

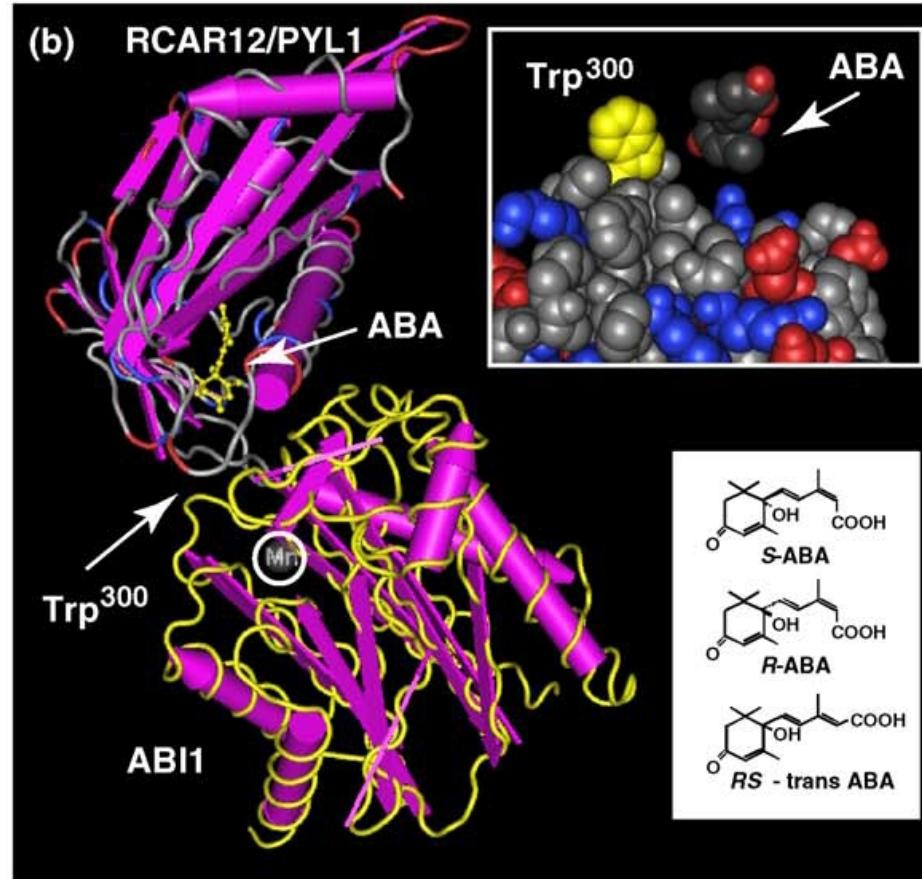


# Natural Variation of Gene Regulatory Networks in *Arabidopsis thaliana*

Arthur Korte

Montpellier Oct 10<sup>th</sup> 2019

# PhD in Molecular Biology

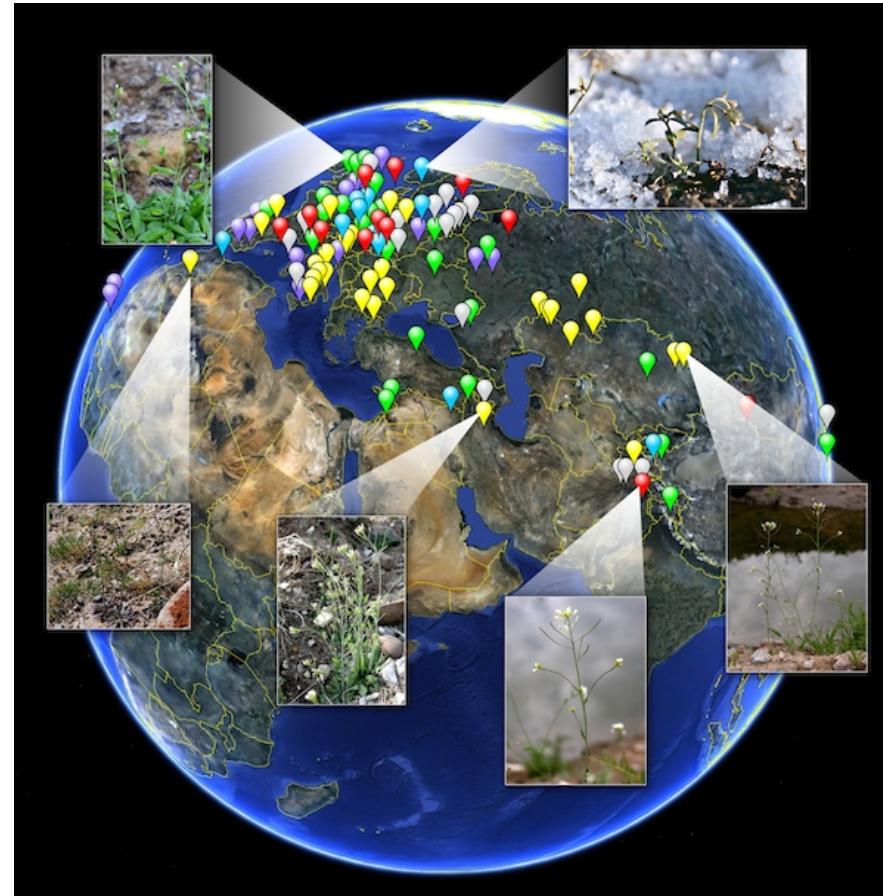


## How do plants react to stress ?

# Postdoc in Population Genetics



By Emmanuel Boutet - Own work, CC BY-SA 3.0,  
<https://commons.wikimedia.org/w/index.php?curid=1437488>

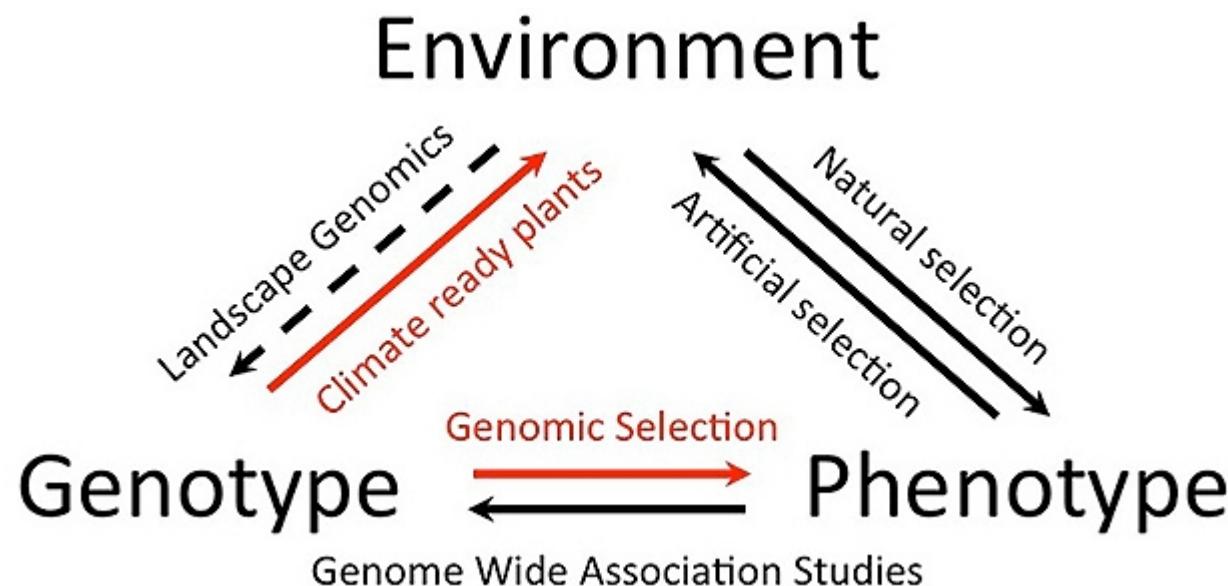


Credit:  
Photographs/images  
were contributed by  
Patrick Gooden,  
Kathleen Donohue and  
Google Earth. Graphic  
design: Jamie Simon,  
Salk Institute.

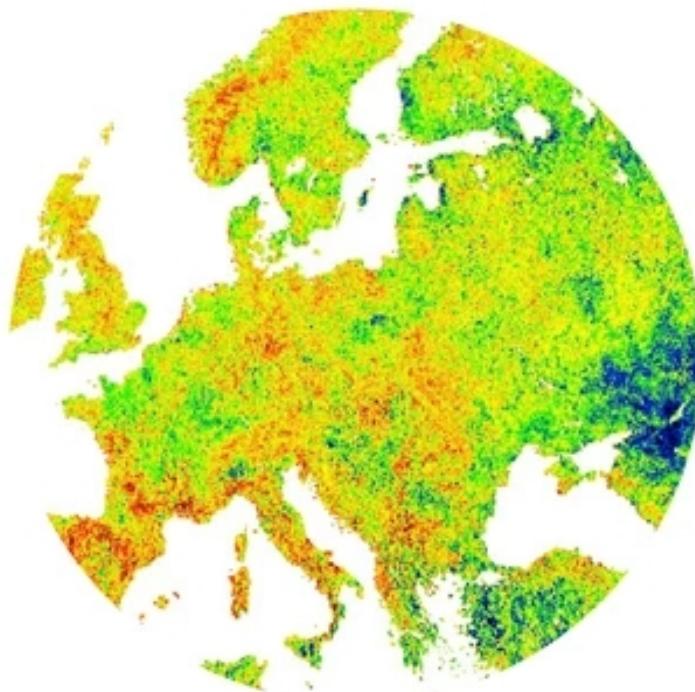
## How do populations adapt to the environment ?

# **Quantitative Genetics & Computational Biology**

# Evolutionary Genomics

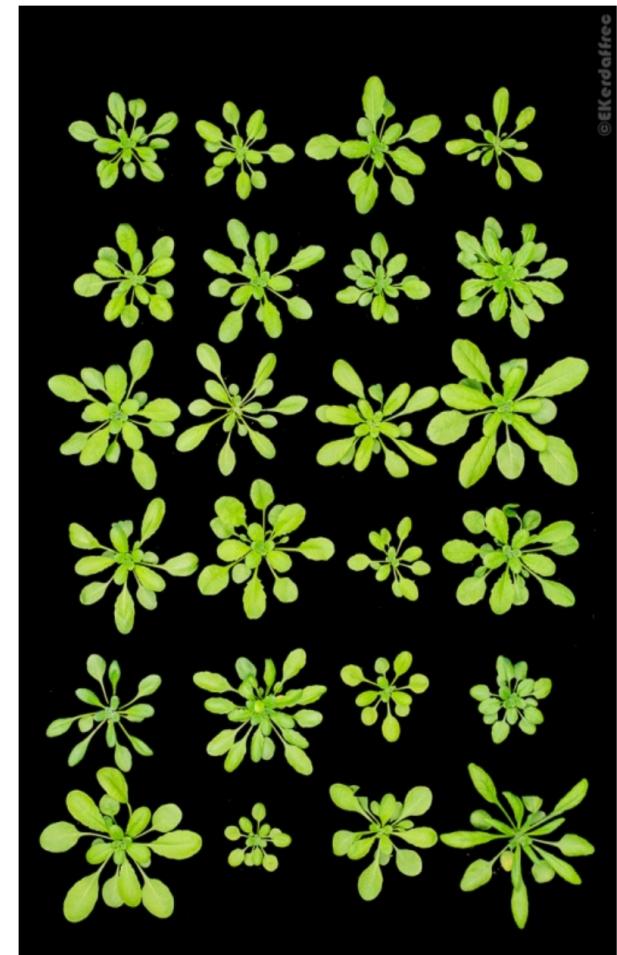
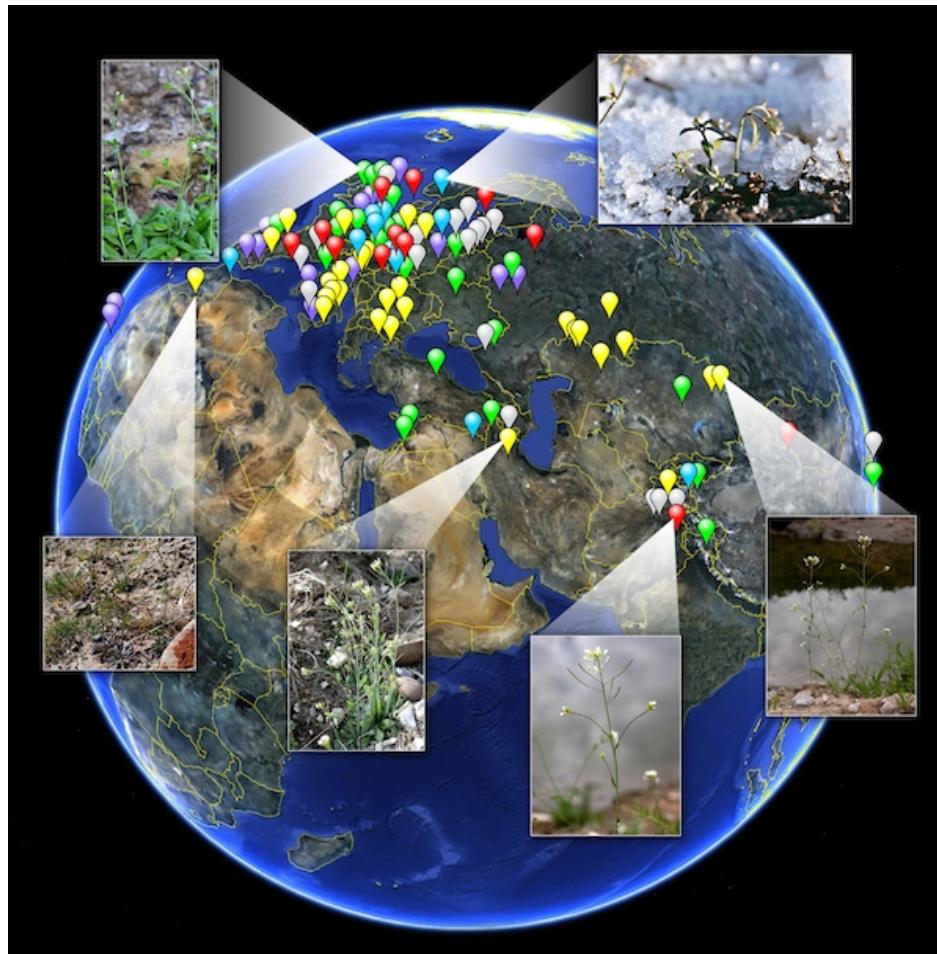


# How do plants adapt to drought



ACATTAA  
GCGCAATTCCAC  
TCAGCATAAGGAAGA  
GTGAGGGTCACATGGC  
CTCAGTTAGCGTACATCCTA  
GTCAAAAGTGCGGGTTTCG  
TTCAGAACATCATGACCTGCAC  
GAGCTCGACATGGCAATAAC  
GAGAACATGTGTAAACGAAGC  
AGGGGTCTGTGAATTGTTA  
GTGGGGGATAGTCACCG  
GAACAGGTTGGTTAA  
TACTGTAGTAA  
^ATCTTTAC'

# *Arabidopsis thaliana*



# The 1001 Genomes Project



[www.1001genomes.org](http://www.1001genomes.org)

## 1001 Genomes

A Catalog of *Arabidopsis thaliana* Genetic Variation

Home

Data Providers

Accessions

Tools

Software

Data Center

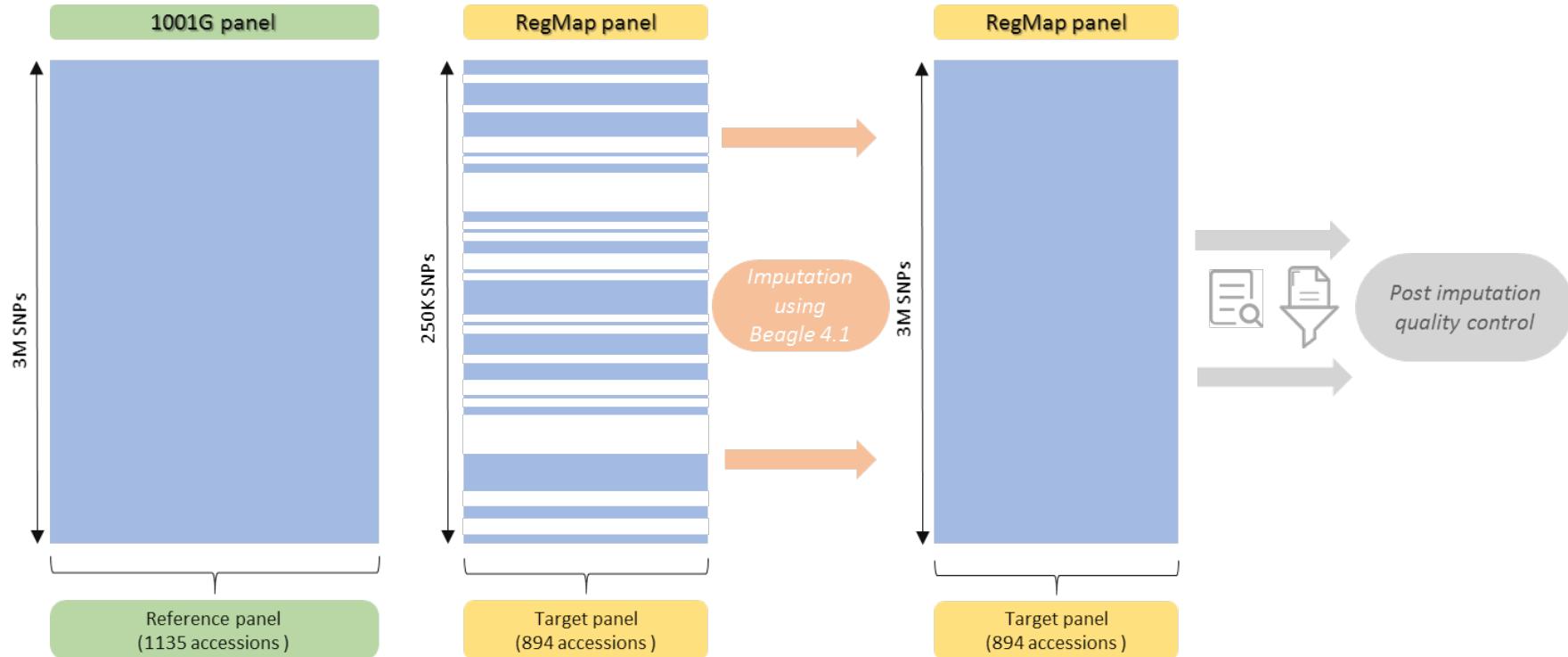
About

Welcome to the 1001 Genomes Project

**Data : 1,135 high quality genomes with more than 10 M SNP and  
500k structural variants**

- Robust (replicated) phenotypes in controlled environments (GxE)
- Follow-up studies

# Imputation of missing data



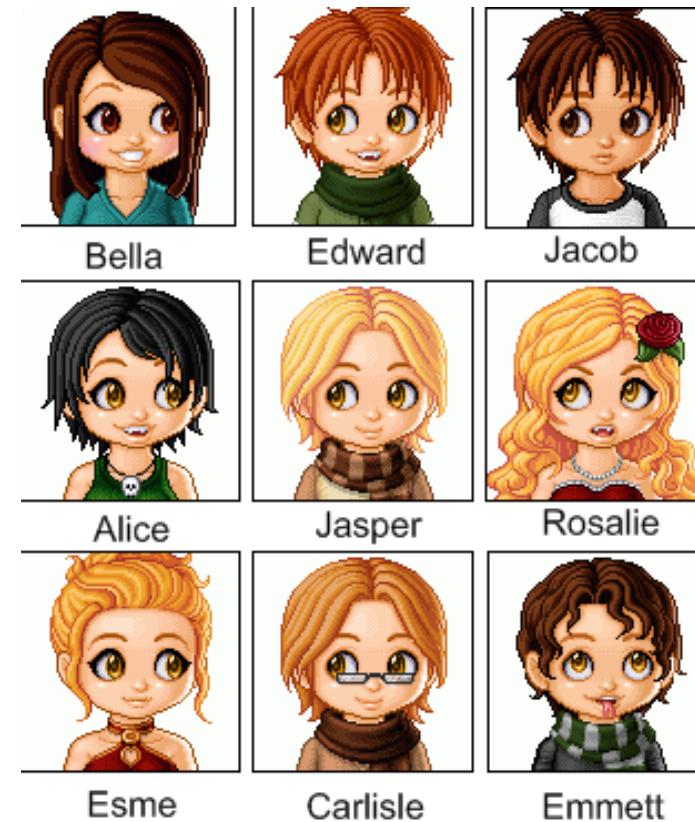
**High confidence data for 2029 accessions on ~2 M SNPs**

# Genome-wide association studies (GWAS)

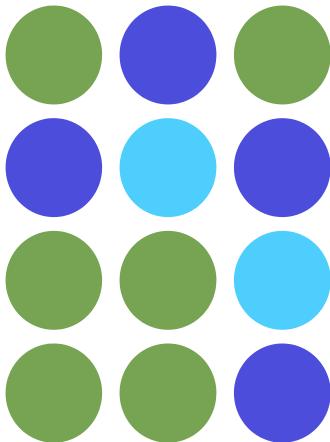


Associations between phenotype and genotype

# Genome-wide association studies (GWAS)



Associations between phenotype and genotype



# Introduction: GWAS

•••AGCCTG-----TGCACTAAGA**Ct**•••

•••AGCCTG-----TGCACTAAGA**Ct**•••

•••AGCCTG-----TGCACTAAGA**Gt**•••

•••AGCCTG-----TGCACTAAGA**Ct**•••

•••AGCCTGAGTGTGCACTAAGA**Gt**•••

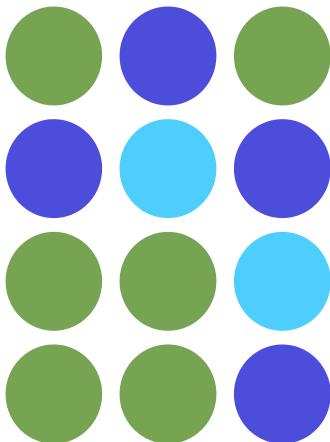
•••AGCCTGAGTGTGCACTAAGA**Gt**•••

•••AGCCTGAGTGTACTAAGA**Ct**•••

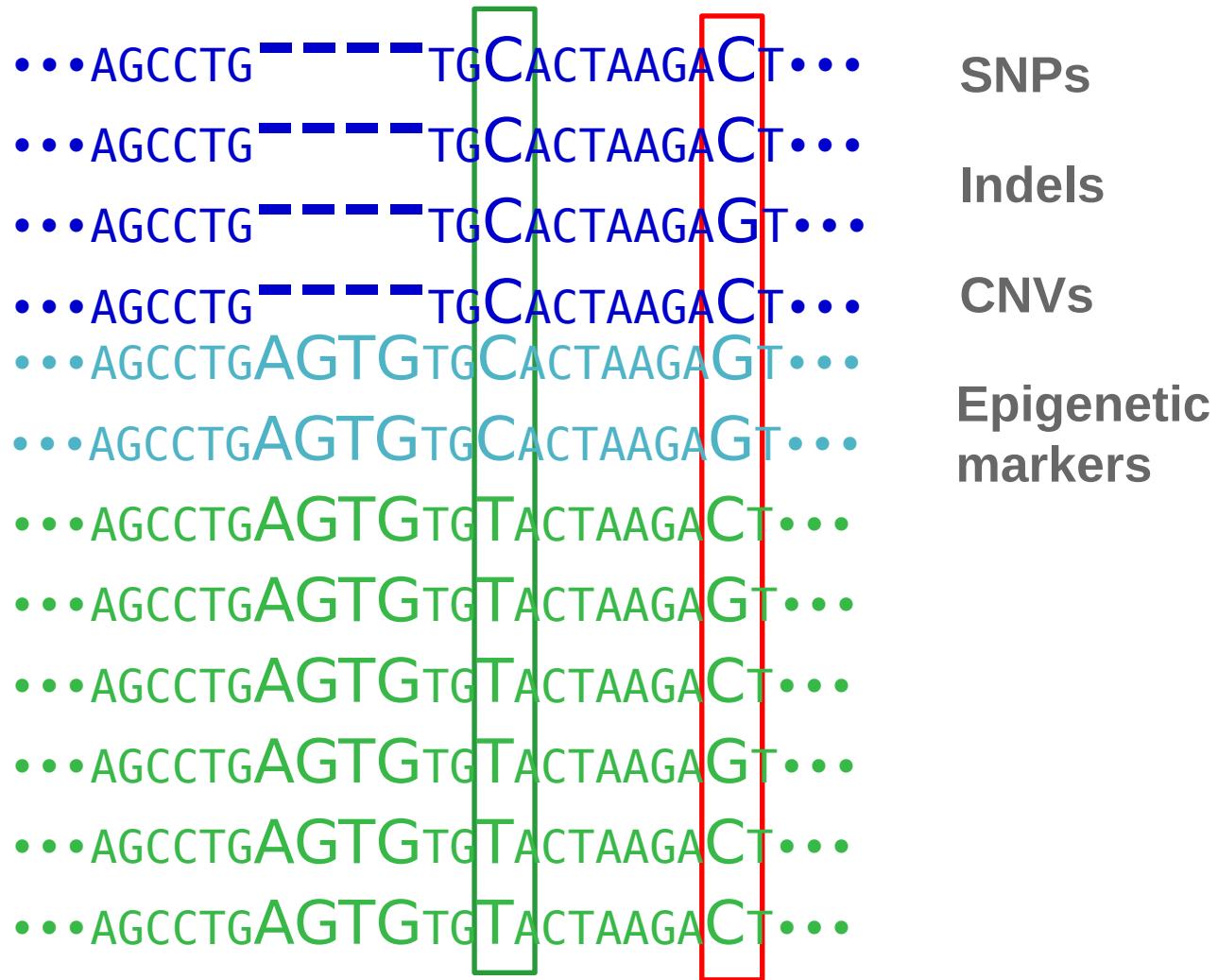
•••AGCCTGAGTGTACTAAGA**Gt**•••

•••AGCCTGAGTGTACTAAGA**Ct**•••

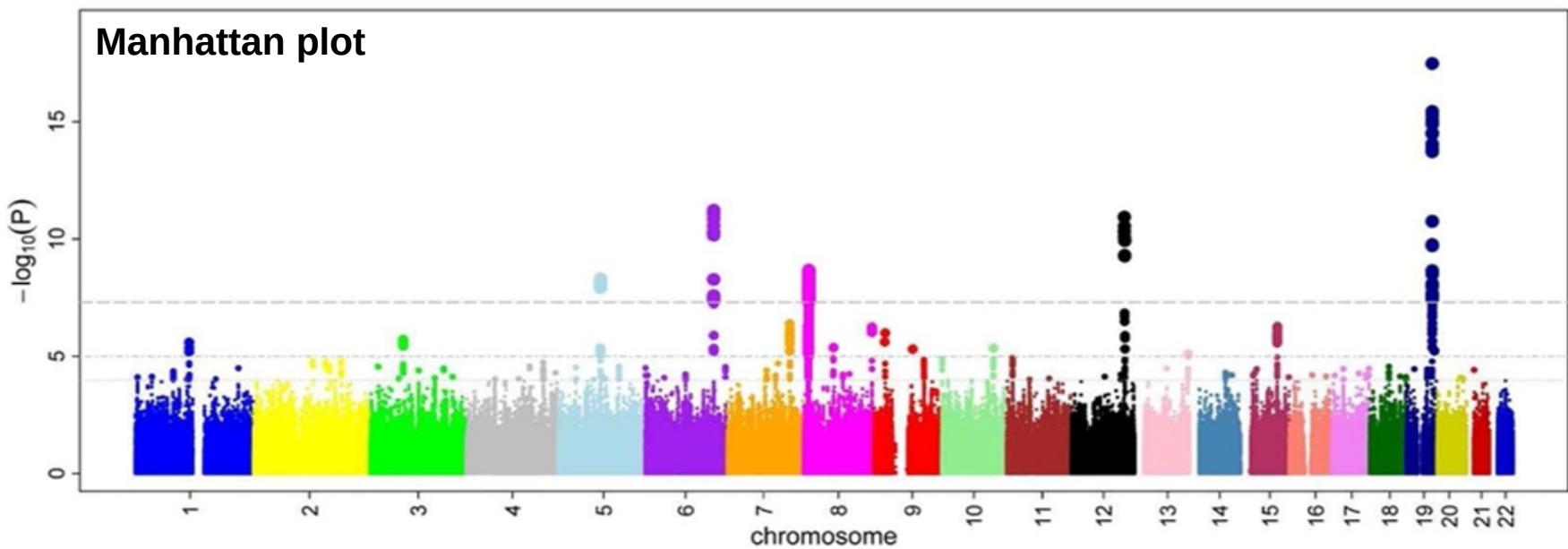
•••AGCCTGAGTGTACTAAGA**Ct**•••



# Introduction: GWAS



# GWAS on blood microcirculation

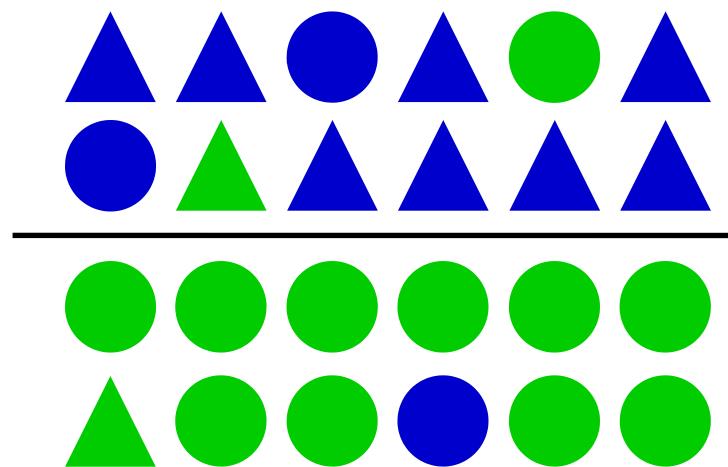


Ikram et al. (2010) PLoS Genet.

# Introduction: Population structure

**Confounding due to population structure may arise if it correlates with the trait in question**

Sub-population 1



Sub-population 2

**Mixed Models (introduced by R.A. Fisher (1918) and further developed by Henderson in the 1950s can handle different levels of relatedness**



## Introduction: GWAS

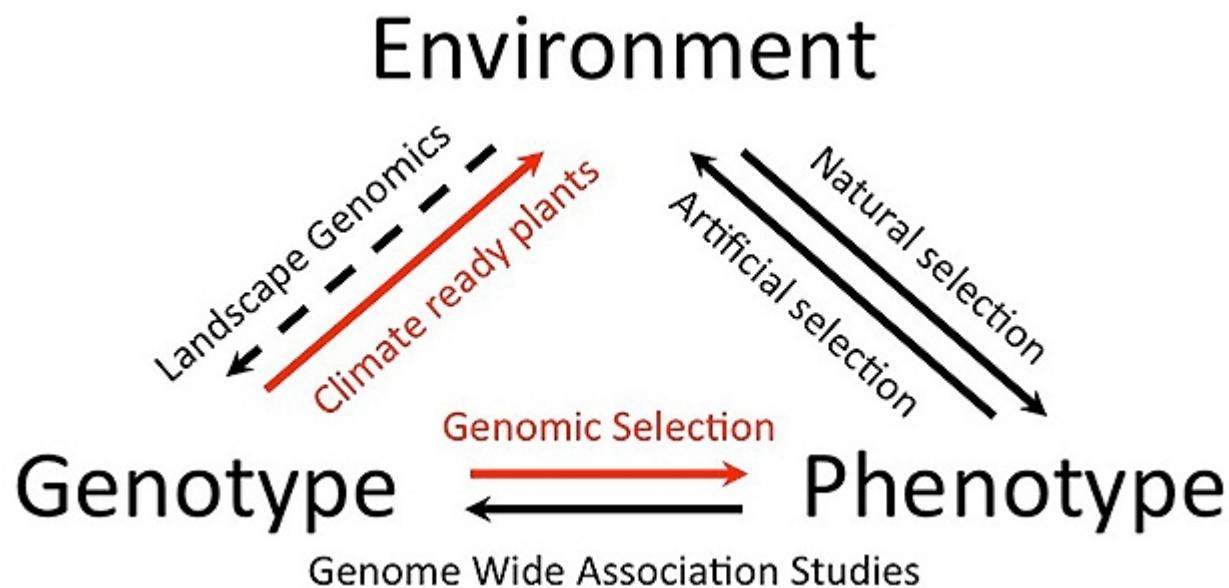
**GWAS is a powerful tool to detect genotype-phenotype association which are valuable for functional follow-up studies or trait prediction,**

**but commonly**

- only marginal effects of markers a tested
  - trait correlation is ignored
  - model assumptions might be violated

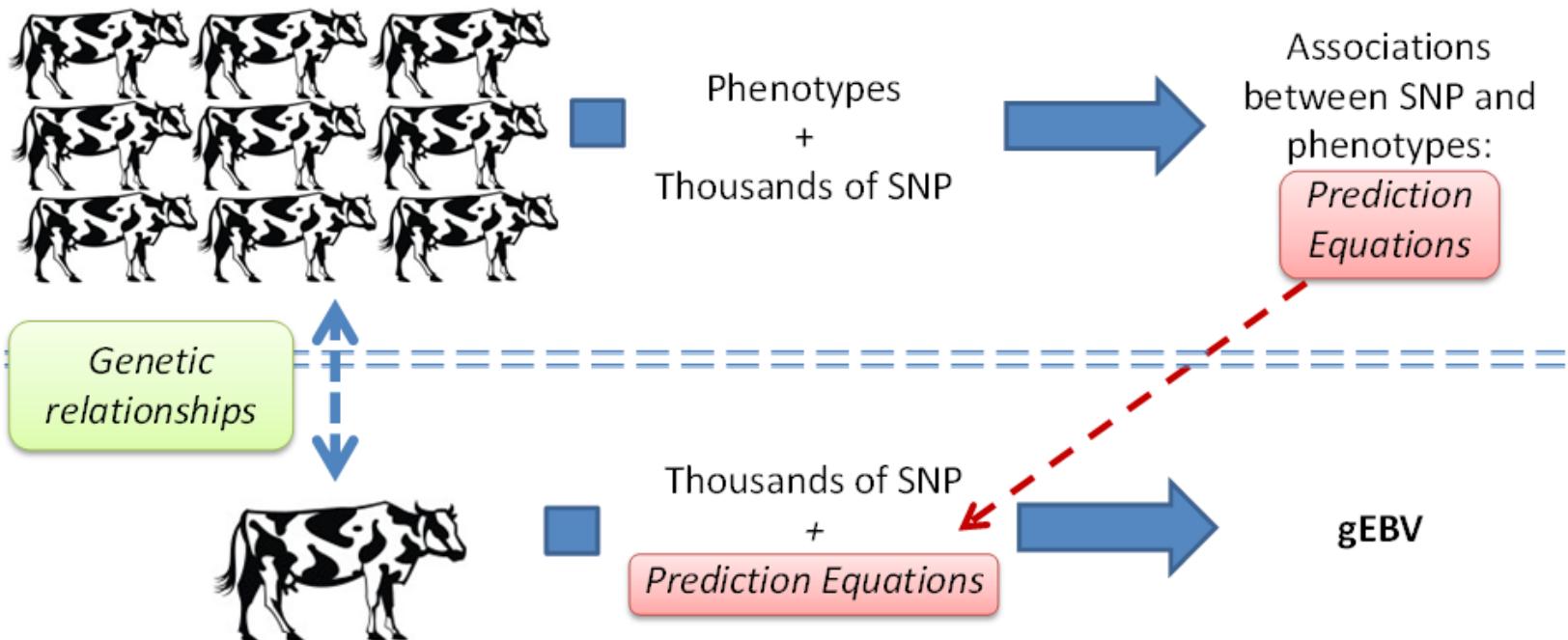
...

# Evolutionary Genomics



# Genomic Prediction

Reference population: Development of prediction equations



Main population: Application of prediction equations

# Genomic Prediction

1	1	1	1	0	0	1	1	1	1	1
0	1	1	1	1	1	1	1	0	0	1
1	1	1	1	1	1	1	1	1	1	0
1	1	0	1	0	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	1	1

→ phenotype value (e.g. height, flowering time ...)

SNP Matrix

$$\text{GWAS : } \mathbf{Y} = \beta_0 + \mathbf{X}\beta_1 + \mathbf{u} + \boldsymbol{\varepsilon}, \mathbf{u} \sim N(\mathbf{0}, \sigma_g \mathbf{K}), \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_e \mathbf{I})$$

$$\text{GP : } \beta_0 + \mathbf{X}\beta_1 + \mathbf{uZ} + \boldsymbol{\varepsilon} = \mathbf{Y}$$

Might be possible to learn the betas instead of using a Gaussian distribution

## Decomposition of the phenotypic variance

$$\sigma_P = \sigma_G + \sigma_E + \sigma_{GxE}$$

$$\sigma_G = \sigma_A + \sigma_D + \sigma_I$$

## Decomposition of the phenotypic variance

$$\text{GP : } \beta_0 + X\beta_1 + uZ + \varepsilon = Y$$

$$\sigma_G = \sigma_A + \sigma_I$$

# Decomposition of the phenotypic variance

$$\text{GP : } \beta_0 + X\beta_1 + uZ + \varepsilon = Y$$

$$\sigma_G = \sigma_A + \sigma_I$$



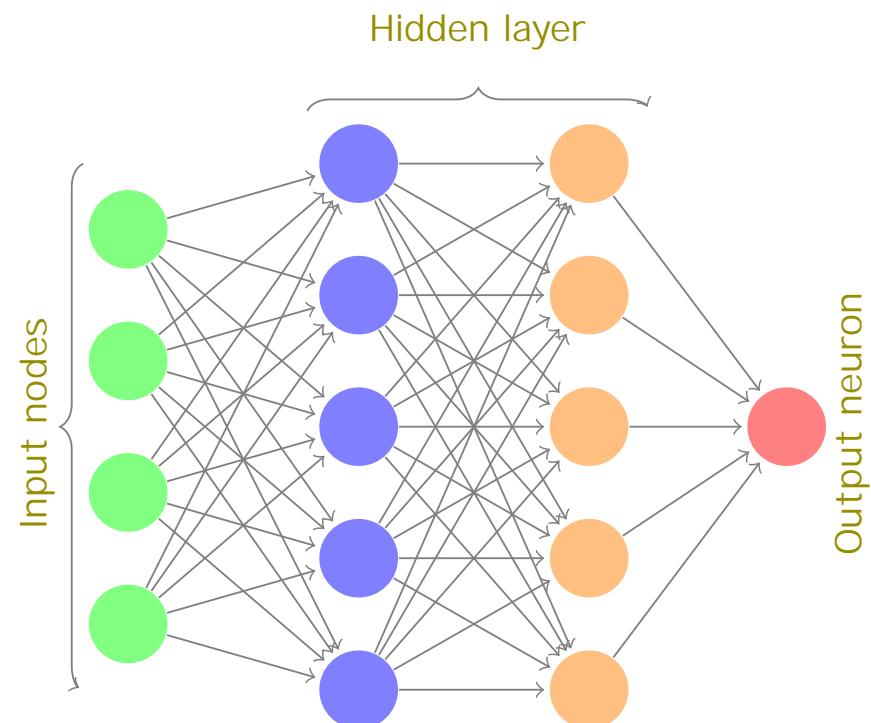
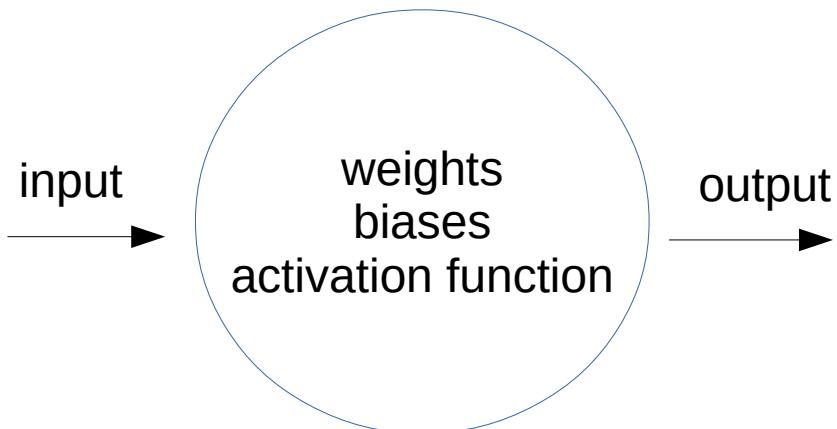
**perfect for linear models**

**Non-linear interaction,  
perfect for ML**

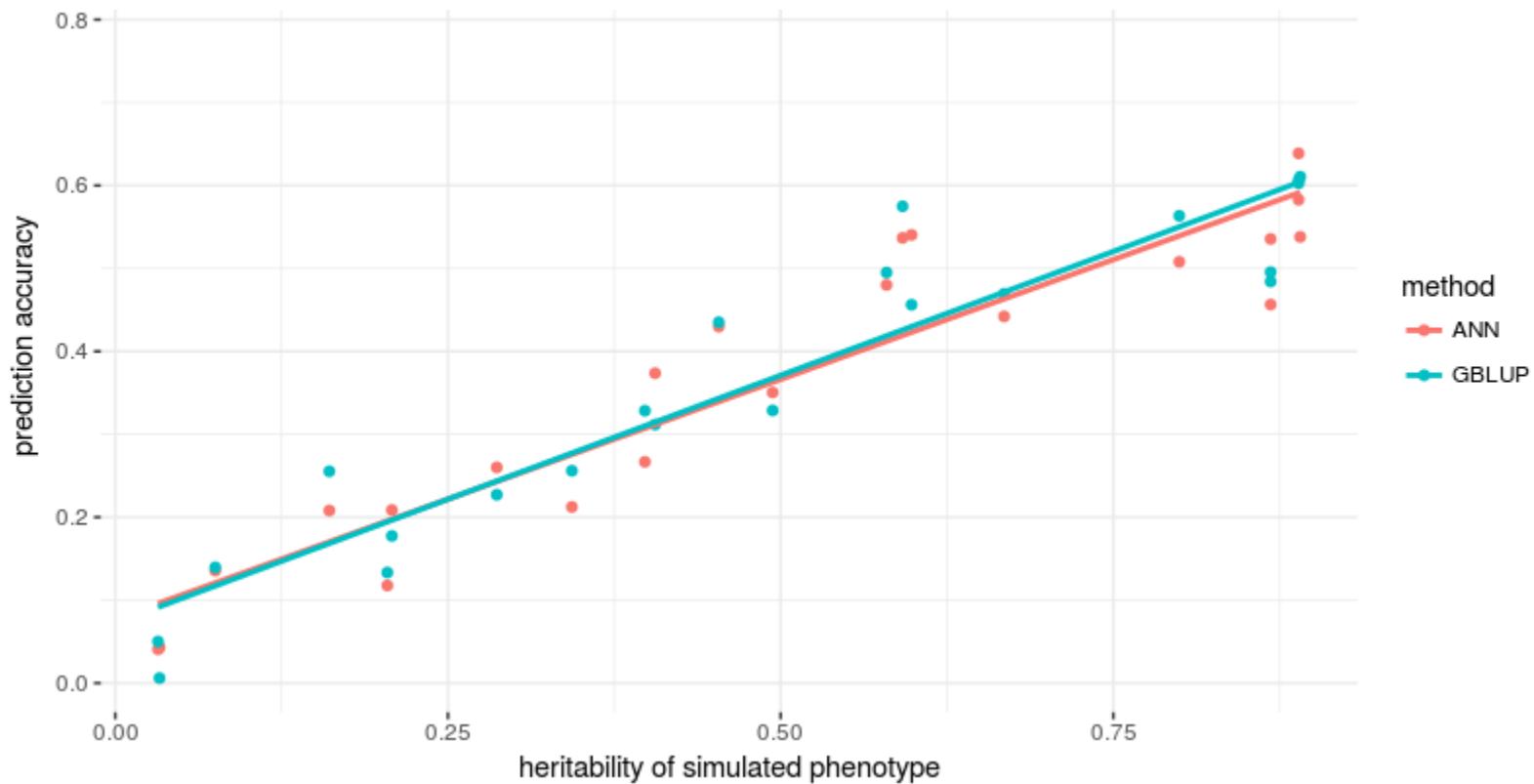
# ML for Genomic Prediction



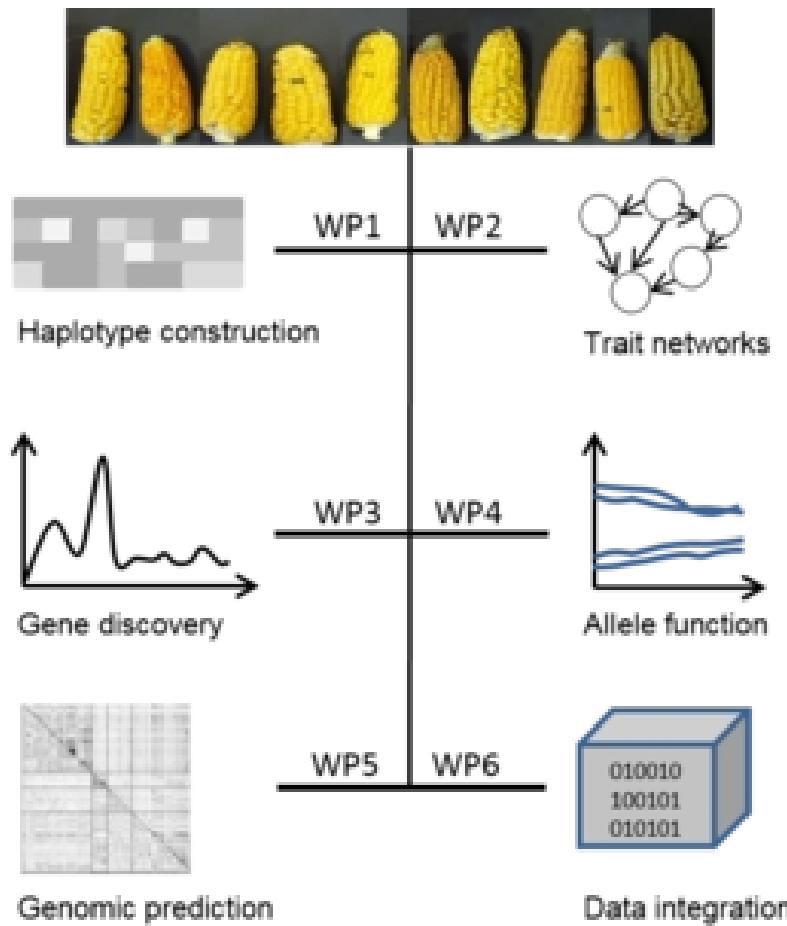
**Artifical neuron**



# ML for Genomic Prediction



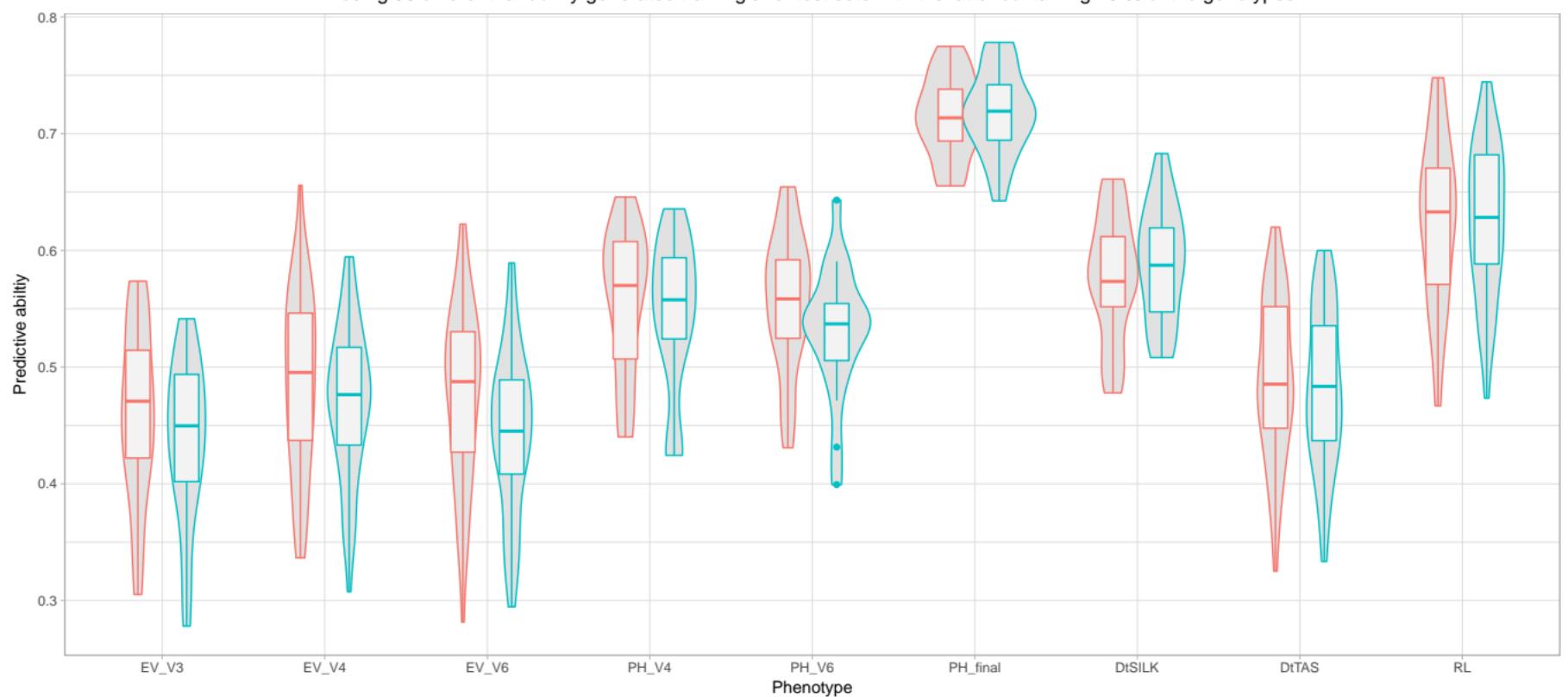
# MAZE - Assessing the genomic and functional diversity of maize to improve quantitative traits



SPONSORED BY THE

# ML for Genomic Prediction

Predictive abilities of bayesian ridge regression (BRR) and artificial neural networks (ANN)  
in a DH-population derived from Kemater landmaize  
using 50 different randomly generated training and test sets with the latter containing 20 % of the genotypes



# Using TensorFlow for GWAS



Cold  
Spring  
Harbor  
Laboratory



THE PREPRINT SERVER FOR BIOLOGY

HOME | A  
| CHANNELS

Search

New Results

Comment on this paper

## GWAS-Flow: A GPU accelerated framework for efficient permutation based genome-wide association studies

Jan A. Freudenthal, Markus J. Ankenbrand, Dominik G. Grimm, Arthur Korte

doi: <https://doi.org/10.1101/783100>

This article is a preprint and has not been certified by peer review [what does this mean?].

Abstract

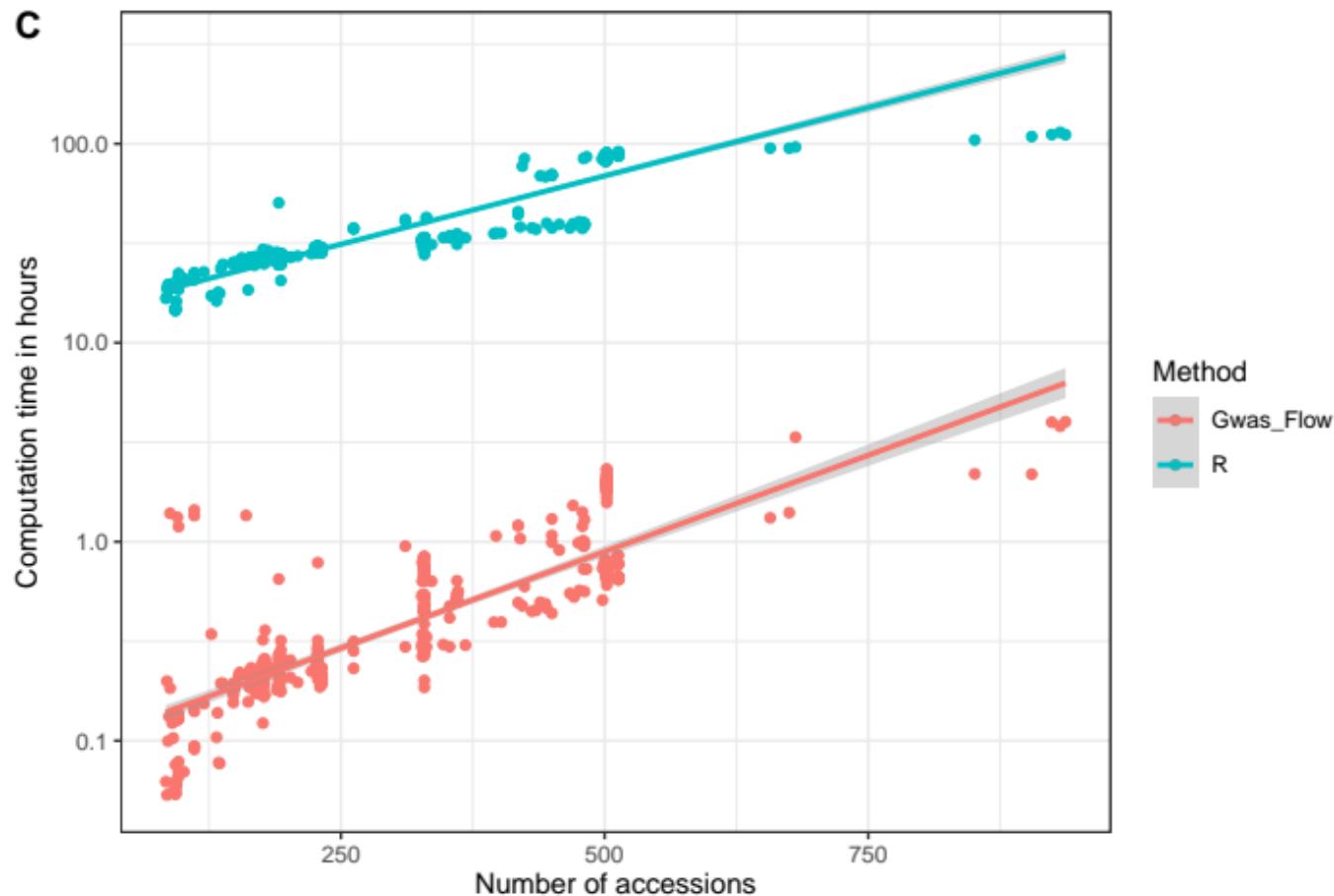
Full Text

Info/History

Metrics

Preview PDF

# GWAS-Flow



# AraPHENO and AraGWAS

AraPheno  AraRNASeq [Data](#) [Tools](#) [About](#)

AraPheno

AraPheno is a public database collection of *Arabidopsis thaliana* phenotypes. This Database allows to search and filter for public phenotypes and to obtain additional meta-information.

Search a phenotype, study, trait ontology (e.g. type FRI for phenotype, Atwell for study, or concentration for ontology) or accession name

**Quick Stats**

- 22 Studies
- 471 Phenotypes total
- 462 Phenotypes published
- Last Update: Sep/04/2019

## AraGWAS Catalog

[TAKE A TOUR?](#) [FAQ](#) [ABOUT](#)

AraGWAS Catalog

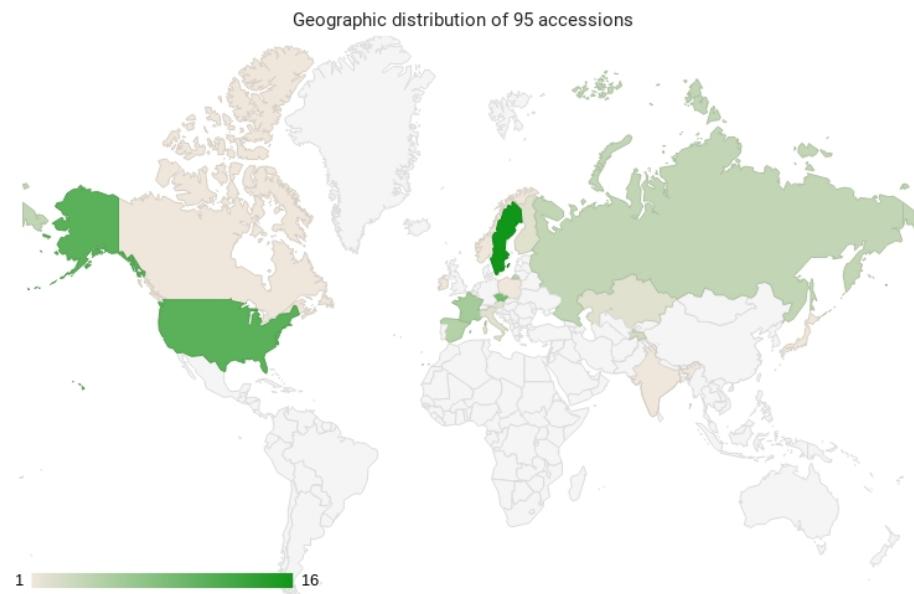
The AraGWAS Catalog is a public and manually curated database for standardised GWAS results for *Arabidopsis thaliana*. This Database allows to search and filter for public GWASs, phenotypes and genes and to obtain additional meta-information. All GWASs were recomputed following a uniformed methodology to allow for comparable results.

# Permutation-based thresholds

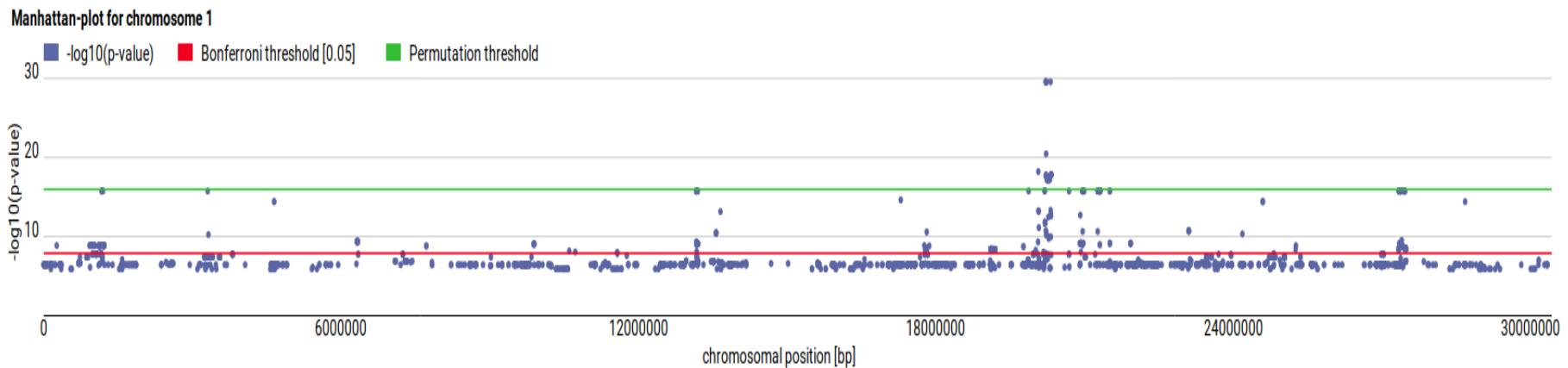
Phenotype: YEL



ID:	28
Study:	<a href="#">Atwell et. al, Nature 2010</a>
DOI:	<a href="#">10.21958/phenotype:28</a> ( Cite)
AraGWAS	<a href="#">AraGWAS:28</a>
Associations:	
Scoring:	Yellowing is a precocious, more diffused leaf chlorosis. Accessions were divided in three subgroups (mild, intermediate and severe) according to the severity of the yellowing.
Growth conditions:	None
Trait-Ontology:	<a href="#">leaf chlorosis (TO:0006060)</a>
Environment ontology:	<a href="#">growth chamber study (EO:0007269)</a>
Unit:	<a href="#">dimensionless unit (UO:0000186)</a>
Shapiro-Wilk*:	'4.46e-19' (p-value)

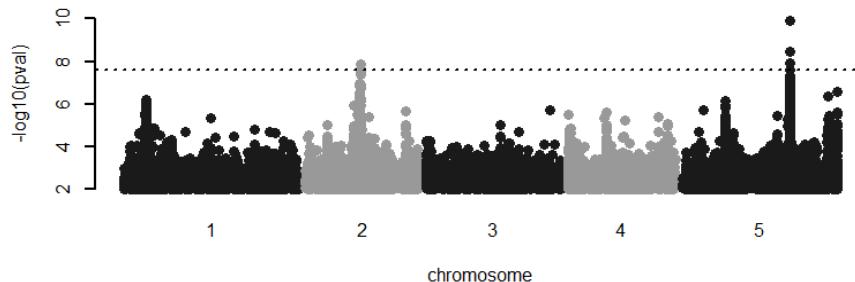
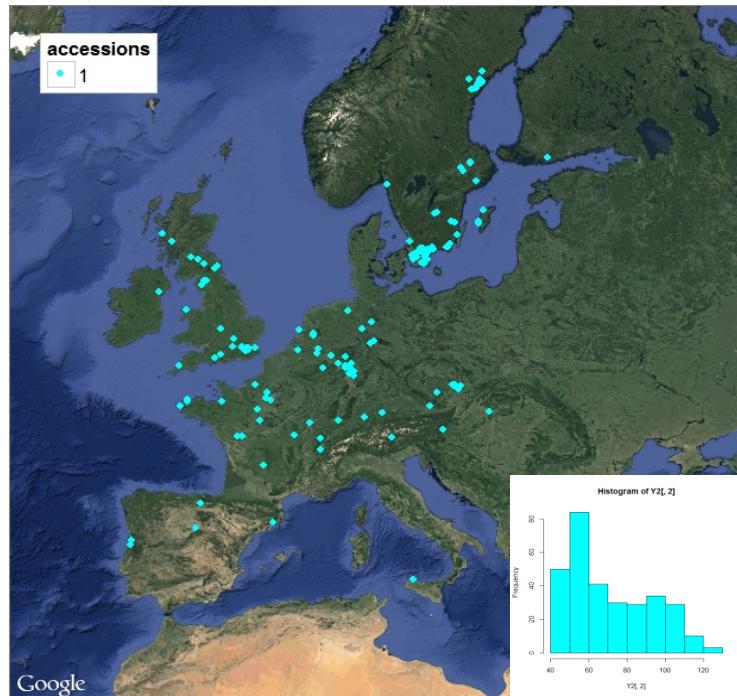


# Permutation-based thresholds

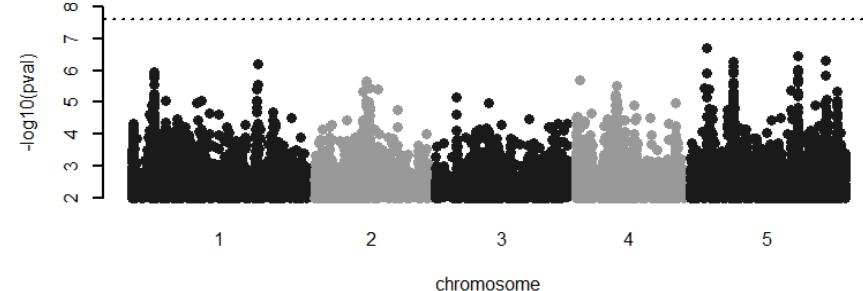
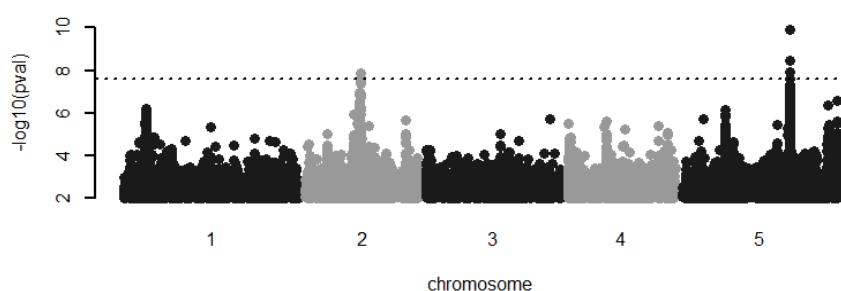
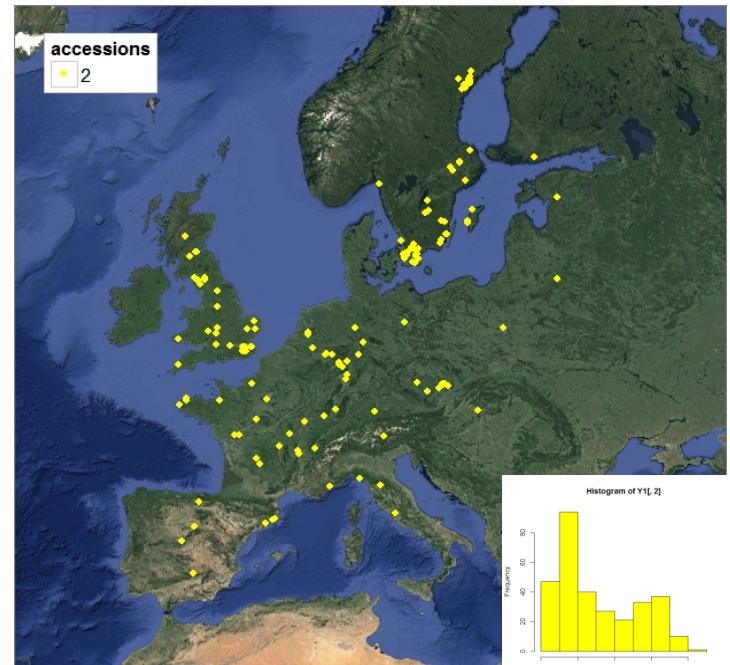
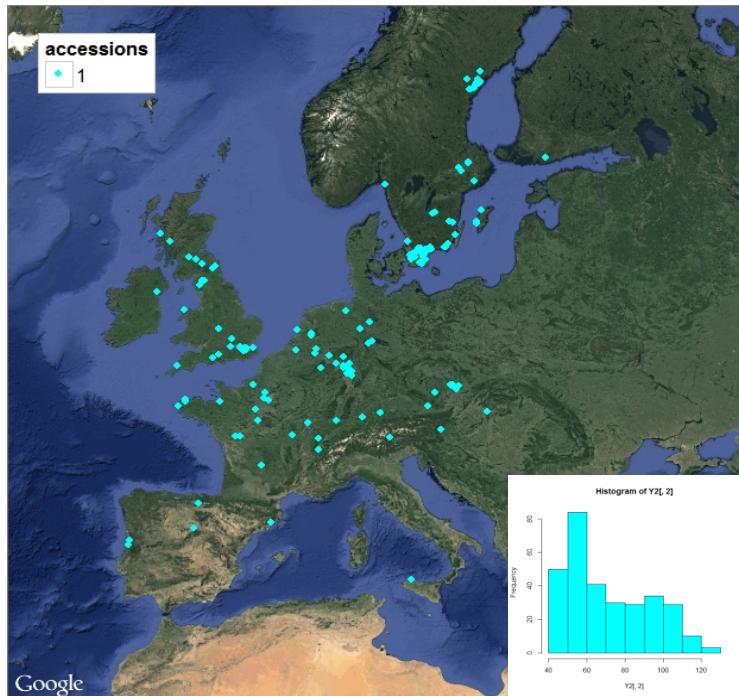


Permutation-based threshold can provide a more realistic significance threshold that depend on the phenotypic distribution

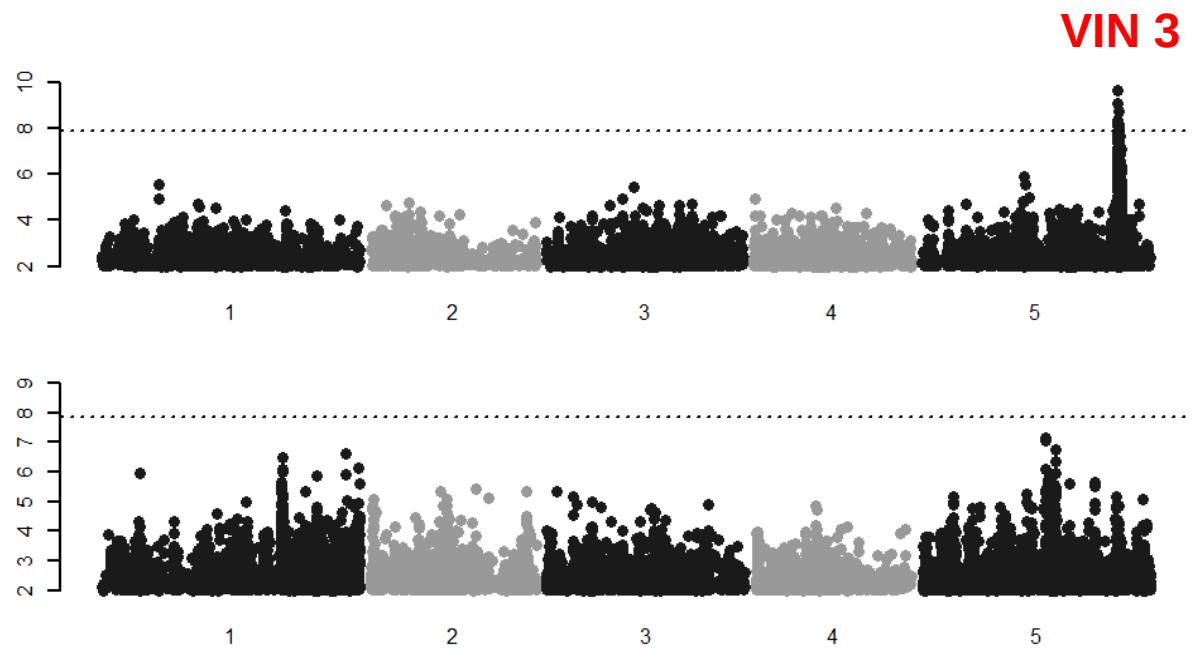
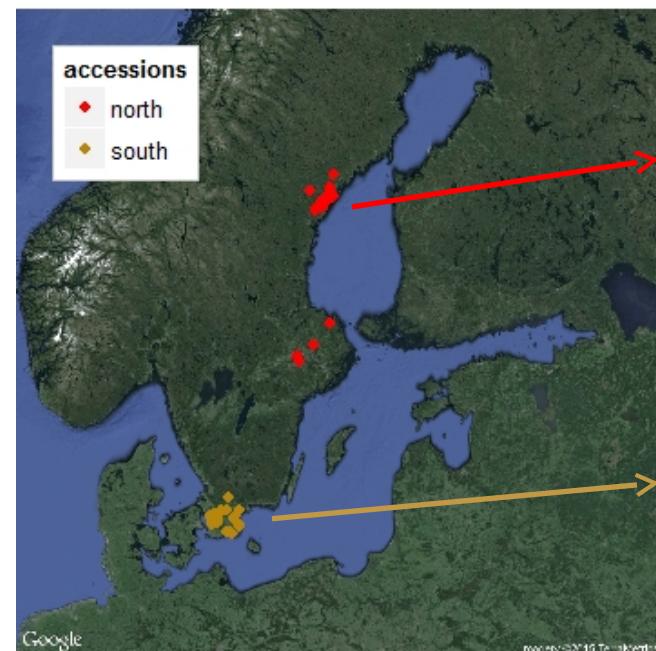
# GWAS on flowering time in *A.thaliana*



# GWAS on flowering time in *A.thaliana*

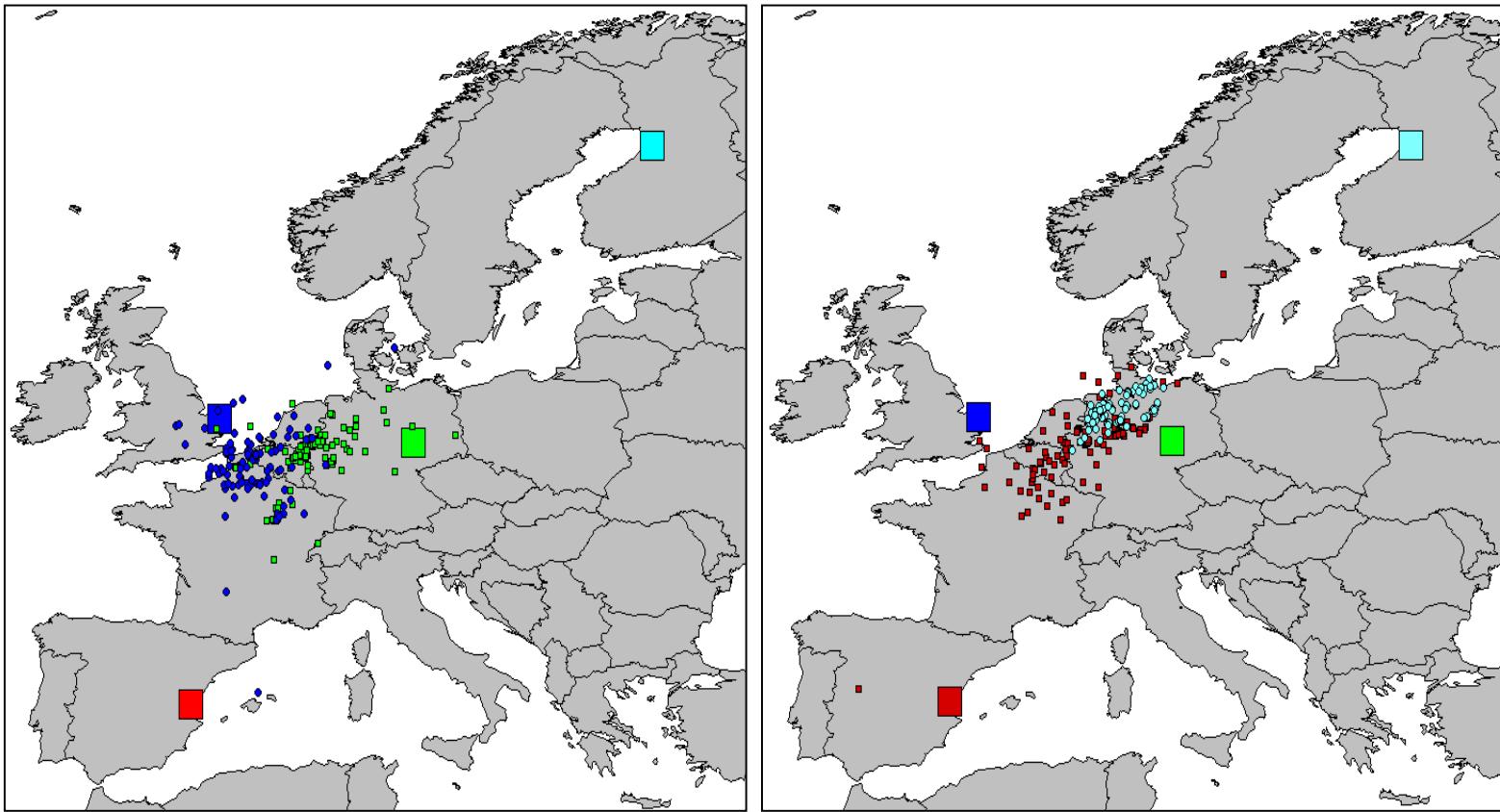


# GWAS in local subsets



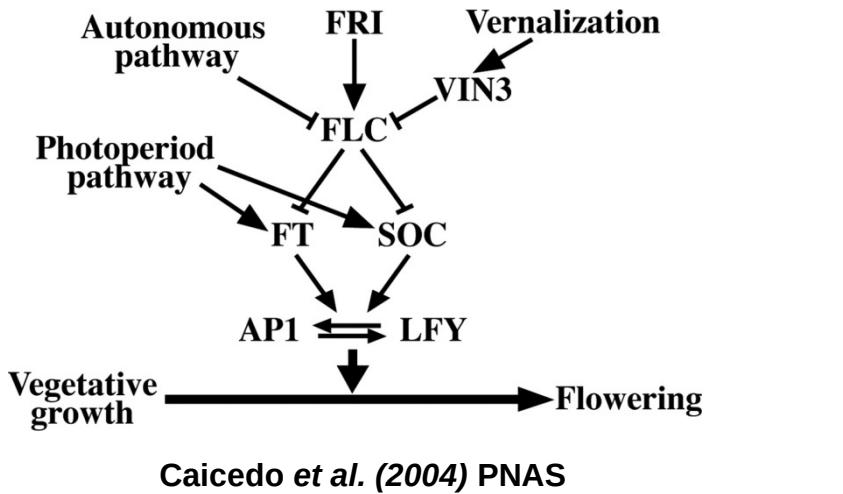
Different results in different subsets

# Local Adaptation



Genes providing fitness under natural conditions are mostly local

# GWAS in local subsets



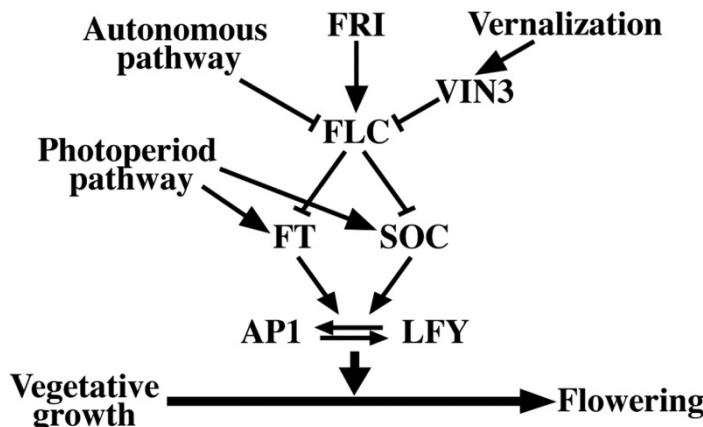
VIN 3



## Reasons

- Differences in Allele frequency of the causative marker
- The associated SNP is tagging different haplotypes
  - Epistasis

# GWAS in local subsets



Caicedo et al. (2004) PNAS

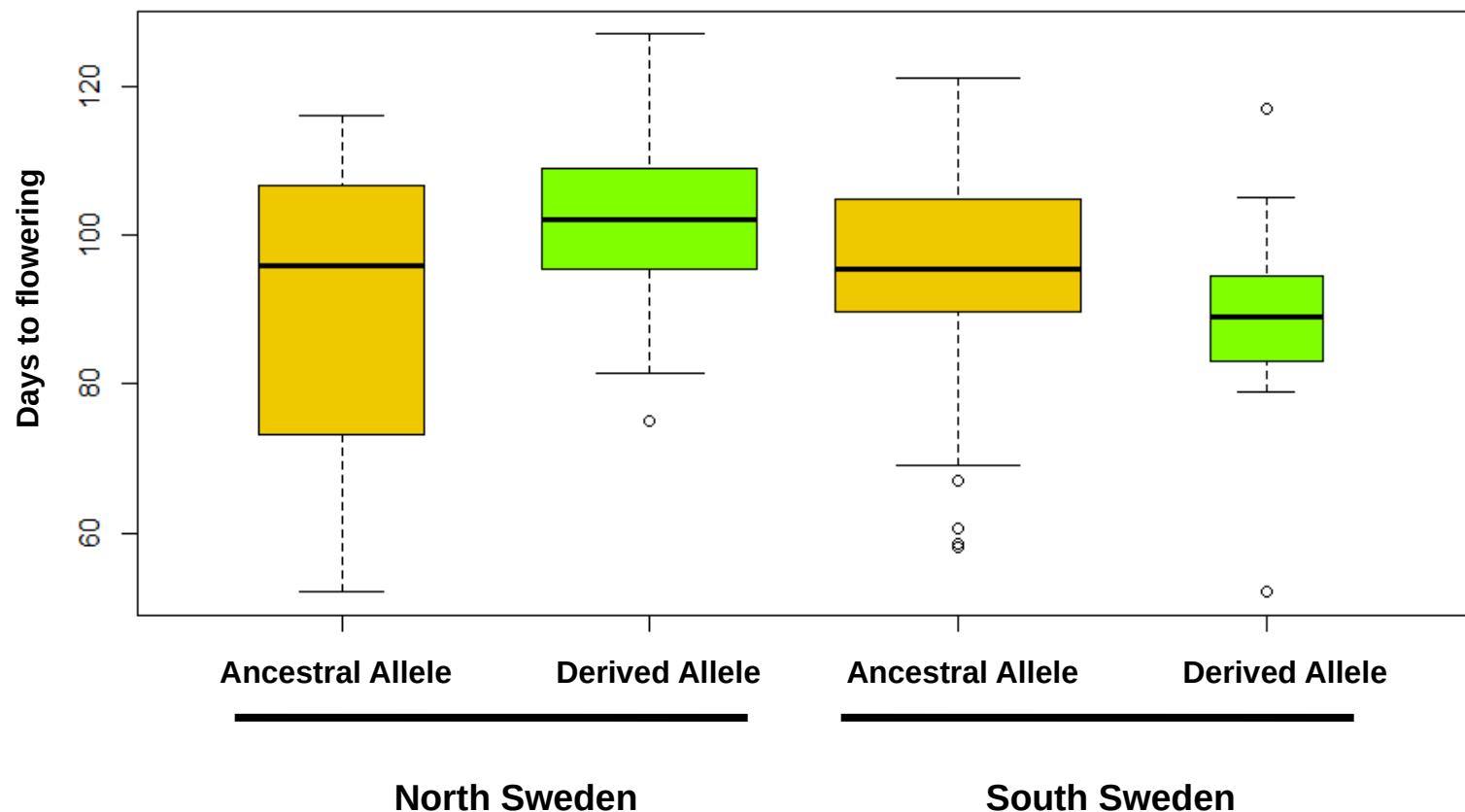
## Reasons

- Differences in Allele frequency of the causative marker
- The associated SNP is tagging different haplotypes
- Epistasis

$$\text{GWAS} : Y = \beta_0 + X\beta_1 + u + \epsilon$$

$$\text{GWAS} : Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \dots + X_1X_2\beta_{12} + u + \epsilon$$

# Effect of the *VIN3* Allele on Flowering time



**The effect of the respective allele depends on the genetic background**

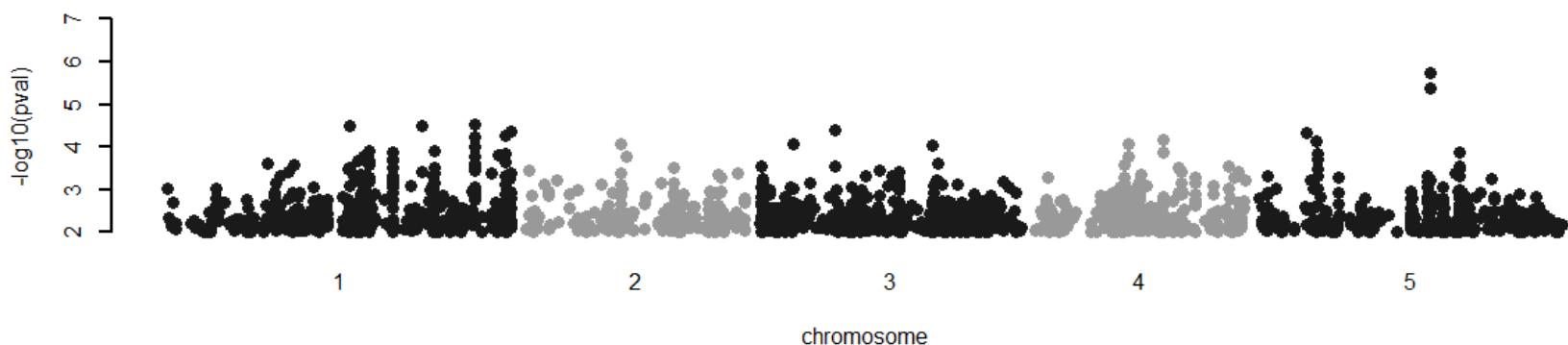
# GWAS models to test for epistasis

**SNP by SNP epistasis**

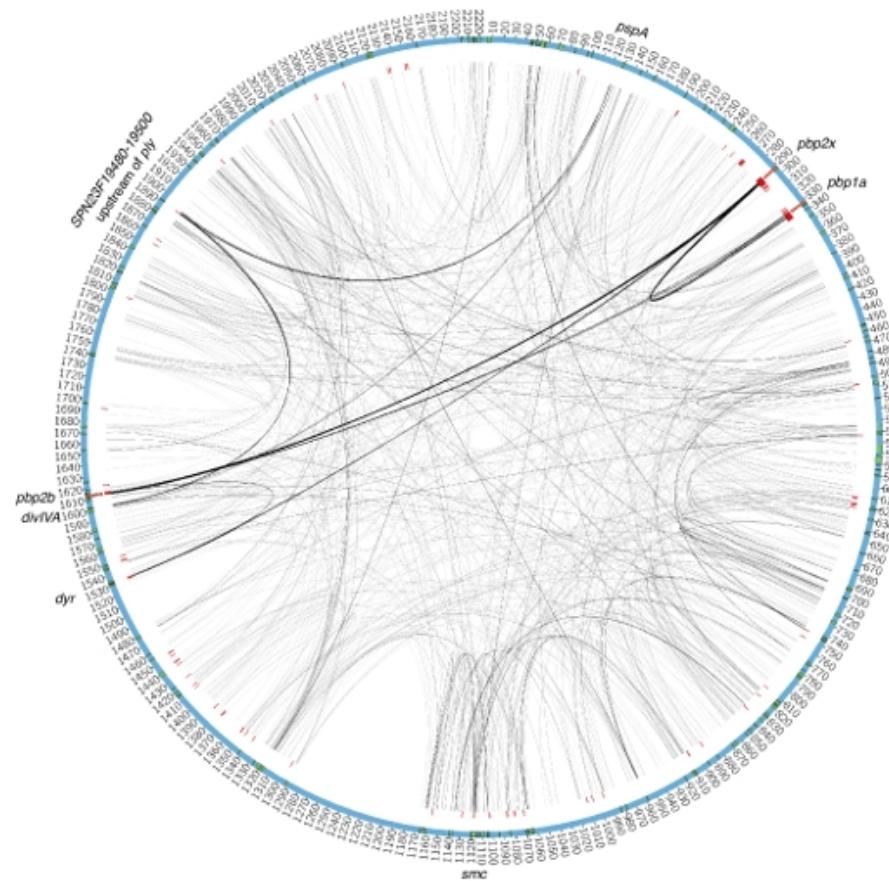
**Genome-wide epistasis with the lead SNP in *VIN3***

**3 different models with two homozygous SNPs**

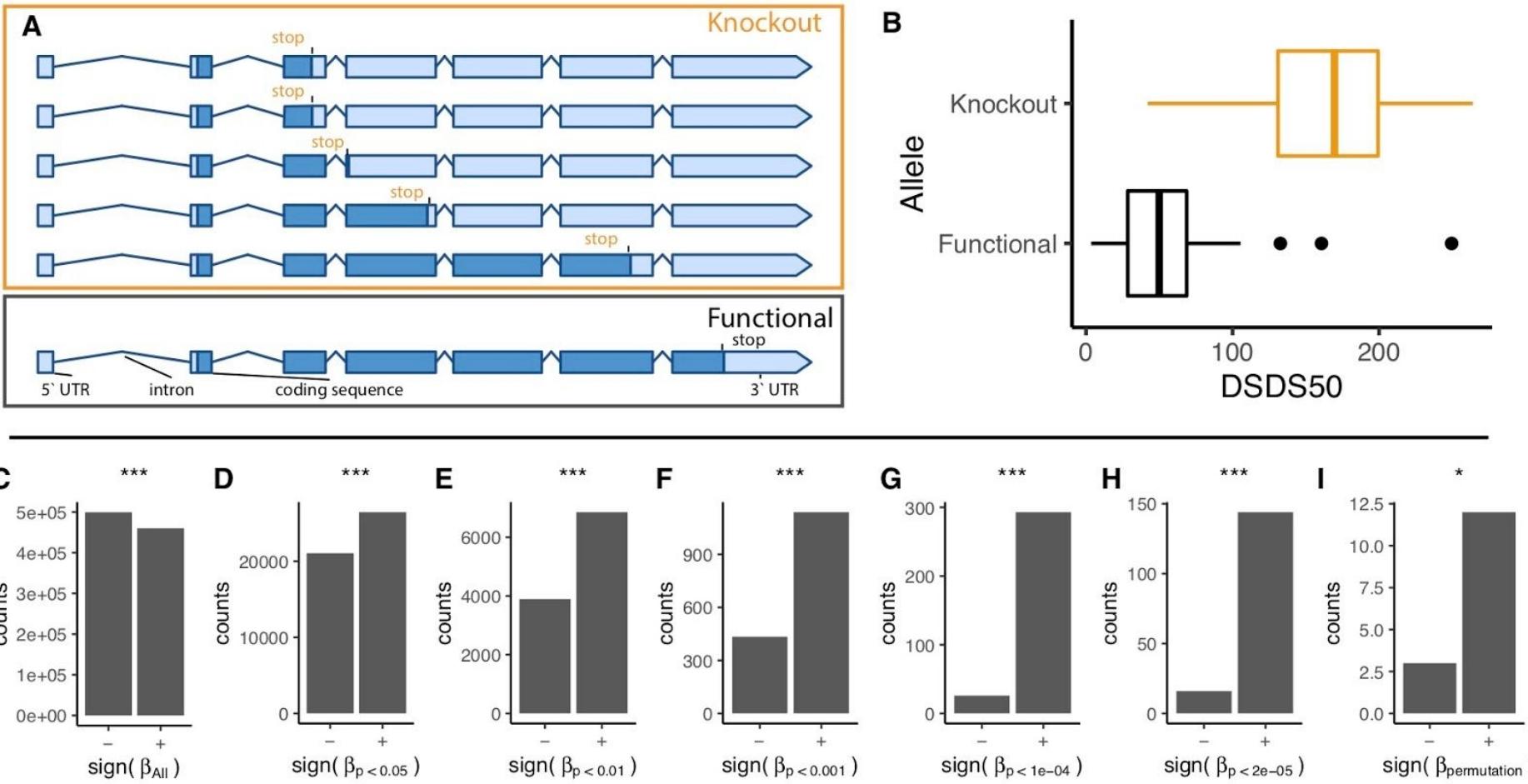
**AND, OR and XOR**



# Epistatic interactions in a bacterial genome

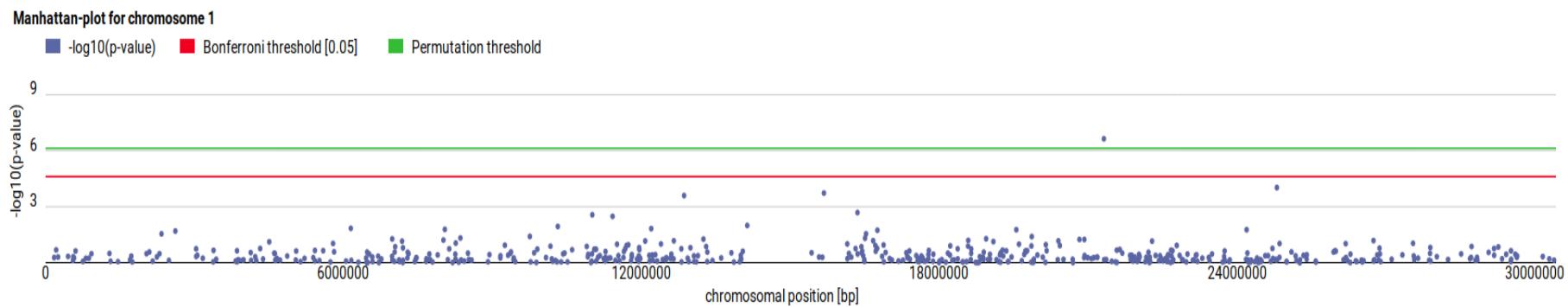


# Association with ko-mutations



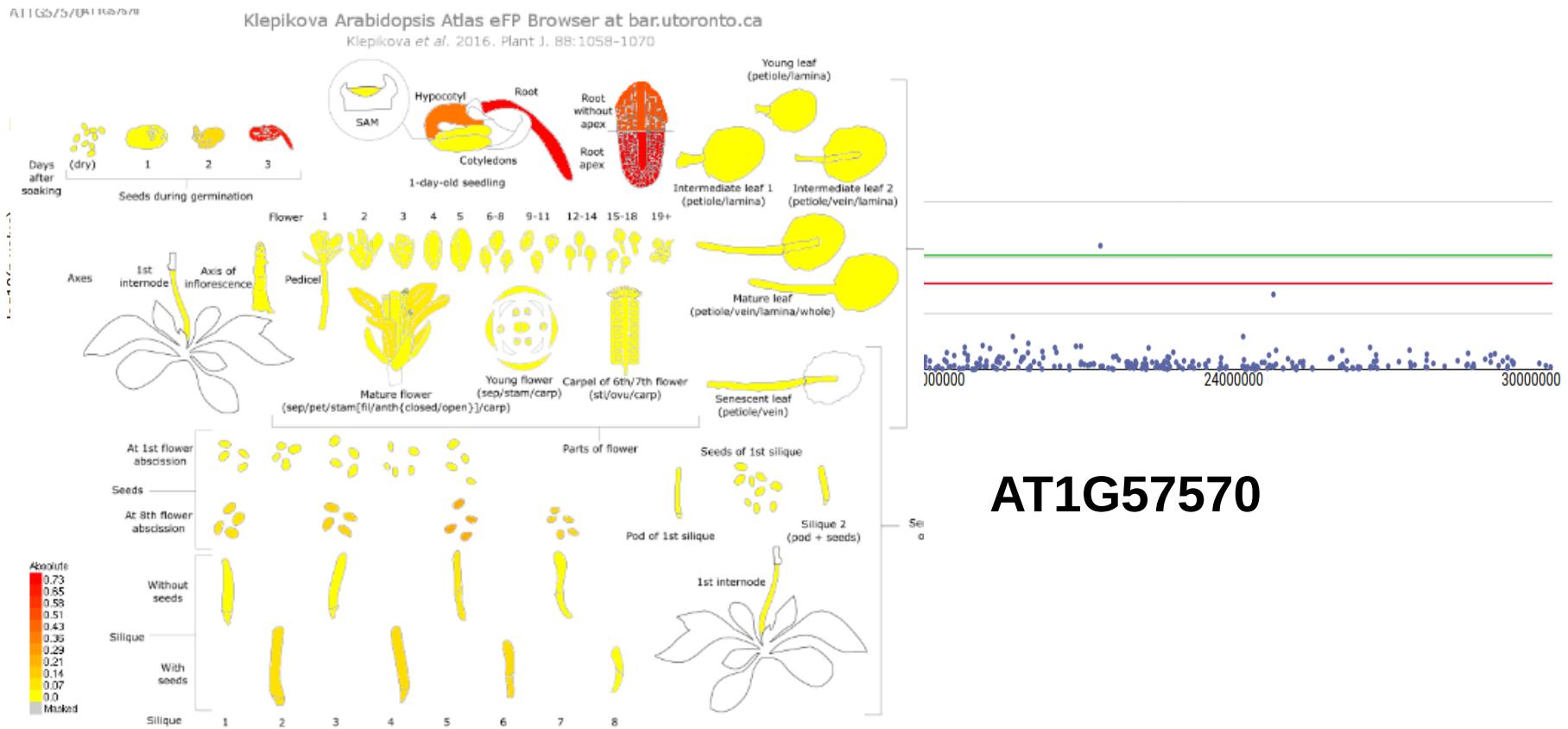
# Association with ko-mutations

**DSDS50:**Number of days of seed dry storage required to reach 50% germination (Alonso-Blanco et al., 2003)



# Association with ko-mutations

**DSDS50:**Number of days of seed dry storage required to reach 50% germination (Alonso-Blanco et al., 2003)



# Genomic variation in the *A.thaliana* population (1135 accessions)

10,709,466 SNPs segregate in the population

1,854,599 SNPs are located in coding regions

28,148 SNPs lead to a premature STOP codon



Nearly 10,000 genes are knock-out  
in at least 1 accession

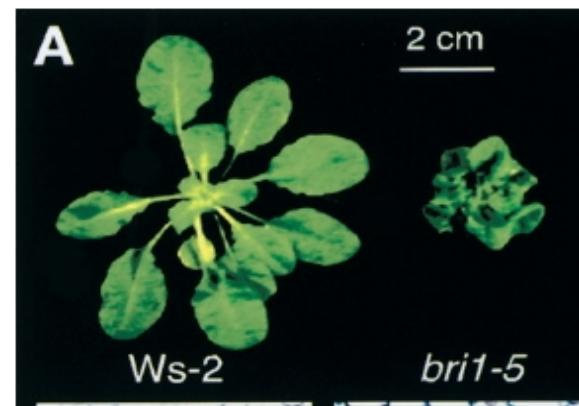
# Natural BRI1 knock out in an accession from Portugal (IP-Alo-0)



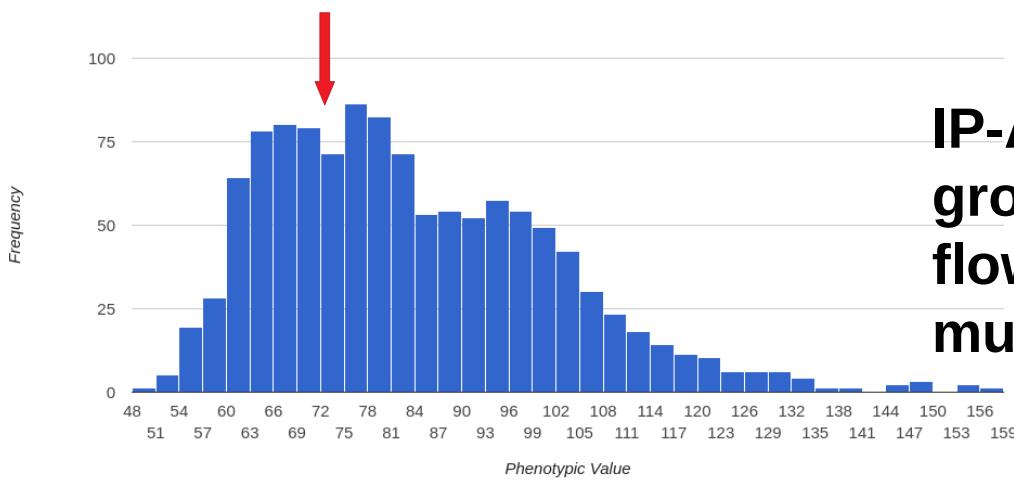
# Natural BRI1 knock out in an accession from Portugal (IP-Alo-0)



# Natural *BRI1* knock out in an accession from Portugal (IP-Alo-0)

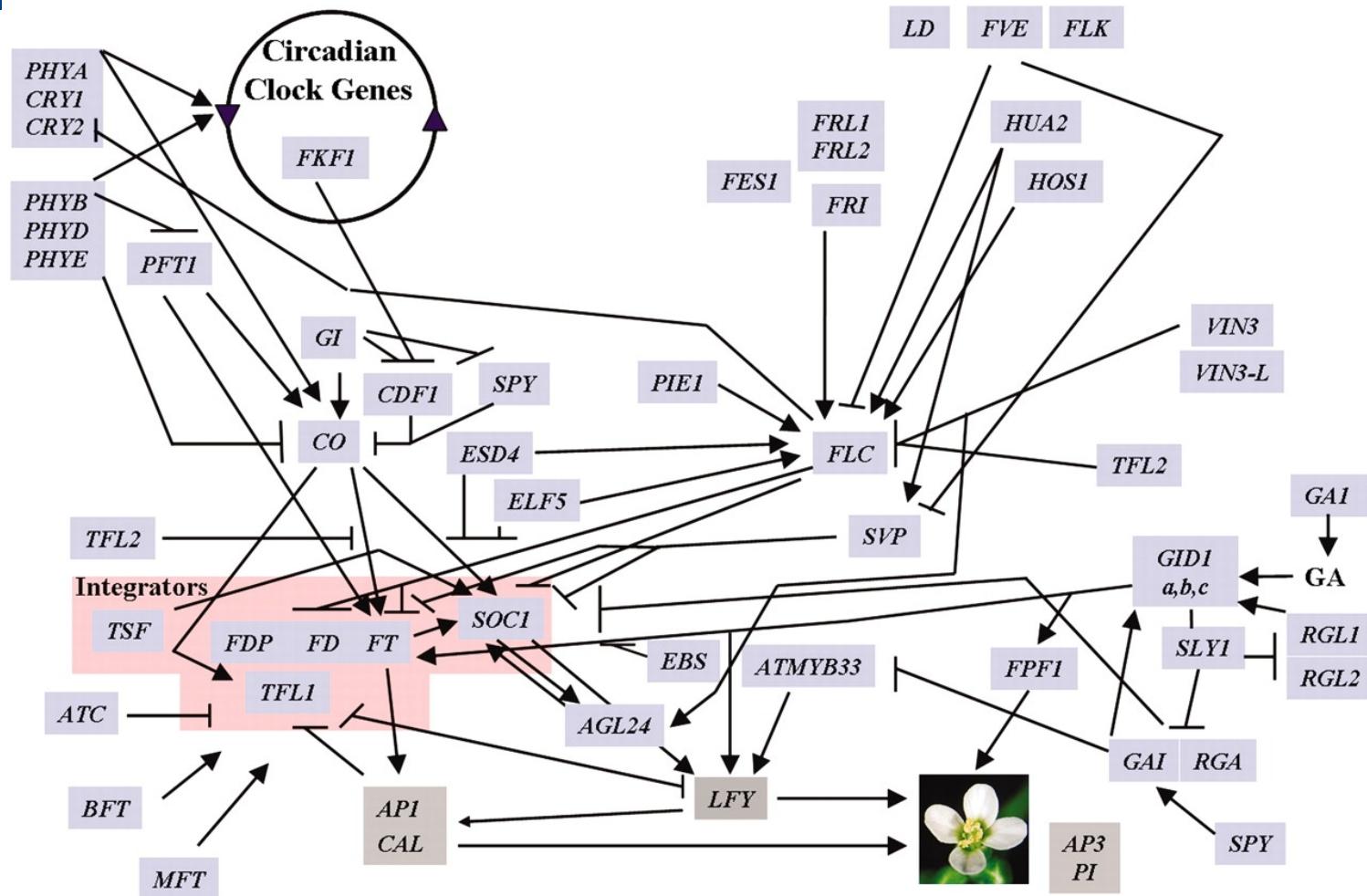


Santiago Mora-García et al. 2004, Genes Dev.

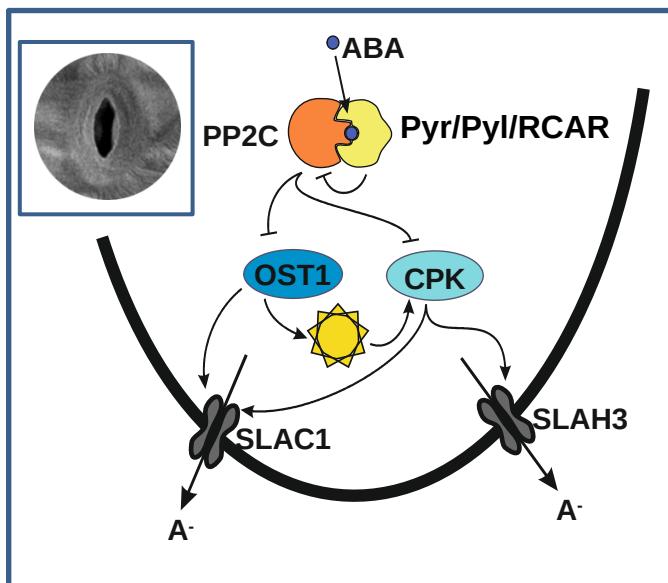


**IP-Alo-0 does neither show the growth defect nor the delay in flowering time observed in *bri1* mutants**

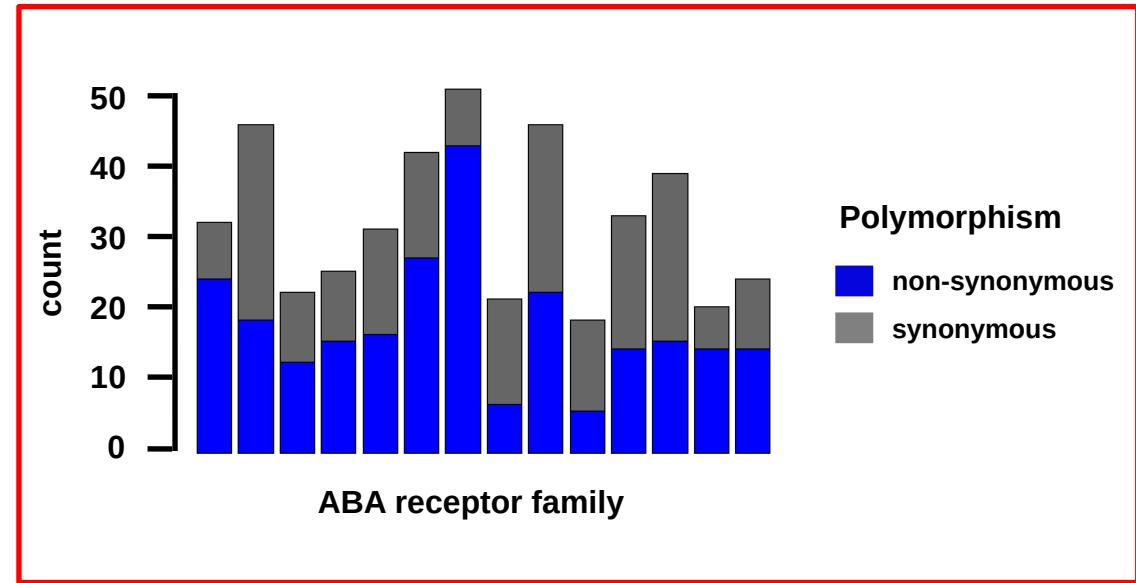
# Flowering time pathway in *A.thaliana*



# Guard Cell signalling



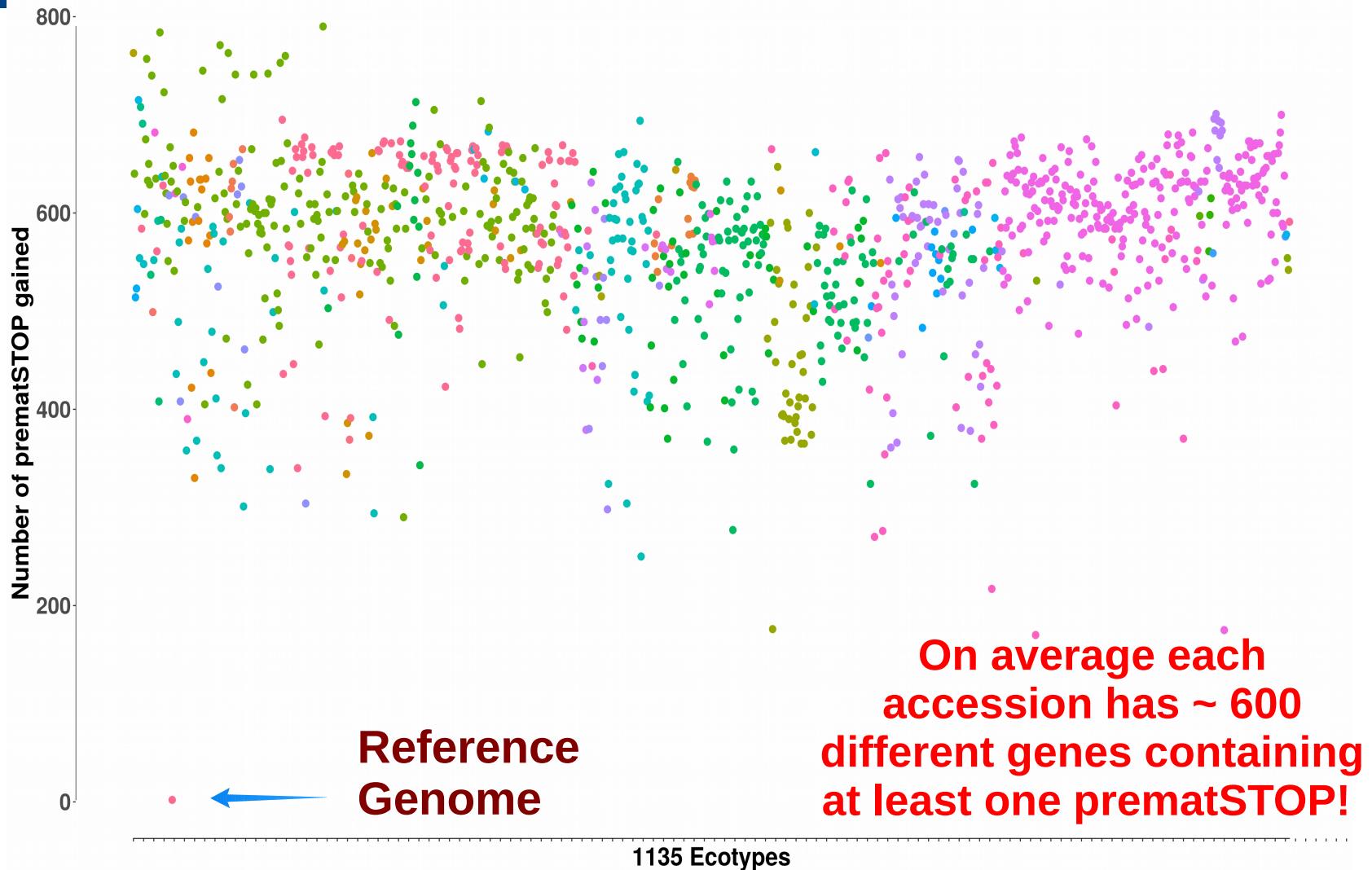
Scherzer et al. Mol Plant 2015



Korte et al. unpubl.

Gene	OST1	SLAC1	SLAH3
Synonymous	23	19	49
Non-synonymous	15	22	42
Splice sites	9	1	8
Alternative Start	4	0	2
Premature Stop	0	0	0

# Frequency of prematSTOP codons by accessions



# Co-occurrence of premature Stop codons

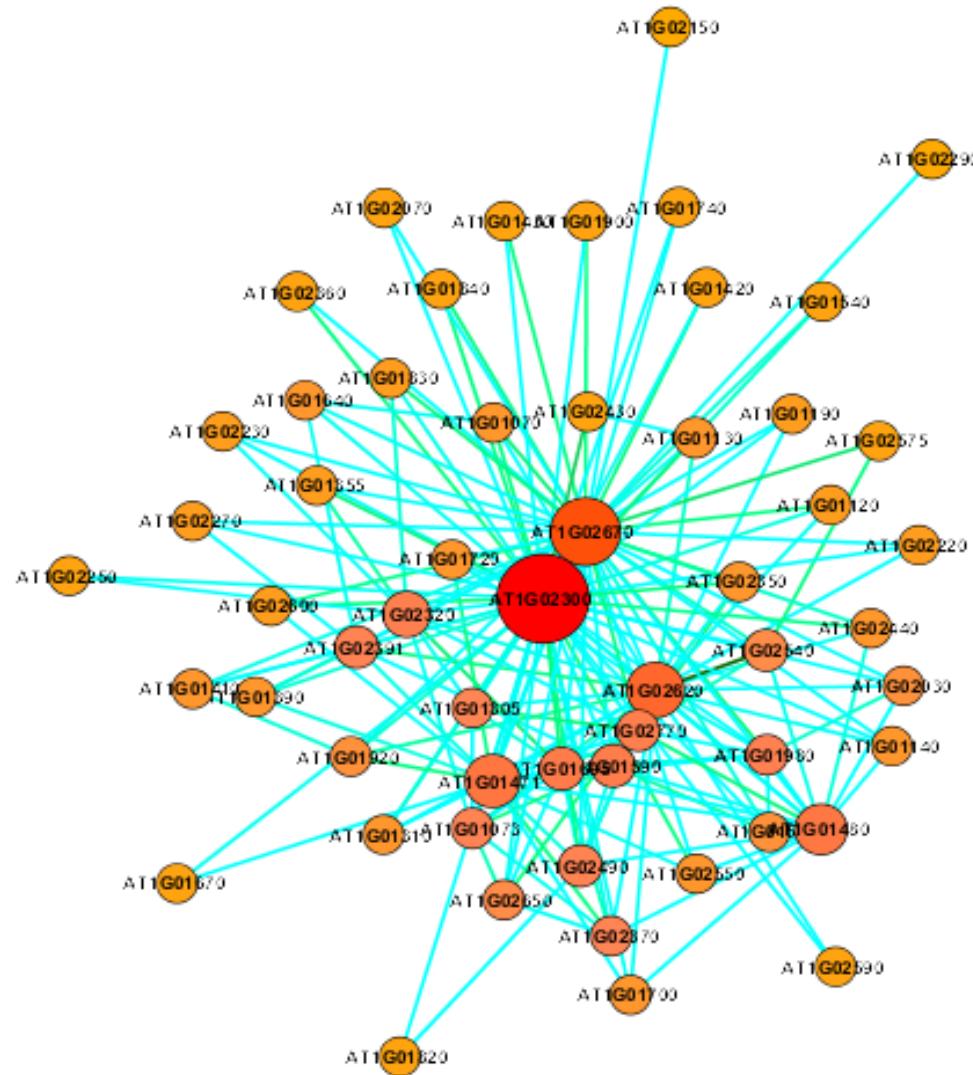
Filter for functional prematSTOP by using available RNAseq data (Kawakatsu *et al.* 2016)

 665 accessions (6.000 genes / 16.000 prematSTOP)

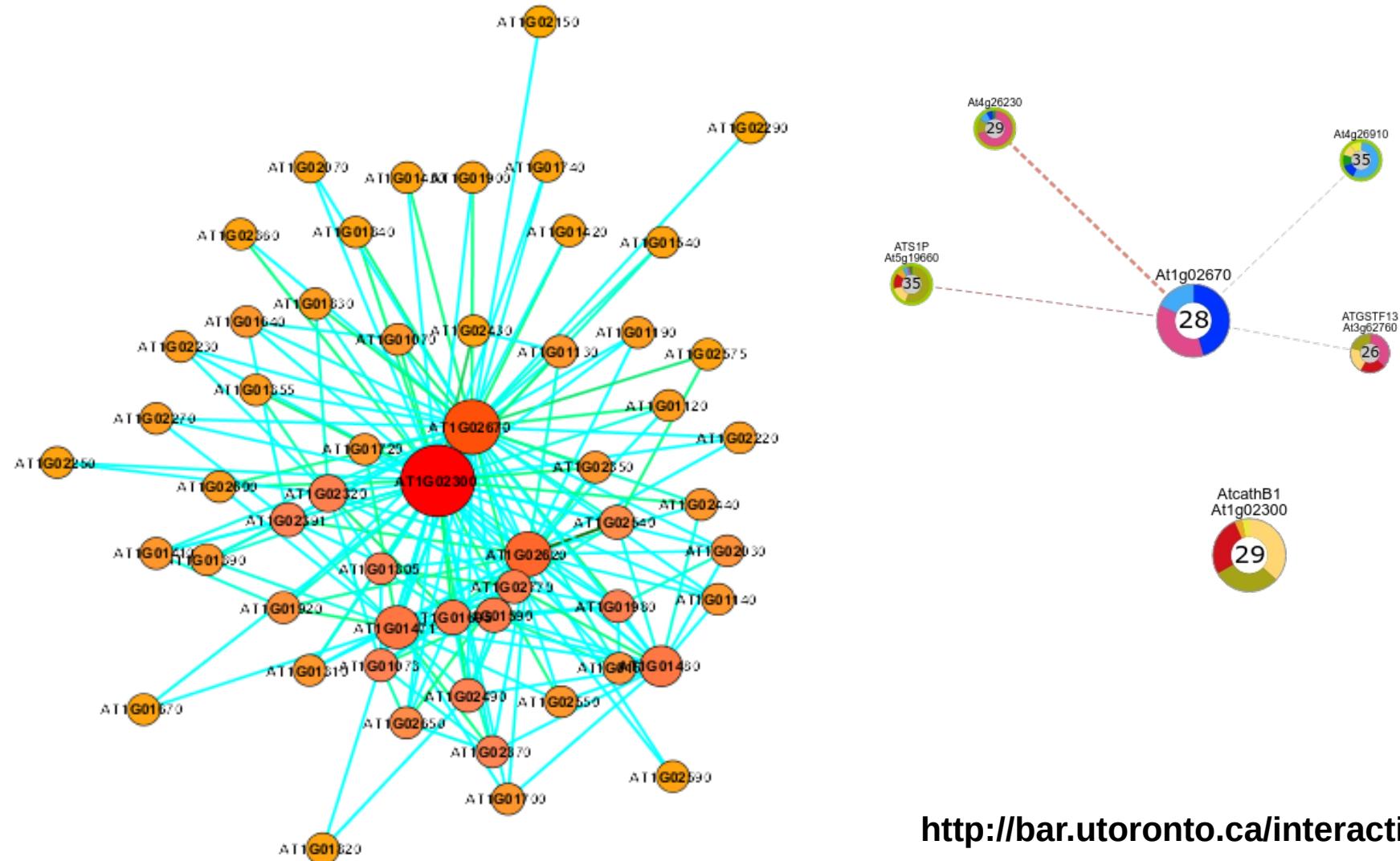
Gene 1	Gene 2	gene1_count	gene2_count	Co - occurrence	Lower border	Upper border	P-value
AT1G66920	AT3G59750	341	251	122	100	157	0.03
AT4G36140	AT5G45150	312	356	266	138	196	1.7e-57
AT3G05685	AT3G63320	444	114	38	55	96	5.6e-16

Many genes are knocked-out together more (or less) often than expected  
If we use synonymous SNPs as a control we don't observe this

# Co-occurrence of premature Stop codons

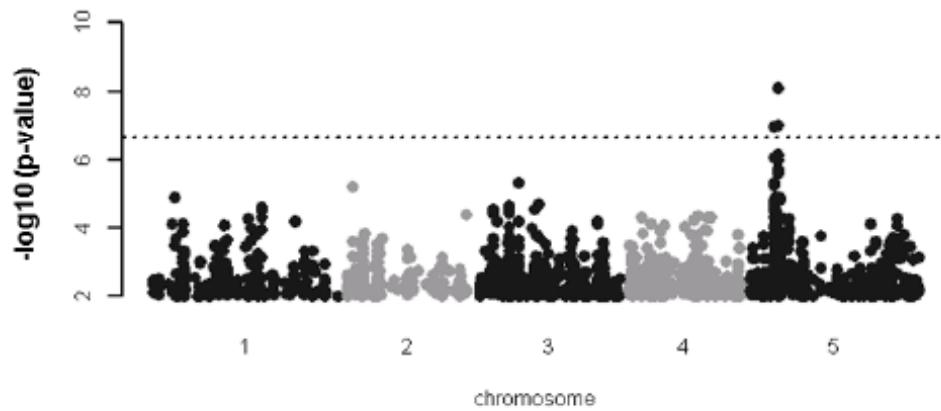


# Co-occurrence of premature Stop codons



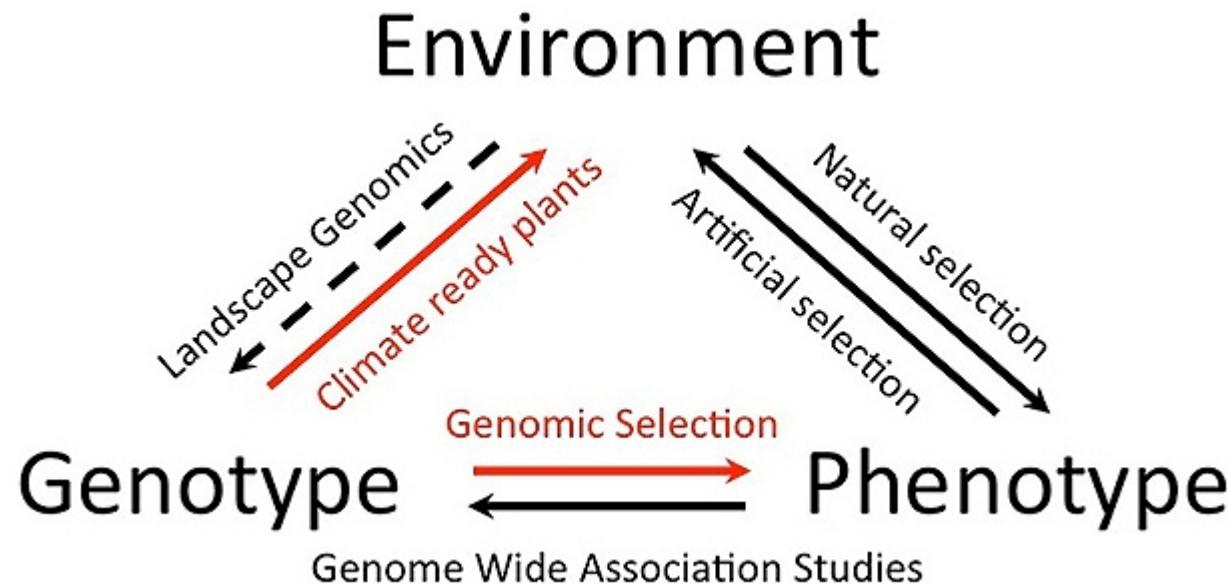
# GWAS on the Co-occurrence of prematSTOP

Gene 1	Gene 2	gene1_count	gene2_count	Co - occurrence	Lower border	Upper border	P-value
AT4G36140	AT5G45150	312	356	266	138	196	1.7e-57



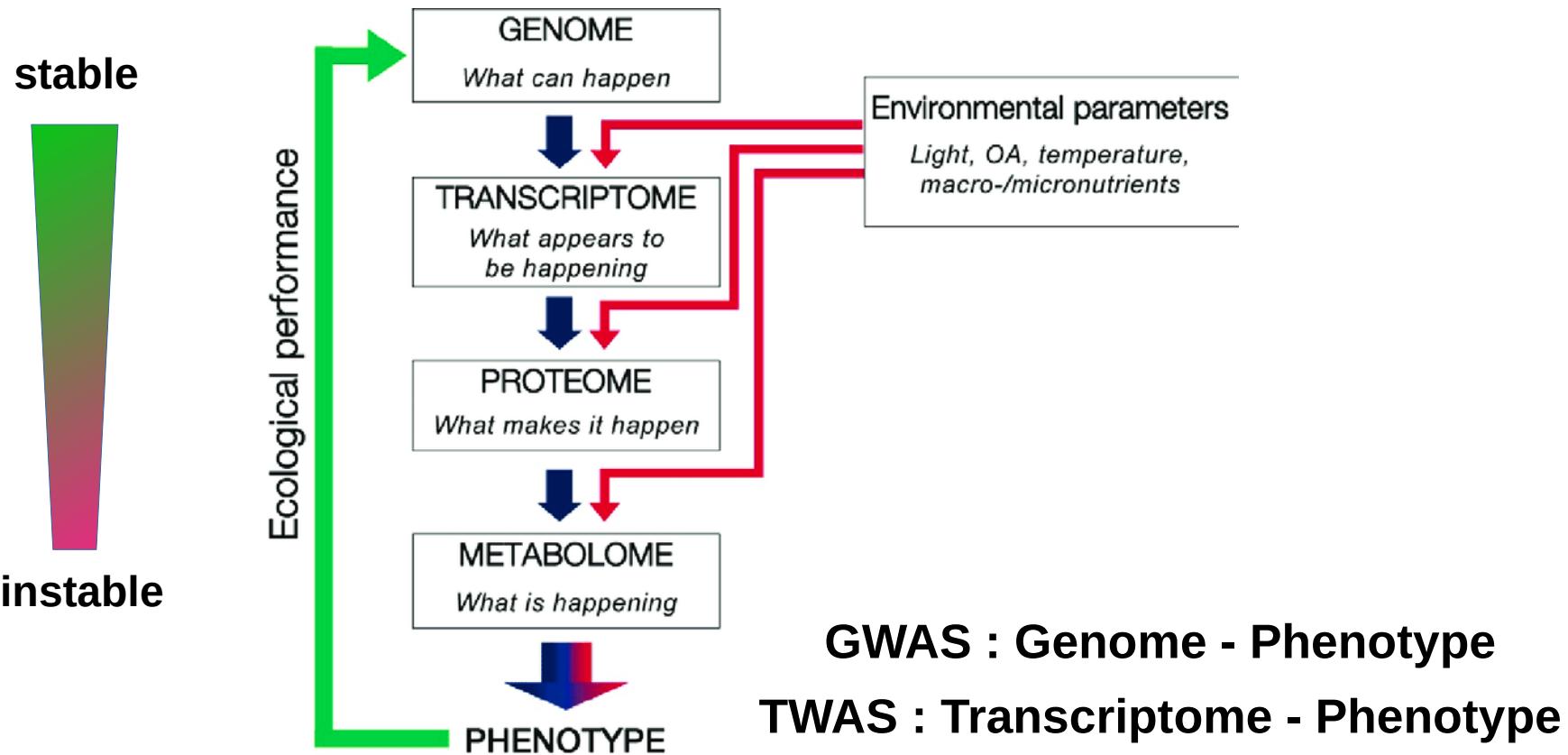
We can map genes that are responsible for the co-occurrence

# Evolutionary Genomics



If we want to understand G-P relationship, understanding of networks and interactions is essential

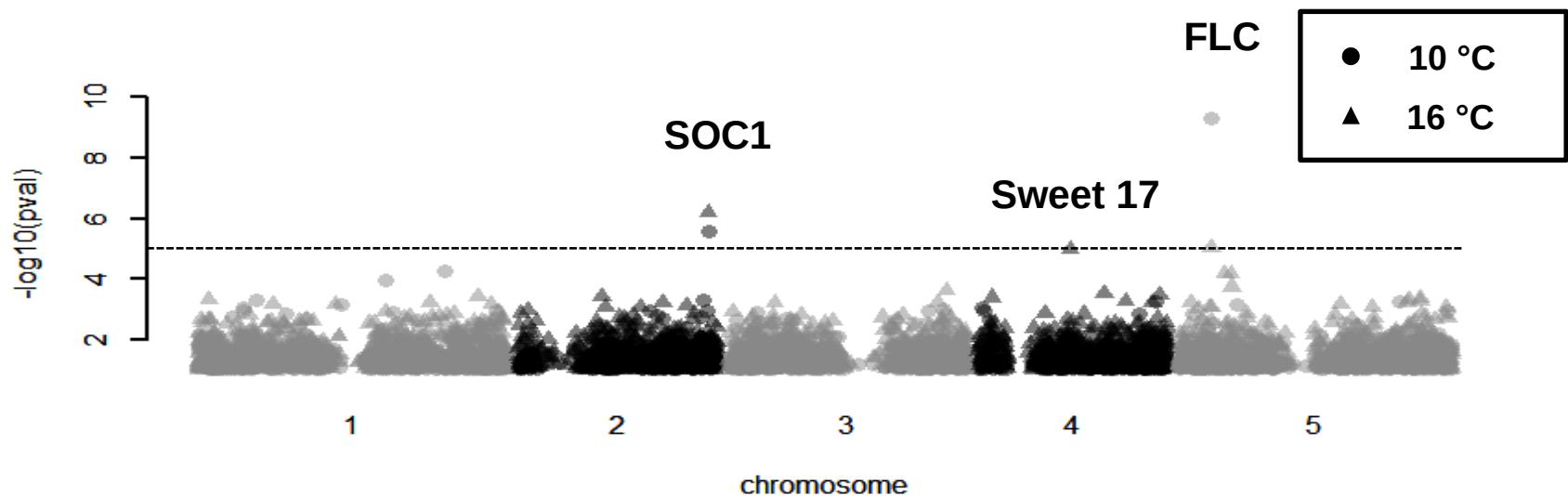
# The Omnics cascade



Dettmer et al. (2007) Mass Spect Rev

# TWAS on Flowering Time

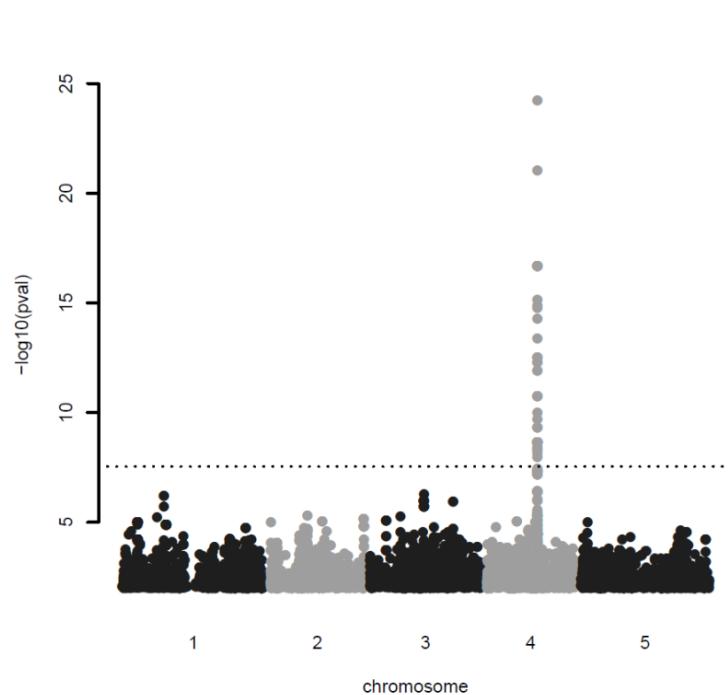
Using RNA expression to explain the phenotype



Beyond GWAS: Understanding the function of pathways

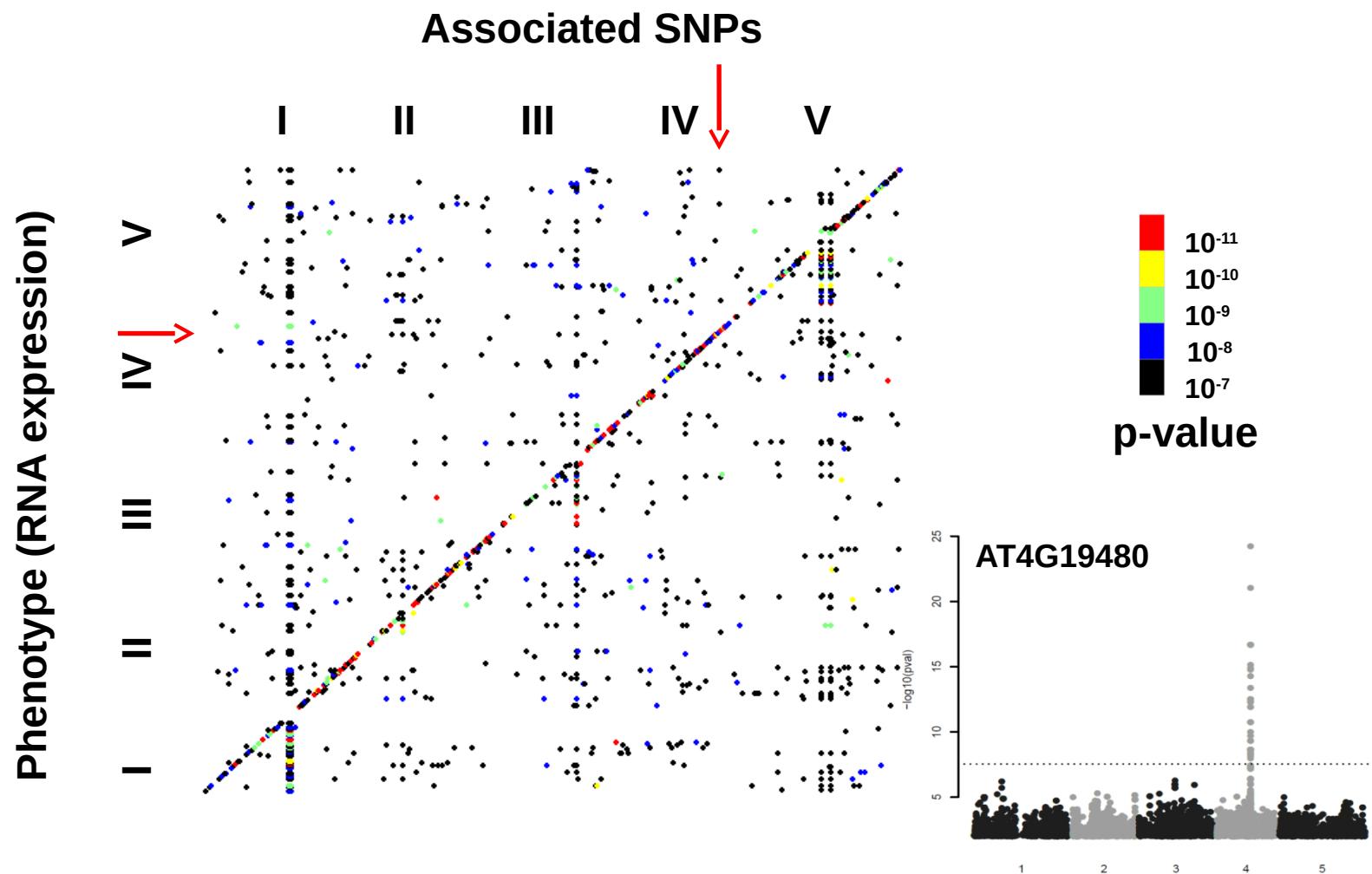
# eGWAS

**Data: 80 accessions grown under drought and control conditions (VIB Ghent), RNAseq for all of them**

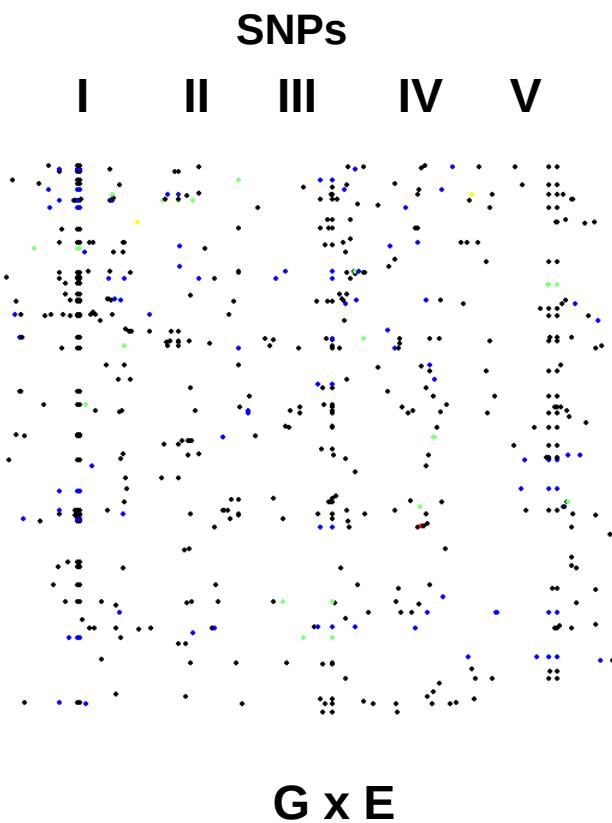
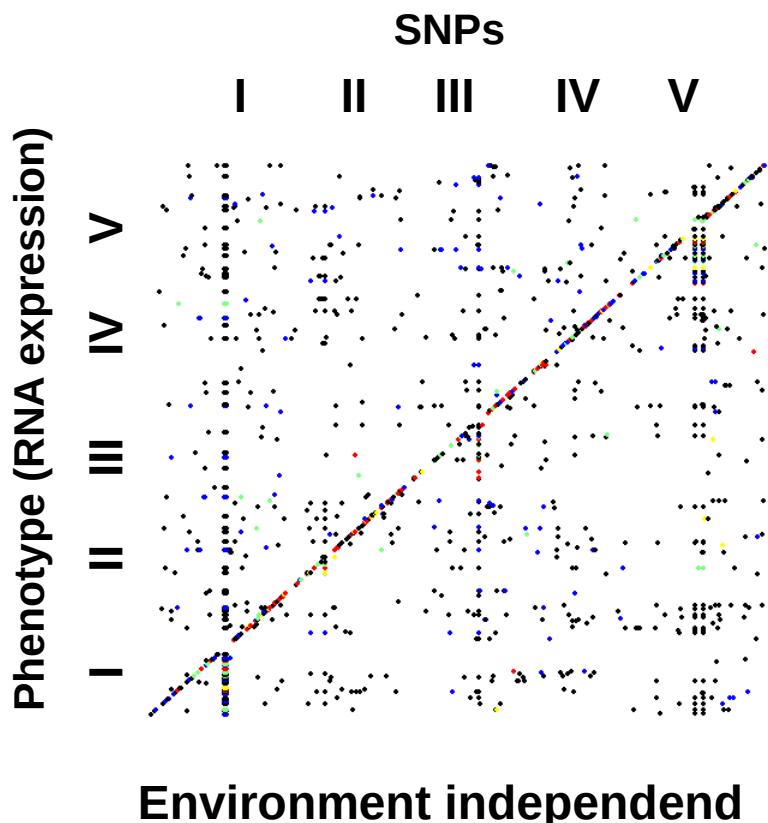


**A SNP on Chromosome 4 affects the expression of the gene AT4G19480  
How does this affect the phenotype ?**

## eGWAS

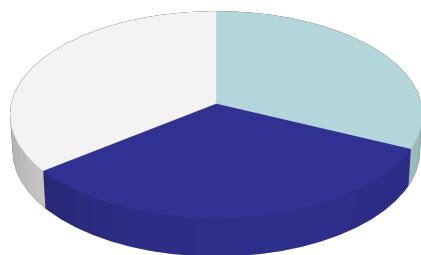


# Summary of MTMM on RNAseq data

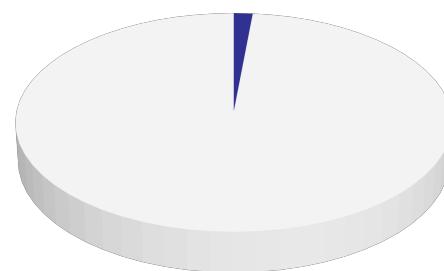


# Summary of MTMM on RNAseq data

Environment independent



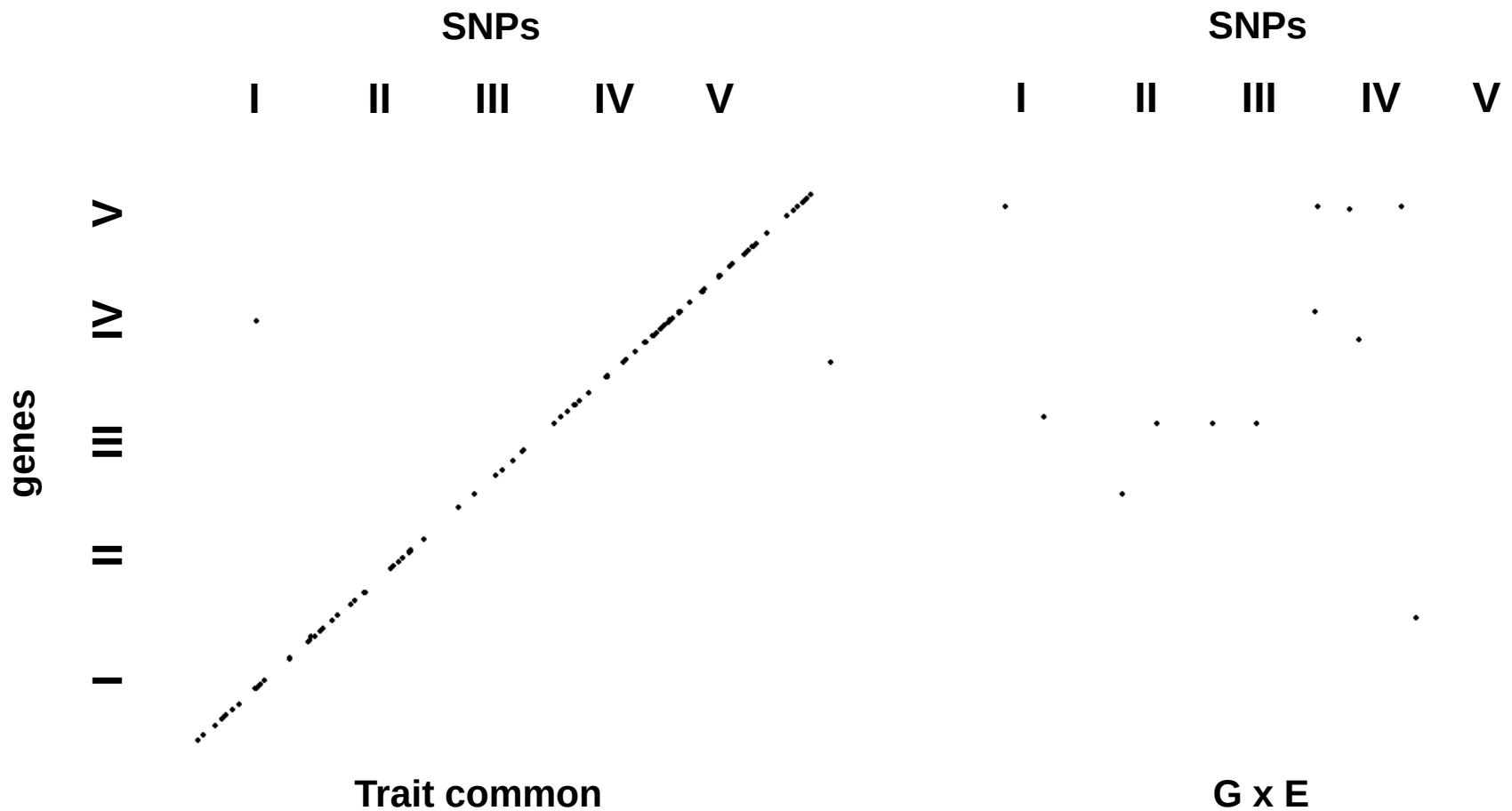
G x E



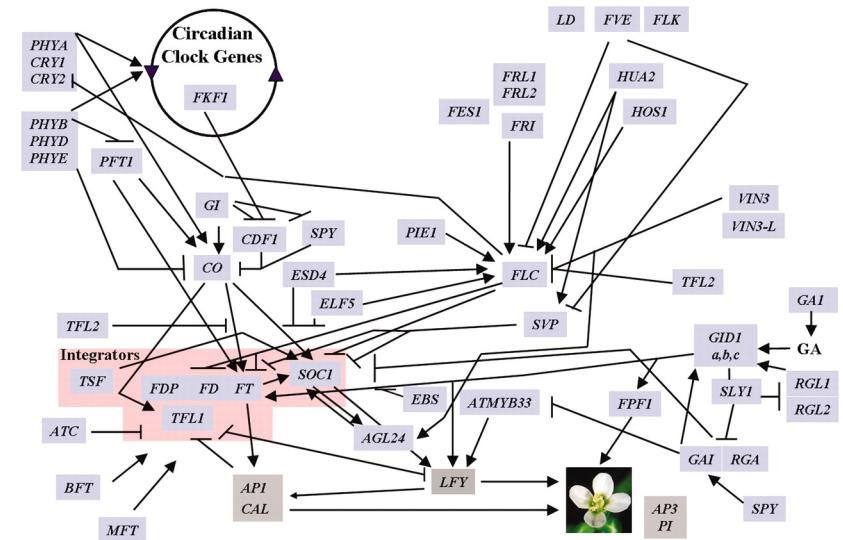
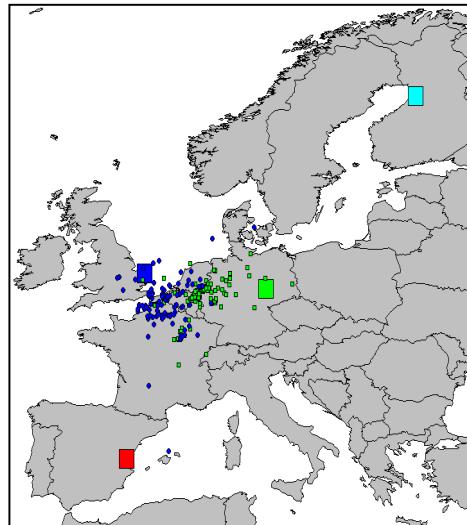
- only cis
- cis and trans
- only trans

Beyond GWAS: Understanding the function of pathways

# Dataset after very stringent filtering



# Summary



Tremendous amount of Natural Variation in signaling pathways

If we want to understand G-P, we need to understand this networks

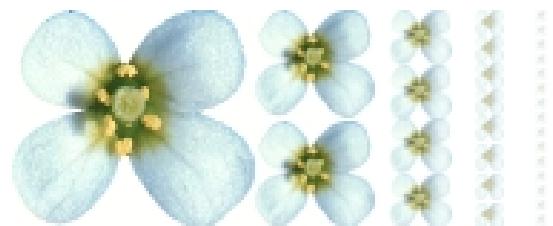
# Acknowledgments



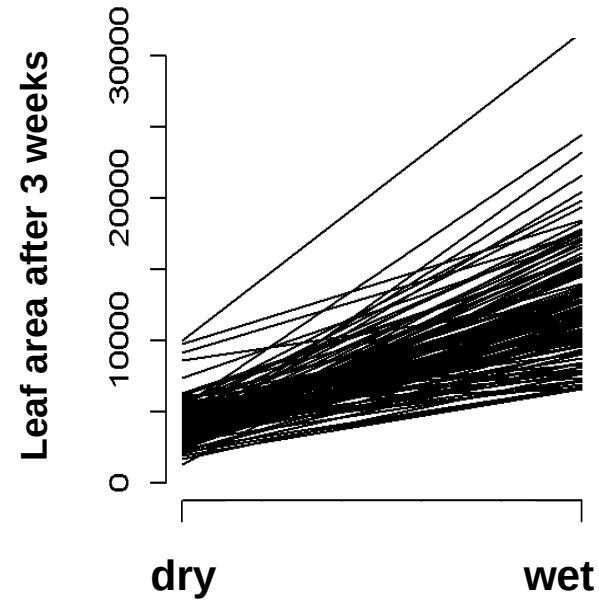
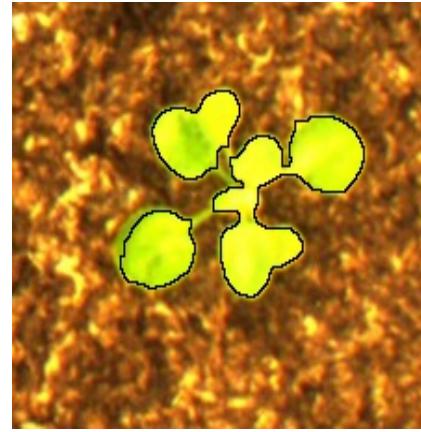
**Uni Würzburg**  
Jan Freudenthal  
Ammarah Anwar  
Willia Lopez

**GMI Vienna**  
Magnus Nordborg  
Ümit Seren

**VIB Ghent**  
Dirk Inze  
Frederik Coppens  
Pieter Clauw

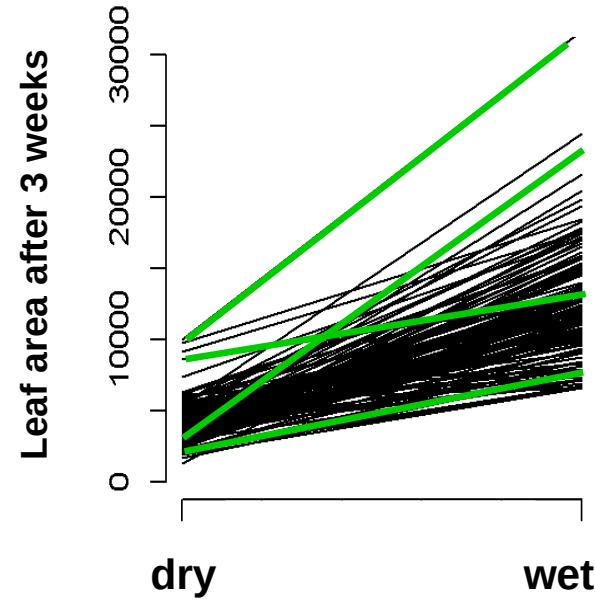
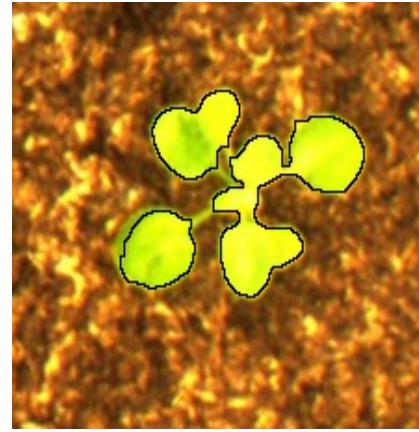


# Drought experiment



**Phenotypic differences in how drought affect seedling growth**

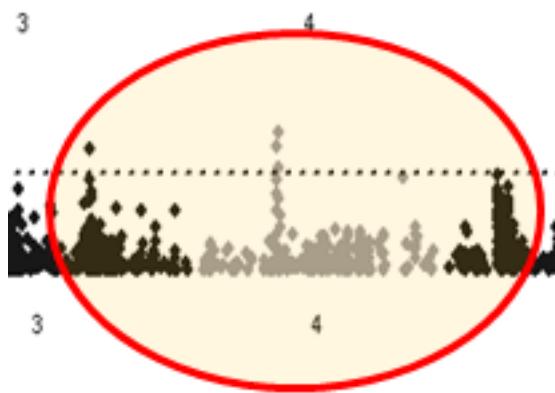
# Drought experiment



Phenotypic differences in how drought affect seedling growth

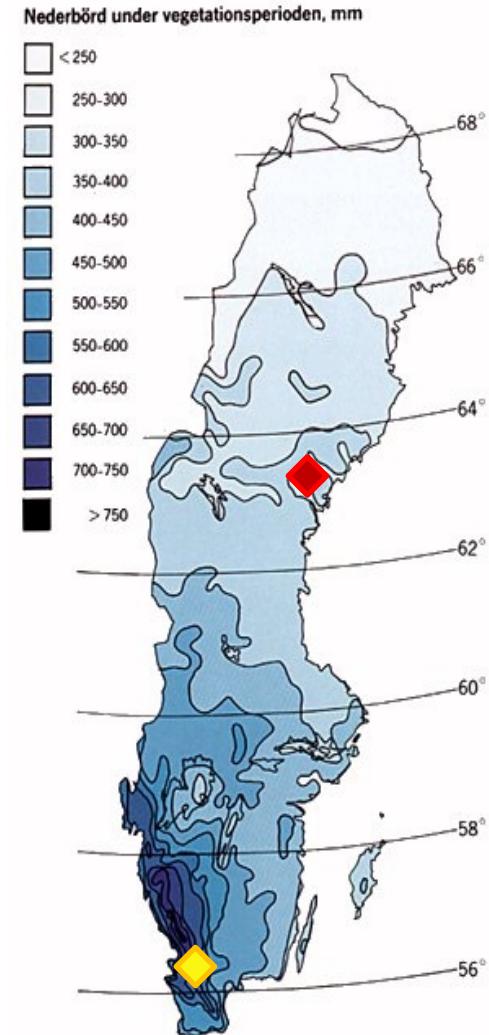
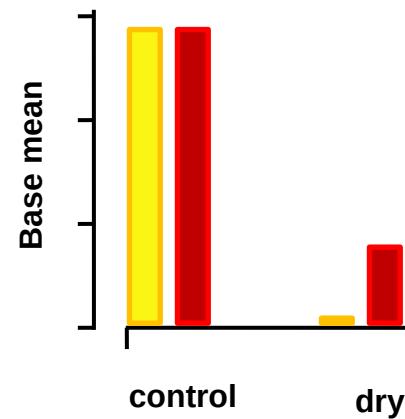
# Example : MTMM on Leaf area day 21

MTMM  
Trait specific test

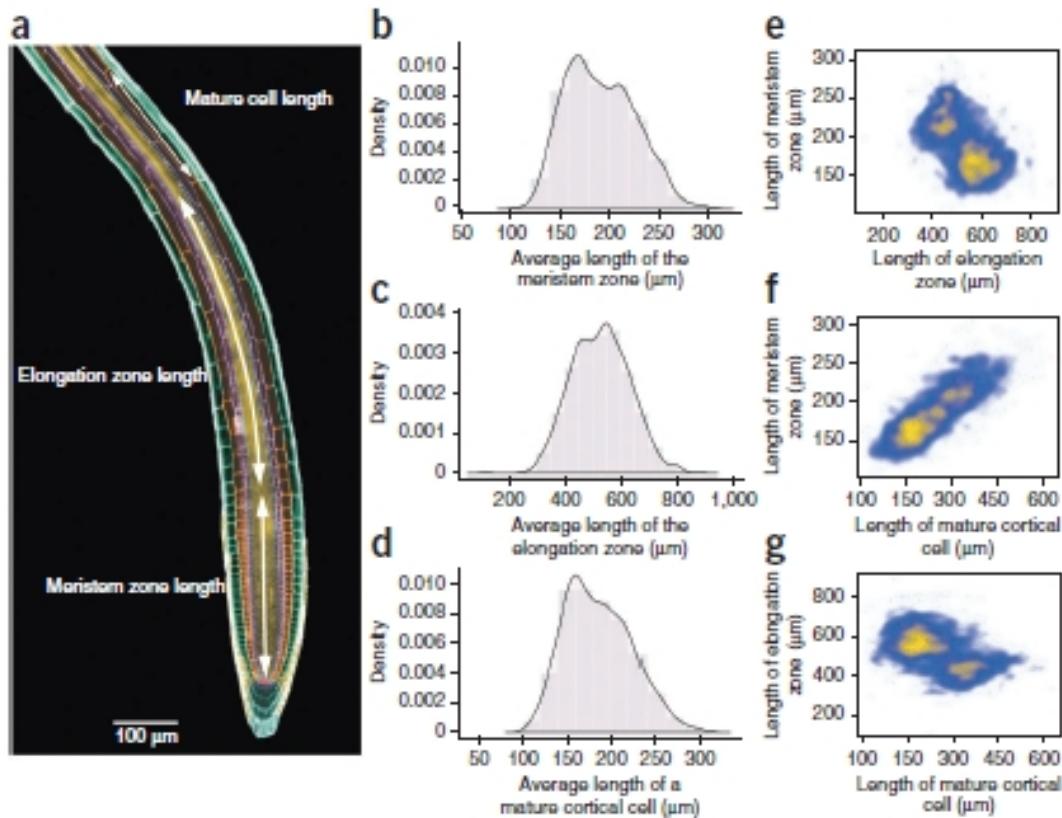


AT4G08950 EXO  
(response to  
brassinosteroids)

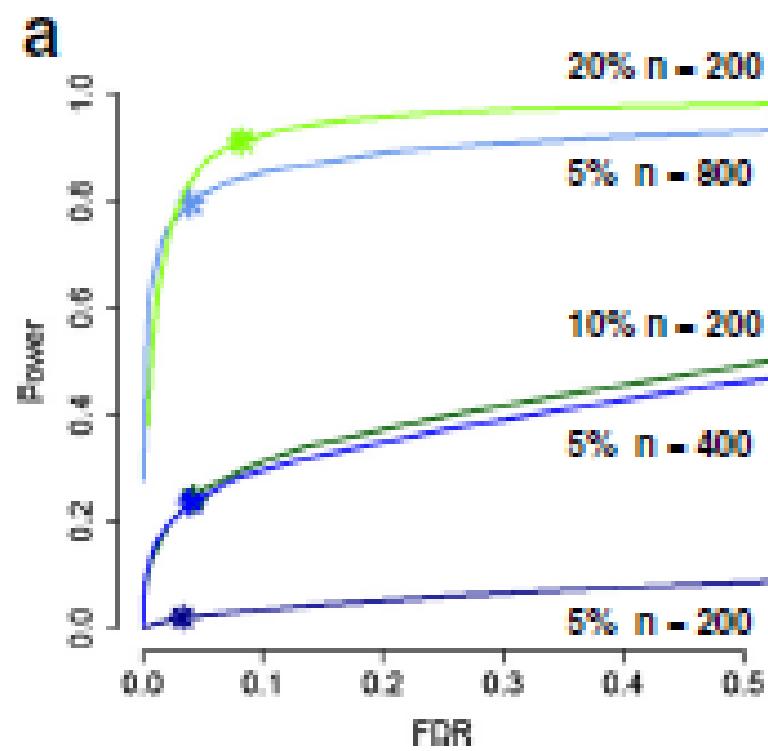
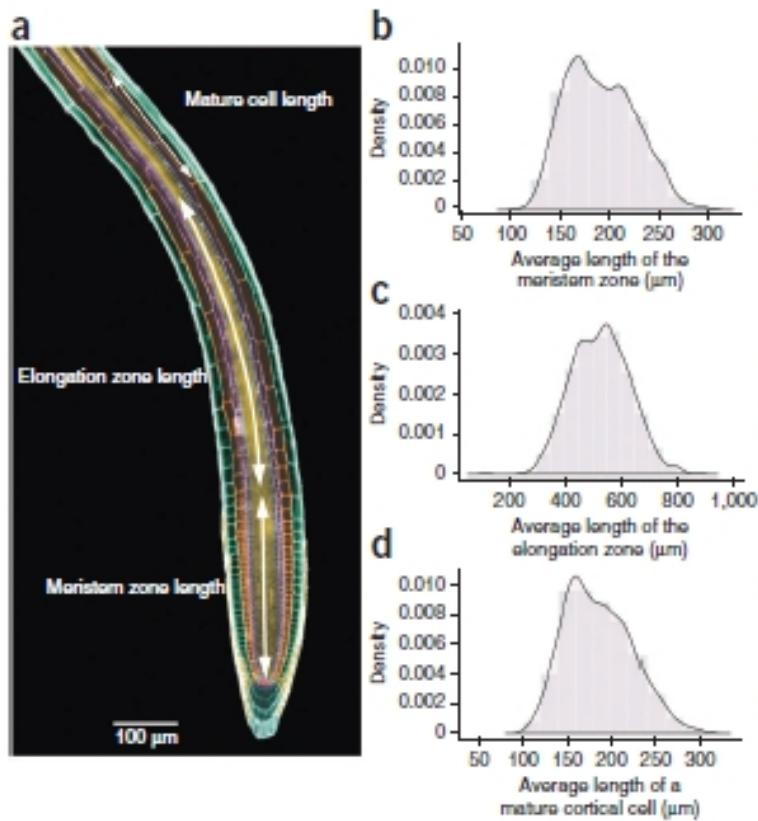
RNA expression  
of AT4G08950



# GWAS on cellular traits

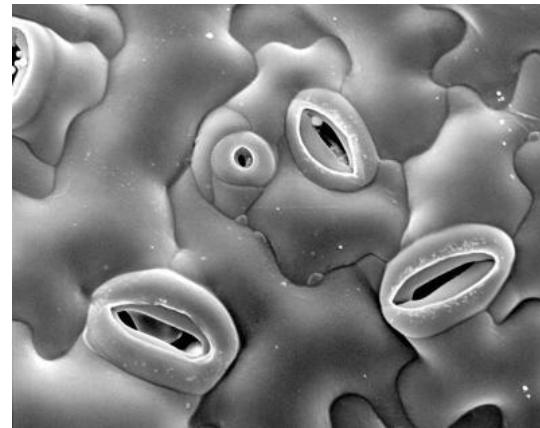
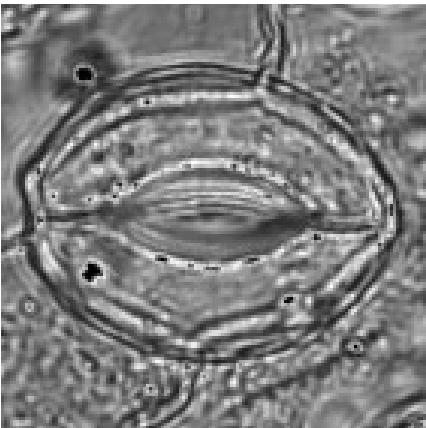


# GWAS on cellular traits

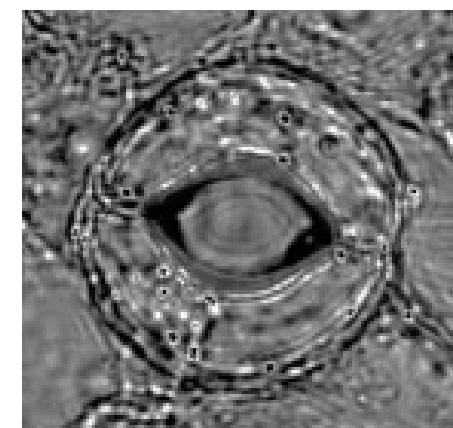


# Natural variation in guard cell signalling

Low humidity  
High CO<sub>2</sub>  
Darkness



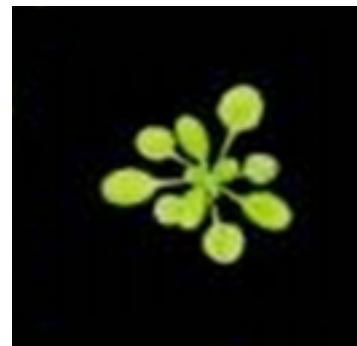
Heat  
Low CO<sub>2</sub>  
Light



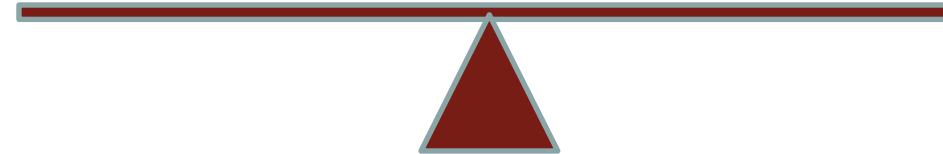
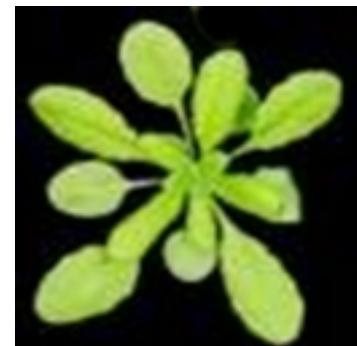
Guard cells directly sense environmental cues and balance water loss and CO<sub>2</sub> uptake

# Natural variation in guard cell signalling

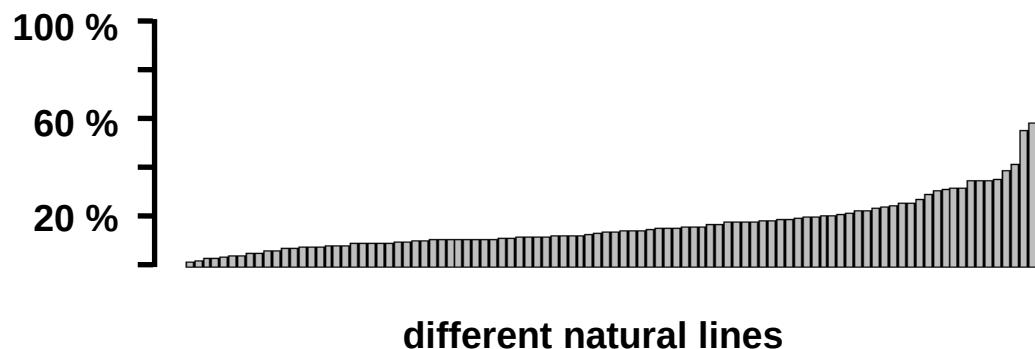
Drought tolerance



Biomass



relative leaf growth under progressive drought



different natural lines

Korte et al. unpubl.

# Natural variation in guard cell signalling

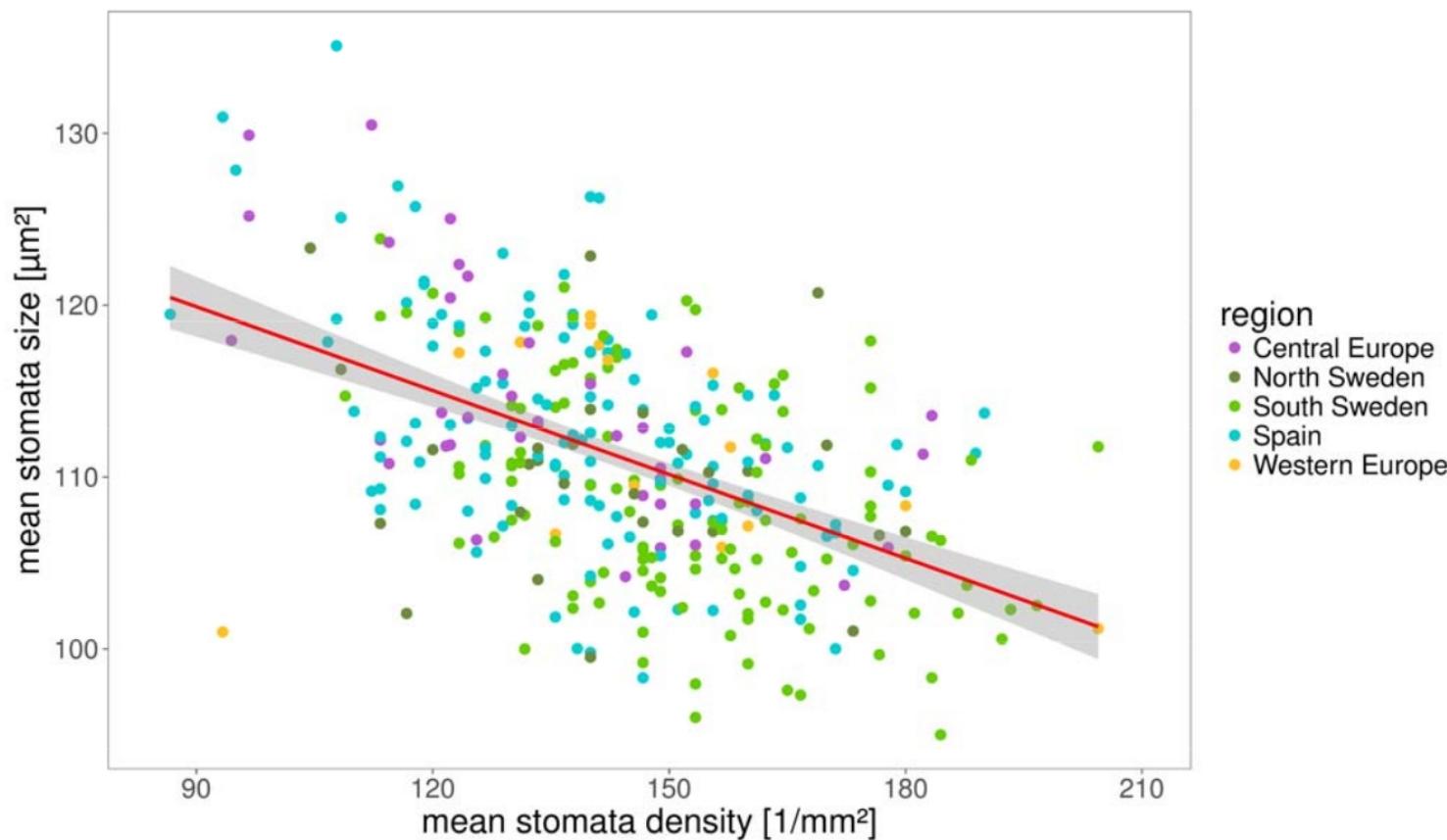


Figure 1: Natural variation in stomata patterning