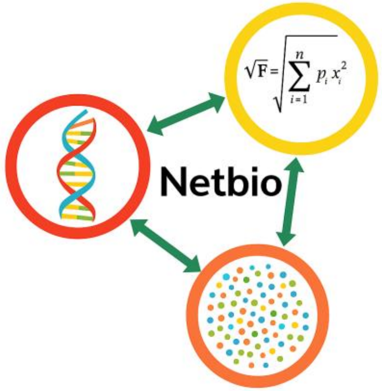
Exploration de la valeur interprétative des modèles multi-omiques pour la prédiction phénotypique chez le peuplier noir

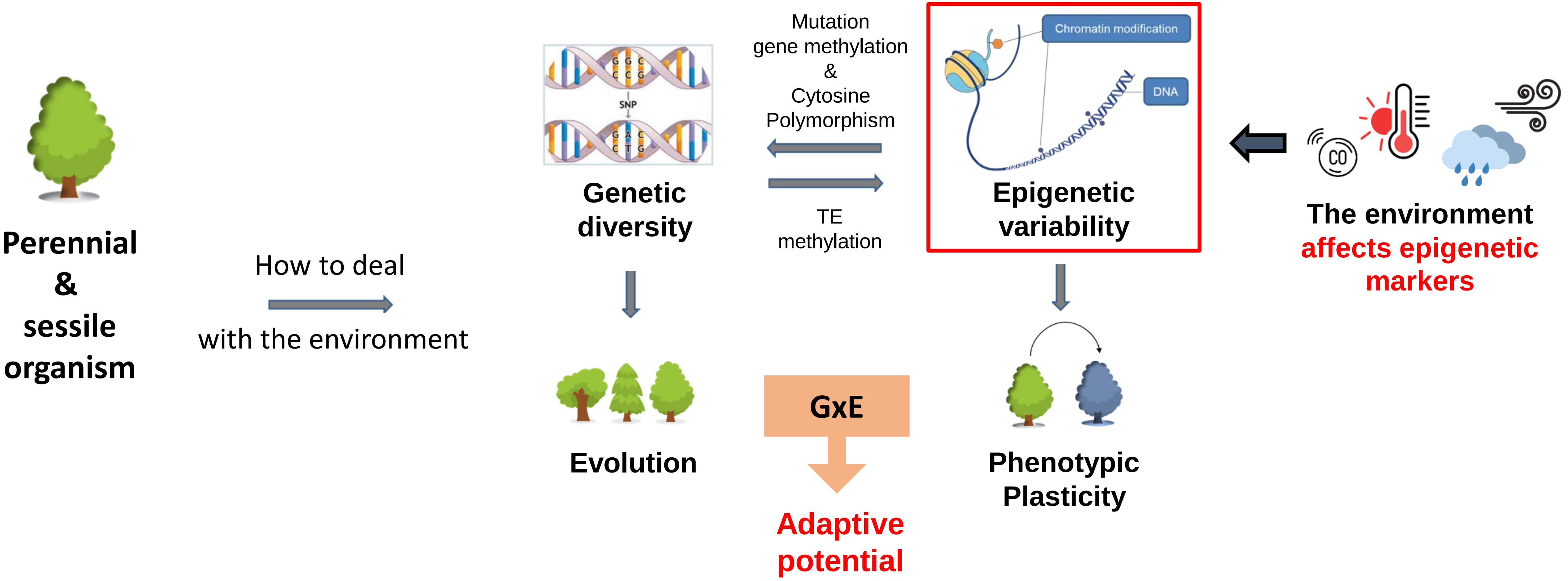
Alexandre DUPLAN

NETBIO
Orléans 2025



Pr. **Stéphane MAURY** (P2E)
Dr. **Leopoldo SANCHEZ** (BioForA)
Dr. **Harold DURUFLÉ** (BioForA)

Mechanism of adaptation to the environment for a fixed organism



Perennial
&
sessile
organism

How to deal
with the environment

Genetic diversity

Mutation
gene methylation
&
Cytosine
Polymorphism

TE
methylation

Epigenetic variability

The environment
affects epigenetic markers

Evolution

GxE

Adaptive potential

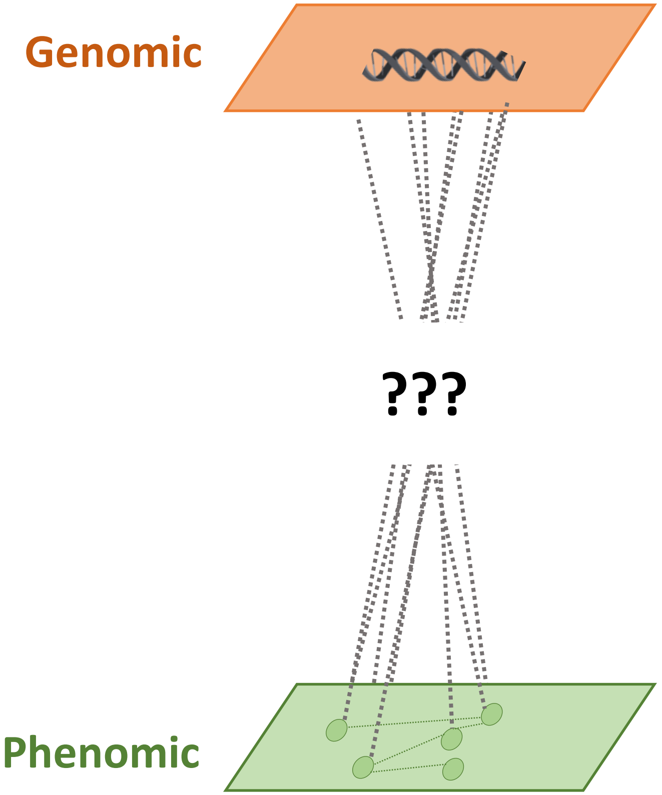
Phenotypic Plasticity

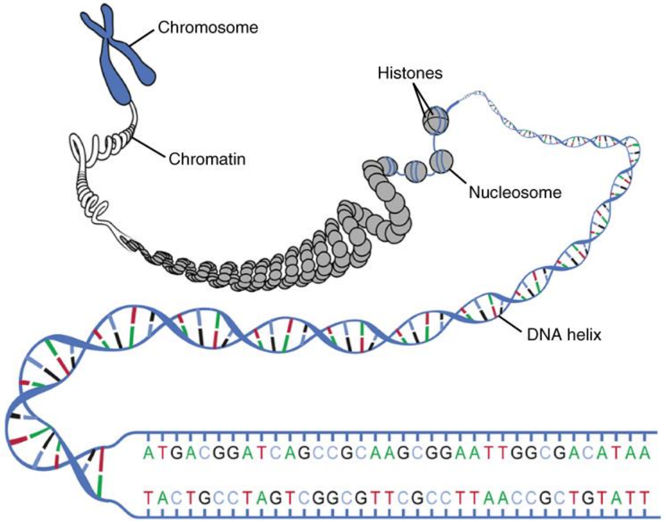
Understanding **how epigenetic variability** plays a role in **ecological memory** and **local adaptation**

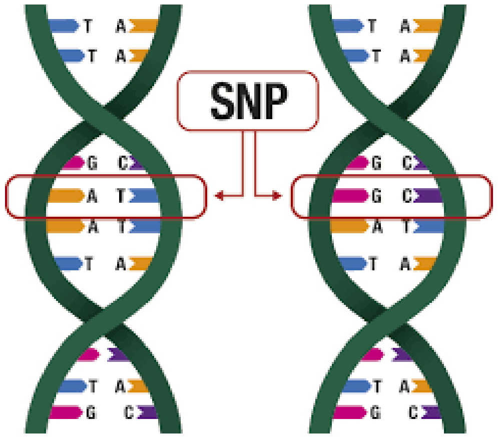
→ Integration of multi-omics data to investigate biological processes involved in this mecanisms

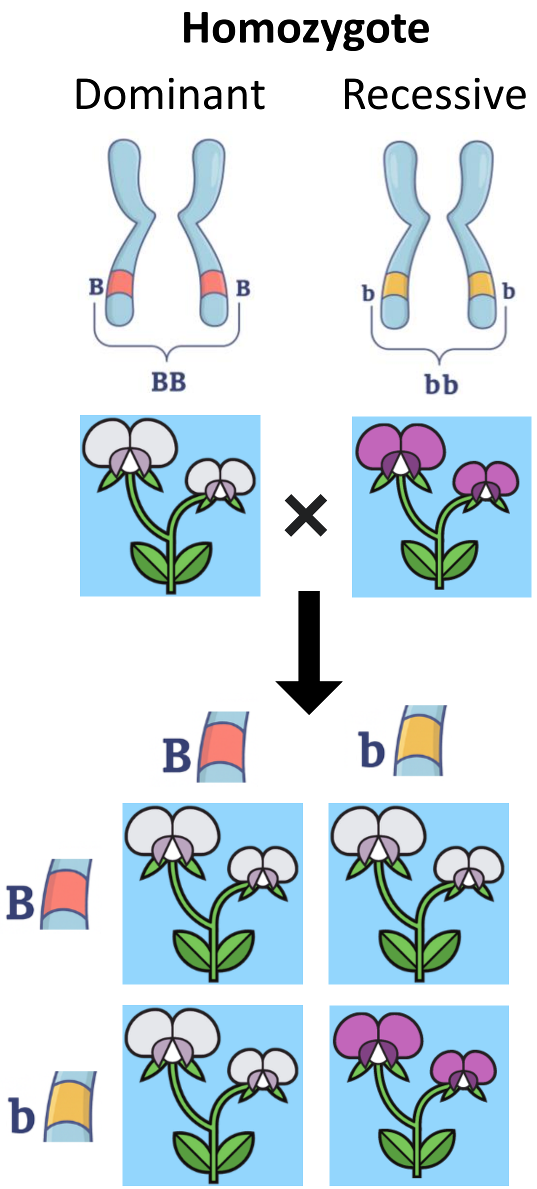
→ What is multi-omics?

The molecular carrier of biological information: the genome





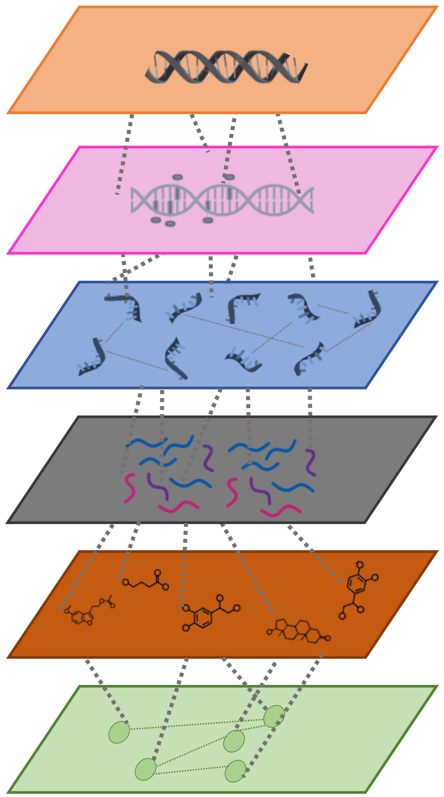




→ **Genetic diversity can affect phenotypic trait**

→ **But what happens to complex phenotypic traits ?**

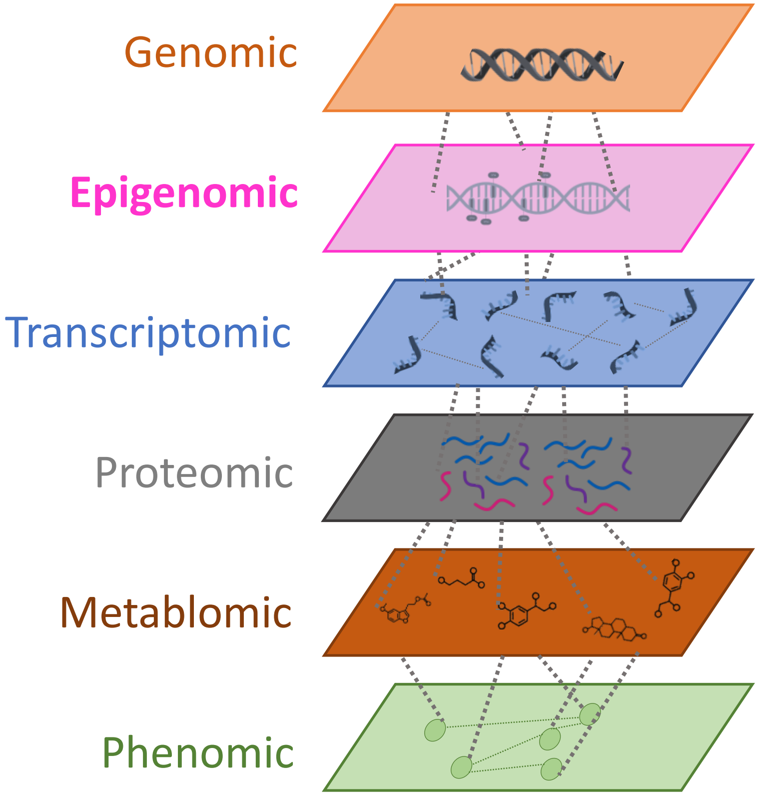
What is an omics data?

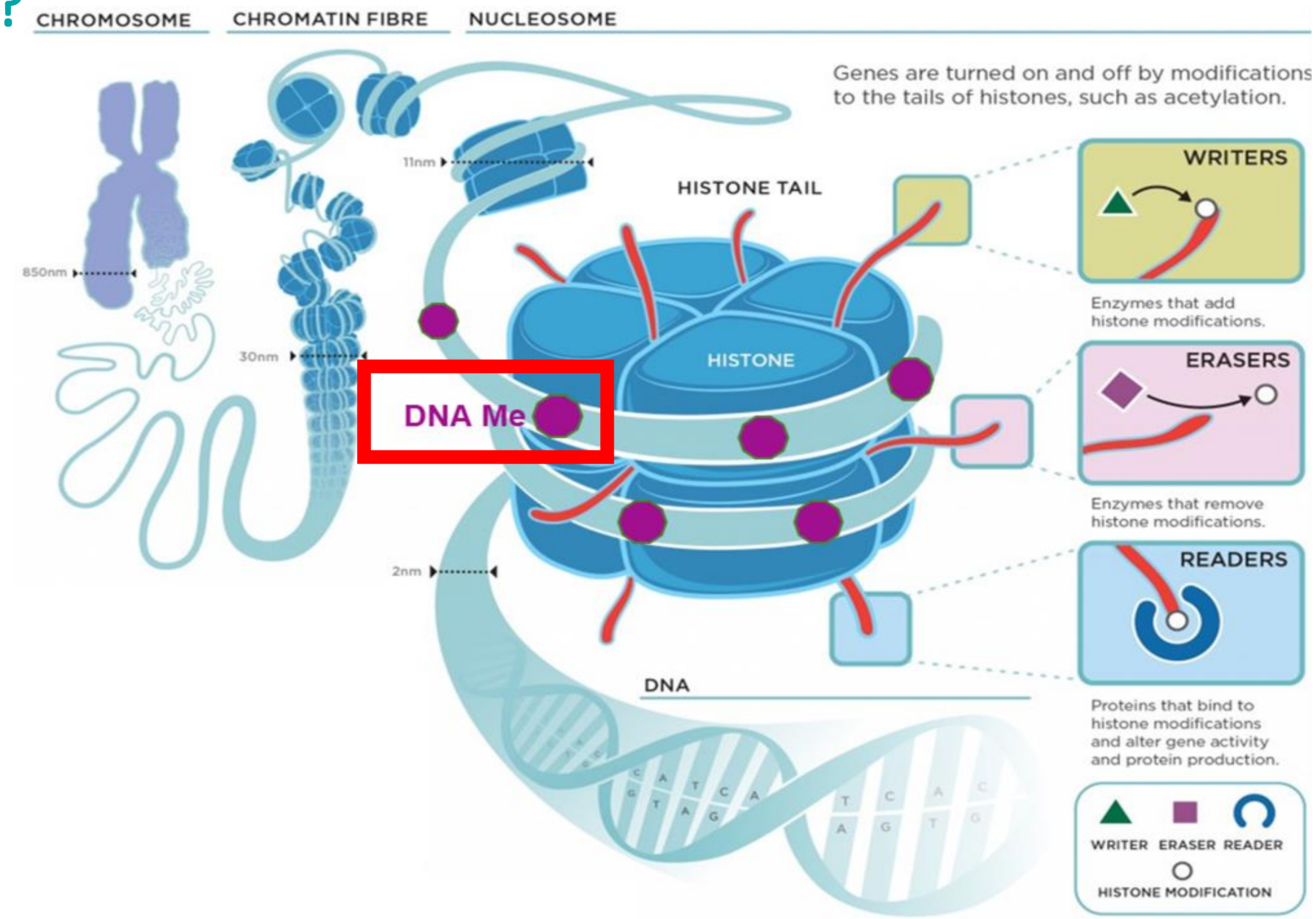


	Supported by
Genomic	DNA
Epigenomic	Chromatine Cytosines
Transcriptomic	RNA
Proteomic	Proteins
Metablomic	Small Molecule *(Sugar, lipid, AA, nucleotid, ...)*
Phenomic	Measurable characteristic *(Growth, resistance, forms, …)*

→ **Each omics layer can interact with another omics layer**

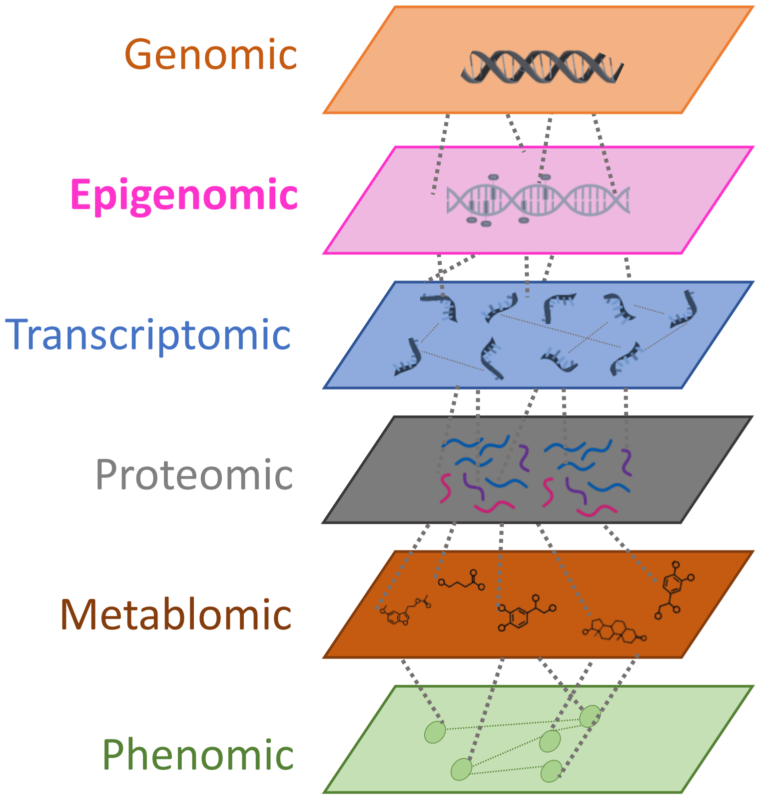
What is the epigenetics?

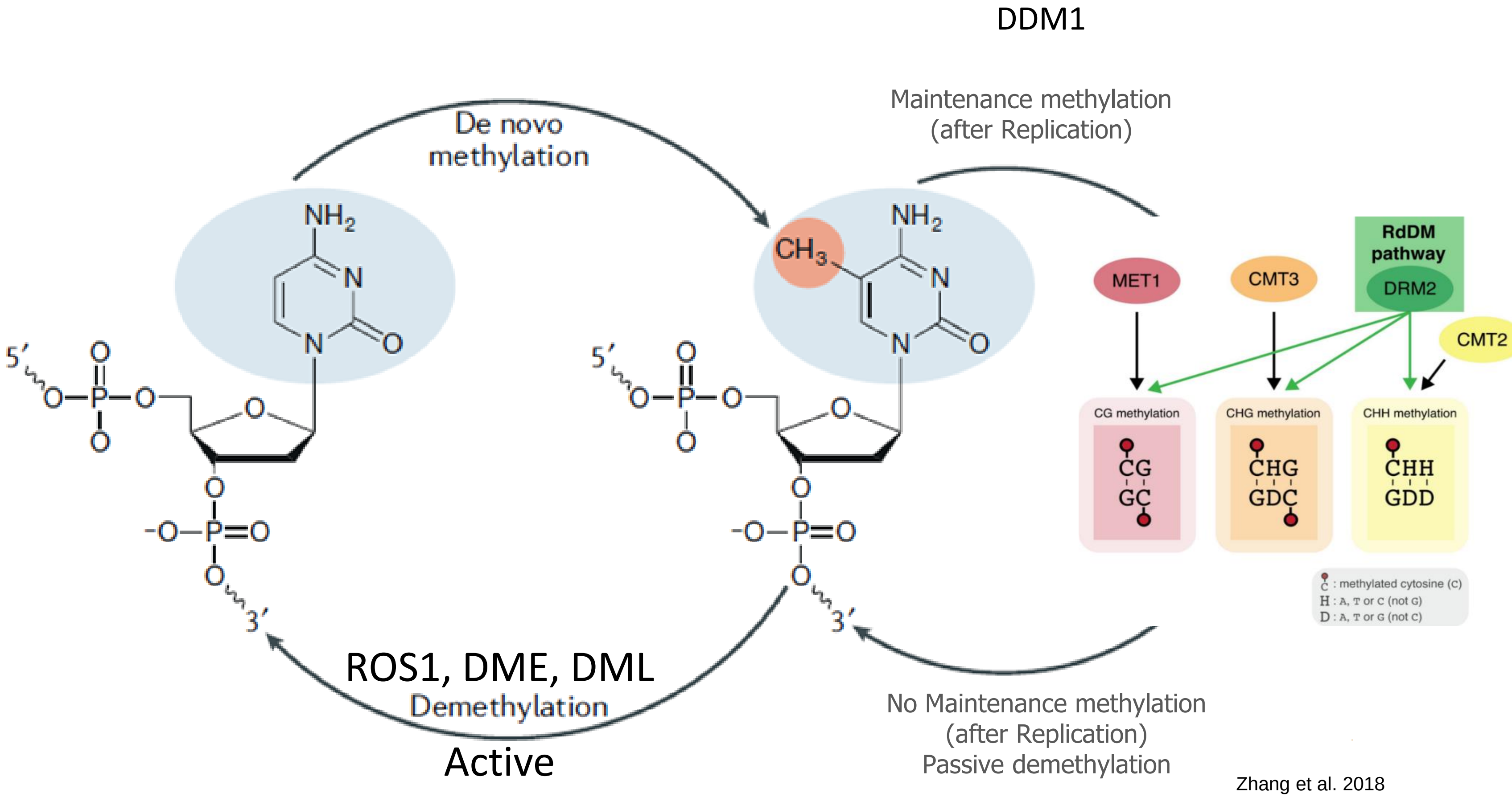




→ **Study of chromatin-level processes affecting gene expression and/or the activity of transposable elements (TEs) without altering the DNA sequence and heritable**
→ **Affected by the environment**

What is the epigenetics?





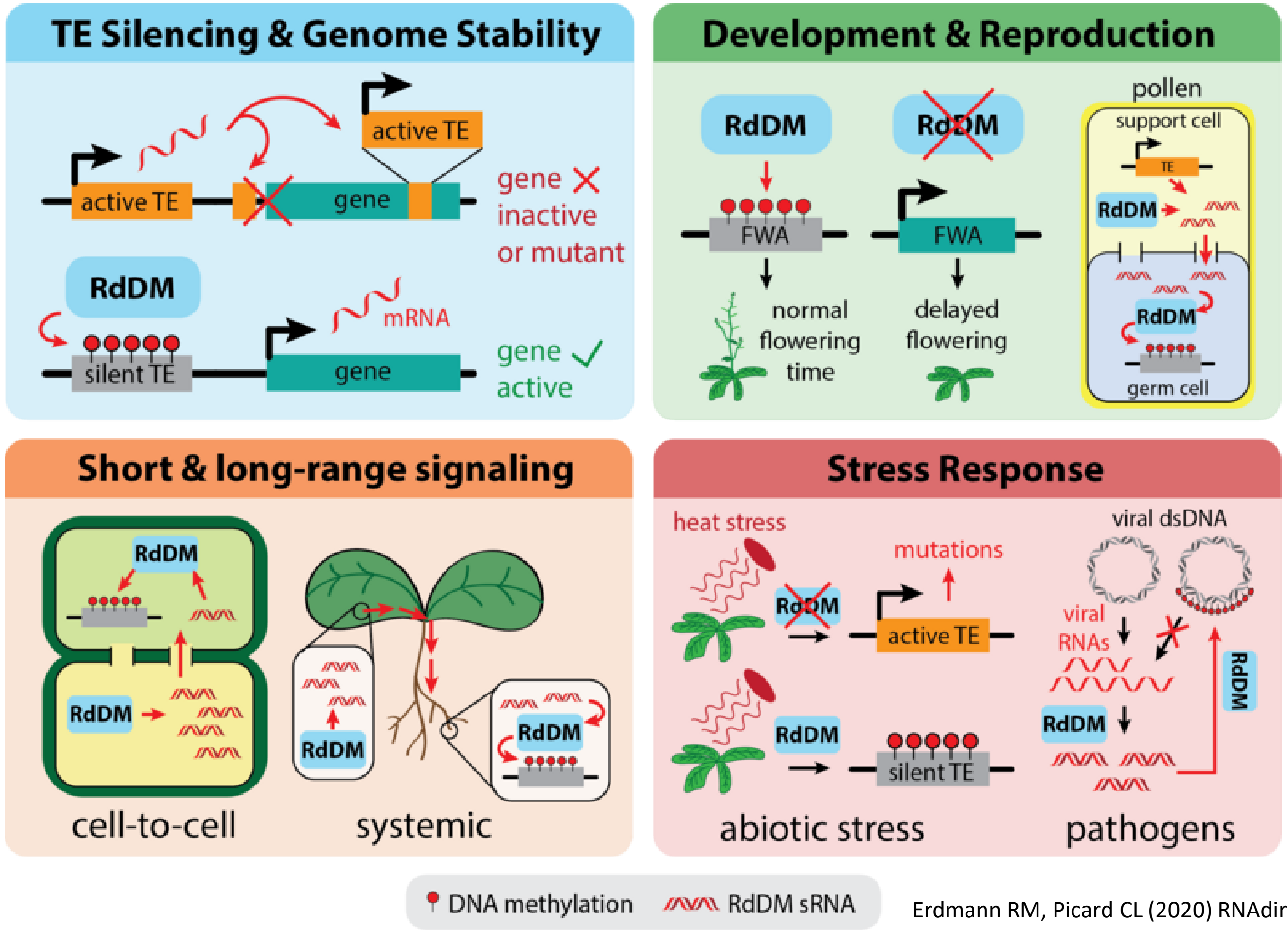
Zhang et al. 2018
Erdmann RM, Picard CL
2020

→ ***de novo* and maintenance DNA methylation**

→ **Methylation and demethylation**

→ **3 methylation contexts: CpG, CHG & CHH (H=A, T or C)**

DNA methylation and Plant Biology



Erdmann RM, Picard CL (2020) RNAdirected DNA Methylation. PLoS Genet 16(10): e1009034. https://doi.org/10.1371/journal.pgen.1009034

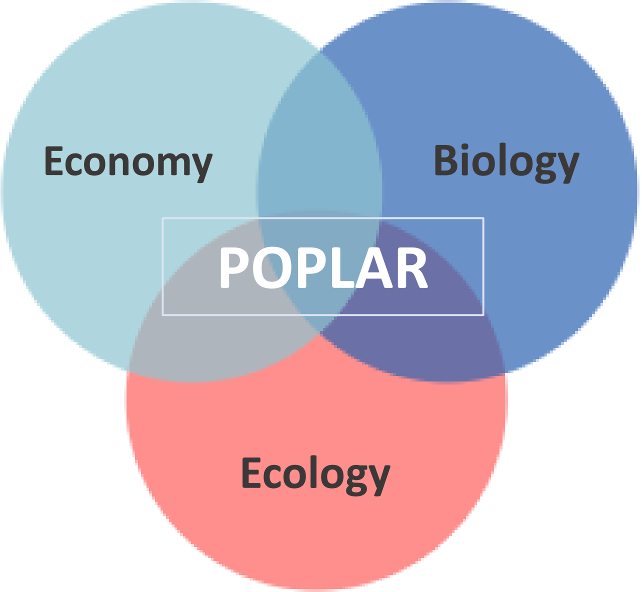
The power of epigenetics

Same genome
Different epigenome

→ Epigenetic (DNA methylation) play many roles in plants
→ Which trees to study this mechanism

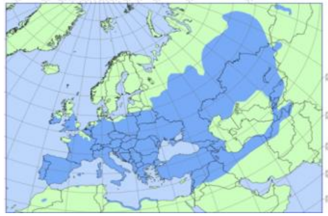
The model tree: the poplar

- **Cultivated interspecific hybrids** (P. nigra (♂) X P. deltoides (♀))
- 9.6 million ha worldwide
- **Timber production**




M. Villar

- **Rapid growth**
- **Clonal propagation**
- Dioecious (controlled crosses)
- High levels of **genetic diversity**
- Transformation capacity (**transgenic**)
- **First tree genome sequenced** (2006)
- **Small diploid genome** (485 ± 10 Mb)

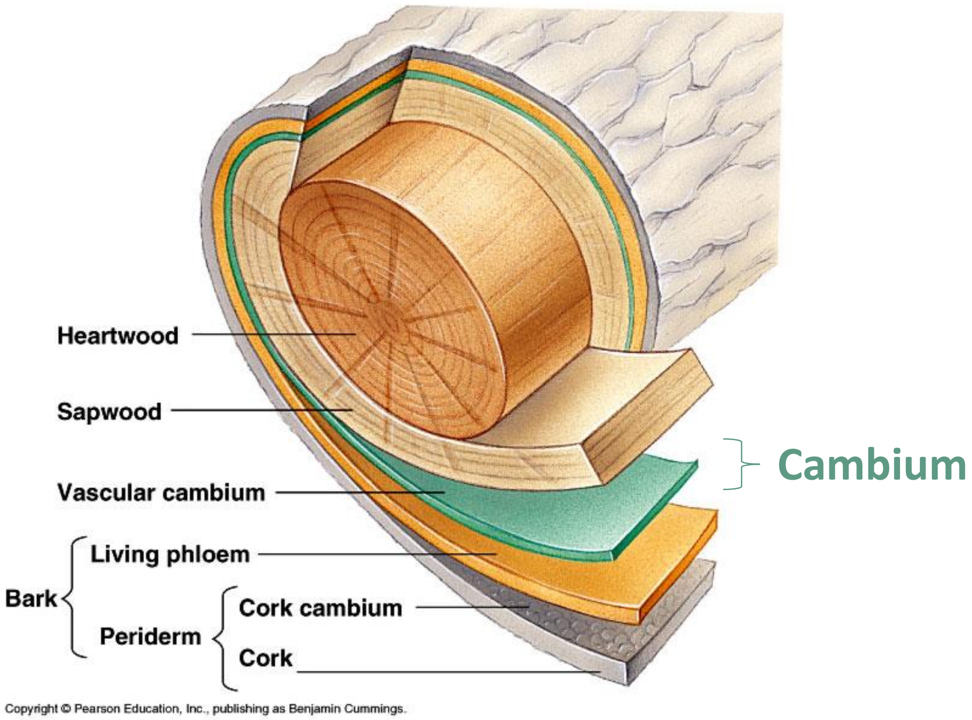

EUFORGEN, *Populus nigra*

- **Large distribution area**
- **Riparian species**
- **Indicator of biodiversity dynamics**

→ Poplar is the tree model

→ But which tissues : leaf, root, trunk?

A meristematic tissue: the Cambium



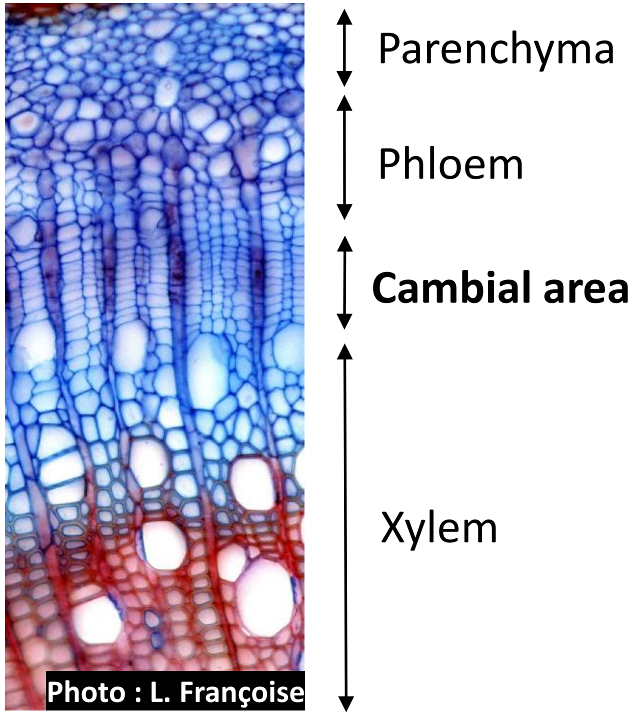


Photo : L. Françoise

Meristem

Cell division (epigenetic transmission)
Cell differentiation
Histogenesis
→ Epigenetic rearrangement

Cambium

- **Wood production (radial growth)**
- **Sensitive to environmental variation**
- **Numerous regulatory factors**
- → **Hormones** (development and stress response)
- → **Genes, transcription factors** (Kim et al., 2019)
- → **Epigenetics** (Maury et al., 2019)

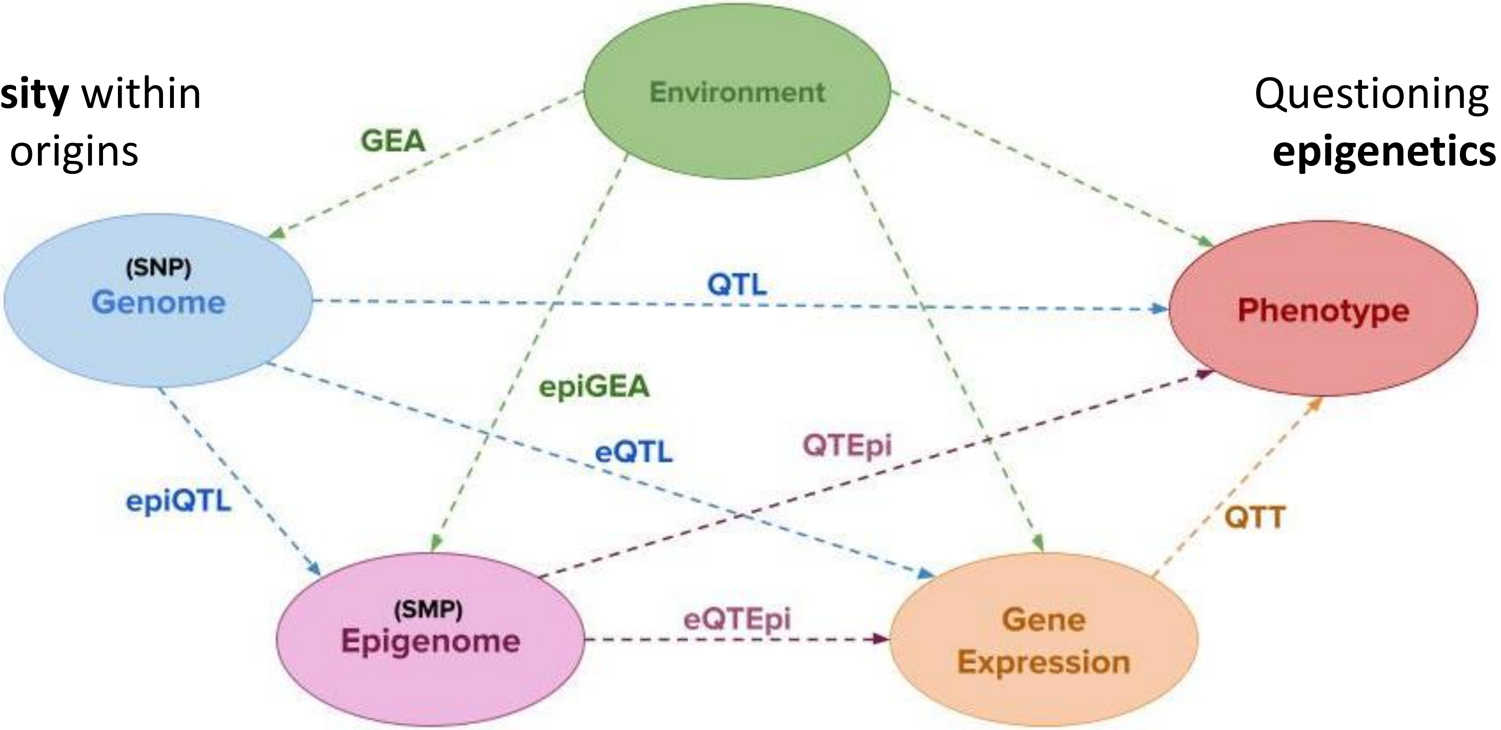
→ A model tissue to better understand the role of epigenetics in trees

Strategy and objective

Search for **epigenetic signatures** and assess their **impact** on the **phenotype**

Studying **epigenetic diversity** within populations of diverse origins

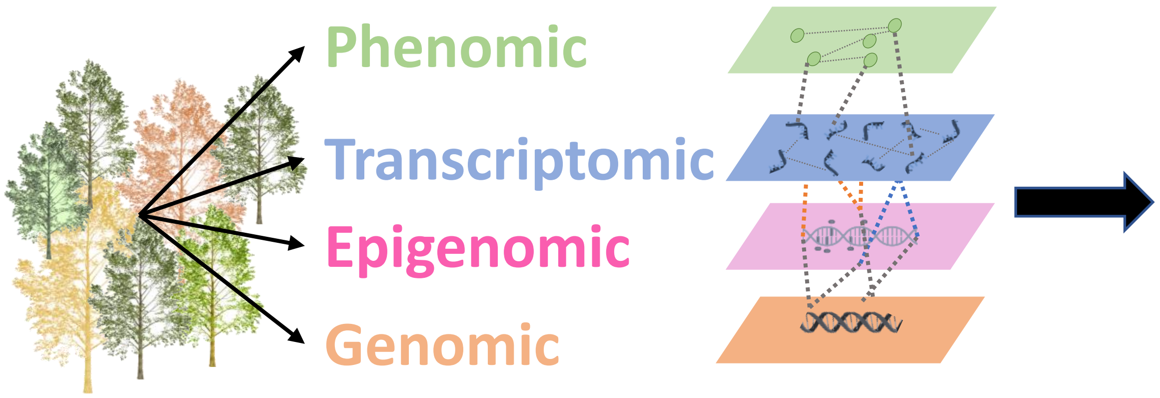
Questioning the potential **role of epigenetics** in **local adaptation**



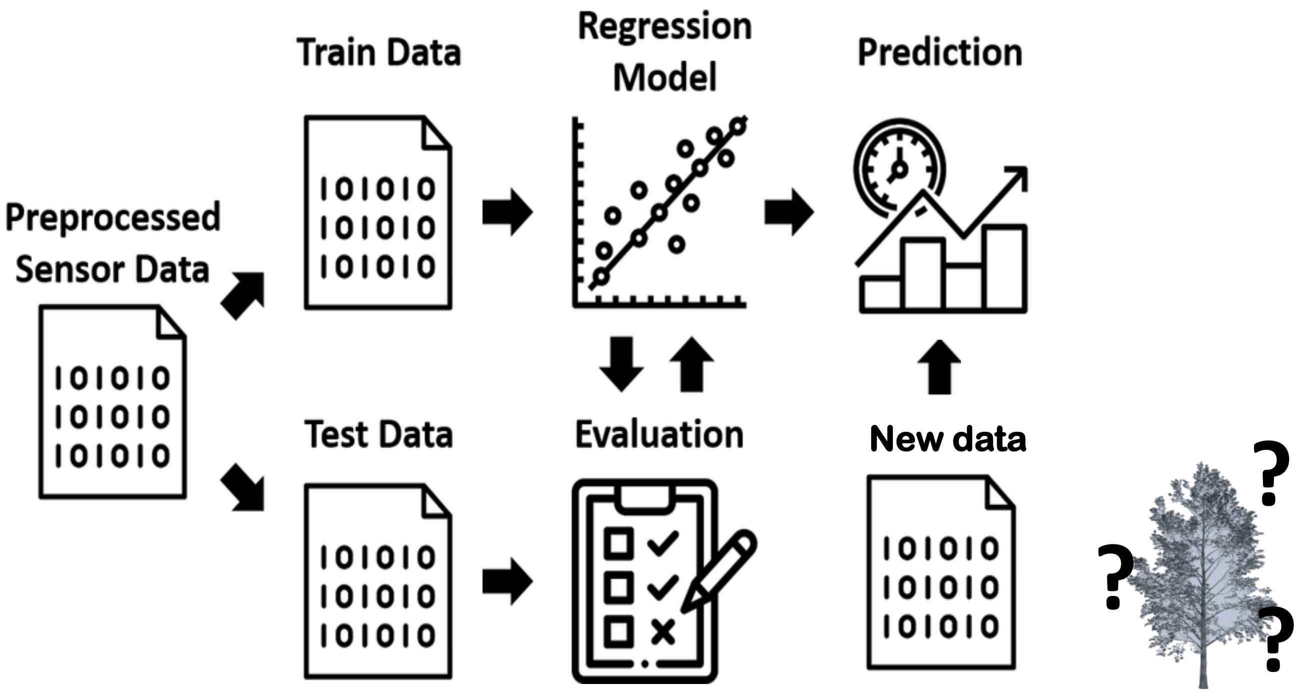
Evaluate the impact of **epigenetic data** on **phenotypic prediction**

→ Multi-omics associations of various loci

Multi-omics and predictive model



Predictive model



Wunderlich *et al*., 2023 (foods)

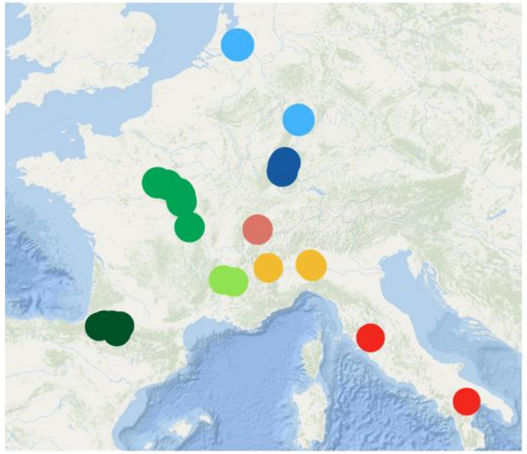
→ **A multi-omics approach (including epigenetics) could improve prediction accuracy**

→ **Variable selection using a predictive model could enhance our understanding of biological processes in local adaptation**

Experimental setup

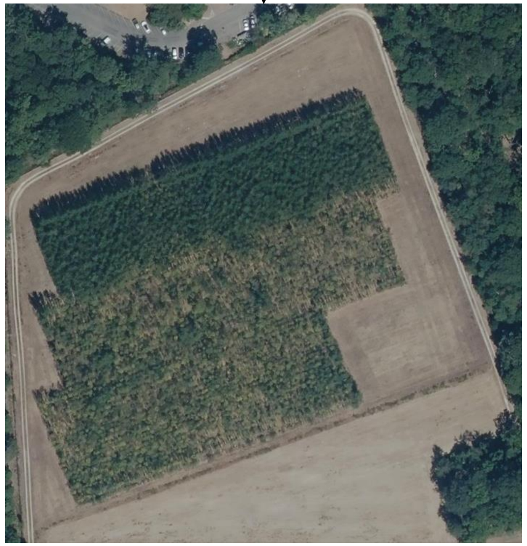
anr agence nationale de la recherche
ANR-13-JSV6-0001
EPITREE

>200 genotypes of black poplar *(P. nigra)*



8 Biogeographical Origin

- 🟦 **Neth_Kuh**
- 🟦 **Rhin**
- 🟥 **Dranse**
- 🟩 **Loire_Allier**
- 🟩 **Ramieres**
- 🟩 **Adour**
- 🟨 **Ticino**
- 🟥 **Pag_basen**



Common garden

Phenotype

20 Phenotypic traits

Orléans (🇫🇷) Savigliano (🇮🇹)

Hauteur, Tolérance à la rouille, angle des branche, composition du bois..

Gene expression

~25k Transcript

Mapping RNAseq V4.1 (Odile ROGIER)

Genomic

~217K SNPs

Environment

37 Climatic variable x 2 scenario

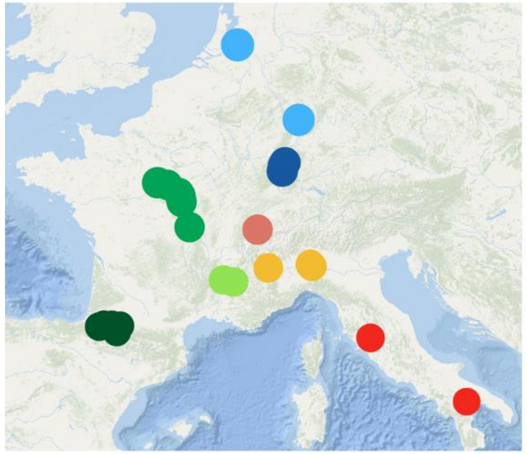
Climate DT B4est

Epigenetics
SeqCap bisulfite

CG : 313K
CHG : 201K
CHH : 360 K

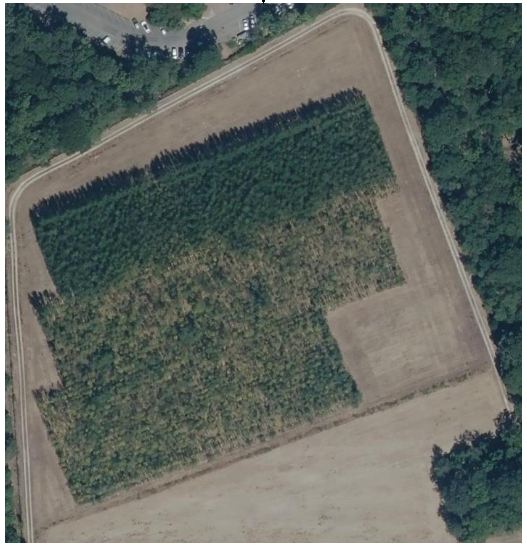
→ **Study of the multi-omics data variability**

Observable variability in the multi-omics data of our individuals

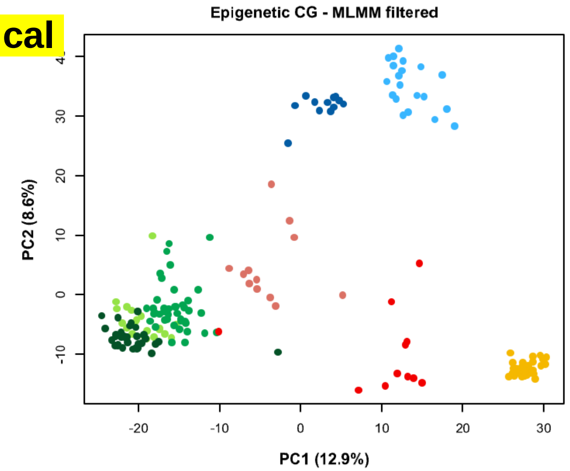


8 Biogeographical Origin

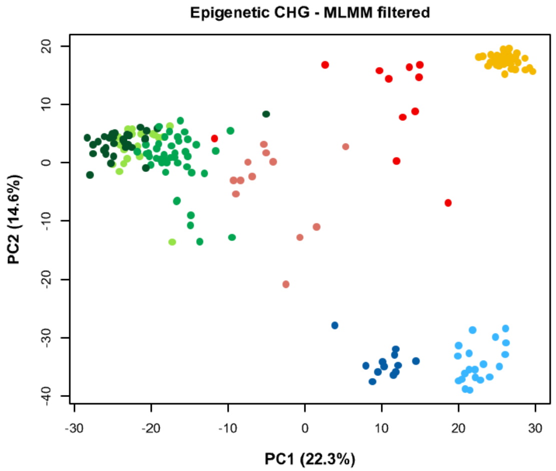
- 🟦 Neth_Kuh
- 🟦 Rhin
- 🟧 Dranse
- 🟩 Loire_Allier
- 🟩 Ramieres
- 🟩 Adour
- 🟨 Ticino
- 🟥 Pag_basen



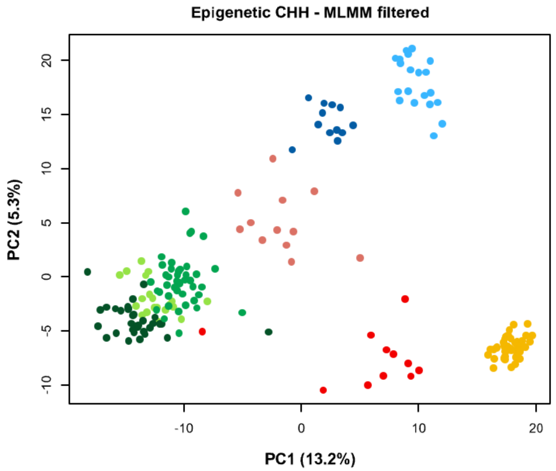
Common garden



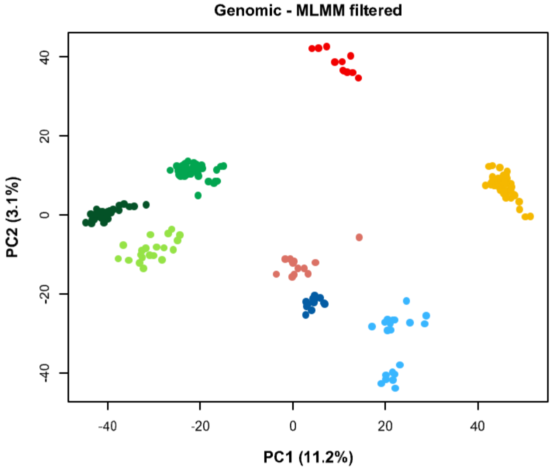
Epigenetic CG - MLMM filtered



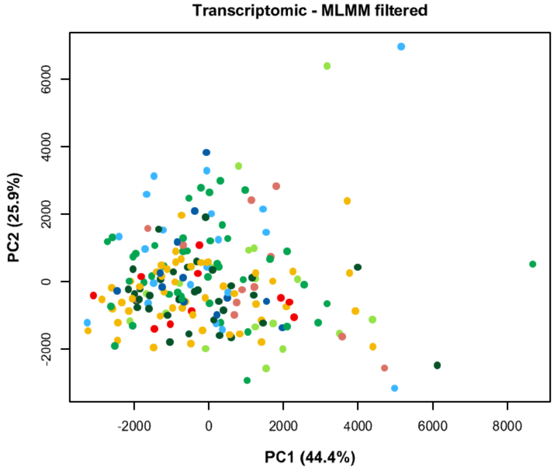
Epigenetic CHG - MLMM filtered



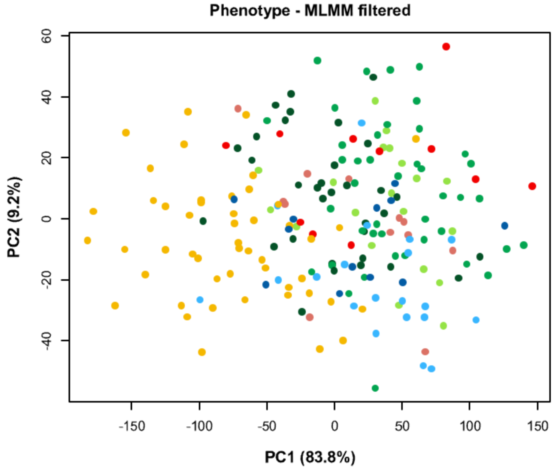
Epigenetic CHH - MLMM filtered



Genomic - MLMM filtered



Transcriptomic - MLMM filtered

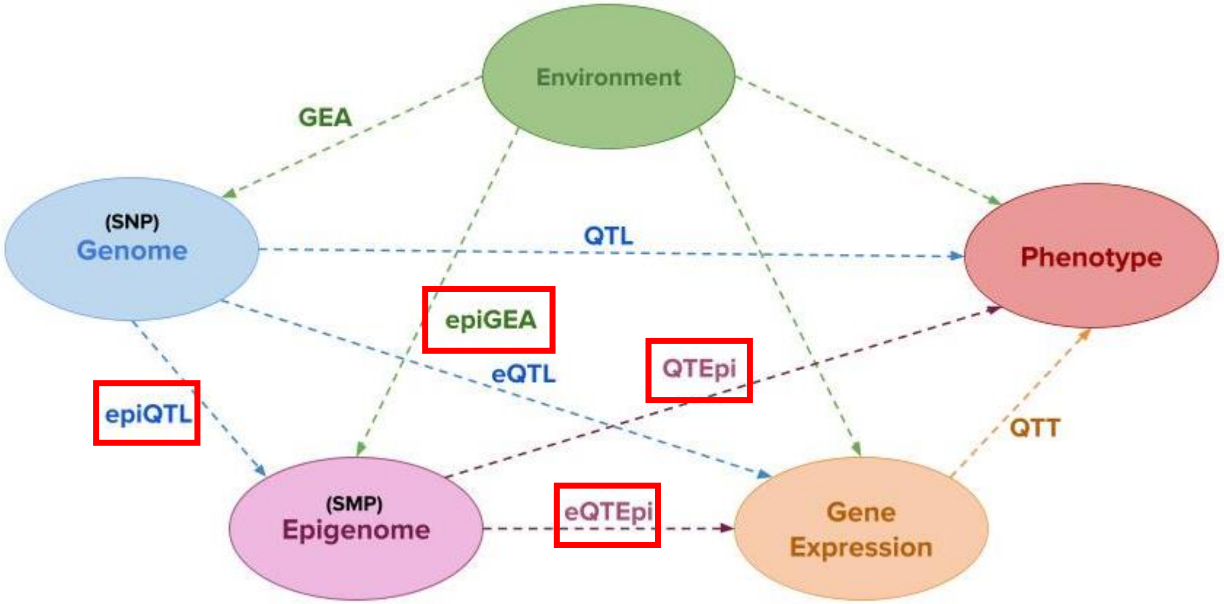


Phenotype - MLMM filtered

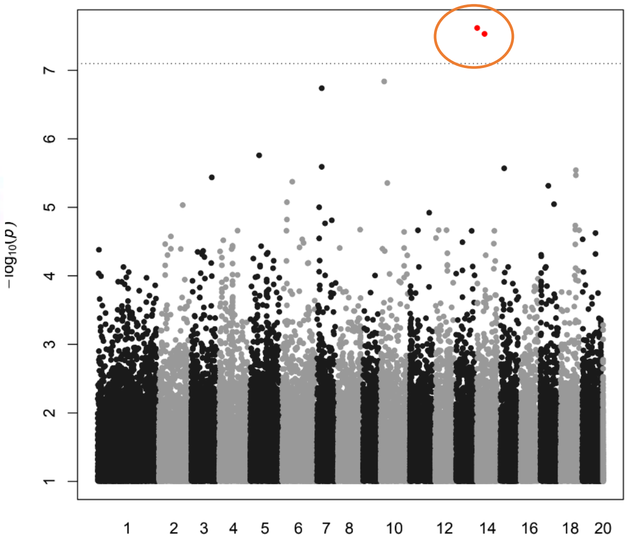
→ Inter- and intra-population genetic and epigenetic variability

→ Climate gradient?

Locus correlation at various omics levels

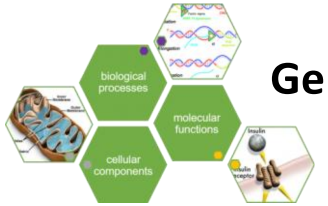


~140K SMPs associated to other omics



Correction Bonferroni

Features

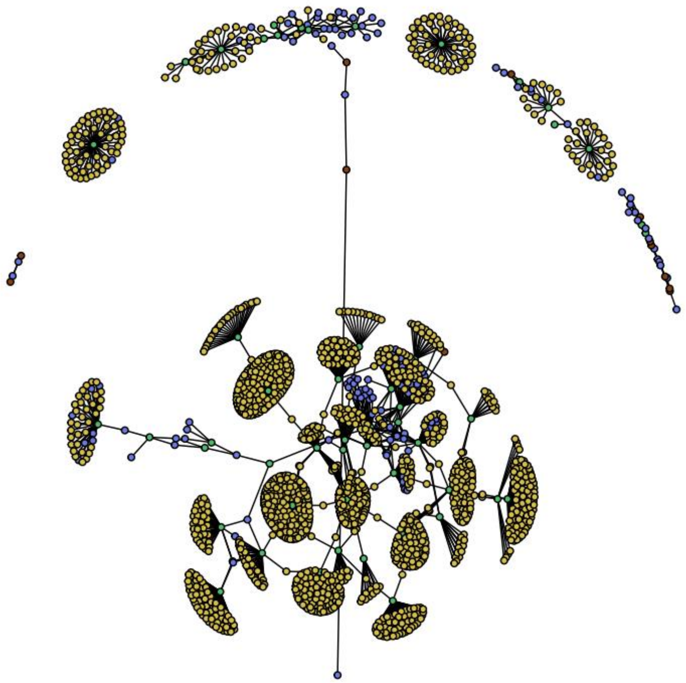


Gene ontology

Biblio

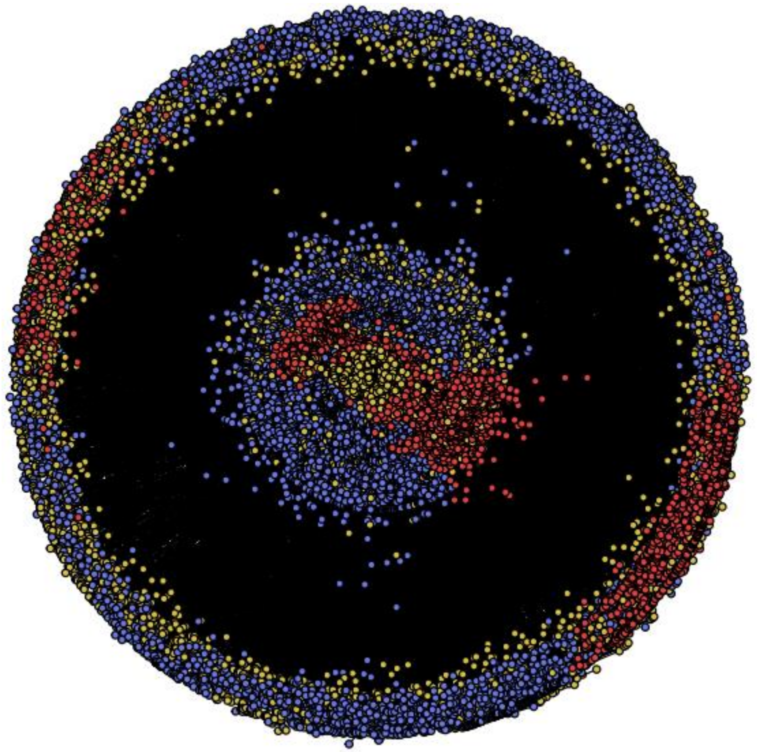
→ What reveal this multiple associations test

Visualisation of omics associations



Highlighting certain pathways

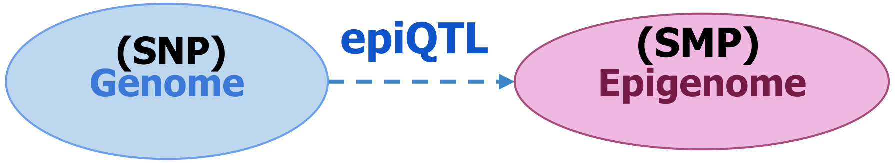
🟥 Mutation
🟦 Epimutation
🟨 Gene expression

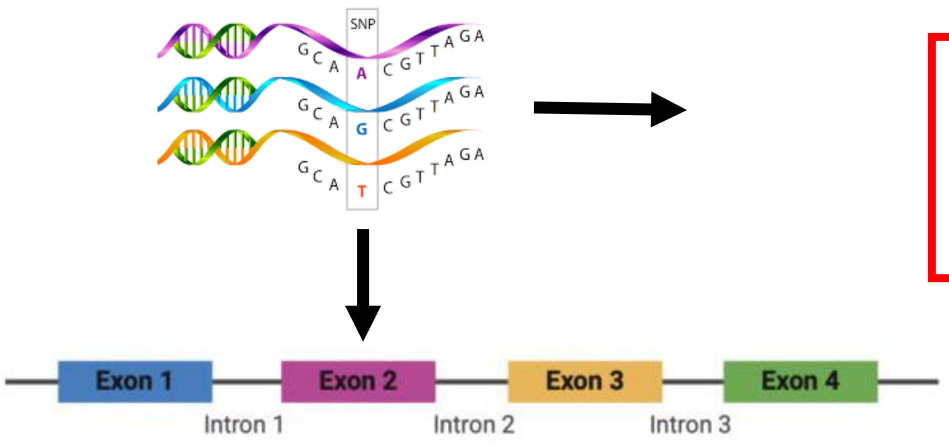


All the relationships highlighted by the associations

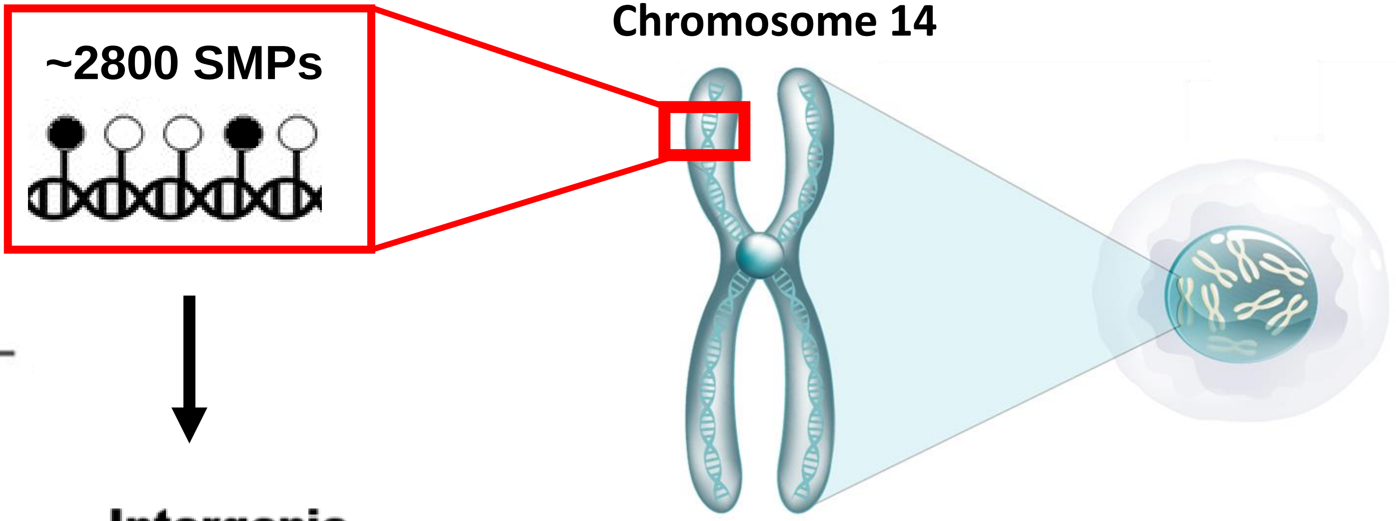
→ **Visualization issues with so many associations**
→ **Best candidate analysis**

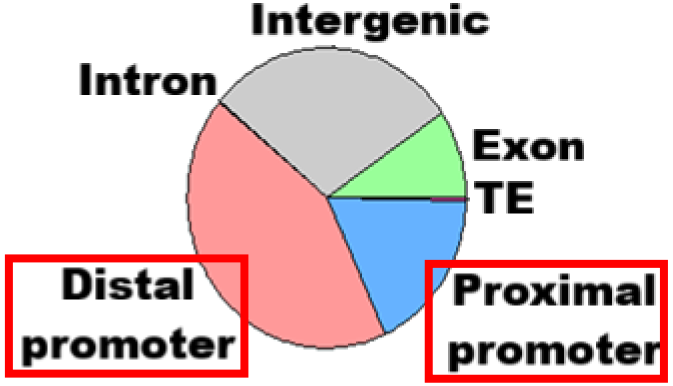
Best candidates found in association





SBP1
Selenium binding protein 1



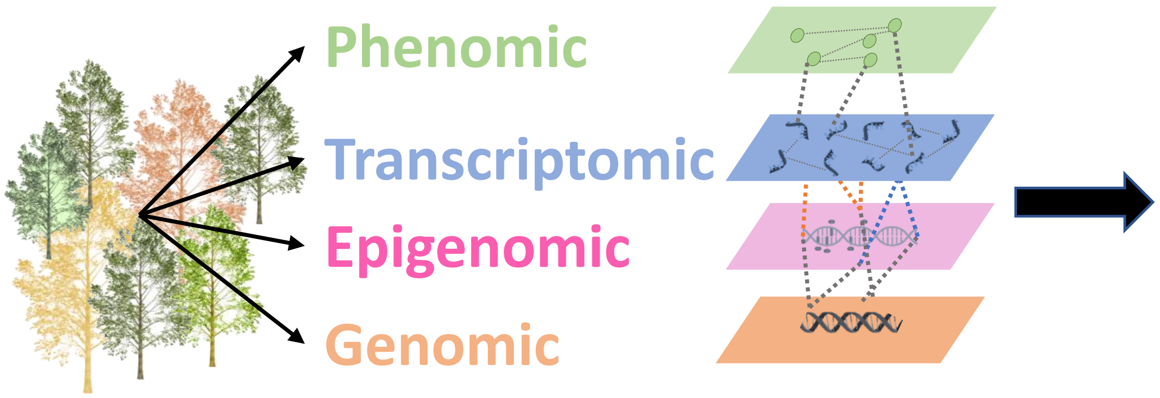


~60% in PROMOTER

→ ????

13 genes involved to stress reponse
with same function

Multi-omics and predictive model



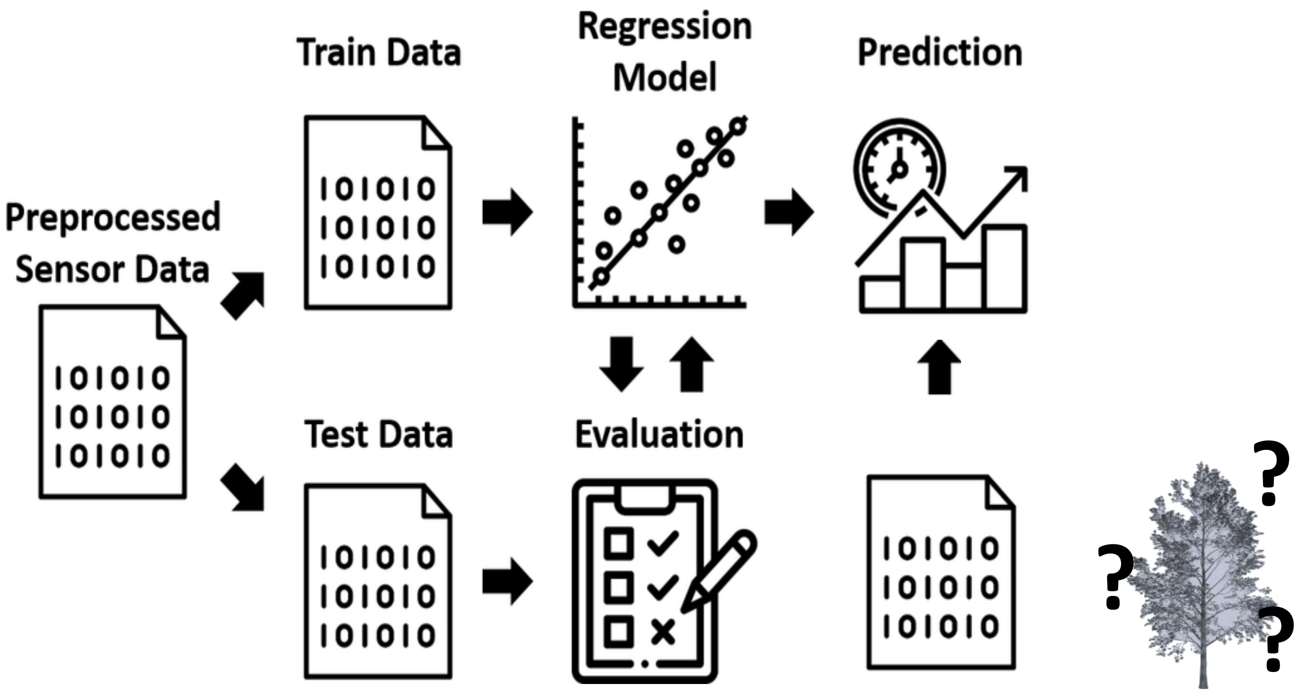
Phenomic

Transcriptomic

Epigenomic

Genomic

Predictive model



Wunderlich *et al*., 2023 (foods)

→ **Epigenetics could improve accuracy of predictive models**

→ **Predictive model using multi-omics data could reveal biological pathways related to phenotype**

Phenotypic prediction model using our data

Research article | Open access | Published: 28 June 2022

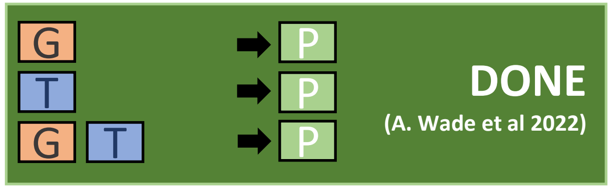
eQTLs are key players in the integration of genomic and transcriptomic data for phenotype prediction

Abdou Rahmane Wade, Harold Duruflé, Leopoldo Sanchez & Vincent Segura

BMC Genomics **23**, Article number: 476 (2022) | Cite this article

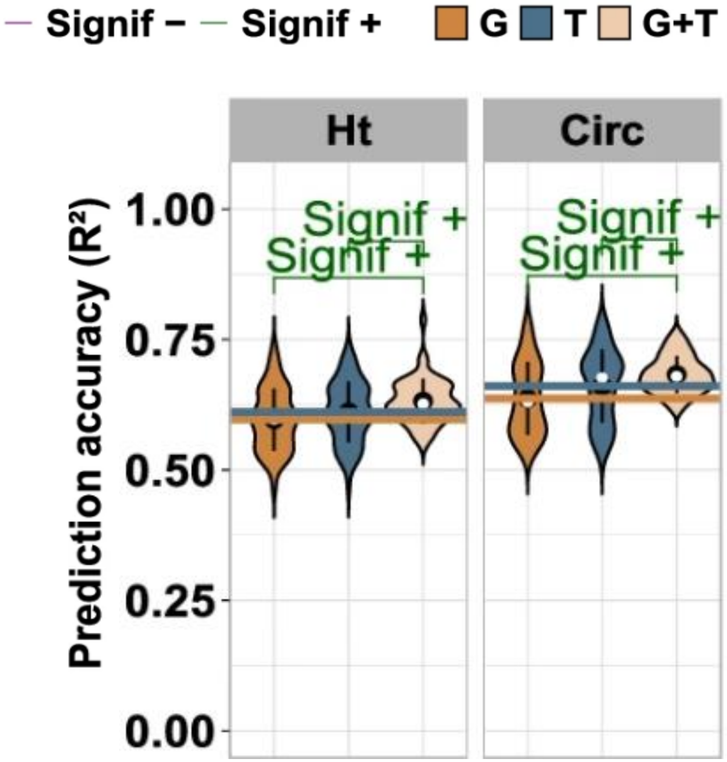
2108 Accesses | **4** Citations | **3** Altmetric | Metrics

Ridge Regression



DONE
(A. Wade et al 2022)

Reveal a set of eQTLs of interest



Wade *et al.*, 2022

→ Using multiple layers of omics data can improve accuracy

Strategy applied to multi-omics prediction

Research article | Open access | Published: 28 June 2022

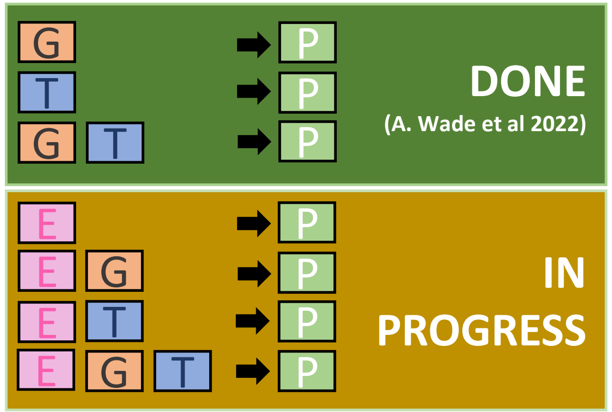
eQTLs are key players in the integration of genomic and transcriptomic data for phenotype prediction

Abdou Rahmane Wade, Harold Duruflé, Leopoldo Sanchez & Vincent Segura

BMC Genomics **23**, Article number: 476 (2022) | Cite this article

2108 Accesses | **4** Citations | **3** Altmetric | Metrics

Applying **epigenetic** data to the **Ridge Regression** model



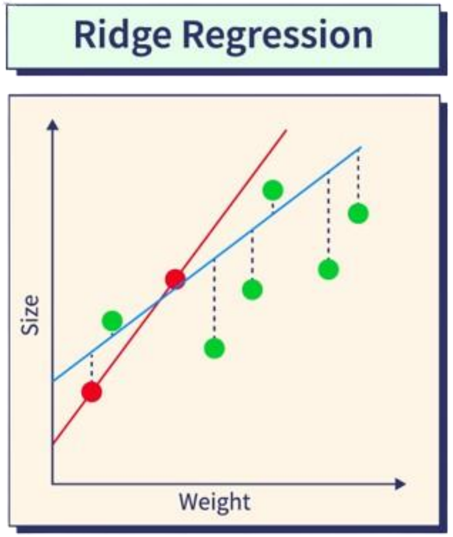
G ➡ P
T ➡ P
G T ➡ P

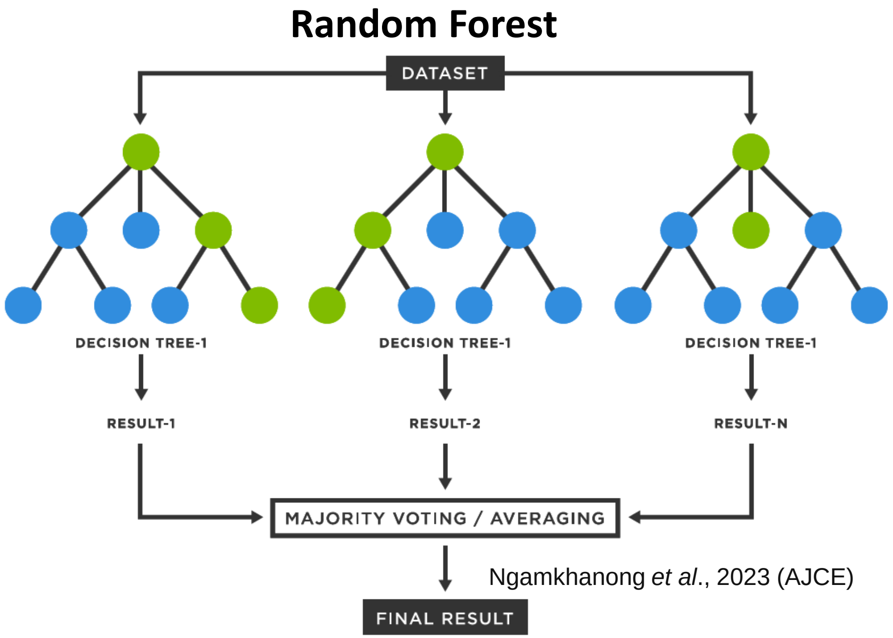
DONE
(A. Wade et al 2022)

E ➡ P
E G ➡ P
E T ➡ P
E G T ➡ P

IN PROGRESS

Other ML/DL models?



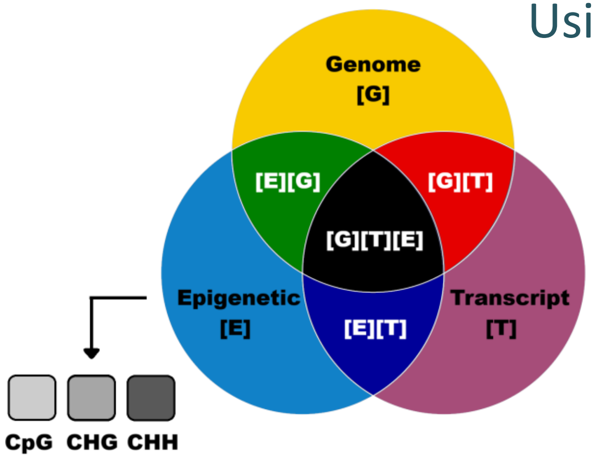


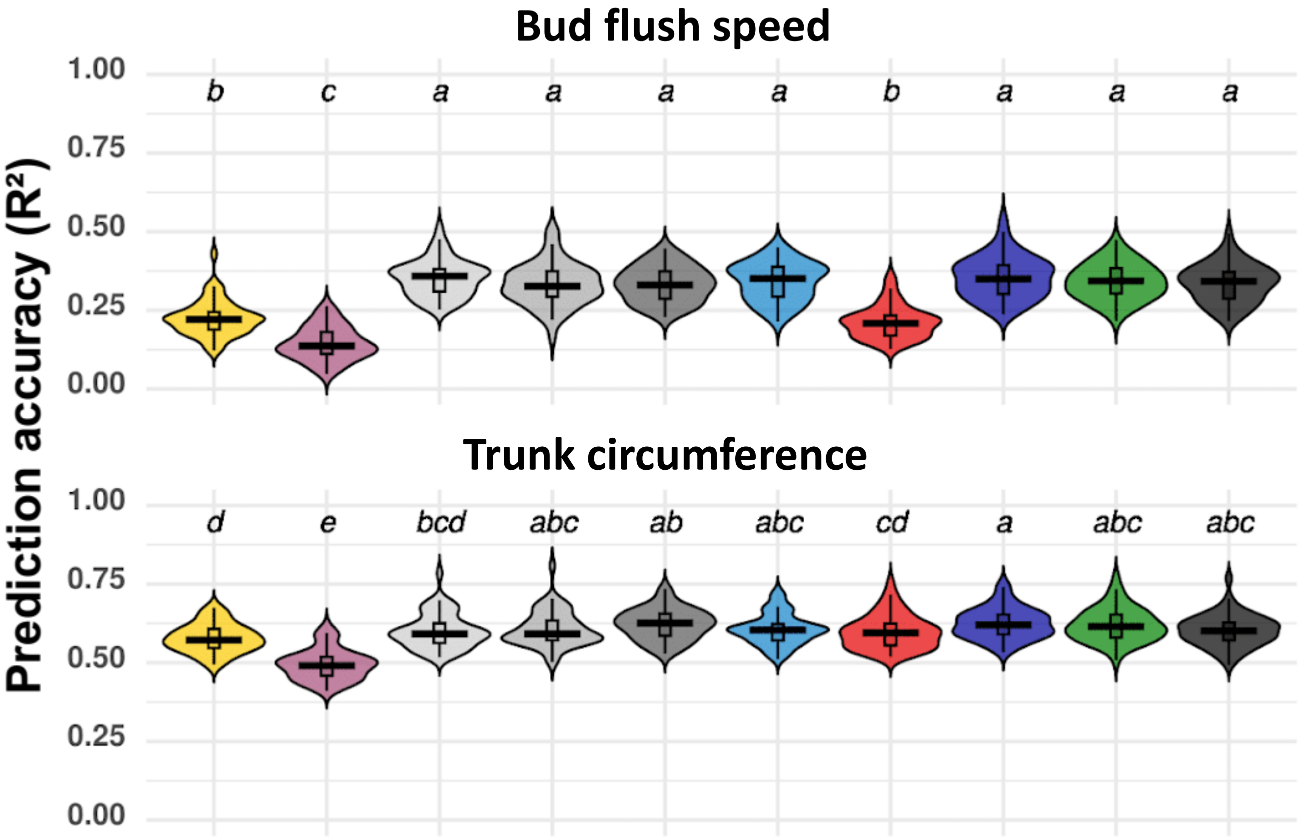
Ngamkhanong *et al.*, 2023 (AJCE)

→ **Evaluate the predictive capacity of epigenetic data**

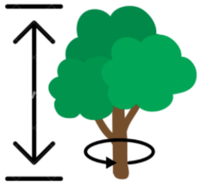
Prediction accuracies of methylation data

Using the concatenation method and ridge regression model




Bud flush speed

Trunk circumference



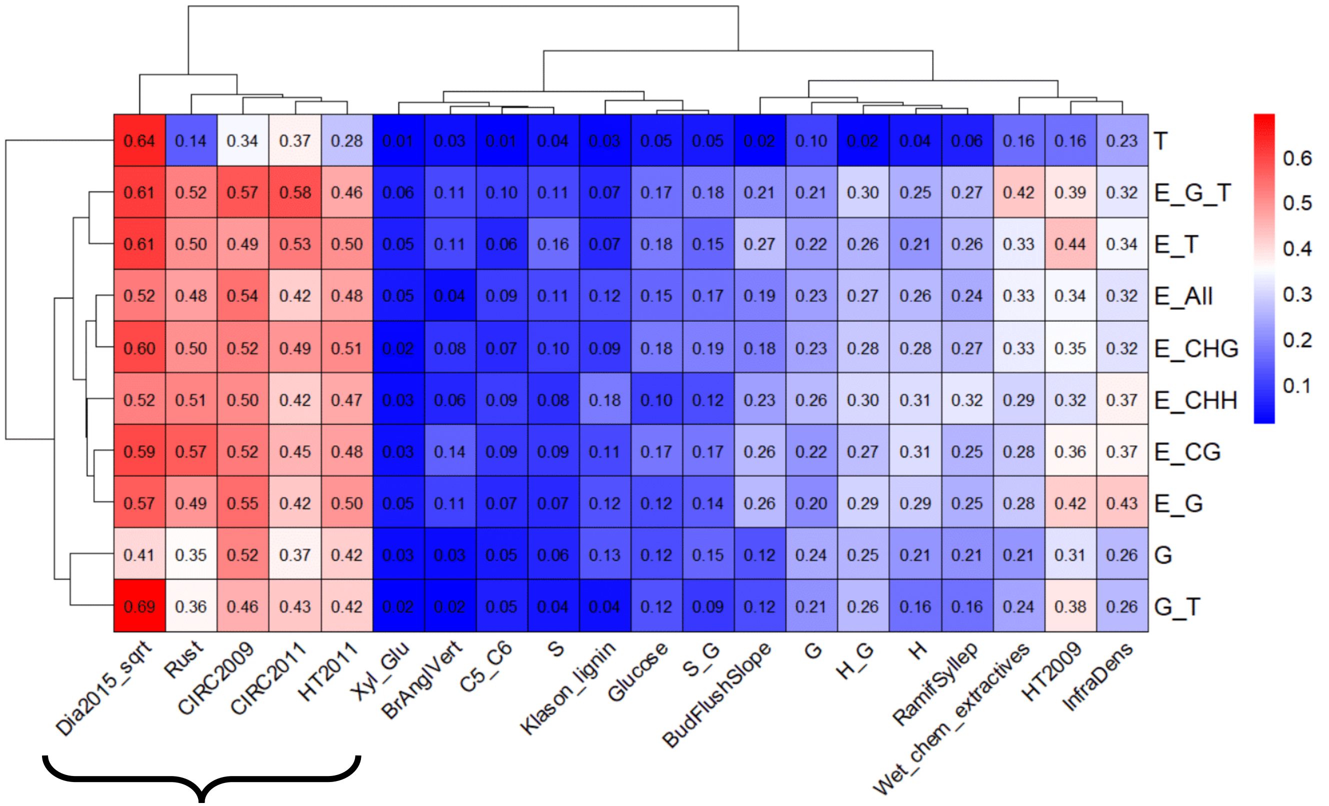
Duplan et al. In prep

→ DNA methylation as a new potential source of predictors

Prediction accuracies are phenotypic- and dataset-dependent

- Accuracy with R² between **0,01** and **0,69**
- The combinason of all omics don't increase the accuracy

Accuracy (R²) of phenoypic prediction

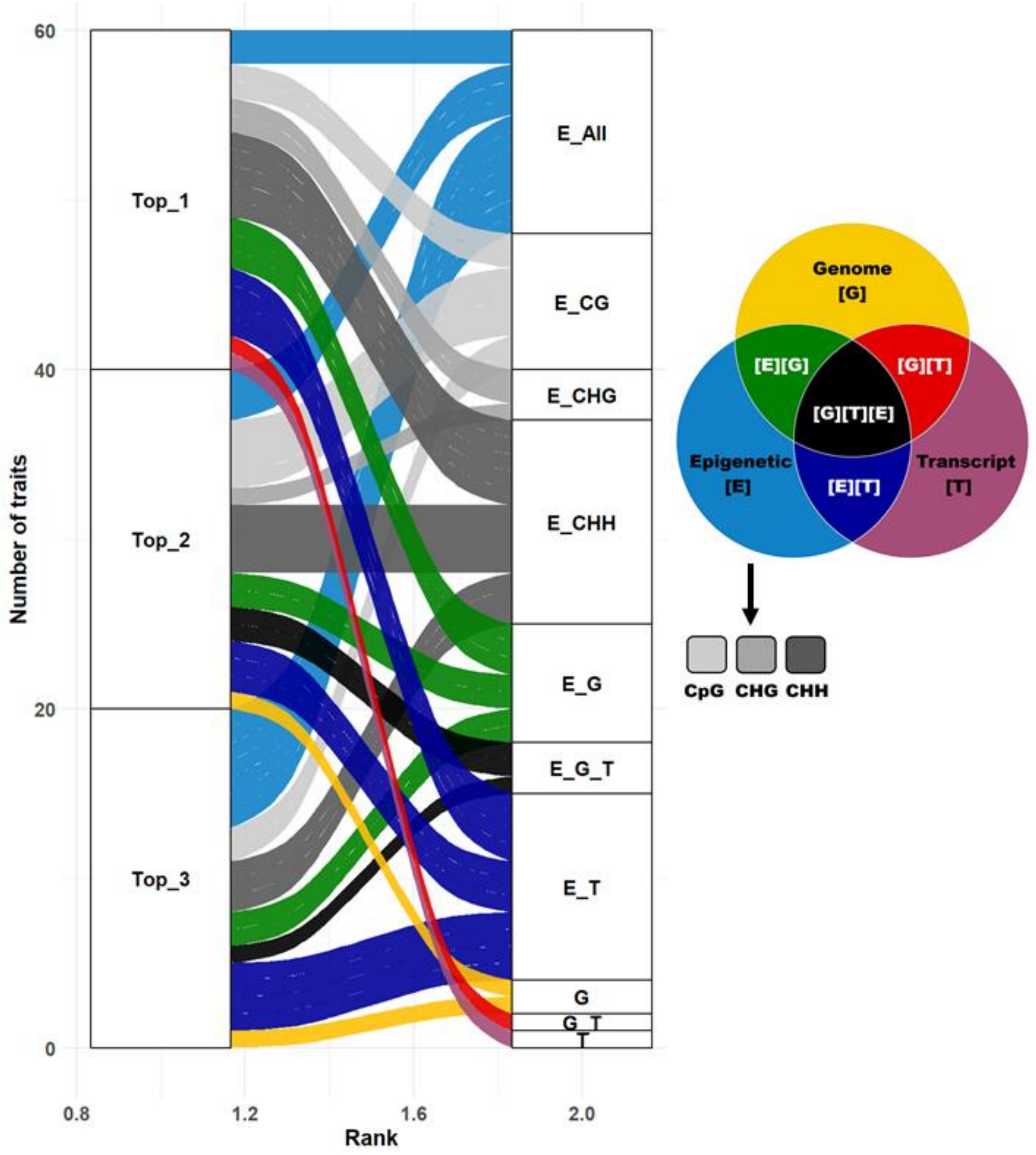


	Dia2015_sqrt	Rust	CIRC2009	CIRC2011	HT2011	Xyl_Glu	BrAnglVert	C5_C6	S	Klason_lignin	Glucose	S_G	BudFlushSlope	G	H_G	H	RamifSyllep	Wet_chem_extractives	HT2009	InfraDens
T	0.64	0.14	0.34	0.37	0.28	0.01	0.03	0.01	0.04	0.03	0.05	0.05	0.02	0.10	0.02	0.04	0.06	0.16	0.16	0.23
E_G_T	0.61	0.52	0.57	0.58	0.46	0.06	0.11	0.10	0.11	0.07	0.17	0.18	0.21	0.21	0.30	0.25	0.27	0.42	0.39	0.32
E_T	0.61	0.50	0.49	0.53	0.50	0.05	0.11	0.06	0.16	0.07	0.18	0.15	0.27	0.22	0.26	0.21	0.26	0.33	0.44	0.34
E_All	0.52	0.48	0.54	0.42	0.48	0.05	0.04	0.09	0.11	0.12	0.15	0.17	0.19	0.23	0.27	0.26	0.24	0.33	0.34	0.32
E_CHG	0.60	0.50	0.52	0.49	0.51	0.02	0.08	0.07	0.10	0.09	0.18	0.19	0.18	0.23	0.28	0.28	0.27	0.33	0.35	0.32
E_CHH	0.52	0.51	0.50	0.42	0.47	0.03	0.06	0.09	0.08	0.18	0.10	0.12	0.23	0.26	0.30	0.31	0.32	0.29	0.32	0.37
E_CG	0.59	0.57	0.52	0.45	0.48	0.03	0.14	0.09	0.09	0.11	0.17	0.17	0.26	0.22	0.27	0.31	0.25	0.28	0.36	0.37
E_G	0.57	0.49	0.55	0.42	0.50	0.05	0.11	0.07	0.07	0.12	0.12	0.14	0.26	0.20	0.29	0.29	0.25	0.28	0.42	0.43
G	0.41	0.35	0.52	0.37	0.42	0.03	0.03	0.05	0.06	0.13	0.12	0.15	0.12	0.24	0.25	0.21	0.21	0.21	0.31	0.26
G_T	0.69	0.36	0.46	0.43	0.42	0.02	0.02	0.05	0.04	0.04	0.12	0.09	0.12	0.21	0.26	0.16	0.16	0.24	0.38	0.26

Phenotypic trait well predicted is related to growth

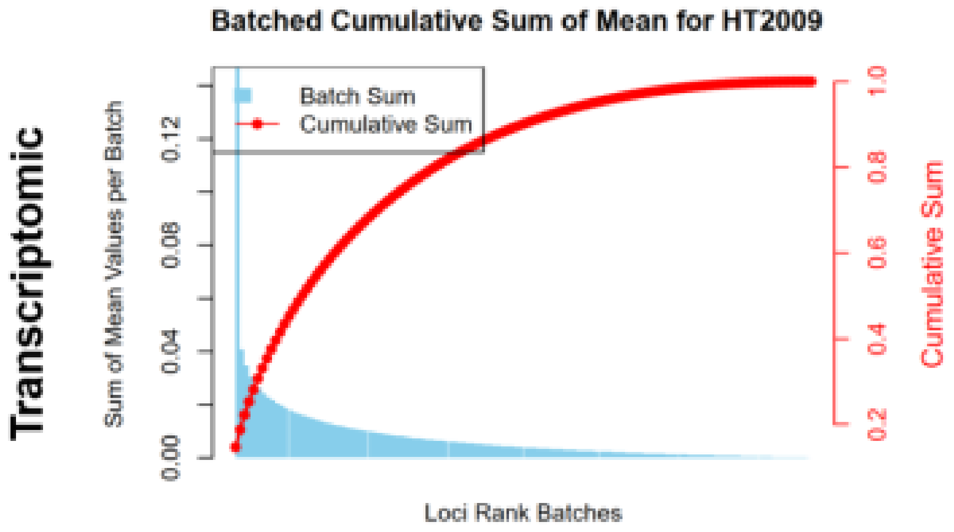
Prediction accuracies are phenotypic- and dataset-dependent

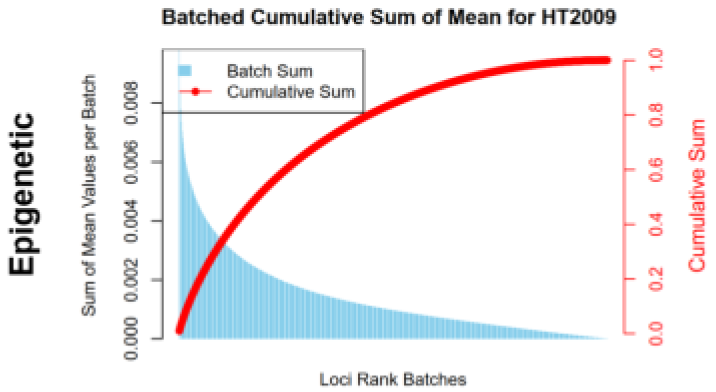
- Accuracy with R^2 between **0,01** and **0,69**
- The combinason of all omics don’t increase the accuracy
- **Epigenetic** is the **best dataset predictor**

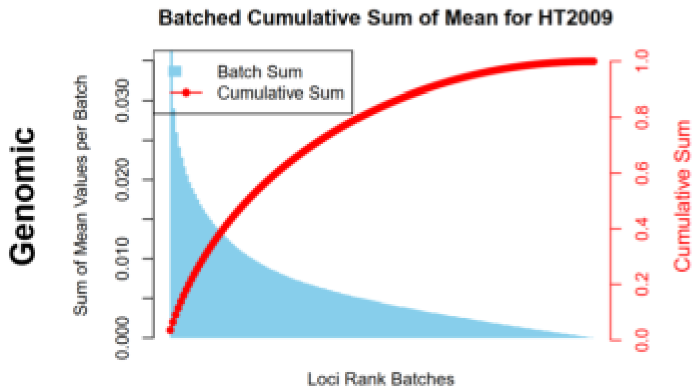


Prediction accuracies are phenotypic- and dataset-dependent

- Accuracy with R^2 between **0,01** and **0,69**
- The combinason of all omics don’t increase the accuracy
- **Epigenetic** is the **best dataset predictor**
- A batch of 200 transcriptions seems to be much more productive than the rest
- Top 20 explicative variable show a reccurent transcript related to the same function as omics associations
- Kernel polynomial and problem of interpretability







Take home message

Epigenetic is an interesting predictor dataset for phenotypic trait and contributes to the understanding of biological processes.

Research on new methods for integrating omics data and relationships discovered by omics associations

Articles in progress
1 - Brandt & Duplan et al. in prep
2 - Duplan et al. in prep

Jérôme Salse
Mamadou Dia Sow

Ludovic Duvaux
Shannon Brandt

Research will continue with a phd student

Thomas Wacquet
(PhD 2025-2028)

Thanks to the Thesis supervisors

Stéphane Maury

Leopoldo Sanchez-Rodriguez

Harold Duruflé

→ Thesis defense on December 18, 2025

ECOLE DOCTORALE SSBCV
Santé Sciences Biologiques Chimie du Vivant

Université d'ORLÉANS

INRAE

BioForA

P2e
PHYSIOLOGY ECOLOGY ENVIRONMENT

ANR
AGENCE NATIONALE DE LA RECHERCHE

EPITREE

Thank’s for your attention

Q A