# Plant metabolome analysis

## Bioinformatic tools

Corentin Maslard

October 19, 2026

Université de Neuchâtel
Faculty of Science




# Quick poll: pe.app/cmaslard4

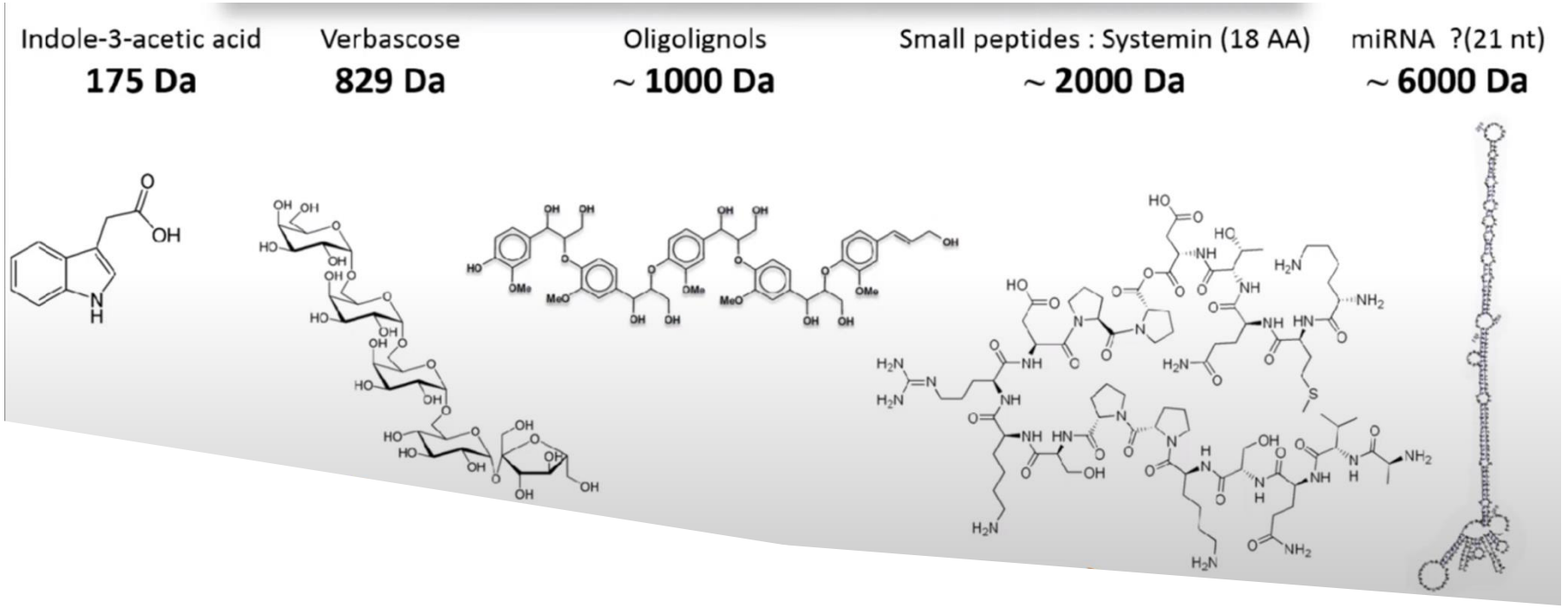
One question from Session 4 before we move on.

# What is metabolomics?

# The metabolomics



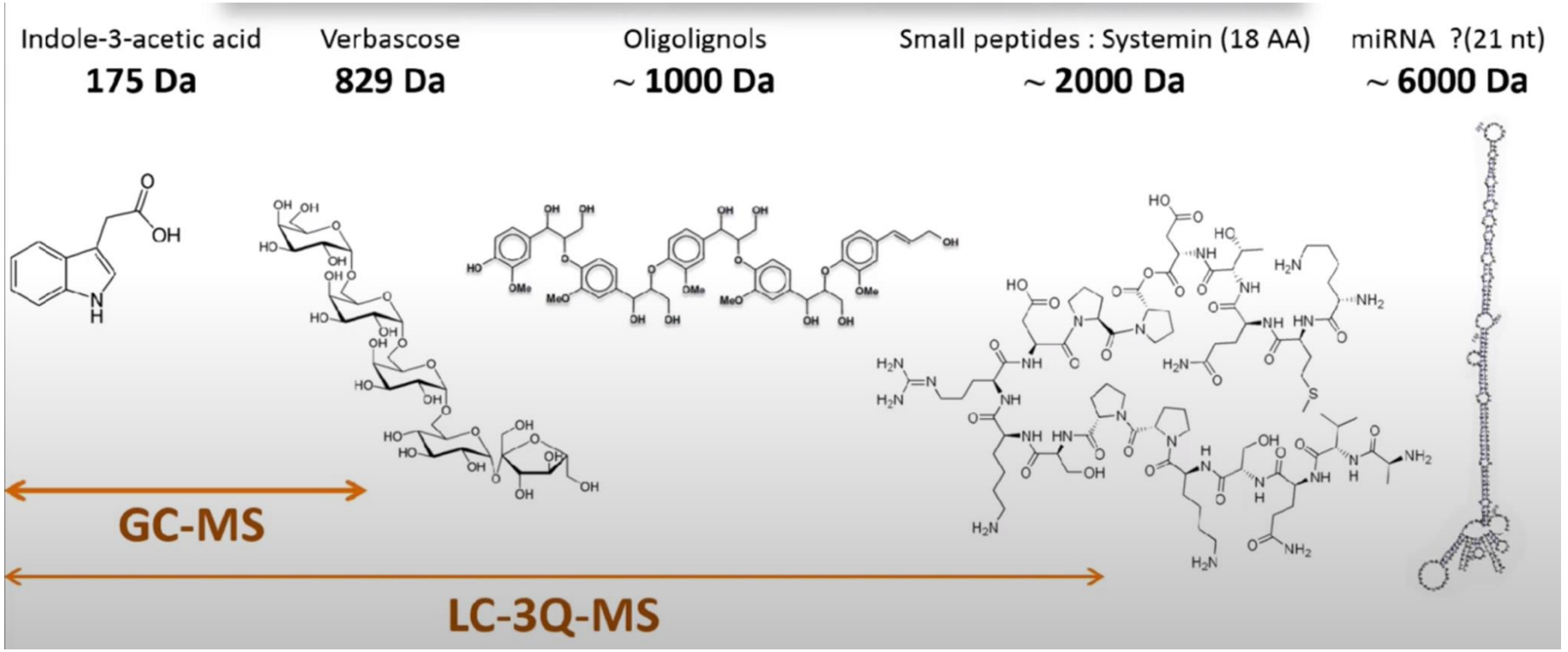

Term often associated with metabolomics



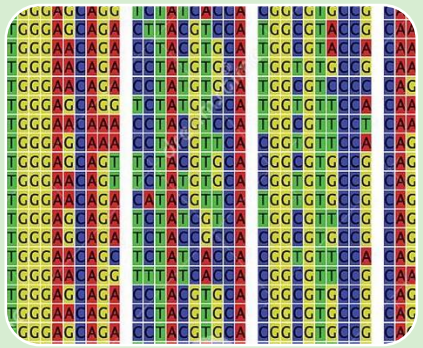
# The metabolomics



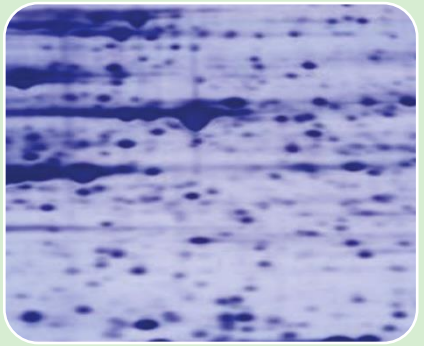

Term often associated with metabolomics

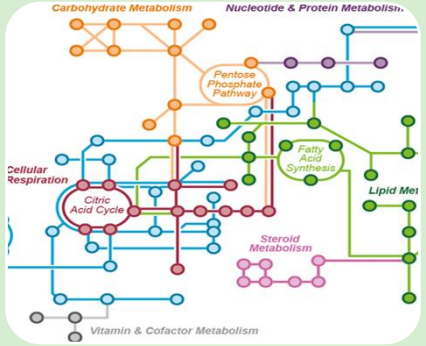


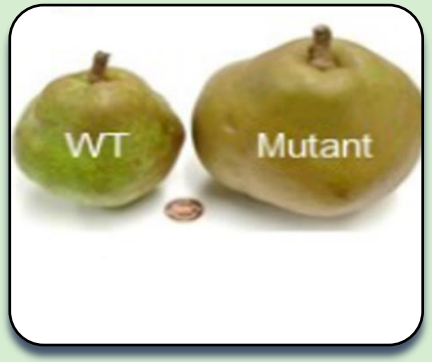
# Overview of Omics Technologies










## Genomics

- All genes



## Transcriptomics

- All messenger RNAs (mRNAs)



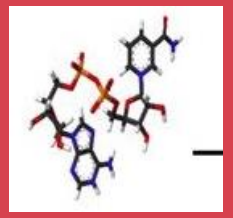
## Proteomics

- All proteins



## Metabolomics

- All metabolites

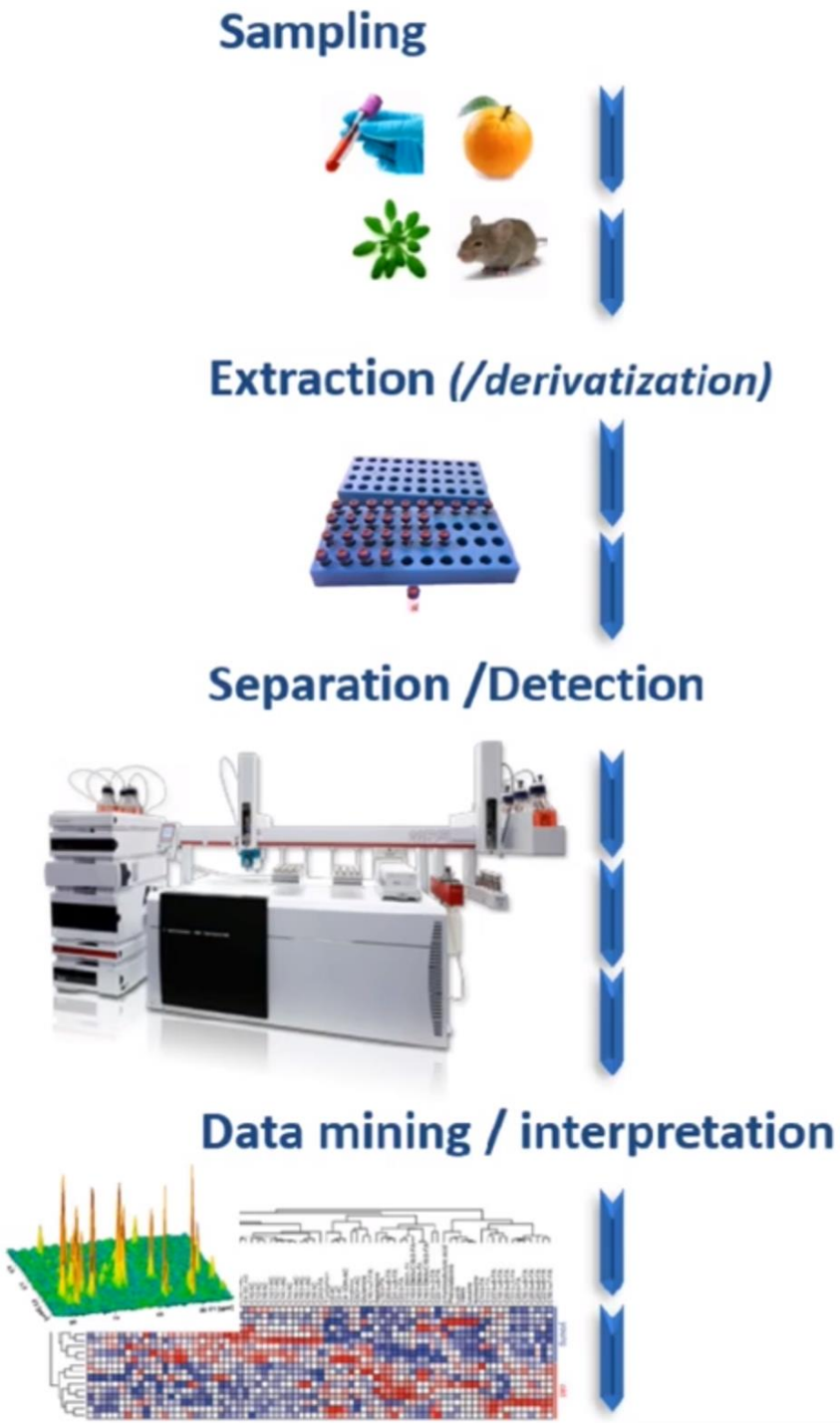


## Phenomics

- All phenotypes



## Overview of the metabolomics workflow & Separation and detection techniques in metabolomics




- **Separation**
  - **LC**
    - Liquid Chromatography
  - **GC**
    - Gaz chromatography
  - **CE**
    - Capillary electrophoresis
- **Detection (Quantification & Identification)**
  - **MS**
    - Mass spectrometry
  - **NMR**
    - Nuclear magnetic resonance
  - **Other Detectors**
    - UV, Fluo, FID …

# Today, the data comes from this type of analysis

## Separation

### LC
- Liquid Chromatography

### GC
- Gaz chromatography

### CE
- Capillary electrophoresis

## Detection (Quantification & Identification)

### MS
- Mass spectrometry

### NMR
- Nuclear magnetic resonance

### Other Detectors
- UV, Fluo, FID …








# Today, the data comes from this type of analysis

Separation



LC

- Liquid Chromatography

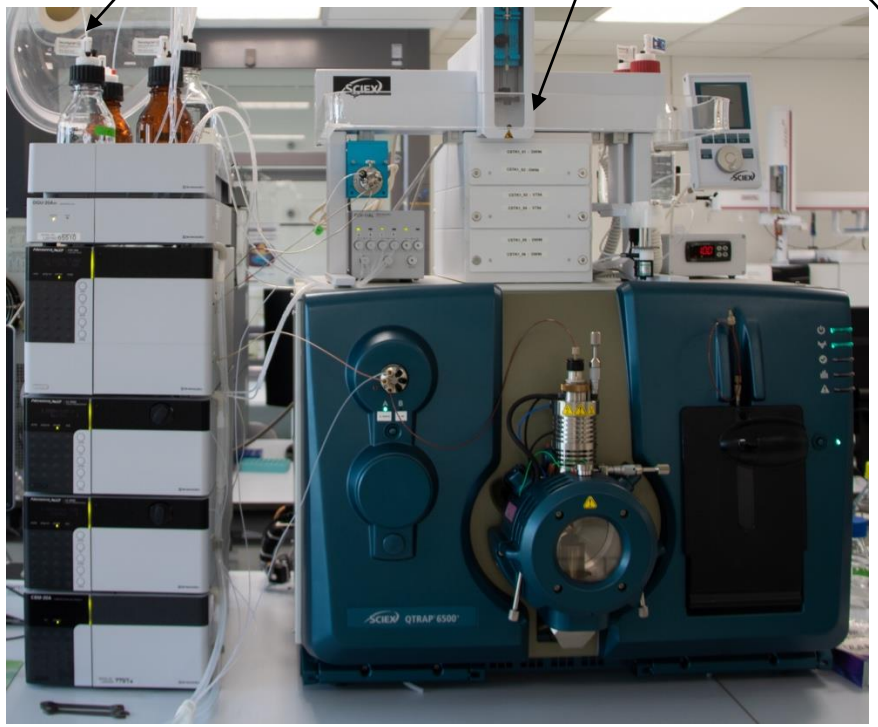
**Column C18**



**binary pump**

**Sample Feeder / Injector**

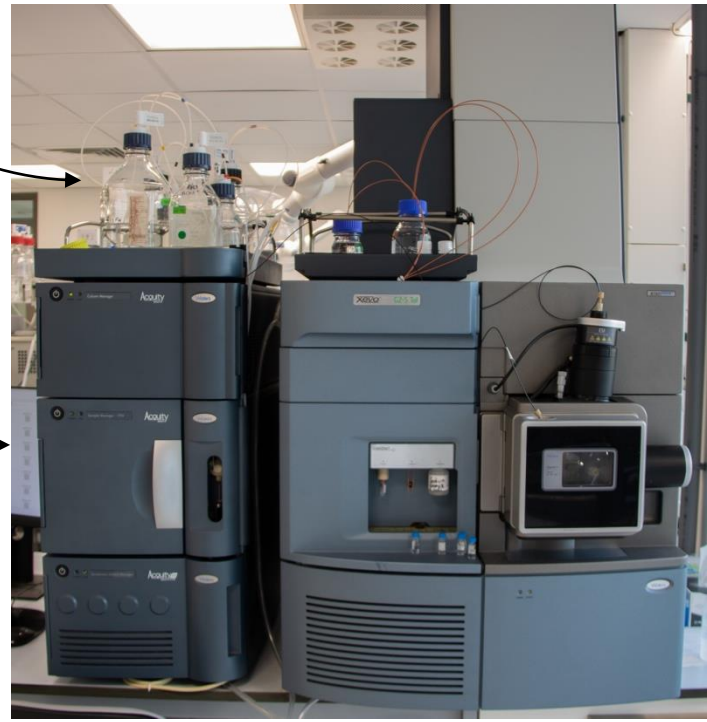
**Mobile phase:**
$H_2O$, Methanol
and Formic Acid, Acetonitrile



Q-TRAP

**Column C18**

**Binary pump**



**Sample Feeder / Injector**

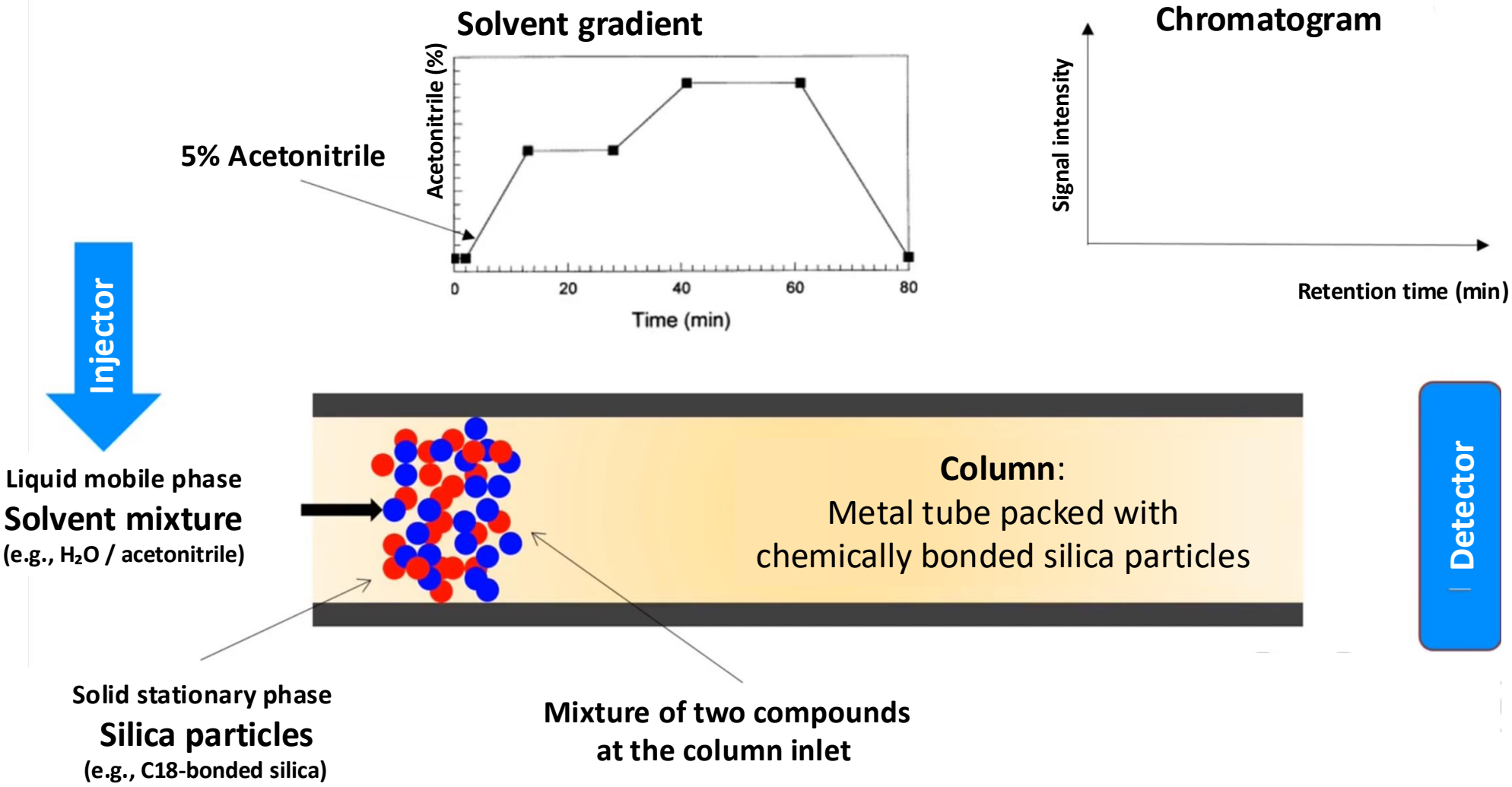
Q-TOF

# Liquid Chromatography: Separation Based on Polarity

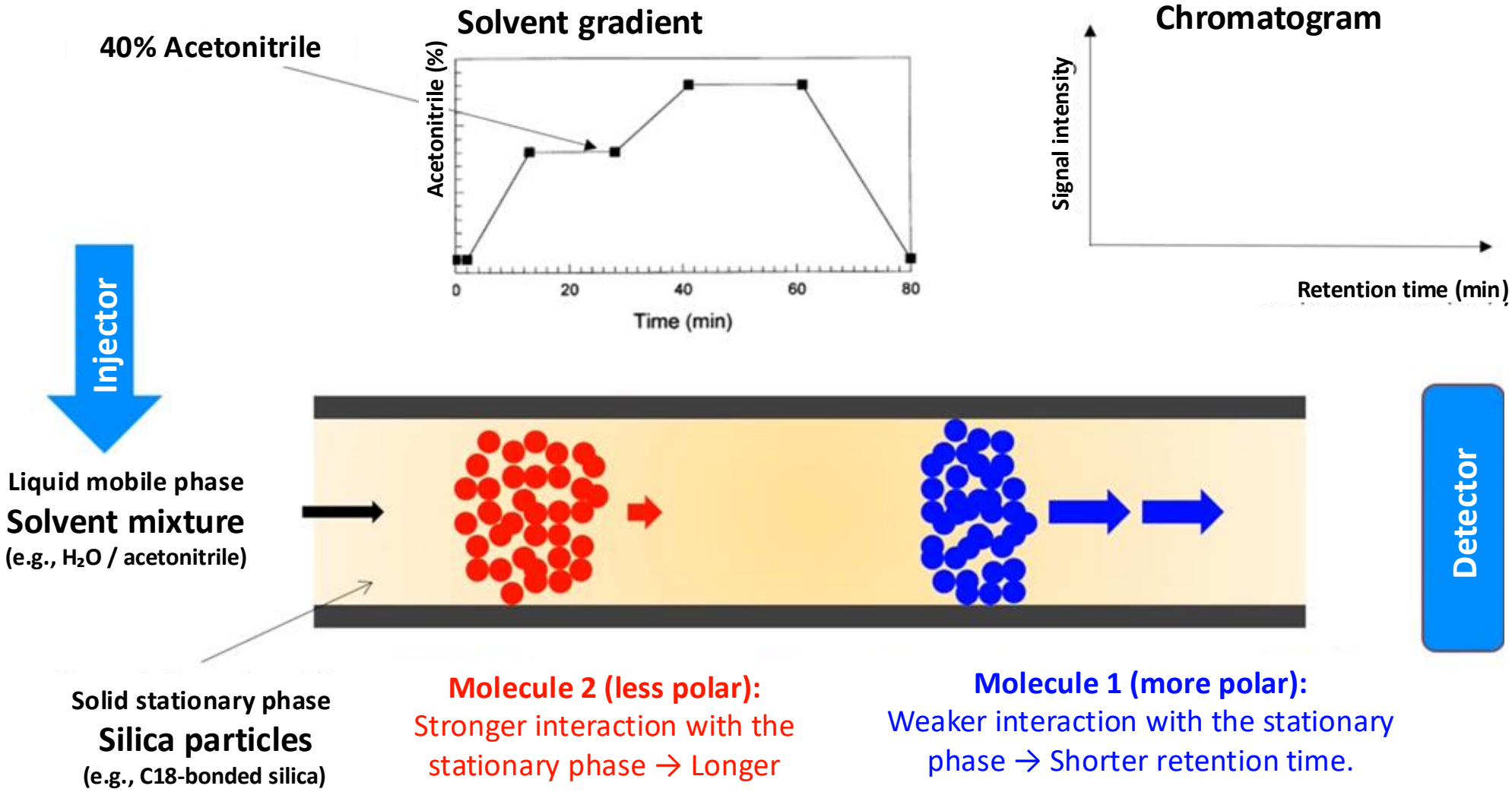



- Hydrophilic interactions, hydrogen bonding, net charge, hydrophobic interactions, etc.
- HPLC, LC, UPLC, LC-UV, LC-MS/MS, LC-DAD-MS/MS, LC-QTOF, LC-3QMS, LC-ESI-MS, etc.
- Columns packed with chemically bonded silica microparticles
- Aqueous and organic solvents ($H_2O$, MeOH, ACN, isopropanol, etc.)
- Soluble compounds

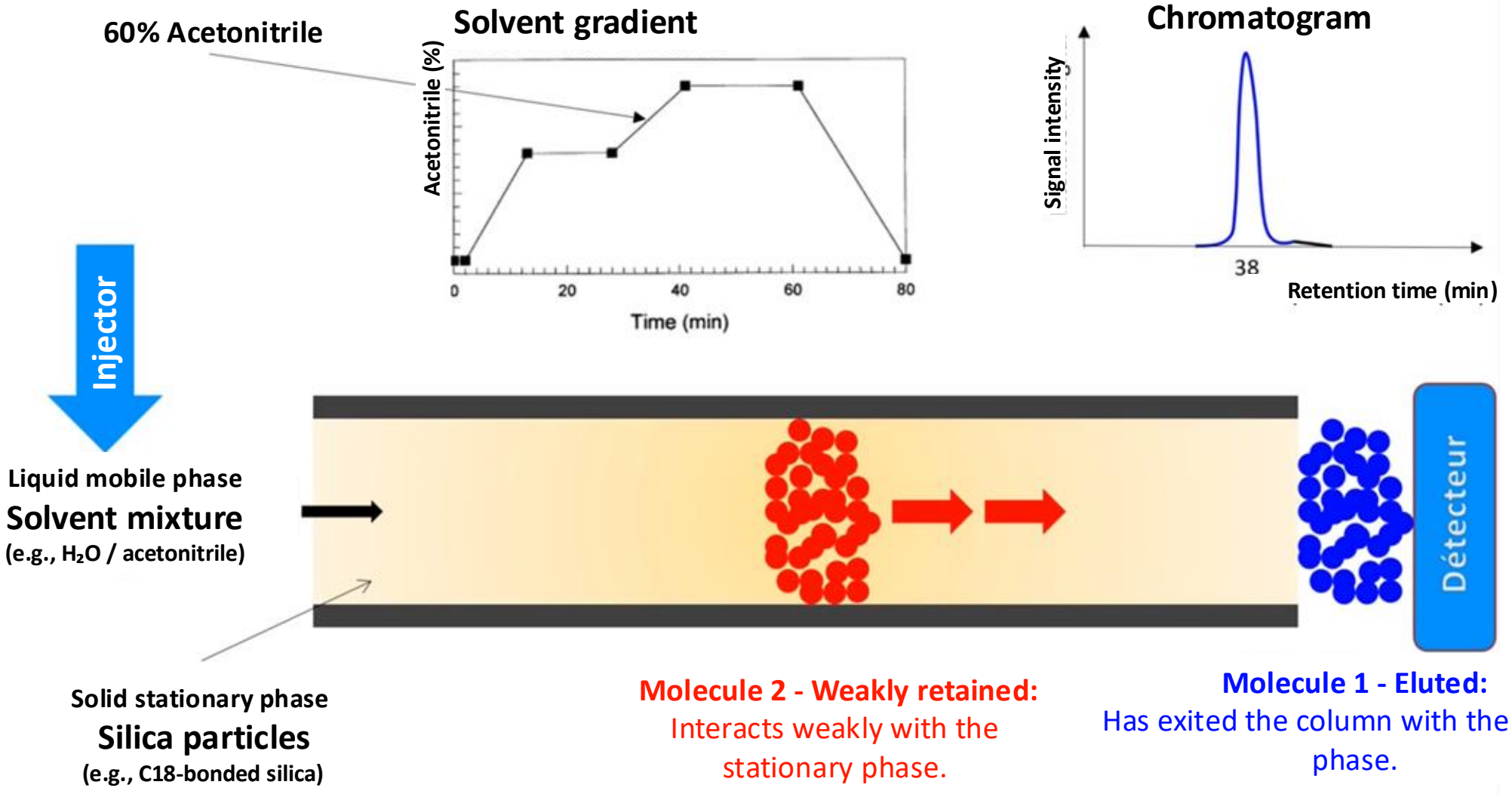
# A chromatographic column ?



# A chromatographic column ?



# A chromatographic column ?

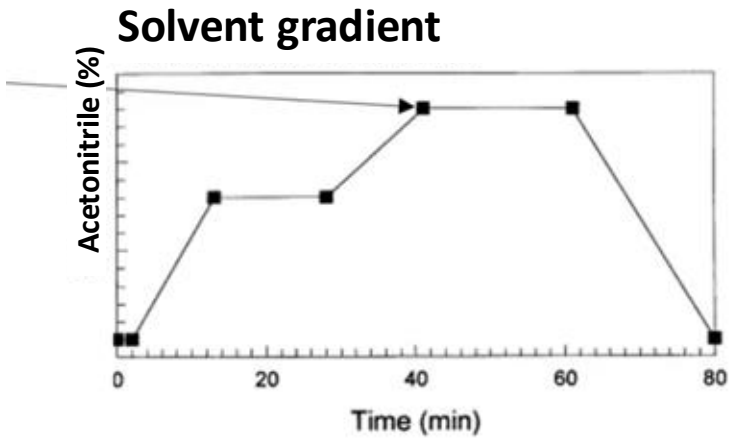


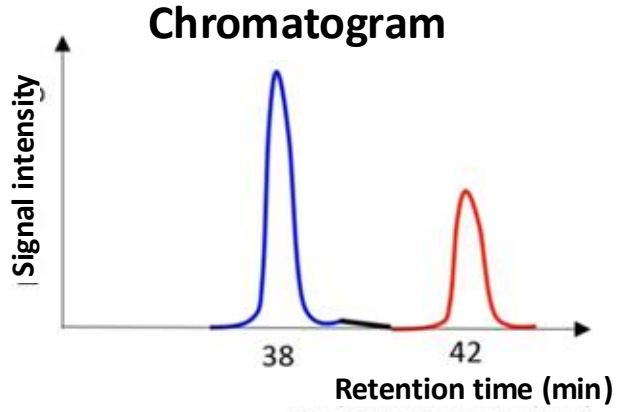

**Molecule 2 - Weakly retained:**
Interacts weakly with the stationary phase.

**Molecule 1 - Eluted:**
Has exited the column with the mobile phase.

# A chromatographic column ?

**80% Acetonitrile**




Injector

**Liquid mobile phase**

**Solvent mixture**

**(e.g., $H_2O$ / acetonitrile)**




**Solid stationary phase**

**Silica particles**

**(e.g., C18-bonded silica)**

**Molecule 2 - Eluted:**
**Has exited the column with the mobile phase.**

# Liquid Chromatography: Separation Based on Polarity & Gas Chromatography: Separation Based on Volatility

**Separation**

- **LC**
  - Liquid Chromatography
- **GC**
  - Gaz chromatography
- **CE**
  - Capillary electrophoresis







- Hydrophilic interactions, hydrogen bonding, net charge, hydrophobic interactions, etc.
- HPLC, LC, UPLC, LC-UV, LC-MS/MS, LC-DAD-MS/MS, LC-QTOF, LC-3QMS, LC-ESI-MS, etc.
- Columns packed with chemically bonded silica microparticles
- Aqueous and organic solvents ($H_2O$, MeOH, ACN, isopropanol, etc.)
- Soluble compounds

**<-- Gas Chromatography: Separation Based on Vaporization Temperature**

# Detection (Quantification & Identification)

## Separation

### LC
- Liquid Chromatography

### GC
- Gaz chromatography

### CE
- Capillary electrophoresis

## Detection (Quantification & Identification)

### MS
- Mass spectrometry

### NMR
- Nuclear magnetic resonance

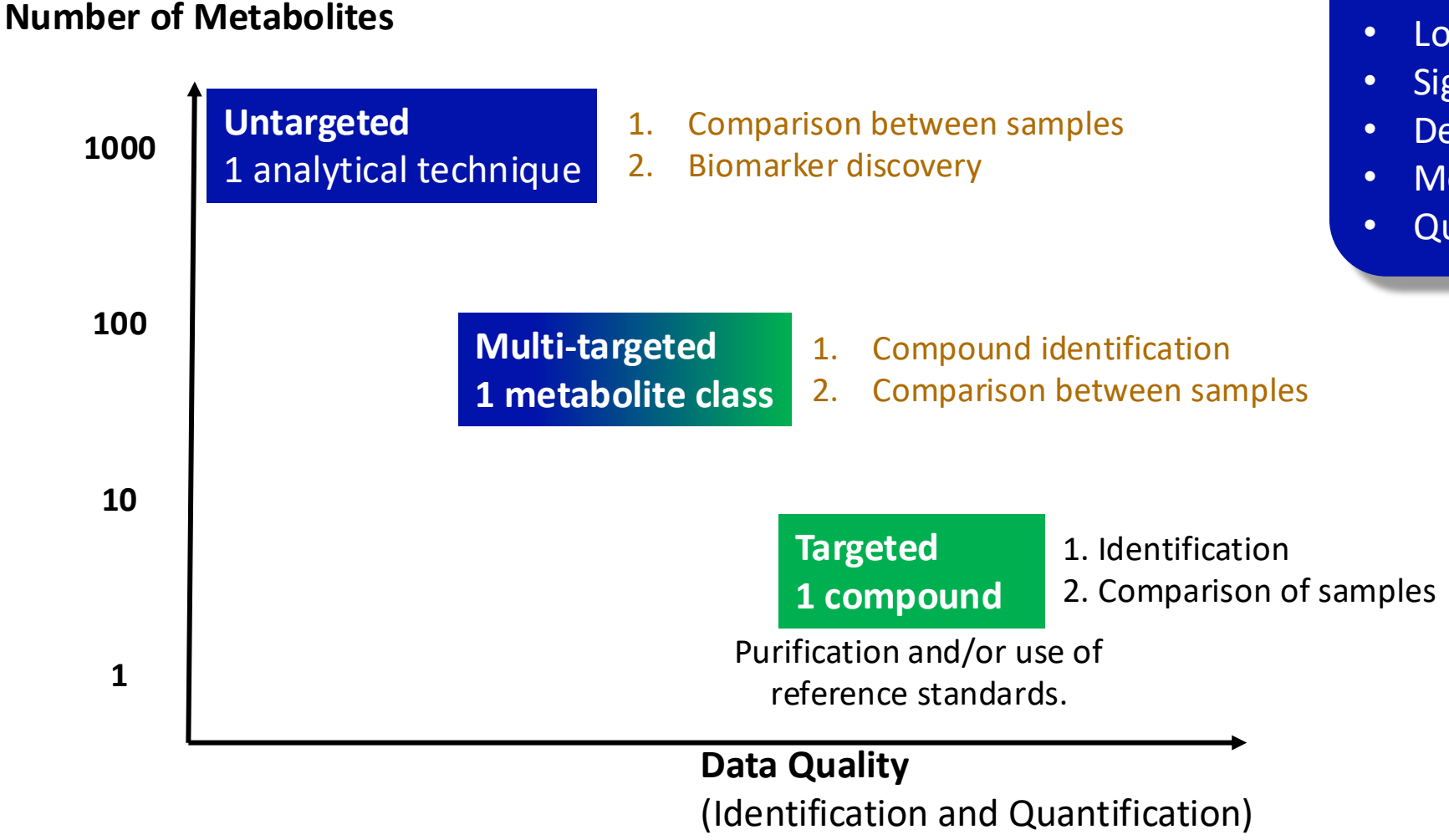
### Other Detectors
- UV, Fluo, FID …








# Metabolomics Approaches: Coverage vs. Data Quality



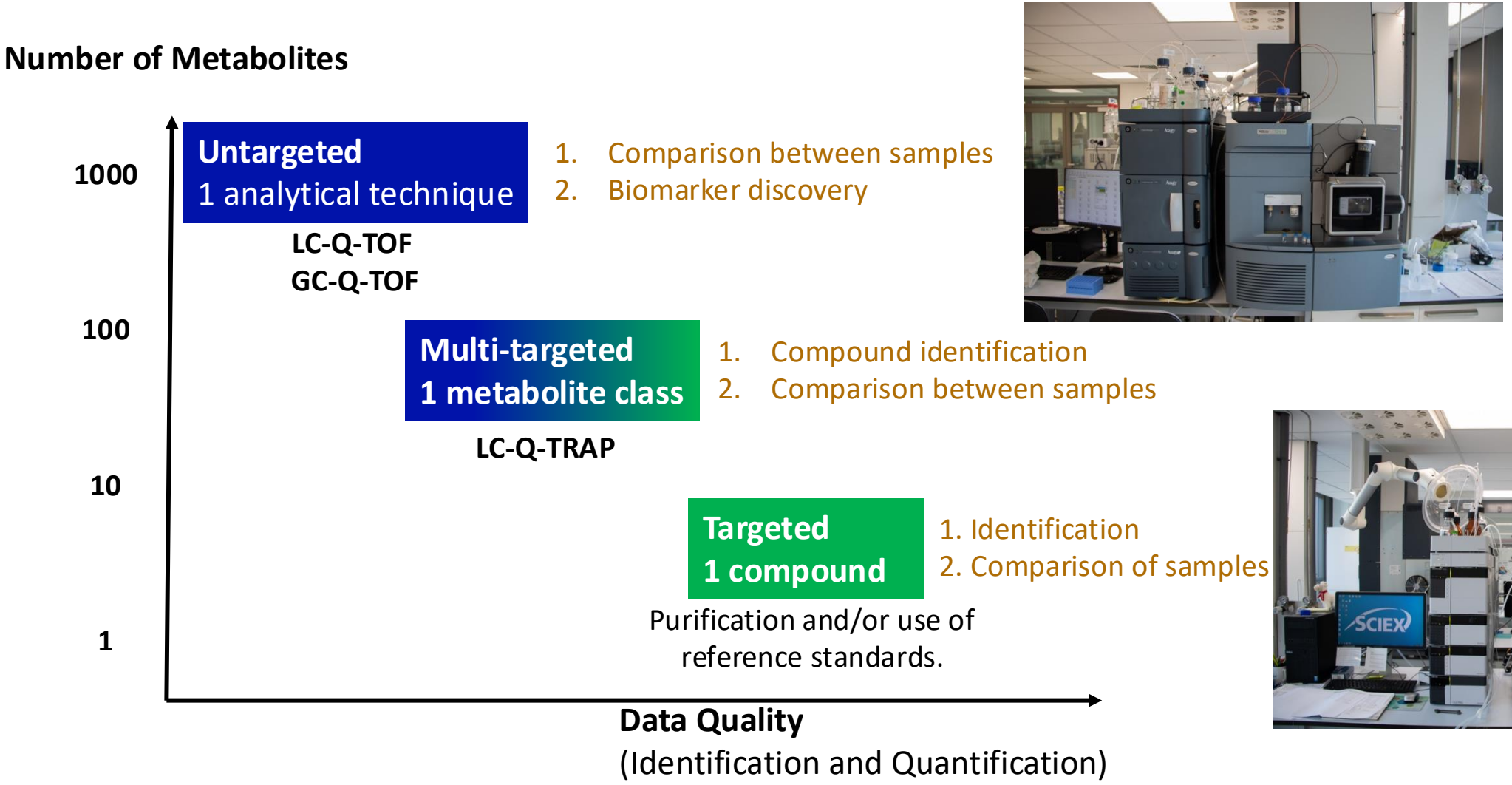

**Caution:**

- Loss of compartmentalization
- Signal too weak
- Device sensitivity
- Molecule identification
- Quantification

**Why? :**

- Metabolites are highly diverse (hundreds of thousands)
- Concentrations vary widely (glucose vs. hormones)
- Metabolism is highly dynamic
- Metabolism is compartmentalized (often disrupted)

# Metabolomics Approaches: Coverage vs. Data Quality








# The mass spectrometry, Measures the total mass of the molecules

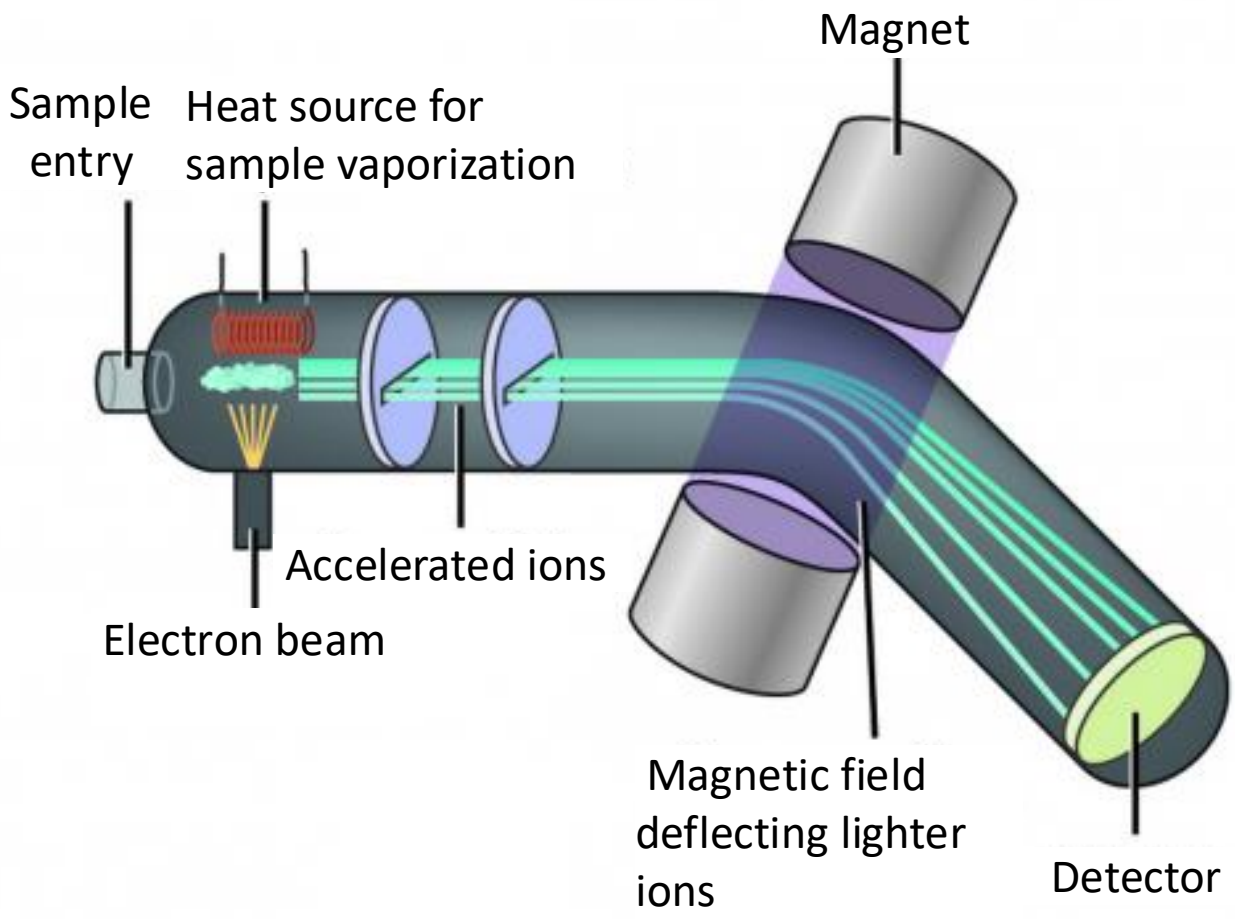
**Mass spectrometer**





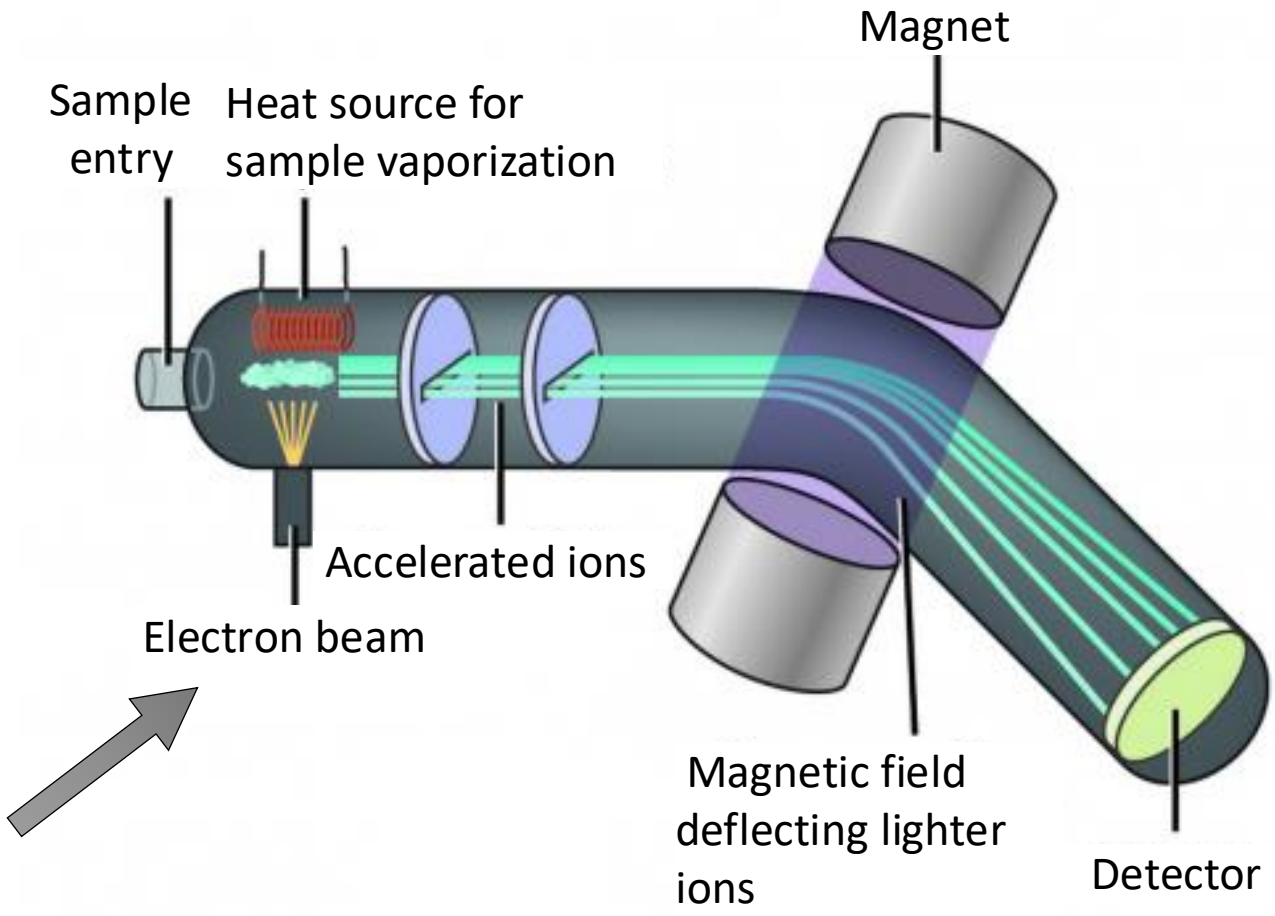

**Measurement of the mass of charged molecules by studying their motion in a calibrated electromagnetic field**

# Mass Spectrometry, Ionization, and Analysis




- Mass spectrometry measures the mass-to-charge ratio of an ion: m/Z
- In metabolomics, molecules are by definition small, and the probability that they will be multiply charged is virtually zero

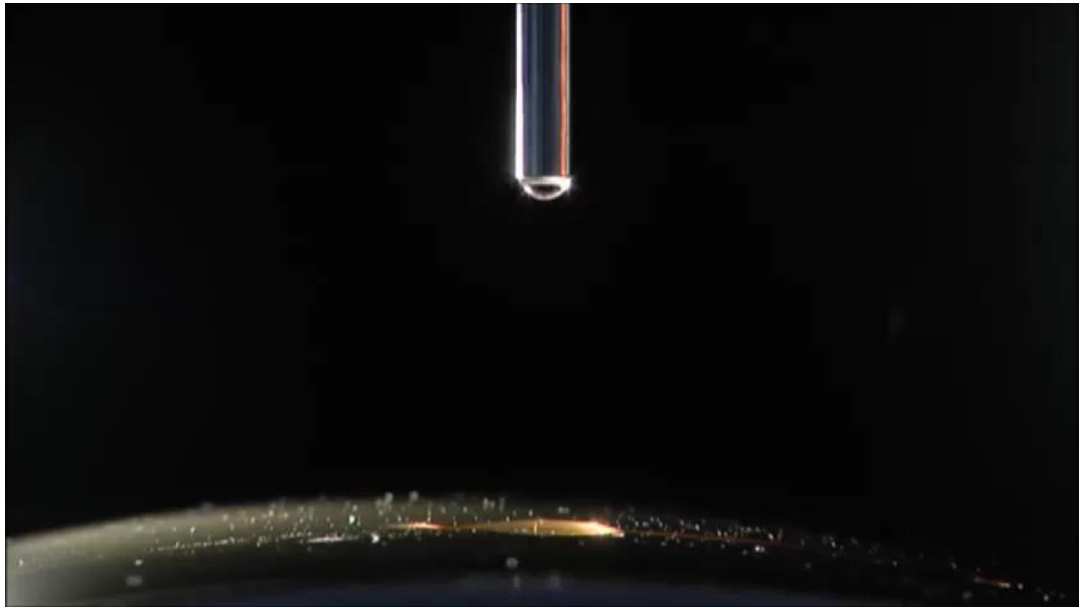
# Mass Spectrometry, Ionization, and Analysis



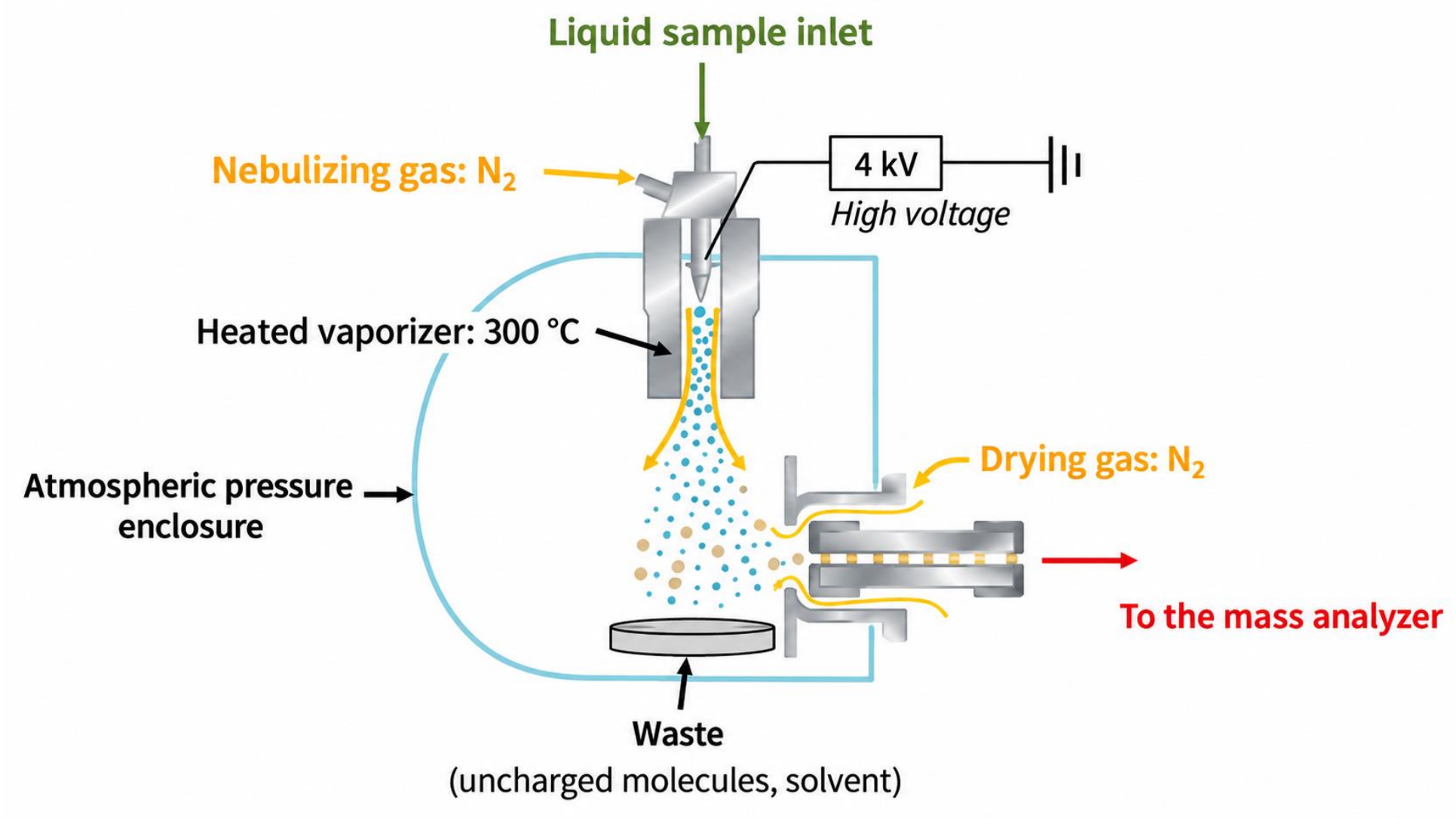

- Mass spectrometry measures the mass-to-charge ratio of an ion: m/Z
- In metabolomics, molecules are by definition small, and the probability that they will be multiply charged is virtually zero







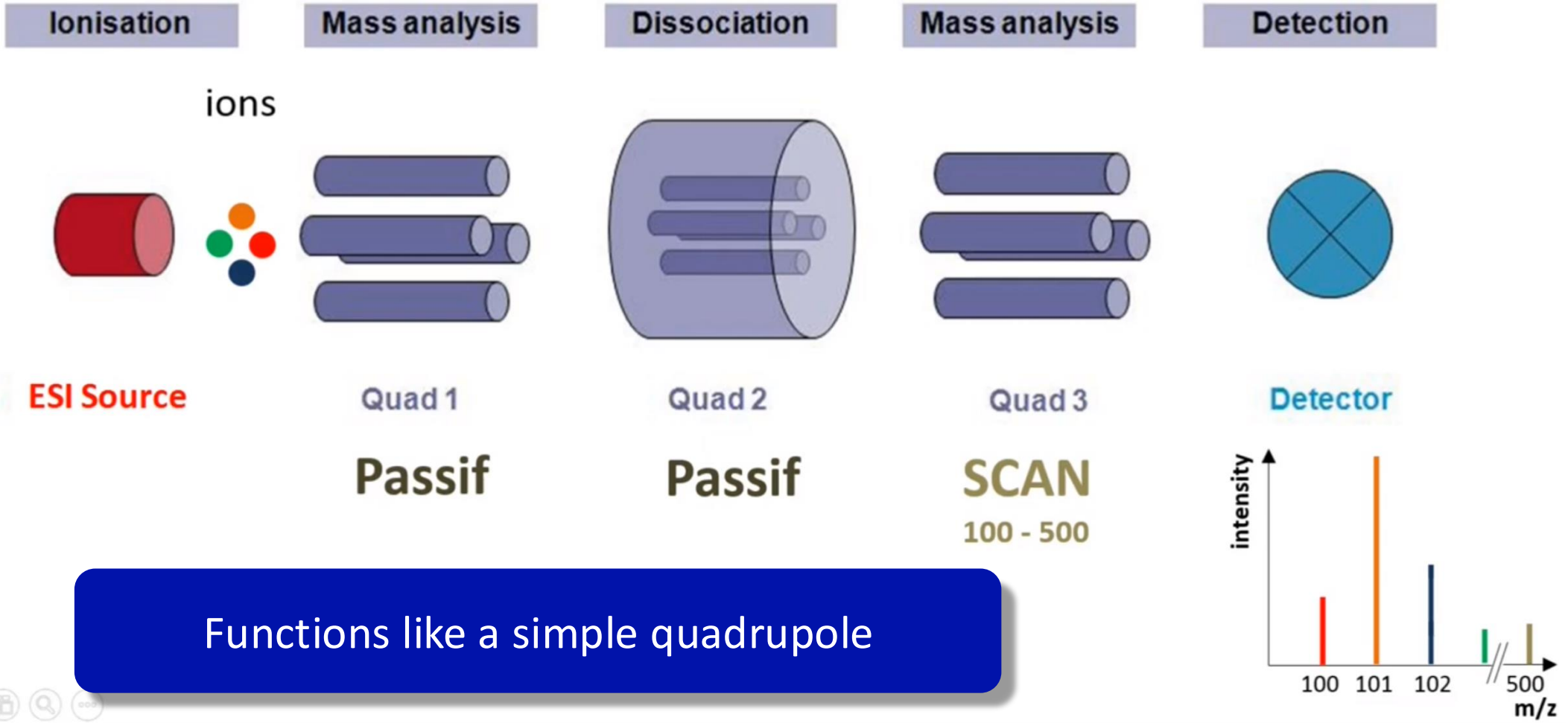
# The Q-TRAP


Liquid sample inlet
Nebulizing gas: $N_2$
4 kV
High voltage
Heated vaporizer: 300 °C
Drying gas: $N_2$
Atmospheric pressure enclosure
To the mass analyzer
Waste
(uncharged molecules, solvent)

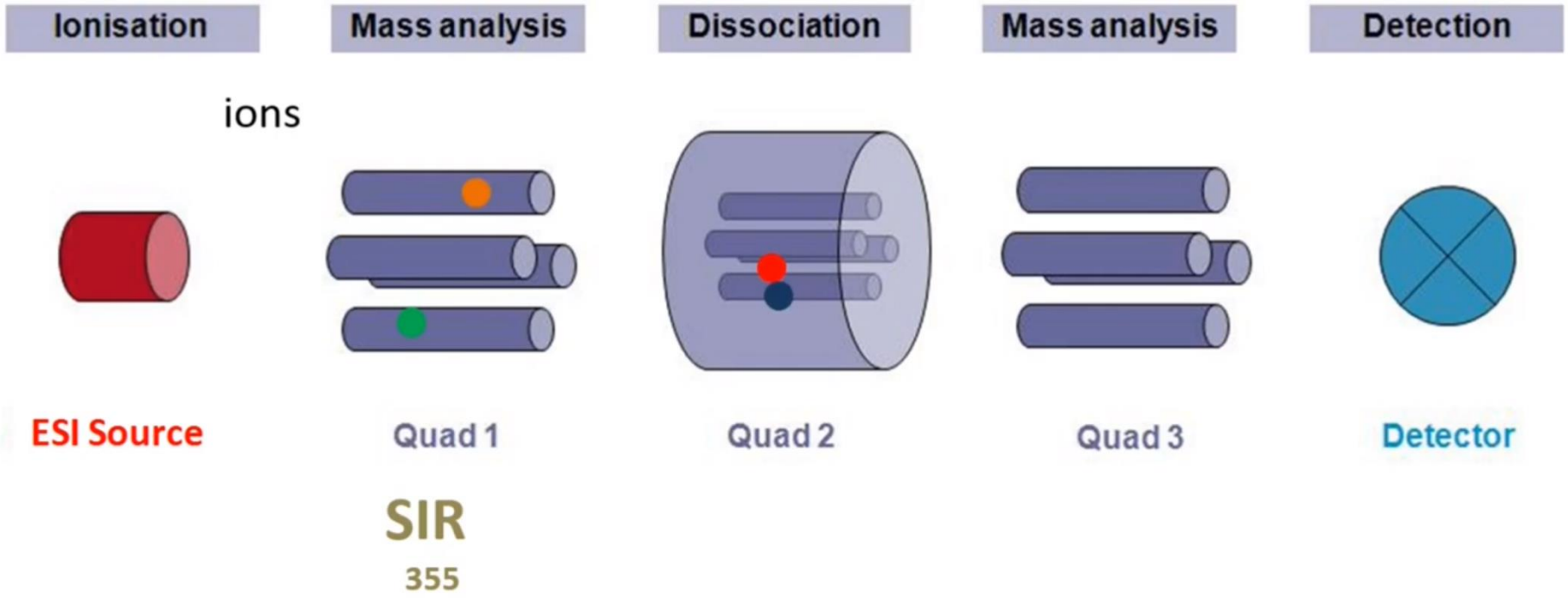
# The Q-TRAP



# How a Triple Quadrupole Mass Spectrometer Works in Full Scan Mode




Functions like a simple quadrupole

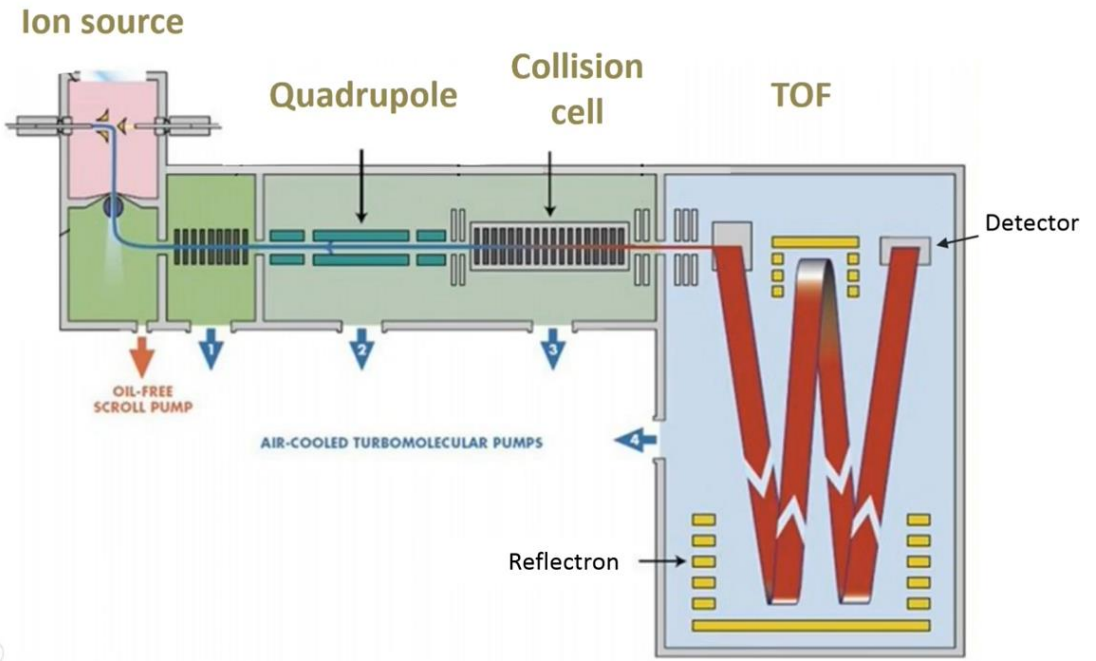



Selection of parent ions with m/z = 355. Here, two different ions have an m/z of 355.

# The Q-TOF, Time-of-Flight (TOF) Mass Spectrometer





# Challenges and Limitations of Metabolomics

**1. Technical Challenges**

- Wide range of metabolite concentrations
- Limited instrument sensitivity
- Loss of spatial and cellular information
- Highly dynamic metabolism

**2. Data Analysis & Interpretation**

- Complex data processing and software
- Difficult metabolite identification and annotation
- Sample-specific databases required
- Challenging biological interpretation

**3. Targeted Metabolomics**

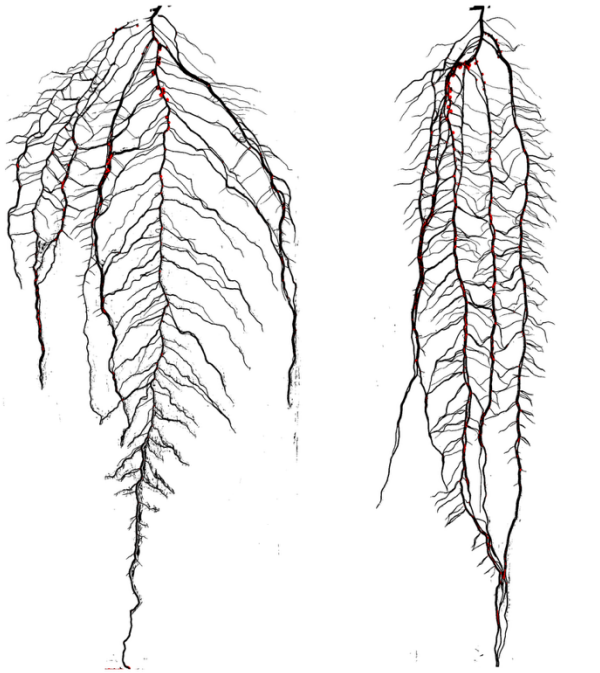
- Requires method optimisation for each metabolite or compound family
- Reference standards needed for reliable identification and quantification
- Time-consuming and expensive

Metabolomics is a powerful but technically demanding approach, requiring continuous methodological development and careful biological interpretation.

# From one variable to many

- Session 4: one variable (Shannon diversity), one test at a time.
- Today: 34 metabolites in the root, 2,500 features in the exudates, measured on the same samples.
- Two problems: we cannot look at them one by one, and testing each one gives false positives.

# Experimental design of the experiment

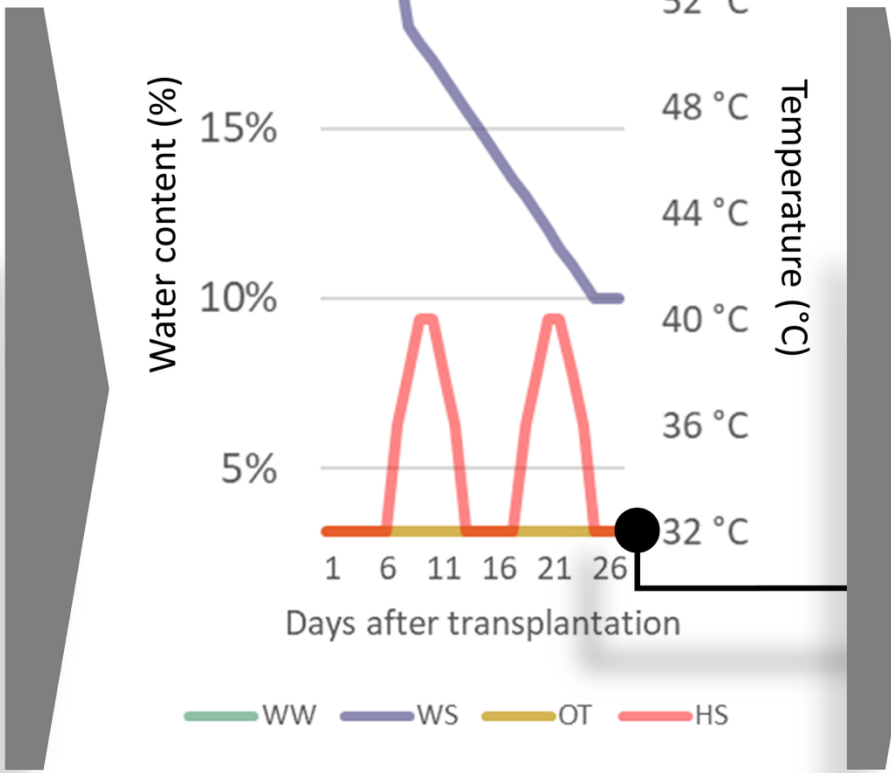


Stocata

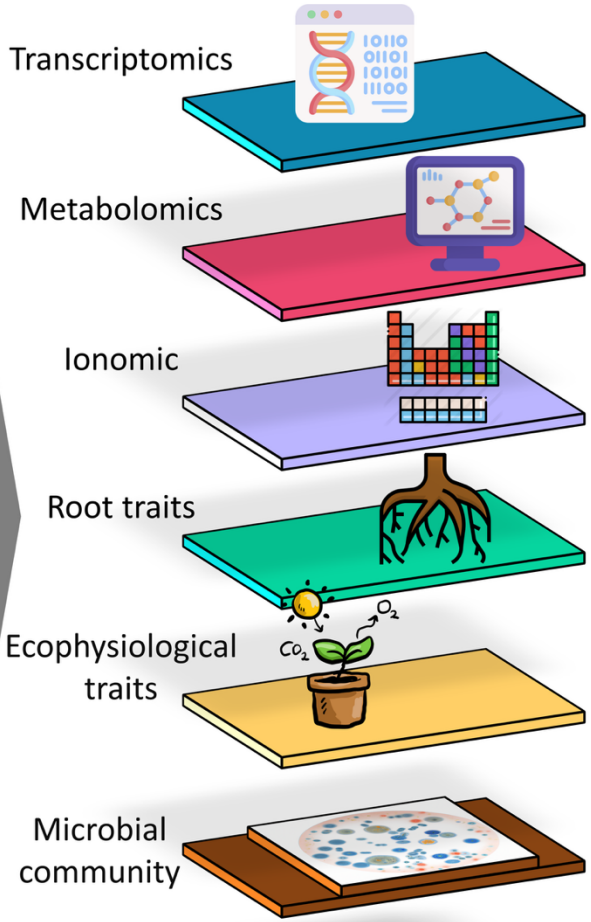


Wendy

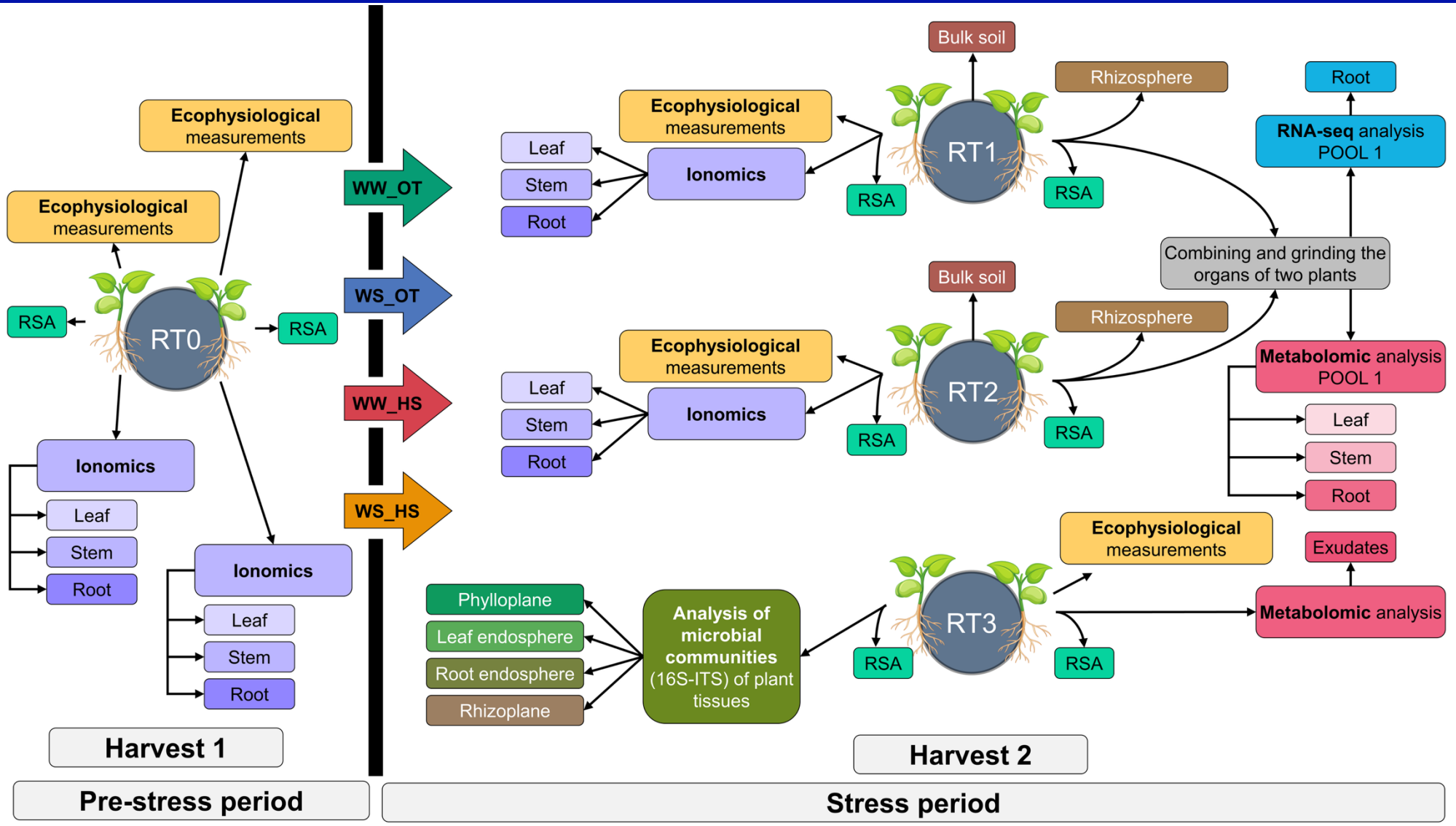
**Two genotypes :**
Use of two soybean genotypes with contrasting architectures




**4 climatic conditions :**
WW_OT / WS_OT / WW_HS / WS_HS



# Harvest diagram for the experiment



# Results with QTRAP, real quantification

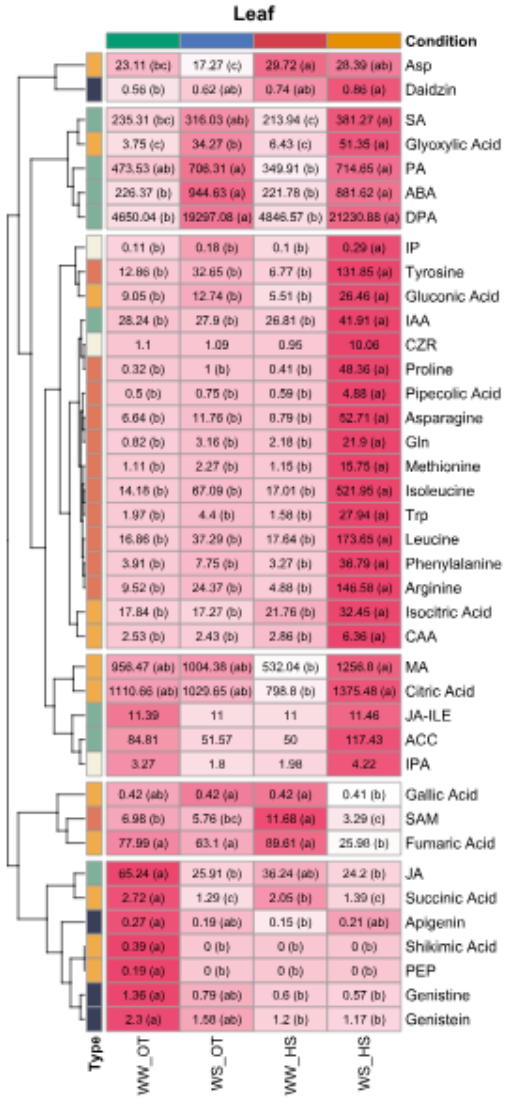
**Condition**

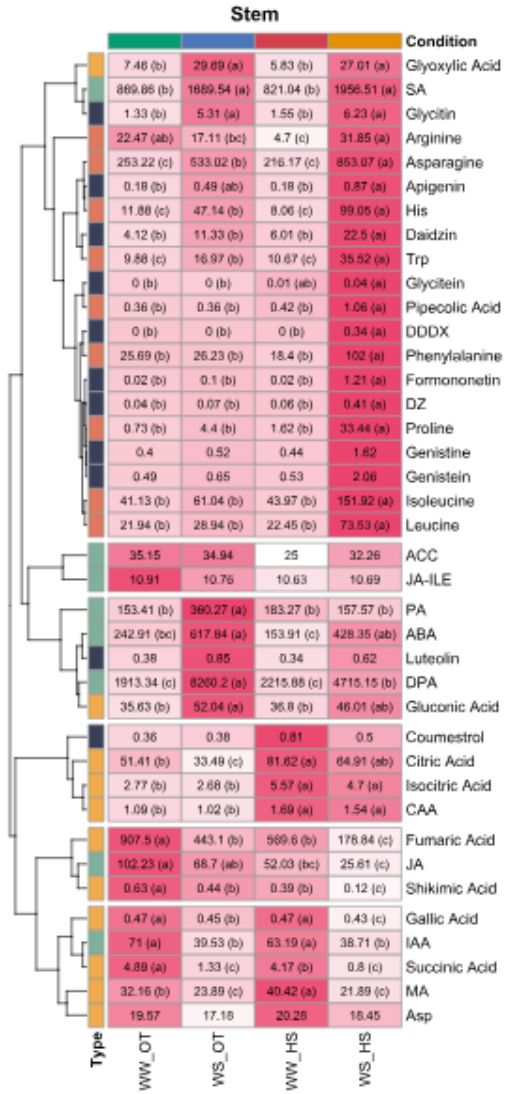
- WS_HS
- WS_OT
- WW_HS
- WW_OT

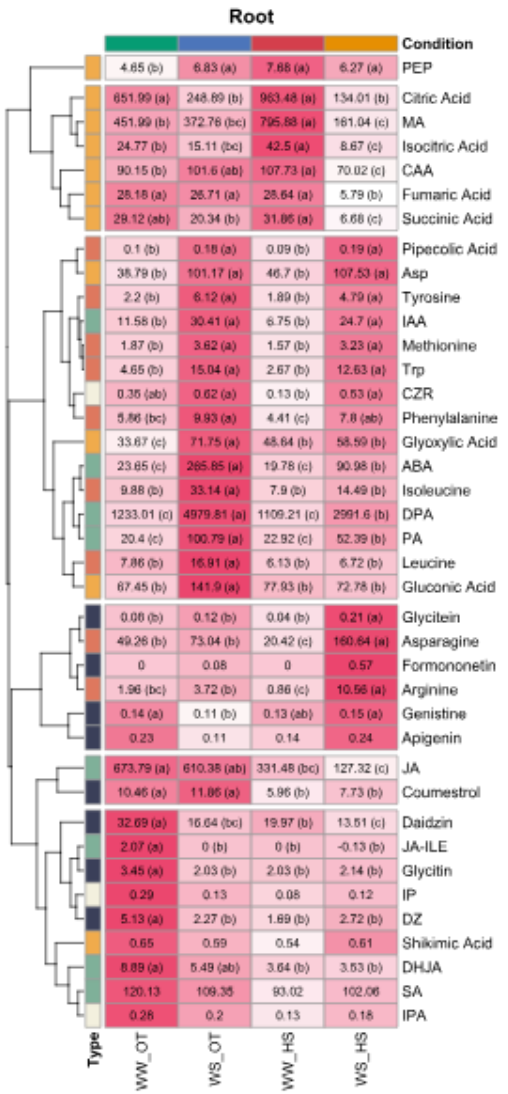
**Type**

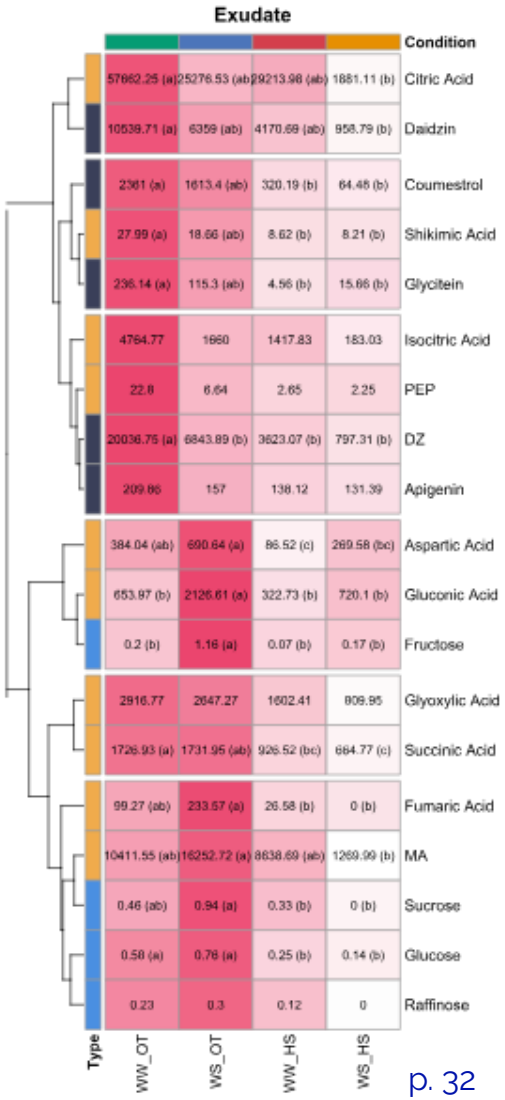
- Amino acid
- Cytokinine
- Flavonoid
- Organic acid
- Phytohormone
- Sugar



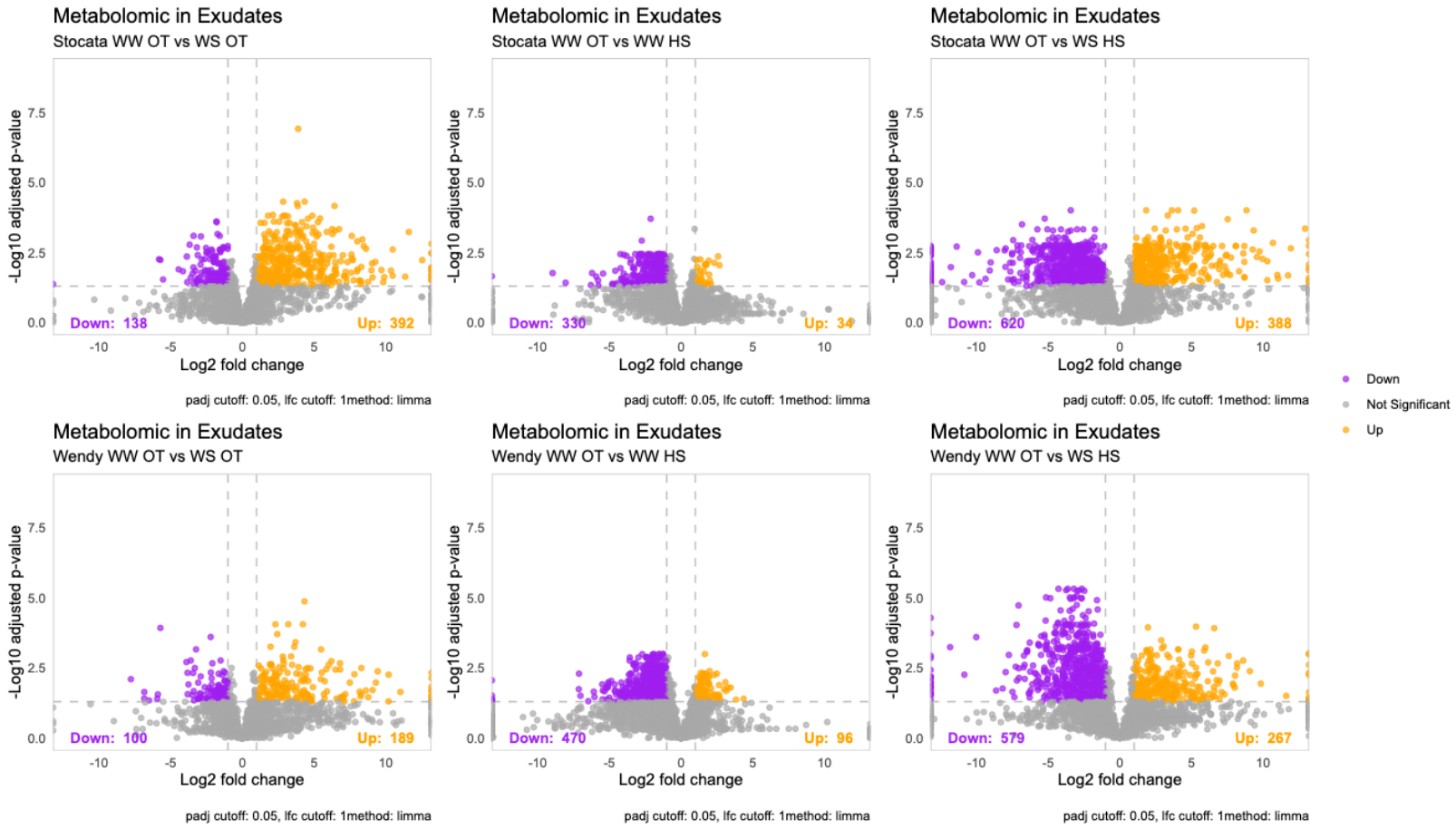


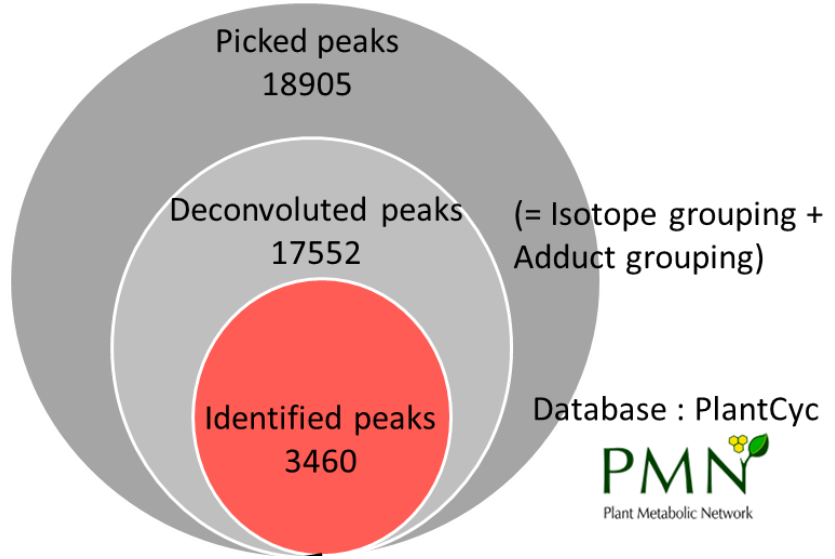



# Results with Q-TOF






# The data you already have

**One row per sample**

- Your table is long: one row per sample × metabolite.
- A multivariate method needs a matrix: one row per sample, one column per variable.
- pivot_wider(names_from = metabolite, values_from = log_conc)
- Then numbers only: as.matrix(), sample names as row names

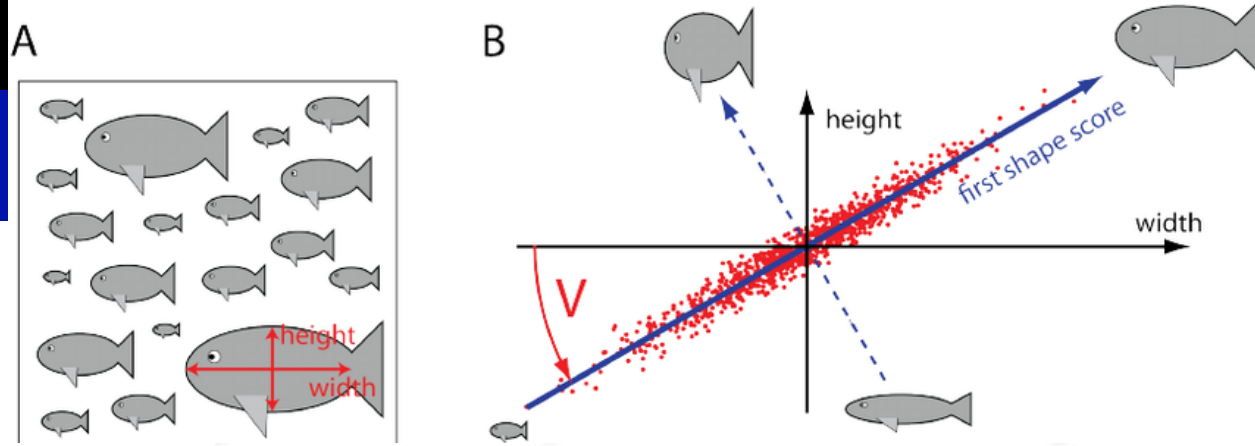
**log, then scale**

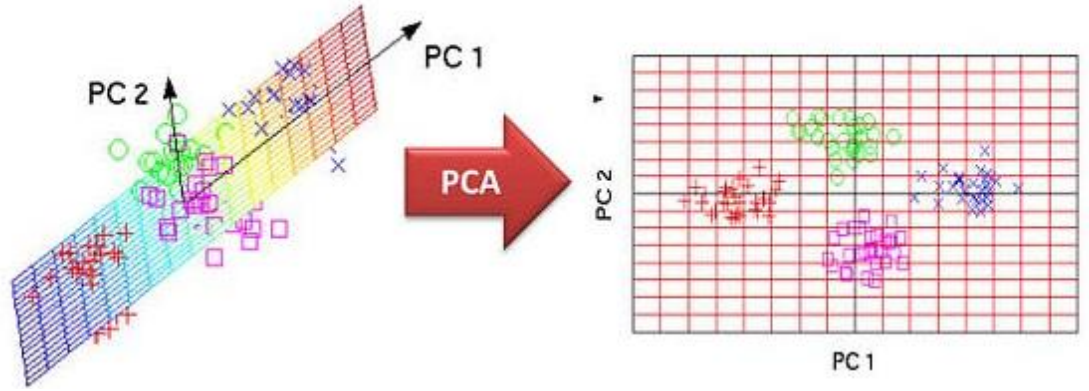
- Concentrations range from 0.02 to 190,000: the big molecules would decide everything.
- log10() compresses the range: a ×10 change becomes +1 whatever the molecule.
- scale. = TRUE: each variable gets mean 0 and standard deviation 1, so every variable has the same weight.
- You will see why in the warm-up: athletes measured in seconds and in metres.

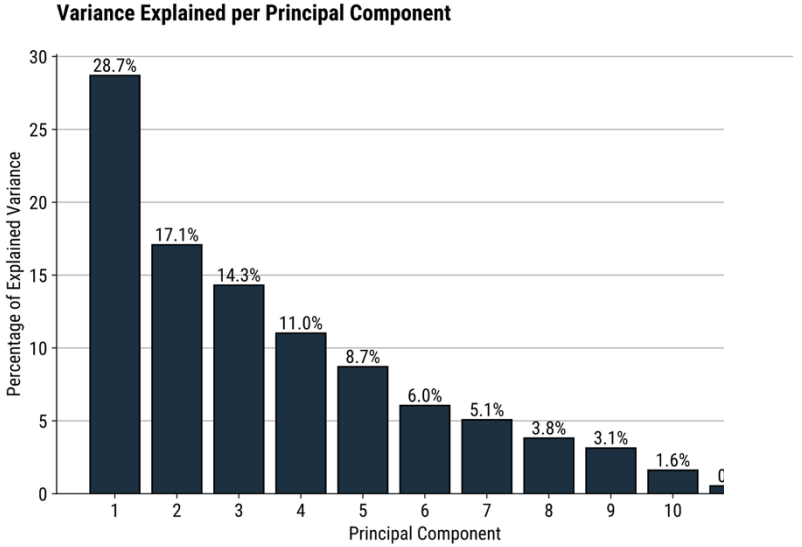
# PCA: the idea

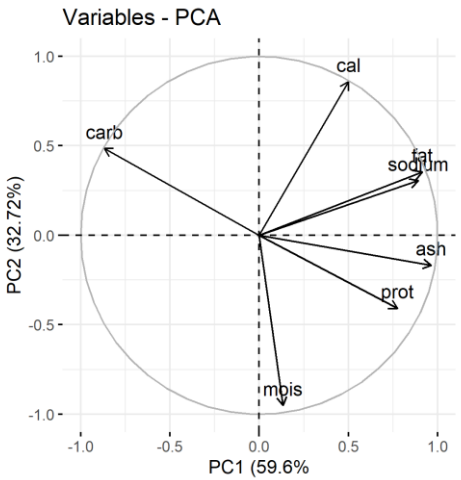
- Many variables, but they are correlated: they tell partly the same story.
- PC1 = the direction in which the samples differ the most. PC2 = the next one, at right angles.
- Each axis has a share of variance: how much of the differences it captures.
- prcomp(matrix, scale. = TRUE)

- Scores (pca$x): one point per sample. Close points = similar samples.
- Loadings (pca$rotation): one value (or arrow) per variable. They say which variables push the samples along each axis.
- Unsupervised: the PCA never sees your groups. Colour by condition afterwards: if the groups separate, the condition is a major source of variation.



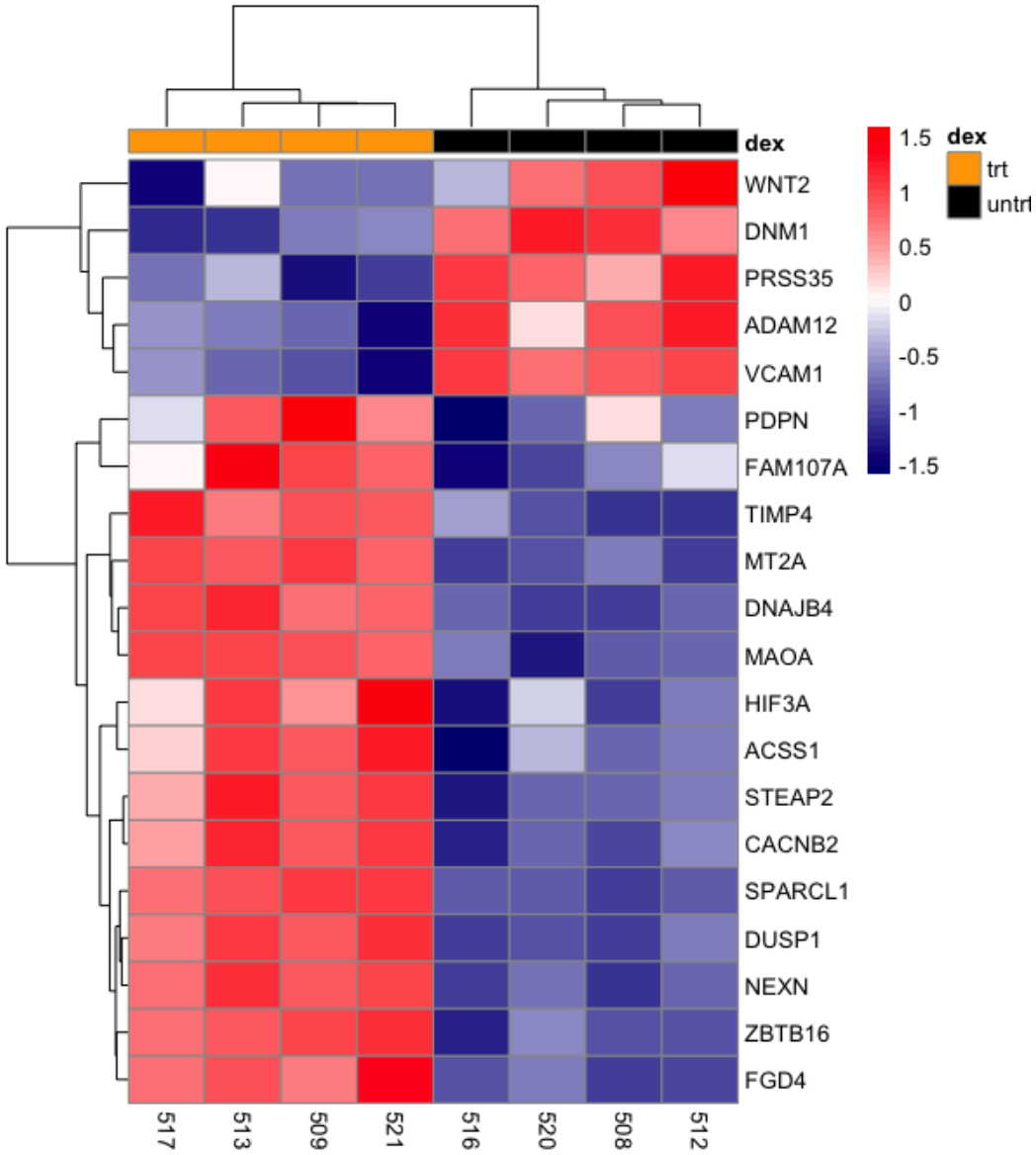
# Heatmaps & Hierarchical Clustering

**1. Heatmap**
Visualises data using a colour scale.
Rows: genes, metabolites or proteins.
Columns: samples or experimental conditions.
Colours represent relative abundance or expression levels.

**2. Hierarchical Clustering**
Groups samples or molecules based on their similarity.
A dendrogram illustrates relationships between groups.
Similar profiles are clustered together.

**3. Number of Clusters**
Depends on the similarity threshold (dendrogram cut-off).
Can be determined using statistical methods (e.g., silhouette analysis).
More clusters reveal finer patterns but may reduce interpretability.

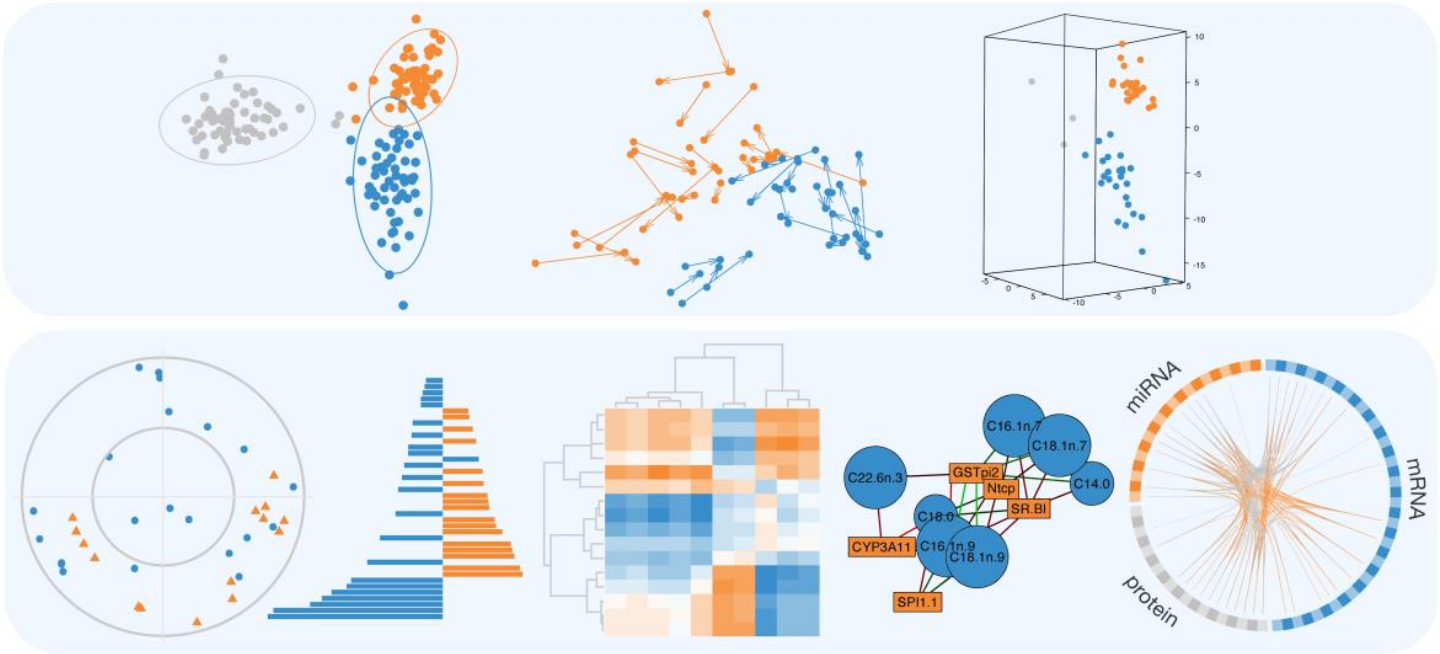
Heatmaps reveal patterns, while clustering identifies groups with similar biological profiles

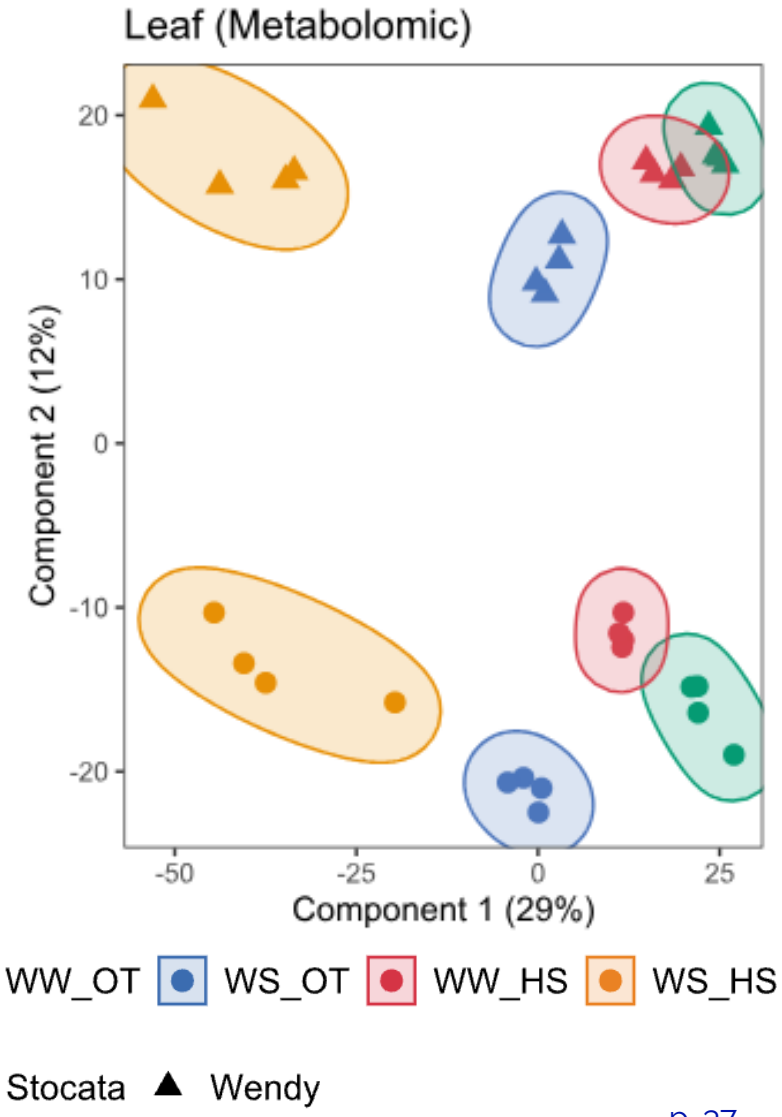


# Supervised: PLS-DA

- Now we tell the method the groups (drought / watered).
- It finds the combination of variables that best separates them.
- plsda(matrix, groups, ncomp = 2)
- Problem: it always finds something.

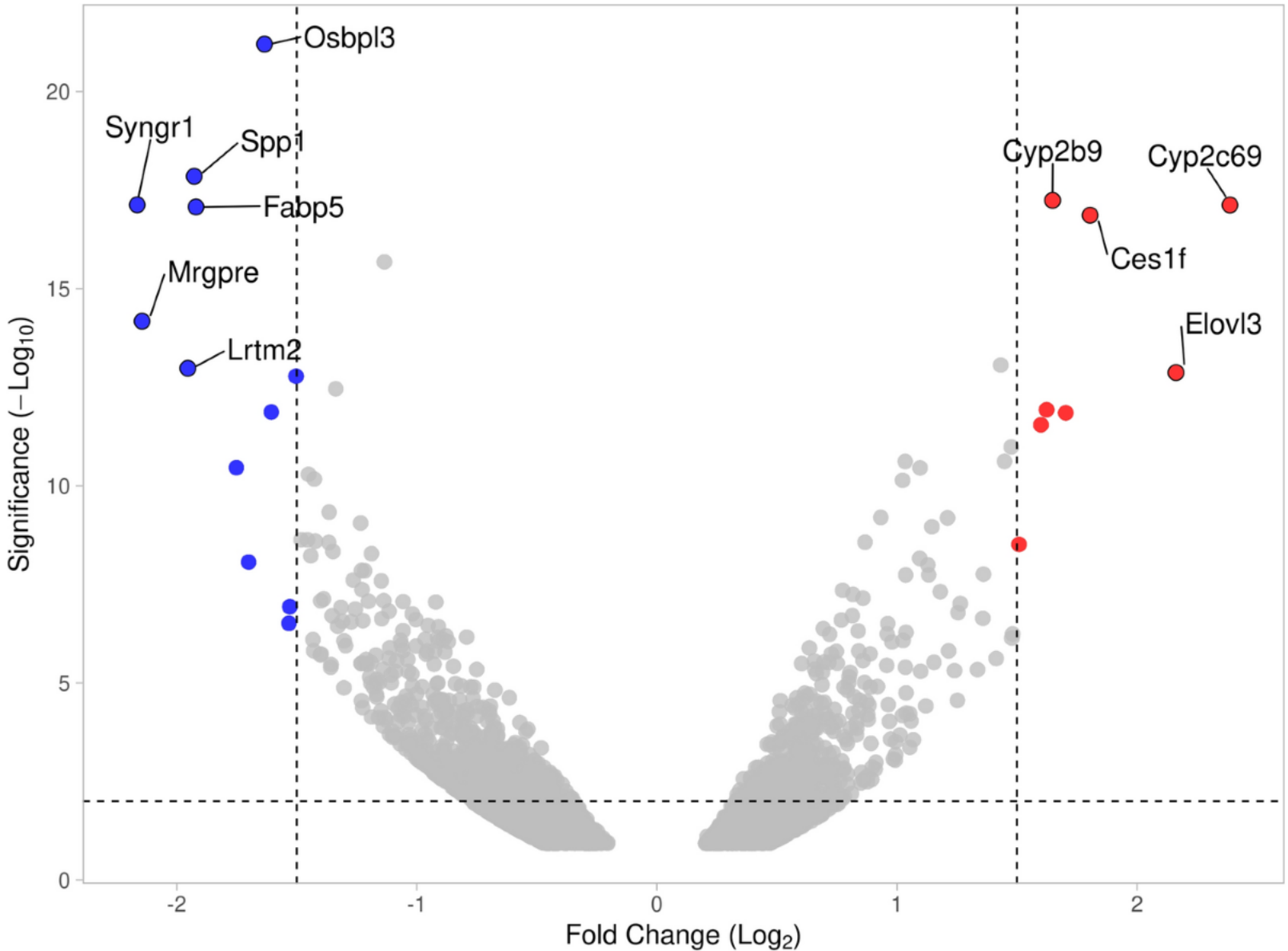
https://mixomics.org/



# Many tests at once: volcano and FDR

- One t-test per metabolite, as in Session 4. But 2,500 tests at 5% give about 125 "significant" results by chance alone.
- Correct the p-values: p.adjust(p_value, method = "BH") (false discovery rate).
- Volcano plot: x = effect (log2 fold change), y = evidence (−log10 adjusted p). Top left and top right = strong and reliable changes.



# Which method, when?

| Your question | Method | Supervised? |
|---|---|---|
| What are the main patterns? Do samples group? | PCA | no |
| Which samples and variables look alike? | heatmap + clustering | no |
| Which combination separates known groups? | PLS-DA + cross-validation | yes |
| Which variables change, one by one? | tests + FDR, volcano | (a test) |

# Today's practical

- First, in the Console: install mixOmics (3 lines on the page, 1-2 min). Start it now, then do the Terminal part while it runs.
- B: single molecules. Warm-up: a PCA on decathlon athletes. C-D: the root, guided, in pairs.
- E (leaf) and F (exudates, targeted then untargeted): on your own, same reasoning.
- Fill every ___, click Render: it now makes a PDF.
- Submit WORKBOOK.pdf on Moodle by **Sunday 25 October**, evening. If you see errors, also submit your .qmd and your results/ folder.