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, 27. Januar 2020, 10:00 Uhr

Ort: Seminarraum 009

(Fachbereich Mathematik und Informatik, Arnimallee 6, 14195 Berlin)

Disputation über die Doktorarbeit von

Herrn Marten Jäger

Thema der Dissertation:

Annotation und Interpretation von Varianten und Polymorphismen im humanen Genom

Thema der Disputation:

Advantages of graph genome representations

Die Arbeit wurde unter der Betreuung von **PD Dr. P. N. Robinson** durchgeführt.

Abstract: Next generation sequencing techniques produce (short) linear sequencing reads, that first need to be aligned to a reference genome before allowing to call variations. The selection of the most fitting reference genome is a crucial step in the analysis of such data, since aligning to the wrong scaffold will lead to detect false positive variants. Medical practices until now mostly rely on linear consensus haplotypes of the human genome provided by the Genome Reference Consortium, whose latest version is GRCh38. Those references provide standardized coordinate systems, allowing genome comparison and functional elements annotation. This presentation will give a short introduction to the structure of the human genome, and to traditional approaches to assemble its linear consensus representations. It will then describe a method to implement a graph reference genome, which enables the integration of (population) variability [1]. The proposed pipeline builds a graph representation of the genome using all 16 million single-nucleotide polymorphisms (SNPs) and short insertion/deletions (indels), as well as structural variants, from 2,800 diploid genomes. It allows computationally efficient processing of a single whole-genome sequencing sample, from alignment to genotyping, in a few hours. It improves the read mapping rate and sensibility, leads to a higher variant calling recall, and allows a more accurate genotyping of structural variations. Using graph-based pan-genomes as references hence shows promising outlooks towards more accurate genomics analyses, without precision loss. The current efforts invested in national sequencing projects will further improve the genotyping ability for new, well defined haplotypes. Finally, I will present the results of my dissertation. It covers the genotyping of alternative locus haplotypes based on detection of recurrent variants fingerprints. Then, I present a java-based library for the annotation of next generation sequencing derived SNP/indel variants.

[1] Rakocevic, G., Semenyuk, V., Lee, W. et al. Fast and accurate genomic analyses using genome graphs. Nat Genet 51, 354-362 (2019)

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. K. Reinert