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

Montag, 9. Oktober 2017, 10.00 Uhr

Ort: Seminarraum S1, Turm 3 EG, 0.3.73,
Max-Planck-Institut für Molekulare Genetik,
Ihnestraße 63-73, 14195 Berlin.

Disputation über die Doktorarbeit von

Herrn Matthias Lienhard

Thema der Dissertation:
Computational Analysis of Genome-wide Methylation Enrichment Experiments

Thema der Disputation:
Statistical Assessment of Differential Expression

Die Arbeit wurde unter der Betreuung von **Prof. Dr. M. Vingron** durchgeführt.

Abstract: Assessing differential gene expression between groups of samples from RNA sequencing experiments depends on statistical modeling of the observed read counts. Common approaches are based on the negative binomial (NB) distribution, which has two parameters: a group mean parameter and an overdispersion parameter, accounting for the variability within the group. In this context, Generalized Linear Models (GLMs) provide an efficient and powerful framework to fit the model parameters to the observed read counts, and to assess the evidence for differential expression between groups. In recent years this approach has became standard practice, and has been applied in more than 10.000 studies. However, despite its popularity, the approach depends on the potentially problematic assumption that the transcriptional variability for a given gene is the same for both groups, which is often not the case, in particular when comparing tumor and normal tissues.

In order to critically review the impact of this violated assumption, I suggest an alternative method to fit the parameters to the observed data, based on Bayesian Estimation and Gibbs sampling. By simultaneously estimating the posterior distributions of the model parameters, the approach provides a flexible and versatile alternative to the point parameter estimation using GLMs. It allows me to gradually adjust the model complexity, from (1) fixing the dispersion parameter to a point estimate, over (2) a free common parameter for the two groups to (3) independent dispersion parameters for both groups. This setup enables me to test the impact of the problematic assumption using both simulated data as well as RNA seq data from cancer and adjacent normal tissues.

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

Interessierte werden hiermit herzlich eingeladen

Der Vorsitzende der Promotionskommission
Prof. Dr. M. Vingron