Spotting the difference: rare cell characterization by multiplex mRNA *in situ* hybridization and machine learning in the context of liquid biopsy-based therapy resistance monitoring

Lilli Bonstingl[1]
Katja Sallinger[2], Christina Skofler[3], Christine Ulz[3], Margret Zinnegger[4], Elisabeth Pritz[2], Karin Pankratz[2], Corinna Odar[2], Michael Gruber[2], Laurin Herbsthofer[4], Thomas Bauernhofer[5], Thomas Kroneis[2] and Amin El-Heliebi[2]
[1] Division of Cell Biology, Histology and Embryology, Gottfried Schatz Research Center, Medical University of Graz, Austria; Center for Biomarker Research in Medicine (CBmed), Graz, Austria
[2] Division of Cell Biology, Histology and Embryology, Gottfried Schatz Research Center, Medical University of Graz, Austria
[3] Center for Biomarker Research in Medicine (CBmed), Graz, Austria; Diagnostic and Research Center for Molecular BioMedicine, Diagnostic and Research Institute of Pathology, Medical University of Graz, Austria
[4] Center for Biomarker Research in Medicine (CBmed), Graz, Austria
[5] Division of Oncology, Department of Internal Medicine, Medical University of Graz, Austria

Background & objectives

In situ padlock probe technology can be used to enumerate circulating tumor cells (CTCs) and detect mRNA-based resistance markers. Until now, the main limitations of this approach were time-intensive manual evaluation of *in situ* data and meeting the high sensitivity and specificity-requirements entailed by rare cell analyses. To boost sensitivity, specificity, and informative value, we developed a novel multiplex *in situ* approach to visualize CTC and hematopoietic markers and implemented automated image analysis combined with machine learning-based cell classification to increase feasibility.

Methods

We visualized a panel of epithelial, cancer-specific, neuroendocrine, mesenchymal, and hematopoietic mRNA markers using *in situ* hybridization and a customized CellProfiler pipeline for image analysis. Based on peripheral blood mononuclear cells (PBMCs) of healthy donors and two prostate cancer (PC) cell lines, we trained a random forest classifier to discriminate CTCs from hematopoietic cells with supervised machine learning in CellProfiler Analyst. Finally, we isolated CTCs with the Smart Biopsy Cell Isolator (CytoGen, KR) and analyzed CTC enriched blood samples of patients with advanced metastatic PC. For manual revision of the cell classification, we implemented an interactive image gallery.

Results

In the dataset of healthy controls' PBMCs and PC cell lines, the classification based on *in situ* data reached a sensitivity of 0.94 and precision of 0.96 for CTCs. In the patient samples, CTCs were identified with a sensitivity of 0.67-0.75 and precision of 0.05-0.27. False-positive CTCs were easily identified and removed during subsequent expert revision in the interactive image gallery. The workflow successfully identified CTCs with epithelial, mesenchymal, and/or neuroendocrine markers, revealed intra-patient CTC heterogeneity, and captured both single CTCs and CTC clusters.

Conclusion

Our approach improves the robustness, feasibility, and informative value of CTC analysis and can be used for longitudinal monitoring of therapy resistance in cancer patients. While low precision is readily counteracted by our interactive image gallery, which allows for fast and easy browsing over candidate cells and removal of false-positive CTCs, sensitivity must be increased by including additional cell lines and blood samples in the training dataset.