A U S H A N G

FREIE UNIVERSITÄT BERLIN

Fachbereich Mathematik und Informatik

Promotionsbüro, Arnimallee 14, 14195 Berlin

D I S P U T A T I O N

Freitag, 5. Juli 2024, 9:30 Uhr

Ort: WebEx

Disputation über die Doktorarbeit von

Guido Pacini

Thema der Dissertation:

Transcriptome regulation during the X chromosome inactivation process

Thema der Disputation:

Integration analysis methods for single-cell RNA-sequencing data

Die Arbeit wurde unter der Betreuung von **Prof. Dr. A. Marsico** durchgeführt.

Abstract: Recent technological advances considerably increased both the complexity and throughput of single-cell RNA-sequencing experiments, that are currently capable of profiling the transcriptomes of thousands of cells at a time. Combining multiple experiments enhances the comprehension of the biological process under investigation, and enables the exploration of lowly abundant cell types or states that would otherwise be difficult to identify. Nonetheless, dataset integration comes at a cost of strong batch effects caused by the combination of experiments collected from different samples, donors, sequencing time point and sequencing technologies. These effects result in large variability of gene expression levels across the combined datasets that strongly affects downstream analyses. Adopting effective integration analysis procedures is therefore crucial to remove such technical effects while preserving the biological variability across the batches, enabling the unbiased and comprehensive analysis of merged datasets. Integration analysis methods developed for microarray and bulk RNA-sequencing data analysis employ genewise linear regression models to estimate and regress out the difference in average gene expression between the batches. This approach assumes an identical composition across batches, and that the observed difference in gene expression levels is purely technical. Importantly, both these assumptions are not generally met in single-cell RNA-sequencing data. Indeed the population composition usually differs across such independent experiments, while any difference in cell type abundance leads to biological variations of the gene expression levels. Employing gene-wise regression procedures to remove batch effects in single-cell RNA-sequencing datasets would thus yield inaccurate results and affect all downstream analyses. To overcome this issue, integration procedures specifically designed for the analysis of single-cell RNA-sequencing data have been developed. These methods rely on the projection of every profiled cell onto a lower-dimensional space, estimation of cell-wise correction vectors based on the subsets of cells showing the most similar transcriptomic profiles across batches, and return batch-corrected gene expression matrices that can be used for downstream analyses such as dimensionality reduction, cell clustering and pseudotime analyses. In this presentation I will first provide an overview of the most widely used single-cell integration analysis procedures, then I will show an application of such methods for the integration of human Hematopoietic Stem and Progenitor Cells (HSPCs) single-cell RNA-sequencing datasets.

Die Disputation besteht aus dem o. g. Vortrag, danach der Vorstellung der Dissertation einschließlich jeweils anschließenden Aussprachen.

Interessierte werden hiermit herzlich eingeladen

Die Vorsitzende der Promotionskommission
Prof. Dr. A. Marsico