Comparison of open-source data batch normalization tools for the use on imaging mass cytometry data

Henrike Salié[1], Ana Marta Sequeira[2,3], Marieke Ijsselsteijn[2], Felix Röttele[1], Yuan Suo[1], Lara Wischer[1], Thomas Longerich[4], Noel F. C. C. de Miranda[2], Bertram Bengsch[1,5,6]

[1]Clinic for Internal Medicine II, University Medical Center Freiburg, Freiburg, Germany.

[2]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands.

[3]CEB-Centre Biological Engineering, University of Minho, 4710-057 Braga, Portugal.

[4]Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany.

[5]Signalling Research Centres BIOSS and CIBSS, University of Freiburg, Freiburg, Germany.

[6]German Cancer Consortium (DKTK), Partner Site Freiburg, Germany

Introduction

Batch effects can be a significant hurdle for high-dimensional imaging mass cytometry (IMC) data analysis due to difficulty to discriminate procedural and experimental variability from biological differences. Batch-effect correction strategies aim to reduce non-biological sample-specific signals and improve overall data quality and may operate on the segmented single-cell or pixel level data. So far no formal comparison on IMC data between published methods exists.

Methods

We performed imaging mass cytometry with subsequent cell segmentation on 12 hepatocellular carcinoma (HCC) patients. To reduce batch-effects between single patients, semi-automated background removal, image-based channel normalization and single-cell data based batch correction tools (fastMNN, harmony and Seurat) were performed separately. After phonograph clustering, we compared the performance of these approaches against each other and non-corrected data based on the ability to identify expected cell types in the HCC and liver microenvironment, their numeric distribution as well as patient specificity of clusters.

Results

Application of batch correction tools led to a reduction of sample-specific clusters. Most expected cell types were identified after fastMNN, harmony, channel normalization and semi-automated background removal. We observed varying numbers of identified CD8 T cells in some patients with dense immune infiltrates. Inferring test quality criteria from ground truth comparison allowed for a comparison of sensitivity and specificity for CD8 T cell detection, with a favourable balance observed after image-based channel normalization and harmony.

Conclusion

Batch correction may enhance IMC data performance by limiting non-biological patient-specific variability and ensuring robust cell type detection using clustering algorithms in tissue microenvironments. Comparing approaches with predefined endpoints allows for optimal decision-making regarding the application of batch correction and normalization tools. Our work suggests that both strategies to normalize batch effects on a primary data level and on a post-segmentation level can be successfully applied.