

## T-61.5050 High-throughput bioinformatics

Exam 19.05.2014

Dr. Elena Czeizler, Ali Faisal, Suleiman Khan

There are 6 questions (remember to turn the page). You can reach 30 points in total, 5 points for each question.

### Question 1 (5p): Term Definitions

Define the following concepts (1-2 phrases for each concept):

- interactomics (1p)
- paired reads (1p)
- metagenomics (1p)
- p-value (1p)
- alternative splicing (1p)

### Question 2 (5p): Sequencing

- Describe the difference between sequencing by synthesis and sequencing by ligation? Name one example technology for each! (2p)
- Briefly define the following terms: contig, scaffold, coverage. (1p)
- Briefly describe the Solexa/Illumina Genome Analyzer pipeline. (2p)

### Question 3 (5p): Burrows-Wheeler transform

- Construct the suffix array and the Burrows-Wheeler transform for the string "CCATAGAT\$". Describe the construction procedure in 1-2 sentences. (3p)
- Construct the original sequence, knowing that its Burrows-Wheeler transform is "R\$KEPKOOBEE" (show and explain the intermediary steps of this back-transformation). (2p)

### Question 4 (5p): Transcriptomics

- Name 2 of the main types of approaches that are nowadays used to measure gene expression on the transcriptomic level. What kind of information do you get from the measurement and how do the approaches differ in that respect? (2p)
- Choose one of the 2 approaches named in part (a) and list the general steps in its pipeline from biological samples to gene expression values. (Focus on the description of the main steps and use relevant keywords, rather than going into details.) (2p)
- Explain the difference between biological replicates and technical replicates, as well as their purposes! (1p)

### Question 5 (5p): Enrichment analysis

- Define the concept "enrichment of a gene set S". (1p)
- Name at least 2 resources that can be used for assigning genes to biological processes. (1p)