UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Exam in INF-BIO9121/INF-BIO5121 - High Throughput Sequencing technologies and bioinformatics analysis
Day of exam: 06.11.2024
Exam time: 10.00-12.00
The exam set consists of 2 pages
No attachments
Allowed materials: none

Teacher Karin Lagesen can be reached on 915 75 916

Ensure that the exam set is complete before you start answering questions.

Please use separate sheets to answer question 1, 2 and 3.

Note: PhD students should answer all questions, while master students should not answer those noted as PhD students only. You need at least 50 points (MSc student) or 70 points (PhD student) to pass the this written exam.

Question 1 - Genome assembly - 25 or 40 points

a. 5 pts. Define the following terms, and describe how they relate to each other:
    i. read
    ii. contig
    iii. scaffold
b. 5 pts. After an assembly has been done, we have scaffolds with the following lengths: 51, 70, 93, 25, 123, 80, 5, 30, 150. Calculate the scaffold N50 value.
c. 5 pts. Describe how the length of reads can affect the lengths of contigs in an assembly, and also explain why longer reads are generally more desirable than short ones.
d. 10 pts. What is the difference between mate pairs and paired end reads, and how can mate pairs help an assembly?
e. (PhD students only) 15 pts. Describe briefly the difference between how Overlap Layout Consensus and De Bruijn Graph based approaches use reads to assemble a genome.

Question 2 - RNA-seq - 25 pts

a. 10 pts. Describe briefly the steps involved in a differential gene expression analysis, from fastq file to a list of differentially expressed genes.
b. 5 pts. What is coverage, and why is high coverage desirable in an RNAseq experiment?

c. 10 pts. What is a splice isoform, and how can isoforms be discovered using RNAseq? What effect can read length have on experiments that aim at discovering isoforms?

Question 3 - Variant calling - 25 or 35 pts

a. 5 pts. What is the goal in a variant calling experiment? Describe briefly what kind of variants that can be discovered in variant calling.
b. 5 pts. What is “base quality” and and how can it it affect variant quality? Give a heuristic explanation.
c. 15 pts. In variant calling, the reads are first mapped before an alignment process happens. What is mapping and alignment, and why do variant calling pipelines use both types?
d. (PhD students only) 10 pts. Explain the concept of variant annotation and explain how alternative splicing can make the task of variant annotation challenging.