Machine Learning for Metabolomics: Module Introduction

Dr Maria Anastasiadi
(m.anastasiadi@cranfield.ac.uk)

MSc Applied Bioinformatics 2024-2025

13-17 January 2025

www.cranfield.ac.uk

Machine Learning for Metabolomics

Module Aims

- To cover the main aspects related to the analysis of the metabolic profile in living organisms.
- To explore statistical and machine learning techniques that are central to the field of metabolomics.
- To introduce the main concepts of predictive modelling.
- To provide practical examples of metabolomics applications in different fields (health, plant science, food safety and food quality).

Machine Learning for Metabolomics

The Module will Cover

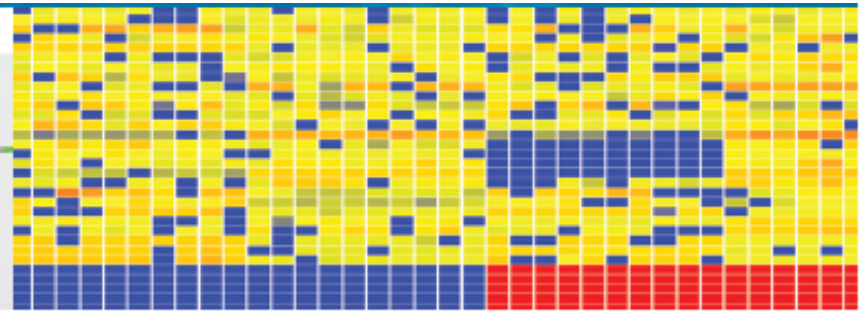
- Metabolomics workflow
- Instrumentation for metabolomics
- Advanced topics in R
- Multivariate classification (e.g. PLS-DA, Machine Learning)
- Biomarker discovery
- Regression methods for numerical prediction (e.g. MLR, PLS, Machine Learning)
- Advanced topics in Machine Learning
- Introduction to Image Analysis
- Case Studies (Soil, Computer Vision, Food Safety)

Metabolomics

Metabolites are the end products of cellular processes and represent the ultimate reflection of the response of biological systems to genetic or environmental changes.

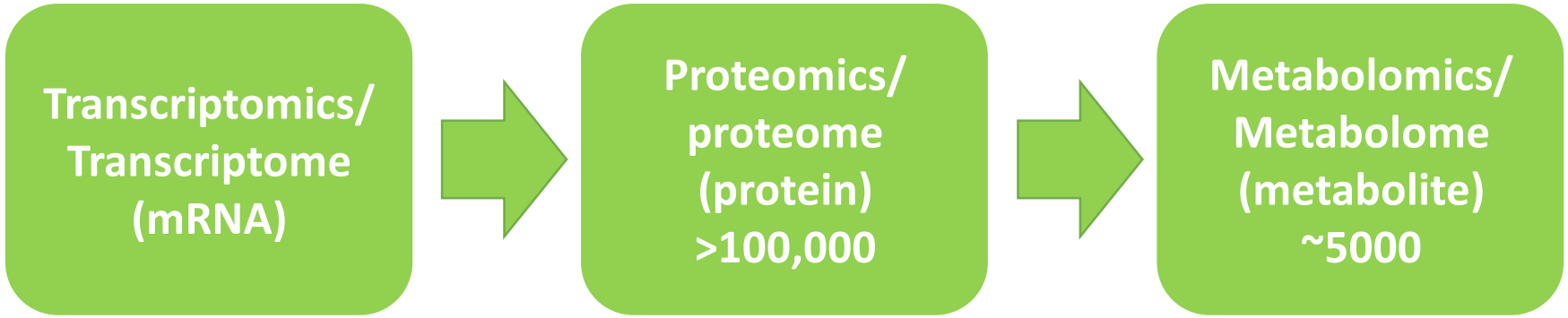
Metabolomics is the technology geared towards providing an overview of global changes in an entire metabolite set.

A powerful, emerging discipline with a broad range of applications including basic research, drug development, disease biomarker discovery, crop optimization and food science.



Metabolomics

Relationship between different “omic” techniques:



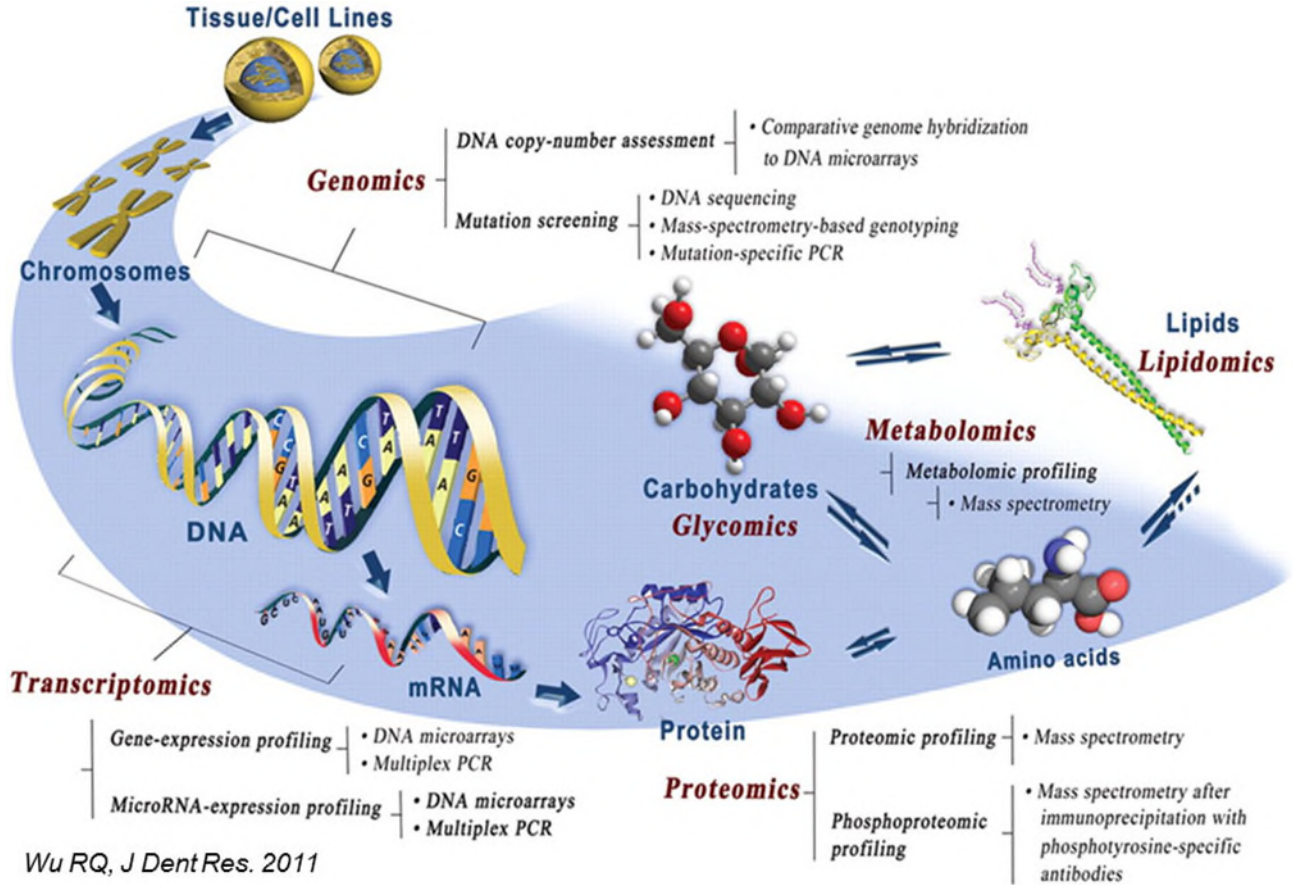
- **Transcriptomics:** observation of alterations in gene expression usually over time (e.g. drug exposure or disease stages).
- **Proteomics:** semi-quantitative measurement of the production of cellular proteins in response to stimuli.
- **Metabolomics:** provides evidence of metabolic changes inside an organism following treatment.

⚠ Metabolomics also accounts for lifestyle, diet and environment of an individual.

Moving towards a holistic approach

Systems biology (multi-omics)

- **Aim**: understand whole systems, by studying the effect of altered external factors on the genome, transcriptome, proteome and metabolome simultaneously.



Metabolomics

Pros:

- The metabolome is the final downstream product of gene transcription; changes in the metabolome are amplified relative to changes in the transcriptome and the proteome.
- As the downstream product, the metabolome is closest to the phenotype of the biological system studied.

Cons:

- Although the metabolome contains the smallest domain, it is more diverse, containing many different biological molecules, making it more physically and chemically complex than the other “omes”.

Metabolomics: Workflow

1. Study

- Biological question
- Experimental Design

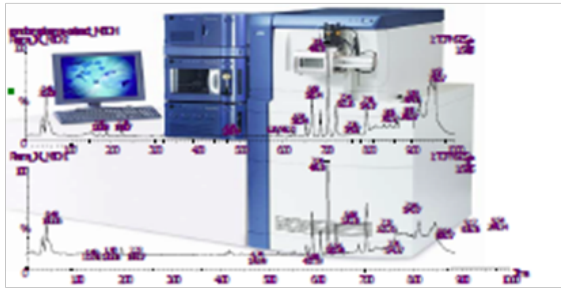
2. Sampling

- Sample collection
- Sample Storage

3. Sample preparation

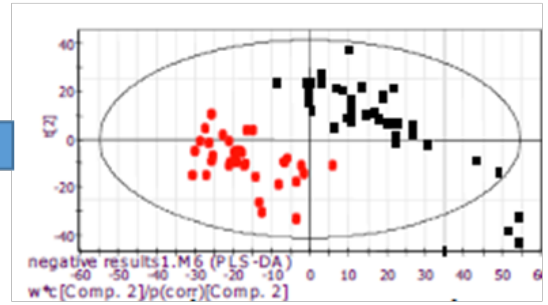
- Sample preparation
- Pre-treatment

4. Metabolite profiling



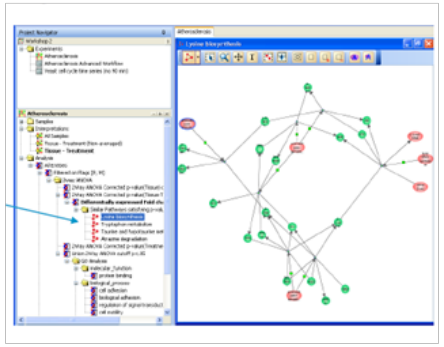
- Data acquisition
- Data pre-processing
- List of peaks/metabolites

5. Data mining



- Data analysis
- Modelling
- Biomarker discovery

6. Biological interpretation

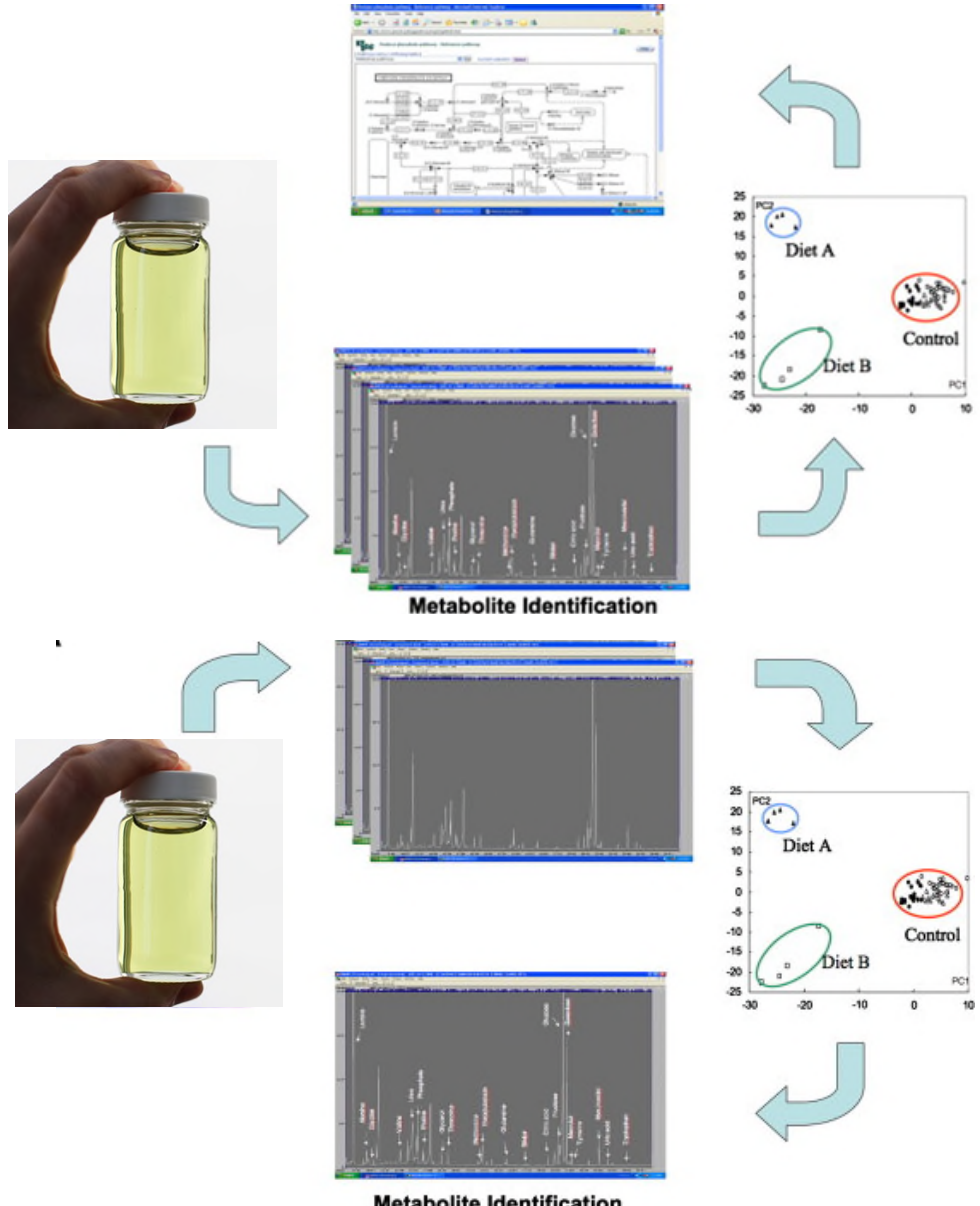


- Database search
- Metabolic pathways

Targeted vs Untargeted Metabolomics

Targeted metabolomics:
Identification/quantification of specific groups of metabolites. Authentic standards required.

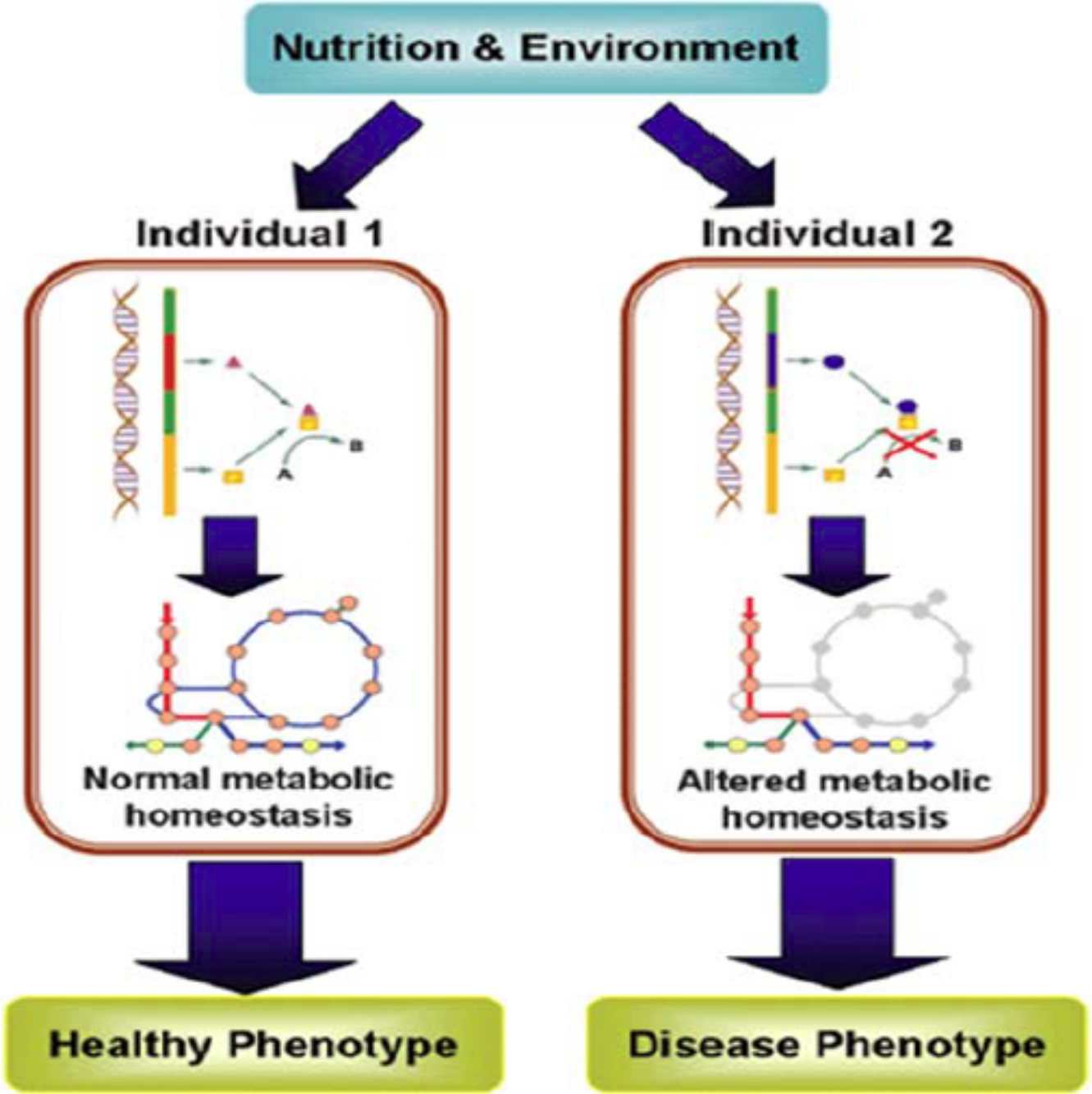
Untargeted metabolomics: Detection of many groups of metabolites to create a metabolic fingerprint of biological processes. Identification and use of standards not required but desirable.



Machine Learning for Metabolomics

- Metabolomics: Applications

- Mechanistic understanding of factors influencing the phenotype.
- Investigate metabolic effects due to nutritional, toxicological or biochemical stimulus.
- Investigate relationship between metabolism and disease.
- Biomarker ID.



Machine Learning for Metabolomics

Learning Outcomes

On successful completion of this module a student should be able to:

- Critically assess various metabolomics analytical and spectral platforms.
- Develop critical awareness of various classification and regression models available for metabolomics data analysis acquired from various analytical platforms.
- Apply state-of-the-art best practices in machine learning to fit the purpose of the analysis and derive biological relevant information from multivariate metabolic datasets.

Machine Learning for Metabolomics

MSc Applied Bioinformatics 2024/25

Module 6: Machine Learning for Metabolomics

Module manager: Maria Anastasiadi (m.anastasiadi@cranfield.ac.uk)

Day	Monday		Tuesday		Wednesday		Thursday		Friday	
Date	13th January	Room	14th January	Room	15th January	Room	16h January	Room	17th January	Room
		Staff		Staff		Staff		Staff		Staff
09:00-09:30	Module Introduction		Introduction to ML		Introduction to Artificial Neural Networks (A)		Multivariate Regression		Ensemble Methods	
09:30-10.00		MA		MA				MA		MA
10:00-10:15	Break		Break		Break				Break	
10:15-10:30	Metabolic Data Acquisition Methods		Classification with ML (Part A)		Introduction to Artificial Neural Networks (B)				ML Applications: Food Safety	
10:30 -11:00							Break			
11:00 - 11:30	Break		Break		Break					FM
11.30:12.00	Multivariate Classification		Practical: Classification with ML Part A		Practical: Intro to ANN		Practical: Multivariate Regression		Break	
12.00:12.30									Practical: Ensemble Methods	
12.30:13.00		MA		MA		MA		TK		
13:00 - 13:30	Lunch		Lunch		Lunch		Lunch		Lunch	
13:30 - 14:00										
14:00 - 15:30	Practical: Multivariate Classification		Classification with ML (Part B)		Introduction to Image Analysis		SMIS: ML Application in Soil Management		Assignment	
14.30:15.00										
15.00:15:30		MA		MA		MA		TK	Break	
15:30-16:00	Break		Break		Break		Break			
16:00-16:30	Practical: Multivariate Classification		Practical: Classification with ML Part B		Practical: Image Analysis		Computer Vision Applications in AgriFood		Practical: Ensemble Methods	
16.30:17.00										
17.00:17.30		MA		MA		MA		DL		MA
VENUE:										

Practical
Lecture
Assignment

STAFF:

MA= Maria Anastasiadi
TK= Tomazs Kurowski
FM= Fady Mohareb
DL= Derek Long (guest Lecturer)