P27

A Versatile Automated Detection System for Erythrocytes in Human Tissue Sections

A. Heindl[1,2], T. Thalhammer[1], G. Bises[1], M. Schepelmann[1,3], R. Rogojanu[1,4], R. Ecker[4], A. Seewald[2], I. Ellinger[1]

[1]Department of Pathophysiology and Allergy Research, Center for Pathophysiology, Immunology and Infectiology, Medical University of Vienna, Austria; [2]Seewald Solutions, Vienna, Austria; [3]Division of Pathophysiology and Repair, School of Biosciences, Cardiff University, Cardiff, United Kingdom; [4]TissueGnostics GmbH, Vienna, Austria

Introduction

Microscopic analysis of biological tissue samples is indispensable in medical research and diagnosis. However, the visual evaluation of microscopic images by the human analyst is biased and measures remain semi-quantitative and are difficult to reproduce. As a consequence, automated (computer-based) methods evolved for a robust and reliable quantitative evaluation of images which are based on automated segmentation of the tissue into functional cell groups and subsequent extraction of associated parameters such as area or protein expression levels. Advantages of automated image-analysis (AIA) are speed, reproducibility and processing of large amount of data in short time. Since environmental influences combine the genotype with the result in individual cellular phenotypes, single-cell AIA helps to collect a maximum of molecular information to reflect the full heterogeneity of cellular systems (organs). This cellular heterogeneity contains not only diagnostic information. AIA can help to extract predictive information on the future disease course in individual patients and therefore is an essential tool for a predictive medicine.

In case AIA is performed on immunofluorescence microscopic images, tissue-associated auto-fluorescent objects can interfere significantly with automated segmentation and quantification of fluorescence-labeled target molecules or structures. An example are red blood cells (erythrocytes), which can be found in many mammalian tissue sections, and which exhibit high levels of auto-fluorescence (see Fig.1-3). Two strategies can be considered to remove auto-fluorescence from tissue preparations: (1) Elimination/reduction of auto-fluorescence during tissue processing by chemical treatments [1] or (2) *in silico* elimination of auto-fluorescence during automated analysis following automated recognition of the auto-fluorescent structures.

Aim

We aimed to develop and validate a machine-learning based computer algorithm that identifies reliably erythrocytes in human tissue sections.

Results

We have developed and currently validate a machine-learning based computer algorithm (EryQuest) that identifies erythrocytes in human tissue sections .

Proposed applications of this algorithm include:

1/ Detection and *in silico* elimination of the erythrocyte-associated auto-fluorescence in the given image as a prerequisite for subsequent erythrocyte-independent tissue-associated parameter analysis.

2/ Detection and quantification of erythrocytes-covered area or associated auto-fluorescence levels for future diagnostic purposes. Examples for applications are: detection of changes in erythrocyte-associated protoporphyrin fluorescence that might serve as marker for e.g. diabetes [2] or may accompany dysplastic progression [3]. Erythrocyte accumulations may represent the first step of vascular pathology in cerebral small vessel disease (CMVD) and might be a promising target for implementing prophylactic and therapeutic strategies in human CMVD [4].

Summary

We have developed an approach for automated detection of erythrocytes in human tissue sections. The presented method identifies erythrocytes in transmission images without the need for a specific erythrocyte staining. Currently, the performance asymptotically converges to human performance.

Literature

[1] Baschong W, Suetterlin R, Laeng RH, J Histochem Cytochem. 2001 Dec;49(12): 1565-72
[2] Fauaz G, Miranda AR, Gomes CZ, Courrol LC, Silva FR, Rocha FG, Schor N, Bellini MH. Appl Spectrosc. 2010 Apr;64(4):391-5.
[3] Gomes de Góes Rocha FG, Chaves KC, Gomes CZ, Campanharo CB, Courrol LC, Schor N, Bellini MH. J Fluoresc. 2010 Nov;20(6):1225-31. Epub 2010 May 18.
[4] Schreiber et al, J Cereb Blood Flow Metab. 2011 Aug 31. doi:10.1038/jcbfm.2011.122
[5] Lienhart et al, IEEE ICIP 2002, An Extended Set of Haar-Like Features for Rapid Object Det.

Automated Image Analysis Approaches

Preparation of virtual slides

Paraffin-embedded sections from human placenta, colon, prostate, lung and ovary were processed for immunofluorescence microscopy to detect the proteins of interest. Fig.1 shows the section of a human placenta stained with antibodies against cytokeratin 7 (shown in white/Cy5, see also Fig.3), a receptor protein (RAGE, shown in red/cy3, see also Fig.2) as well as DAPI to detect nuclei (blue). In these fluorescence channels (Cy3 and Cy5) erythrocytes in fetal vessels (yellow arrows) and maternal circulation (orange arrows) exhibited high levels of autofluorescence.

The tissue was scanned and virtual images were produced (Fig.5a), using an automated microscopy system (TissueFAXS, TissueGnostics GmbH). As the designed algorithm for erythrocyte-detection should make use of the characteristic structure/texture of erythrocytes, transmission images were acquired (Fig. 5b) along with images of the fluorescence channels of interest.

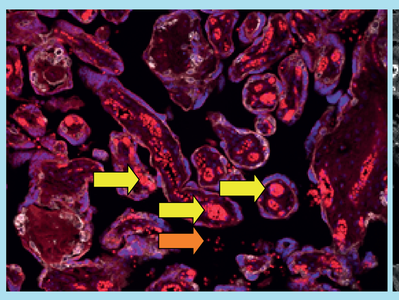

Figure 1

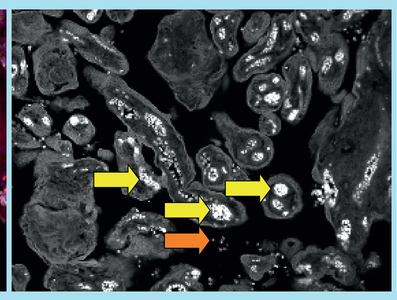

Figure 2

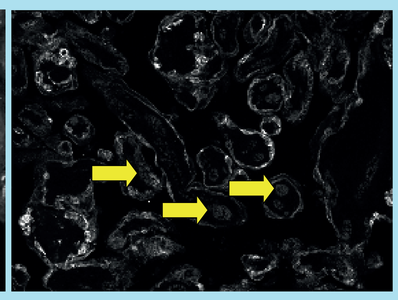

Figure 3

Design of EryQuest

The **classifier is** built of a **cascade of boosted decision trees (Fig. 4) using Haar-like features [5]**. Each weak classifier of the cascade represents a boosted decision tree that rejects a small fraction of non-erythrocytes. Samples that are not rejected after the n^{th} level are classified as an erythrocyte. Haar-like features encode the existence of oriented contrasts between regions in the acquired images and can be computed efficiently by employing integral images.

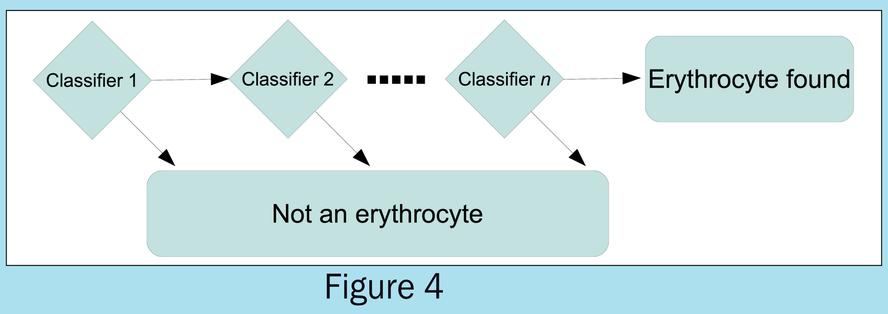

Figure 4

For the **training of the classifier cascade, 3764 positive samples** (Fig. 5c) containing erythrocytes as well as **2000 negative samples** (not shown) were cut out of images. To build the cascade OpenCV's haartraining was used (Fig. 5d) which implements the Adaboost technique and results in an xml cascade file that is later used for erythrocyte recognition. The training step requires features therefore an extended set of Haar-like features ([4], Fig. 5d) was employed.

Acknowledgement This project is supported by the FFG Bridge-Project [818094].

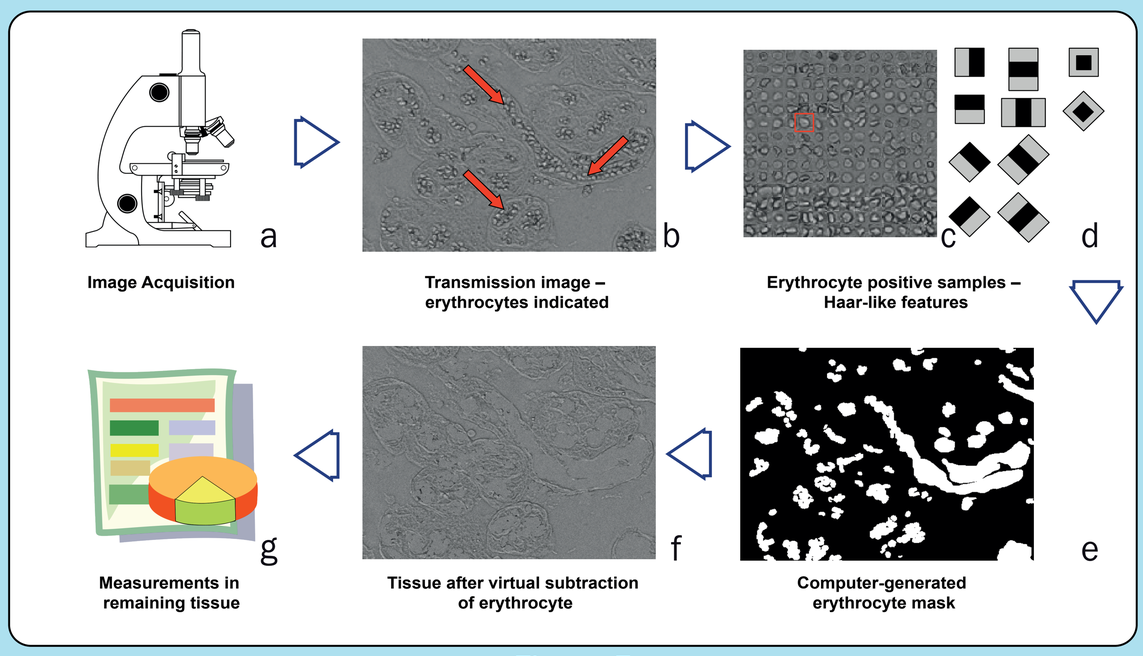

Figure 5

During haartraining, these rectangles are moved in different scales over the full input image and the difference of the sum of intensity values of both sub-rectangles (sum of gray minus sum black) is computed. This process results in a huge set of features even for small rectangles. Each of this subregion is then classified as erythrocyte or non-erythrocyte area. The result of this step is a binary output mask (Fig. 5e) for example which can then be used to virtually eliminate erythrocytes from the image (Fig. 5f). Additionally, it is also possible to compute various measurements (Fig. 5g) such as area covered by the detected erythrocytes for diagnostic/predictive applications.

Validation of EryQuest

Evaluation of the system was done on pixel level by comparing expert markups (generated by 2 experts) with the result mask of EryQuest and computing the F1 score. Figure 6 illustrates these results in various tissue sections tested. The blue bar represents the mean F1 score of the human experts (n>=2) grouped by tissue whereas the red bar shows the corresponding performance of EryQuest.

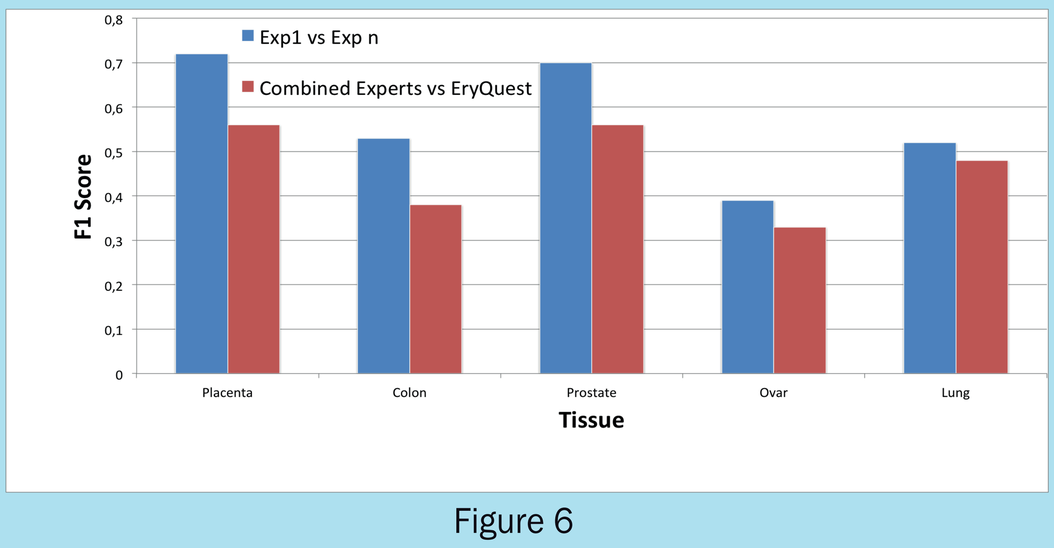

Figure 6