



HAL
open science

Multiblock omics data fusion using the Consensus OPLS R package Upgrades Conclusions

Céline Bougel, van Du T. Tran, Julien Boccard, Marie Tremblay-Franco,
Florence Mehl

► To cite this version:

Céline Bougel, van Du T. Tran, Julien Boccard, Marie Tremblay-Franco, Florence Mehl. Multiblock omics data fusion using the Consensus OPLS R package Upgrades Conclusions. Réseau Francophone de Métabolomique et Fluxomique (RFMF), Jun 2024, Saint-Malo (France), France. hal-04623802

HAL Id: hal-04623802

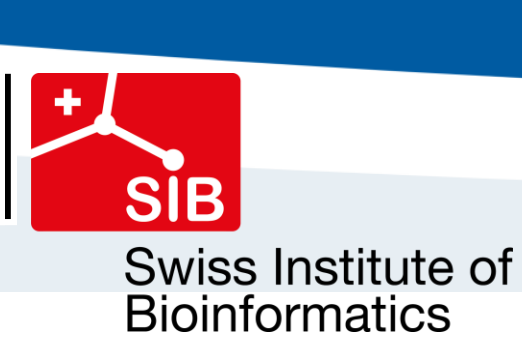
<https://hal.science/hal-04623802>

Submitted on 25 Jun 2024

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.

Multiblock omics data fusion using the Consensus OPLS R package



Céline Bougel^{1,2}, Van Du Tran³, Julien Bocard⁴, Marie Tremblay-Franco^{1,2}, Florence Mehl³

¹ Toxalim (Research Center in Food Toxicology), Toulouse University, INRAE, ENVT, INP-Purpan, UPS, Toulouse, France

² MetaboHUB-Metatoul, National Infrastructure of Metabolomics and Fluxomics, Toulouse, France

³ Vital-IT Group, SIB Swiss Institute for Bioinformatics, 1015 Lausanne, Switzerland

⁴ School of Pharmaceutical Sciences, University of Geneva, 1211, Geneva 4, Switzerland



16^{èmes} Journées Scientifiques du Réseau Francophone de Métabolomique et Fluxomique

4-6 juin 2024, Saint-Malo

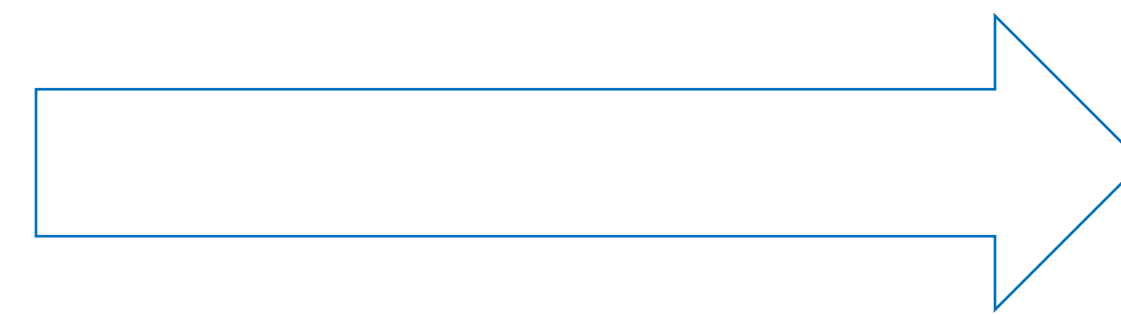
Introduction

Omics approaches have proven their value in providing a broad monitoring of biological systems. However, despite the wealth of data generated by modern analytical platforms, the analysis of a single dataset is still limited and insufficient to reveal the full biochemical complexity of biological samples. The fusion of information from several data sources constitutes therefore a relevant approach to assessing biochemical events more comprehensively. However, inherent problems encountered when analyzing

single tables are amplified with the generation of multi-block datasets. Finding the relationships between data layers of increasing complexity constitutes a challenging task. Here we propose an extension to the versatile methodology combining the strength of established data analysis strategies, multi-block approaches with Orthogonal Partial Least Squares Discriminant analysis (OPLS-DA) framework [1], to offer an efficient tool as an R package for the fusion of Omics data obtained from multiple sources.



MATLAB in-house scripts available at [GitHub Unige](#) with demo_data [2]



R package available in [CRAN](#) New features



Upgrades

Kernels non-linear

Multiple predictive components

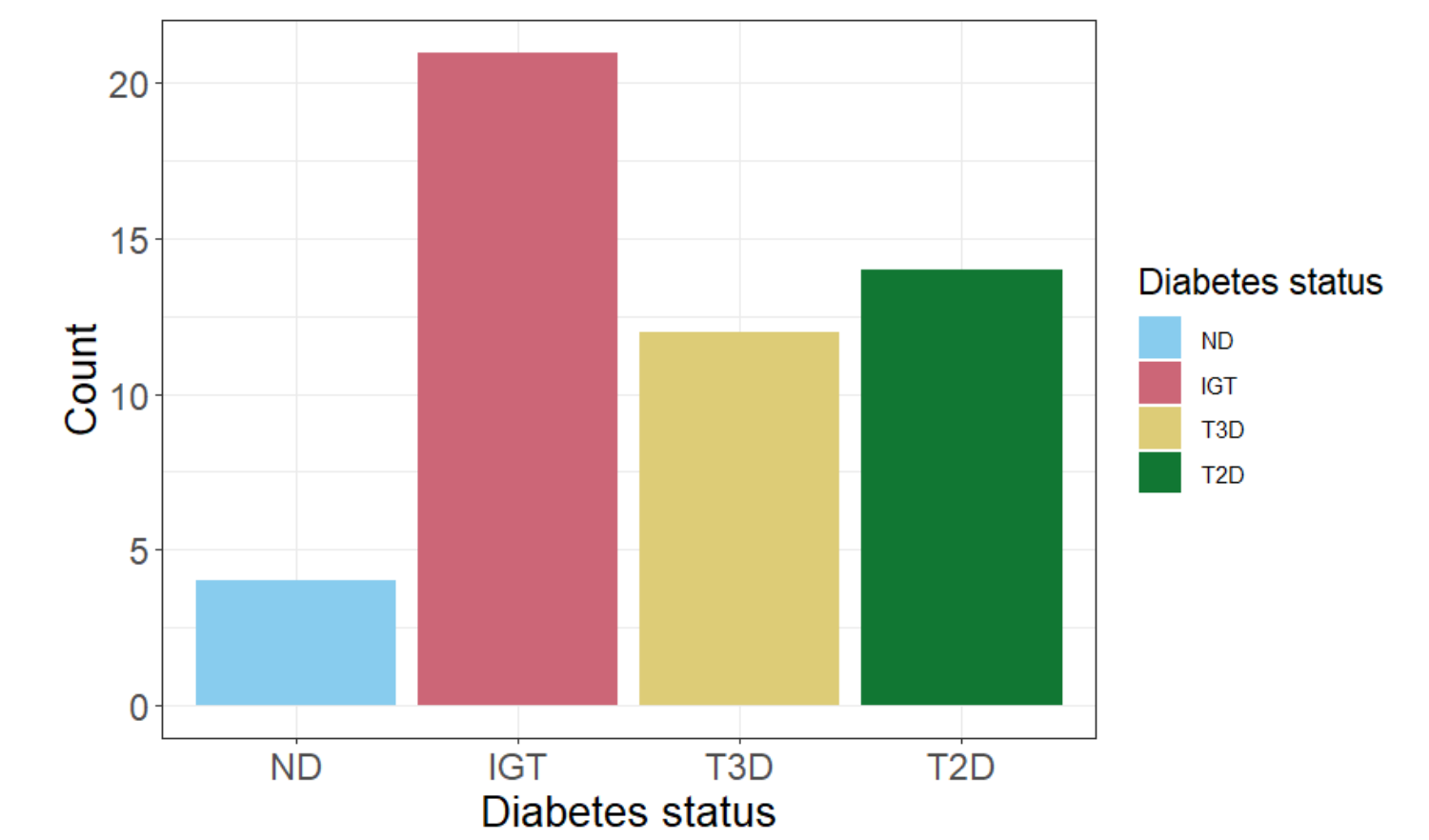
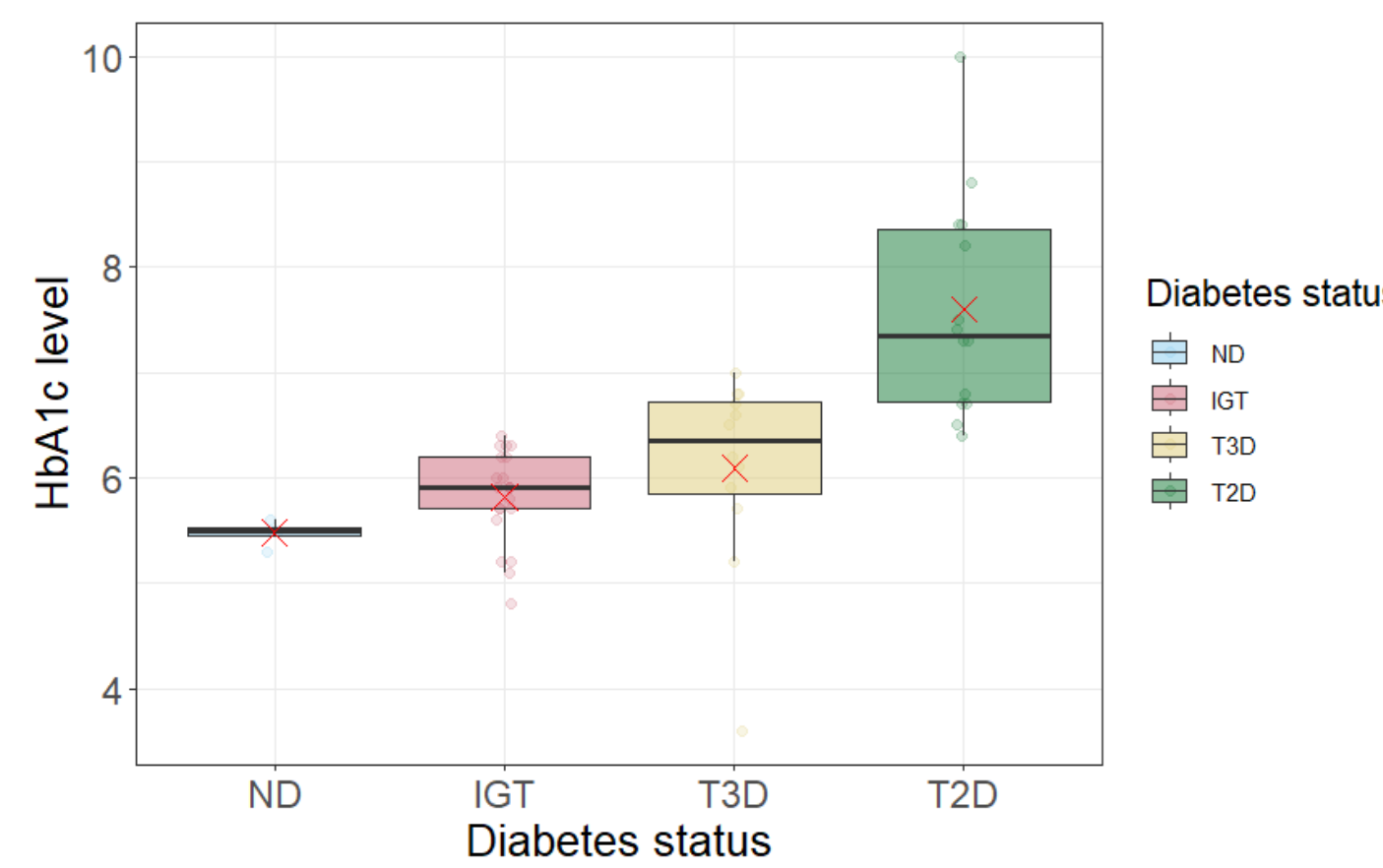
Variable Importance in Projection (VIP)

Prediction of new samples

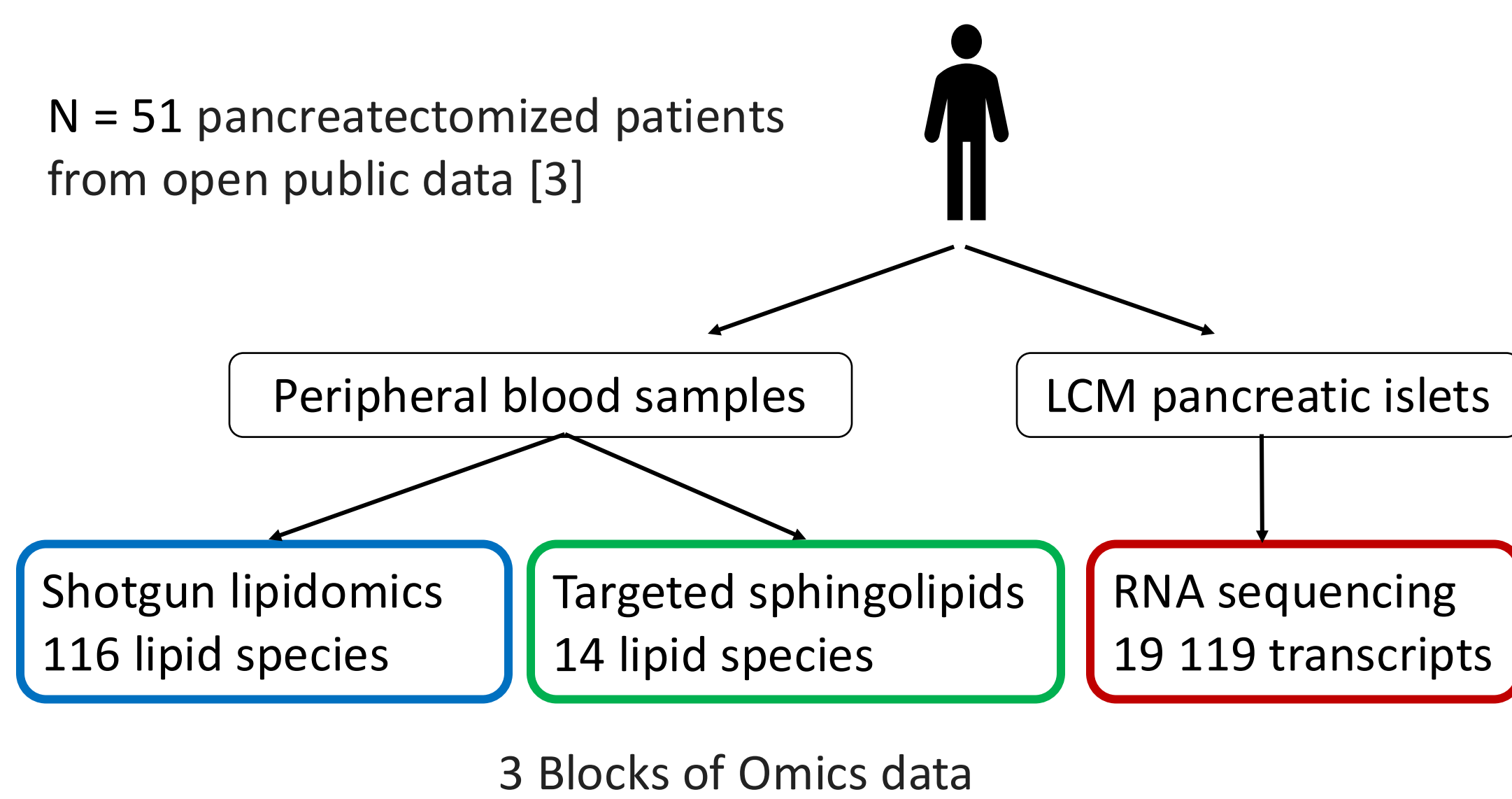
Parallel computing

Use case

4 groups of patients based on glycemic values at fasting and at the 2-hour timepoint of an oral glucose tolerance test, using the thresholds defined in the guidelines of the American Diabetes Association : Non-diabetic (ND); impaired glucose tolerance (IGT); type 2 diabetes (T2D); type 3c diabetes (T3D).



N = 51 pancreatized patients from open public data [3]



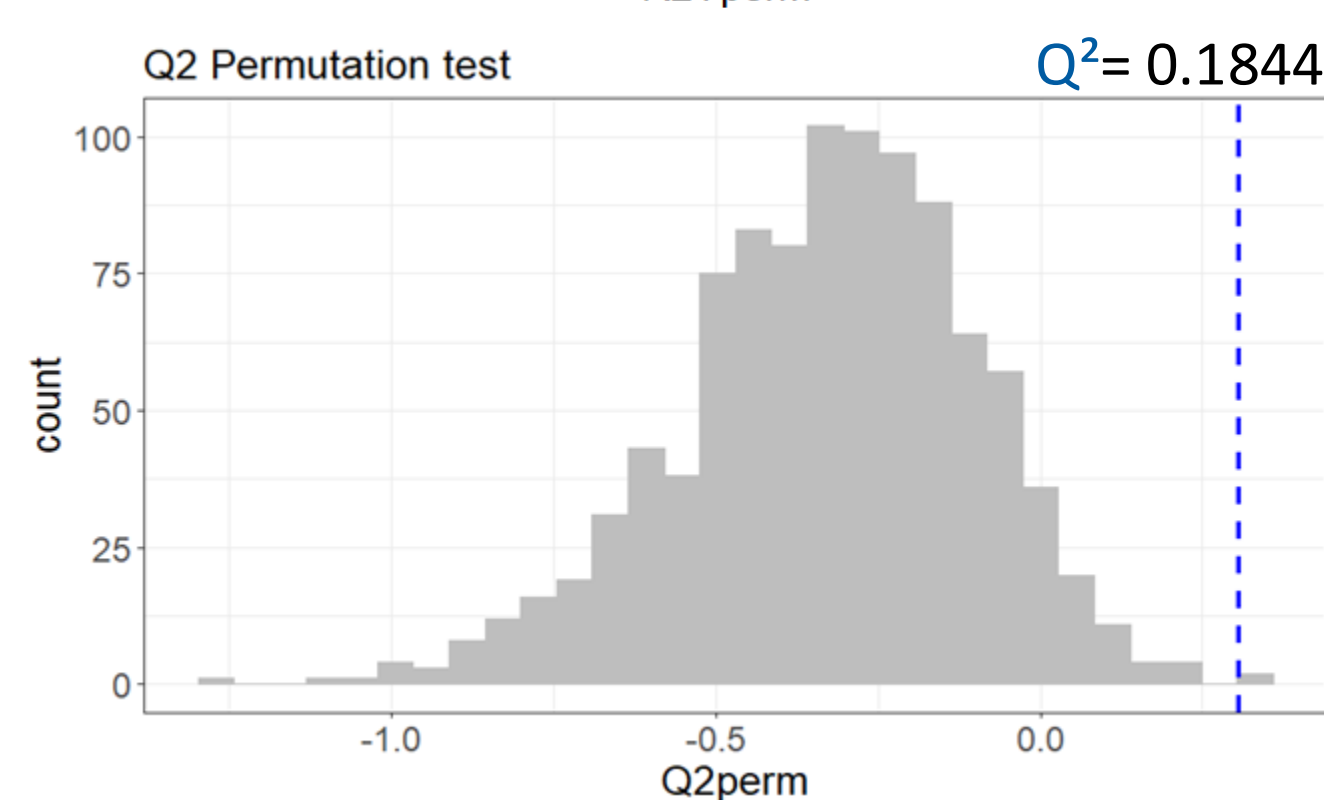
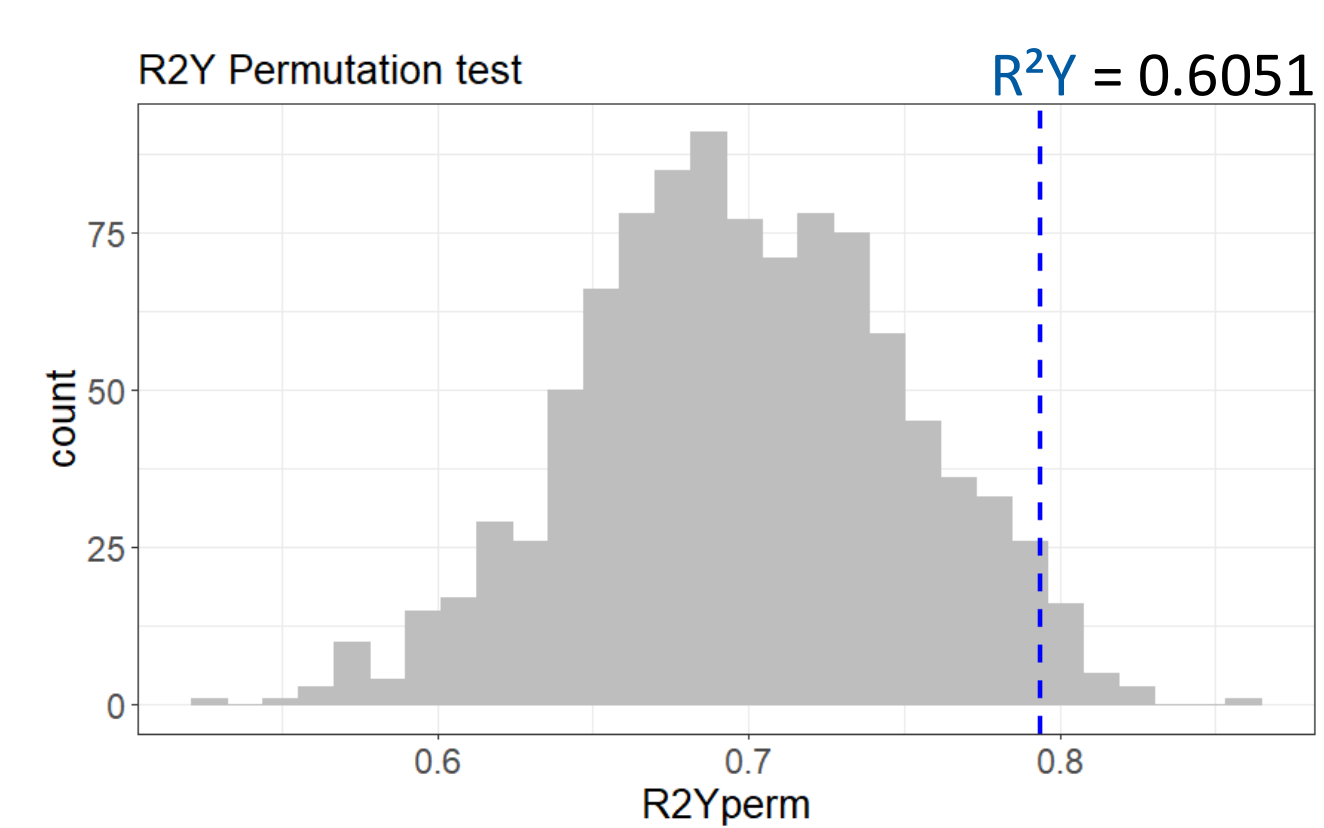
Results and discussion

Model computation

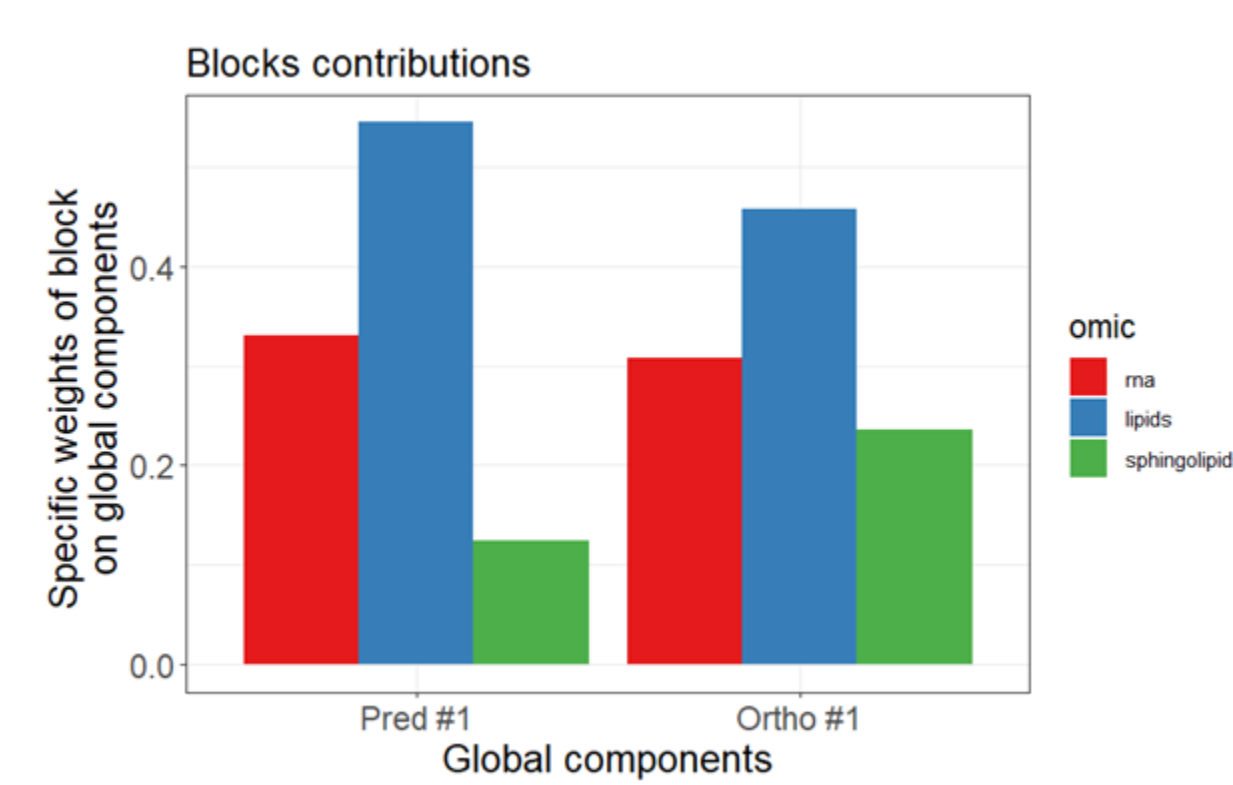
```
COPLS_results <- ConsensusOPLS(  
  data = my_data_with_3_blocks_of_omics_data,  
  Y = response_HbA1c_variable,  
  maxComp = 1, #one predictive component  
  maxOrtho = 1, #one orthogonal component  
  modelType = "reg", #regression model  
  nperm = 1000, #number of permutations  
  cvType = "nfold", #type of cross-validation  
  nfold = 51, #number of subjects = leave-one-out  
  kernelParams = list(type = "p",  
    params = c(order = 1)),  
  #linear polynomial kernel  
  #how many cores for parallelization  
  mc.cores = 1  
)
```

Quality assessment

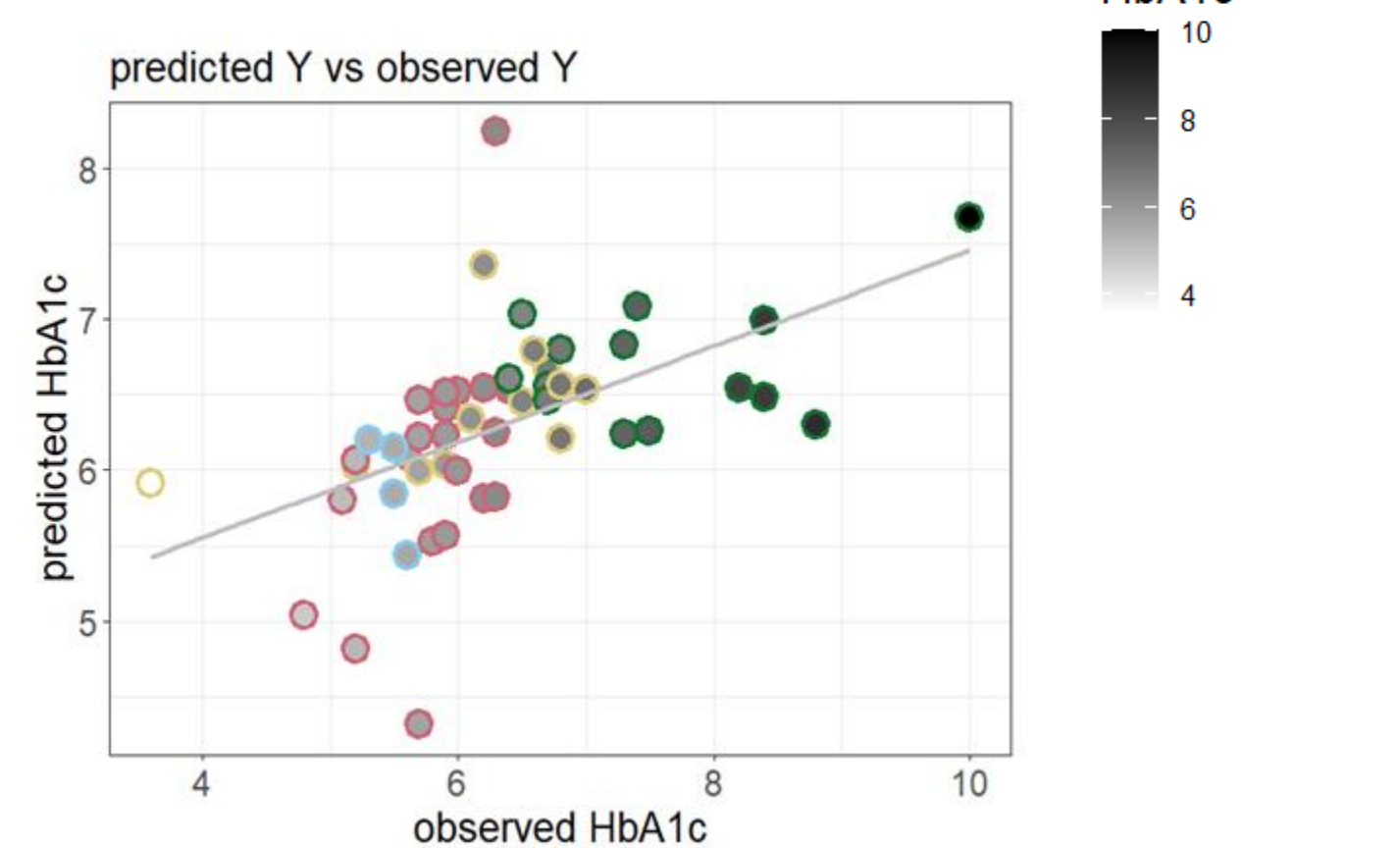
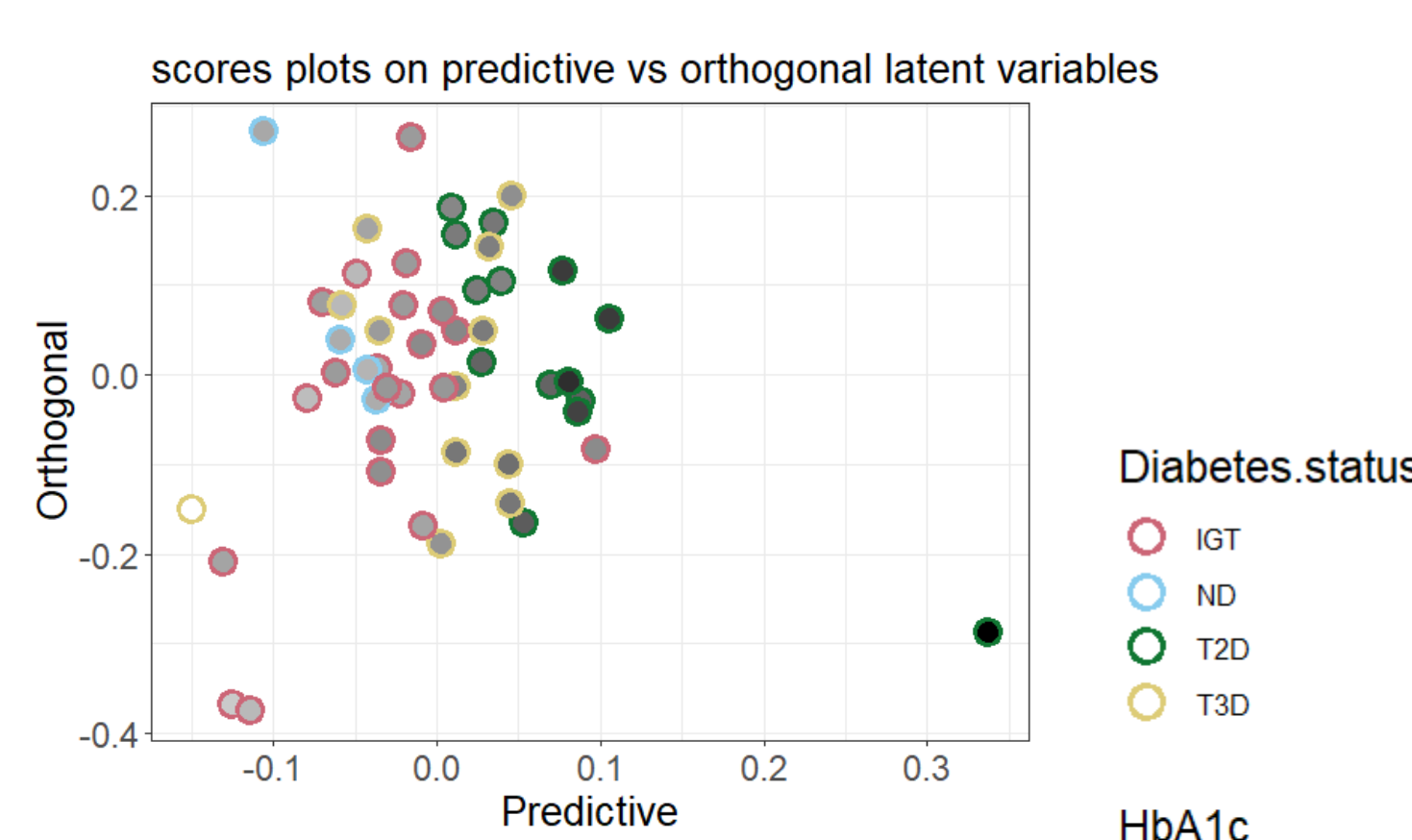
1000 models are computed with permuted Y values. The results of the optimal model are significantly different from the results of the permuted models.



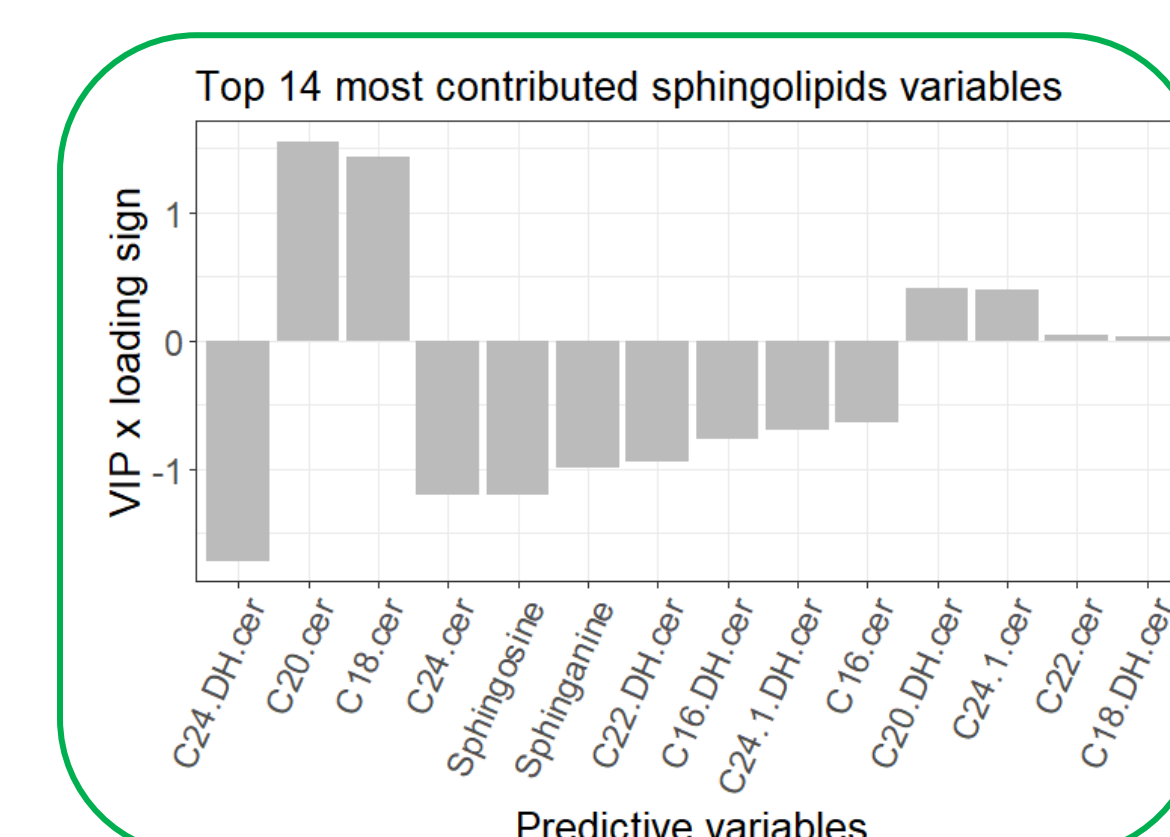
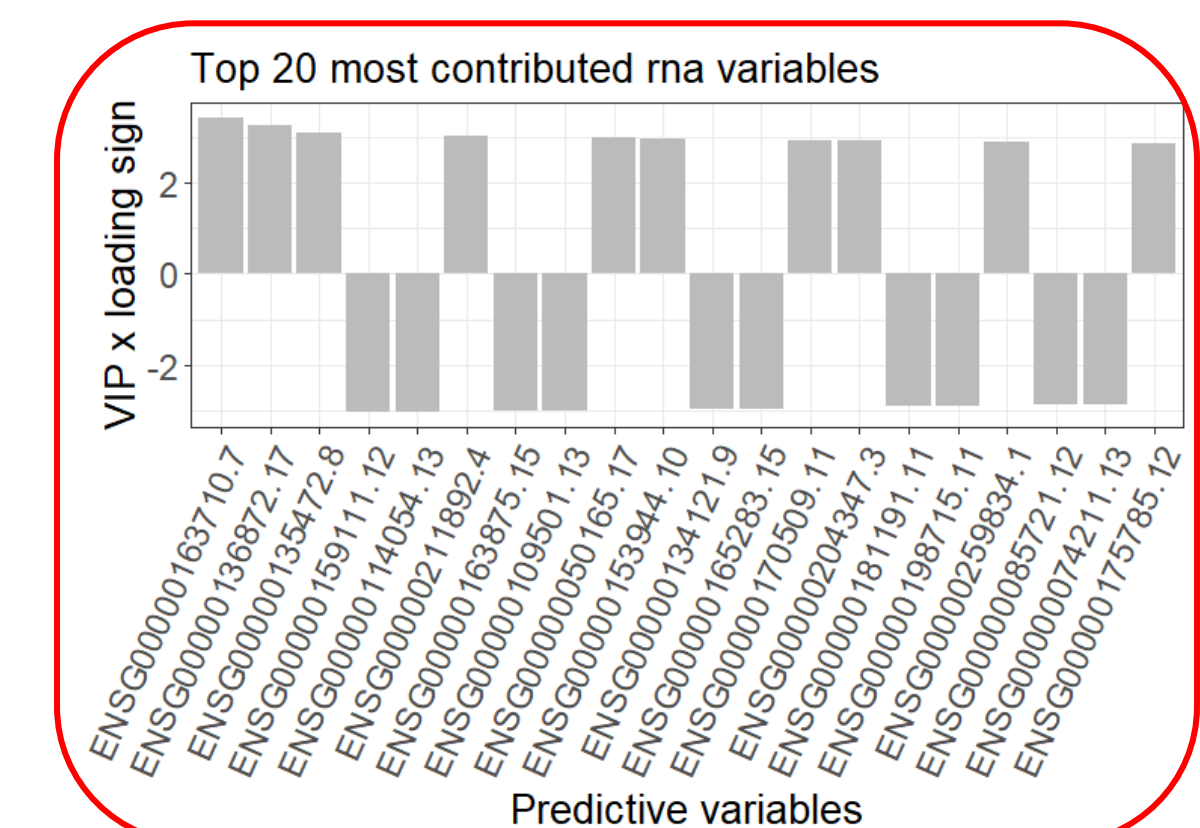
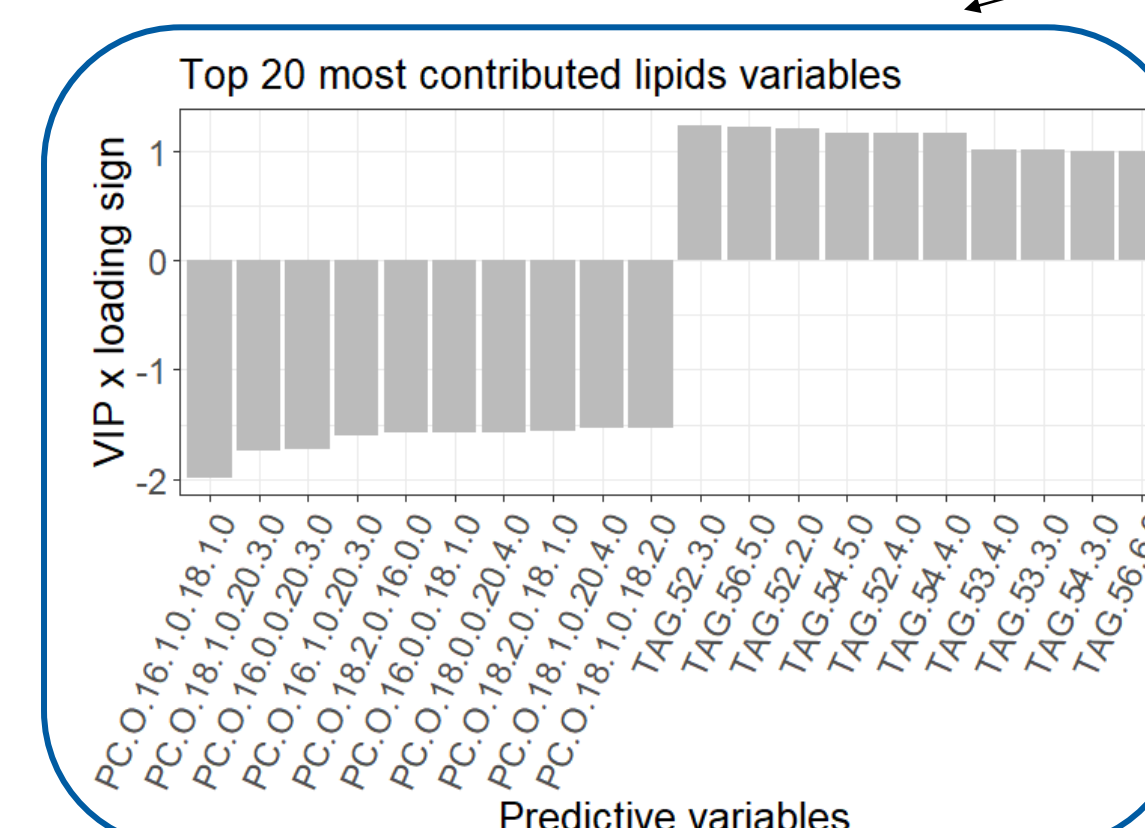
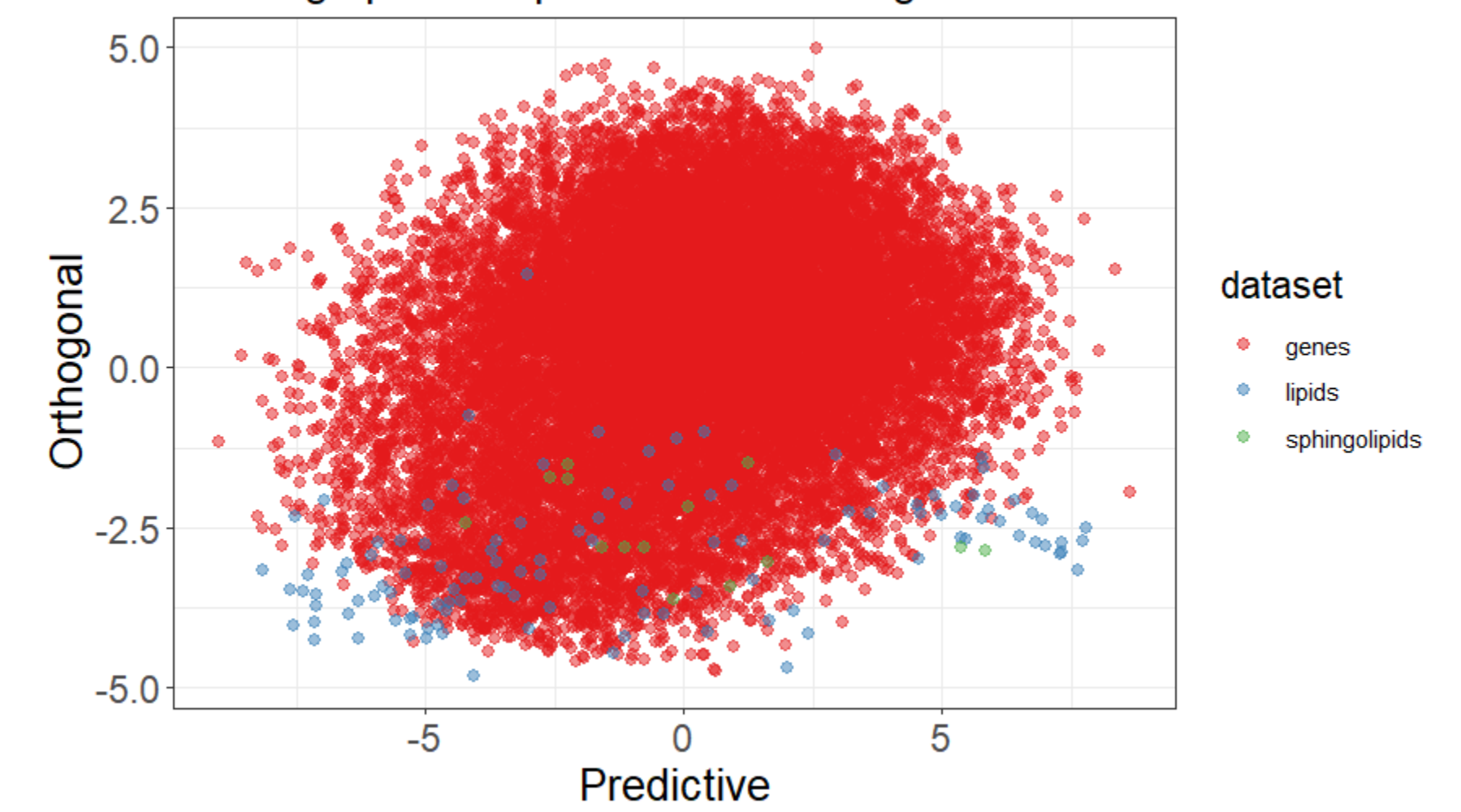
Regression Analysis (ConsensusOPLS-reg) of HbA1c levels



The 3 omics contribute to the modelization of HbA1c levels. The untargeted lipidomics dataset brings the largest part of the information.



Loadings plots on predictive vs orthogonal latent variables



Conclusions

The **ConsensusOPLS R package** proposes a **free, open-source, and easy-to-use software** for the Consensus OPLS method, which was demonstrated as a **relevant and widely applicable** method for the **horizontal integration of omics data**, with multiple functionality upgrades.

References

The **ConsensusOPLS R package** is available at <https://CRAN.R-project.org/package=ConsensusOPLS>.

[1] J. Bocard, D.N. Rutledge. 2013. "A Consensus Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) Strategy for Multiblock Omics Data Fusion." Analytica Chimica Acta 769 (March): 30–39. <https://doi.org/10.1016/j.aca.2013.01.022>.

[2] R.H. Shoemaker. 2006. "The NCI60 Human Tumour Cell Line Anticancer Drug Screen." Nature Reviews. Cancer 6 (10): 813–823. <https://doi.org/10.1038/nrc1951>.

[3] Wigger L, Barovic M, Brunner AD, et al. 2021. "Multi-omics profiling of living human pancreatic islet donors reveals heterogeneous beta cell trajectories towards type 2 diabetes." Nat Metab. 2021 Jul;3(7):1017-1031. PMID: 34183850. <https://doi.org/10.1038/s42255-021-00420-9>.

Acknowledgments

- This research was funded by the French National Infrastructure for metabolomics and fluxomics MetaboHUB ANR-11-INBS-0010.
- Vital-IT Group, SIB Swiss Institute for Bioinformatics.
- School of Pharmaceutical Sciences, University of Geneva.

Contacts

- marie.tremblay-franco@inrae.fr
- florence.mehl@sib.swiss