor section images from 245 breast cancer patients. In [8], three classes of interactions between tumor and immune cells were defined: surveillance (S), indicating a low chance, combat and surveillance (CS), indicating intermediate chance of interaction, and combat (C), indicat-

ing high chance of interaction. Immune cell clusters were classified using a supervised learning algorithm based on the abundance, the distance to tumor cells and clustering behavior (computed using K-means algorithm).

These approaches are based mainly on spatial statistics for measuring the heterogeneity of immune cell infiltration in tumors. We hereby propose a different approach where the spatial information is given straight by the morphology of the tissue using graph-based mathematical morphology. Graph-based methods were proven to be effective in many applications in image processing and analysis the recent decades [9,10]. Nowadays, with the expanding field of digital pathology, they are gaining large popularity in histopathology image analysis, as they describe spatial characteristics and neighborhood relationships that are visually interpreted by the pathologist during the examination of a tissue specimen [11-16].

In this framework, we propose a new graph-based approach to characterize the spatial relationships existing in the cellular environment of tumors using sparse sets' mathematical morphology (MM). The tools of morphology on graphs were first used in [17] to study the neighborhood relationships between cells in germinal centers from lymph nodes, then in [18,19] for semantic spatial configuration modeling in histopathology. We continue these efforts by focusing on the characterization of the spatial distribution of tumor cells and the interactions with surrounding normal cells. TME is composed of a wide range of biological structures but this work is dedicated to the characterization of the immune-cancer interactions.

The next section provides a description of the proposed approach. First, we present our nuclei segmentation and classification algorithm. Then, we introduce our technique for the identification of significant spatial clusters of cells, before we describe our framework for the characterization of the spatial interactions between tumor cell and immune cell aggregates. Section 3 describes the experiments and results of this study. Finally, a conclusion summarizes our main contribution and further works.

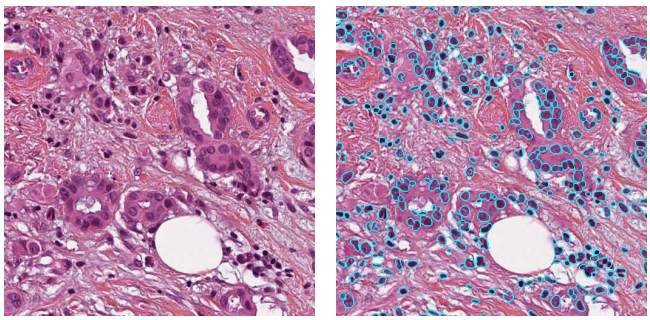
2. METHOD

2.1. Nuclei segmentation and classification

In digital pathology image analysis, the extraction of meaningful information describing the relationships between the tumor and its microenvironment relies on an accurate cell identification technique. In this paragraph, we present our approach for nuclei segmentation and classification from HES (Hematoxylin-Eosin-Saffron)-stained breast cancer tissue. In the following, a tissue image is subdivided into a series of tiles of size 1024×1024 at $0.495\mu m$ per pixel. Nuclei segmentation and classification is performed on each tile separately.

2.1.1. Nuclei segmentation

First, we apply a fast superpixel segmentation algorithm, called SLIC [20], that clusters pixels in the five-dimensional color and image plane space to efficiently generate compact, nearly uniform small regions. Each region is expected to represent a specific biological structure. The number of desired superpixels and the compact factor of SLIC algorithm were set to 3500 and 35, respectively. Then, a stain separation is performed using a color deconvolution algorithm [21] in order to distinguish nuclear regions. To circumvent the problem of variability in hematoxylin concentration, a histogram stretching is employed to the hematoxylin channel, output of the color deconvolution algorithm, which contains real values in [0 1]. Nuclear objects are then extracted by image thresholding, i.e. where the hematoxylin intensity is greater than 0.4, followed by morphological filtering and small object removal. At this step, some segmented nuclei may be overlapped. Therefore, we use the output of the superpixel segmentation algorithm to make nuclei separation and delineate the nuclei boundaries. It should be noted that this separation technique, i.e. using SLIC segmentation, gives better results than watershed-based methods. The parameter values of this algorithm were chosen based on optimization of the detection F-score on the dataset provided in [22]. Qualitative results are shown in figure 1.



(a) (b)

Fig. 1. (a) Example of a 512×512 image of HES-stained breast cancer tissue. (b) Result of nuclei segmentation

2.1.2. Nuclei classification

The purpose of this step is to classify the nuclei detected during the previous step into three classes: epithelial cells, immune cells and fibroblasts. Therefore, we use a supervised learning algorithm based on color and texture features. In order to generate a ground truth for the learning algorithm, we have manually annotated 2533 regions. Each region contains one or more nuclei of the same class. The regions were chosen from 1005 tiles, extracted from 17 whole slide images. The learning and evaluation of the classification algorithm were restricted to the annotated area. The total number of nuclei detected within the annotated regions is 112125 (77676 cancer nuclei, 31037 immune cell nuclei, and 3412 fibrob-

lasts).
The color features extracted were calculated within the segmented nucleus area, but also in the area surrounding the nucleus, as it gives contextual information. The area is defined by a morphological dilation with a radius $r = 20$ of the segmented nucleus object. The color features are therefore the mean, the standard deviation and the median of each of the RGB channels, hematoxylin channel, the gray-intensity, the blue ratio BR and the red ratio RR, given by the following equations:
$BR = \frac{255 \times B}{(1+R+G)(1+R+G+B)}$ and $RR = \frac{R}{1+G+B+\frac{(G-B)^2}{1+G+B}+R}$

The texture features are calculated from a 100×100 frame centered at the nuclear centroid. Four statistics are calculated (Contrast, Correlation, Energy, and Homogeneity) from the Gray-Level Co-occurrence Matrix (GLCM) with 8 levels and a step of 1 pixel in 4 directions ($0°, 45°, 90°$, and $135°$), applied to both grey-level and blue ratio images. Also five statistics were extracted (mean, standard deviation, skewness, kurtosis and entropy) from the results of the convolution of the blue ratio image with 5-sized Laws' masks [23].
The proposed features were chosen using Fisher Score for feature selection after fine-tuning the different parameters, techniques and channels. The total number of the selected features is 147 features, chosen out of 1425 features. The average accuracy of the classification using Random Forest classifier based on the selected features is 0.9612. Qualitative results are shown in figure 2.b.

2.2. Sparse sets' mathematical morphology

Mathematical morphology on graphs was first explored by Vincent et al. [26], where the morphological operators perform on the graph nodes rather than the image pixels. This theory provides a great number of powerful tools for studying graphs that can be defined on a given set of objects depending on a desirable neighborhood relationship. In this study, we perform morphological operations on the indices (labels) of the vertices of the graph. Let $G = (V, E)$ be a simple graph comprising a set V of vertices and a set $E \subseteq V \times V$ of edges, we associate for each vertex v a value $I(v)$, that represents its cell type, i.e. cancer cell, immune cell, fibroblast, and we denote $G = (V, E, I)$.
The morphological dilation of G is defined by the graph $\delta(G) = (V, E, I_d)$ such that $I_d(v) = max\{I(u), u \in N_G(v)\}, \forall v \in V$. The morphological erosion of G is defined by the graph $\epsilon(G) = (V, E, I_e)$ such that $I_e(v) = min\{I(u), u \in N_G(v)\}, \forall v \in V$.
We define dilation of order n as
$\delta^n(G) = \delta \circ \delta \circ \cdots \circ \delta(G), n\ times.$
Similarly, we define erosion of order n as
$\epsilon^n(G) = \epsilon \circ \epsilon \circ \cdots \circ \epsilon(G), n\ times.$

2.3. Tumor and lymphoid aggregates detection

To identify the location of significant cell aggregates, state-of-the-art approaches use mainly spatial statistics methods. In [5], the authors used a density-based clustering (DBC) algorithm [25] that groups cells that are close to each other in Euclidean distance into clusters. In [7], Getis-Ord analysis algorithm [26] was used to detect statistically significant cancer and immune hotspots. In our approach, we use graph-based mathematical morphology to detect hotspots.
The nodes that we have to this step correspond to the locations of detected nuclei. Another TME component that should also be considered is the connective tissue as it reflects spatial delimitation of between tumor aggregates (TA). Therefore, collagen fibers are segmented by thresholding the red ratio image (section 2.1.2), followed by morphological filtering. Then, nodes are extracted from the centroids of the superpixels of collagen.
The first step consists of setting up a neighborhood relationship between the different cells. In our study, we have chosen Delaunay graph [11] where its built on all nuclei and collagen nodes and refined by alpha-shape filter [27] to circumvent the border effects. We denote $G(V, E, I)$ the labeled graph obtained after filtering. We denote $G_c(V, E, I^c)$ the subgraph of all nuclei of type c, where $I^c(v) = 1$ if the node v is of class c and zero otherwise. Each connected component of the graph G_c represents cells of the same class c that are connected to each other without being interfered with other cell types. Hence, it is a first result of cell aggregate detection. Since we are interested in significant clusters of cells, we perform a morphological filtering to remove small, morphologically unstable aggregates of cells. Immune system (IS) hotspots are detected using the operation $\delta^3(\epsilon^2(G_{IS}))$ and tumor patterns are detected using the operation $\delta^1(\epsilon^1(G_{IS}))$. We have chosen larger radius values for IS cell filtering because lymphoid aggregates are compact and dense, when they exist in general. Qualitative results are shown in figure 2.c The morphological operations permit also to filter the errors that occurred during the nuclei detection and classification steps.

2.4. Quantification of the immune-cancer interactions

To investigate the spatial interactions between tumor and immune cells, we first generate the morphological distance map (MDM) of the immune system from its hotpots. The MDM associates to each node of the graph G its geodesic distance from the IS hotspots. MDM is calculated using an algorithm based on recursive morphological erosions of the graph $G_{I\bar{H}}(V, E, I^{I\bar{H}})$, where $I\bar{H}$ is the set of all nodes that dont belong to immune hotspots. The geodesic distance value at iteration i of the vertex v is calculated by addition of $(+1)$ after the erosion i. The process is repeated until all nodes $G_{I\bar{H}}$ are null. Figure 2.d shows the MDM value at each tumor cell. The approach offers a comprehensive visual representation about the degree of interaction of each tumor cell with immune system hotspots. Reddish-colored cancer cells have great spatial interactions with the inflammatory microenvironment. While bluish-colored cancer cells are distant

from the immune infiltration.

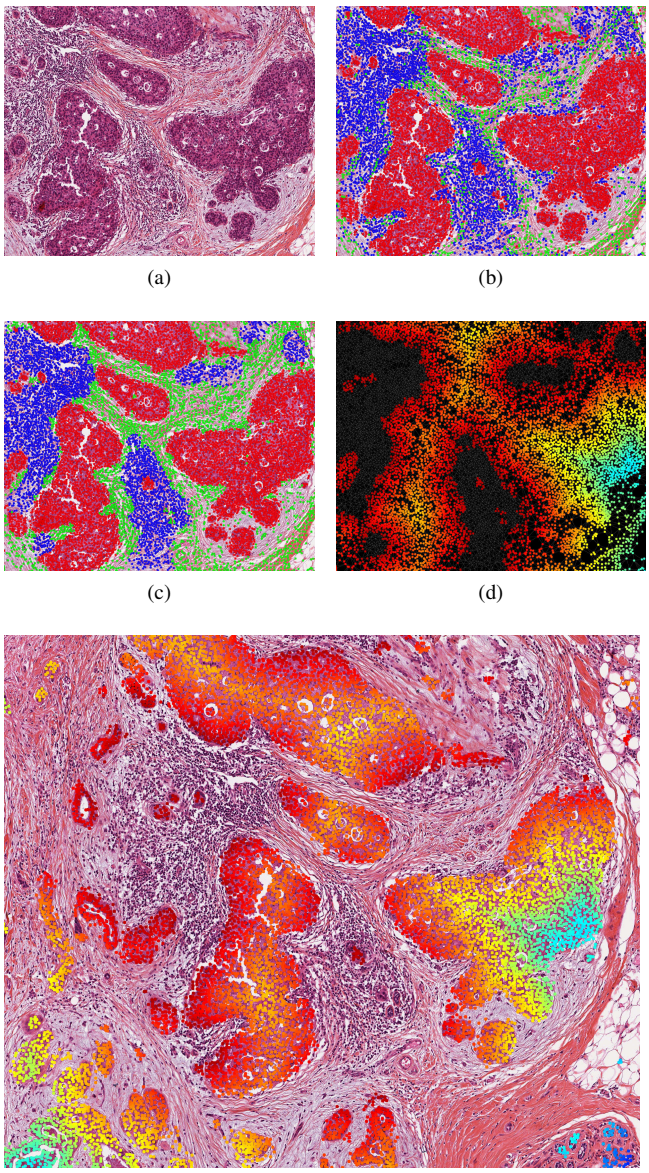


Fig. 2. (a) Original HES image. (b) Result of nuclei classification. (c) Morphological Filtering output. (d) Morphological Distance Map. (e) Cancer-immune spatial interaction representation

3. RESULTS AND DISCUSSION

In order to assess the validity of our approach we have defined and annotated three tumor-immune interaction scores: (i) TA entirely surrounded by immune cells, (ii) TA close to immune aggregates, and (iii) TA far from lymphoid infiltration, and might be protected from the immune attack. Figure 3 shows the mean of MDM values of tumor cells of TAs from different types. Tumor nodes that belong to a unique TA are calculated using Depth-First Search algorithm for connected component labeling. For a quantitative evaluation, we have annotated 340 tumor aggregates with one of the 3 classes from 15 whole slide images of breast cancer tissues. The diagram in figure 3 demonstrates the characterization of the 3 interaction scores defined previously. The mean MDM value of TAs that are entirely surrounded by immune cells are very low comparing to those of TAs that are just in contact with immune hotspot, which in turn have lower values than TA that were evaluated distant from immune hotspots. More features, such as enclosing, could be calculated from only the edges of the TAs rather than all nodes. The flexibility and the scalability of the approach are some of its most important advantages. In fact, this technique allow us to use morphological operations on an entire whole slide image at the nuclear level, which wouldnt be possible with classical mathematical morphology due to the very large size of the image at low resolution. The technique is adjustable to different application in digital pathology image analysis, i.e. tumors of different organs and to tissue with different stains. This approach works identically in the three-dimensional space as in the bi-dimensional, seeing the progress being made in the development of the 3D Histopathology. The approach gives also a visual interpretation of the spatial interactions in the inflammatory tumor microenvironment, which can help the pathologist during the evaluation of the immune response.

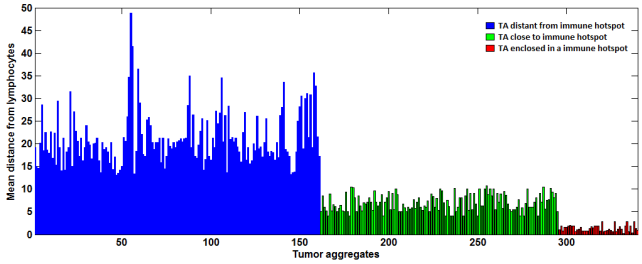


Fig. 3. Interpretation of the morphological parameters

4. CONCLUSION

In this work, we have presented a conceptual framework for analyzing the spatial interactions between cancer and immune system in histopathology images using graph-based mathematical morphology. Features that could lead to a better understanding of immune changes in the tumor microenvironment were proposed. In our future works, we will study the heterogeneity of the spatial interactions described in this work on a large dataset in order to uncover interdependencies with clinical outcome and/or genomic features.

ACKNOWLEDGMENTS

This study has been done with the support of the European Institute of Innovation & Technology (EIT) Health - PAPHOS project and the French Foundation for Medical Research (FRM) - convention DBS20131128436.

5. REFERENCES

[1] J. K. Adama, B. Odhavb, K. D Bhoola, "Immune responses in cancer," *Pharmacol Ther*, 2003.

[2] T. F. Gajewski, H. S. and Y. X. Fu, "Innate and adaptive immune cells in the tumor microenvironment," in *Nature Immunology*, 2013.

[3] Ali H. R., A. Dariush, Provenzano E., Bardwel H., E. Abraham J., Iddawela M., Vallier A. L., Hiller L., A. Dunn J., J. Bowden S., Hickish T., McAdam K., Houston S., J. Irwin M., D. P. Pharoah P., D. Brenton J., A. Walton N., M. Earl H. and Caldas C., "Computational pathology of pre-treatment biopsies identifies lymphocyte density as a predictor of response to neoadjuvant chemotherapy in breast cancer," in *Breast Cancer Research*, 2016.

[4] Yinyin Yuan, "Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer," *R Soc Interface*, 2016.

[5] A. Y. Chang, and al, "Spatial organization of dendritic cells within tumor draining lymph nodes impacts clinical outcome in breast cancer patients," *Journal of Translational Medicine*, 2013.

[6] A. F. Setiadi, and al, "Quantitative, Architectural Analysis of Immune Cell Subsets in Tumor-Draining Lymph Nodes from Breast Cancer Patients and Healthy Lymph Nodes," *Plos One*, 2010.

[7] N. Sidra, H. Andreas, K. Konrad, Y. Yinyin, "Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer," *Modern Pathology*, 2015.

[8] J. M. Kruger, and al, "Combat or surveillance? Evaluation of the heterogeneous inflammatory breast cancer microenvironment," *J. Pathol*, 2013.

[9] L. Grady, J. R. Polimeni, "Discrete Calculus: Applied Analysis on Graphs for Computational Science," *Springer*, 2010.

[10] O. Lezoray, L. Grady, "Image Processing and Analysis with Graphs: Theory and Practice," *CRC Press*, 20012.

[11] H. Sharma, and al, "A review of graph-based methods for image analysis in digital histopathology," *Diagnostic pathology*, 2012.

[12] V.T. Ta, O. Lezoray, A. Elmoataz, S. Schupp, "Graph-based tools for Microscopic Cellular Image Segmentation," *Pattern Recognition Special Issue on Digital Image Processing and Pattern Recognition Techniques for the Detection of Cancer*, 2009.

[13] S. Doyle, and al, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," *ISBI (IEEE)*, 2008.

[14] P. Khurd, and al, "Network Cycle Features: Application to Computer-Aided Gleason Grading of Prostate Cancer Histopathological Images," *ISBI (IEEE)*, 2011.

[15] C. Gunduz-Demir, M. Kandemir, A.B. Tosun, and C. Sokmensuer, "Automatic segmentation of colon glands using object-graphs," *Medical Image Analysis*, 2010.

[16] D. Altunbay, C. Cigir, C. Sokmensuer, and C. Gunduz-Demir, "Color graphs for automated cancer diagnosis and grading," *IEEE Transactions on Biomedical Engineering*, 2010.

[17] E. Raymond, and al, "Germinal Center Analysis with the Tools of Mathematical Morphology on Graphs," *Cytometry*, 1993.

[18] N. Lomenie and D. Racoceanu, "Point Set Morphological Filtering and Semantic Spatial Configuration Modeling: application to microscopic image and bio-structure analysis," *Pattern Recognition*, 2012.

[19] N. Lomenie and D. Racoceanu, "Spatial Relationships over Sparse Representations. Spatial Relationships over Sparse Representations," *Image and Vision Computing*, New Zealand. 2010.

[20] R. Achanta, and al, "SLIC Superpixels," *EPFL Technical Report no. 149300*, 2010.

[21] A.C. Ruifrok, and D.A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analytical and Quantitative Cytology and Histology*, 2001.

[22] K. Sirinukunwattana, and al, "Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images," *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 2016.

[23] Laws KI, "Textured Image Segmentation," *University of Southern California*, 1980.

[24] Luc Vincent, "Graphs and Mathematical Morphology," *Signal Processing*, 1989.

[25] Chen DZ, Smid M, Xu B, "Geometric algorithms for density-based data clustering," *Int J Comput Geometry Appl*, 2005.

[26] Getis A, Ord JK, "The analysis of spatial association by use of distance statistics," *Geogr Anal*, 1992.

[27] Lomenie N., Gallo L., Cambou N., and Stamon G., "Morphological operations on Delaunay triangulations," *ICPR*, 2000.