



Using machine-learning on metabolomics data to predict complex phenotypes

Sylvain Prigent, Yves Gibon

► To cite this version:

Sylvain Prigent, Yves Gibon. Using machine-learning on metabolomics data to predict complex phenotypes. JOBIM2022, Jul 2022, Rennes, France. hal-03760581

HAL Id: hal-03760581

<https://hal.inrae.fr/hal-03760581>

Submitted on 25 Aug 2022

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.

Using machine-learning on metabolomics data to predict complex phenotypes

Sylvain PRIGENT^{1,2} and Yves GIBON^{1,2}

¹ Univ. Bordeaux, INRAE, UMR1332 BFP, 33882 Villenave d'Ornon, France

² Bordeaux Metabolome, MetaboHUB, PHENOME-EMPHASIS, 33140 Villenave d'Ornon, France

Corresponding author: `sylvain.prigent@inrae.fr`

The metabolome is often seen as the ultimate cellular phenotype, resulting from all biochemical processes taking place in an organism. Using a robotized high throughput platform to prepare samples and modern mass spectrometers coupled with chromatography, it is nowadays possible to obtain both targeted and untargeted metabolomic data on hundreds of samples in a few days. Those experiments lead to thousands of metabolic features representing a precise description of the cell state at a given time. This state is a good proxy of the past, the present and the future of a given tissue, combining the result of the previous developmental stages, the current status and the bricks enabling the future development of the cells. Moreover, by applying different types of stresses to an organism, it is possible to increase the diversity of the metabolome, hence displaying broader characteristics.

By combining metabolomic and phenotypic data and using machine-learning approaches, we were able to accurately predict several complex phenotypes such as relative growth rate [1], grafting success [2], plant elevation [3] or resistance to pathogens [4] on various plant matrices. Besides, enabling prediction of those complex phenotypes, those techniques permit the discovery of biomarkers that give insights into mechanisms controlling the phenotypes.

Here we will present the different cases studied recently by our laboratory and metabolomic platform. The accuracies of the predictions and the physiological knowledge acquired thanks to those models will be detailed. Perspectives about the integration of knowledge acquired through the use of those machine-learning models into more mechanistic models will also be discussed.

Acknowledgements

The authors want to thank all partners involved in the acquisition of the data, the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing computing resources as well as the GenOuest bioinformatics core facility (<https://www.genouest.org>) for providing the computing infrastructure.

References

- [1] Léa Roch, Sylvain Prigent, Holger Klose, Coffi-Belmys Cakpo, Bertrand Beauvoit, Catherine Deborde, Laetitia Fouillen, Pierre van Delft, Daniel Jacob, Björn Usadel, et al. Biomass composition explains fruit relative growth rate and discriminates climacteric from non-climacteric species. *Journal of Experimental Botany*, 71(19):5823–5836, 2020.
- [2] Grégoire Loupfit, Josep Valls Fonayet, Sylvain Prigent, Duyen Prodhomme, Anne-Sophie Spilmont, Ghislaine Hilbert, Céline Franc, Gilles De Revel, Tristan Richard, Nathalie Ollat, et al. Identifying early metabolite markers of successful graft union formation in grapevine. *Horticulture Research*, 2022.
- [3] Thomas Dussarrat, Sylvain Prigent, Claudio Latorre, Stéphane Bernillon, Amélie Flandin, Francisca P Díaz, Cédric Cassan, Pierre Van Delft, Daniel Jacob, Kranthi Varala, et al. Predictive metabolomics of multiple atacama plant species unveils a core set of generic metabolites for extreme climate resilience. *New Phytologist*, 234(5):1614–1628, 2022.
- [4] Estrella Luna, Amélie Flandin, Cédric Cassan, Sylvain Prigent, Chloé Chevanne, Camélia Feyrouse Kadiri, Yves Gibon, and Pierre Pétriacq. Metabolomics to exploit the primed immune system of tomato fruit. *Metabolites*, 10(3):96, 2020.