Helmut-Schmidt-Universität, Postfach 700822, 22008 Hamburg

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

# Student/Master Project:
# ALRA + HarmonizR coupling for dealing correctly with scRNA seq data (+ ComBat-seq)

*Description:* To account for internal biases and shifts in measurements within biomedical data, the HarmonizR framework [1] has been developed to adjust for these so-called 'batch effects' while tolerating missing data points. Single cell RNA-sequencing data naturally comes with a lot of values marked with zero, usually treated as missing. Some of these are actual, biological zeroes and others are filled in due to technical measurement limitations; technical zeroes. The ALRA algorithm [2] was developed to differ between these two types. Coupling the HarmonizR with ALRA can potentially improve the amount of conserved, meaningful data within these datasets and properly adjust it. Additionally, a ComBat-seq version of the underlying ComBat adjustment algorithm is available and could benefit the single cell RNA-seq adjustment pipeline.

*Prerequisites:* Knowledge about the R programming language. Furthermore, interest in working with biological and medical data would be beneficial, but is not mandatory.

*Contact:* Chair for High Performance Computing
Simon Schlumbohm, schlumbohm@hsu-hh.de
Prof. Dr. Philipp Neumann, philipp.neumann@hsu-hh.de

*References:*

[1] Voß, H., Schlumbohm, S., Barwikowski, P. et al. HarmonizR enables data harmonization across independent proteomic datasets with appropriate handling of missing values. Nat Commun 13, 3523 (2022). https://doi.org/10.1038/s41467-022-31007-x

[2] Linderman, G.C. et al. Zero-preserving imputation of single-cell RNA-seq data. Nat Commun 13, (2022). https://doi.org/10.1038/s41467-021-27729-z