# ASICS: an R package for the identification and quantification of metabolites in $^1$H NMR spectra

G. Lefort, L. Liaubet, C. Canlet, N. Vialaneix and R. Servien

7$^e$ Rencontres R - Rennes 2018

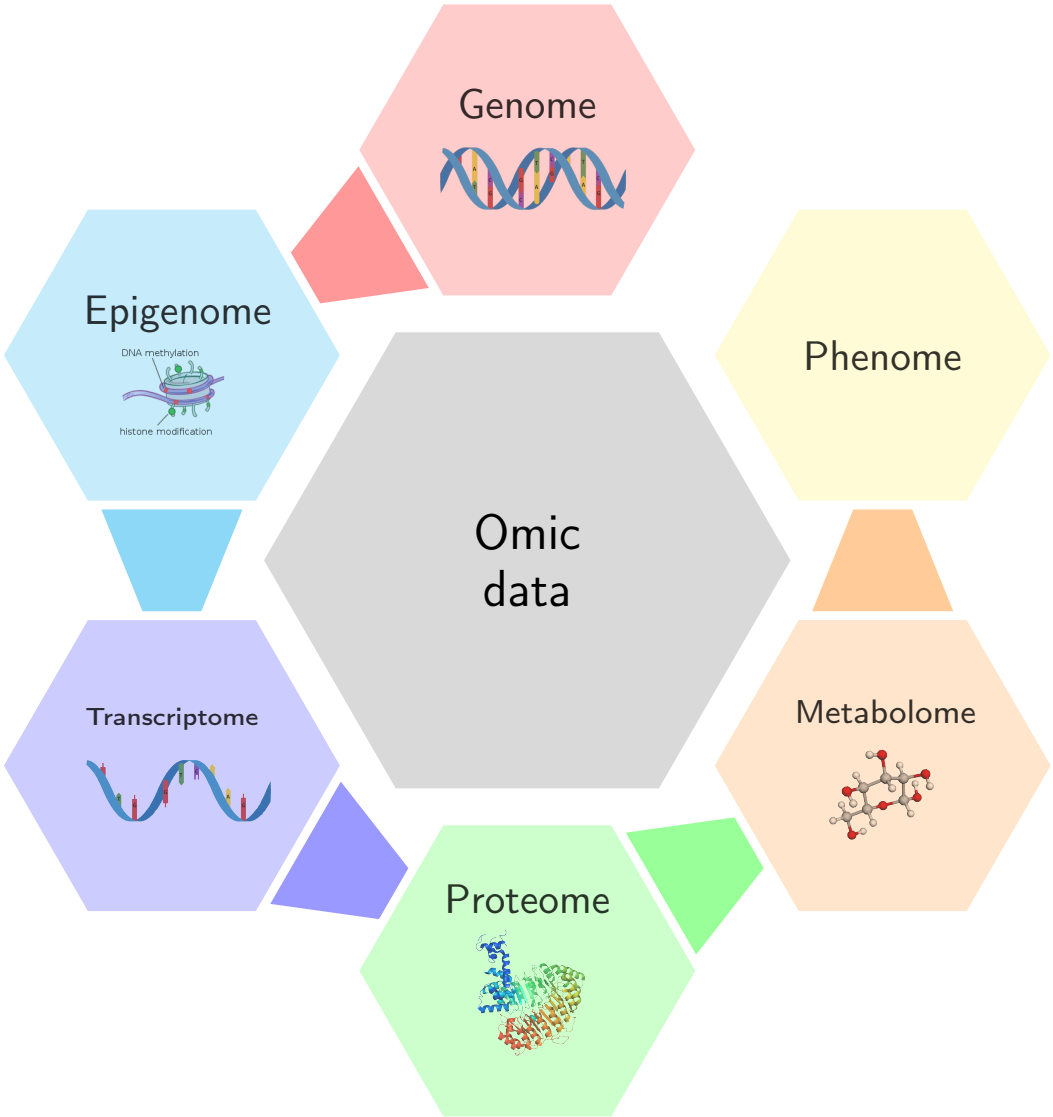
July 2018



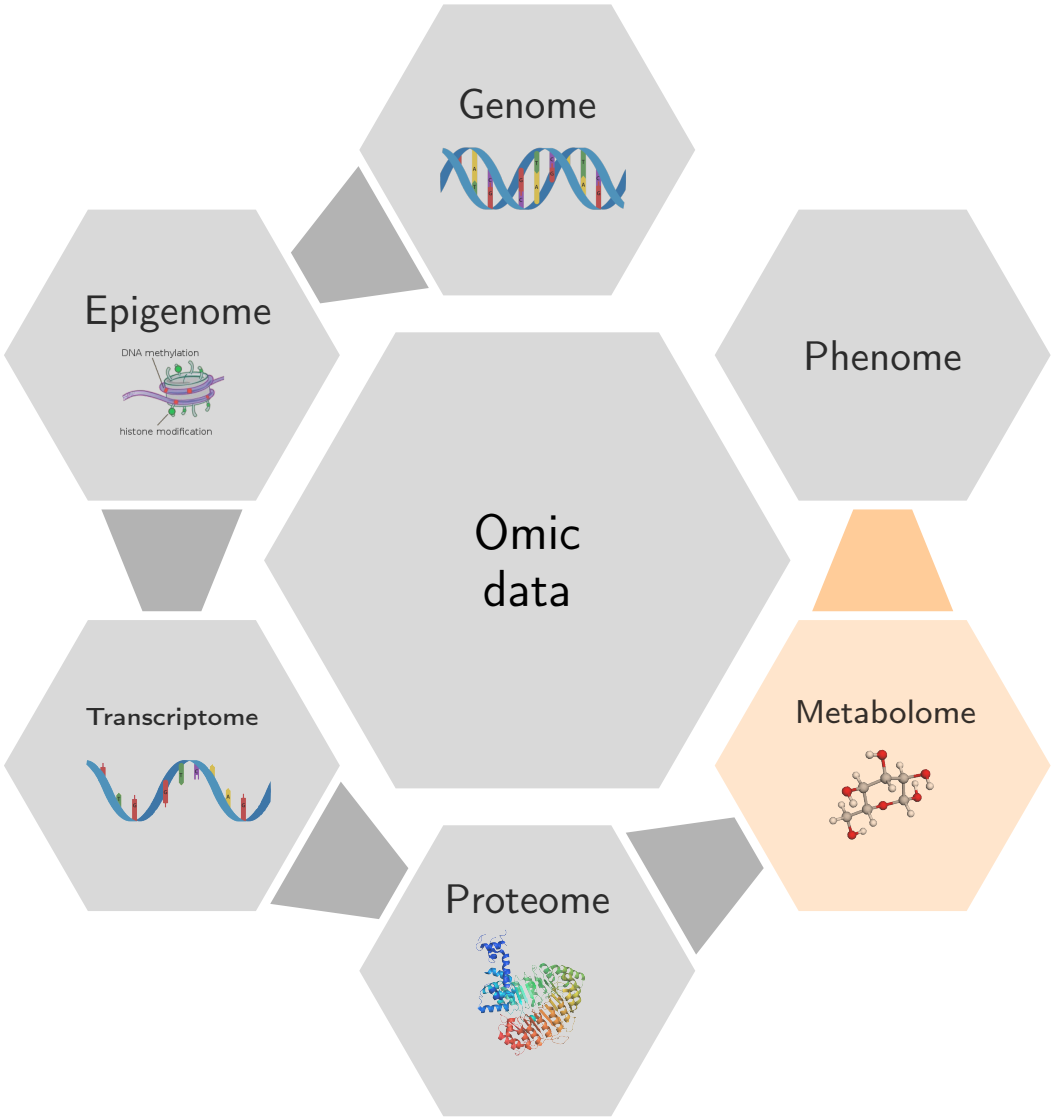




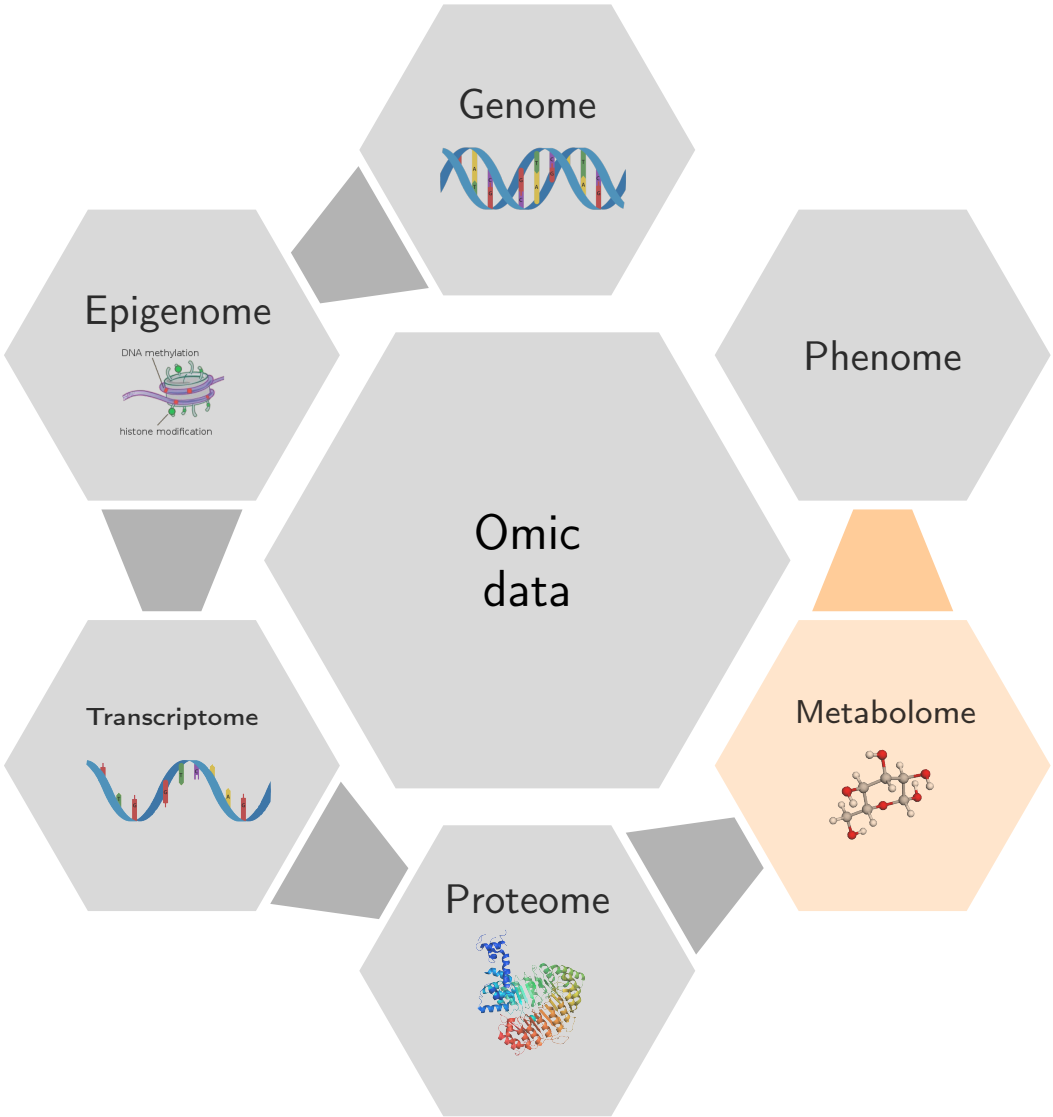
# Omic data and metabolome



# Omic data and metabolome



# Omic data and metabolome




## Advantages

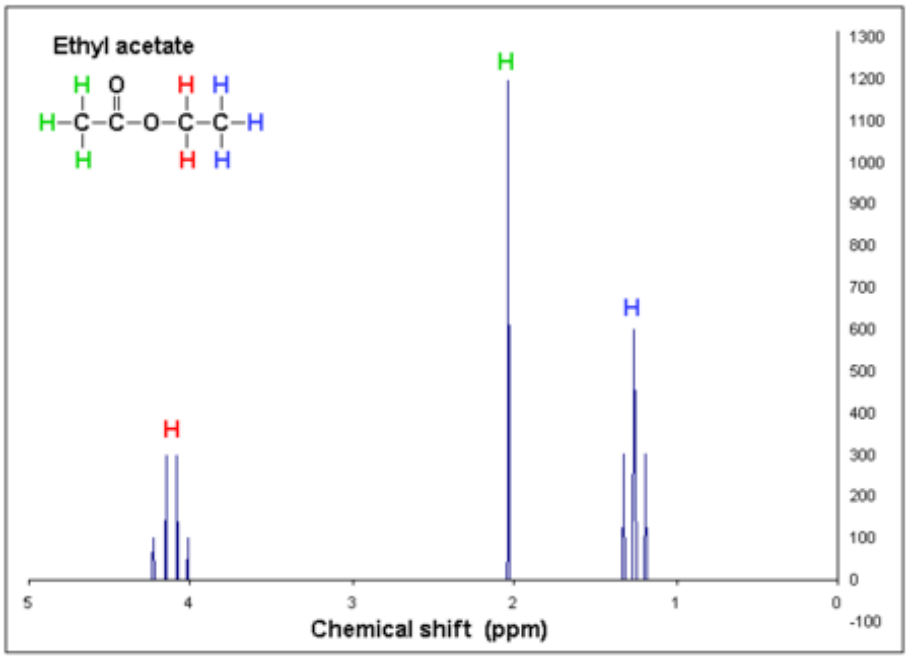
- Close to final phenotypes
- Relatively inexpensive for Nuclear Magnetic Resonance (NMR)

## Drawback

- Difficult to interpret

# Nuclear magnetic resonance (NMR)

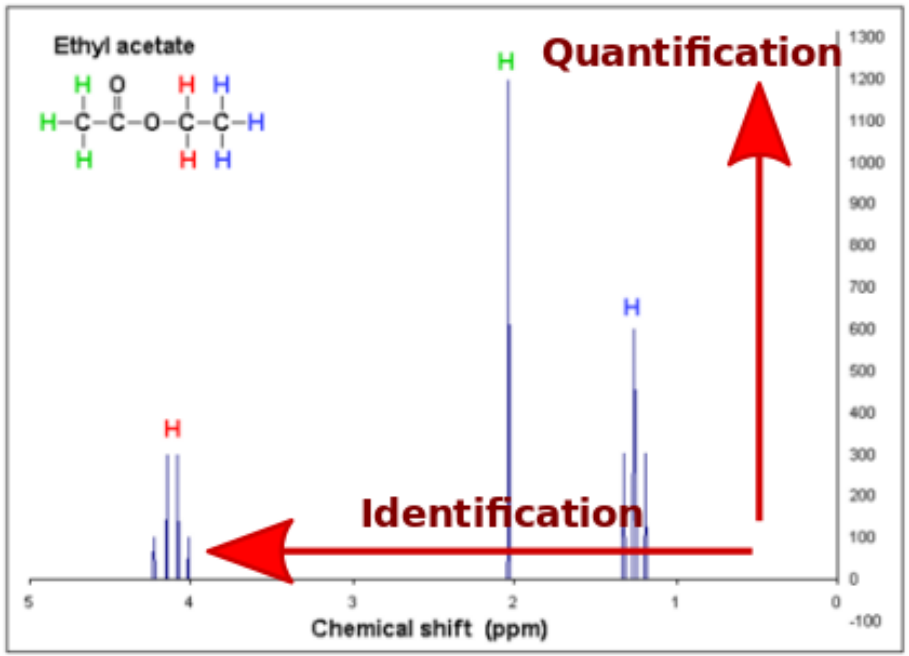
Spectrum of a pure metabolite (*i.e.* standard)




1 metabolite $\Longleftrightarrow$ Several peaks

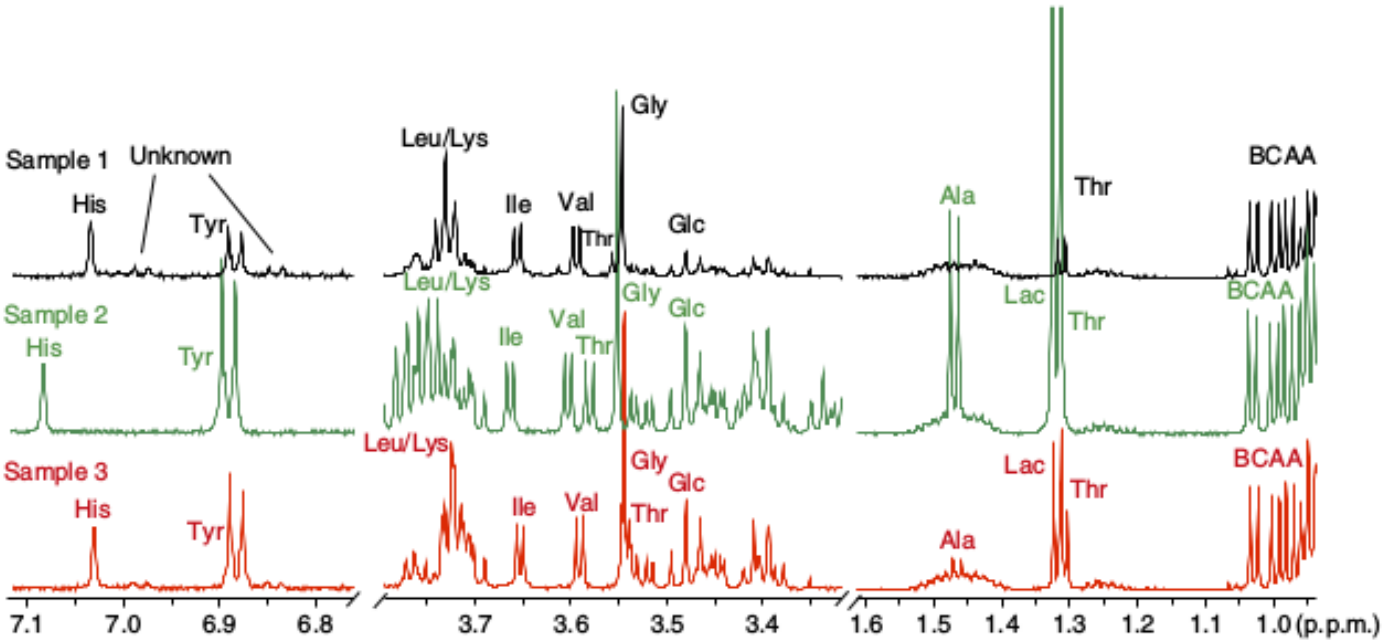
# Nuclear magnetic resonance (NMR)

Spectrum of a pure metabolite (*i.e.* standard)



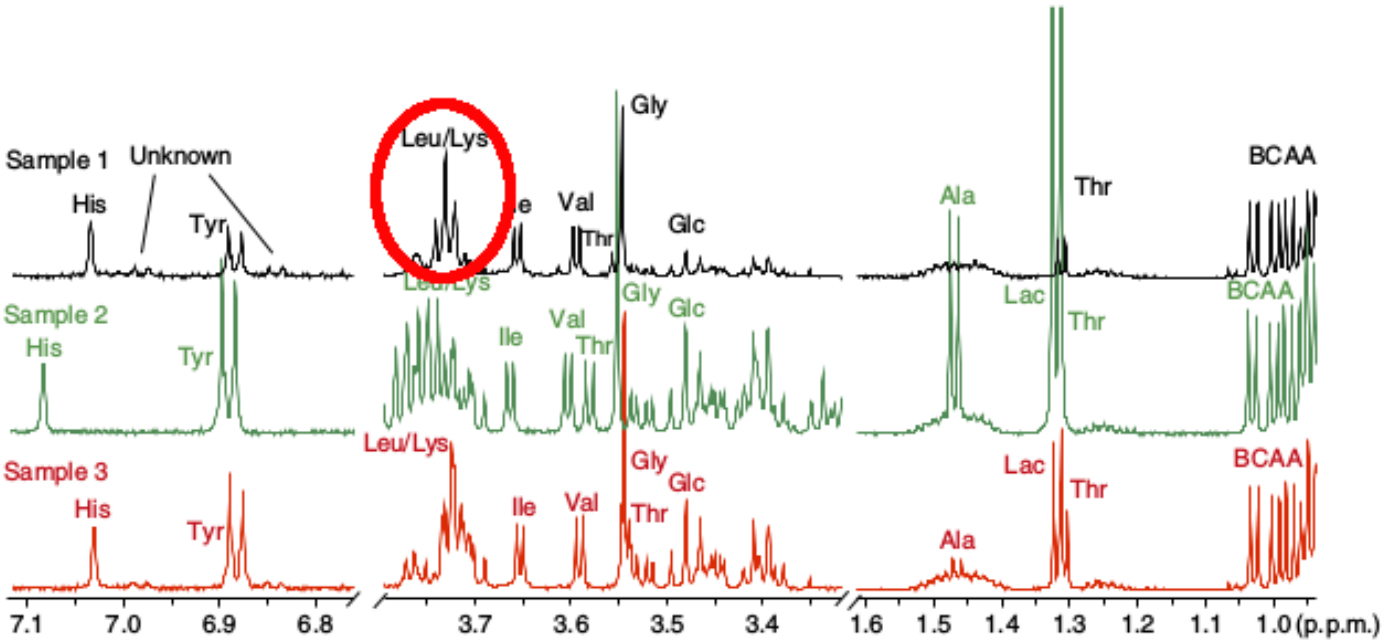

1 metabolite $\Longleftrightarrow$ Several peaks

# Spectra of complex mixtures




[Hao et al., 2012]

# Spectra of complex mixtures

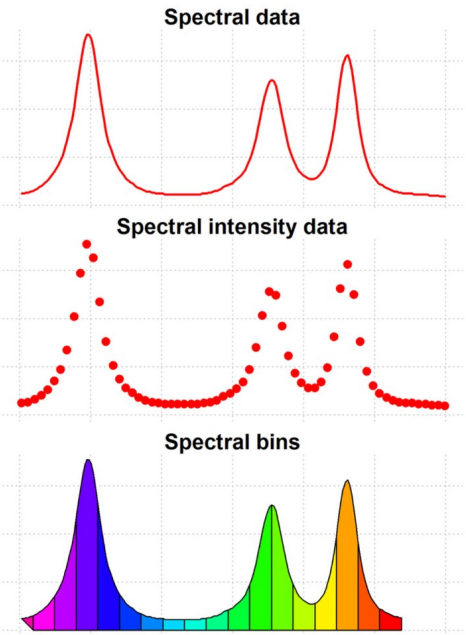



1 peak $\Longleftrightarrow$ Several metabolites

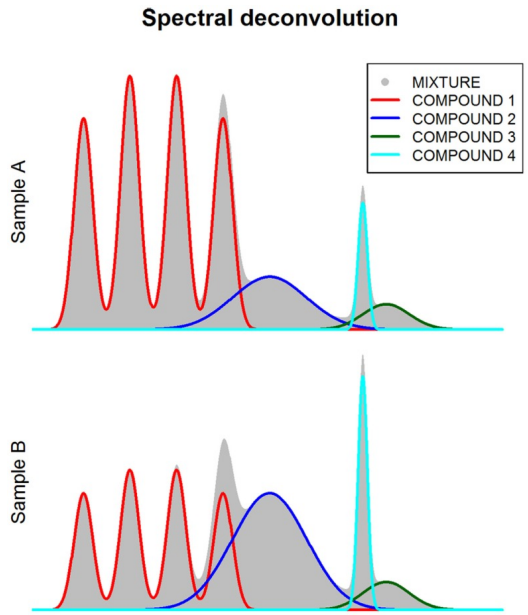
[Hao et al., 2012]

# How to use a spectrum in an analysis?

## Bucketing + identification




## Quantification




[Alonso et al., 2015]

# How to use a spectrum in an analysis?

## Identification

- Manual identification by an expert
- MetaboHunter [Tulpan et al., 2011]

## Quantification

- Autofit [Weljie et al., 2006]
- batman [Hao et al., 2012]
- Bayesil [Ravanbakhsh et al., 2015]
- rDolphin [Cañueto et al., 2018]
- ASICS [Tardivel et al., 2017]

# 1- ASICS: Automatic Statistical Identification in Complex Spectra

# ASICS: Automatic Statistical Identification in Complex Spectra

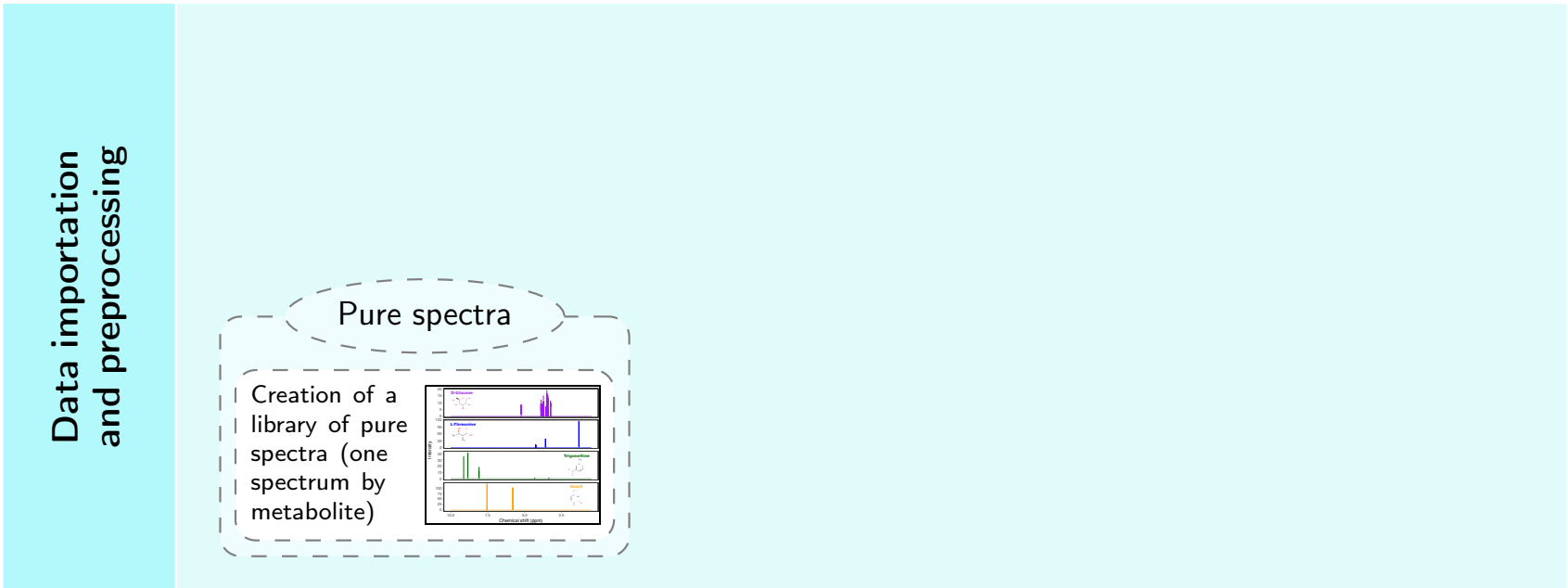
First idea From a library with spectra of pure metabolites, quantify metabolite concentrations of a complex mixture

$$g(t) \simeq \sum_{i=1}^{p} \beta_i f_i(t) \qquad \text{with } \beta_i \geq 0$$

with:

- $g$ a complex mixture
- $f_i$ a set of $p$ pure metabolite spectra
- $\beta_i$ the metabolite concentrations

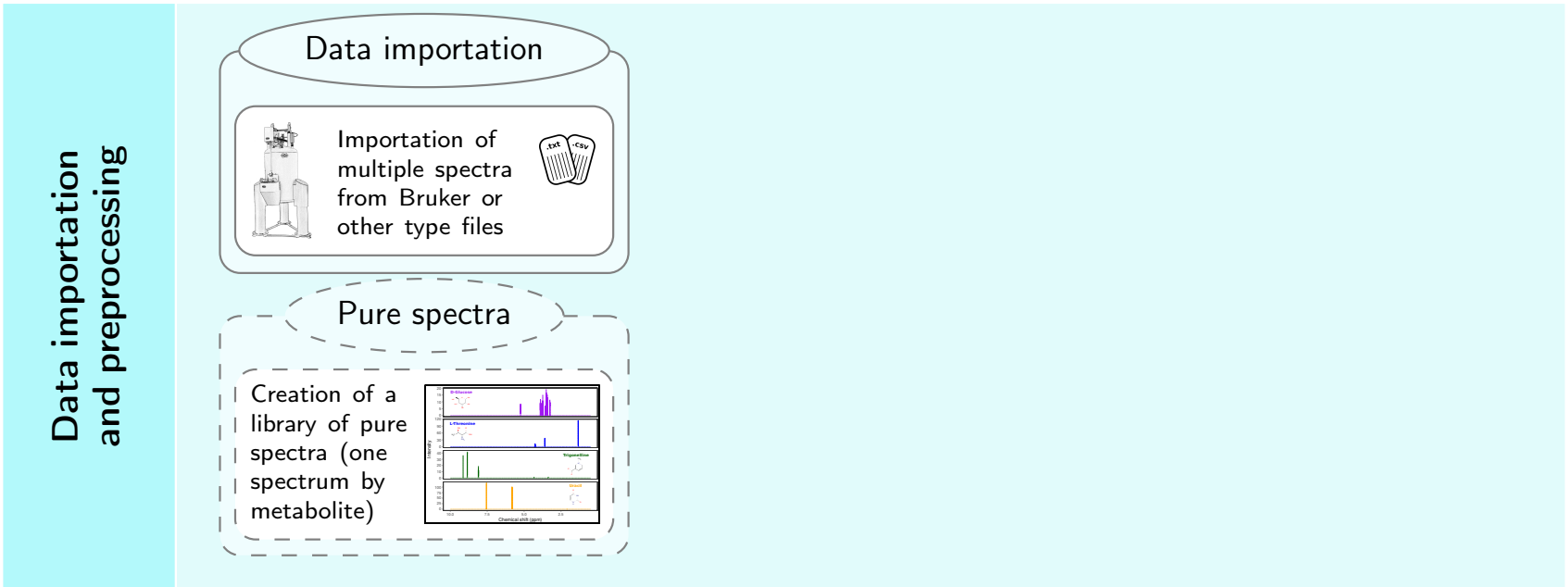
# Data importation and preprocessing



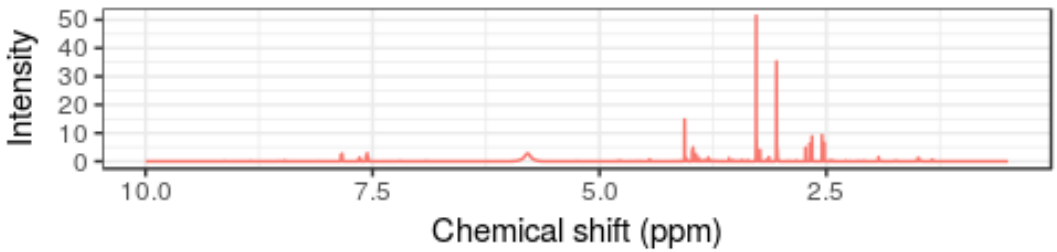

## Pure library

Create a pure library or use the one available in the package (191 pure spectra)

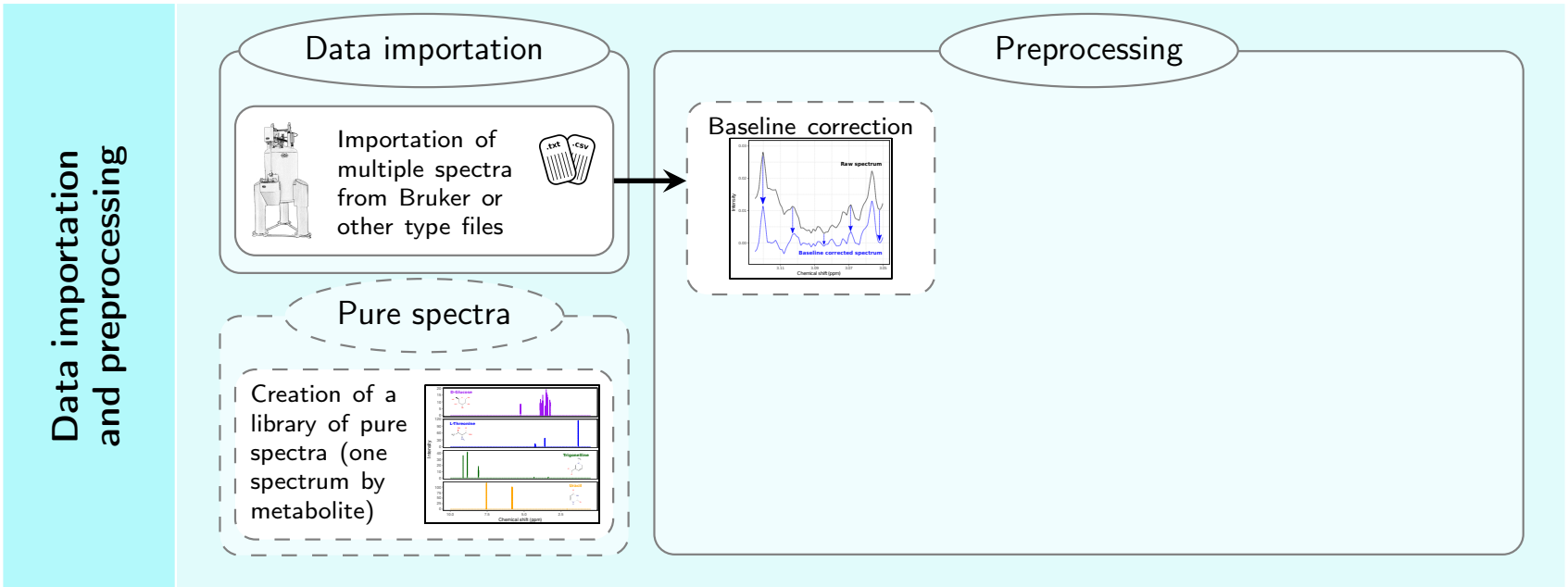
# Data importation and preprocessing




## Data importation



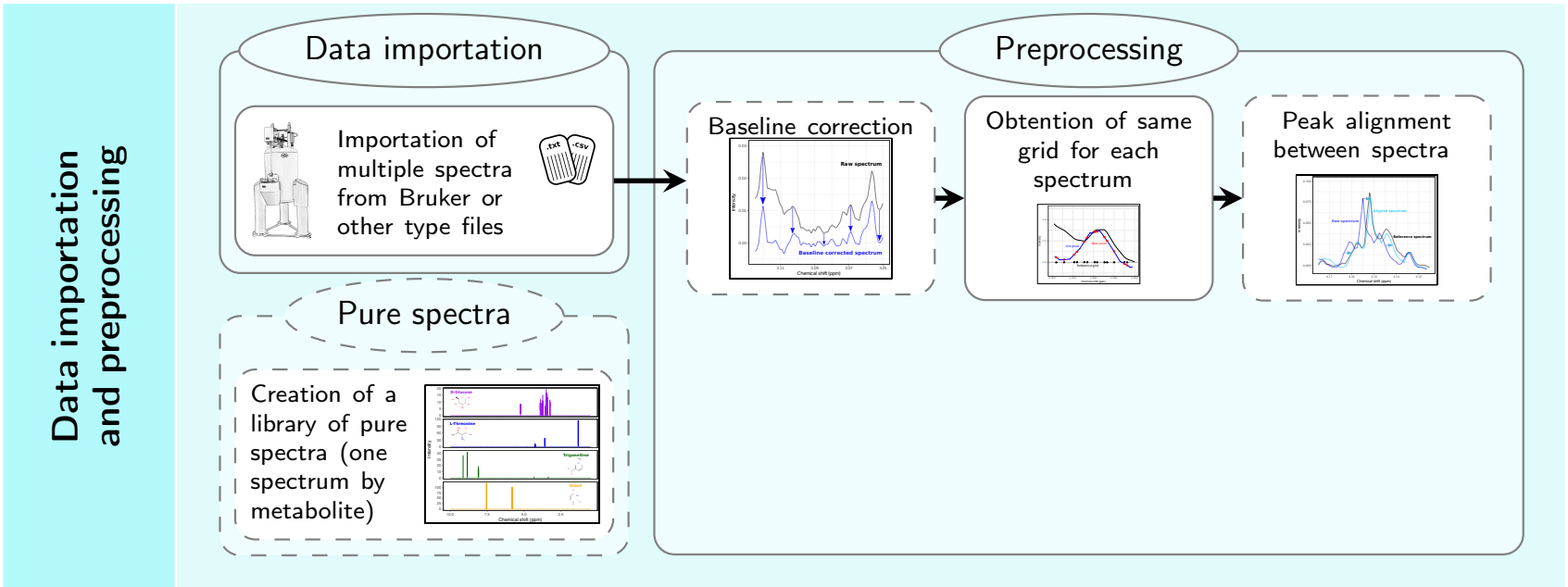
# Data importation and preprocessing




## Baseline correction

Estimate the baseline and substract it from the spectrum [Wang et al., 2013]

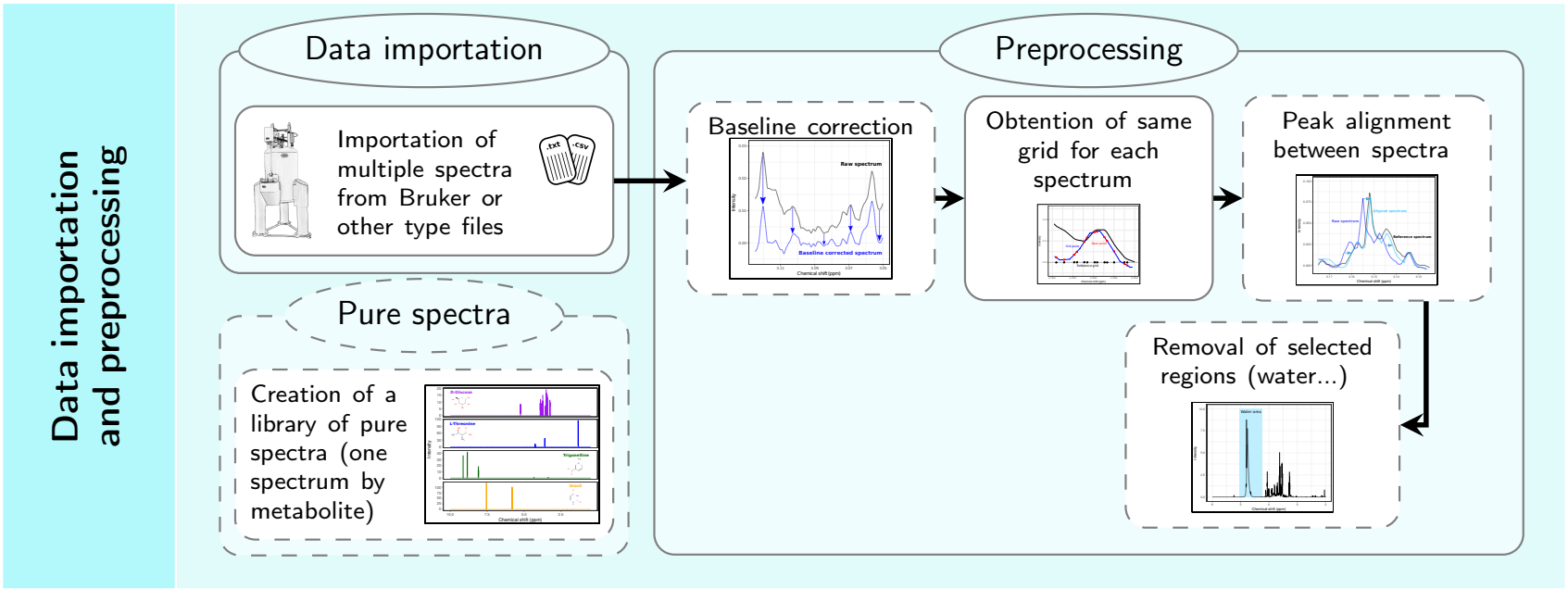
# Data importation and preprocessing




## Peak alignment

Align all spectra on a reference one with a hierarchical clustering (R package speaq; [Vu et al., 2011])

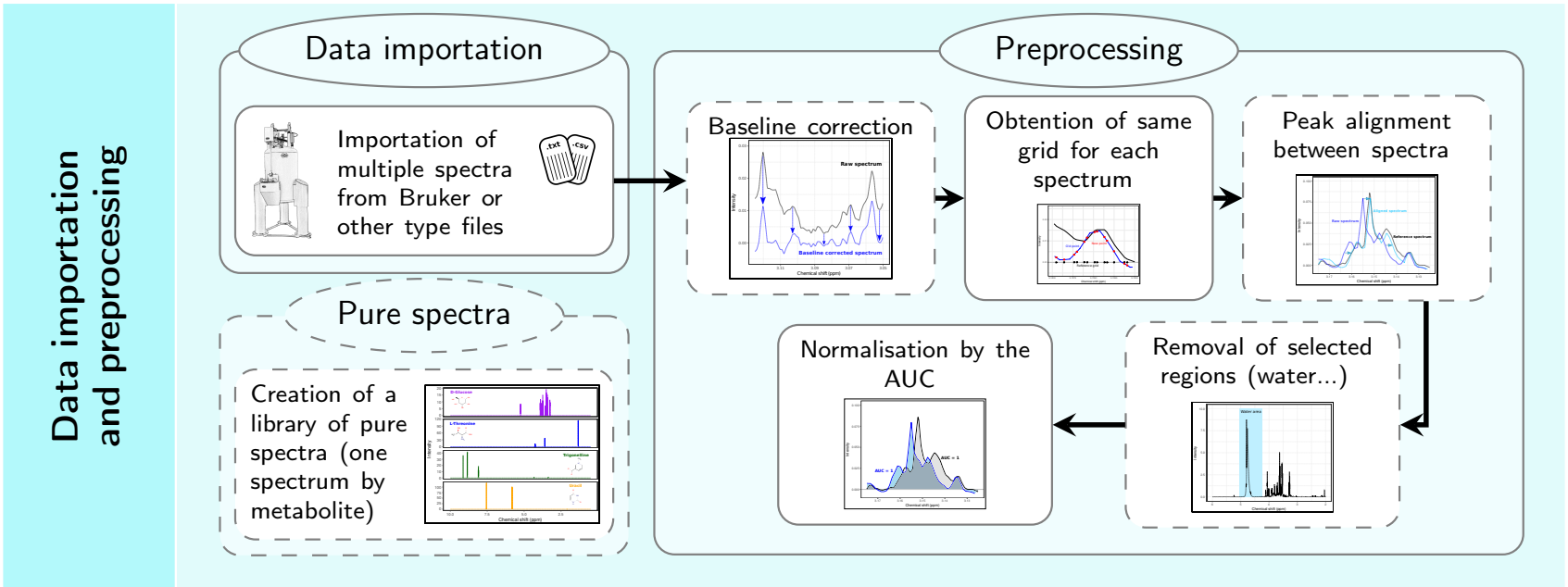
# Data importation and preprocessing




## Removal of unwanted regions

Remove regions of no interest like water region (4.5-5.1 ppm)

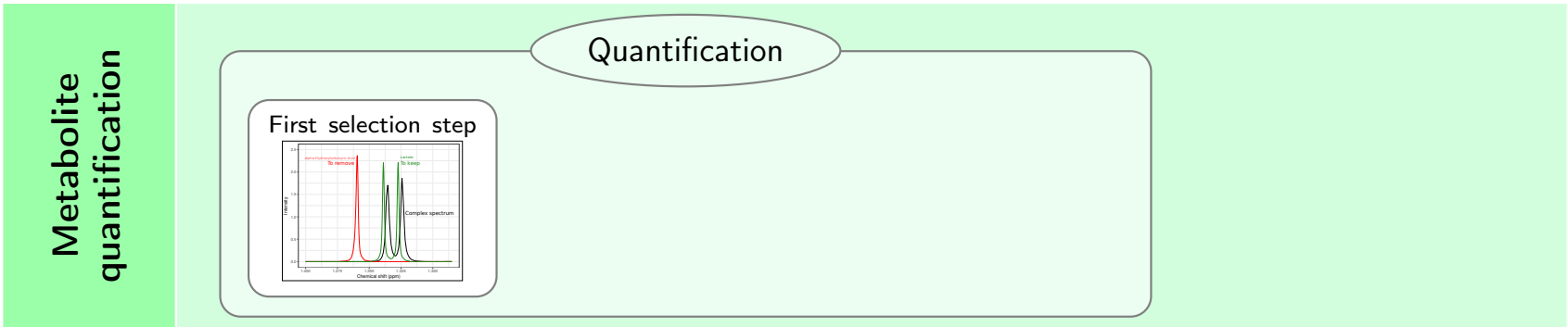
# Data importation and preprocessing



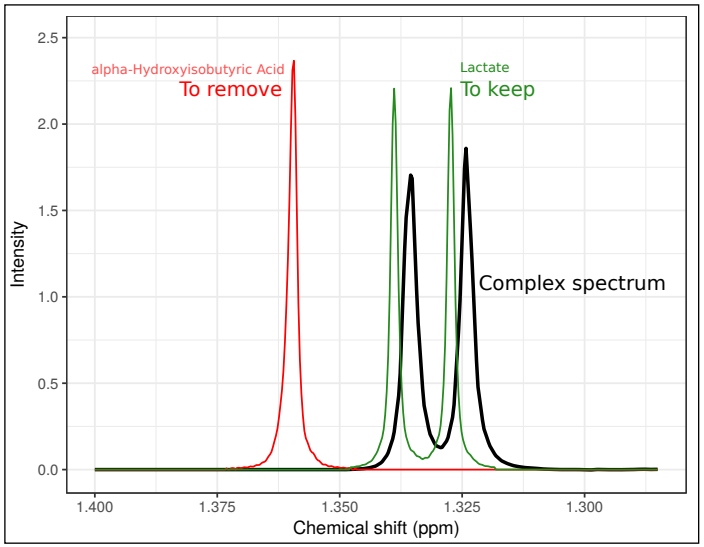

## Normalisation

Normalise each spectrum to a constant sum [Craig et al., 2006]

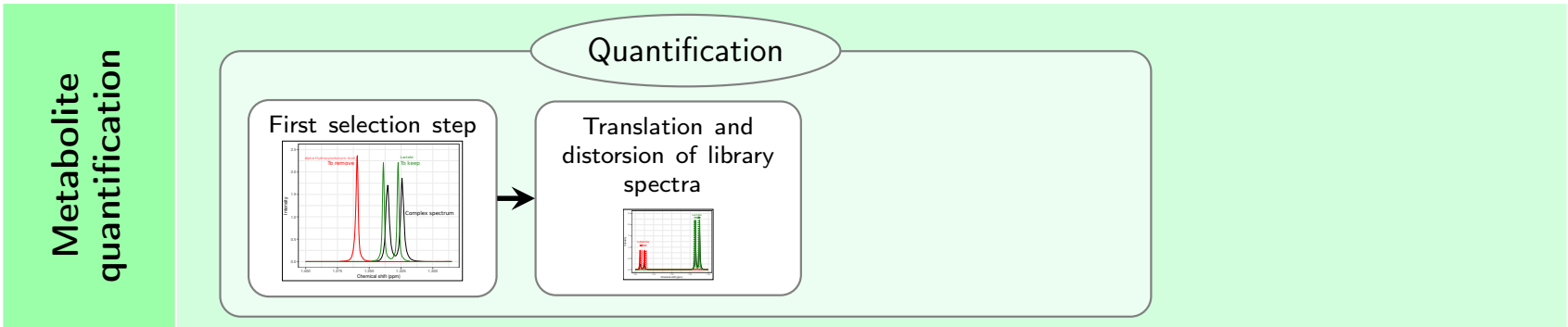
# Metabolite quantification



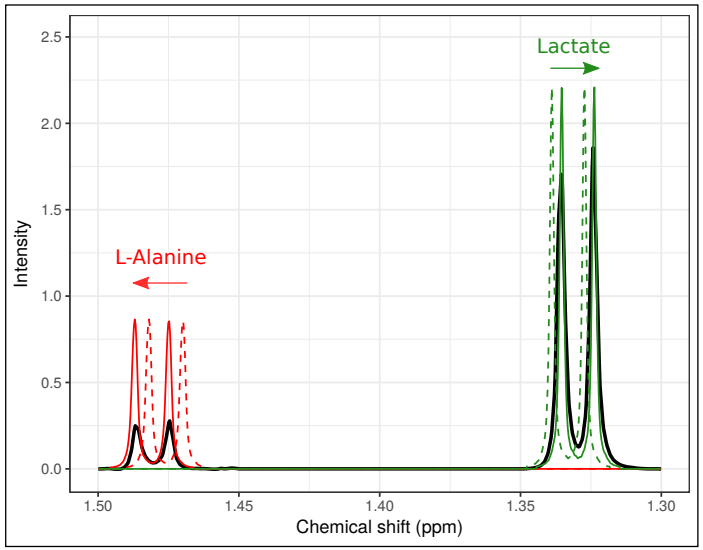

## First selection step



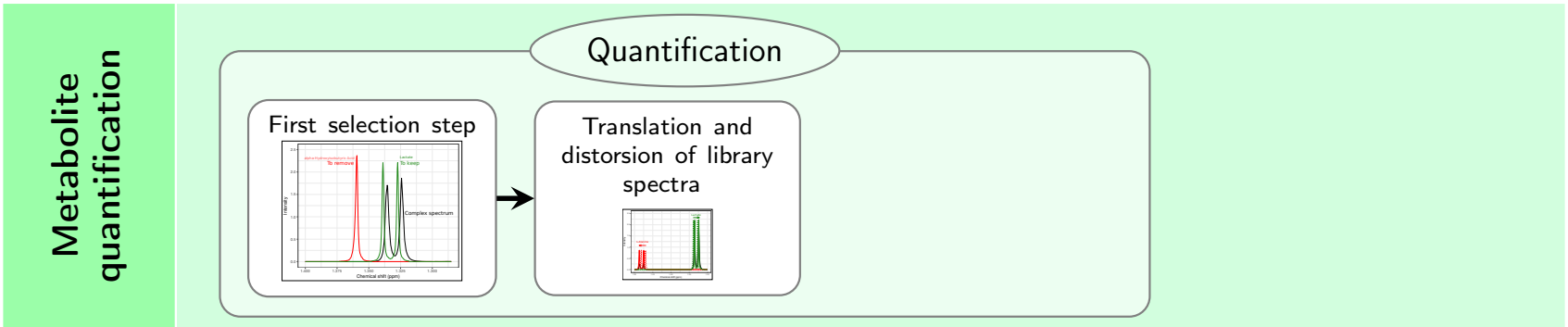
# Metabolite quantification




## Library alignment



# Metabolite quantification

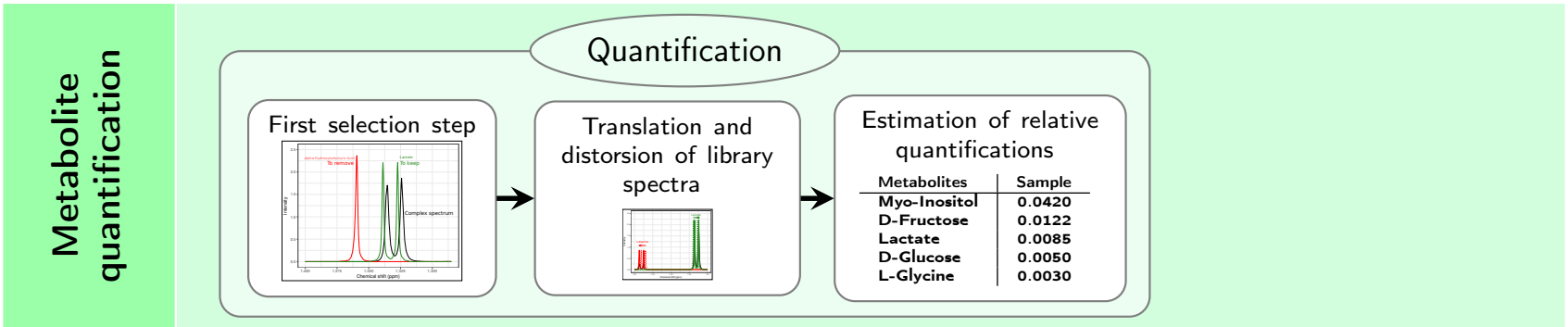



## Library alignment

1. Global alignment with the Fast Fourier Transform Cross-Correlation [Wong et al., 2005] for each metabolite
2. Localized distorsion of each peak grid with the function:

$$\phi(x) = ax(1 - x) + x$$

# Metabolite quantification

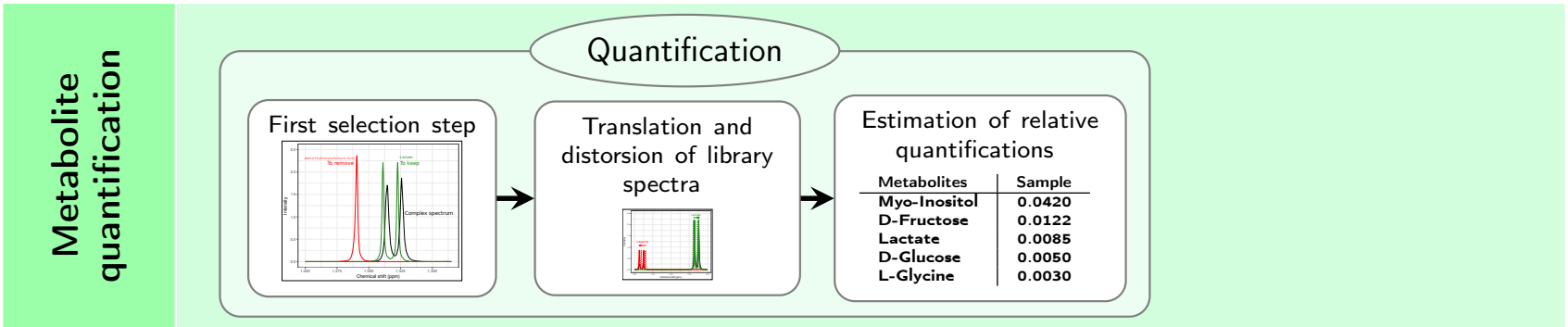



## Metabolite quantification

1. Estimate the coefficient $\beta_i$ :

$$g(t) = \sum_{i=1}^{p} \beta_i f_i(\Phi_i(t)) + \epsilon(t) \qquad \text{with } \beta_i \geq 0$$

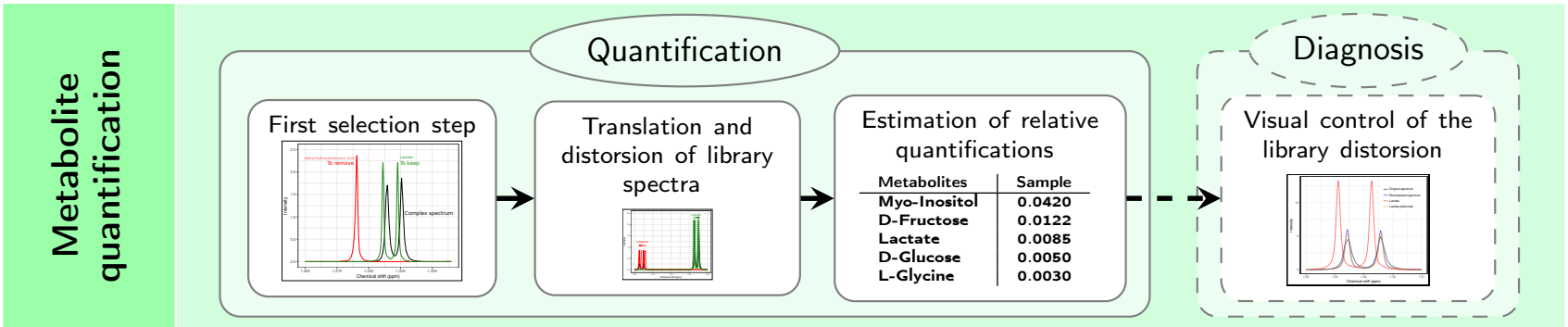
# Metabolite quantification



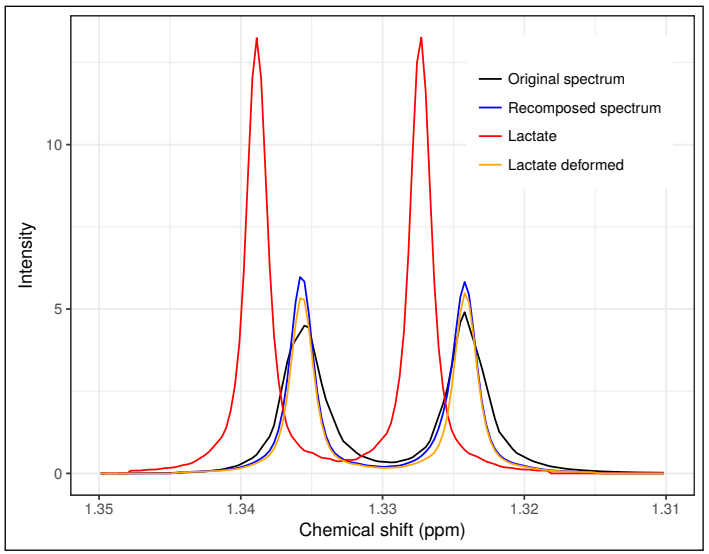

## Metabolite quantification

1. Estimate the coefficient $\beta_i$
2. Test $\mathcal{H}_i : \beta_i = 0$, $\mathcal{H}_i$ is rejected when $\hat{\beta}_i > s_i^*$ [Tardivel et al., 2018]
   - Avoid to wrongly identify too many metabolites by controlling the FamilyWise Error Rate (FWER)
   - Handle peak overlapping

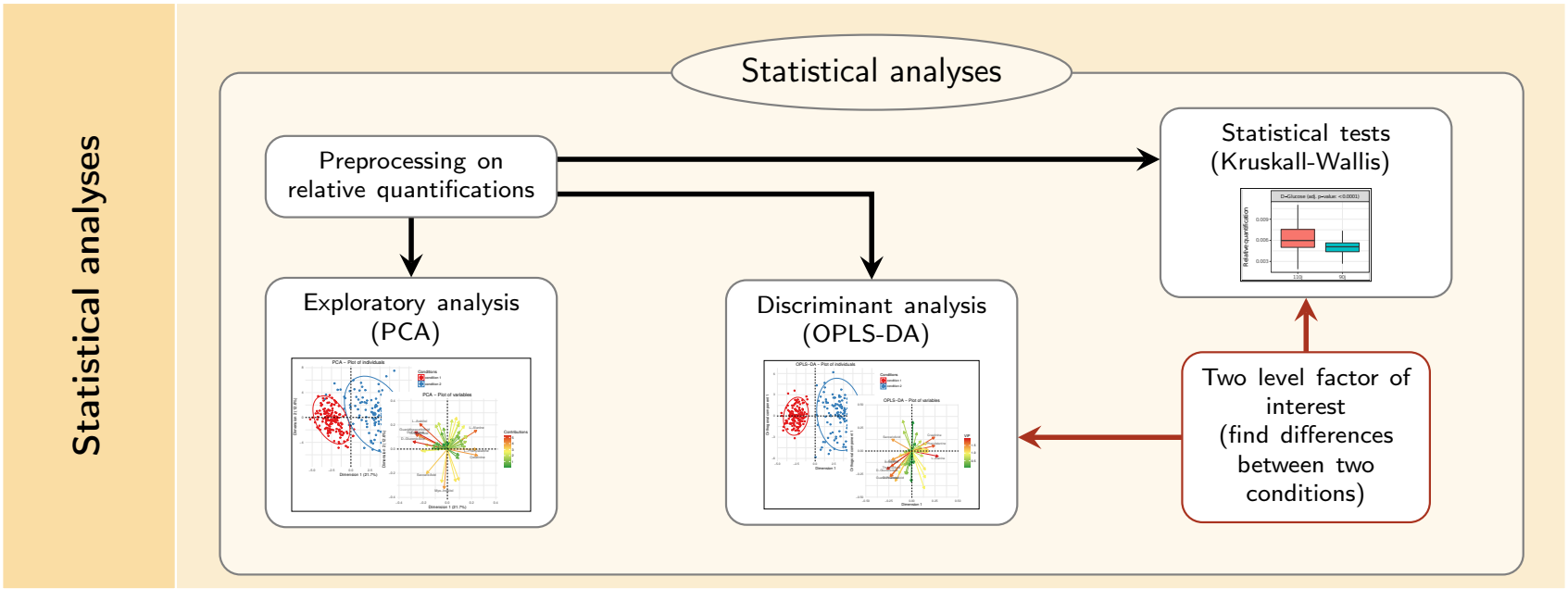
# Metabolite quantification




## Visual control



# Statistical analyses



# 2- **ASICS performances**

## 2.1- Validation 1: Comparison on a mixture with known metabolites

# Comparison of identified metabolites

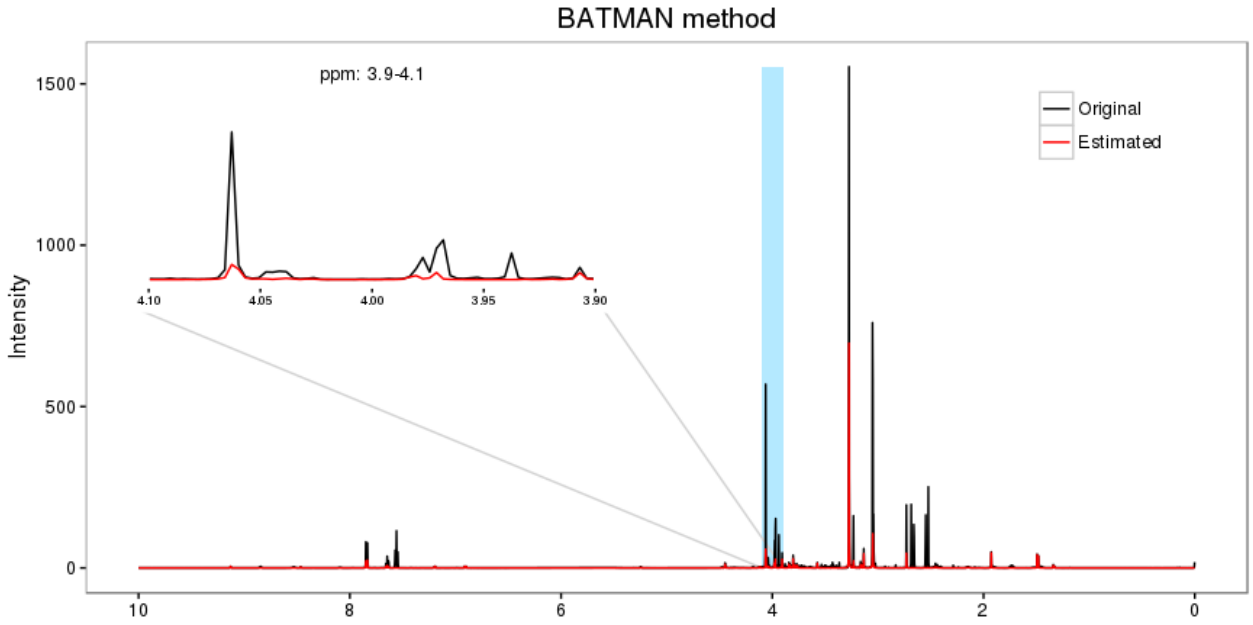
**Complex mixture** Mixture of 21 metabolites in known concentrations

**Quantification results**

# Comparison of identified metabolites

**Complex mixture** Mixture of 21 metabolites in known concentrations

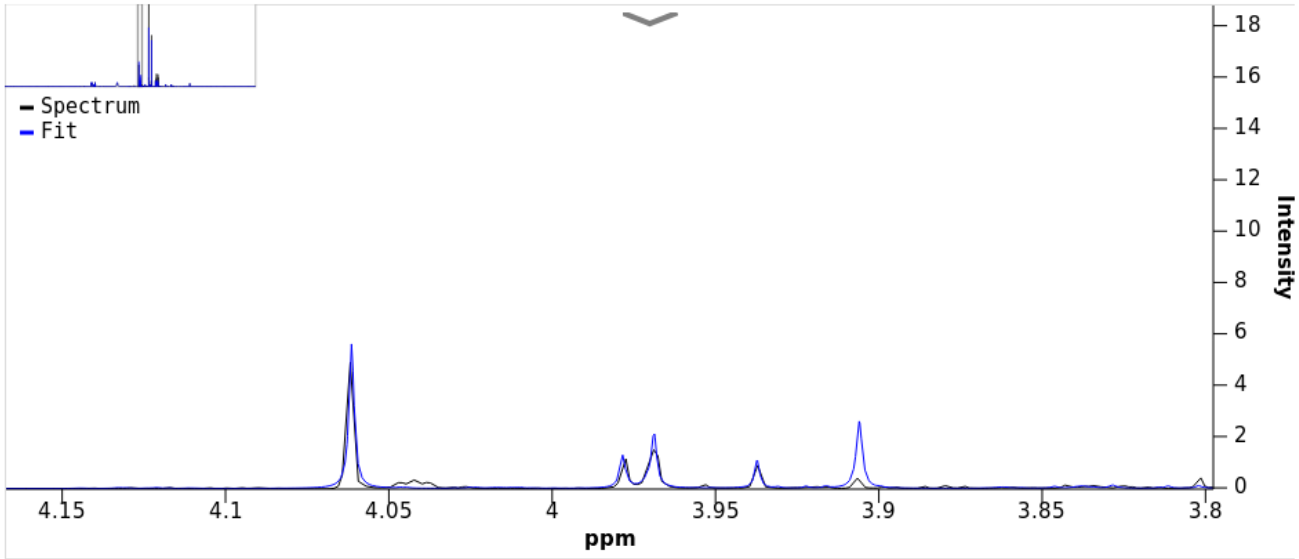
**Quantification results** With batman :



# Comparison of identified metabolites

**Complex mixture** Mixture of 21 metabolites in known concentrations

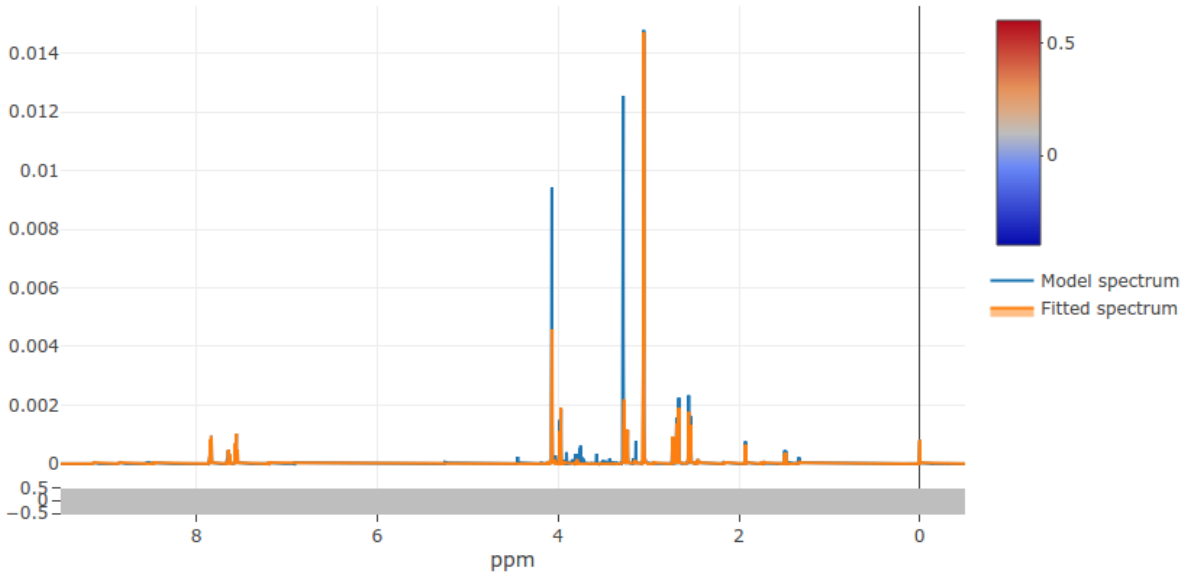
**Quantification results** With Bayesil :



# Comparison of identified metabolites

Complex mixture Mixture of 21 metabolites in known concentrations

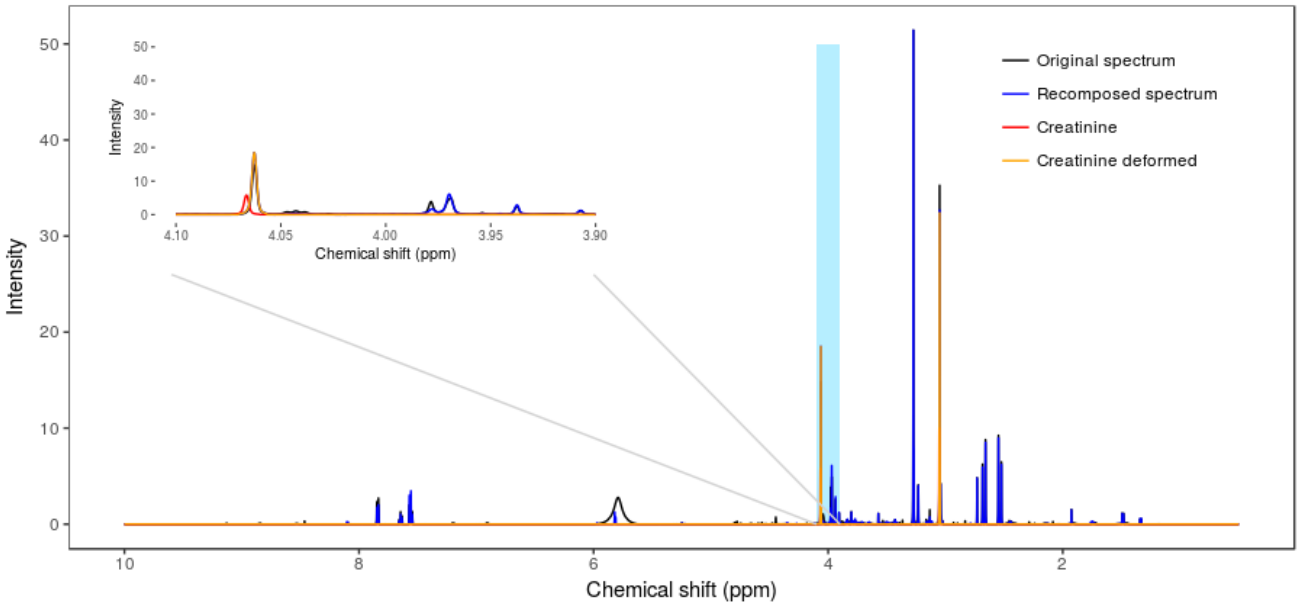
Quantification results With rDolphin :



# Comparison of identified metabolites

**Complex mixture** Mixture of 21 metabolites in known concentrations

**Quantification results** With ASICS :



# Comparison of identified metabolites

Complex mixture Mixture of 21 metabolites in known concentrations

Quantification results

Comparison of the identification of the five methods

| **Method** | **True positive (/Identified metabolites)** | **Accuracy (%)** | **Compounds in library** | **Computing time** |
|---|---|---|---|---|
| ASICS | 18/27 | 93 | 191 | 1min30 |
| Autofit | 15/63 | 54 | 338 | <1min |
| MetaboHunter | 4/55 | 92 | 867 | <1min |
| Batman | 21/146 | 18 | 147 | 74h |
| Bayesil | 17/56 | 54 | 90 | 30min |
| rDolphin | 17/48 | 56 | 89 | 1min30 |

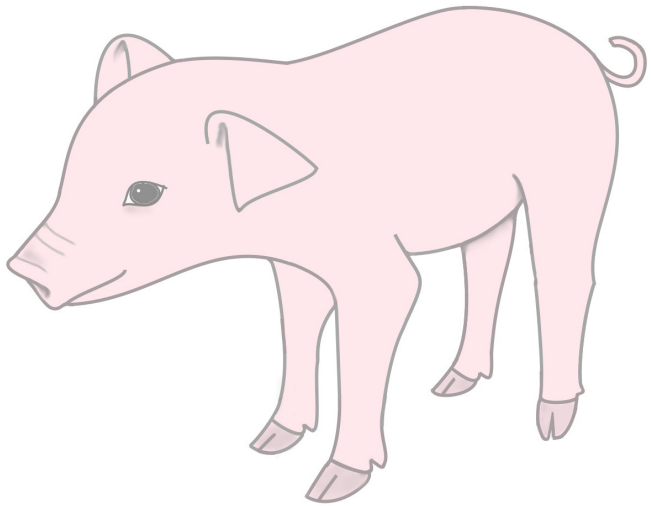
# 2- **ASICS performances**

## 2.2- Validation 2: Comparison on real data (several complex mixtures)

# PORCINET

**Goal** Understand the mechanisms of maturity and perinatal survival in pigs

**Animals** 283 Large White fetuses at two stages of gestation (90 and 110 days)



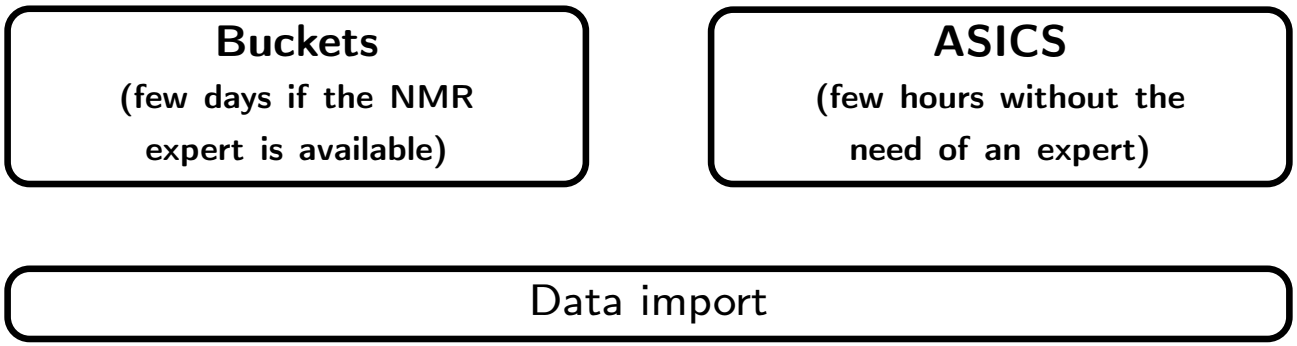
**Data**

- NMR spectra in plasma
- Biochemical dosages of some target metabolites
  $\Rightarrow$ to validate quantification

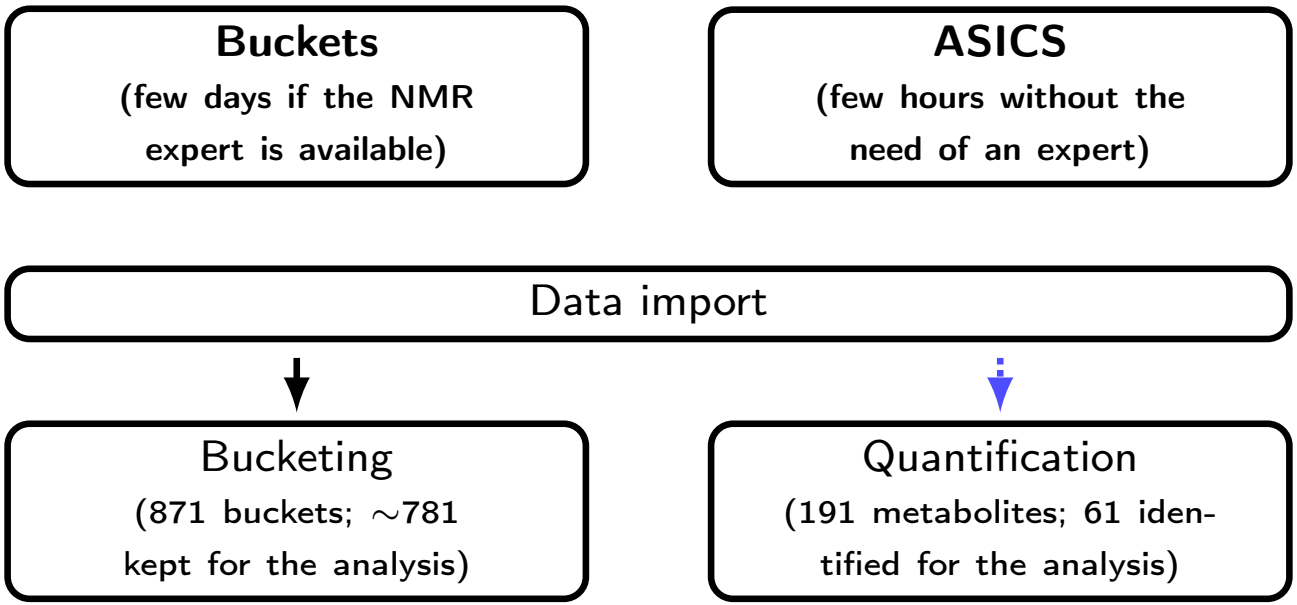
# Comparison with biochemical dosages

| | Lactate | Fructose | Glucose |
|---|---|---|---|
| ASICS | 0.93 | 0.95 | 0.90 |
| Autofit | 0.52 | 0.74 | 0.75 |
| batman | 0.46 | 0.56 | 0.22 |
| rDolphin | 0.82 | Not available | 0.77 |
| Buckets | 0.93 | 0.95 | 0.90 |

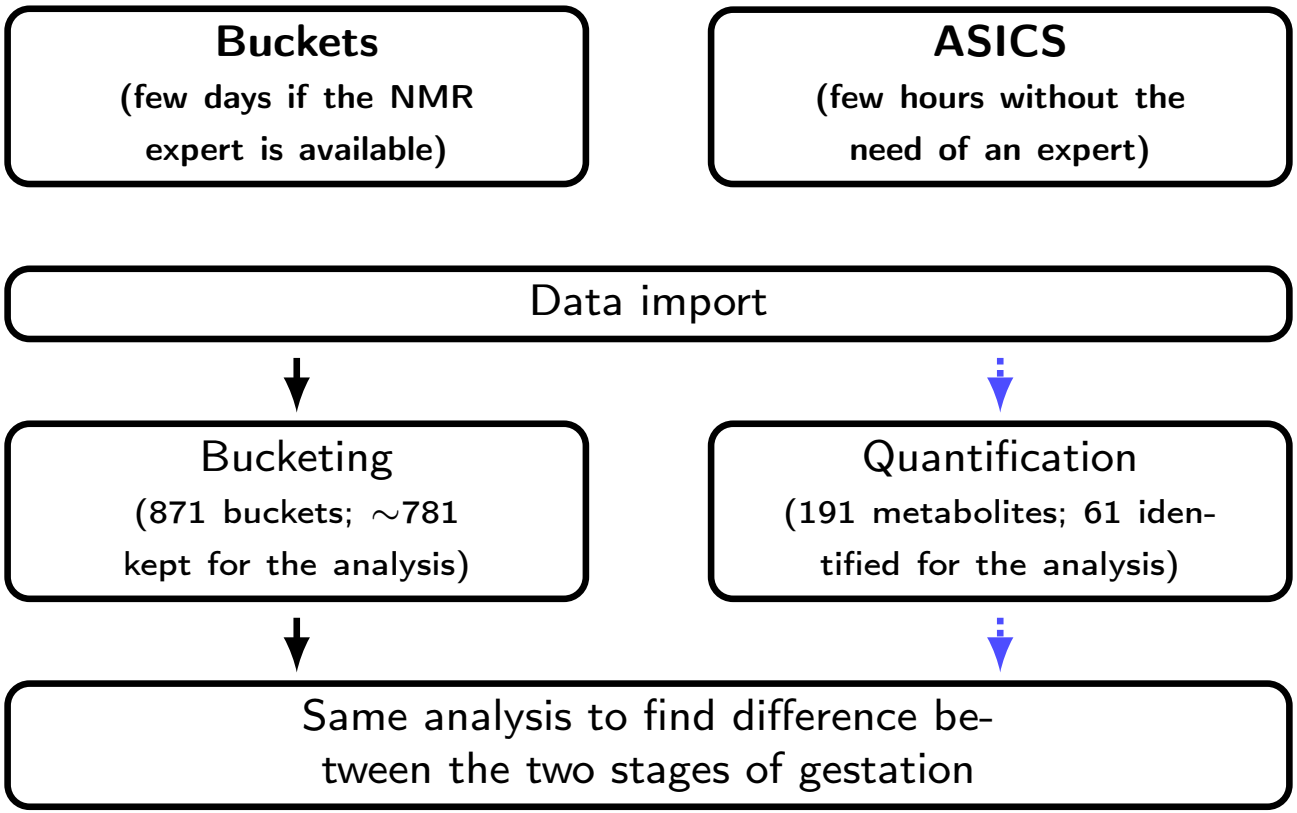
# Analysis workflow



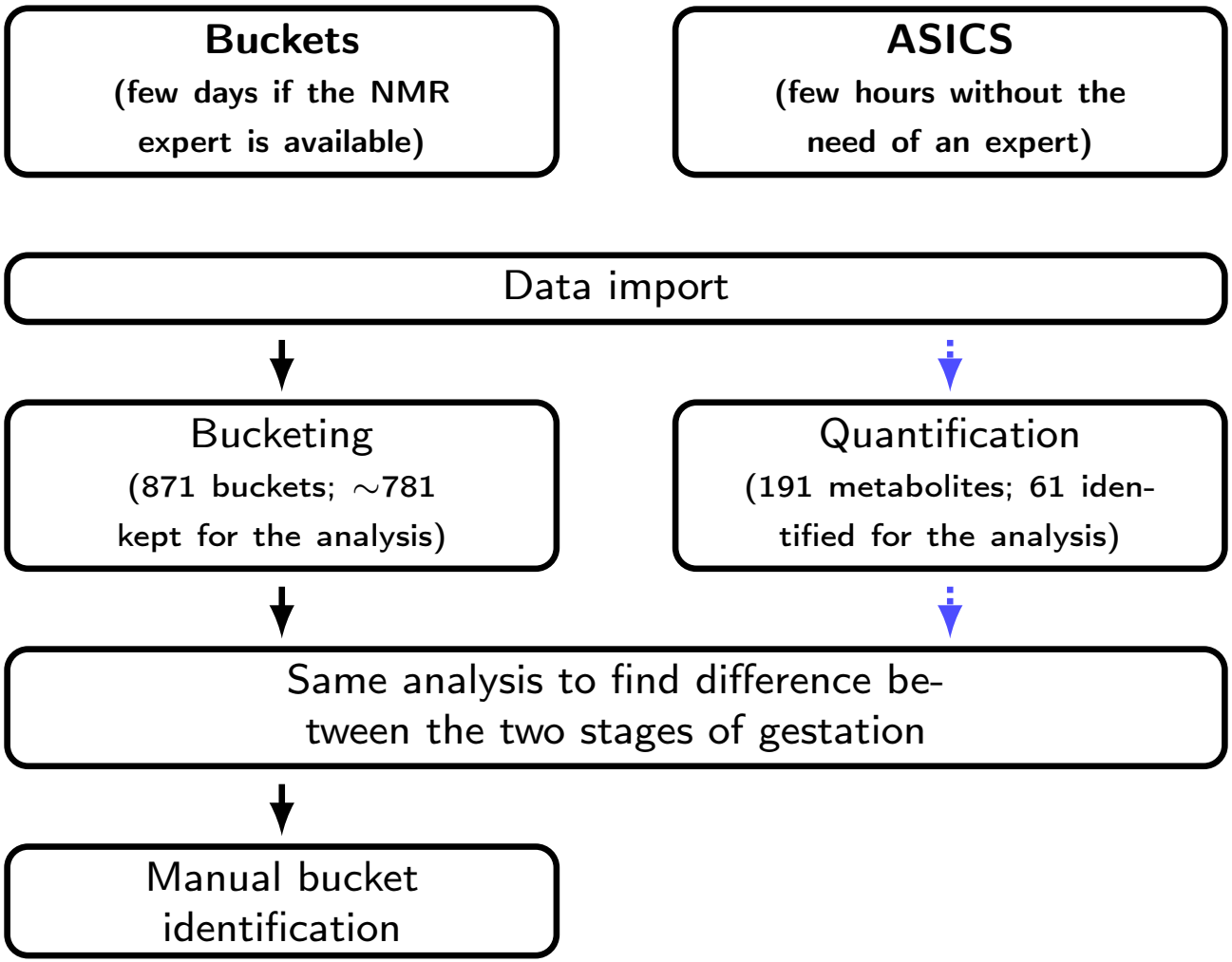
# Analysis workflow



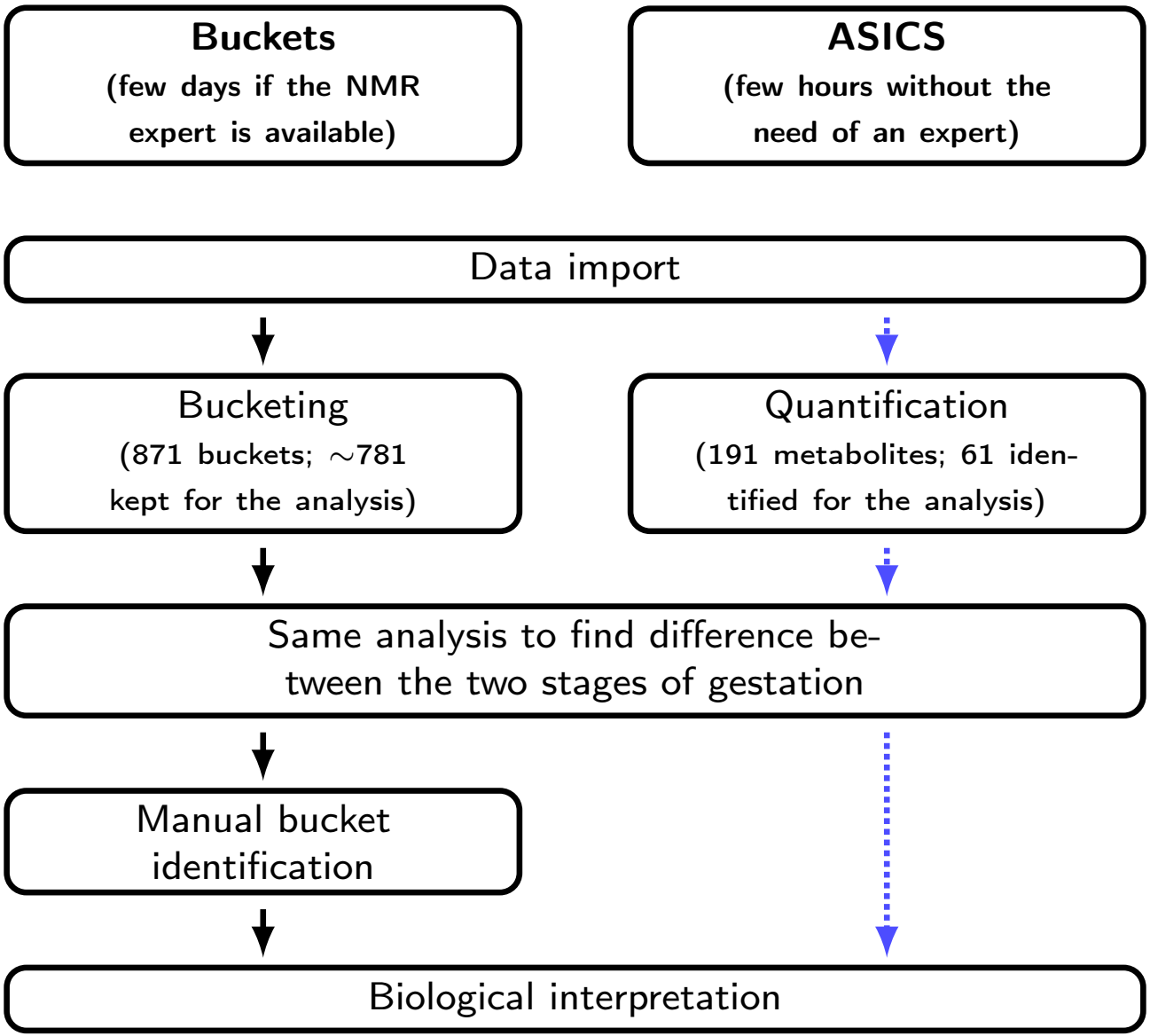
# Analysis workflow



# Analysis workflow



# Analysis workflow



# Comparison of analysis results

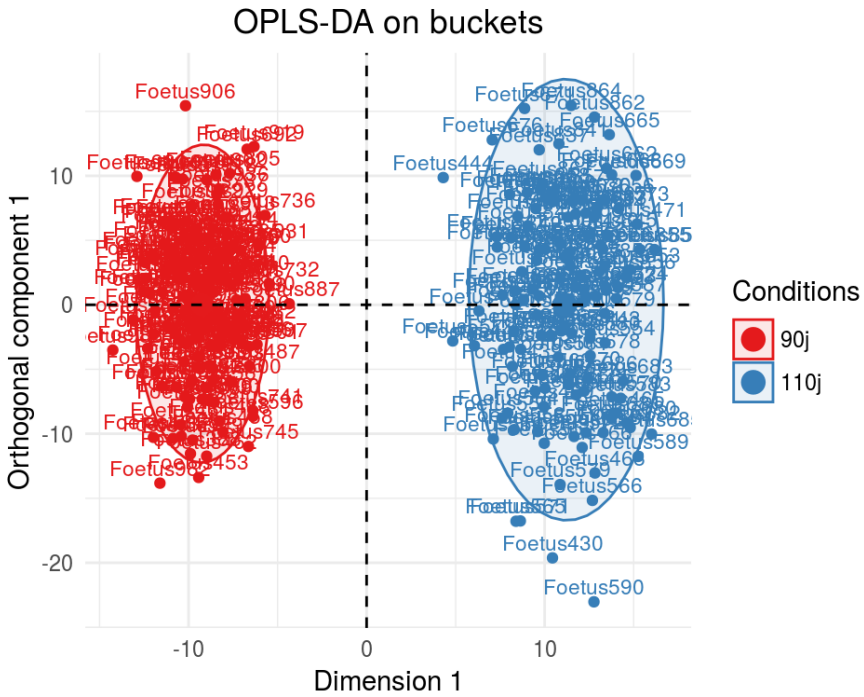
**Method** Orthogonal Partial Least Squares - Discriminant Analysis (OPLS-DA)

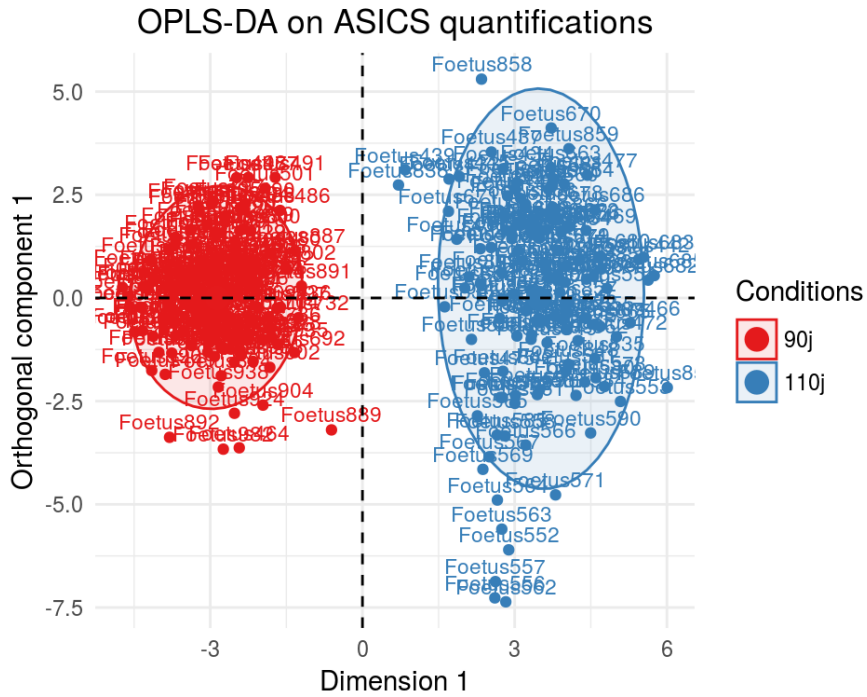
- Supervised classification method
- First component correlated with the variable of interest and a second orthogonal
- The Variable Importance in Projection (VIP) indicator is used for feature selection

# Comparison of analysis results

## Discriminating power

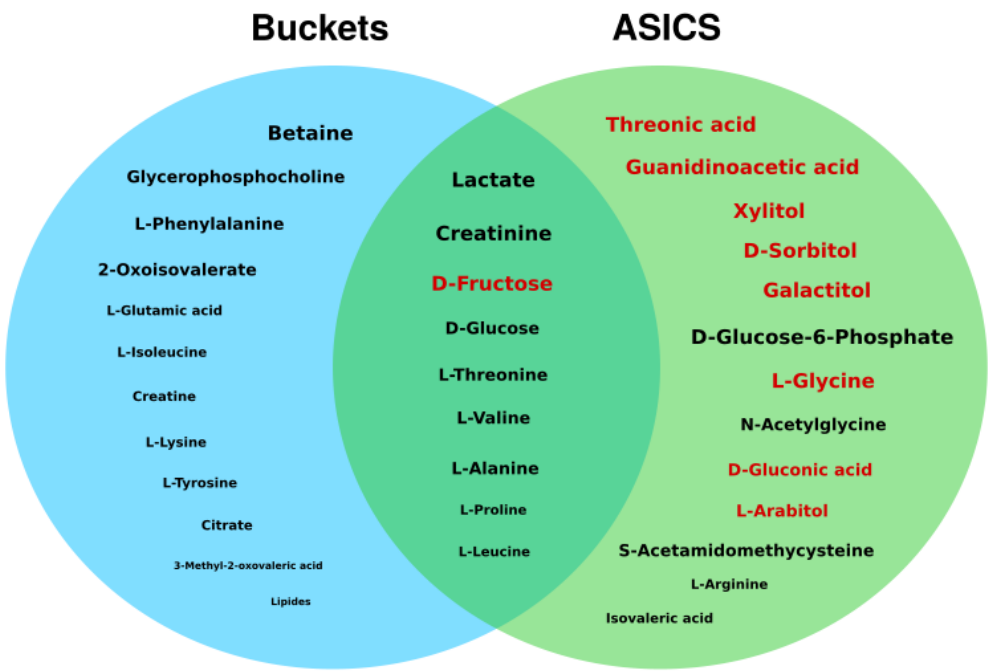
| Analysis | Buckets | ASICS |
|---|---|---|
| Prediction error | 0% | 0% |



# Comparison of analysis results

## Comparison of influential metabolites obtained with each analysis



# Conclusions and perspectives

- Quantification with ASICS on many spectra is automatic and can be executed in a parallel environment
- Reliable quantifications and almost as accurate than an analysis on buckets
- Results directly interpretable without expert NMR knowledge

- Package available on Bioconductor (`http://bioconductor.org/packages/ASICS/`)

- Improve identification and quantification using the information coming from several spectra
- Integrated quantification/differential analysis method

Thank you for your attention!

# Bibliography (1)


Alonso, A., Marsal, S., and JuliÃ , A. (2015).

Analytical Methods in Untargeted Metabolomics: State of the Art in 2015.
*Frontiers in Bioengineering and Biotechnology*, 3:23.

Cañueto, D., Gómez, J., Salek, R. M., Correig, X., and Cañellas, N. (2018).

rDolphin: a GUI R package for proficient automatic profiling of 1D 1H-NMR spectra of study datasets.
*Metabolomics*, 14(3):24.

Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., and Lindon, J. C. (2006).

Scaling and normalization effects in NMR spectroscopic metabonomic data sets.
*Analytical Chemistry*.

Hao, J., Astle, W., De Iorio, M., and Ebbels, T. M. D. (2012).

BATMAN–an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model.
*Bioinformatics*, 28(15):2088–2090.

Ravanbakhsh, S., Liu, P., Bjordahl, T. C., Mandal, R., Grant, J. R., Wilson, M., Eisner, R., Sinelnikov, I., Hu, X., Luchinat, C., Greiner, R., and Wishart, D. S. (2015).

Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics.
*PLOS ONE*, 10(5):e0124219.

# Bibliography (2)


Tardivel, P., Canlet, C., Lefort, G., Tremblay-Franco, M., Debrauwer, L., Concordet, D., and Servien, R. (2017).
ASICS: an automatic method for identification and quantification of metabolites in complex 1D 1H NMR spectra.
*Metabolomics*, 13(10):109.

Tardivel, P., Servien, R., and Concordet, D. (2018).
A powerful multiple testing procedure in linear Gaussian model.

Tulpan, D., Léger, S., Belliveau, L., Culf, A., and Čuperlović-Culf, M. (2011).
MetaboHunter: an automatic approach for identification of metabolites from 1H-NMR spectra of complex mixtures.
*BMC Bioinformatics*, 12(1):400.

Vu, T. N., Valkenborg, D., Smets, K., Verwaest, K. A., Dommisse, R., Lemière, F., Verschoren, A., Goethals, B., and Laukens, K. (2011).
An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data.
*BMC Bioinformatics*, 12(1):405.

Wang, K.-C., Wang, S.-Y., Kuo, C.-h., and Tseng, Y. J. (2013).
Distribution-Based Classification Method for Baseline Correction of Metabolomic 1D Proton Nuclear Magnetic Resonance Spectra.
*Analytical Chemistry*, 85(2):1231–1239.

# Bibliography (3)


Weljie, A. M., Newton, J., Mercier, P., Carlson, E., and Slupsky, C. M. (2006).

Targeted Profiling: Quantitative Analysis of 1 H NMR Metabolomics Data.
*Analytical Chemistry*, 78(13):4430–4442.

Wong, J. W. H., Durante, C., and Cartwright, H. M. (2005).

Application of Fast Fourier Transform Cross-Correlation for the Alignment of Large Chromatographic and Spectral Datasets.
*Analytical Chemistry*, 77(17):5655–5661.