



1SC4210 – Biology and statistics

Instructors: Laurent Le Brusquet, Arthur Tenenhaus

Department: DOMINANTE - VIVANT, SANTÉ, ENVIRONNEMENT

Language of instruction: FRANCAIS

Campus: CAMPUS DE PARIS - SACLAY

Workload (HEE): 60

On-site hours (HPE): 34,50

Description

The cell is a complex multi-scale system that biologists have been studying for several centuries. The development of new technologies now makes it possible to explore it and measure its activity more and more finely. The objective of the first part of the course is to present the different facets of the cell to better understand its overall functioning. Measurement technologies will also be discussed.

The second part of the course presents the statistical methods widely used in the biomedical field. At the end of this section, students will have a useful toolbox to answer the various questions raised by the analysis of biomedical data.

This course will be widely illustrated on medical applications to enable students to acquire the knowledge and vocabulary necessary to understand the applications proposed during the EI (basics in biology, genomics, etc.), to understand the origin of the huge volume of data related to medical applications. For this reason, approximately one third of the total volume of this course will be reserved for our physician/biologist partners.

Quarter number

ST4

Prerequisites (in terms of CS courses)

Statistics and Learning

Syllabus

PART I : INTRODUCTION TO CELLULAR and MOLECULAR BIOLOGY (12h)

The cell is a complex multi-scale system that biologists have been studying for several centuries. The development of new technologies now makes it possible to explore it and measure its activity more and more finely. The objective of this module is therefore to present the different facets of the cell in order to better understand its overall functioning. Measurement technologies will also be discussed.



I.1. CELL BIOLOGY & MOLECULAR GENETICS

- *General organization of the eukaryotic cell
- *Nucleic acids (DNA and RNA)
- *Proteins
- *The cell cycle of the eukaryotic cell
- *DNA replication in eukaryotes:
- *The regulation of gene expression in eukaryotes
- *Genetic engineering tools

I.2 GENOMIC

I.2.1. Structural genomics

- *Definition of genomics - the rise of bioinformatics
- *Inventory of genome sequencing
- *Genome structure and organization
- *Genome sequencing methods

I.2.2. Functional Genomics: Gene expression

- * The transcriptome :
- * Transcriptional & epigenetic regulation:

I.2.3 Data Integration

- * Omics: interests in biology and difficulties in bioinformatics
- * Post-analysis integration: supervised or unsupervised integration
- * Pre-analysis integration: towards network inference
- * Applications in basic research and personalized medicine

I.2.4. Conclusion & perspectives : evolution of the notion of gene & modifications of classical models of transcription regulation. Information provided by the single-cell genomic.

Each part of the course will be associated with practical work under R

PART II : BIOSTATISTICS (12h)

At the end of Part II, students will have a useful toolbox to answer the various questions raised by the analysis of biological data.

- * Univariate approach and multiple tests (Bonferonni, False Discovery Rate...);
- * Linear model, ANOVA and mixed effects models.
- * GLM (regression, logistic regression, ...)
- * Principe Component Analysis & Hierarchical Clustering

PART III : DIRECTED WORK/TP (9h)

A large part of the concepts covered in the biostatistics section will be implemented in R.

Class components (lecture, labs, etc.)

8 lectures (24h) and 3 TDs (9h)

Grading

written exam (1H30)



Resources

- Teaching team : Arthur Tenenhaus and Marie-Anne Debily (MCF Evry)
- Size of the TDs: 25 for the TDs with computer.
- Software : R
- TP rooms : Signal and Statistics department (reception capacity: 100 students if no interaction with other courses).

Learning outcomes covered on the course

Two main objectives :

1. Introduction to molecular and cellular biology
2. Biostatistics: to provide additional statistical information (useful for the analysis of bomb data) in addition to the "Statistics and Learning" course.