

Statistical Methods and Software for mRNA-Seq and ChIP-Seq

Centro de Investigación Príncipe Felipe
Valencia, Spain

November 8-10, 2010

[[Home](#)] [[Lectures](#)] [[Labs](#)] [[References](#)]

Home: [[Instructors](#)] [[Description](#)] [[Program](#)]

Instructors

[Dr. Laurent Jacob](#), Genentech Innovation Fellow, Center for Computational Biology and Department of Statistics, UC Berkeley

Oleg Mayba, Department of Statistics, UC Berkeley

Course developed in collaboration with [Professor Sandrine Dudoit](#), Division of Biostatistics and Department of Statistics, UC Berkeley

Description

The short course will provide an overview of statistical methods and software for the analysis of high-throughput sequencing (HTS) data, with emphasis on transcriptome analysis (mRNA-Seq) and the study of DNA-protein interactions (ChIP-Seq). The morning lectures on statistical methodology will be followed by afternoon lab sessions demonstrating the application of the methodology to mRNA-Seq and ChIP-Seq data using R software packages from the Bioconductor Project (<http://www.bioconductor.org>, <http://www.bioconductor.org/help/workflows/high-throughput-sequencing>).

For registration and other practical matters: <http://bioinfo.cipf.es/RNA-seq2010>

Requirements

We expect the participants to have some familiarity with the R programming environment, but not necessarily with the Bioconductor suite of packages. For the lab portion of the course, we require that the latest version of R (currently R-2.12.0) and the latest version of Bioconductor (currently version 2.7) be installed. These can be obtained from <http://www.r-project.org> and <http://www.bioconductor.org>, respectively.

In addition to Bioconductor core packages, we will also be using packages

- aroma.light
- ChIPpeakAnno
- GenomeGraphs
- Genominator
- rGADEM
- rtracklayer
- ShortRead
- BSgenome (**NEW**)
- BSgenome.Hsapiens.UCSC.hg19
- BSgenome.Scerevisiae.UCSC.sacCer2 (**NEW**)
- yeastRNASeq

Please make sure that these are installed and can be loaded into R.

NEW: As part of the first lab, we would like you to do short read alignment using Bowtie (<http://bowtie-bio.sourceforge.net/index.shtml>). Please install this aligner and the pre-built *S.cerevisiae* genome available from Bowtie's main page. The reads to be aligned come from a yeast data set that we will discuss and can be obtained from GEO (<http://www.ncbi.nlm.nih.gov/geo/>) with accession number GSM298523 (Isogenic wild-type Rep 1). You will need to download the .fastq file containing reads and their quality scores from <ftp://ftp.ncbi.nlm.nih.gov/sra/static/SRX003%2FSRX003157> and unzip it.

Note: this list of requirements and packages is subject to change as the course syllabus is finalized, so please check back frequently.

Program

Day 1 - November 8th

Lecture 1: Introduction to mRNA-Seq and ChIP-Seq

- High-throughput gene expression assays
- Experimental design
- Pre-processing: Image analysis, base-calling, and read-mapping

Lab 1: Introduction to statistical software for HTS

- Bioconductor basics
- Overview of Bioconductor packages for HTS
- Exploratory data analysis (EDA)

Day 2 - November 9th

Lecture 2: Statistical methods for mRNA-Seq

- Exploratory data analysis
- Controls
- Normalization and expression quantitation
- Differential expression

Lab 2: Statistical software for mRNA-Seq

Day 3 - November 10th

Lecture 3: Statistical methods for ChIP-Seq

- Experimental setup
- Statistical aspects of peak-finding
- Use of control samples and normalization
- Peak validation

Lab 3: Statistical software for ChIP-Seq

- Introduction to relevant Bioconductor tools
- Peak annotation
- Motif finding
- Data visualization

