EUROPEAN INSTITUTE FOR BIOMEDICAL IMAGING RESEARCH

VPHPRISM

Theme: ICT-2011.5.2 - Virtual Physiological Human

VPH-PRISM

Virtual Physiological Human: Personalized Predictive Breast Cancer Therapy through Integrated Tissue Micro-Structure Modeling

Grant Agreement Number: 601040

Quantitative description and classification of microstructure

Deliverable 5.4

Lead Partner:	STICHTING KATHOLIEKE UNIVERSITEIT		
Author(s):	Babak Ehteshami Bejnordi, Jeroen van der Laak, Henning Kost, André Homeyer		
Work Package No:	5		
Estimated delivery date:	August 31, 2015	**Actual delivery date:**	August 31, 2015
Nature:	**Please choose one:** Report (R)		
Dissemination level:	**Please choose one**: Public (PU)		

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 601040. This information reflects only the authors views and the European Union is not liable for any use that may be made of the information contained therein.

INTRODUCTION

In Task 5.4 quantitative features have been developed to extract meaningful information from the regions defined in Task 5.3. These features serve as an important source of information for correlating histopathology with multimodal radiological imaging, which is the topic of research in WP8 (Task 8.1).

TASK 5.4: QUANTITATIVE DESCRIPTION AND CLASSIFICATION OF MICROSTRUCTURE

A pathologist uses a diversity of visual clues for reaching a diagnosis. Important visual information consists of nuclear size and shape, tissue architecture, presence of mitoses, necrosis, and many other sources of information. Some of features used by the pathologist to draw diagnosis are, however, not even fully known to the observer themselves but have been acquired by years of looking at microscopic tissue sections. Task 5.4 defines a set of quantitative features that can be accurately and objectively extracted from scanned tissue sections. In section 5.4.1 we focus on the extraction of features describing individual cell nuclei (i.e. size and shape). These may be used directly by calculating statistics over a larger number of cells, or may be used to classify cells into distinct classes (e.g. fibroblasts, lymphocytes, tumour cells). In section 5.4.2 we derive quantitative features describing the tissue architecture in terms of the spatial layout of nuclei. This is especially meaningful in (pre)malignant regions, as it expresses the level of differentiation of a tumour which is an important diagnostic feature. Lastly, section 5.4.3 focuses on a specific diagnostic feature, which has gained huge interest in the last years: presence of lymphocytic infiltrate within or adjacent to the tumour. It was shown in multiple studies that the extent to which a tumour succeeds in downregulating the hosts' immune reaction is an important prognostic factor. These features will serve as an important source of information for correlating histopathology with multimodal radiological imaging, which is the topic of research in WP8 (Task 8.1).

5.4.1 Quantitative nuclear features

Based on nuclear masks we can extract morphometric features describing the geometry (shape, size, position and boundary) of the nucleus, resembling the assessment of nuclear pleomorphism by a pathologist. Moreover, several texture features can be extracted which describe the spatial variation of gray-levels within the nucleus. Such features can be broadly classified into statistical features and structural features and have shown strong discriminatory power for characterising tumour cells among normal ones. Some of the most widely used texture analysis methods are Haralick texture features from Co-occurrence matrix, Statistics of grey-scale histogram, Local binary patterns, and features from filter banks (Gabor, Laplacian of Gaussian, etc.).

All these features have been developed and will be extracted from the segmented nuclei. Several statistics can then be calculated to describe a population of cell nuclei in a particular ROI. The extracted features can also be used for cell nuclei classification.

5.4.2 Graph based analysis to describe spatial layout of nuclei

An important factor to consider when studying the characteristics of a neoplasm is the level of differentiation, which may be assessed by studying nuclear pleomorphism (see above) but also by assessment of the architecture of the tissue. In the widely used 'modified Bloom Richardson breast

cancer grading', for instance, the level of tubule formation is one of three prognostic factors. Tissue architecture may be studied applying techniques from mathematical graph analysis. By first identifying centres of cell nuclei and next constructing a graph consisting of edges connecting neighbouring nuclei applying strict criteria, a large number of quantitative features describing the spatial layout of cells may be derived. A number of different graphs (e.g. Voronoi, Delaunay, Gabriel, minimum spanning tree) may be constructed which all allow different characteristics to be extracted. Figure 1 shows an example of an immune-fluorescent double staining of a breast cancer tissue section.

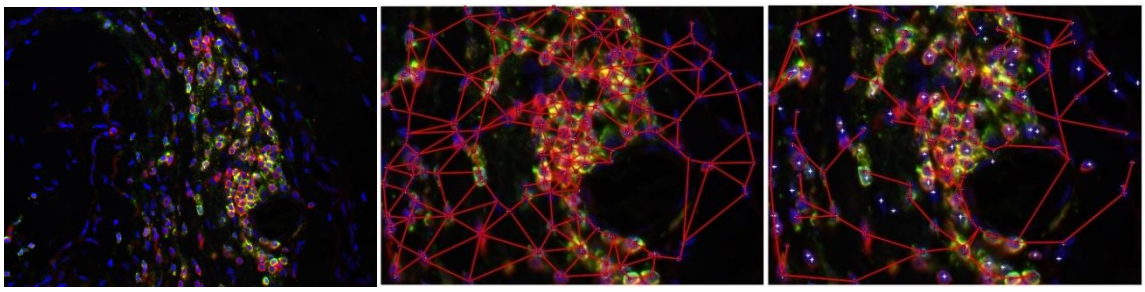

Figure 1: a. Immunofluorescence stained breast cancer tissue section with nuclei in blue, CD3 in red and CD45RO in green. b. So-called Gabriel graph describes spatial layout of cells. c. Minimum spanning tree, showing the shortest path connecting all cell nuclei.

We recently implemented a series of almost 30 graph based features describing the average number of neighbours of a cell and the average distance between a cell and its neighbours, the area of influence of cells and the stability of the graph structure. These algorithms may be used to assess the overall tissue architecture, but also to quantify the relationship between different subpopulations of cells (e.g. whether CD3 positive cells are clustered in groups or dispersed throughout the tumour).

5.4.3: Location-specific quantification of lymphocytic infiltrates

The extent in which a tumour succeeds in downregulating the host's immune reaction is an important prognostic factor. In multiple studies, it has been shown that the presence of lymphocytic infiltrates within or adjacent to the tumour correlates with patient survival and outcome (Alexe et al. 2007).

In Hematoxylin-and-Eosin-stained (H&E) histological images, lymphocyte nuclei appear as small roundish spots of an intense dark purple color. The lymphocyte density, that is, the number of lymphocytes per area, is a common quantitative measure for the amount of lymphocytic infiltration.

Since the manual determination of the lymphocyte density requires the counting of large numbers of lymphocytes, it is a very time consuming process. When the lymphocyte density is supposed to be assessed at multiple locations, like within or adjacent to the tumour, the manual counting quickly becomes infeasible. To solve this problem, we have combined image analysis methods developed in Tasks 5.2 and 5.3 in order to enable the automatic location-specific quantification of lymphocytes.

In Task 5.4, we have adapted the nuclei detection method developed in Task 5.3 to lymphocyte nuclei. For this, we have manually annotated thousands of example pixels within and outside of lymphocyte nuclei. The annotations were directly performed in whole-slide images of breast tissue in order to capture the variability of appearances to the greatest extent.

A main component of the nuclei detection method is a machine-learning algorithm which generates a nucleus probability map. By training this machine-learning algorithm with the example pixels, the

algorithm was adapted to produce a probability map for the presence of lymphocyte nuclei. Afterwards, an optimized region extraction and merging algorithm was applied in order to delineate the individual lymphocyte nuclei.

We have evaluated the resulting lymphocyte detection method on 10 images of H&E-stained breast tumour tissue. The images showed individual field-of-views at 0.5 micron/pixel resolution. For every image, ground truth annotations of the depicted lymphocytes were provided by an expert pathologist. In total, 3645 lymphocytes were annotated. Compared with the ground truth annotations, the analysis results produced by the lymphocyte detection method proved to be accurate and achieved precision and sensitivity values of 79% and 87%, respectively. Fig. 2 shows an example detection result by our proposed algorithm.

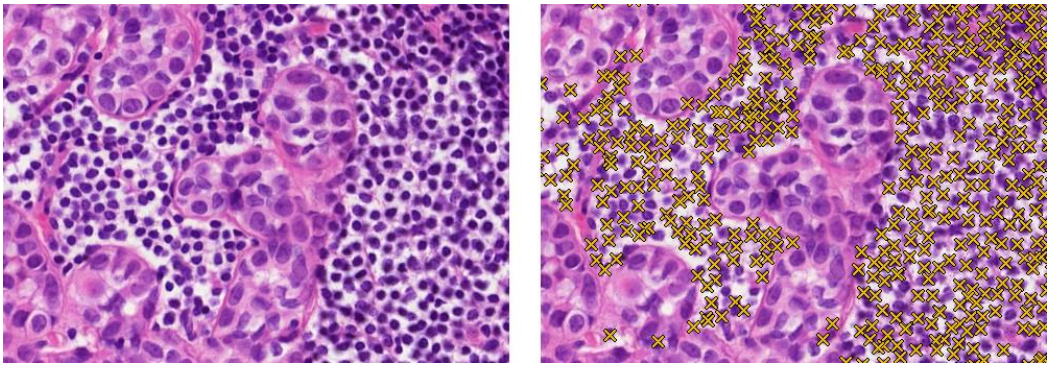

Figure 2: Detected lymphocytes in a ductal carcinoma in situ.

In order to enable the location-specific quantification of lymphocytes, we have integrated the lymphocyte detection method with a region segmentation method developed in Task 5.2. The region segmentation method enables the automatic segmentation of breast tissue images into potential tumour tissue, stroma, and other regions. For this, the region segmentation method divides the image into a lattice of tiles. Every tile is classified as either potentially tumour, stroma, other tissue or background with respect to color statistics and texture features extracted at different scales.

We have integrated the C++ implementations of the lymphocyte detection and region segmentation methods, so that the lymphocyte detection is performed for every tile individually. Since we're only interested in lymphocytic infiltrates within or adjacent to a tumour, the lymphocyte detection is only performed in tiles classified as potentially tumour or stroma. By skipping other tissue and background tiles we save a considerable amount of computation time.

The tile size is chosen to be large enough to be statistically meaningful for the computation of the lymphocyte density and small enough to enable a high spatial resolution of tiles. As a result, we have settled on a tile size of 0.064x0.064 mm, that is, 128x128 pixels at 0.5 micron/pixel resolution.

All analysis results are automatically stored in a unified relational data model in the open-source sqlite3 database. This includes the positions of the detected lymphocyte nuclei as well as the bounds and classifications of the individual tiles. Locally-connected tiles that are classified as potentially tumour are represented as individual tumour regions in the data model. In this manner, it becomes possible to quantify the lymphocytic infiltration within or in the vicinity of tumour regions by simple database queries.

We have implemented the method for the location-specific quantification of lymphocytic infiltrates into the graphical software developed in Task 5.1. This software enables the application of the method to arbitrary slide image files and the interactive evaluation of the results. The software

features a simple user interface, in which the user can pan and zoom through digital slides, just like with an ordinary microscope. A visualization of the analysis result is displayed as a semi-transparent overlay on top of the original image.

The visualization shows a heat map of the lymphocytic densities of the individual tiles (see Fig. 3). In addition, the identified tumour regions are delineated by lines. In this manner, it becomes possible to assess spatial patterns of lymphocytic infiltration within tumour and the heterogeneity of lymphocytic infiltration across the tissue. The software also allows the user to interactively query the lymphocyte density values of individual tumour regions by clicking on them with the mouse.

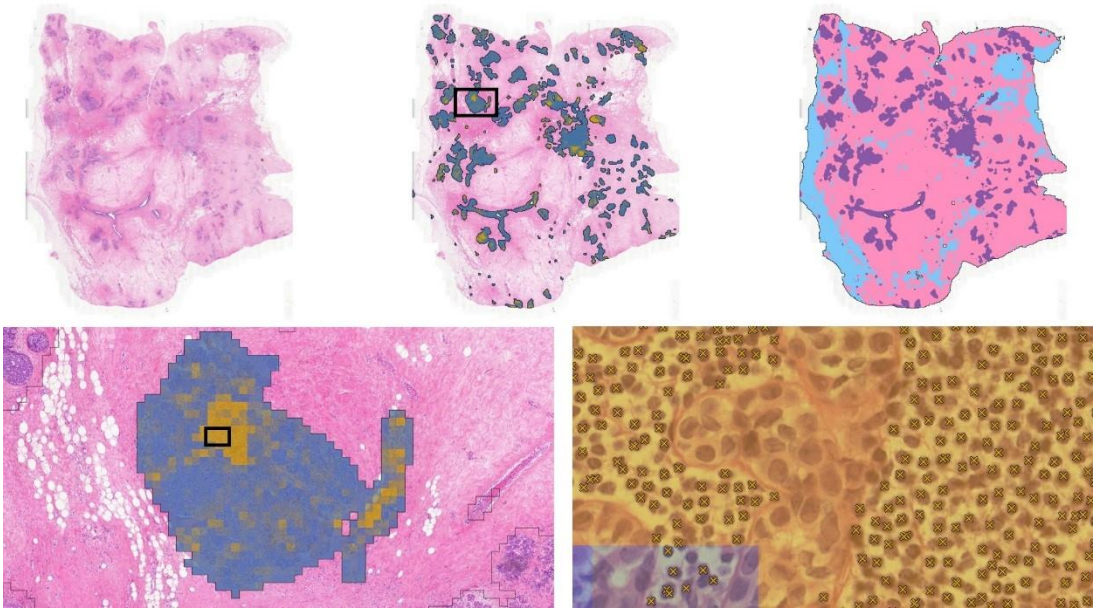

Figure 3: Location-specific quantification of lymphocytic infiltrates across entire tissue sections

References

Alexe, Gabriela, Gul S. Dalgin, Daniel Scanfeld, Pablo Tamayo, Jill P. Mesirov, Charles DeLisi, Lyndsay Harris, et al. 2007. “High Expression of Lymphocyte-Associated Genes in Node-Negative HER2+ Breast Cancers Correlates with Lower Recurrence Rates.” Cancer Research 67 (22) (November): 10669–10676.