Introduction to Bioinformatics

Introduction to Bioinformatics and Sequencing Technologies

- Introduction to DNA and RNA sequencing
- Raw reads and fastq files
- Databases: Ensembl, Genecode, ENA, and GEO
- Exercise: Download raw reads from ENA database

Introduction to UNIX and Linux

- How a UNIX PC is organized (root, directories, files ecc…)
- The directory structure (ls, mkdir, cd, pwd)
- Editing and creating files (cp, mv, rm, rmdir, top, cat, less, head, tail, vim, touch)
- Searching the content of a file (grep)
- Introduction to regular expression and wildcards
- Pipes
- Handling PC resources, background and foreground applications
- Compiling software
- Run software
- Exercise: installing common bioinformatics software (es. fastqc, STAR, htseq counts)

Mapping files and common genomic arithmetics

- Genomic mapping files
- SAM and BAM format, samtools
- Mapping reads to a reference genome DNA-seq
- Mapping reads to a reference genome RNA-seq
- Quantify gene expression
- Common enrichment analysis
- Common genomics arithmetic
- Excercise: Find overlapping genomic regions between 2 datasets

Introduction to R

- Introduction to R (installing R and Rstudio)
- Rstudio
- Biocondutor
- Vector, matrices, lists, dataframes
- Data input (read.tables)
- Write output (write.table)
- Data filtering
- Simple statistics
- Plotting data
- Exercise: Differential gene expression analysis

Practical exercise: From fastq to differential gene expression analysis

NOTE: in order to optimize the class, it is compulsory to have a working Linux distribution before the start of the first lesson. It is also recommended to have a word processor and spreadsheet software (e.g. MS Office and LibreOffice), as well as Notepad++. A preliminary meeting can be organized, upon request, to help set up the students' notepad and to discuss the program of the class.