Automated REcognition of Tissue-associated Erythrocytes (ARETE)
A new Tool in Tissue-Cytometry

Heindl, A. [1], Seewald, AK. [2], Thalhammer, T. [1],Bises, G. [1],
Dekan, S. [3], Mesteri, I. [3], Rogojanu, R. [4], Ellinger, I. [1]

[1]Department for Pathophysiology and Allergy Research; Medical University Vienna, [2]Seewald Solutions; Vienna, Austria,
[3]Department of Clinical Pathology; Medical University Vienna, [4]TissueGnostics GmbH, Vienna, Austria

Introduction

Tissue cytometry is automated analysis of microscopic images derived from tissue sections. When based on immunofluorescence microscopy, localization and quantitative analysis of multiple molecules and parameters are enabled on a single cell basis in the tissue context. This supports dissection of cell-specific functions while taking into account the strong influence of intact cell-cell and cell-matrix interaction on cellular behavior. Automated analysis of large tissue-areas enables quantification with statistical significance and consequently tissue cytometry is an essential tool not only for research, but also for diagnostic, therapeutic and predictive medicine.

In fluorescence microscopic analysis of tissue samples, auto-fluorescent objects such as erythrocytes can interfere with automated quantification of immunofluorescence labeled targets. Erythrocytes are present in almost any tissue even after intense washing steps prior to tissue fixation. This is exemplified in Figure 1a. Multiple erythrocytes (dark grey structures, indicated by asterisks) are visible in a transmission image of human placental chorionic tissue. As shown for a subregion (1b), erythrocytes exhibit strong autofluorescence (asterisks, Figure 1c-e) and are found in proximity to proteins under investigation such as cytokeratin 7 (CK7; 1c), receptor for advanced glycation end product (RAGE; 1d) or CK7+ γ-smooth muscle-actin (1e). Arrows indicate these target stainings.

The erythrocyte-associated fluorescence intensity values often significantly exceed background threshold values. Thus, they might be mistaken and measured as specific fluorescence. In automated image-analysis, this influences image segmentation and/or falsifies the quantification of the target proteins or cells (see application). Therefore, erythrocyte auto-fluorescence represents a source of quantification errors in automated image-analysis that may lead to incorrect scientific conclusions.

Aim

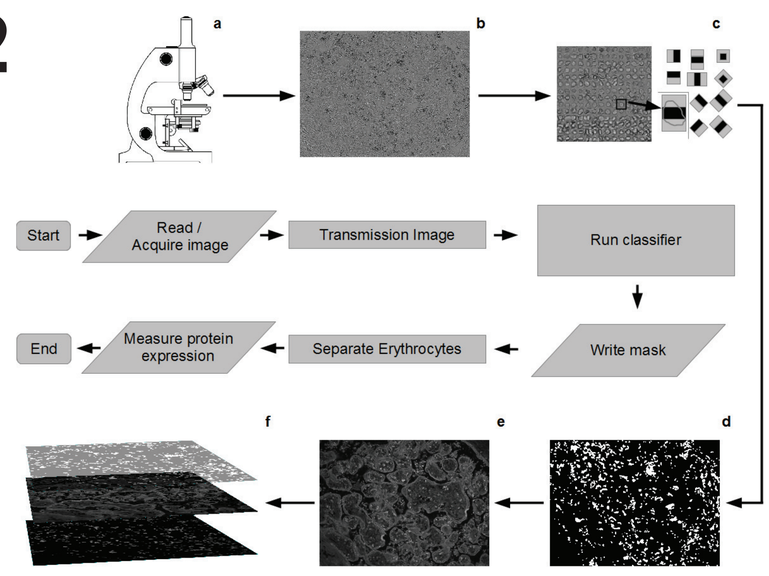
Develop and validate the system ARETE (Automated REcognition of Tissue associated Erythrocytes), which identifies tissue-associated erythrocytes in (co-acquired) transmission images utilizing their distinctive shape and structure. ARETE enables automated *in-silico* exclusion of erythrocyte-associated auto-fluorescence from fluorescence image evaluation of tissue sections.

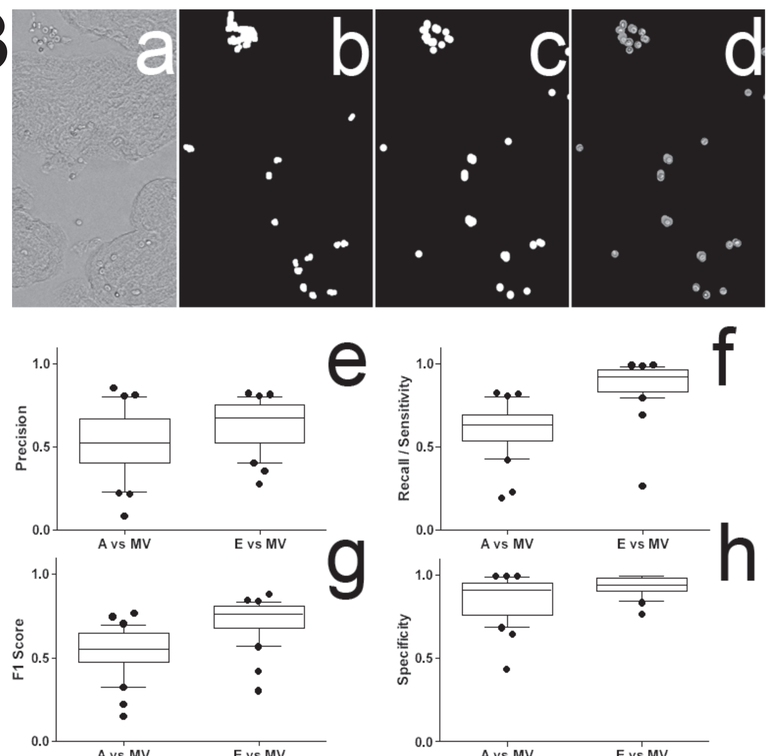
Methods

Formalin-fixed paraffin embedded human tissue (placentas, grade 2 colorectal cancer) was sectioned (4 µm), deparaffinized and rehydrated. Following antigen retrieval, permeabilization and blocking, placenta sections were incubated with specific primary (CK7, RAGE or gsm-actin) and fluorochrome-conjugated secondary antibodies (Figure 1). Colon sections were stained with primary antibodies against Ki-67 and human keratin 8 (CK8) as well as fluorochrome-conjugated secondary antibodies. Nuclei were visualized with Dapi. Tissue sections were acquired with TissueFAXSplus tissue cytometer (TissueGnostics GmbH, Vienna, Austria) equipped with epifluorescence optics. A monochrome camera was used to acquire grayscale images of all samples using extended focus, an integrated feature of the software. Fluorescence and corresponding transmission images were stored with lossless compression in PNG format. Cell segmentation of colon sections in order to identify single epithelial (CK8+) cells positive for Ki-67 staining was done by TissueQuest software version 4.0 (TissueGnostics GmbH).

Results

1. The workflow of ARETE is depicted in Figure 2. Tissue sections are subjected to immunofluorescence microscopy to detect proteins of interest. Images are acquired using an automated microscope (2a), resulting in a set of fluorescence images and the corresponding transmission image (2b). ARETE employs a trained cascade of boosted decision trees of Haar-like features (2c), a machine-learning approach formerly mainly used for pedestrian detection and face recognition. To train this cascade, 8460 positive samples (containing erythrocytes) and 4000 negative samples (areas without erythrocytes) were employed. The cascade of weak classifiers ensures fast execution the classifier must be run on every possible subwindow of the image at multiple resolutions while the boosting (Adaboot) step ensures sufficiently good performance of the ensemble of weak classifiers at each cascade level. The Haar-like features consist of small rectangles that can be efficiently computed by employing integral images (2c). Each pixel located at position (x,y) in an integral image contains the sum of the pixel values above and to the left of (x,y) thus reduces the effort for feature calculation (i.e. differences of arbitrary rectangle sums) to a simple addition of a few values. By using Haar-like features, several hundreds of thousand features can be extracted from small subwindows (15x15 pixel) allowing a reliable extraction even if illumination gradients are present in an image. The output is a binary image indicating erythrocyte positions (2d), which can be combined with the co-acquired fluorescence images to analyse protein expression patterns (2e-f).

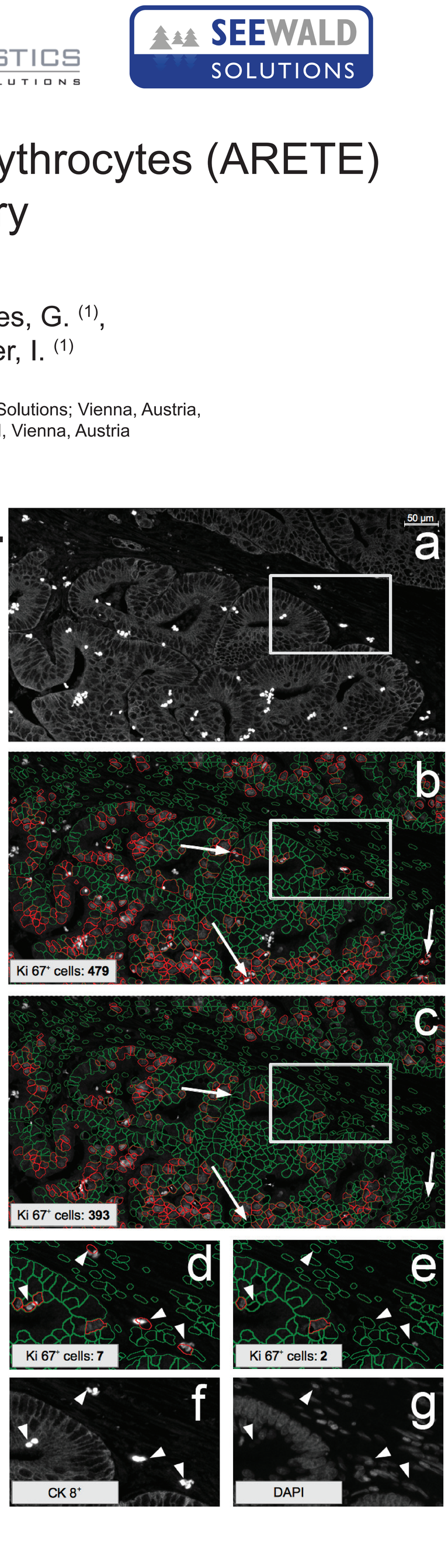




2. Validation of ARETE (Figure 3) was done on pixel level due to the small shape and huge number of erythrocytes. Human experts examined transmission images (n=60; 3a) and indicated the localization of erythrocytes (3b). Human-derived masks were combined using majority voting and compared pixelwise against a mask computed by ARETE (3c). Precision (3e), recall (3f), F1 score (a harmonic mean of precision and recall; 3g), specificity (3h) and Cohen's κ coefficient ([1]; human expert 0.68 vs 0.41 ARETE) were calculated from these comparisons.

The medians obtained from the comparison of ARETE with the majority voting of the experts were lower than the comparison of single human experts with the majority voting of the experts (3e-h), which can be partly explained by the approximation of the erythrocyte shape by circles, which increases dramatically the speed of ARETE. However, overlapping distributions suggested a performance of ARETE close to the human expert. Indeed, when assessed via Cohen's κ coefficient and the scale of Fleiss [2], both ARETE vs. human experts agreement and human interrater agreement were classified to be in the same category (fair to good).

3. An application of ARETE is depicted in Figure 4, which demonstrates the effect of *in-silico* removal of erythrocytes by ARETE on the automated detection of CK8/Ki-67 double positive cells in human colon cancer tissue. Such cells are of interest in analysis of the metastatic potential of a tumor. Figure 4a shows CK8-positive epithelial cells in the colon tissue section. Automated image segmentation based on Dapi staining outlined all cells in green, those expressing CK8 as well as the proliferation marker Ki-67 in red (4b and 4c). Auto-fluorescent erythrocytes were not removed in 4b, resulting in 479 CK7/Ki-67 positive cells. Application of ARETE prior to image segmentation (4c) reduced the amount of Ki-67 positive cells by 18% to 393. Arrows (4b,4c) indicate, where cell touching erythrocytes caused false positive Ki-67 cells. A subregion is magnified in (4d/ARETE) and (4e/+ARETE) with the corresponding fluorescence images demonstrating CK8 (4f) and Dapi (4g) staining. In this subregion, the arrowheads point out false positive cells (n=5) caused by erythrocyte associated auto-fluorescence.



Conclusion

ARETE is a rapidly trainable, fast and autonomous system for erythrocyte detection in tissue transmission images that correlates well with human perception. ARETE is supposed to be used in tissue cytometry to improve results of tissue segmentation and protein as well as cell quantification.

References

[1] J. Cohen: A coefficient of agreement for nominal scales. In:Educational and Psychological Measurement. 20,1960, 37-46.
[2] J. L. Fleiss: The measurement of interrater agreement. In: ders., Statistical methods for rates and proportions. 2. Issue. John Wiley & Sons, New York 1981, S. 212-236, Chapter 13

Acknowledgements

This project is funded by FFG (Bridge 818094)