Analysis of High Throughput Sequencing RNA-Seq Data (1+2+1 credits)

Course code: PNS0208

Application code: SLU-P0055

Application deadline: 2020-11-09

Course period: 23 Nov - 10 Dec 2020

Distance: Yes

Syllabus

Syllabus approved: 2020-10-08 (from)

Subjects: Bioinformatics

Education cycle: Third cycle

Grading scale: Pass / Failed
The requirements for attaining different grades are described in the course assessment criteria which are contained in a supplement to the course syllabus. Current information on assessment criteria shall be made available at the start of the course.

Language: English

Prior knowledge: For the main module, basic knowledge of Unix (specifically, the command line interface) and the programming language R is required. Attending the first optional module is enough to reach this requirement. The course is primarily for SLU PhD students but will be open also for researchers if space allows.

Objective, including learning outcomes:

The aims of this course are to:
- familiarize the participants with advanced RNA-Seq data analysis methodologies and how biological knowledge can be gained from these by illustrating different analyses approaches
- allow the participants to acquire computational competences about latest RNA-Seq analytical approaches and related statistical methods
- expose the participants to modern computing technologies, as the computer hands-on of the course will be performed using cloud computing resources
- develop the participants' awareness to critically reflect onto RNA-Seq analysis strengths and weaknesses
- enable participants to perform the analysis of their own data (in stance of the exemplary dataset otherwise available)

The corresponding Learning Outcomes are:

Module one (optional)
- develop basic skills to navigate and interact with the Unix Command Line Interface (CLI)
- develop basic skills to interact with the RStudio Integrated Development Environment (IDE)
- reproduce basic analysis in the R programming language

Module two
- summarise High Throughput Sequencing (HTS) technologies, past and present
- report advantages and limitations of HTS for expression profiling
- describe the HTS data pre-processing
- perform the HTS data pre-processing
- describe the pseudo-alignment principles
- compare (pseudo-)alignment methods
- perform the pseudo-alignment
- list statistical concepts of importance for RNA-Seq
- perform the biological QA
- describe the principle of Differential Expression (DE)
- enumerate the statistical concepts of DE
- list the assumptions associated with DE
- perform the differential expression analysis

Module three (optional)
- apply the knowledge from module two to one own's data
- interpret and reflect on the results
- present the results to a scientific audience

Content:

High Throughput Sequencing (HTS) and in particular Next Generation Sequencing (NGS) have revolutionized the way we conduct gene expression analysis. In comparison to micro-array-based methods, NGS has unleashed an almost unlimited power to perform gene expression analysis, for a similar price and with faster, more comprehensive, more efficient and more reproducible characteristics. Rapidly, the data generation rate has exceeded the analytical capabilities and data analysis has become the major bottleneck in gene expression studies. This course, aimed at advanced students, has the objective to help addressing this problem by (i) helping participants critically assess the challenges faced in the HTS field, (ii) developing efficient communication skill with bioinformatician, and (iii) training students in analyzing HTS data.

The course has one main module, and two optional modules (two and one ECTS credits each, respectively). The main module covers the analysis of High Throughput RNA-Seq data, while the first optional module is an introduction to the Unix and R environments to give attendees the necessary admission requirements for the main module. The second optional module is a follow-up for attendees to apply the knowledge gained from the main module content on their own data. The course will provide two to four ECTs depending on the selected modules, as follows:

- Module 1 (optional): 1 ECTS credit
- Module 2 (mandatory): 2 ECTS credits
- Module 3 (optional): 1 ECTS credits

Module 1 will be happening ahead of the main module and last two days. Literature reading and pre-course material for module 2 will be made available prior to the course. Active teaching on module 2

and 3 will last 5 and 1 day(s), respectively.

For Module 3, the students who are interested will be able to analyse data from their own project by applying the methods learned in module 2 and with some guidance from the teachers (synchronous and asynchronous). Students will need to organise their access to the high-performance computing facility (probably their PI / supervisor will need to do so), with the help of SLUBI if needed. First, with asynchronous support from SLUBI, the students will perform the pre-processing and initial analysis of their data. Then, during a day, they will meet with the trainers to address issues, discuss the data interpretation and prepare a flash presentation about their project and results. These will then be given to the other participants and trainers in the form of an online mini-symposium.

The course will run for five to eight days; depending on whether optional modules are selected by the students - mixing lectures, interactive lectures, computer hands-on session, literature review, etc. The course will use various teaching environment, including meetings as a classroom, but also as smaller groups. In addition, asynchronous virtual environment will be used, such as virtual classrooms, flipped classrooms, etc. to offer the participants the possibility to review and deepen their understanding of the course material at their own pace and to discuss it among peers.

Requirements for examination

The course will have a continuous assessment, conducted once a day at the end of the day, based on activities including, but not limited to, small group's work with peer-review, literature review, journal club, etc. The continuous assessment and the final exam will provide five marks (six, seven or eight if either or both of the optional modules are also attended). All assessments must be passed in order to pass the course. In addition, attendance at all scheduled activities and approved computer exercises are required. All the material will be made available and students missing a day will be offered the possibility to go through it in self-study and be assessed.

Additional information

Teaching on this course will be provided by the SLU Bioinformatics Infrastructure (SLUBI), namely by Nicolas Delhomme, Juliette Hayer and Domenico Simone. The course is organized on behalf of the SLU research school Organism Biology. There is no tuition fee. Participants are expected to bring their own laptop for the practical computer exercises.

Due to the current COVID-19 situation, the course given in autumn 2020 will be conducted online only. However, if students wish so, they can organise themselves in groups and join using one of the teaching facilities available at SLU.

The maximum number of participants is 20. SLU-registered PhD students are prioritized over other researchers but otherwise admission will be on a first come, first serve basis, provided that the minimal requirements are met.

If you have any question, please contact the course leader or the slack channel #organism-biology-research-school-rna-seq-course of the slubi-workspace.slack.com workspace.

Responsible department: Department of Plant Biology

Course leader: Nicolas Delhomme, E-mail: nicolas.delhomme@slu.se