Metabolomics

Johannes Rainer
Eurac Research, Bolzano, Italy
johannes.rainer@eurac.edu - github: jorainer - twitter: jo_rainer

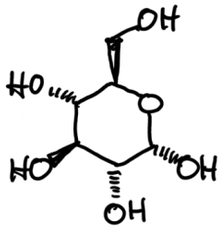
CSAMA 2019

Content

- (Brief) introduction to metabolomics
- Preprocessing of LC-MS data
- Normalization
- Annotation/identification

Metabolite? Metabolism?

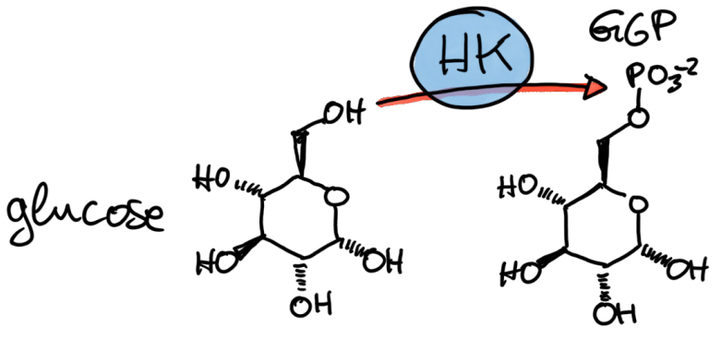
- Glycolysis

glucose 

- Key metabolic pathway common to all cells.
- Creates energy by converting glucose to pyruvate.

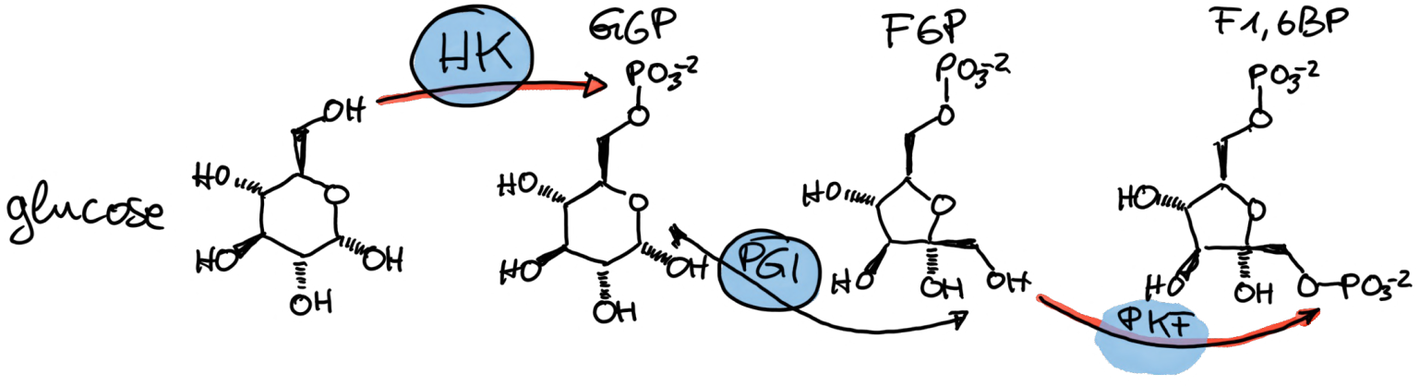
Metabolite? Metabolism?

- Glycolysis



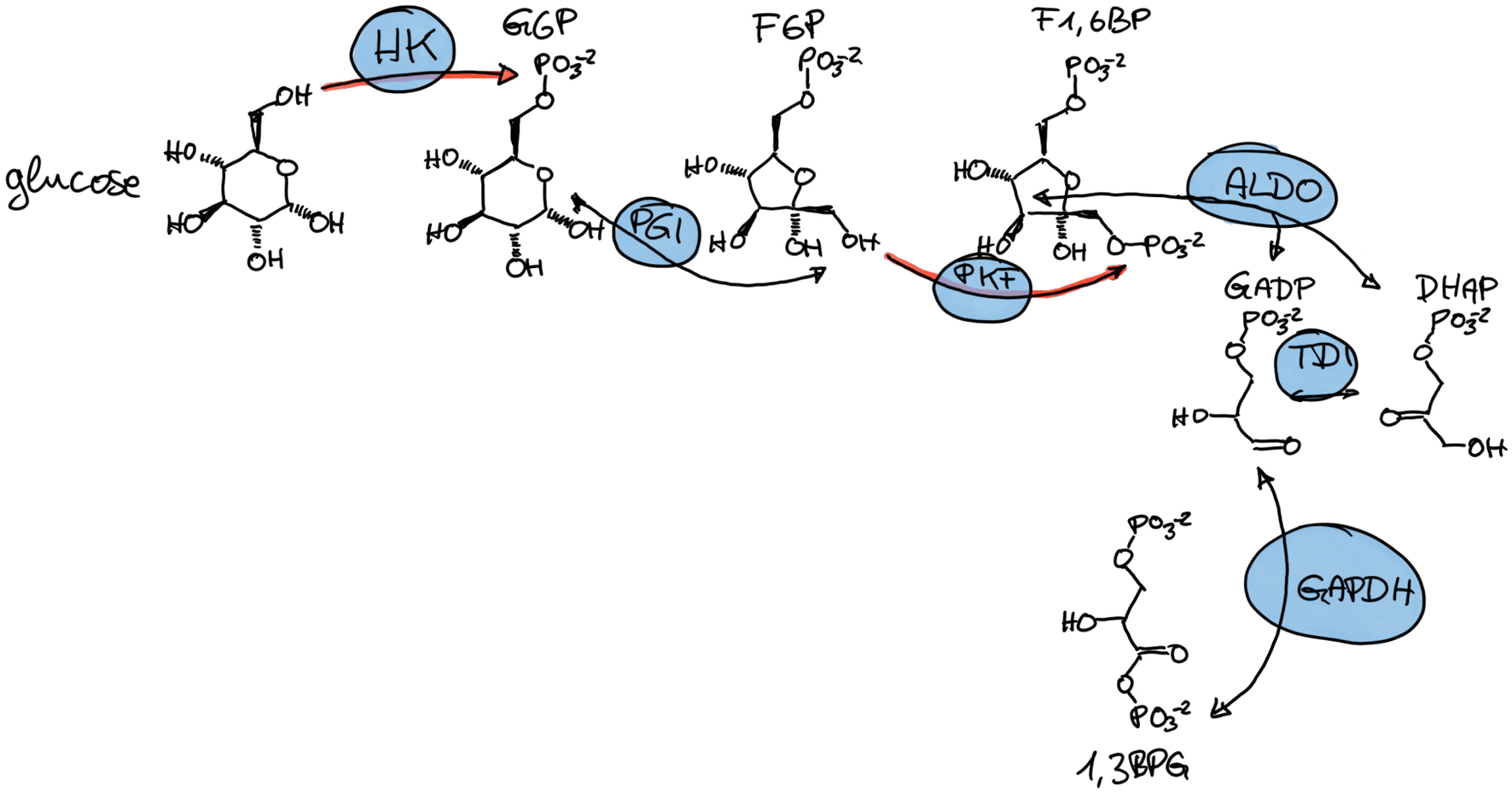
Metabolite? Metabolism?

- Glycolysis



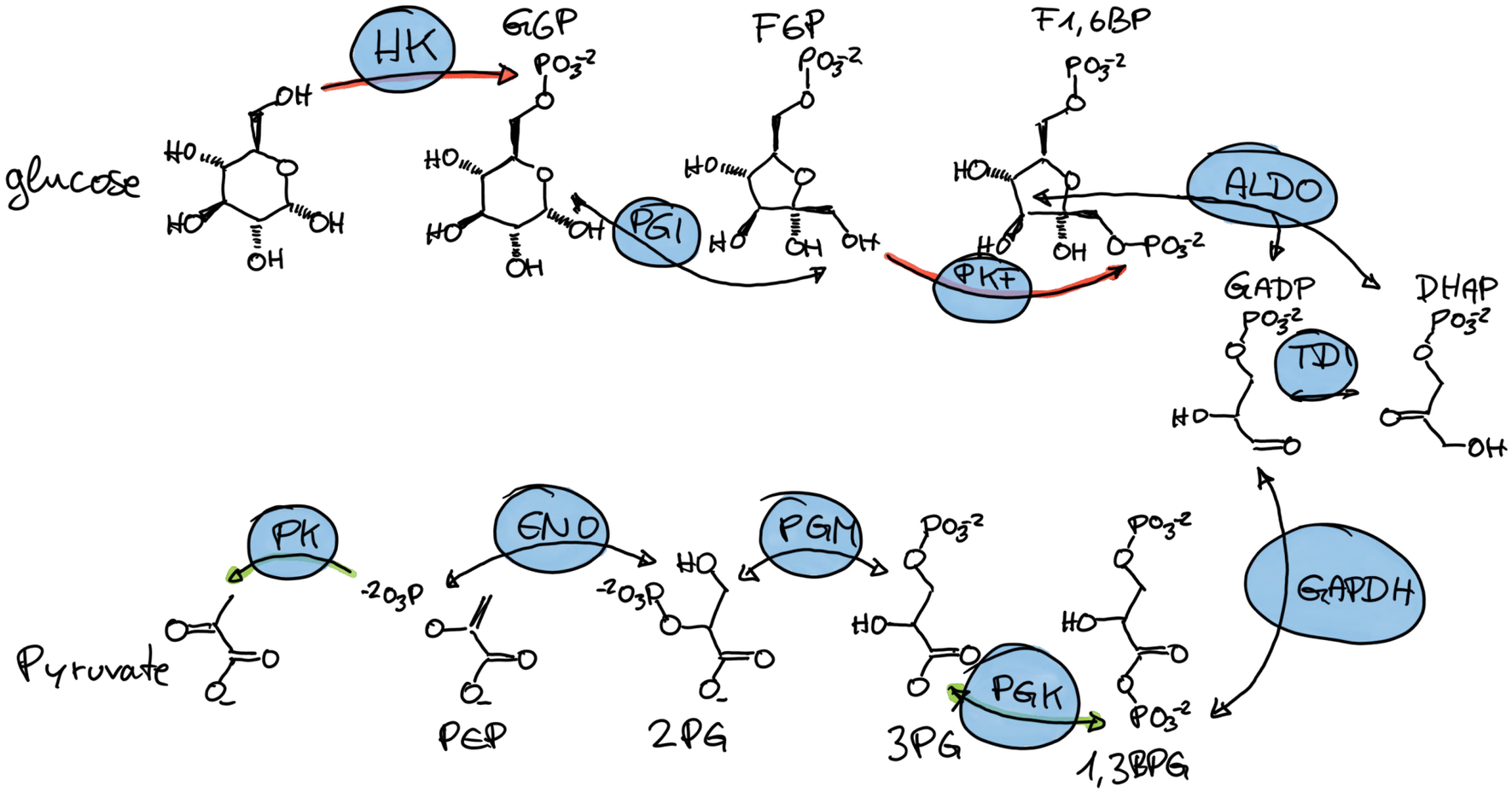
Metabolite? Metabolism?

- Glycolysis



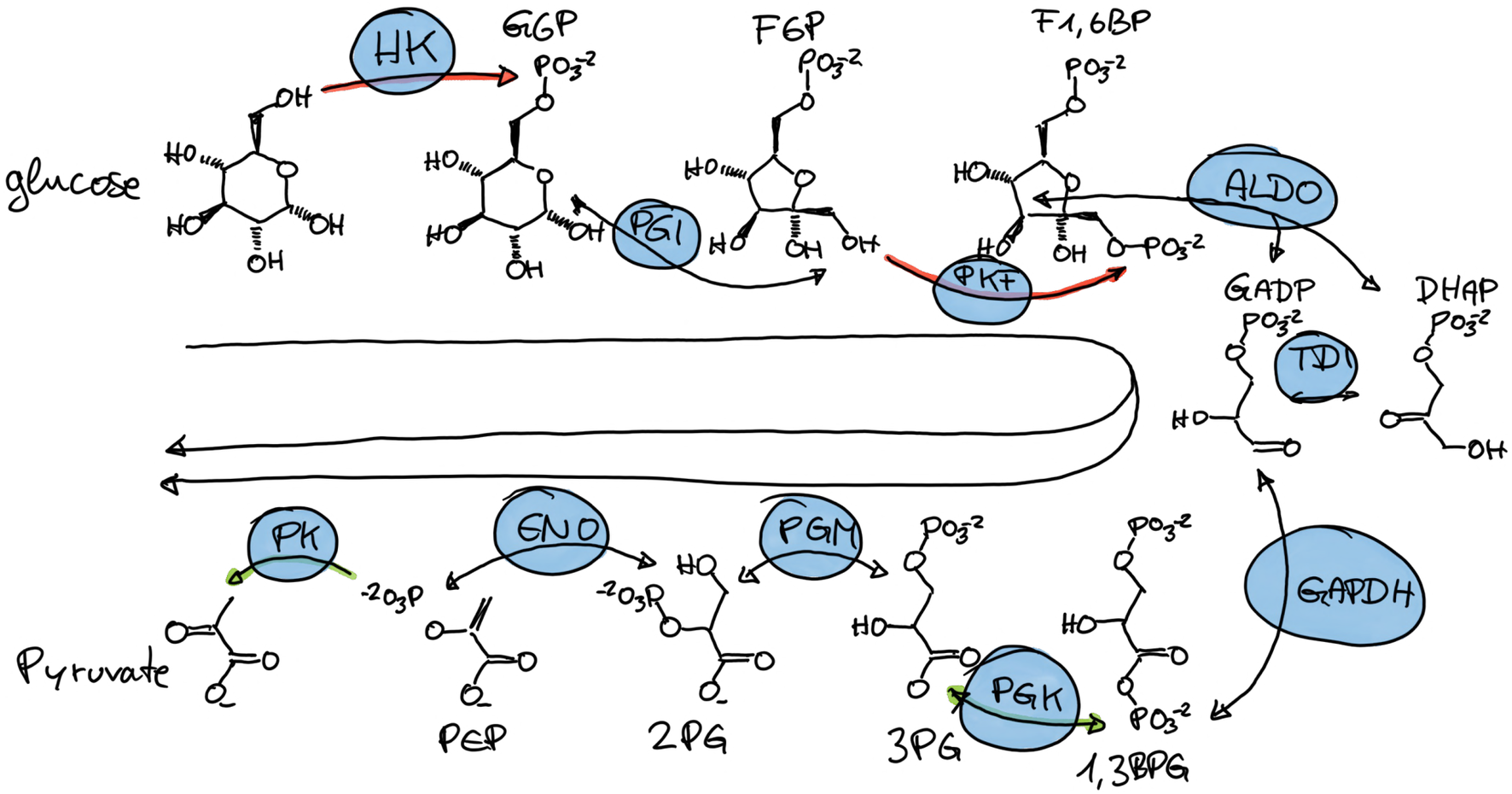
Metabolite? Metabolism?

- Glycolysis



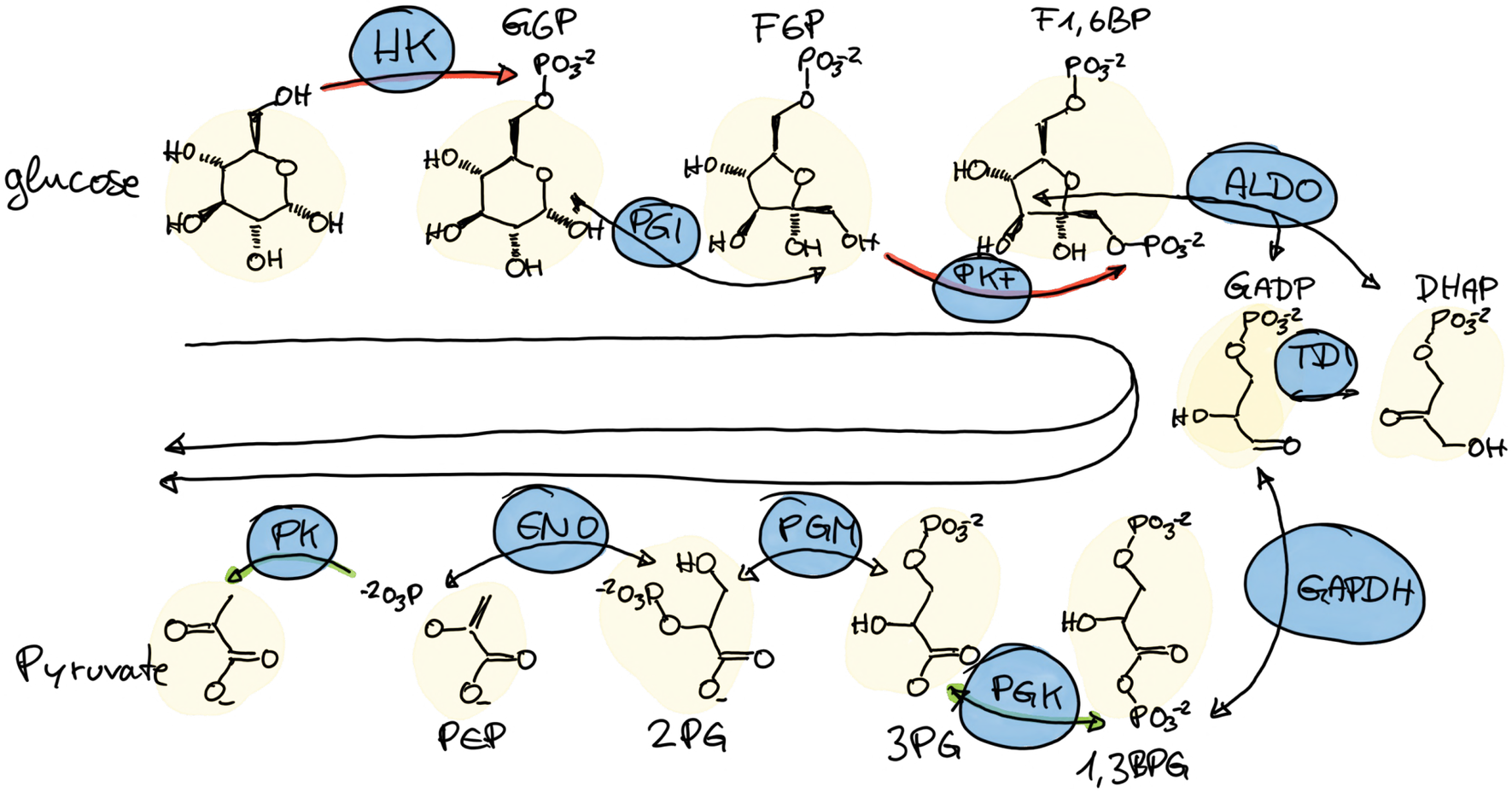
Metabolite? Metabolism?

- Glycolysis



Metabolite? Metabolism?

- Glycolysis



- Metabolites: intermediates and products of cellular processes.

Metabolomics?

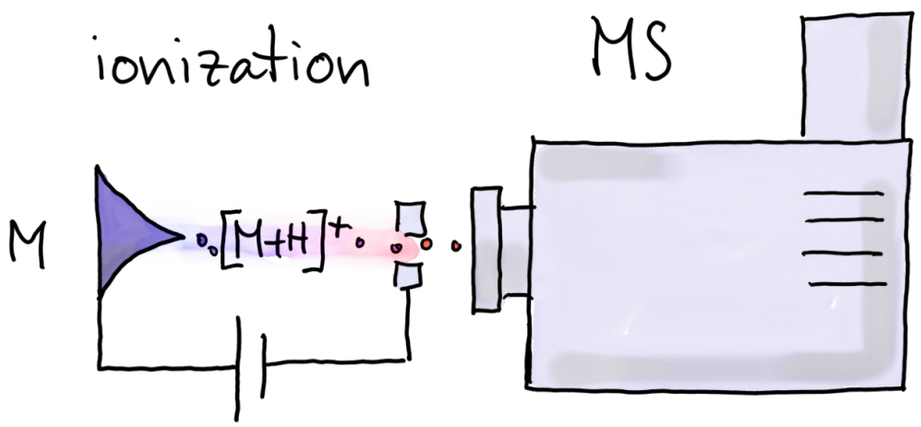
- Large-scale study of small molecules (metabolites) in a system (cell, tissue, organism).
- Comparison of the different -omes:
- **Genome**: what can happen.
- **Transcriptome**: what appears to be happening.
- **Proteome**: what makes it happen.
- **Metabolome**: what actually happened.
- Metabolome influenced by genetic **and** environmental factors.

How can we measure metabolites?

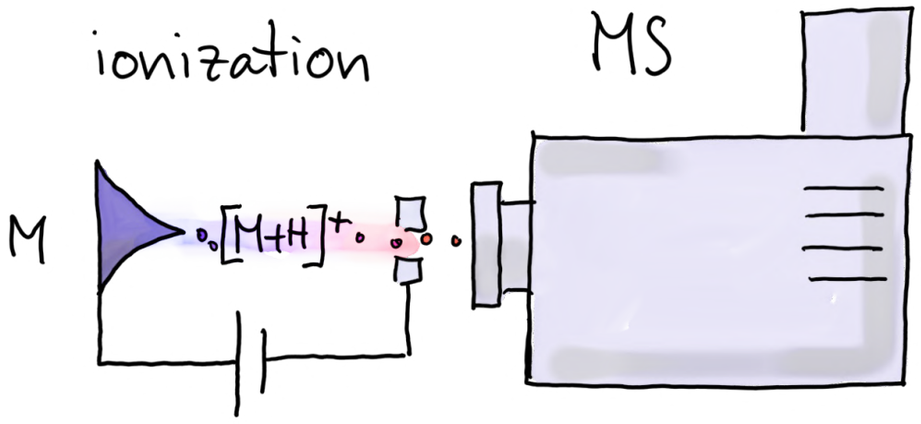
- Nuclear Magnetic Resonance (NMR) - not covered here.
- Mass spectrometry (MS)-based metabolomics.

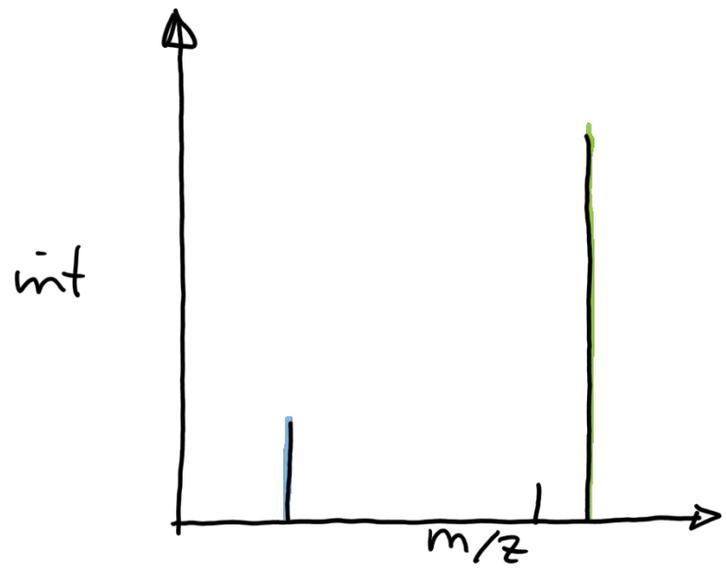
- Metabolites small enough to be directly measured by MS.
- Most metabolites uncharged - need to create ions first.

Mass Spectrometry (MS)



Mass Spectrometry (MS)

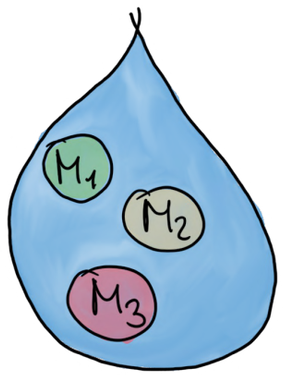




- **Problem:** unable to distinguish between metabolites with the same/similar mass-to-charge ratio (m/z).
- **Solution:** additional separation of metabolites prior to MS.

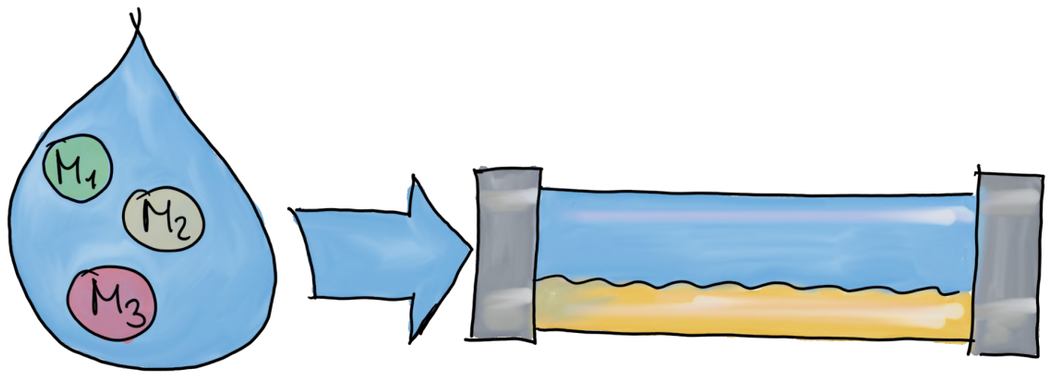
Liquid chromatography

- Sample is dissolved in a fluid (mobile phase).



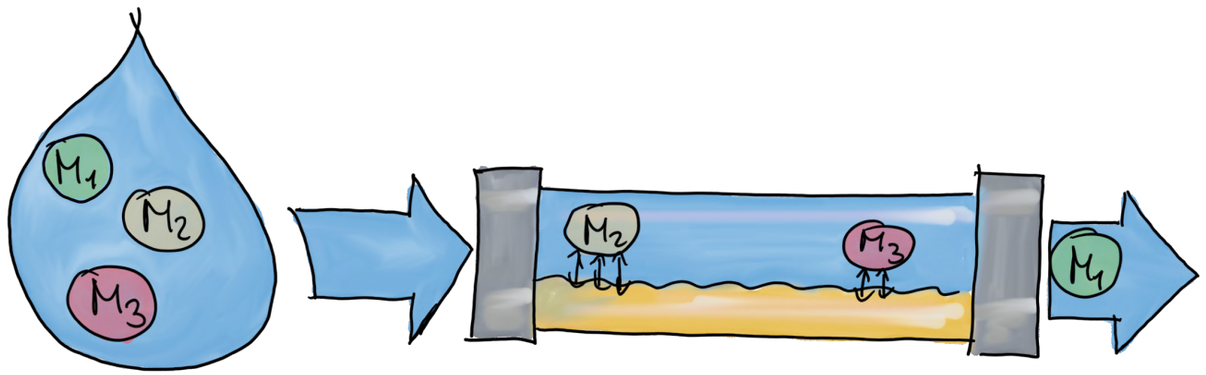
Liquid chromatography

- Sample is dissolved in a fluid (mobile phase).
- Mobile phase carries analytes through column (stationary phase).



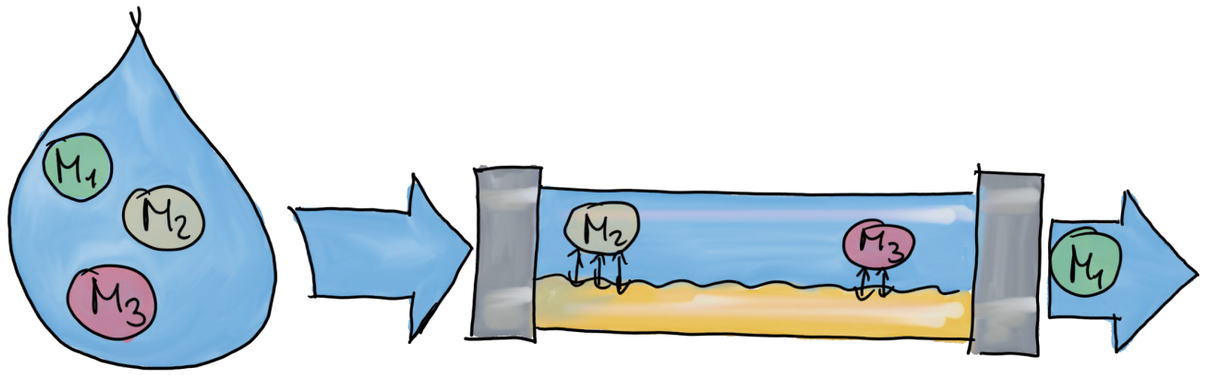
Liquid chromatography

- Sample is dissolved in a fluid (mobile phase).
- Mobile phase carries analytes through column (stationary phase).
- Separation based on affinity for the column’s stationary phase.

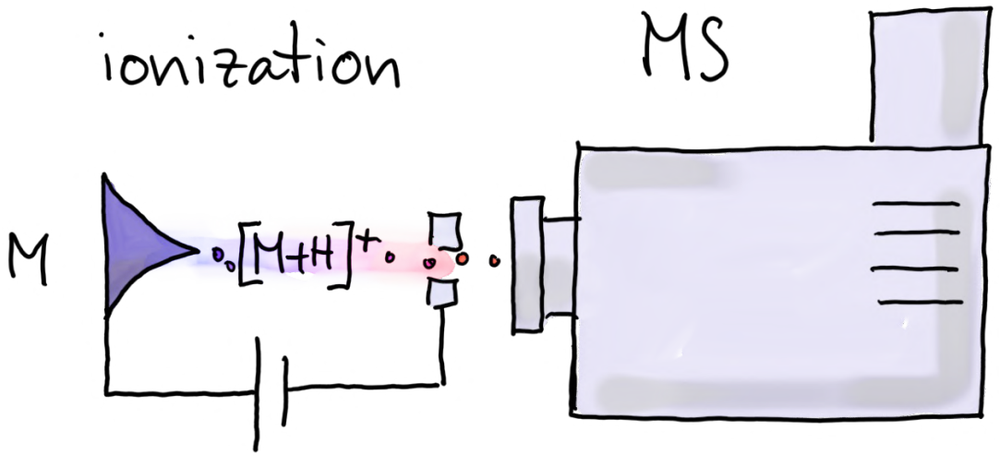


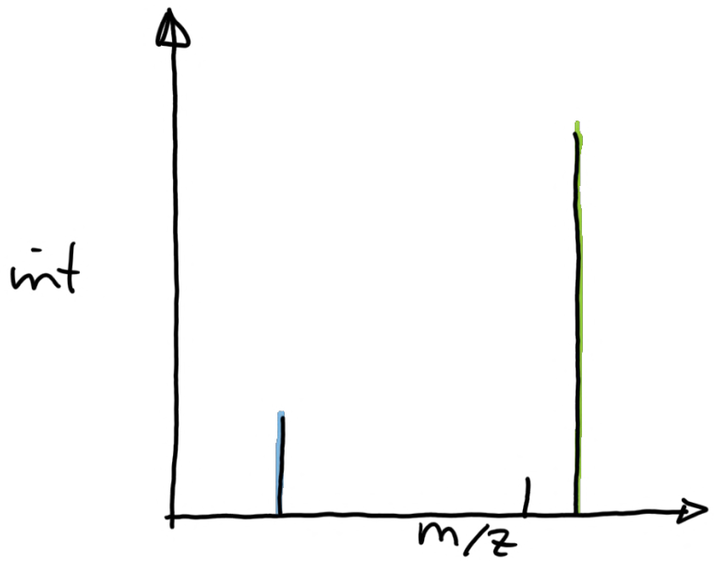
Liquid chromatography

- Sample is dissolved in a fluid (mobile phase).
- Mobile phase carries analytes through column (stationary phase).
- Separation based on affinity for the column's stationary phase.
- HILIC (hyrophilic liquid interaction chromatography):
  - Hydrophilic, polar stationary phase.
  - Analytes solved in mobile phase.
  - Analytes separated by polarity: compounds with low polarity elute first, with high polarity later.

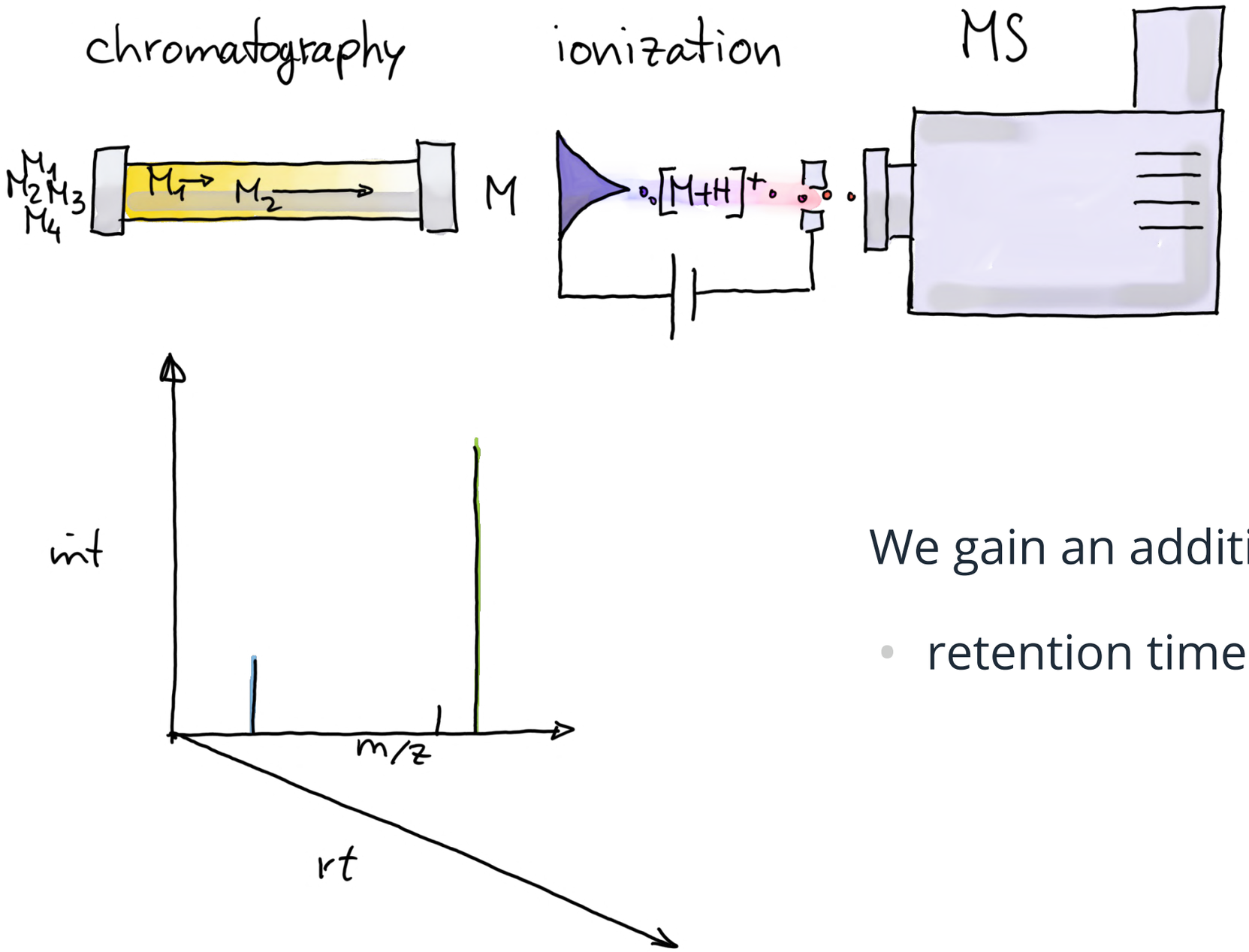


Liquid Chromatography Mass Spectrometry (LC-MS)





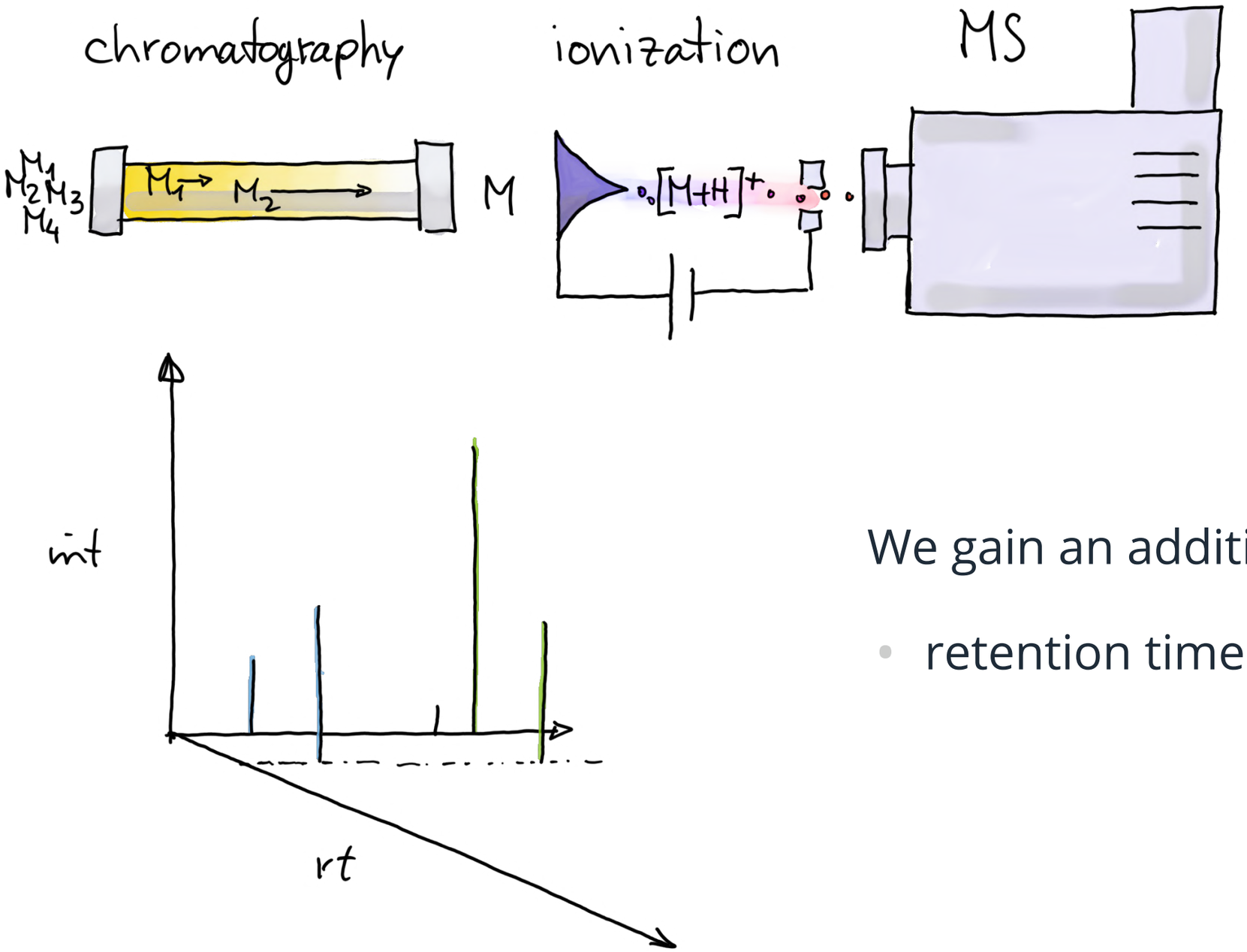
Liquid Chromatography Mass Spectrometry (LC-MS)



We gain an additional dimension:

- retention time.

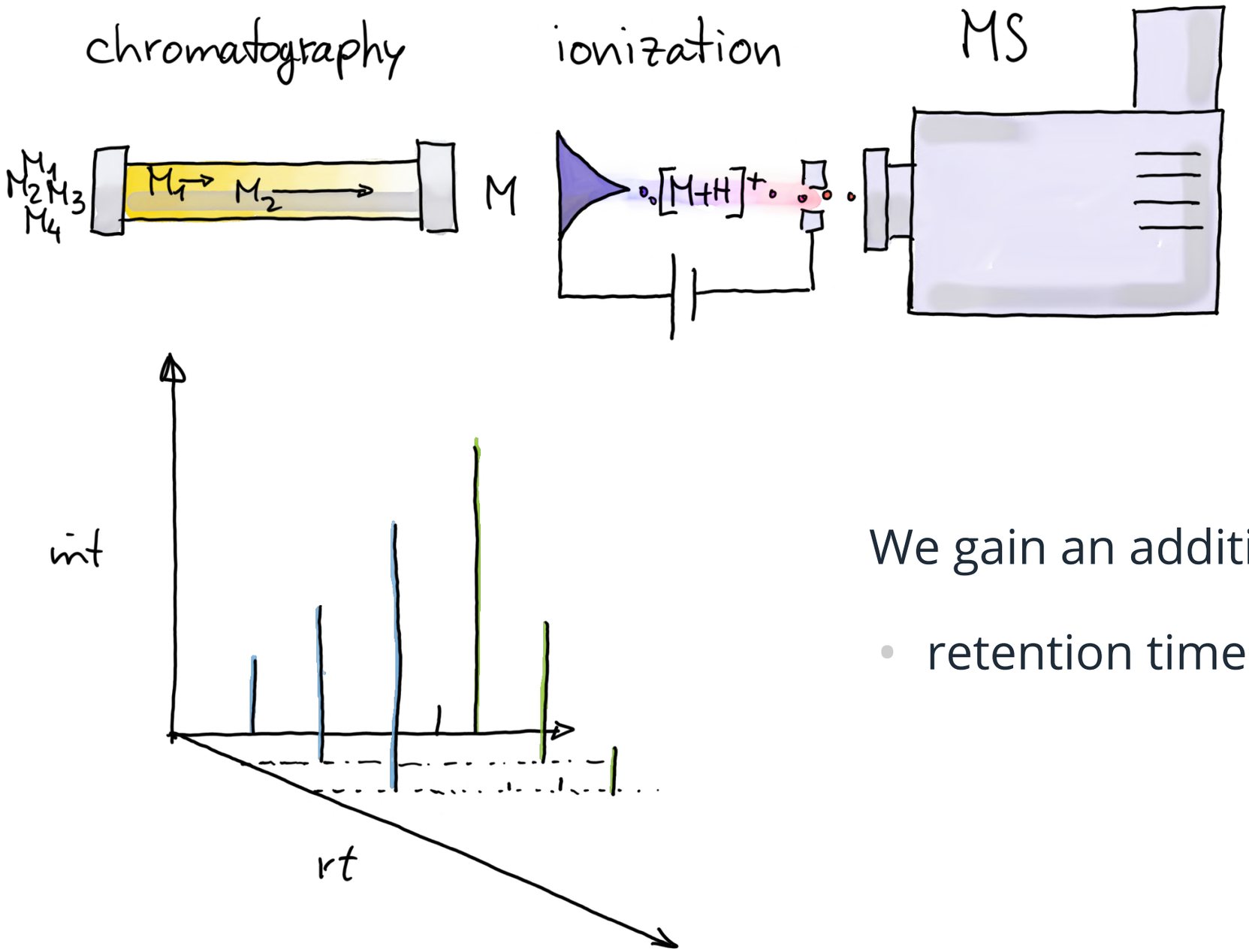
Liquid Chromatography Mass Spectrometry (LC-MS)



We gain an additional dimension:

- retention time.

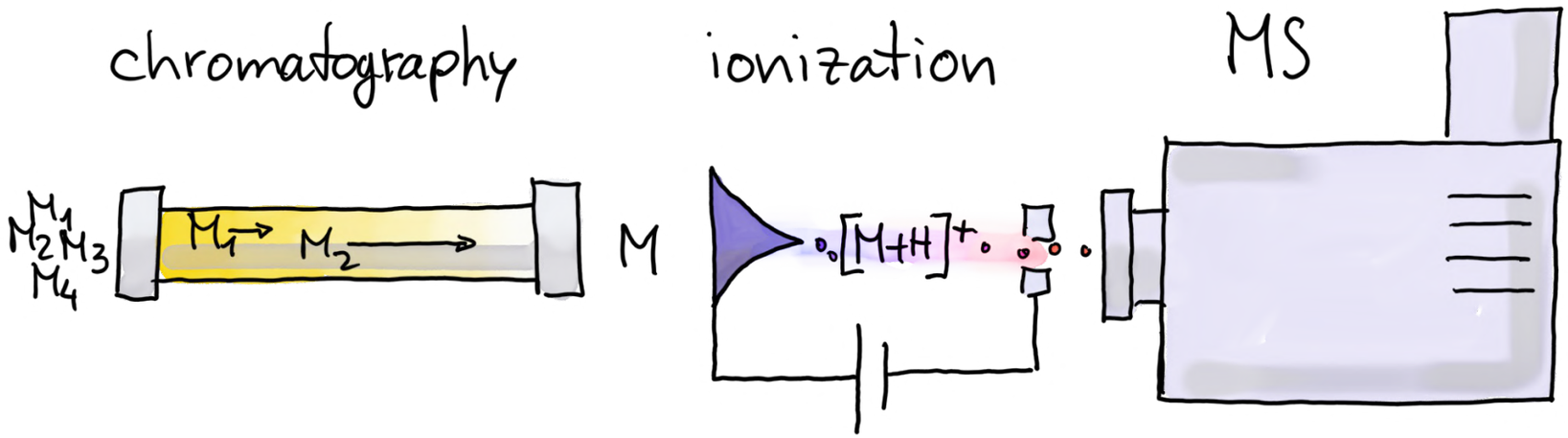
Liquid Chromatography Mass Spectrometry (LC-MS)

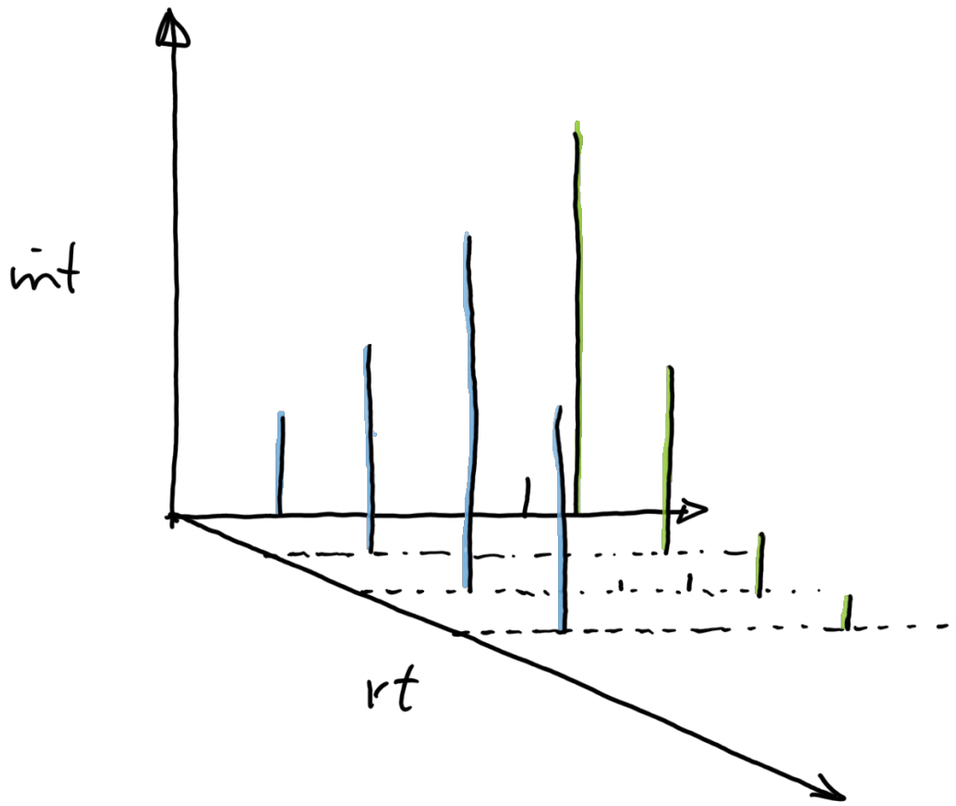


We gain an additional dimension:

- retention time.

Liquid Chromatography Mass Spectrometry (LC-MS)

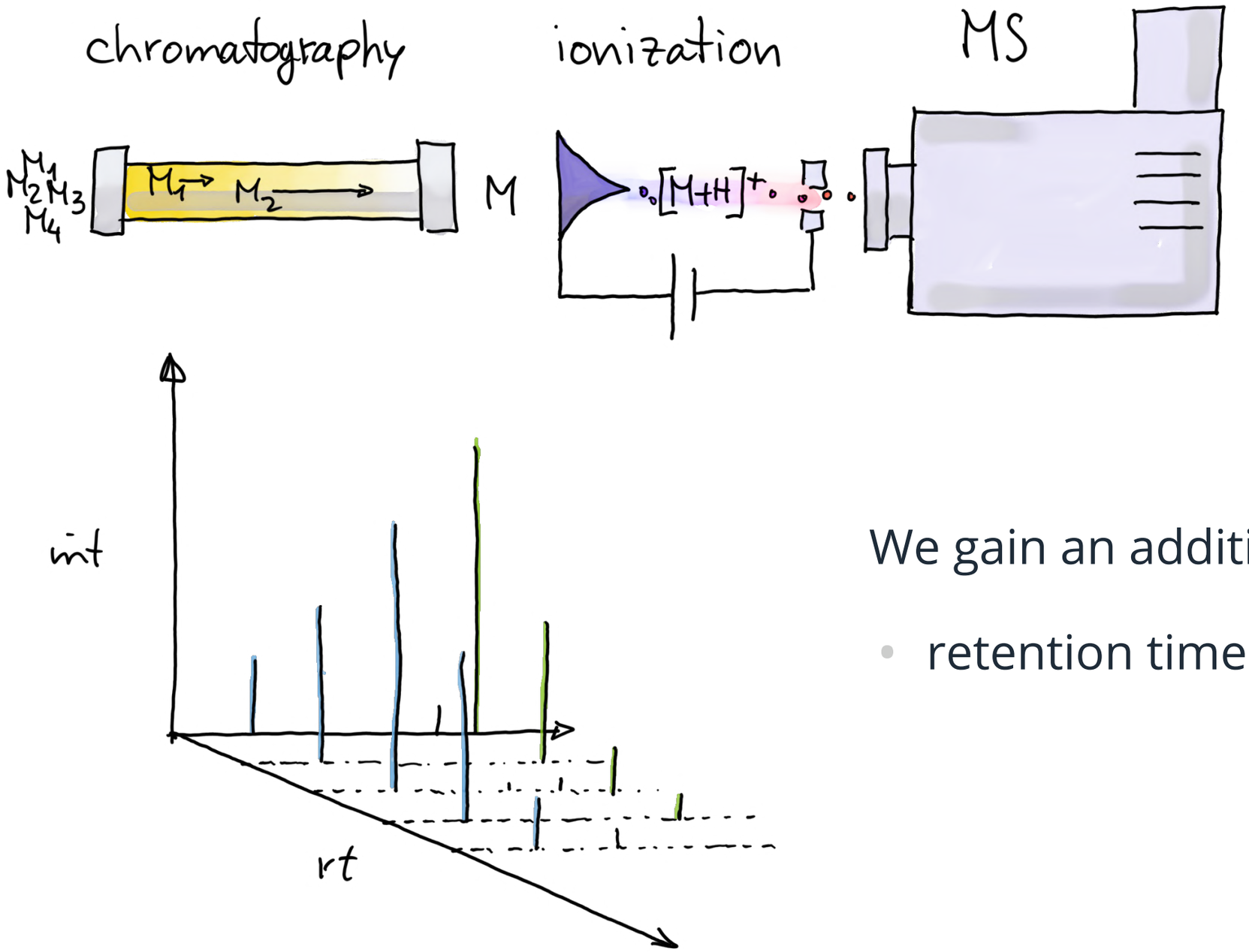




We gain an additional dimension:

- retention time.

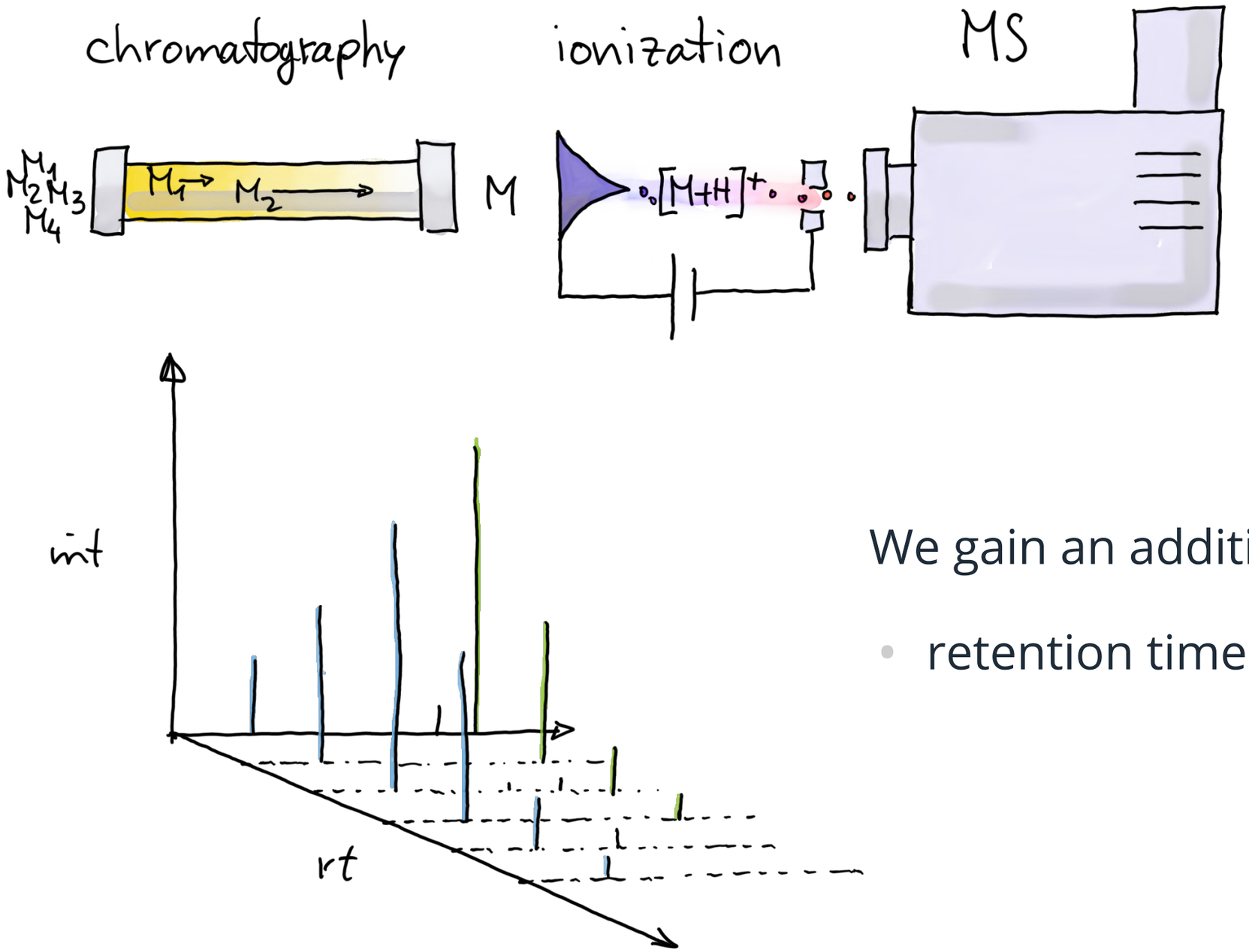
Liquid Chromatography Mass Spectrometry (LC-MS)



We gain an additional dimension:

- retention time.

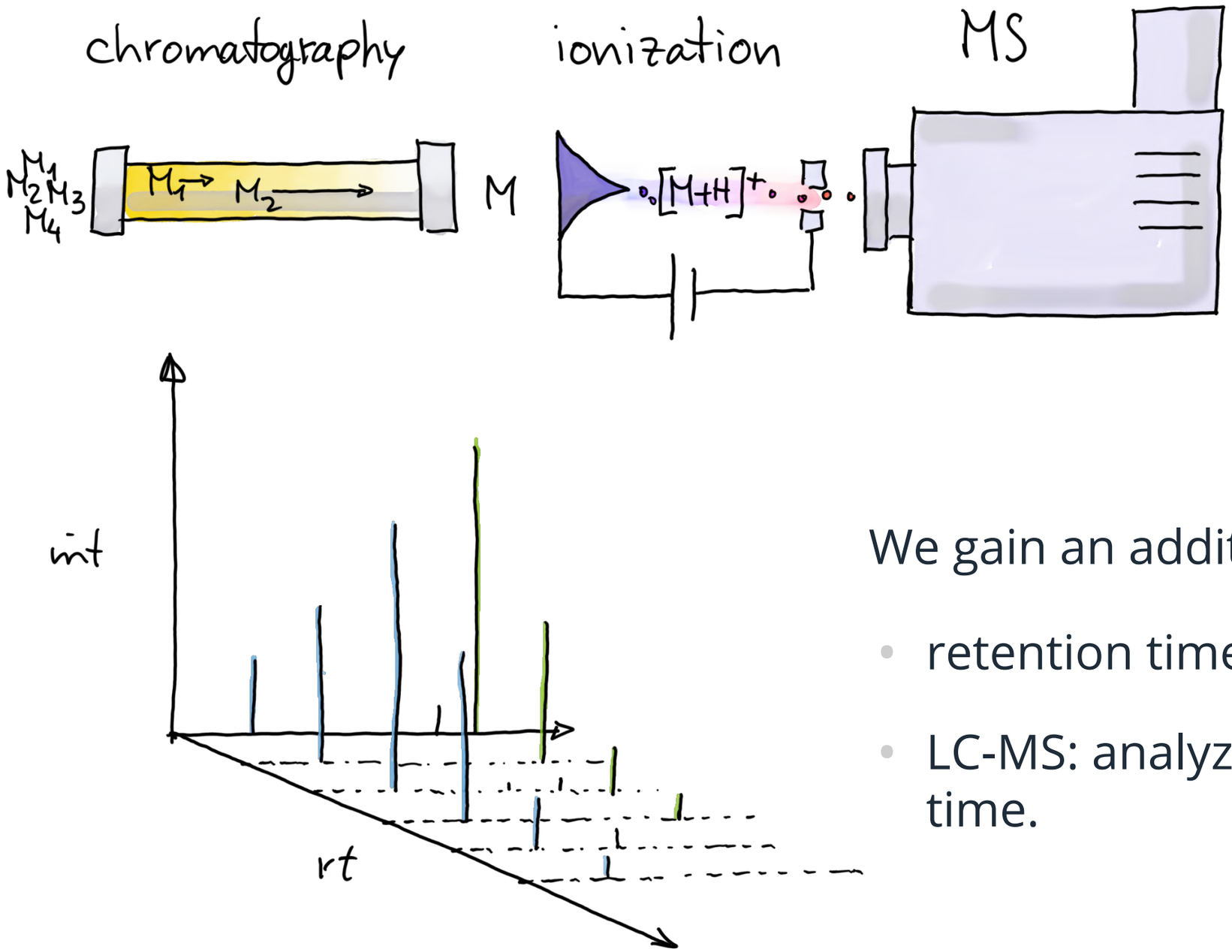
Liquid Chromatography Mass Spectrometry (LC-MS)



We gain an additional dimension:

- retention time.

Liquid Chromatography Mass Spectrometry (LC-MS)



We gain an additional dimension:

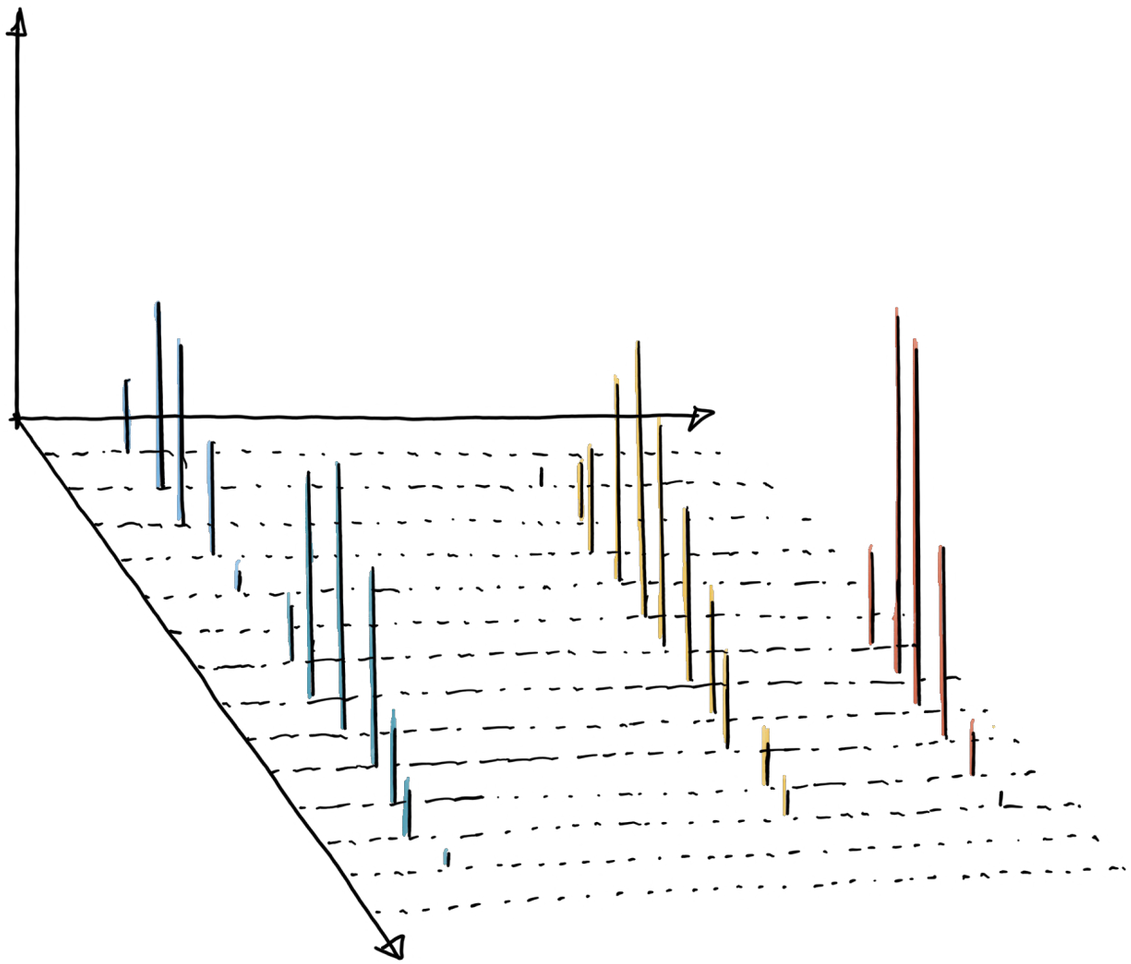
- retention time.
- LC-MS: analyze data along retention time.

LC-MS data preprocessing

- Chromatographic peak detection
- Alignment
- Correspondence

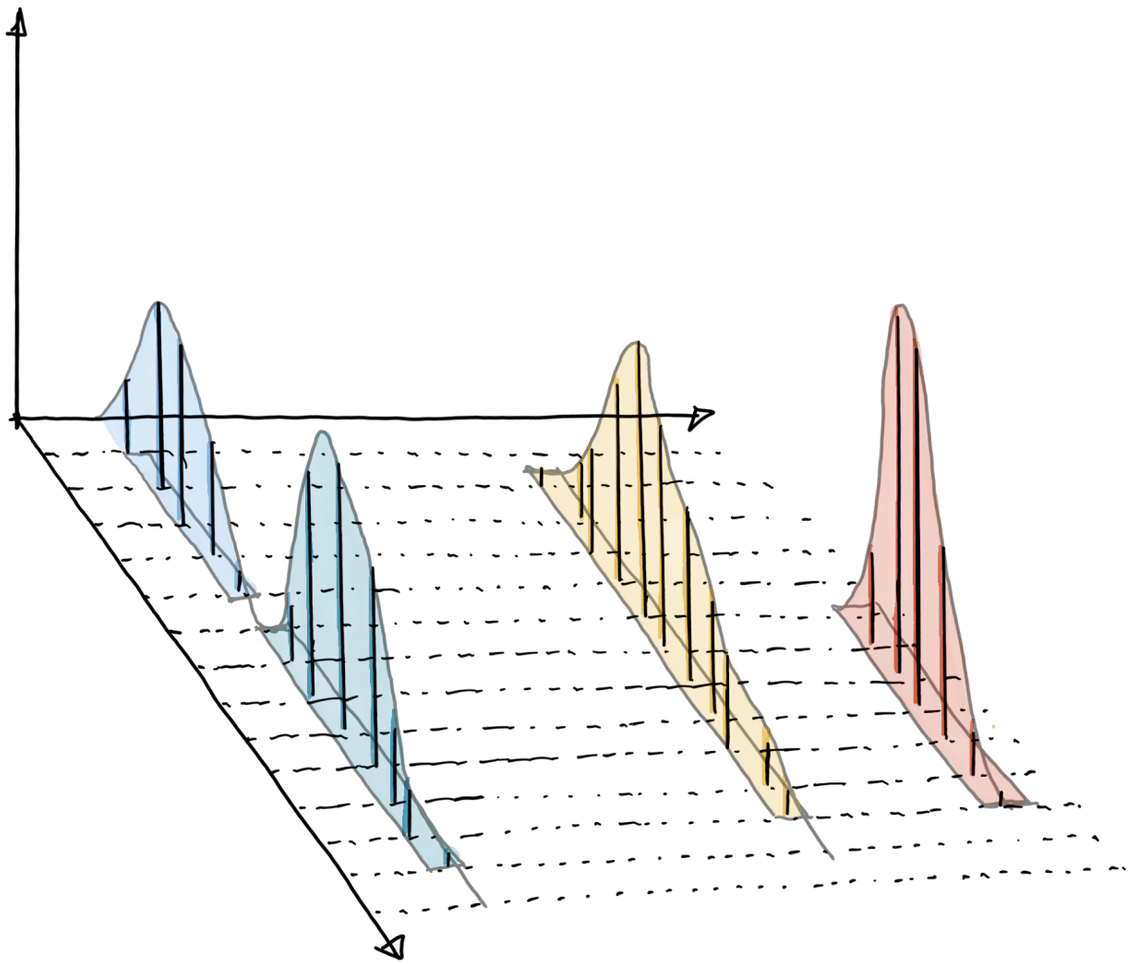
Chromatographic peak detection

- **Aim:** identify chromatographic peaks in the data.



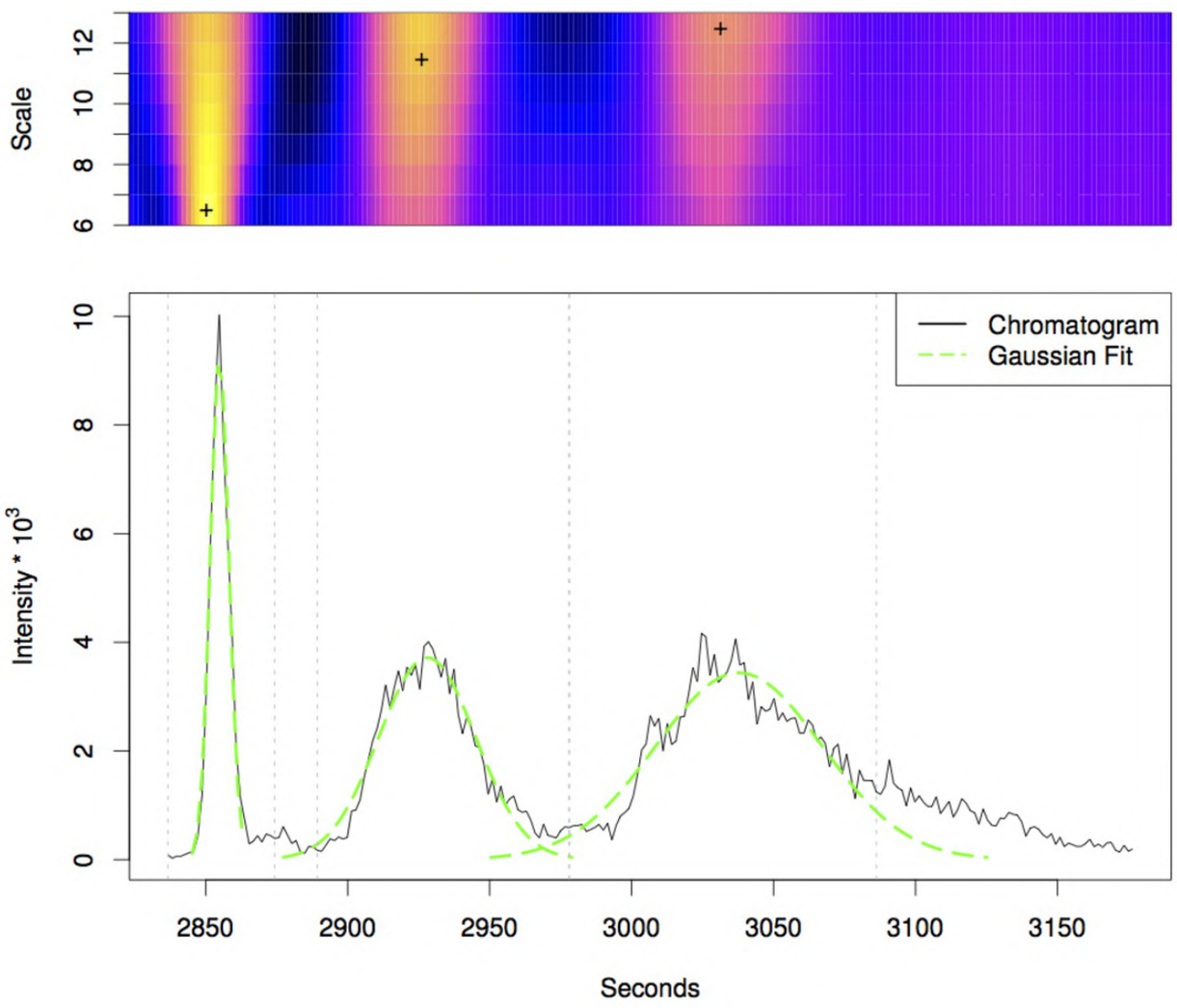
Chromatographic peak detection

- **Aim:** identify chromatographic peaks in the data.



Chromatographic peak detection

- **centWave** [Tautenhahn et al. BMC Bioinformatics, 2008]:
- Allows detection of peaks with different rt widths.



Chromatographic peak detection

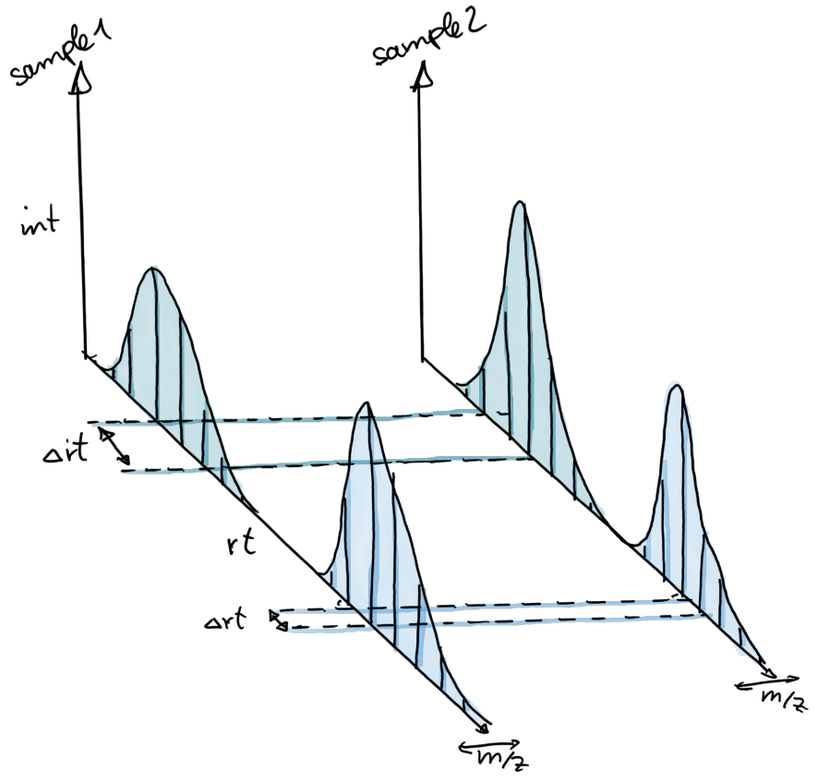
- `MSnbase`: data import with `readMSData`.
- `xcms`: peak detection with `findChromPeaks` and algorithm-specific parameter object.

```
cwp <- CentWaveParam(peakwidth = c(2, 10), snthresh = 5)
data <- findChromPeaks(data, param = cwp)
head(chromPeaks(data), n = 3)
```

```
##              mz    mzmin    mzmax     rt rtmin  rtmax     into     intb
## CP001 114.0907 114.0899 114.0929  1.954 0.280  3.907 1559.829 1555.923
## CP002 114.0913 114.0884 114.0929  5.860 4.465  8.650 1890.221 1885.757
## CP003 114.0914 114.0899 114.0929 10.882 8.650 13.114 1950.953 1946.210
##           maxo  sn sample
## CP001 584.9510 584      1
## CP002 601.8881 601      1
## CP003 691.9580 691      1
```

Alignment

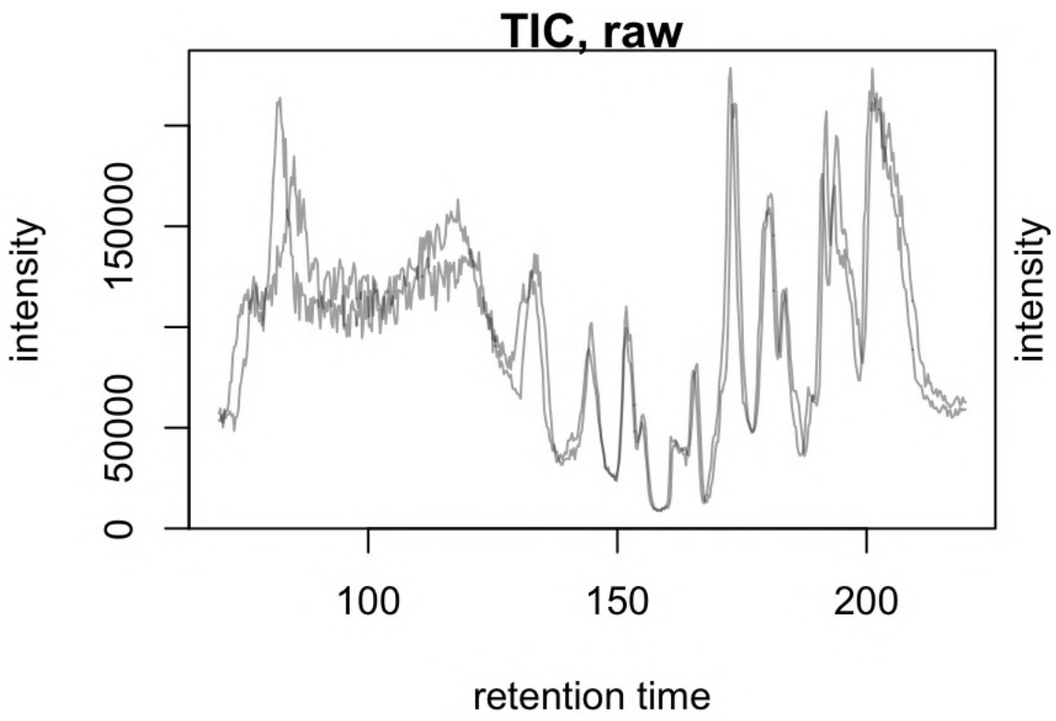
- **Aim:** adjust differences in retention times between samples.
- Same analyte elutes at slightly different time between measurements.

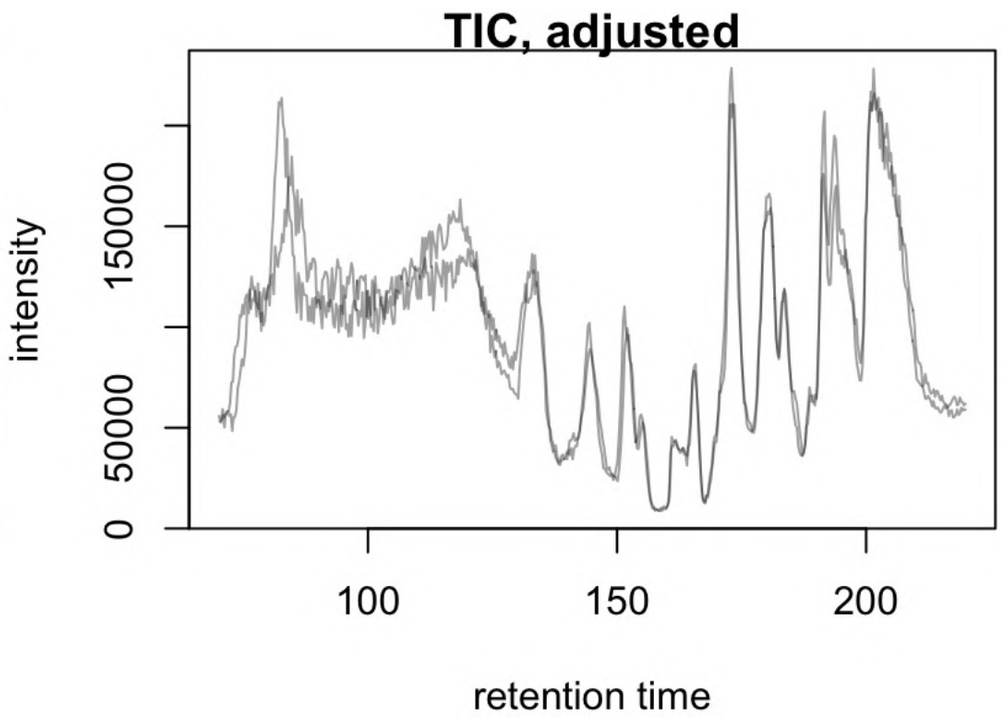


- **Why?** Age of column, temperature …

Alignment

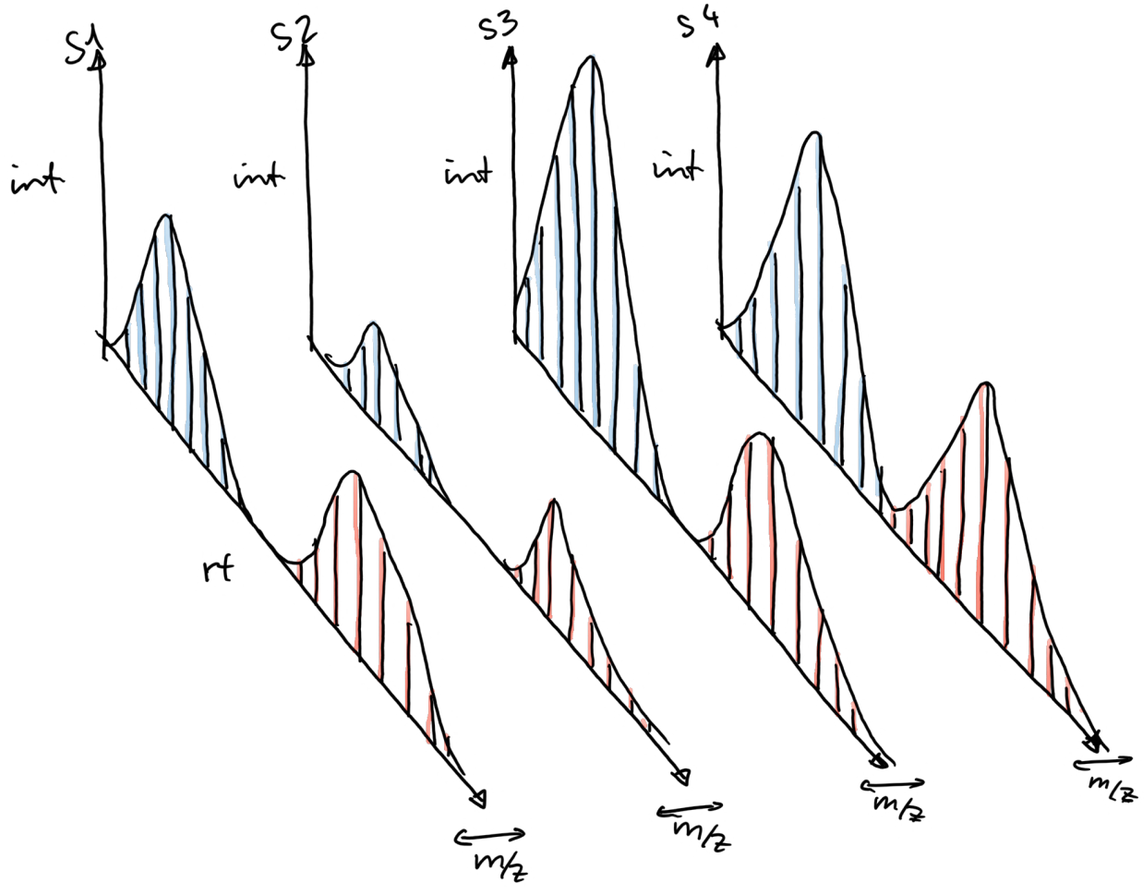
- Many algorithms available [Smith et al. Brief Bioinformatics 2013]
- `xcms`: `adjustRtime` function with `PeakGroupsParam` [Smith et al. Anal. chem. 2006] or `ObiwarpParam` [Prince et al. Anal. chem. 2006].





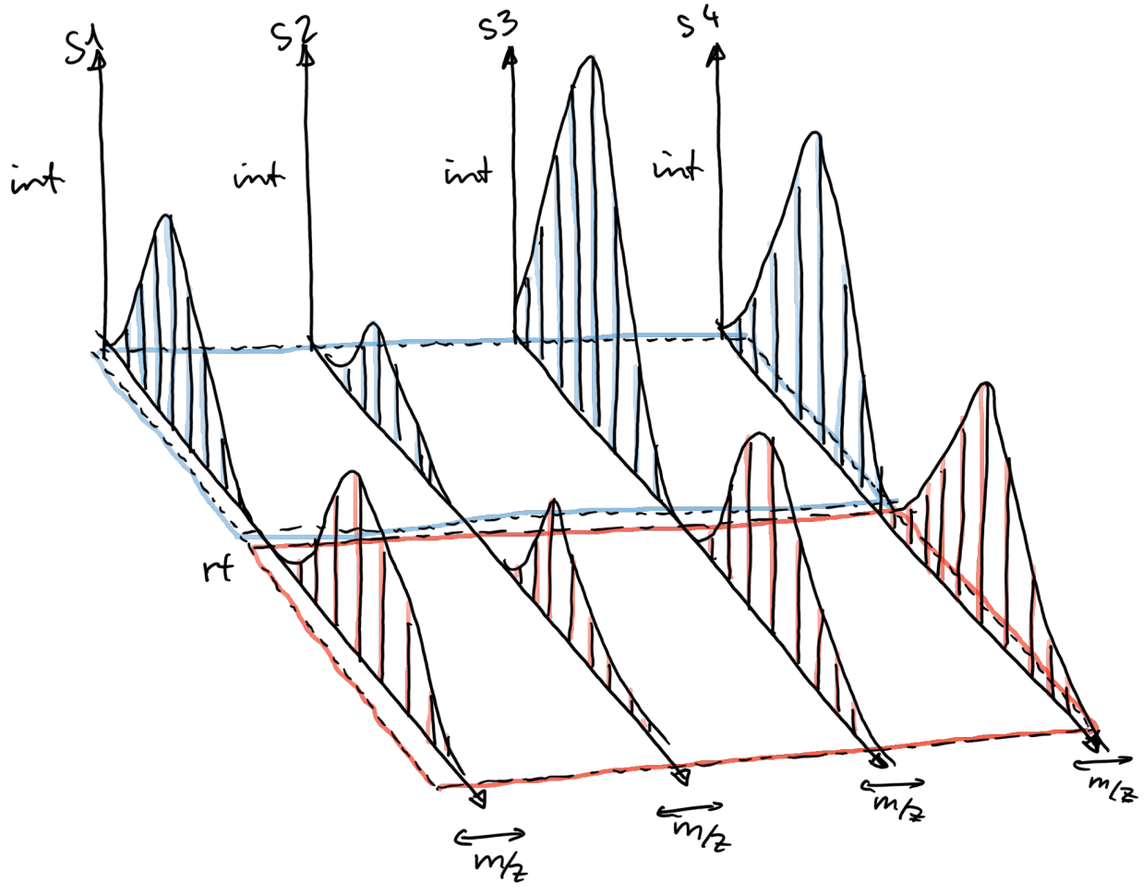
Correspondence

- **Aim:** group peaks representing same ion species across samples.
- **Result:** matrix of abundances, rows *features*, columns samples.



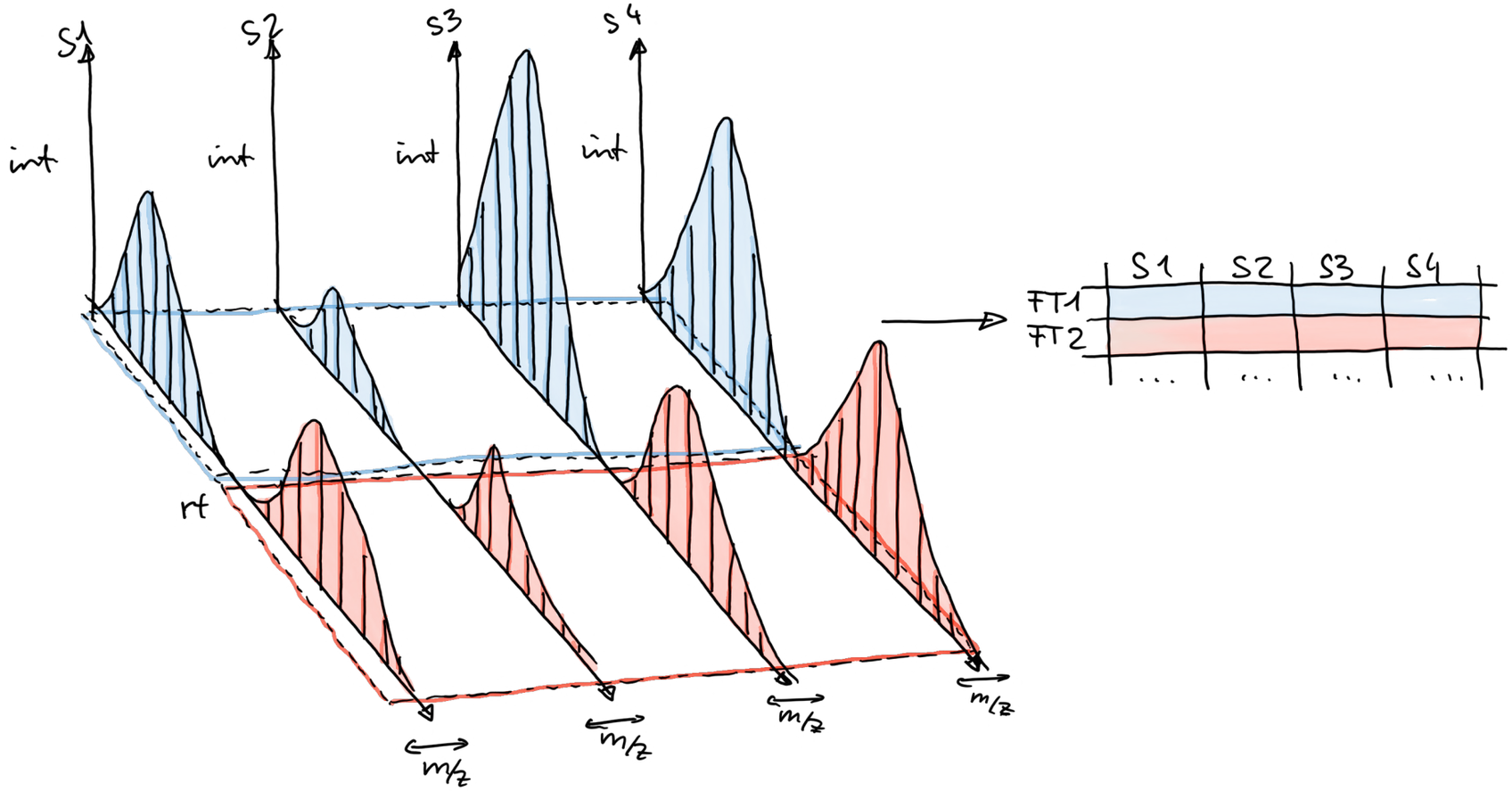
Correspondence

- **Aim:** group peaks representing same ion species across samples.
- **Result:** matrix of abundances, rows *features*, columns samples.



Correspondence

- **Aim:** group peaks representing same ion species across samples.
- **Result:** matrix of abundances, rows *features*, columns samples.

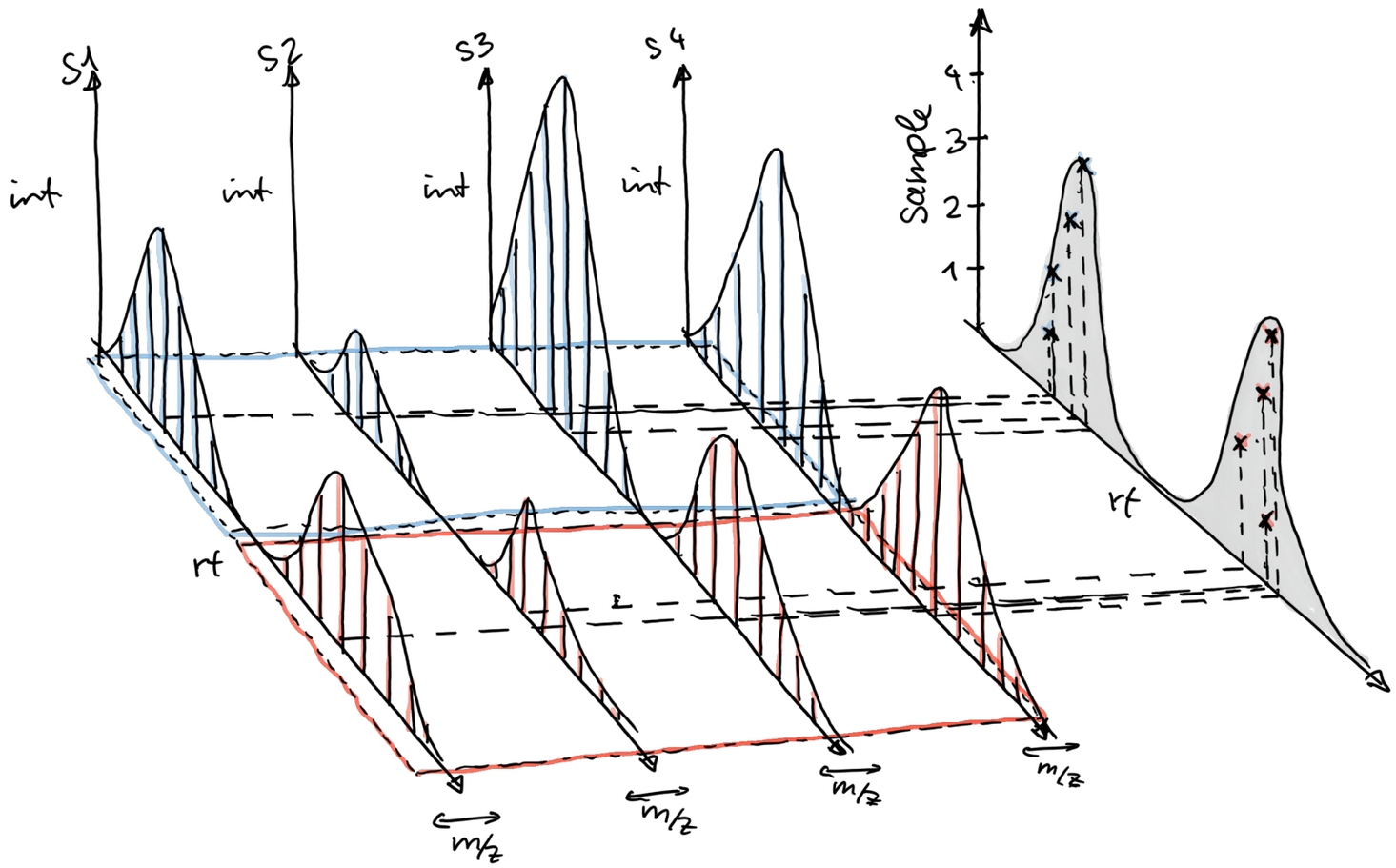


Correspondence

- `xcms`: `groupChromPeaks` with `NearestPeaksParam` [Katajamaa et al. Bioinformatics 2006] and `PeakDensityParam` [Smith et al. Anal. chem. 2006].

Correspondence

- `xcms`: `groupChromPeaks` with `NearestPeaksParam` [Katajamaa et al. Bioinformatics 2006] and `PeakDensityParam` [Smith et al. Anal. chem. 2006].
- *Peak density* approach (for a given m/z slice):
- Identify regions along rt with high peak density, group peaks.



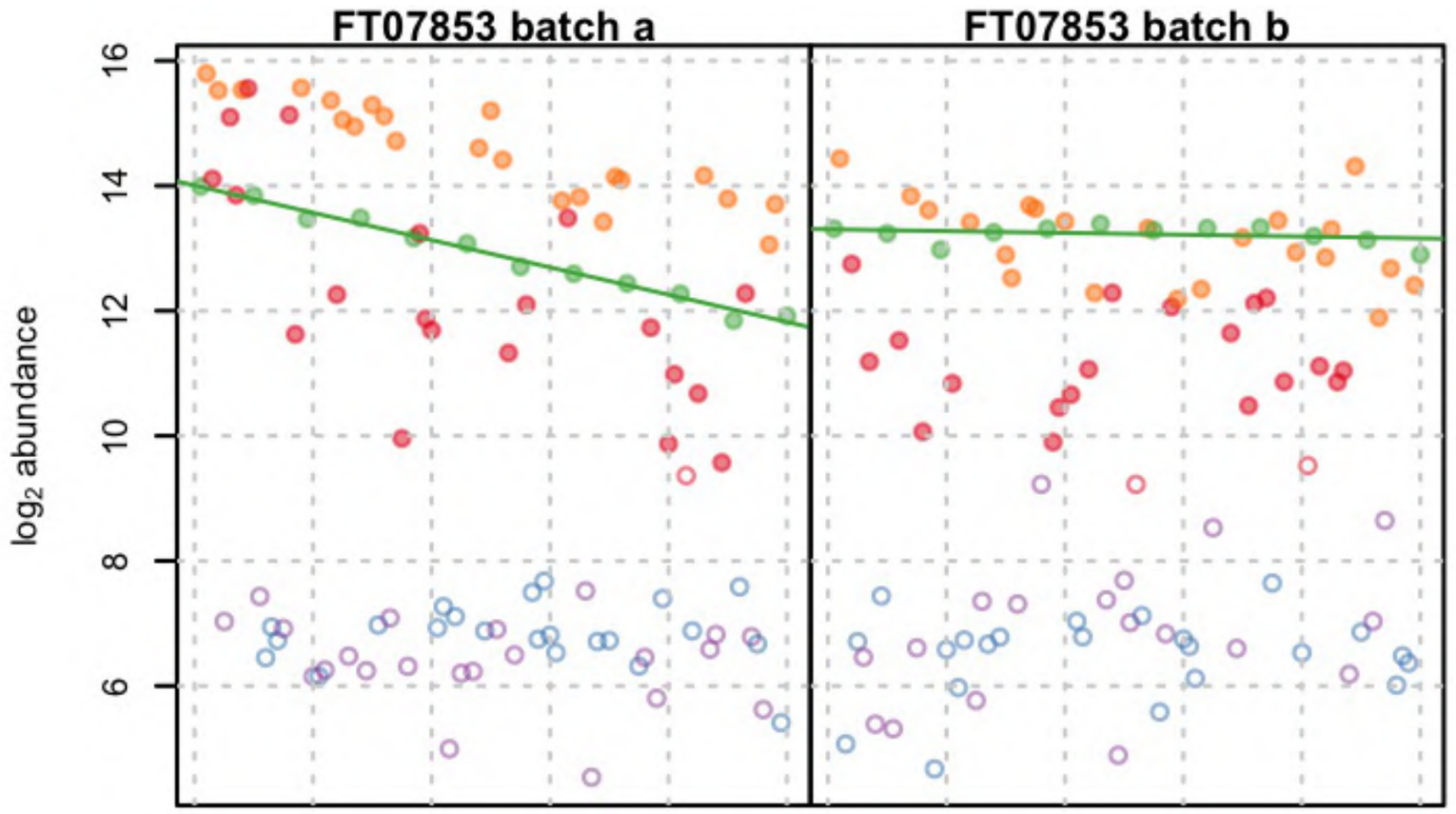
Preprocessing result

- Numeric matrix with abundances.
- Normalization.
- Identification of features of interest.
- Annotation.

Normalization

Account for:

- Sample-specific effects.
- Effects related to batch/measurement run.
- Injection order-dependent effects: specific to metabolite.



Normalization

- Good practice for experimental design:
  - QC samples measured repeatedly.
  - Internal standards.
  - Replicates.
  - Measurement of study samples in randomized order.

- Popular normalization methods:
  - RUV [De Livera et al. Anal. Chem. 2015]
  - linear models [Wehrens et al. Metabolomics 2016]
  - linear and higher order models [Brunius et al. Metabolomics 2016].

Annotation/Identification

- Feature != metabolite.

```
## DataFrame with 4 rows and 4 columns
##                    mzmed             rtmed            POOL_1            POOL_2
##                <numeric>         <numeric>         <numeric>         <numeric>
## FT001 105.041814839707 167.961095453642 229.490739260736 3093.75184315684
## FT002 105.041653033614 157.083057856508 4762.39872227772 6601.45091358641
## FT003 105.069636149683 31.8108067962868 699.723986763237 1033.23232267732
## FT004  105.11027064078 63.7513630255991 20211.2633706294 15839.5504368189
```

- Feature characterized by m/z and retention time.

Annotation based on mass matching

- m/z is **not** the mass.
- Mass of an [M+H]+ ion: m/z - mass of 1 hydrogen.
- Different ions from the same compound: [M+H]+, [M+Na]+, …
- Match mass against database.
  - The Human Metabolome Database (HMDB): https://hmdb.ca
  - Chemical Entities of Biological Interest: https://www.ebi.ac.uk/chebi
  - PubChem https://pubchem.ncbi.nlm.nih.gov
  - …
- Will result in many hits.

Improved Annotation

Annotate features based on m/z **and**:

- **retention time**: requires measurement of compound/standard on the same LC-MS setup.
- **MS2 spectrum**:
  - Requires LC-MS/MS data (DDA or DIA).
  - Reference spectrum has to be available in database.

Afternoon metabolomics lab

- LC-MS data handling (`MSnbase`).
- LC-MS data preprocessing using `xcms`.