Syllabus: Gene expression 2021

Konstantin Zaitsev

3/22/2021

Before we start

In this course, we (Konstantin Zaitsev and Alexander Tkachenko) will try to explain most of the essential things you need to now about gene expression. We will go from explaining ways to capture RNA and measure RNA abundance to discussing how to properly design gene expression study and what the best practices are.

Authors of this course:

1. Konstantin Zaitsev, https://www.researchgate.net/profile/Konstantin_Zaitsev, https://stepik.org/course/512/
2. Alexander Tkachenko, https://publons.com/researcher/3041889/alexander-tkachenko/, https://stepik.org/course/94/

Most of the material updates will be published here.

Disclaimers:

1. All the course materials will be in English
2. By "gene expression" we almost always mean "RNA expression" if not said otherwise

Syllabus

- Central dogma of molecular biology, structure of gene, types of RNA, structure of RNA, transcription, reverse transcription, FACS
- Microarray: quantification, normalization, basic analysis
- RNA-seq: alignment, quantification, QC, normalization, basic analysis
- Overall quality control: PCA, clustering, outlier detection
- Overall quality control: batch correction
- Differential expression (DE): limma for microarray, Deseq2 for RNA-seq
- Downstream analysis: pathway/gene set enrichment analysis
- Downstream analysis: gene expression deconvolution
- Transcriptome assembly, functional annotation
- Single-cell transcriptomics: Seurat basic analysis
- Single-cell transcriptomics: Trajectory analysis, RNA velocity, optimal transport
- Visual data exploration: phantasus, JBR genome browser
- Experimental design of gene expression study

Prerequisites

Most of the lessons will be in format of seminars: lectures and practical part with your laptops. Most of the practices we will be working in R and only for several we will require a command line.

You will be given an access to a remote-server with pre-installed Rstudio and command tools.

Course structure

Course will take place in both offline and Zoom. We will send you the link and also notify you if the zoom link is updated.

Course is split into five days of intensive seminars and homeworks.

1. Day 1: Introduction to course. Transcription. Regulation of transcription.
2. Day 2: Microarray. Differential expression. Secondary analysis in gene expression studies.
3. Day 3: RNA-seq. Normalization of data in RNA-seq. Differential expression in RNA-seq.
4. Day 4: Transcriptome assembly. Gene expression deconvolution. Gene expression study design.
5. Day 5: Single-cell RNA-seq.

Homeworks

There will be 4 projects homeworks during this course:

1. Microarray dataset: processing, QC, analysis, DE
2. RNA-seq dataset: processing (from fastq), QC, analysis, DE, downstream analysis (pathways)
3. Transcriptome assembly and functional annotation
4. Single-cell RNA-seq dataset: analysis with Seurat

Grading the course

Homeworks (during this week):

- 4 homeworks (50 points)

Examination is somewhere in May:

- Dataset processing as a part of examination (25 points)
- Oral examination (25 points)

Grades:

- >= 50 is E (all homeworks will give you E)
- >= 60 is D
- >= 70 is C
- >= 80 is B
- >= 90 is A (that's where you aim)