

**CODE  
MCXA0031**

## **AI for OMICS data integration**

Credits: **6 ECTS**  
Semester: **1**  
Coordinator : **Pascal MOSSUZ**  
e-mail : **pascal.mossuz@univ-grenoble-alpes.fr**

### **Content of the Course Unit**

This course aims to present the different methods of artificial intelligence applied to the exploitation of data produced by Omics techniques (genomics, metabolomics, proteomics, transcriptomics, metagenomics for pathogens). The methods to analyze the data produced by each of the OMICs as well as the AI methods allowing their use in a medical context are discussed. Multi-omics integration for integrative medicine and the principles of modeling biological networks are also offered. The teaching is organized in the form of video batches, interactive tutorial sessions between students and teachers. A practical training in the form of a challenge is offered on a transcriptomic dataset

*Interventions by speakers coming from UGA, IAB, INRIA, CEA, GIN, TIMC, CHUGA*

### **Detailed program**

#### **Introduction (Pascal Mossuz)**

#### **I-Ai for metabolomics**

##### **1) Introduction: The Contribution of Metabolomics for Translational Health Research**

Chronic diseases and drug discovery (Audrey le Gouellec)

Clinical impact in Cancers (Pascal Mossuz)

##### **2) NMR-based metabolomics workflow (Florence Fauvelle)**

**Description:** this presentation will introduce how to find candidate metabolic biomarkers using a NMR-based metabolomics workflow. The complete workflow will be explained, from the biological or clinical issue to experimental design, data acquisition and processing and statistical modelling. Some applications of metabolomics for classification and prediction will be presented.

##### **3) Constraint-based modelling of metabolism (Delphine Ropers)**

Students will be introduced with the simulation of metabolism at genome scale and to genome-scale models of cell metabolism and stoichiometric modelling. Methods included : flux balance analysis; integration of omics data; application in human metabolic modelling; perspectives (alternative FBA approaches, personalized medicine)

#### **II AI for transcriptomic**

##### **1) Challenges in Machine Learning for omics data and possible solutions (Ekaterina Flin)**

**Description:** The objective of this course is to introduce key challenges in Machine Learning for omics data in biomedical research: data processing for learning methods, dealing with insufficient data, robustness and generalization. Main approaches are presented and illustrated with a use case of leukemia classification from transcriptomic data.

##### **2) Concept-driven omics analysis: a cancer marker discovery strategy based on the ectopic expression of tissue-specific genes (Sophie Rousseaux)**

**Description:** This presentation will introduce how biological knowledge can be used as a key to enter omics data analysis and lead to an original biomarker discovery approach which can

be applied to various cancer. This session will also provide a basis on which artificial intelligence approaches can be developed.

3) **Practical training in Histology prediction in lung cancers from transcriptomic data (Florent Chuffart)**

**Description:** This practical training represents a “data challenge” project that the students can achieve individually or in small groups. The students have access to transcriptomic data of several hundreds of patients with lung cancer of two histological types (adenocarcinoma and squamous cell carcinoma) from public TCGA database. The objective of their work is to predict the histology type of lung cancer for a 100 new patients. At the end of the project, they should present a small report summarising their strategy to tackle the problem and showing their results.

**Duration:** project time is not restricted, only the results are evaluated.

**III-AI for proteomics (Delphine Pflieger)**

1) **Global presentation on proteomic analysis:**

**Definitions:** DNA/RNA/proteins, proteome, post-translational modifications (PTMs).

Typical questions that can be answered using proteomics (intro of differential analysis)

**How protein samples are analyzed:** proteolysis; LC-MS coupling; existence of dedicated software for identification and quantification; FDR; False Localization Rate in phosphoproteomics (and in the analysis of other modifications).

**Statistical tests for quantitative proteomics:** state that pairwise comparisons are routinely used, even in the context of a two-factor design; interest of the two-way ANOVA to highlight a possible interaction between factors.

2) **Presentation of papers** from the CPTAC consortium, which demonstrated that phosphoproteomics can provide a complementary classification of cancer patients to what is obtained by transcriptomic data

3) **Proteomic analysis of histones:** characterizing the combination of sequence variants and multiple PTMs.

**IV-AI for modeling of biological network**

1) **Dynamical modeling of biological processes (Eric Fanchon)**

The goal here is to learn the main concepts and to gain some understanding of the challenges and of a few methods in order to be able to communicate with modellers

**1. Introduction to dynamical systems**

Presentation of the basic concepts (rate of a process, steady state, basin of attraction, fluctuations, stability, robustness) and formalisms used in the field (differential equations, stochastic systems, boolean networks, rule based systems). Some general points will be discussed concerning the challenges of mechanistic modeling in biology and medicine, and the usefulness of it.

**2. Methods**

Presentation of a few methods used for model building and analysis. I will present in particular some methods based on constraints, which belong to the field of AI.

**3. Example: modeling histone modifications from ChIP-seq data**

Introduction: epigenetics, histone marks, H3K27, epigenetic drugs (cancer).

Presentation of the stochastic model and simulation schemes.

**4. Perspectives**

Brief presentation of some perspectives (single-cell data, AI used to build a dynamical model)

2) **Molecular signaling identification for therapeutic (Cyril Boyault)**

This course present the *SigNet* spin off Project, which work on knowledge access for healthcare with solutions in drug development, by combining AI, signaling network biology and therapeutic data. To illustrate the benefits of this combination strategy, it will provide some examples of therapeutic leads and their molecular signaling identification.

**VI-AI for Genomic**

- 1) **Introduction to somatic and constitutional genomic (Julien Thevenon)**
- 2) **Methods**  
Séquençage /NGS/machine learning
- 1) **Applications**  
In hereditary disease (Julien Thevenon)  
In cancerology (Anne Mc Leer)

### V-AI fort Metagenomic

- 2) **Methods (Alban Caporossi)**  
Introduction/Quality Control/Classification methods  
Taxonomic profiling/Taxonomic classification  
Database & taxonomy impact on classification
- 3) **Metagenomique for infectious disease ( Yvan Caspar)**

### Organisation

**Total 30 sequences of video recorded**

**1 challenge**

**Sequences of interactive works between student and teacher**

### Rules of validation

*Continuous evaluation (25%),*

*Final exam (75 %),*

### Competencies acquired for MIAI Label

Competencies	Novice	Intermediate	Advanced
1 - Select and use the right tools for structuring, exploring, researching, storing, and using data			x
1.1 - By collecting and consolidating, explaining the data for decision-making assistance (business intelligence)		x	
1.2 - Knowing the sources and the data acquisition to train a model			x
1.3 - By assessing the ethical and regulatory impacts linked to the data and their use	x		
2 - Know and apply learning and symbolic AI technologies		x	
2.1 - Knowing the main models and tools (their context and application conditions, their inputs and outputs)			x
2.2 - By modeling a customer or application problem and identifying the use of AI to solve it			x
3 - Identify, explore and model AI technologies on real applications		x	
3.1 - By having the ability to interact with specialists in the field to identify the problem and specify the needs		x	
3.2 - By understanding the AI architecture dedicated to an application and by making it evolve so that it matches business or customer needs: data (collection, storage, management); learning; decision making; analysis and model relevance.	x		
3.3 - By knowing and mastering the management of an AI project in a company	x		
3.4 - Using AI to transform the company and its management	x		