DECONFOUNDING MICROARRAY ANALYSIS: INDEPENDENT MEASUREMENTS OF CELL TYPE PROPORTIONS USED IN A REGRESSION MODEL TO RESOLVE TISSUE HETEROGENEITY BIAS

A. Ziegler[1], M. Jacobsen[2], D. Repsilber[1,3], A. Gutschmidt[2], A. Neher[4], K. Feldmann[4], H. J. Mollenkopf[2], S. H. E. Kaufmann[2]

[1]*University at Lübeck, Germany*
[2]*Max Planck Institut for Infection Biology, Berlin, Germany*
[3]*University Potsdam, Germany*
[4]*Asklepios Center for Respiratory Medicine and Thoracic Surgery, Munich-Gauting, Germany*

Email: *ziegler@imbs.uni-luebeck.de*

A major limitation of microarray analysis is caused by heterogeneity in the cellular composition of tissue specimens, which frequently confounds the data. We introduce a linear model to deconfound gene expression data from tissue heterogeneity for genes exclusively expressed by a single cell type using highly exact determination of cellular compositions by flowcytometry. Gene expression data are obtained in parallel by real time quantitative polymerase chain reaction (qPCR) and microarray analyses. Verification of deconfounding is done using protein quantification for the respective genes.

Quantification of cell type proportions for peripheral blood mononuclear cells from tuberculosis patients and controls revealed differences in B cell and monocyte proportions between both study groups, and thus heterogeneity for the tissue under investigation. Gene expression analyses reflected these differences in cell-type distribution showing differential expression of B cell and monocyte population specific markers. Fitting an appropriate linear model allowed us to deconfound measured transcriptome levels from tissue heterogeneity effects. In the case of monocytes, additional differential expression on the single cell level could be proposed. Protein quantification verified these deconfounded results.

Deconfounding of transcriptome analyses for cellular heterogeneity greatly improves interpretability, and hence the validity of transcriptome profiling results.