



HAL
open science

Tutti frutti: Metabolomics Meets Machine Learning for Juicy Discoveries

Sylvain Prigent, Millena C. Barros Santos, Giovanni Melandri, Georges Randriafanome- Zantsoa-Radohery, Cédric Cassan, Yves Gibon, Pierre Pétriacq

► To cite this version:

Sylvain Prigent, Millena C. Barros Santos, Giovanni Melandri, Georges Randriafanome- Zantsoa-Radohery, Cédric Cassan, et al.. Tutti frutti: Metabolomics Meets Machine Learning for Juicy Discoveries. JOBIM 2024 - Journées ouvertes en Biologie, Informatique et Mathématiques, Jun 2024, Toulouse, France. hal-04677398

HAL Id: hal-04677398

<https://hal.science/hal-04677398v1>

Submitted on 26 Aug 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.

Équipe META

UMR 1332 Biologie du fruit et pathologie, INRAE Bordeaux

Sylvain Prigent^{1,2}, Millena Barros-Santos^{1,2}, Giovanni Melandri^{1,3}, Georges Randriafanomezantsoa-Radohery¹, Cédric Cassan^{1,2}, Yves Gibon^{1,2} and Pierre Pétriacq^{1,2}

sylvain.prigent@inrae.fr

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

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

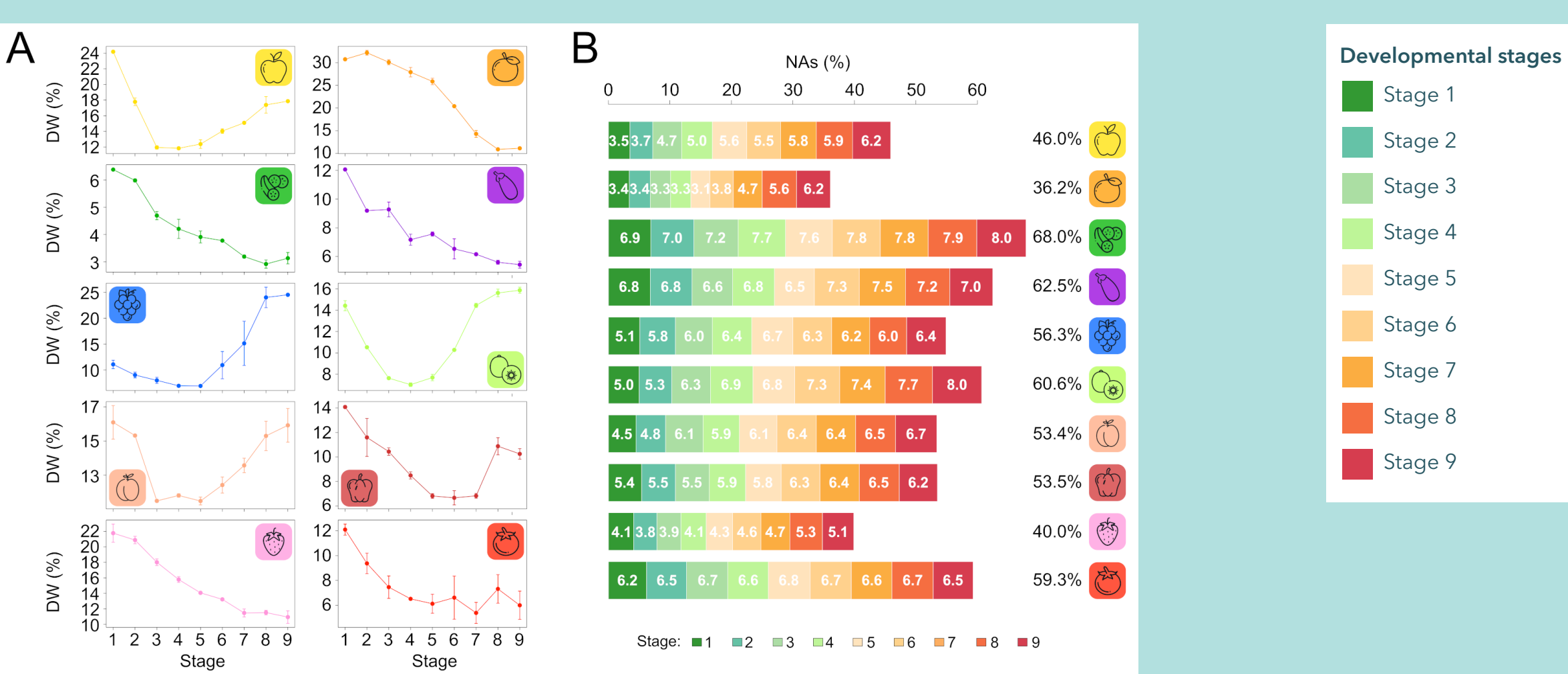
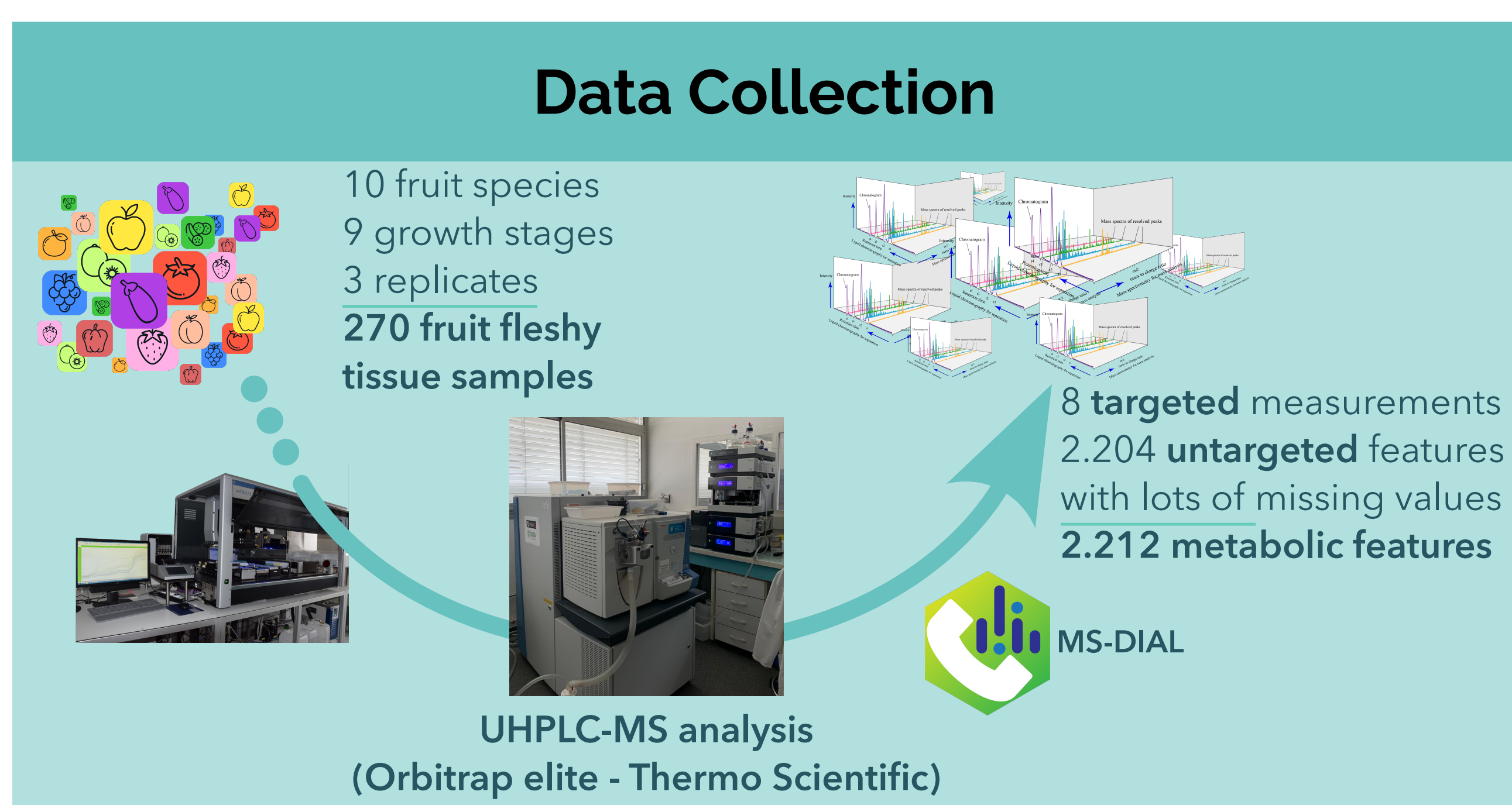
³ School of Plant Sciences, University of Arizona, Tucson, AZ, 85721, USA

Tutti frutti: Metabolomics Meets Machine Learning for Juicy Discoveries

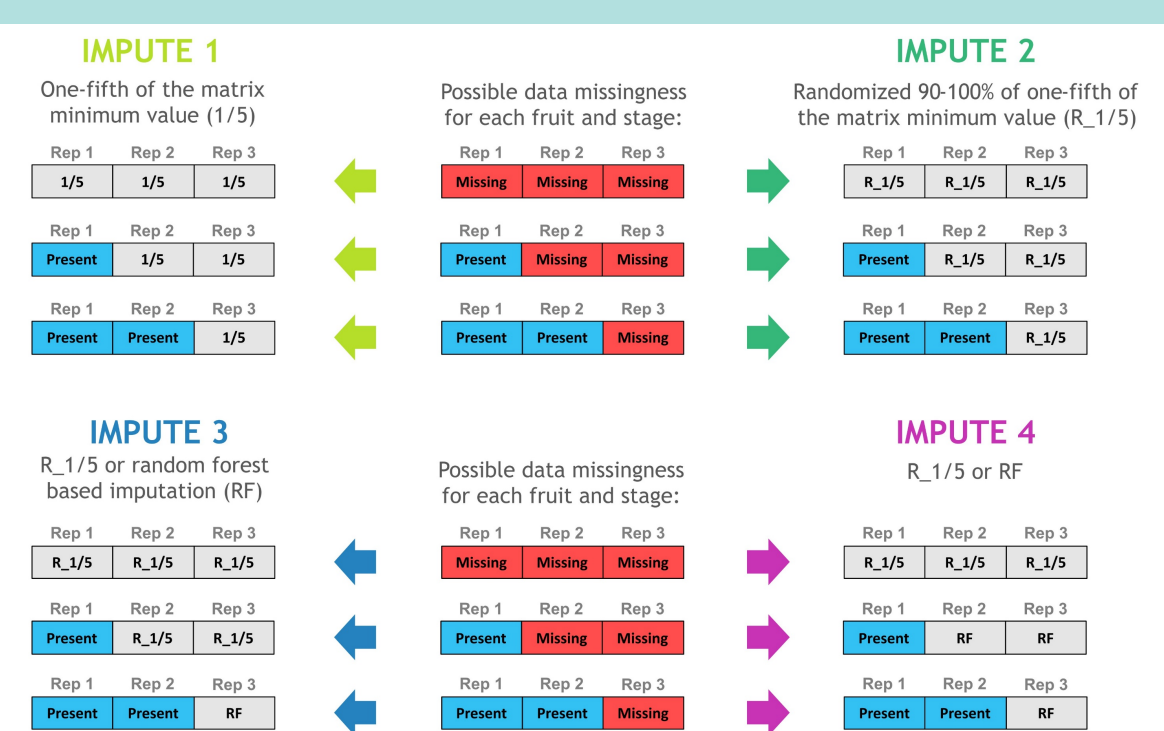
A journey into machine learning and discovery of metabolic biomarkers

The metabolome is often seen as the **ultimate cellular phenotype**:
- resulting from **all biochemical processes** that took place in an organism
- a screenshot of the **present status** of the metabolism
- bricks enabling the **future development** of the cells

The **Bordeaux Metabolome** facility enable the acquisition of metabolomic data on hundreds of samples:
- a **robotized high throughput platform** to prepare samples and **targeted measurements** using microplates
- modern mass spectrometers coupled with chromatography for **untargeted metabolomics**
- metabolomics enables **multi-species studies**
- such studies are key to predict global phenotypes and identify **generic biomarkers of integrative biological processes**

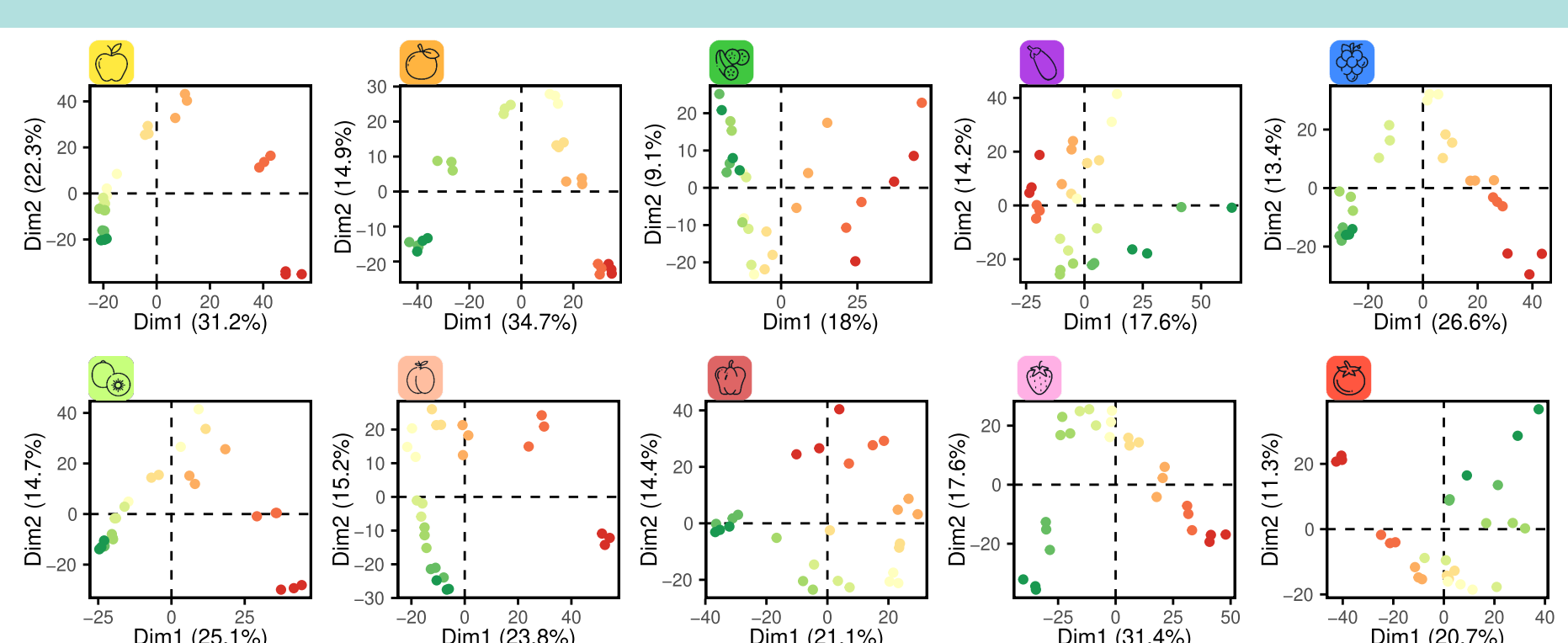


Overview of fruit tissue characteristics and metabolomics data A. Dry weight (DW) percent (n = 3, +/- SD) during fruit development (stage 1 to 9) B. Total and stage-dependent percentage of missing metabolite values

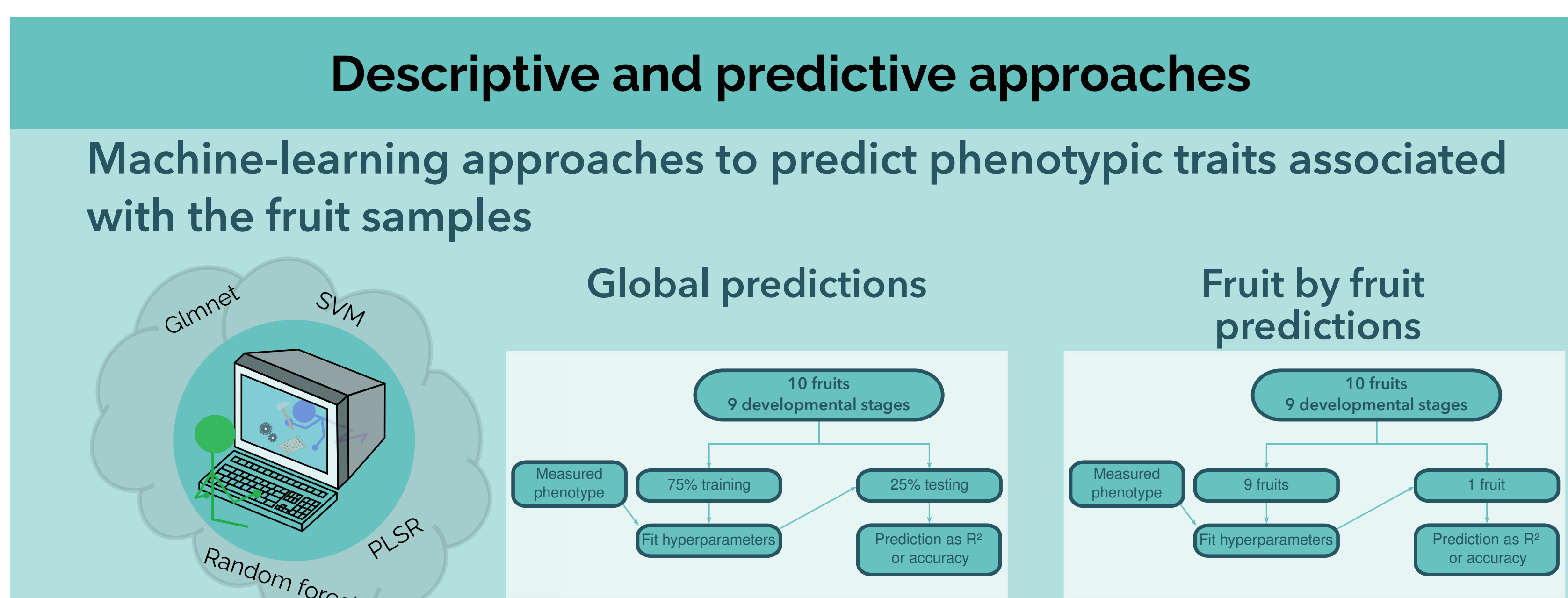


Imputation 1 and 2: missing values are variables not present in the fruit sample

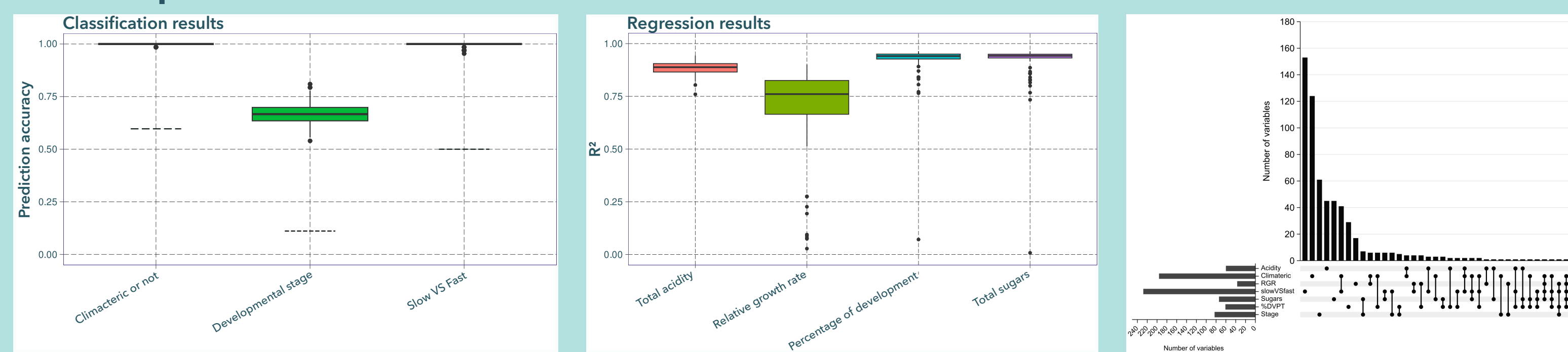
Imputation 3 and 4: if a metabolic feature was identified in at least one replicate, it should be present in the others



Principal component analysis score plots of the ten fruit species based on metabolic features after imputation with strategy 3. Colours correspond to developmental stages while the percentage of variation explained by the first two principal components is reported in each score plot



Global predictions



Results of the predictive approach on the global predictions using Elastic-net or LASSO classification and regression models Box-plots represent the distribution of 100 repeats. For classification, dashed line correspond to the No Information Rate for each phenotype. The upset-plot represents the biomarkers shared for the different predictions, based on their occurrence in the Elastic-net and LASSO models

Fruit by fruit predictions

Model	predicted fruit	predicted phenotype - Metric			
		Imp_3_DW	Imp_4_DW	Imp_3_DW_variables30percent	Imp_4_DW_variables30percent
SVM	Apple	0.80	0.80	0.80	0.80
	Clementine	0.80	0.80	0.80	0.80
	Cucumber	0.80	0.80	0.80	0.80
	Eggplant	0.80	0.80	0.80	0.80
	Grape	0.80	0.80	0.80	0.80
	Kiwi	0.80	0.80	0.80	0.80
	Peach	0.80	0.80	0.80	0.80
	Pepper	0.80	0.80	0.80	0.80
	Strawberry	0.80	0.80	0.80	0.80
	Tomato	0.80	0.80	0.80	0.80
PLS	Apple	0.80	0.80	0.80	0.80
	Clementine	0.80	0.80	0.80	0.80
	Cucumber	0.80	0.80	0.80	0.80
	Eggplant	0.80	0.80	0.80	0.80
	Grape	0.80	0.80	0.80	0.80
	Kiwi	0.80	0.80	0.80	0.80
	Peach	0.80	0.80	0.80	0.80
	Pepper	0.80	0.80	0.80	0.80
	Strawberry	0.80	0.80	0.80	0.80
	Tomato	0.80	0.80	0.80	0.80

Results of the fruit by fruit predictive approach using different machine learning models and datasets Results are displayed as average R² or accuracy over 100 repeats.

Conclusion and perspectives

- Predictive metabolomics enabled the accurate prediction of fruit growth, physiology, and quality traits.
- Identification of biomarkers linked to the predicted traits has been performed
- Need to put the biomarkers in context to identify metabolic pathways in which the markers are present
- Transcriptomics and proteomics data on the same samples are under analysis (check the presentation of Chloé Beaumont on Friday morning!)
- Results provide insights into molecular targets for future breeding efforts to improve fruit traits

