

# Analysis of High Throughput Sequencing RNA-Seq Data

October 5<sup>th</sup> - 16<sup>th</sup> (course P000057); October 23<sup>rd</sup> (course P000058)

## Project details and course format

### Aim of the course

The aim of this applied courses is for you to gain hands-on experience of the steps involved in processing, analysing and interpreting some functional genomics data - in this case RNA-Seq. The steps you will be undertaking during the course directly relate to what you would experience in your working life as a scientist applying functional genomics approaches. We will use example data coming from human T cells, but the same approaches would apply to any other system/species that you might be working on, so these skills are readily transferable.

## Basic information

During the first week of the course you will listen to a short presentation of the characteristics of modern high-throughput sequencing data. After that, you will receive an introduction to using the terminal (working in Linux) and how to run programs and scripts. This will provide essential background and skills to help with your project work. During the rest of the week, you will perform the pre-processing of some RNAseq data coming from human T cells.

In the second week, we will move from the linux terminal to the R programming language. You will first receive an introduction to R, and then you will go on to perform quality assessment and differential gene expression analysis on the provided data. You will also receive a brief introduction about following operations, such as gene ontology enrichment and differential transcript usage.

The 9<sup>th</sup> and 16<sup>th</sup> of October will be catch up days, where you will be able to work on whatever gave you the most problems during the week, with the assistance of the teachers.

In the third week (course P000058) there will be no lectures. Instead, you will work independently on your own data, applying the concepts from the course that you will find most useful to your specific situations. Teachers from the course will be available every day to help you with any problem you might encounter.

In the last day of the course (October 23rd) we will have open discussions about the problems you encountered in the week, and about how to best apply the concepts you learnt to your future work.

## Detailed schedule

### Week 1 - October 5<sup>th</sup> - 9<sup>th</sup> 9:00 - 15:00 every day (course P000057)

#### **Day1: High Throughput Sequencing data processing**

09:00 - 09:30 - Welcome to the course!  
09:30- 09:45 - Daily learning objectives  
09:45 - 10:30 - High Throughput Sequencing Lecture  
10:30 - 10:45 - Coffee break  
10:45 - 11:30 - Data pre-processing and analysis workflow  
11:30 - 12:00 - Workshop data description  
12:00 - 13:00 - Lunch break  
13:00 - 13:30 - Introduction to HPC2N  
13:30 - 14:45 - Checking HPC2N connections  
14:45- 15:00 - Feedback and Daily Assessment

#### **Day2: Introduction to Unix, specifically Linux**

09:00 - 09:30 - Revision session  
09:30- 09:45 - Daily learning objectives  
09:45- 10:30 - Software carpentry 1-2 (guided)  
10:30 - 10:45 - Coffee break  
10:45 - 12:00 - Software carpentry 1-2 (hands-on)  
12:00 - 13:00 - Lunch break  
13:00 - 14:00 - Software carpentry 6 (guided)  
14:00 - 14:45 - Software carpentry 6 (hands-on)  
14:45 - 15:00 - Feedback and Daily Assessment

#### **Day3: Introduction to nextflow and nf-core**

09:00 - 09:30 - Revision session  
09:30- 09:45 - Daily learning objectives  
09:45- 10:30 - Interactive introduction to nf-core  
10:30 - 10:45 - Coffee break  
10:45 - 12:00 - Interactive introduction to nextflow  
12:00 - 13:00 - Lunch break  
13:00 - 13:30 - Introduction to slurm  
13:30 - 14:45 - Setup and start nf-core (hands-on)  
14:45 - 15:00 - Feedback and Daily Assessment

#### **Day4: Running nf-core**

09:00 - 09:30 - Revision session

09:30- 09:45 - Daily learning objectives  
 09:45- 10:30 - Keep trying to start nf-core  
 10:30 - 10:45 - Coffee break  
 10:45 - 12:00 - Continue keeping trying to start nf-core  
 12:00 - 13:00 - Lunch break  
 13:00 - 13:30 - Introduction to multiQC reports  
 13:30 - 14:45 - Exploring and commenting the multiQC report produced by nf-core  
 14:45 - 15:00 - Feedback and Daily Assessment

### **Day5: Catch up day**

09:00 - 09:30 - Revision session  
 9:30 - 15:00 - Working on what was unclear or gave the most problems

## **Week 2 - October 12<sup>th</sup> - October 16<sup>th</sup> - 9:00 - 15:00 every day (course P000057)**

### **Day1: Introduction to R**

09:00 - 09:15 – Daily learning objectives  
 09:15 - 09:45 – Data analysis with R (presentation)  
 09:45- 10:30 - Data analysis with R (exercises)  
 10:30 - 10:45 - Coffee break  
 10:45 - 11:15 – Visualization with R (presentation)  
 11:15 - 12:00 - Visualization with R (exercises)  
 12:00 - 13:00 - Lunch break  
 13:00 - 14:45 – Data analysis learnr tutorial  
 14:45 - 15:00 - Feedback and Daily Assessment

### **Day2: Biological QA**

09:00 - 09:30 – Revision session  
 09:30 - 09:45 – Daily learning objectives  
 9:45 - 10:30 - Exploratory Data Analysis (EDA) - Section Introduction, Metadata and Count data - guided session  
 10:30 - 10:45 - Coffee break  
 10:45 – 12:00 Exploratory Data Analysis (EDA) - Section Introduction, Metadata and Count data – hands on.  
 12:00 - 13:00 - Lunch break  
 13:00 - 14:00 - Exploratory Data Analysis (EDA) - Section Exploratory Data Analysis – guided session  
 14:00 – 14:45 - Exploratory Data Analysis (EDA) - Section Exploratory Data Analysis – hands on  
 14:45 - 15:00 - Feedback and Daily Assessment

### **Day3: Differential expression analysis**

09:00 - 09:30 – Revision session  
 09:30 - 09:45 – Daily learning objectives

9:45 - 10:30 - Differential Expression - Models - guided session  
 10:30 - 10:45 - Coffee break  
 10:45 - 12:00 - Differential Expression - Models – hands-on  
 12:00 - 13:00 - Lunch break  
 13:00 - 14:00 - Differential Gene Expression - DE Results – guided session  
 14:00 - 14:45 - Differential Gene Expression - DE Results – hands-on  
 14:45 - 15:00 - Feedback and Daily Assessment

#### **Day4: GO enrichment**

09:00 - 09:30 – Revision session  
 09:30 - 09:45 – Daily learning objectives  
 9:45 - 10:30 - Gene Set Enrichment Analysis (GSEA) - guided session  
 10:30 - 10:45 - Coffee break  
 10:45 - 12:00 - Gene Set Enrichment Analysis (GSEA) – hands on  
 12:00 - 13:00 - Lunch break  
 13:00 - 13:30 – Official Feedback (Room: SLUBI2023)  
 13:30 – 15:00 – Discussion session

#### **Day5: Catch up day**

09:00 - 09:30 - Revision session  
 9:30 - 15:00 - Working on what was unclear or gave the most problems

### **Week 3 - October 23<sup>rd</sup> (course P000058)**

Day 1 - 9:00 - 17:00

09:00 - 09:30 - Introduction  
 09:30 - 10:30 - Interactive sessions to address analysis issues  
 10:30 - 10:45 - Coffee break  
 10:45 - 12:00 - Interactive sessions to address analysis issues  
 12:00 - 13:00 - Lunch break  
 13:00 - 14:00 - Presentation preparation  
 14:00 - 16:00 - Mini-symposium with talks and peer-review  
 16:00 - 16:45 - Feedback from the trainers on the projects  
 16:45 - 17:00 - Final Feedback