



**HAL**  
open science

## Joint automatic metabolite identification and quantification of a set of $^1\text{H}$ NMR spectra

Gaëlle Lefort, Laurence Liaubet, Nathalie Marty-Gasset, Cécile Canlet,  
Nathalie Vialaneix, Rémi Servien

► **To cite this version:**

Gaëlle Lefort, Laurence Liaubet, Nathalie Marty-Gasset, Cécile Canlet, Nathalie Vialaneix, et al.. Joint automatic metabolite identification and quantification of a set of  $^1\text{H}$  NMR spectra. *Metabolomics* 2020, Oct 2020, online, France. hal-02984360

**HAL Id: hal-02984360**

**<https://hal.science/hal-02984360v1>**

Submitted on 30 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

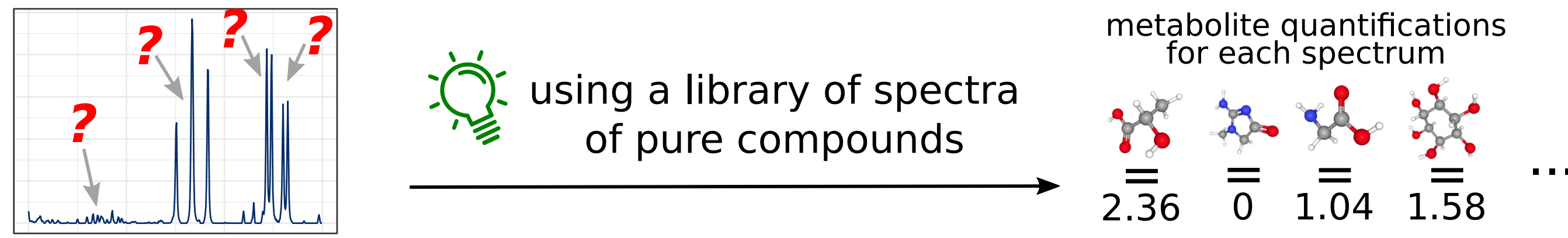


# Joint automatic metabolite identification and quantification of a set of $^1\text{H}$ NMR spectra

G. Lefort<sup>1,2</sup>, L. Liaubet<sup>2</sup>, N. Marty-Gasset<sup>2</sup>, C. Canlet<sup>3,4</sup>, N. Vialaneix<sup>1</sup>, R. Servien<sup>5,6</sup>

<sup>1</sup>INRAE, UR875 Mathématiques et Informatique Appliquées Toulouse, F-31326 Castanet-Tolosan, France, <sup>2</sup>GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet-Tolosan, France, <sup>3</sup>INRAE, Université de Toulouse, ENVT, Toxalim, 31027 Toulouse, France, <sup>4</sup>Axiom Platform, MetaToul-MetaboHUB, National Infrastructure for Metabolomics and Fluxomics, Toulouse, France, <sup>5</sup>INRAE, Univ. Montpellier, LBE, 102 Avenue des étangs, F-11000 Narbonne, France and <sup>6</sup>INTHERES, Université de Toulouse, INRAE, ENVT, Toulouse, France

## From NMR spectra to metabolite quantifications

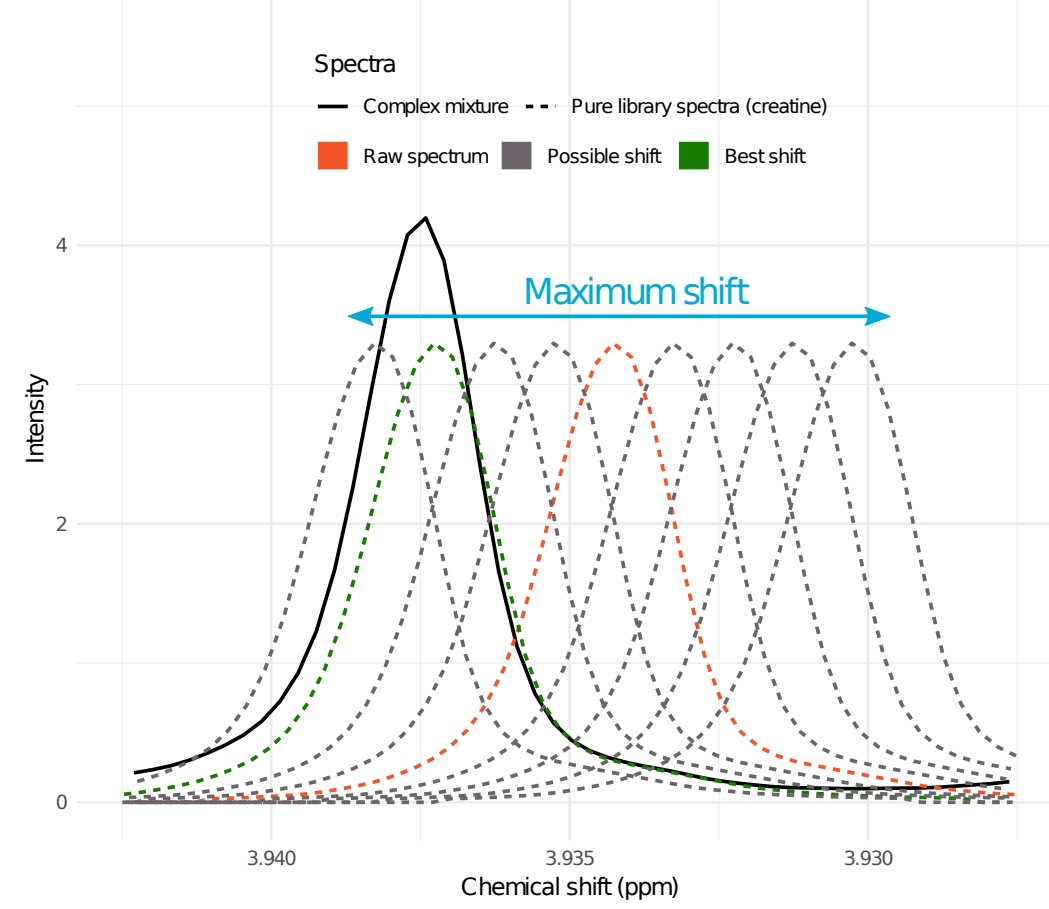


## Main features

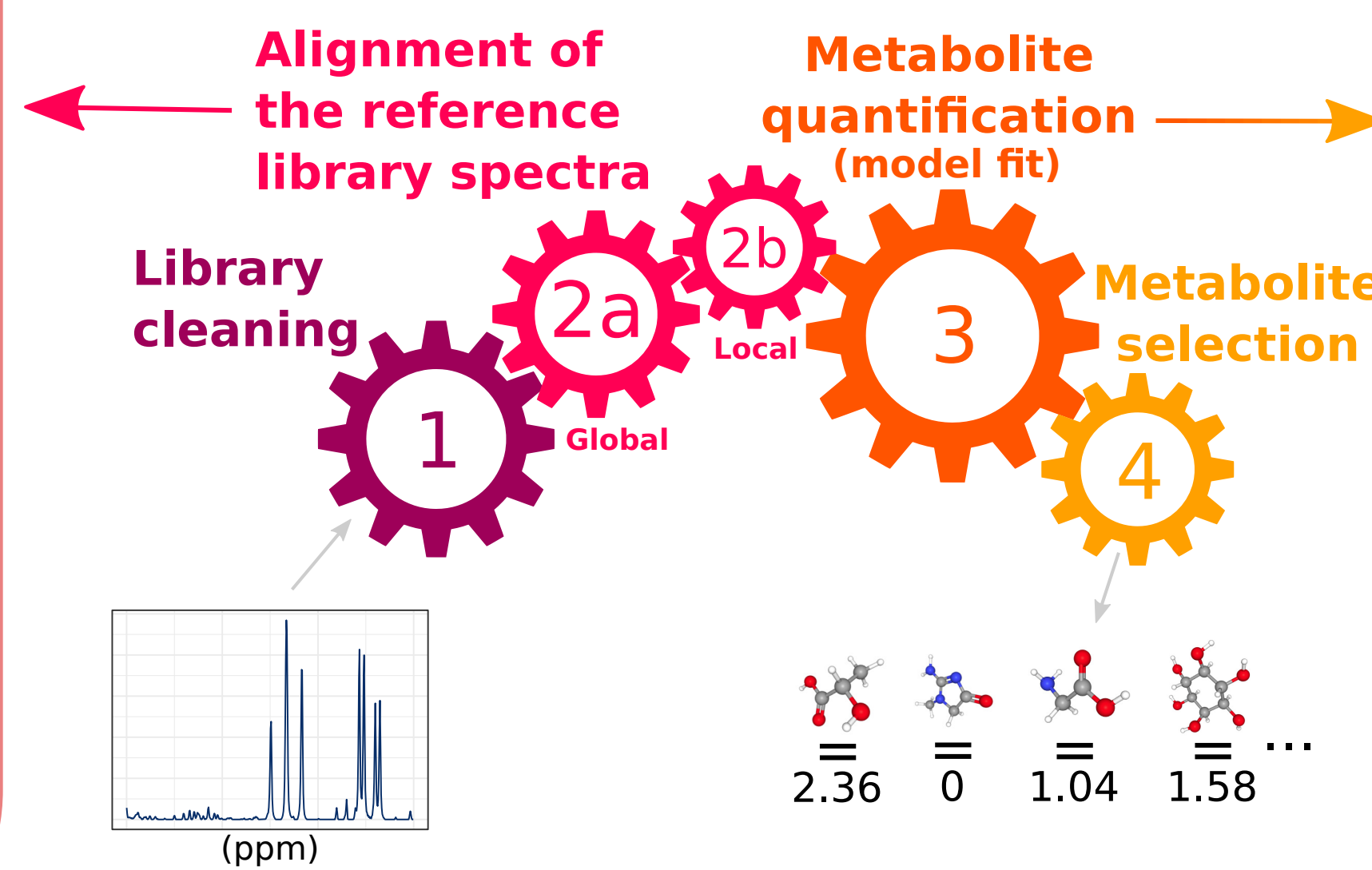
- Joint alignment of pure library spectra by using all complex mixtures together
- Joint quantification and identification of metabolites using multi-response model with a group-Lasso penalty

## Independent vs. joint alignment and quantification

**ASICS v1 [1]**  
**Independent**  
Optimization of the FFT cross-correlation [2] between all possible shifted pure spectra and complex mixture with a common **maximum shift** for all pure spectra



### Quantification workflow

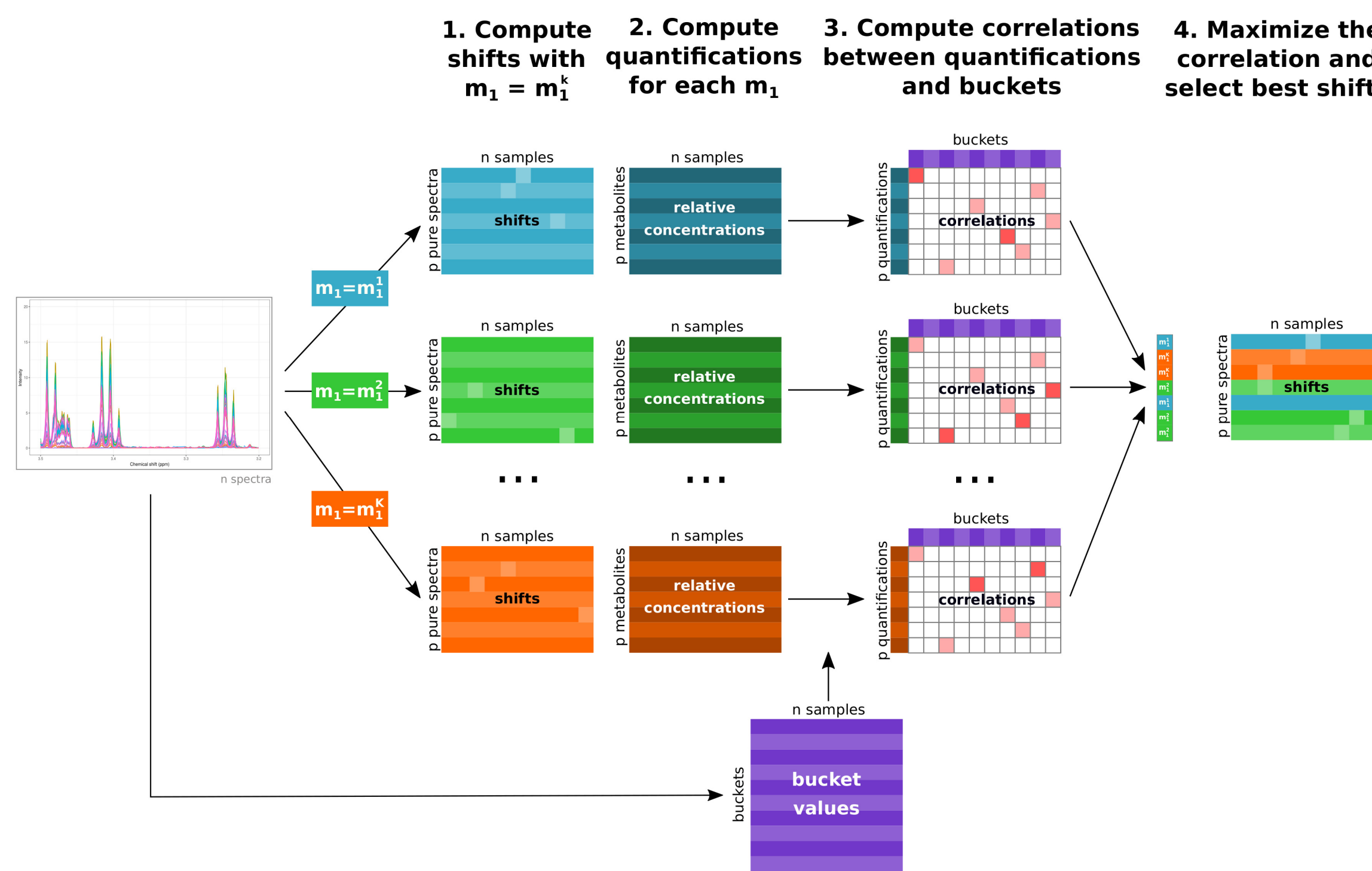


**ASICS v1 [1]**  
**Independent**  
Estimation of metabolite quantifications using linear combination of pure spectra followed by metabolite selection procedure based on FWER [3]

$$f(t) = \sum_{j=1}^p \beta_j g_j(t) + \epsilon(t) \quad \text{with } \beta_j \geq 0$$

Metabolite quantifications

**Joint**  
Optimization of the **maximum shift** for each pure spectrum using all complex mixtures



**Joint**  
Joint quantification of metabolites using a multi-response model [4] with a group-Lasso penalty [5]

$$\arg \min_{\beta \in \mathbb{R}^{p \times n}} \left\{ \frac{1}{2} \|F - rG\beta^T\|_F^2 + \lambda \sum_{j=1}^p \|\beta_j\|_2 \right\}, \quad \text{st } \beta_{ji} \geq 0$$

Metabolite quantifications of all complex mixtures

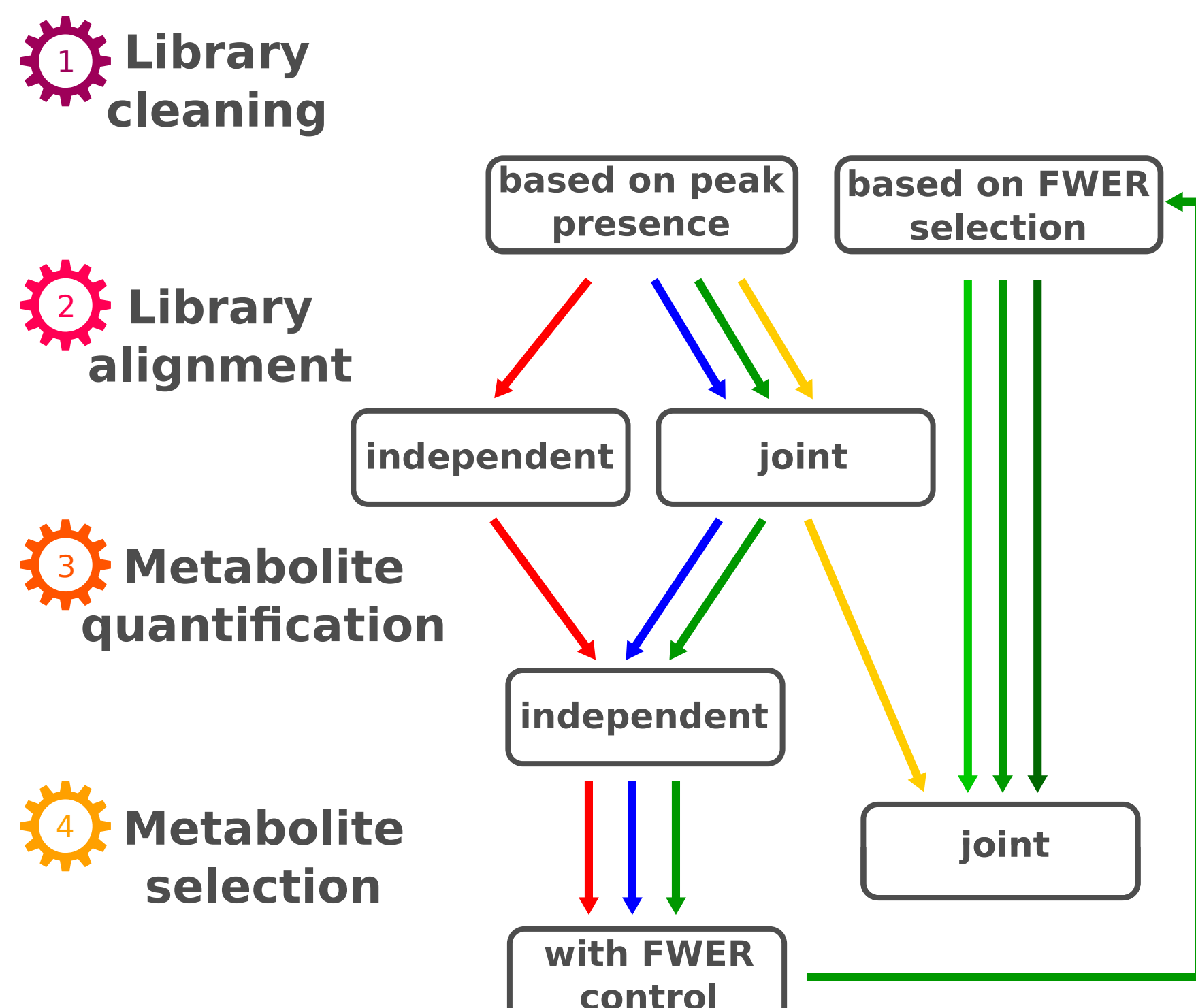
- [1] Lefort, G. et al. (2019). ASICS: an R package for a whole analysis workflow of 1D  $^1\text{H}$  NMR spectra. *Bioinformatics*, **35**, 4356–4363.
- [2] Wong, J. et al. (2005). Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets. *Analytical Chemistry*, **77**, 5655–5661.
- [3] Tardivel, P. et al. (2017). Représentation parcimonieuse et procédures de tests multiples : application à la métabolomique. *PhD thesis*, Université Toulouse 3 Paul Sabatier.
- [4] Friedman, J. H. et al. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**.
- [5] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society*, **68**, 49–67.

## Evaluation of the new algorithm

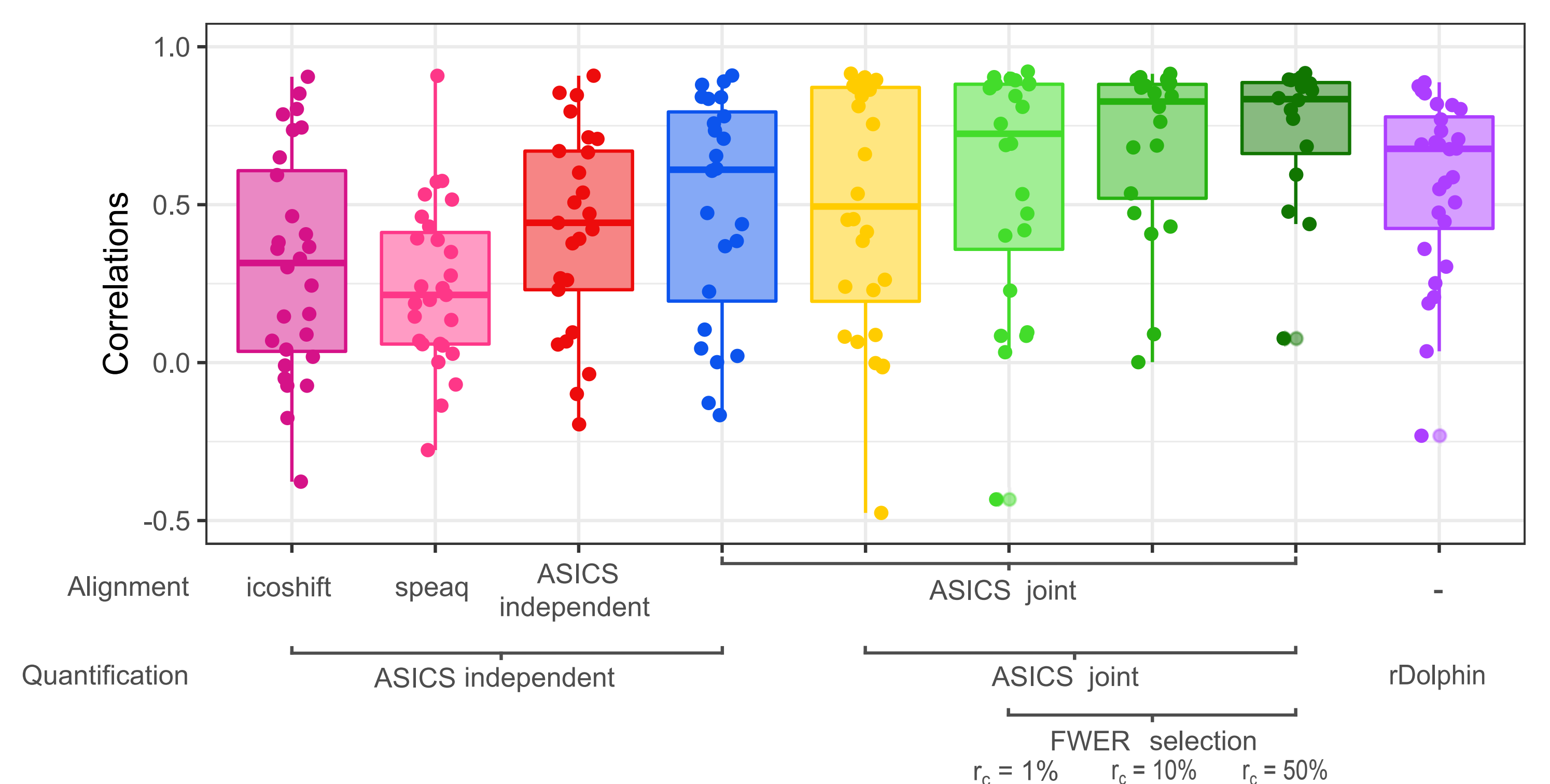
**Data** 97 samples of newborn piglet plasma on which NMR spectra and amino acid UPLC dosages were obtained

**Other tested methods** icoshift [1] and speaq [2] for alignment and rDolphin [3] for quantification

**ASICS** Different scenarios tested with ASICS are represented in the figure below



## Quantification/UPLC dosage correlations



**Figure 1.** Correlations between metabolite quantifications from NMR spectra and UPLC dosages obtained on same samples shown that **ASICS joint alignment** is better than other alignments and **ASICS joint quantification** is better than other methods if it was used with a FWER procedure for metabolite selection ( $r_c \geq 10\%$ ).

## Conclusions

- Better results with joined alignment and quantifications
- Available in R **ASICS** (<https://bioconductor.org/packages/ASICS/>)

## Acknowledgements



- [1] Savorani, F. et al. (2010). icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, **202**, 190–202.
- [2] Beirnaert, C. et al. (2018). speaq 2.0: a complete workflow for high-throughput 1D NMR spectra processing and quantification. *PLoS Computational Biology*, **14**, e1006018.
- [3] Canueto, D. et al. (2018). rDolphin: a GUI R package for proficient automatic profiling of 1D  $^1\text{H}$ -NMR spectra of study datasets. *Metabolomics*, **14**, 24.