

Understanding genotype-phenotype relationships using the model plant *Arabidopsis thaliana*

Arthur Korte

01.03.2021

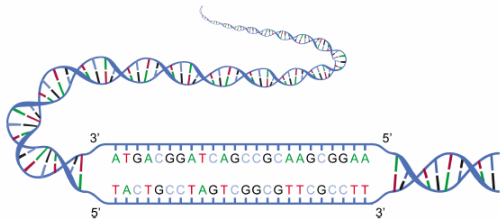


@arthurkorte



arthurkorte

Genotype

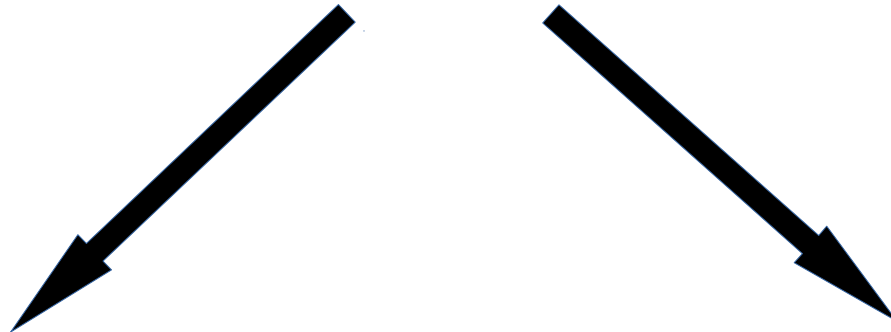


Phenotype





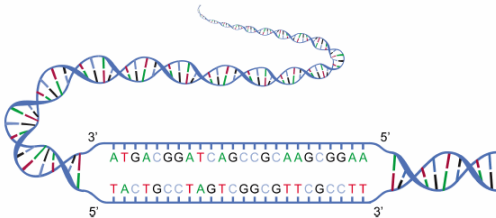
Environment



Genotype



Phenotype





Environment

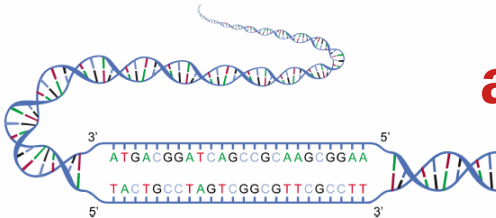
Adaptation

Natural selection

Genotype

Phenotype

**Genome-wide
association studies
(GWAS)**





Environment

Adaptation

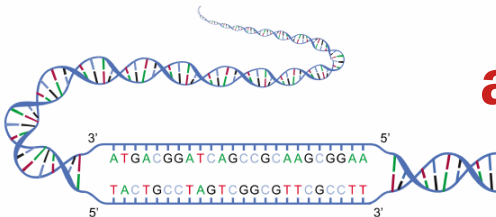
Natural selection

Genomic prediction

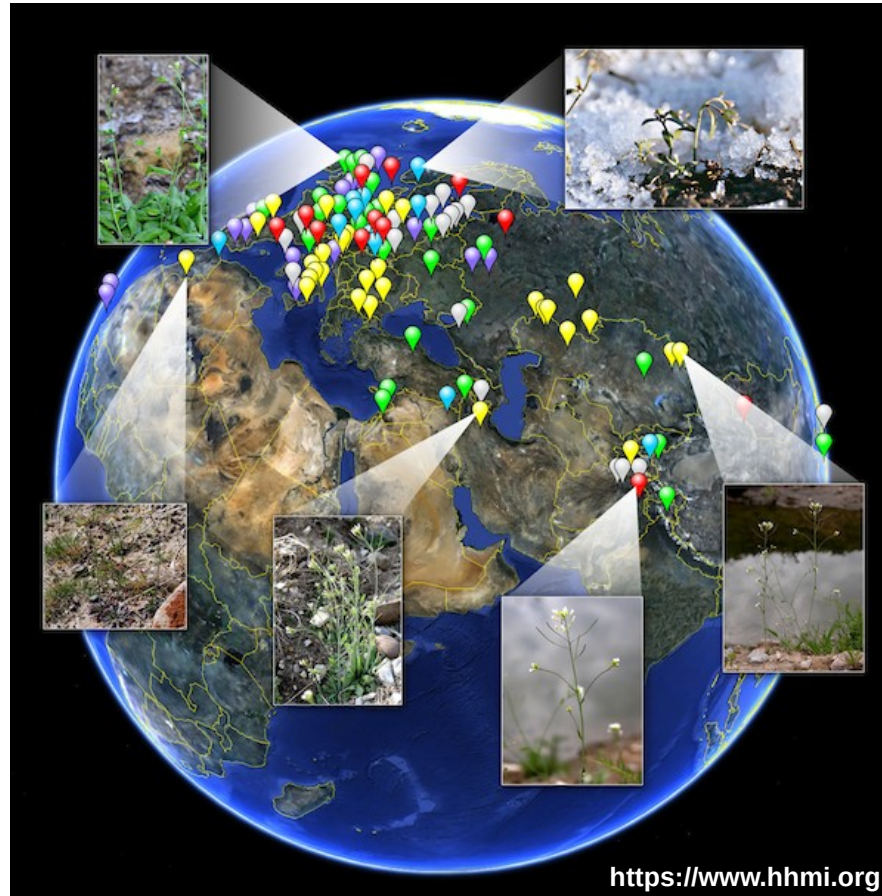
Genotype

Phenotype

**Genome-wide
association studies
(GWAS)**



Arabidopsis thaliana



**Arabidopsis is a prime model for Evolutionary Genomics
(small genome, adapted to the local environment, occurs as natural inbred)**

Arabidopsis thaliana



The 1001 Genomes Project

1001 Genomes

A Catalog of *Arabidopsis thaliana* Genetic Variation

[Home](#)
[Data Providers](#)
[Accessions](#)
[Tools](#)
[Software](#)
[Data Center](#)
[About](#)

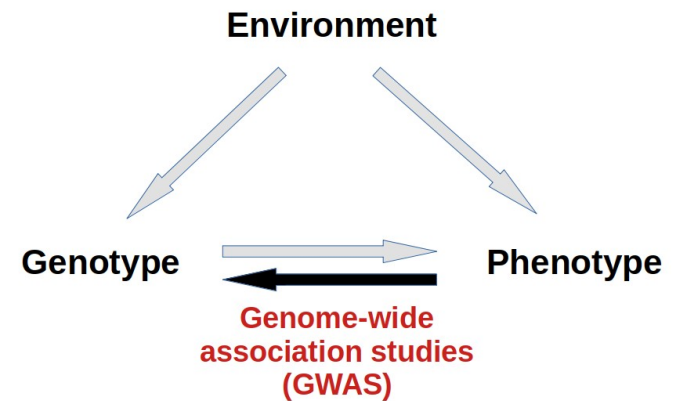
- Full genome sequence of 1,135 different natural ecotypes

10M SNPs and 500k structural variants in a genome with ~125 M Bp



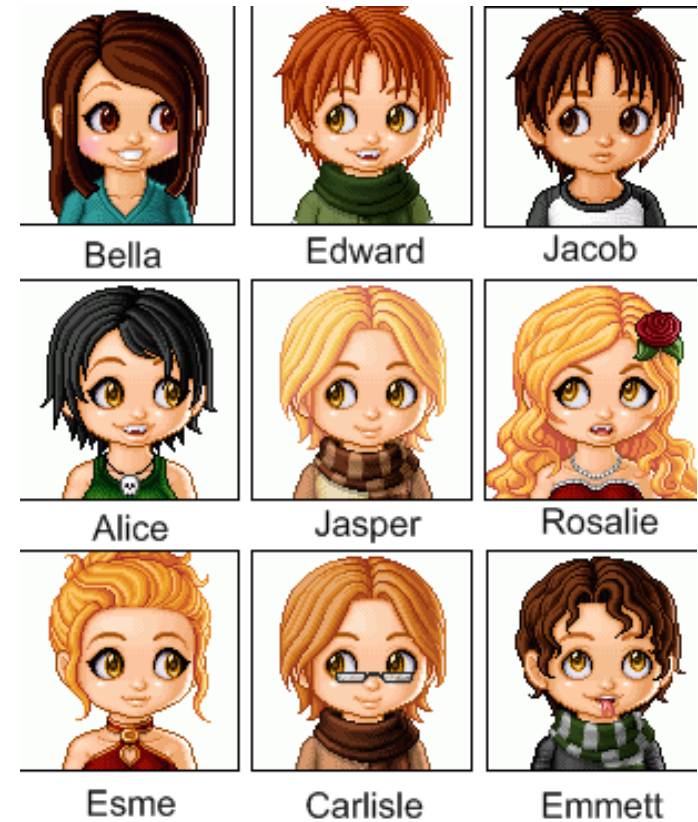
www.1001genomes.org

Genome-wide association studies (GWAS)



Associations between genotype and phenotype

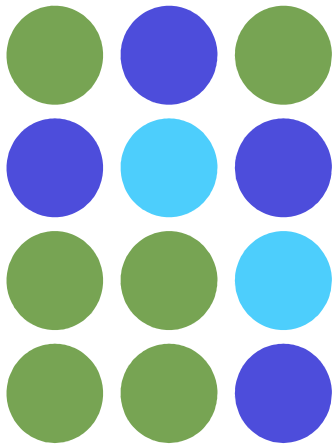
Genome-wide association studies (GWAS)



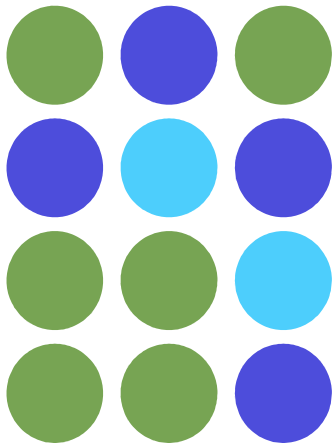
Associations between genotype and phenotype



Introduction: GWAS

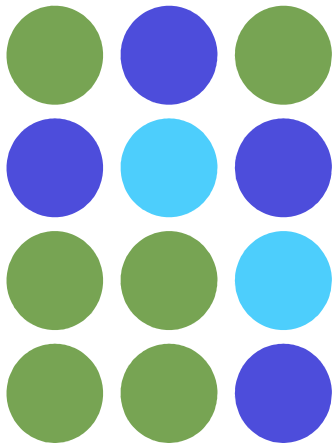


Introduction: GWAS



...AGCCTG-----TGCACTAAGACT...
 ...AGCCTG-----TGCACTAAGACT...
 ...AGCCTG-----TGCACTAAGAGT...
 ...AGCCTG-----TGCACTAAGACT...
 ...AGCCTGAGTGTGCACTAAGAGT...
 ...AGCCTGAGTGTGCACTAAGAGT...
 ...AGCCTGAGTGTGTACTAAGACT...
 ...AGCCTGAGTGTGTACTAAGAGT...
 ...AGCCTGAGTGTGTACTAAGAGT...
 ...AGCCTGAGTGTGTACTAAGACT...
 ...AGCCTGAGTGTGTACTAAGACT...

Introduction: GWAS



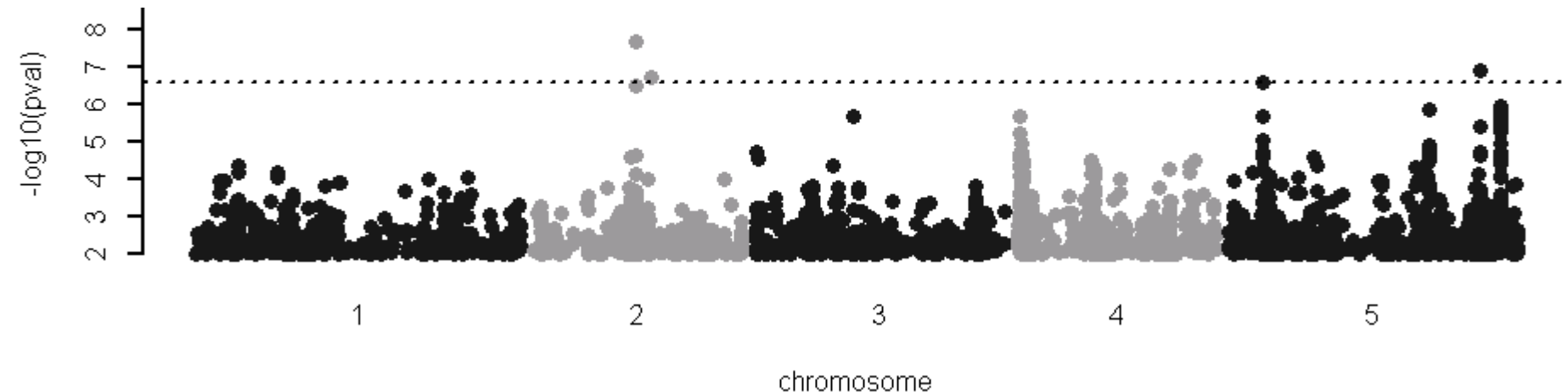
...AGCCTG-----TGCACCTAAGACT...
 ...AGCCTG-----TGCACCTAAGACT...
 ...AGCCTG-----TGCACCTAAGAGT...
 ...AGCCTG-----TGCACCTAAGACT...
 ...AGCCTGAGTGTGCACCTAAGAGT...
 ...AGCCTGAGTGTGCACCTAAGAGT...
 ...AGCCTGAGTGTGTACTAAGACT...
 ...AGCCTGAGTGTGTACTAAGAGT...
 ...AGCCTGAGTGTGTACTAAGACT...
 ...AGCCTGAGTGTGTACTAAGAGT...
 ...AGCCTGAGTGTGTACTAAGACT...
 ...AGCCTGAGTGTGTACTAAGACT...

Introduction: GWAS

Calculate a statistical test for each SNP using linear mixed models

$$Y = X\beta + u + \epsilon, \quad u \sim N(0, \sigma_g K), \quad \epsilon \sim N(0, \sigma_e I)$$

We get a p-value for each SNP that can be plotted in a Manhattan plot



Goal: To identify markers that are significantly associated with a given trait
permutation-based significance threshold (Freudenthal et al. 2019)

AraPHENO and AraGWAS





AraPheno ☐ AraRNASeq Data ▾ Tools ▾ About



AraPheno

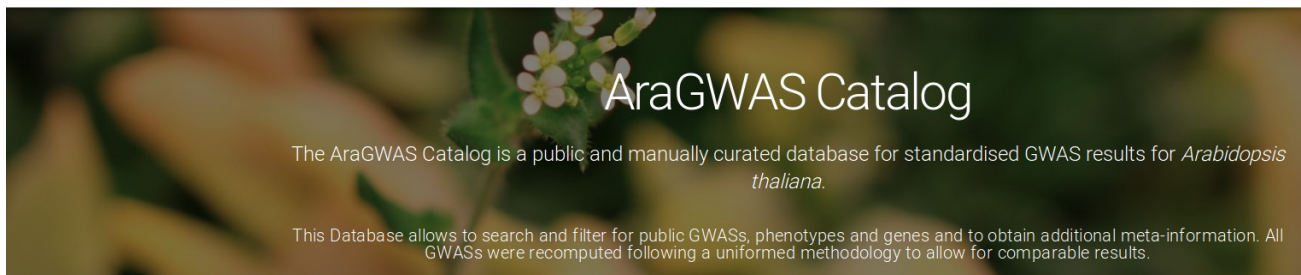
AraPheno is a public database collection of *Arabidopsis thaliana* phenotypes.
This Database allows to search and filter for public phenotypes and to obtain additional meta-information.

Quick Stats

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

AraGWAS Catalog

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

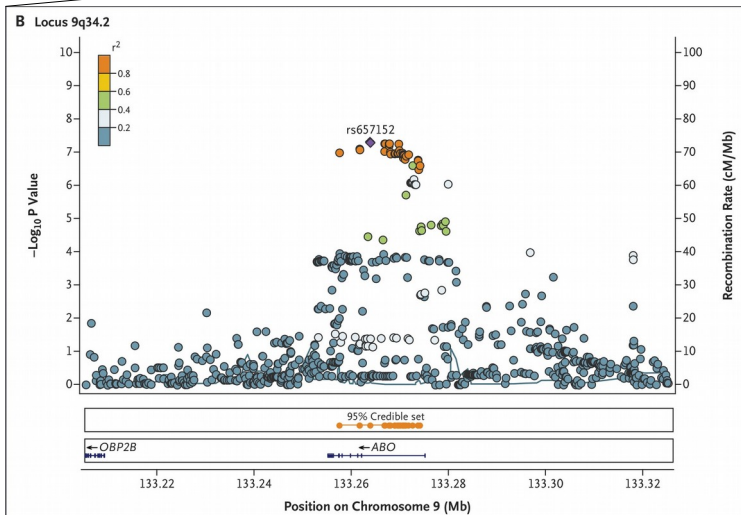
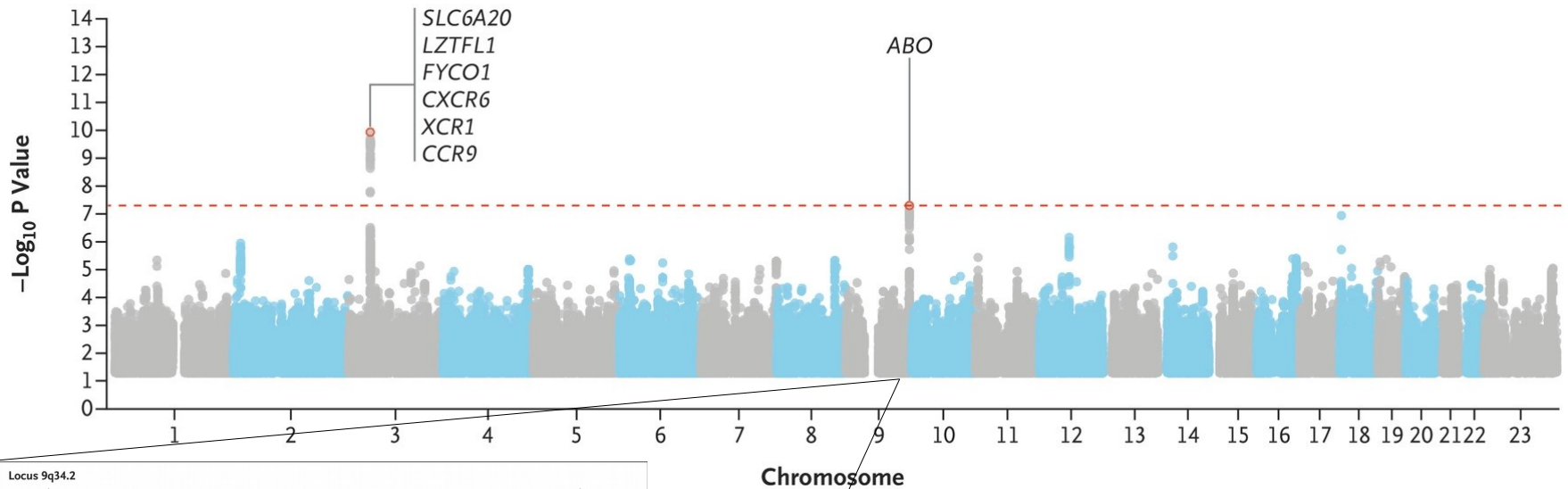


AraGWAS Catalog

The AraGWAS Catalog is a public and manually curated database for standardised GWAS results for *Arabidopsis thaliana*.

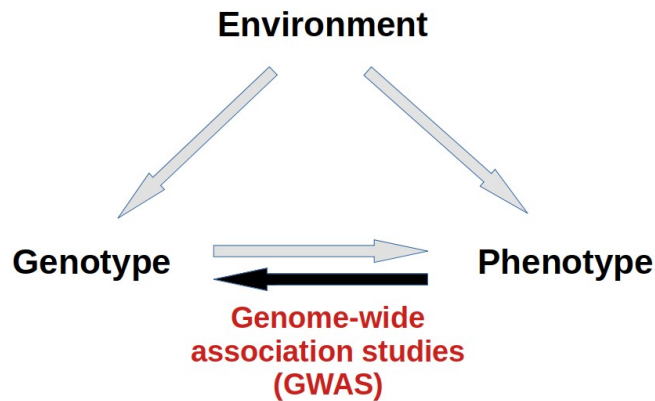
This Database allows to search and filter for public GWASs, phenotypes and genes and to obtain additional meta-information. All GWASs were recomputed following a uniform methodology to allow for comparable results.

GWAS on Covid-19 severeness



The Severe Covid-19 GWAS Group
The New England Journal of Medicine (2020)

Gene x environment interactions



GWAS : Genome - Phenotype



HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2013 March 01.

Published in final edited form as:

Nat Genet. 2012 September ; 44(9): 1066–1071. doi:10.1038/ng.2376.

A mixed-model approach for genome-wide association studies of correlated traits in structured populations

Arthur Korte^{1,*}, Bjarni J. Vilhjálmsson^{1,2,*}, Vincent Segura^{1,3,*}, Alexander Platt^{1,2}, Quan Long¹, and Magnus Nordborg^{1,2}

¹ Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria.

² Molecular and Computational Biology, University of Southern California, Los Angeles, California, United States of America.

³ INRA, UR0588, F-45075 Orléans, France.

Abstract

Genome-wide association studies (GWAS) are a standard approach for studying the genetics of natural variation. A major concern in GWAS is the need to account for the complicated dependence-structure of the data both between loci as well as between individuals. Mixed models have emerged as a general and flexible approach for correcting for population structure in GWAS. Here we extend this linear mixed model approach to carry out GWAS of correlated phenotypes, deriving a fully parameterized multi-trait mixed model (MTMM) that considers both the within-trait and between-trait variance components simultaneously for multiple traits. We apply this to human cohort data for correlated blood lipid traits from the Northern Finland Birth Cohort 1966, and demonstrate greatly increased power to detect pleiotropic loci that affect more than one blood lipid trait. We also apply this to an *Arabidopsis* dataset for flowering measurements in two different locations, identifying loci whose effect depends on the environment.

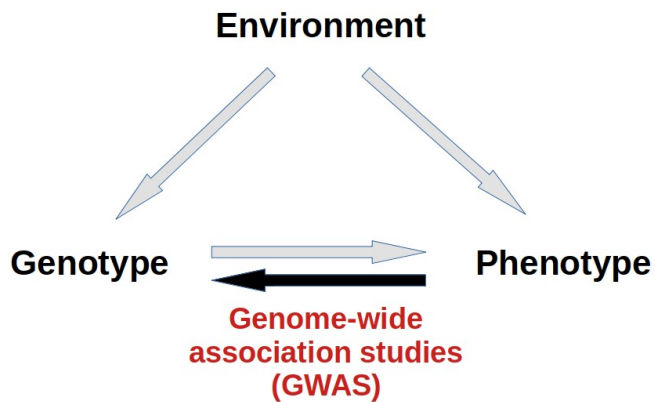
Introduction

Author Manuscript

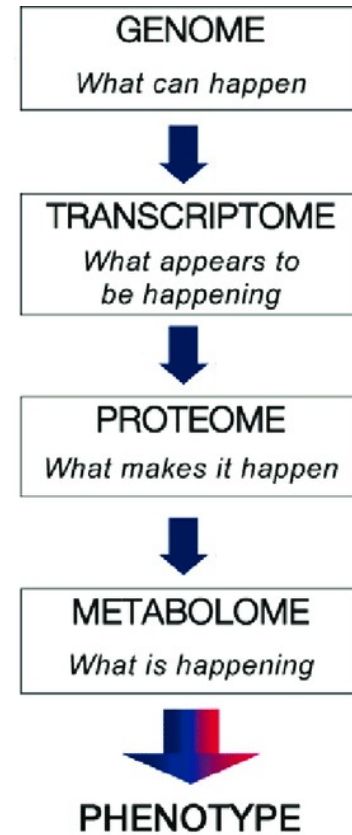
Author Manuscript

Author Manuscript

The Omnics cascade



GWAS : Genome - Phenotype



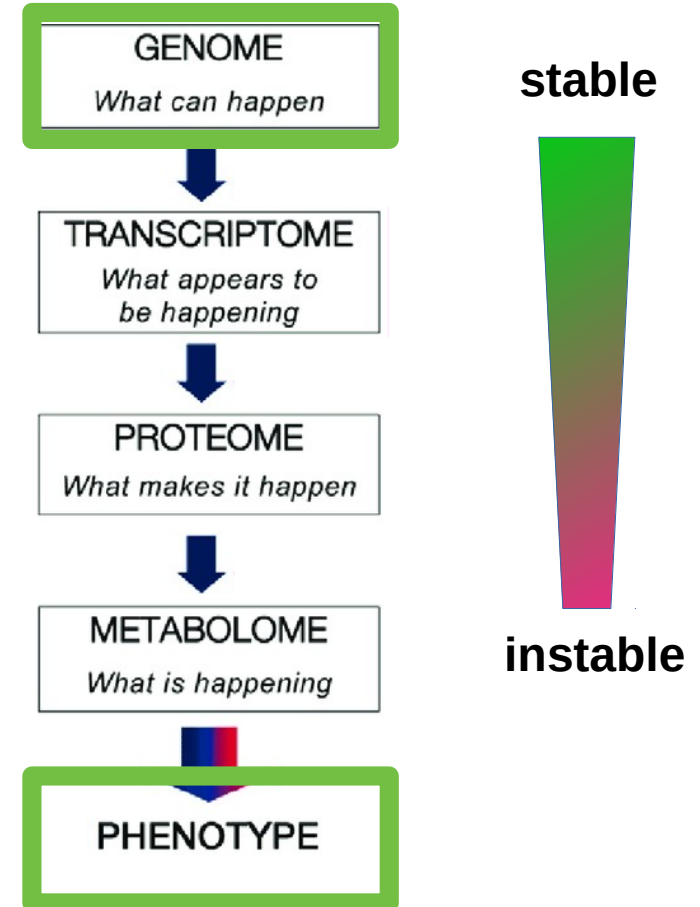
stable

unstable

The Omnics cascade

Full genomes >1000 ecotypes
(1001 genomes consortium, 2016)

> 500 different phenotypes
(Seren et al., 2017 & Togninalli et al. 2020)

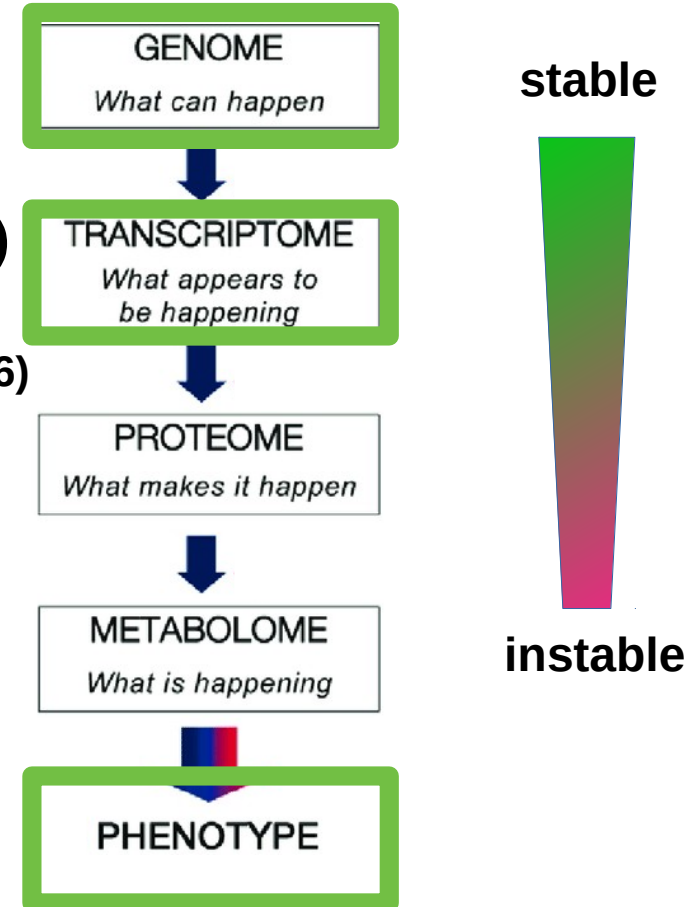


The Omnics cascade

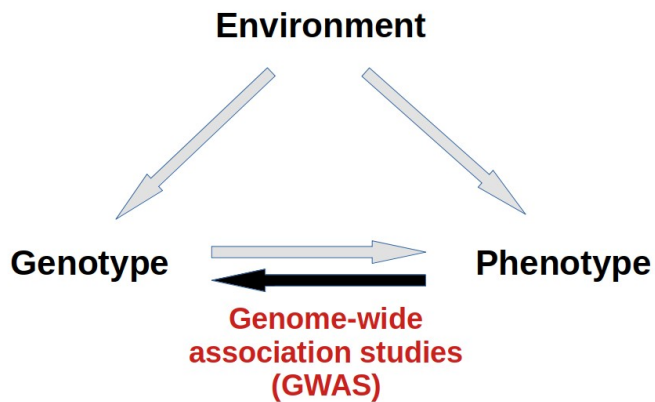
Full genomes >1000 ecotypes
(1001 genomes consortium, 2016)

Complete transcriptomes (28 k genes)
for 80 ecotypes (Clauw et al. 2016)
& for ~ 700 ecotypes (Kawakatsu et al. 2016)

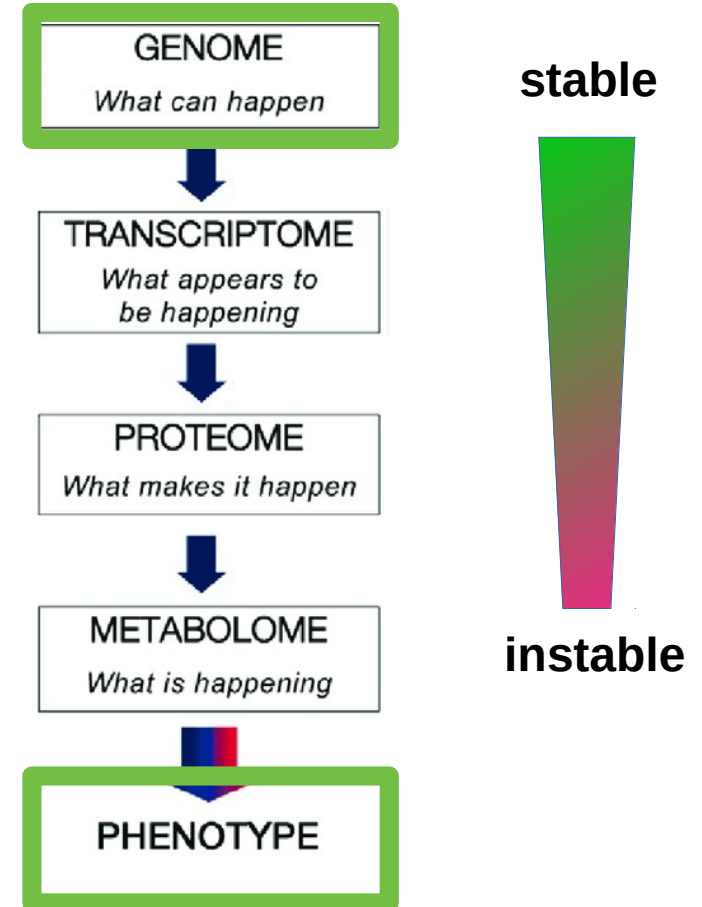
> 500 different phenotypes
(Seren et al., 2017 & Togninalli et al. 2020)



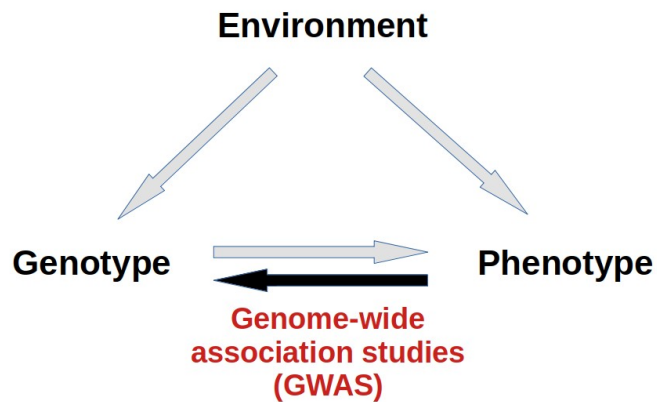
The Omnics cascade



GWAS : Genome - Phenotype

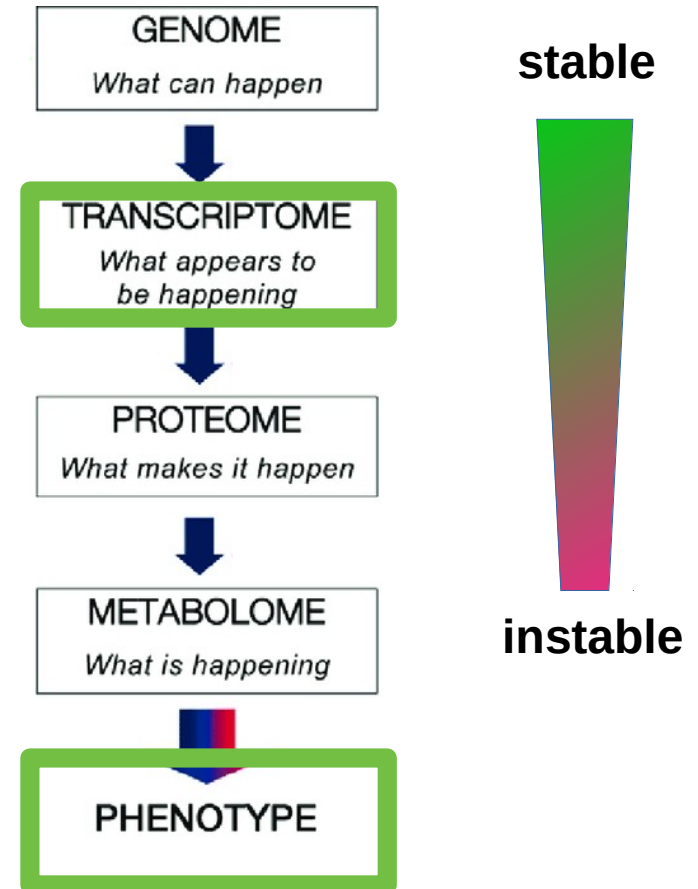


The Omnics cascade



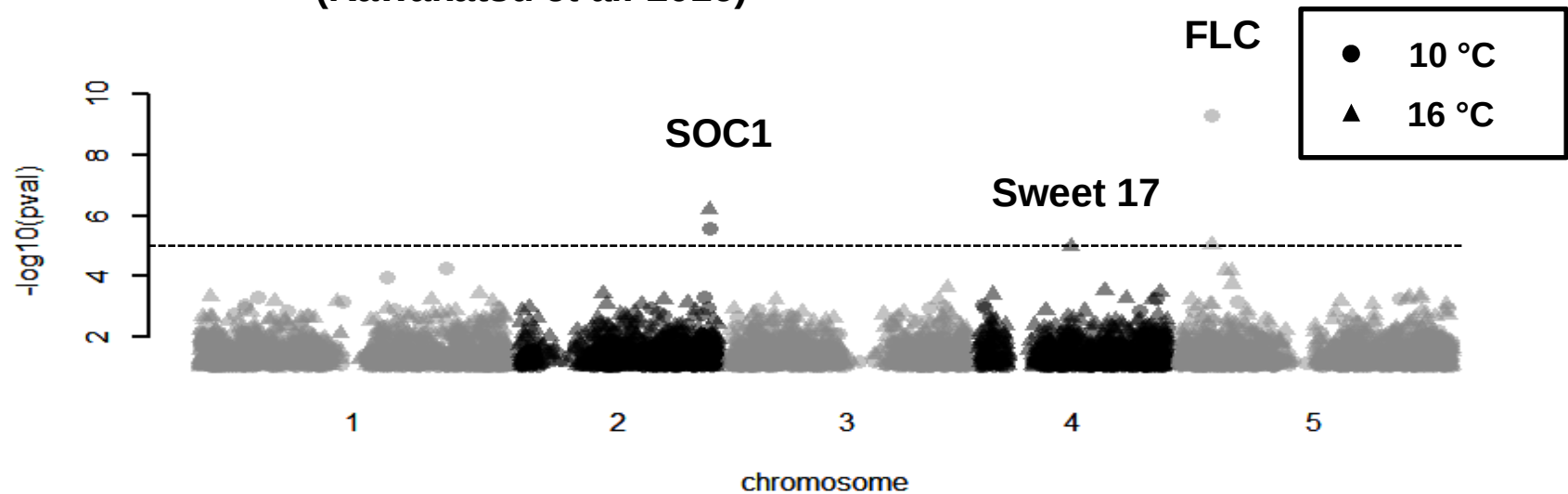
GWAS : Genome - Phenotype

TWAS : Transcriptome - Phenotype



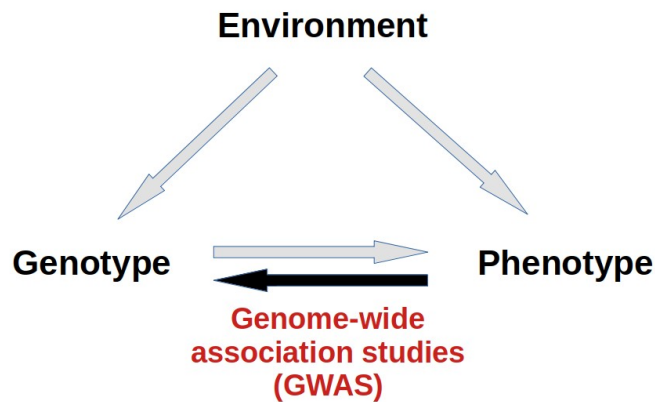
TWAS (Transcriptome – Phenotype) for Flowering Time

Using RNA expression to explain the phenotype
(Kawakatsu *et al.* 2016)



Beyond GWAS: Understanding the function of pathways

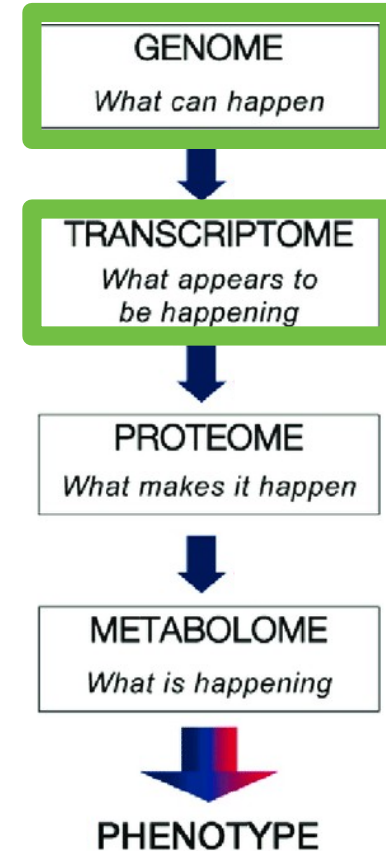
The Omnics cascade



GWAS : Genome - Phenotype

TWAS : Transcriptome - Phenotype

eGWAS: Genome - Transcriptome



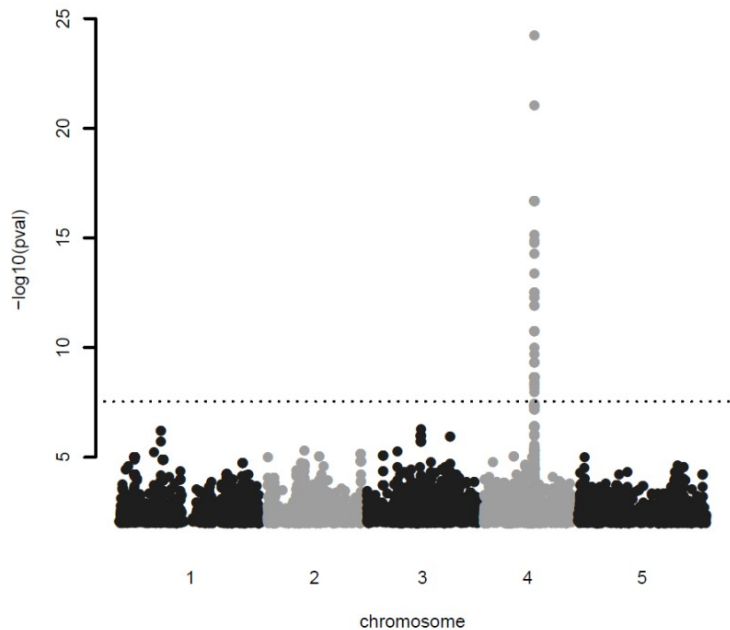
stable

instable

eGWAS (Genome - Transcriptome)

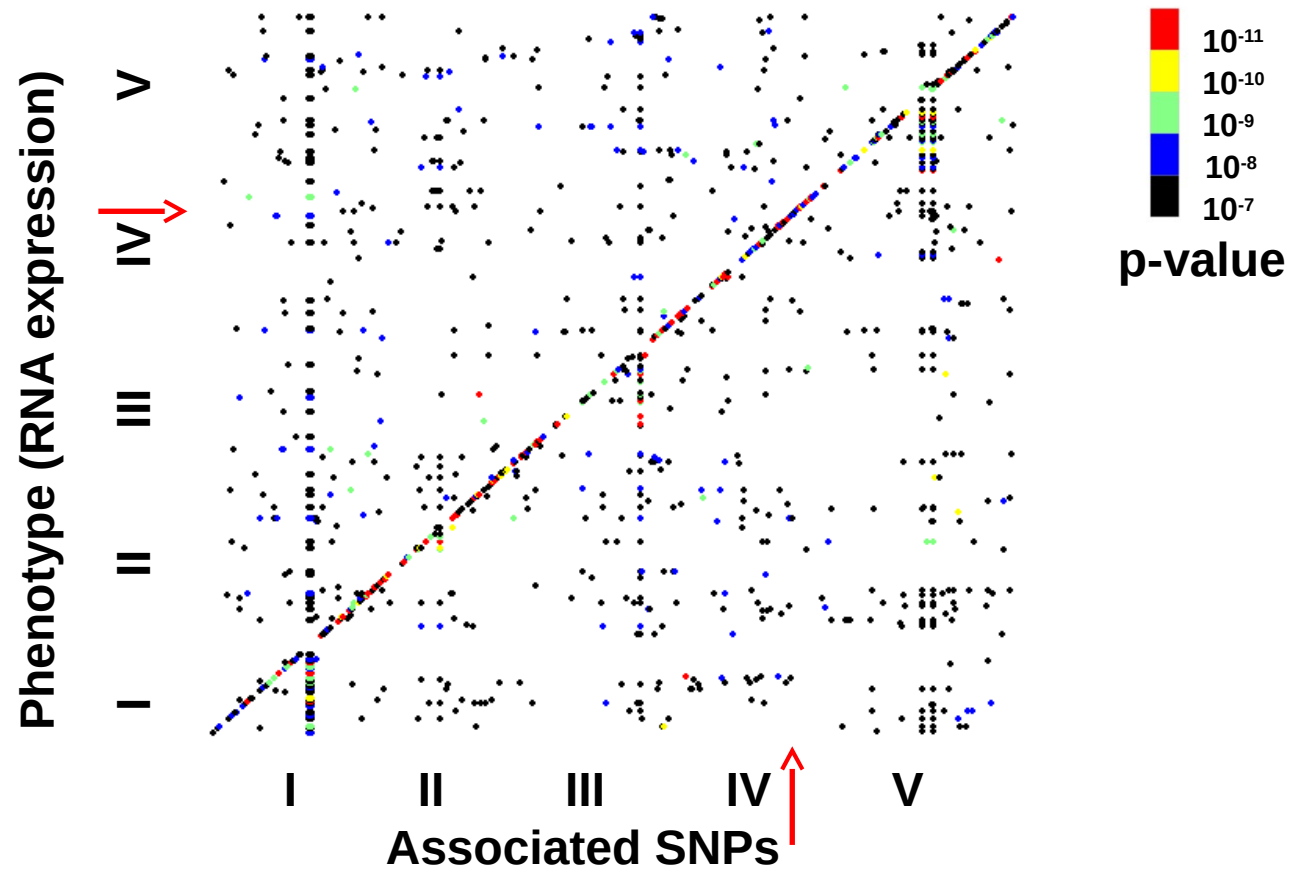
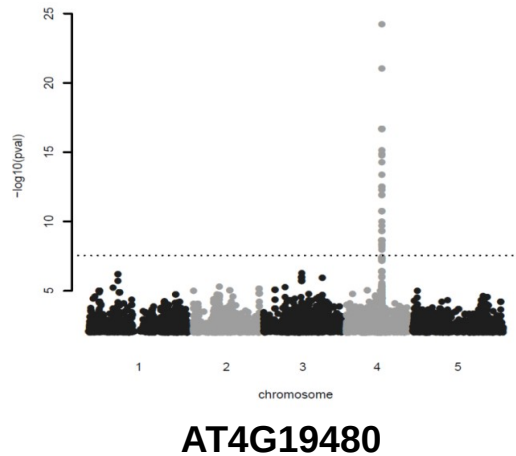
Data: 80 ecotypes grown under drought and control conditions (VIB Ghent), RNAseq for all of them.

Use the expression values of each gene as molecular phenotype



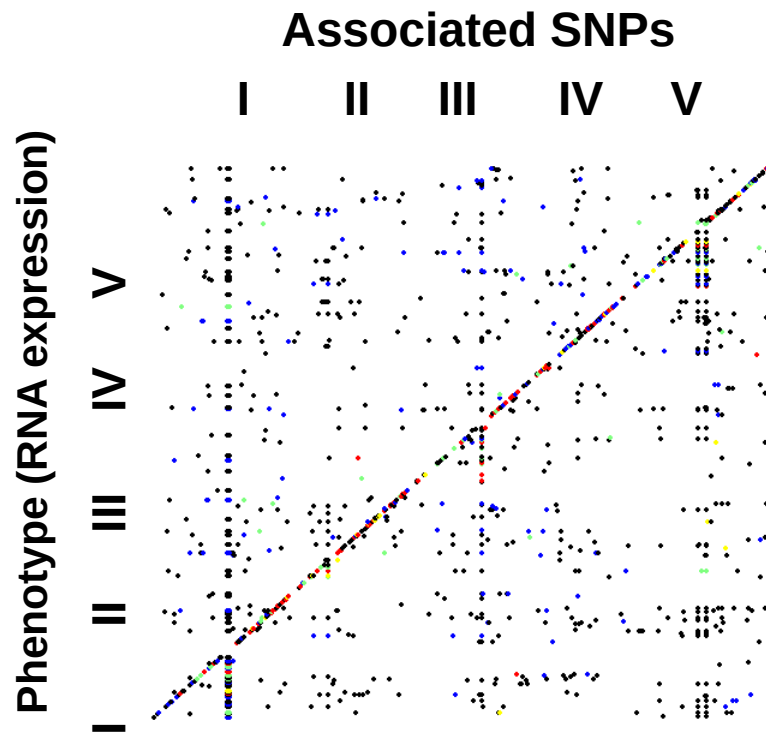
**A SNP on Chromosome 4 affects the expression of the gene AT4G19480
How does this affect the phenotype ?**

eGWAS, summary plot of 28k analyses

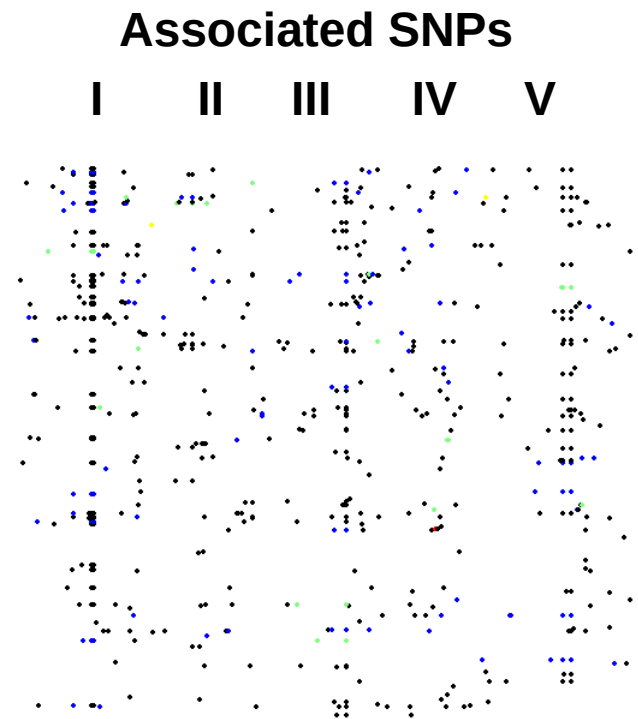


eGWAS with a multivariate model

Joint analyses of RNAseq data from control and drought-stressed samples
Method : Korte *et al.* (2012) NatGen



Environment independent



G x E

Clauw *et al.* (2016) Plant Cell

Global genetic heterogeneity in adaptive traits



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

[HOME](#) | [ABOUT](#) | [SUBMIT](#) | [NEWS & NOTES](#) | [ABOUT](#)
| [CHANNELS](#)

bioRxiv is receiving many new papers on coronavirus SARS-CoV-2. A reminder: these are preliminary reports that have not been peer-reviewed. They should not be used to guide clinical practice/health-related behavior, or be reported in news media as established information.

New Results

[Comment on this paper](#)

[Previous](#)

Global genetic heterogeneity in adaptive traits

Posted February 26, 2021.

William Andres Lopez-Arboleda, Stephan Reinert,  Magnus Nordborg,  Arthur Korte

doi: <https://doi.org/10.1101/2021.02.26.433043>

[Download PDF](#)

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

Abstract

[Info/History](#)

[Metrics](#)

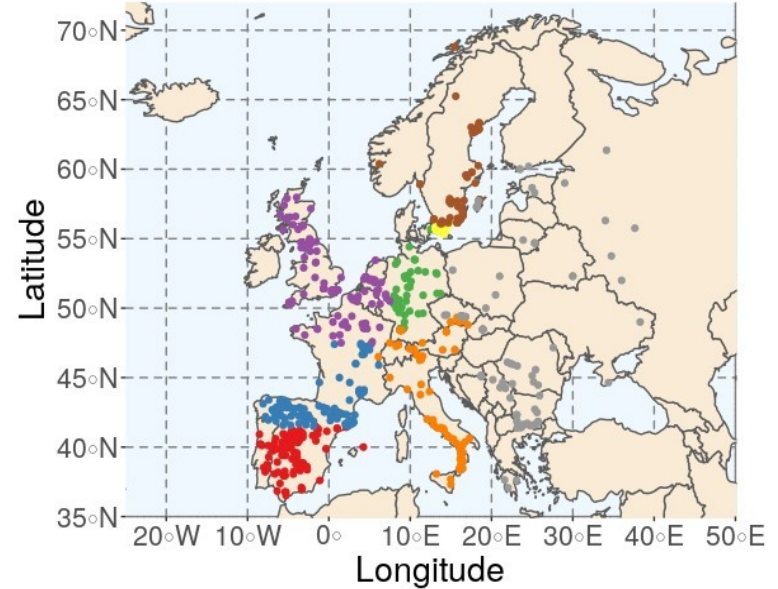
[Preview PDF](#)

[Tweet](#)

[Gefällt mir 0](#)

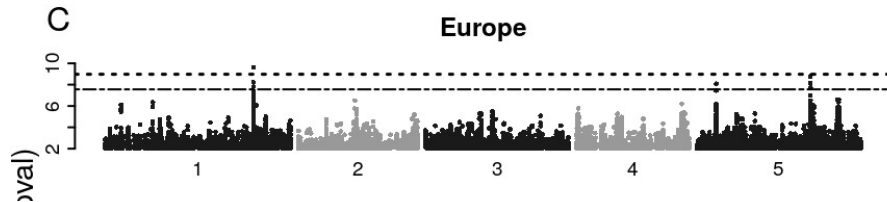
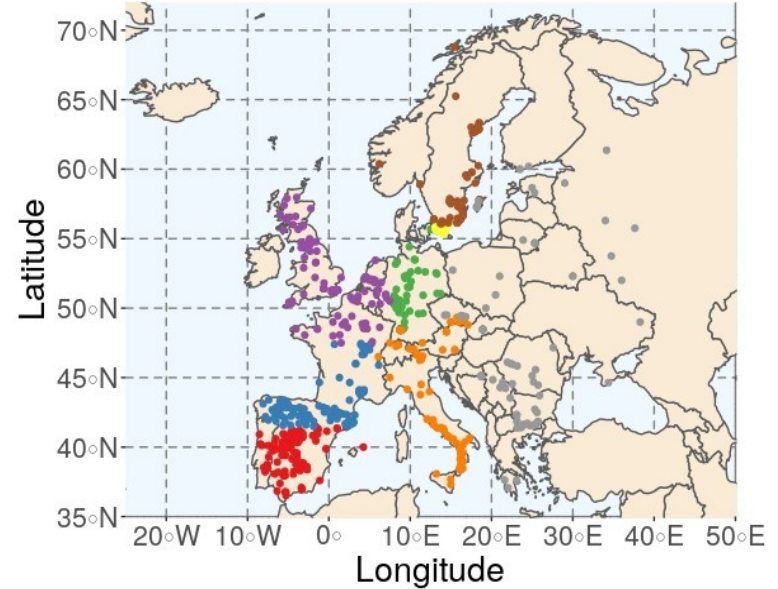
Global genetic heterogeneity in adaptive traits

We measured flowering time in a growth chamber at 10°C for 888 European *Arabidopsis* ecotypes



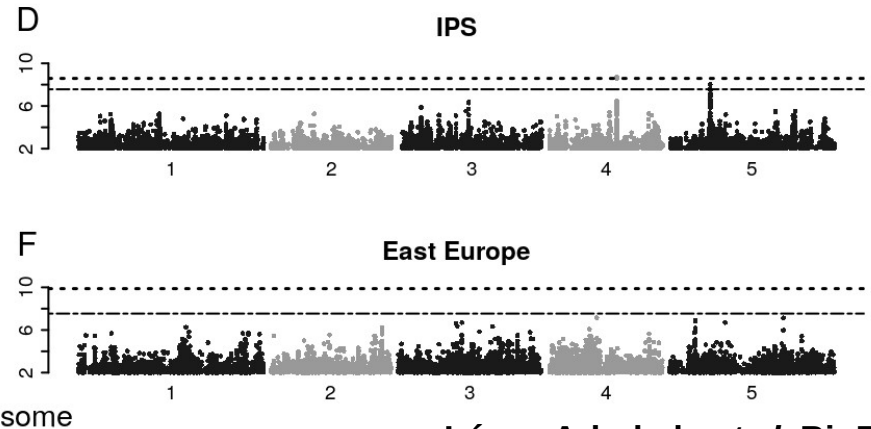
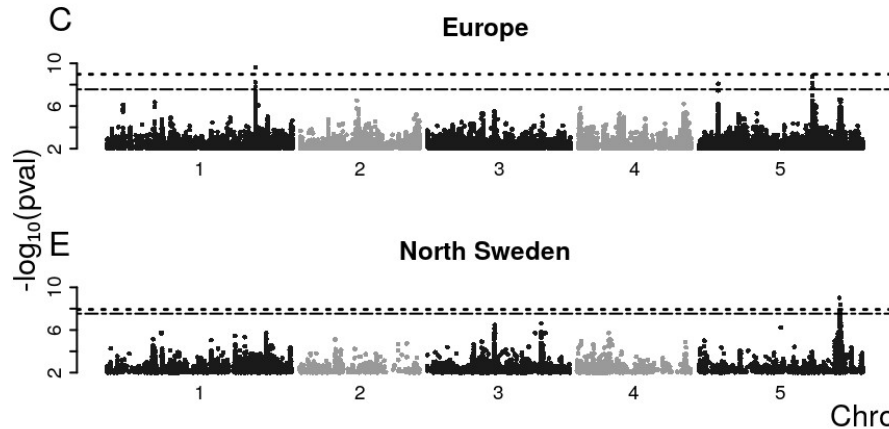
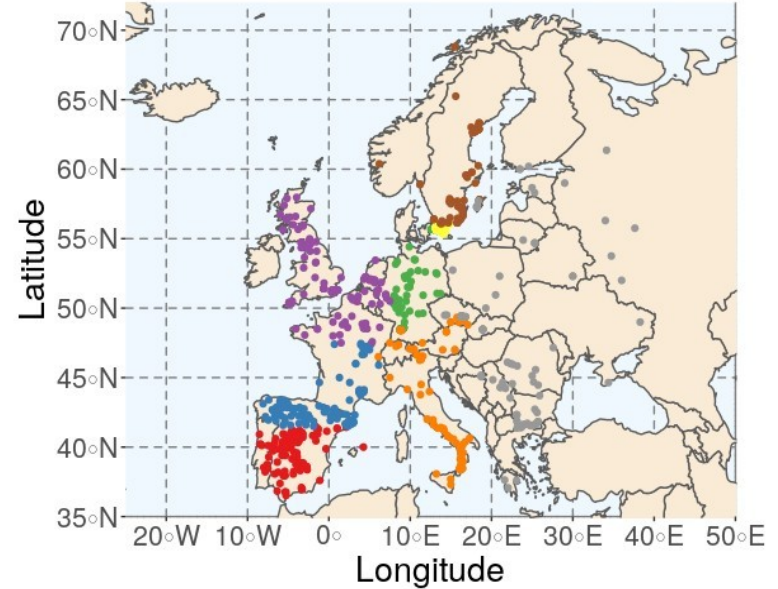
Global genetic heterogeneity in adaptive traits

We measured flowering time at 10°C
In 888 European *Arabidopsis* ecotypes



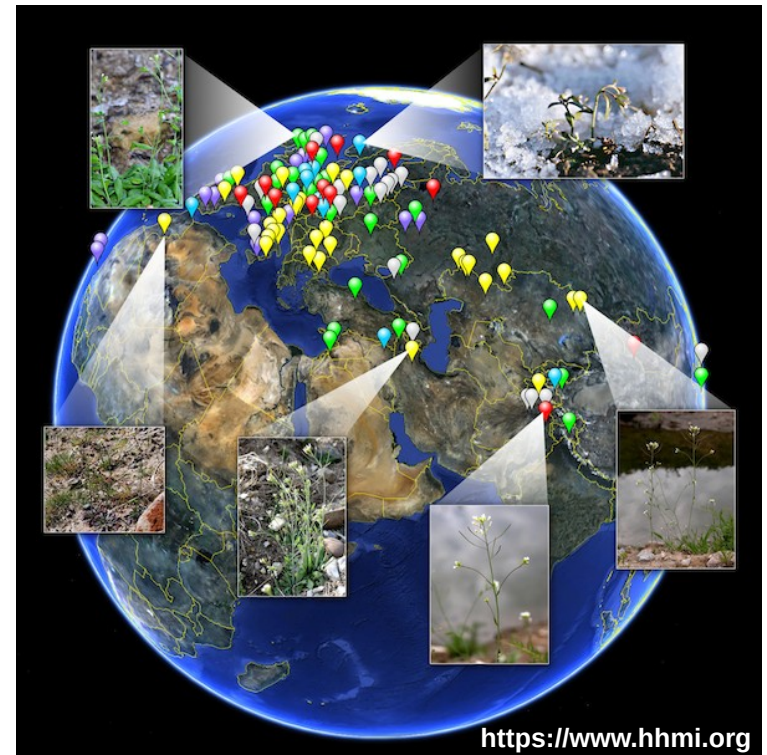
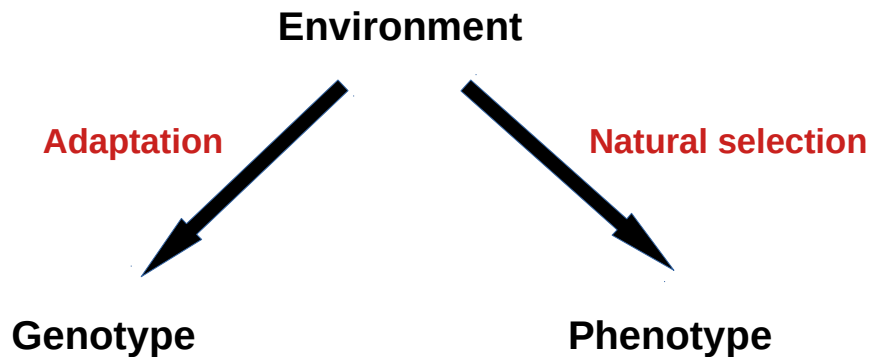
Global genetic heterogeneity in adaptive traits

We measured flowering time at 10°C
In 888 European *Arabidopsis* ecotypes



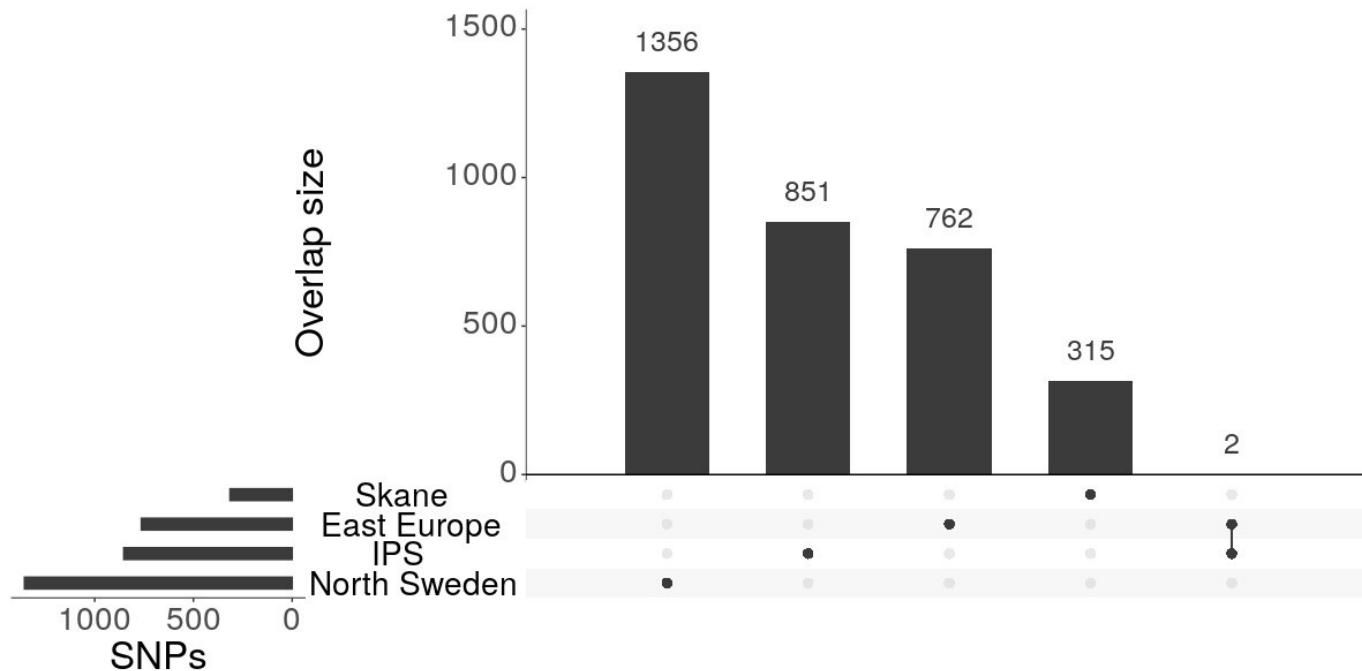
Genotype – Phenotype relationship

Is Flowering time optimized by different genes in different local subsets ?

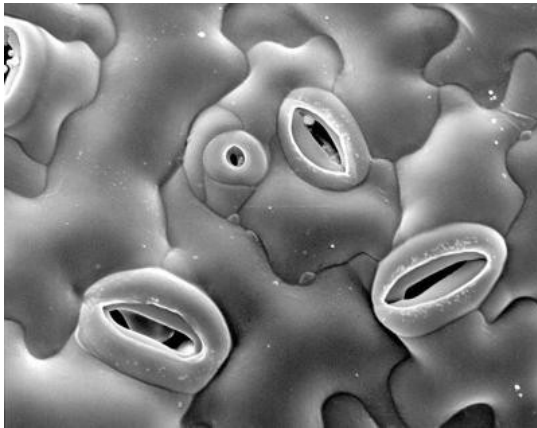


Different genetic architecture of flowering time in different subsets

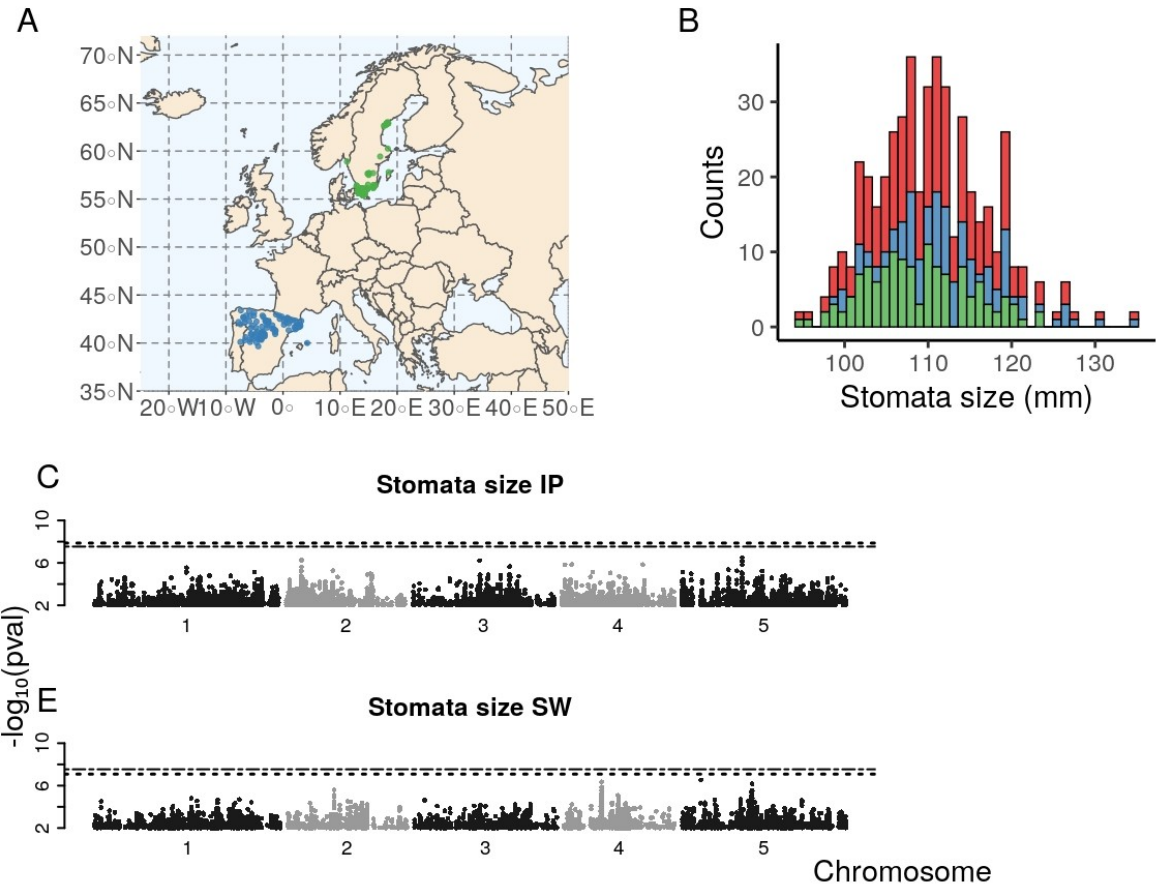
Overlap of SNPs that are associated with a p -value $< 10^{-4}$ in the respective subset



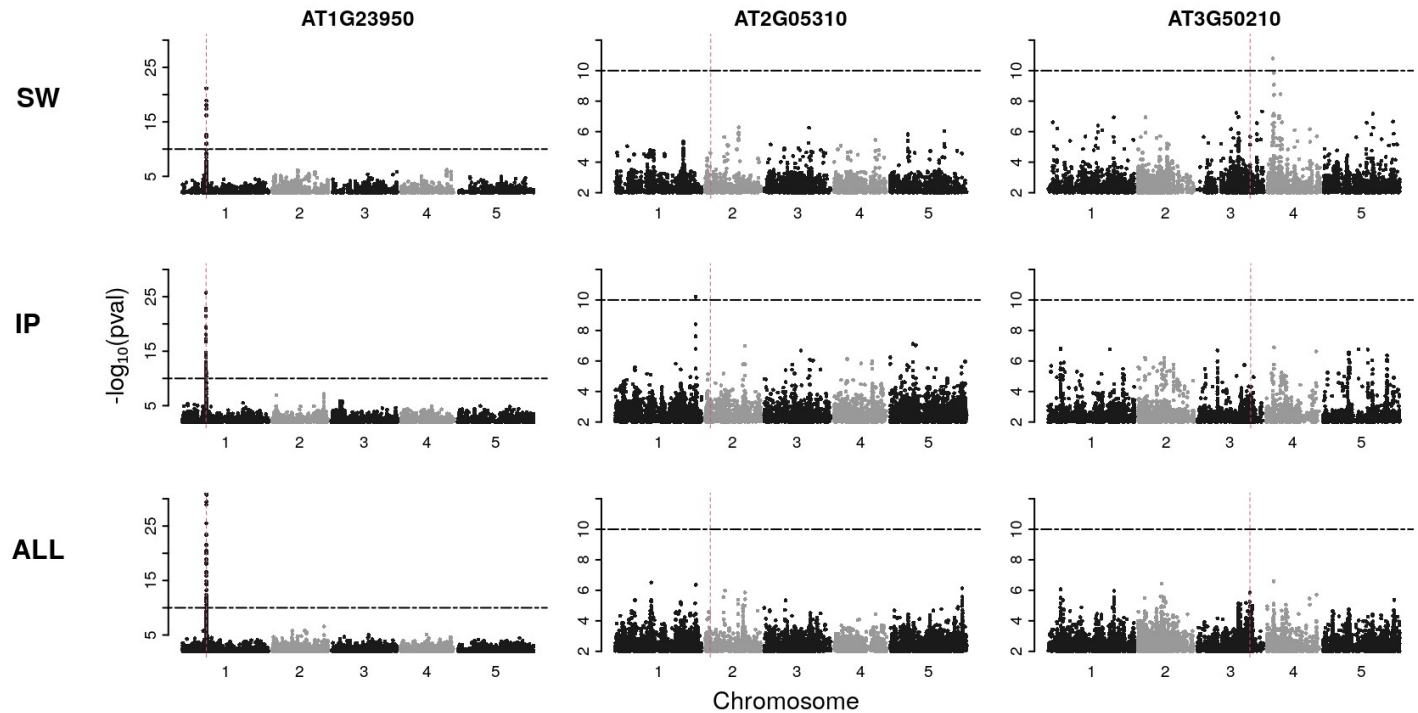
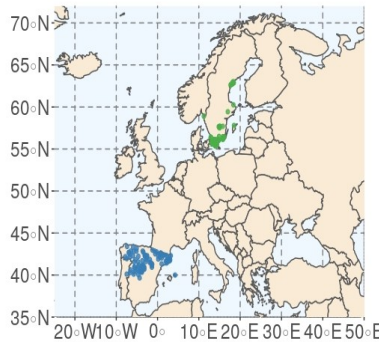
Different genetic architecture in many adaptive traits



Stomata size was measured in a set of 330 European *Arabidopsis* ecotypes
Dittberner *et al.* 2018



eGWAS in different local populations

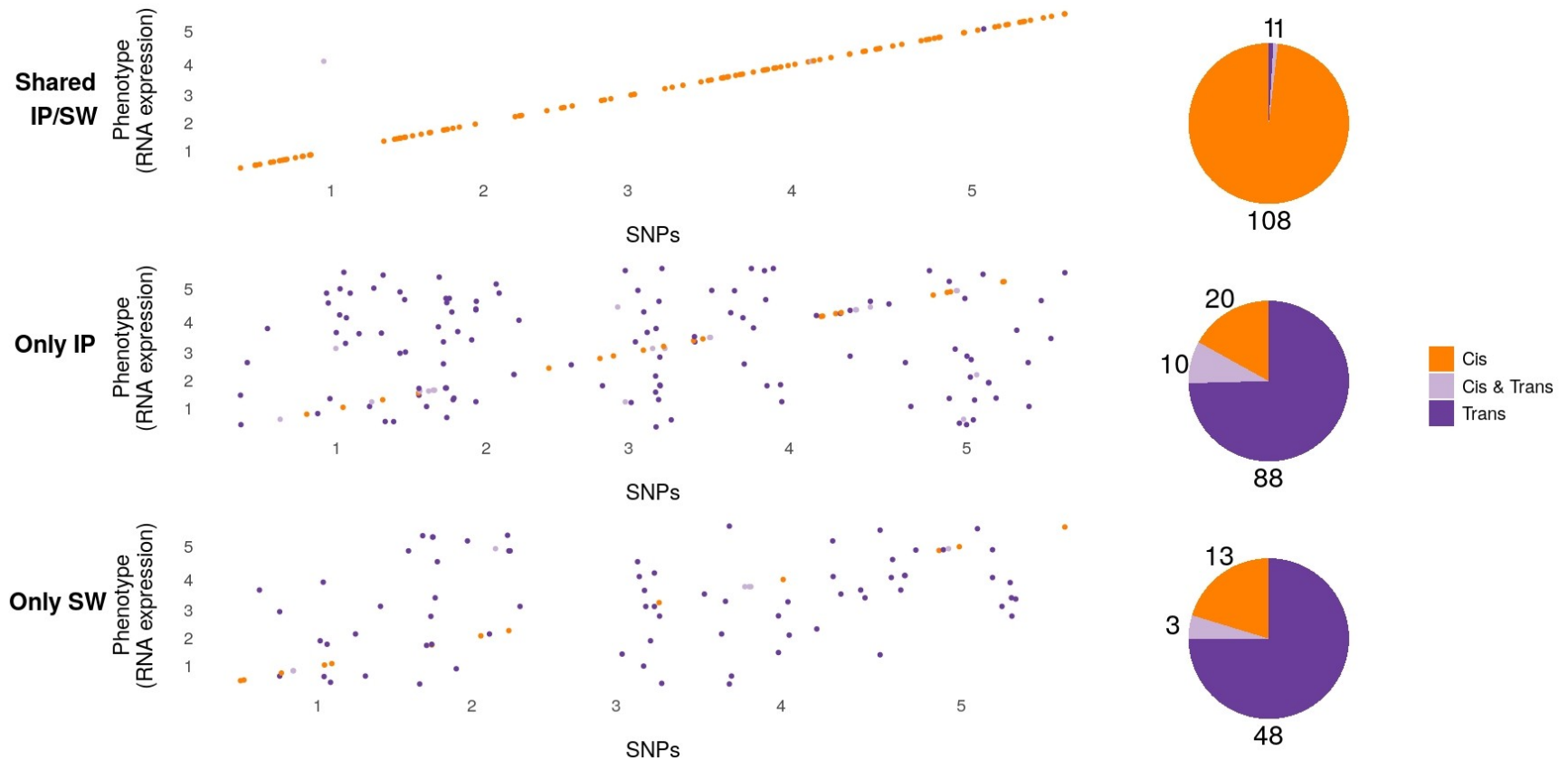


110 genes

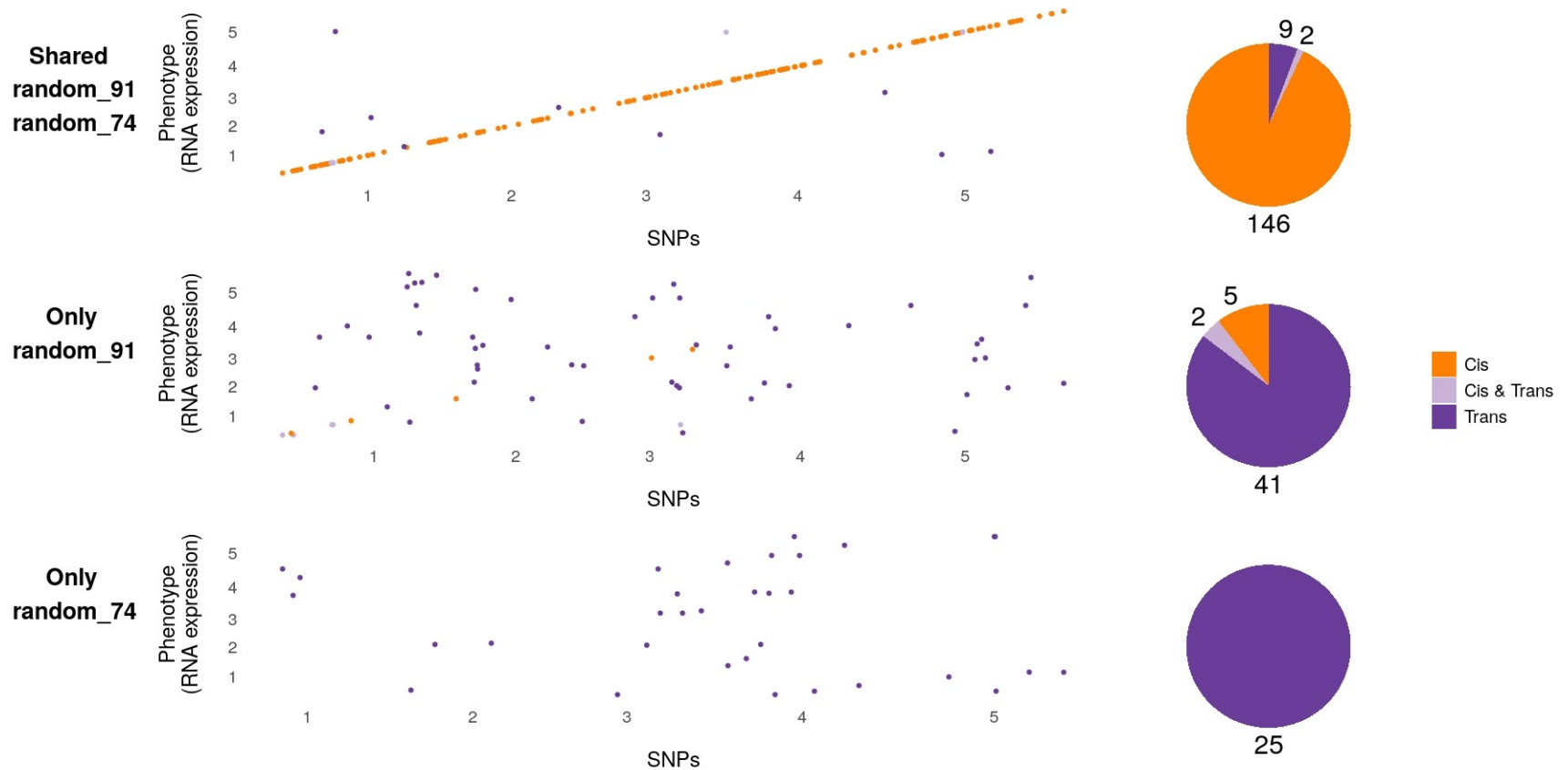
118 genes

64 genes

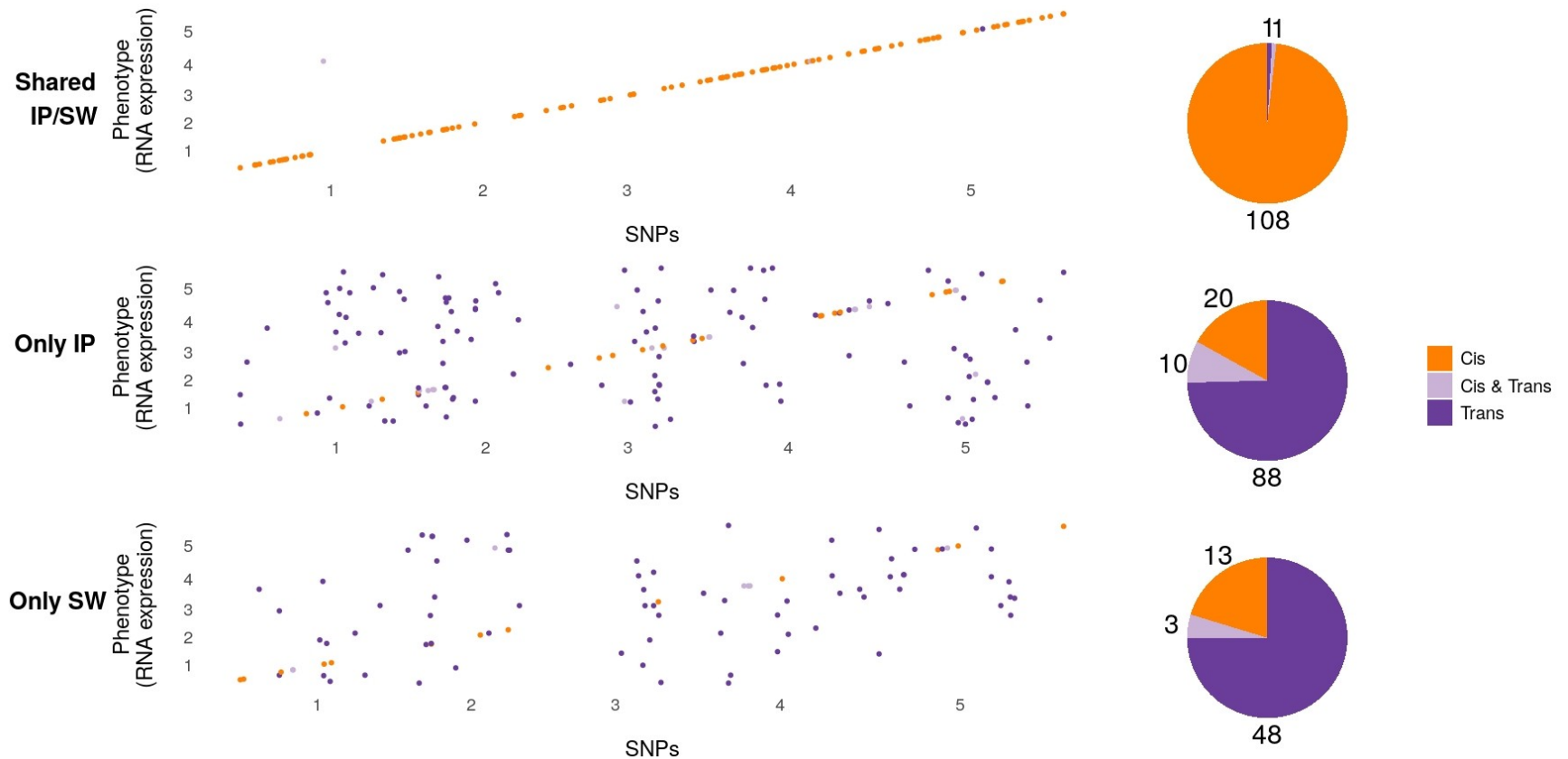
eGWAS in different local populations



eGWAS in different non-local populations



eGWAS in different local populations



Summary I

**What you conclude from GWAS will depend
on where you sample**

**True for humans as well (e.g. genetic
architecture for skin color differ dramatically
between Iceland and Cape Verde Islands)**

Summary II

Can we infer how local adaptation works ?

Putative local adaptive regulatory alleles are mostly regulated in *trans*

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**

Co-occurrence of premature Stop codons

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

Co-occurrence of premature Stop codons

	prematSTOP
Knocked-out together more often (over)	2917
Knocked-out together less frequently (under)	138

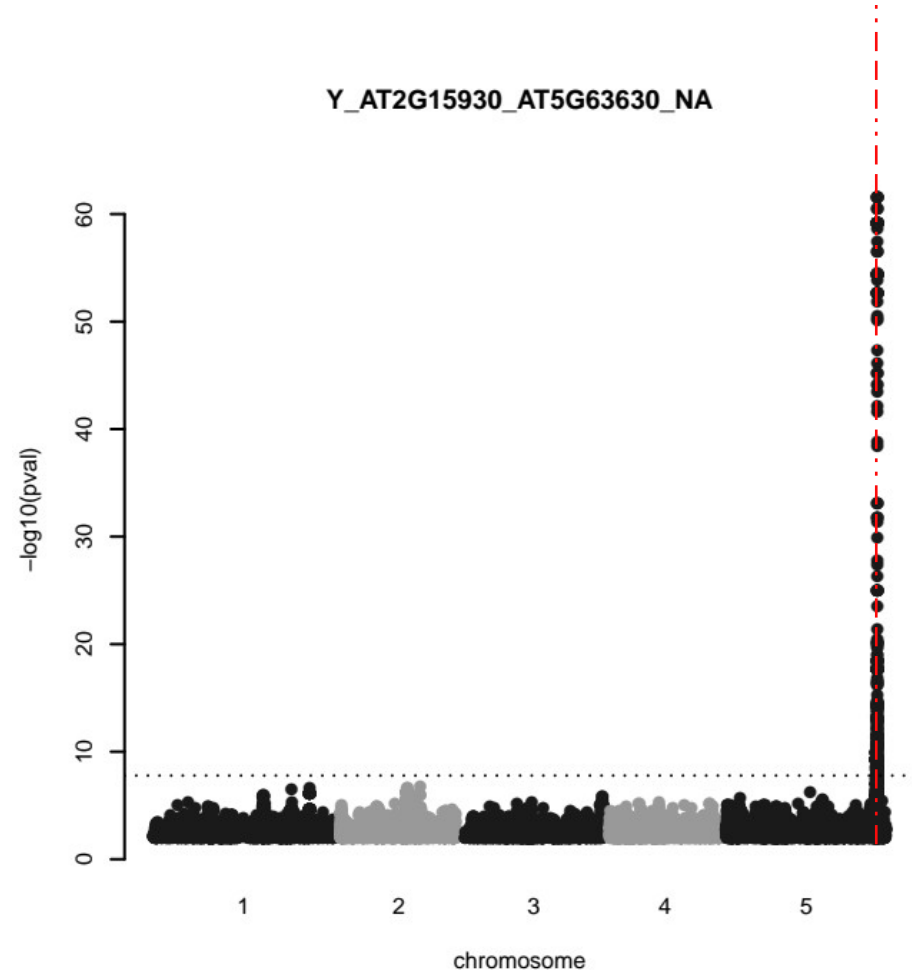
Co-occurrence of premature Stop codons

	prematSTOP	SYN	NON_SYN
Knocked-out together more often (over)	2917	1216	1086
Knocked-out together less frequently (under)	138	46	59

More connections for functional variants compared to (potential) neutral variants

GWAS on the Co-occurrence

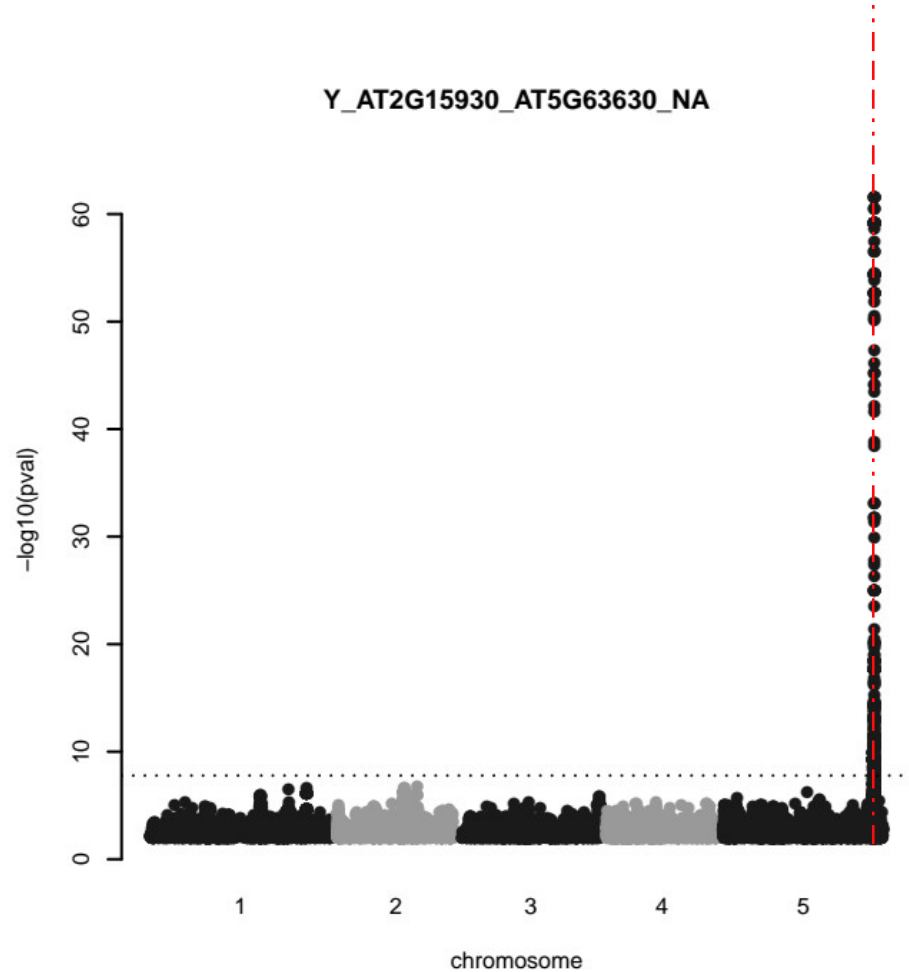
Create a binary trait (0|1)
which indicates if in a
given accessions both
genes contain functional
premature STOP
codons or if only one
gene is knocked out



GWAS on the Co-occurrence

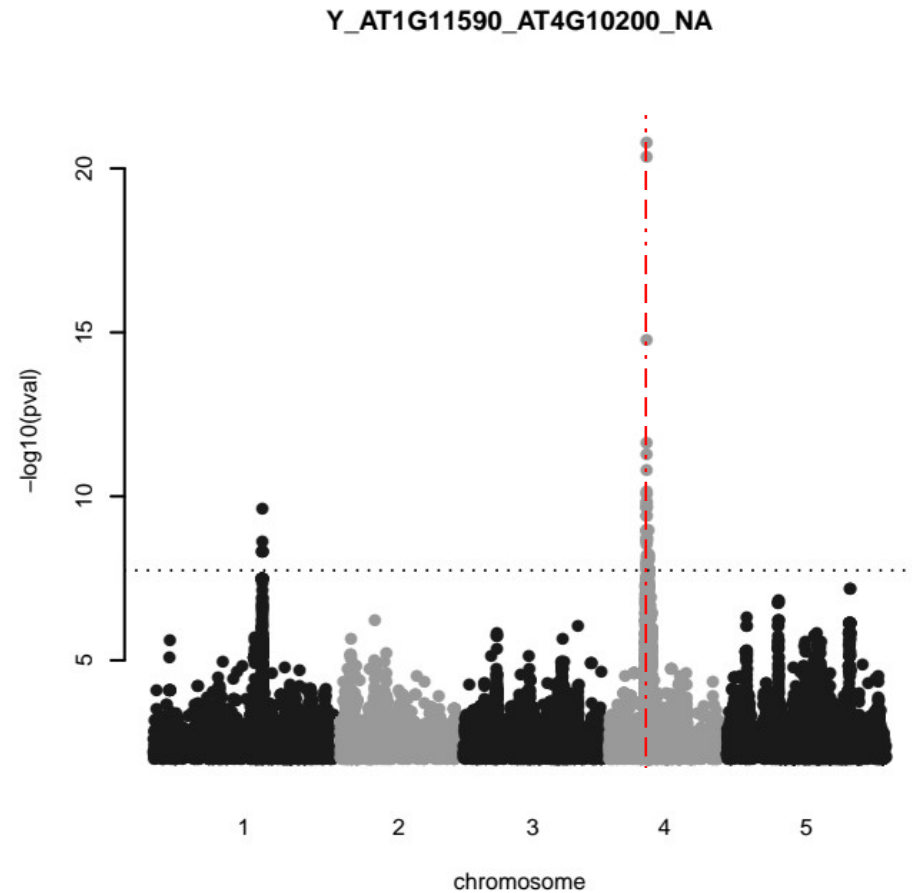
Create a binary trait (0|1)
which indicates if in a
given accessions both
genes contain functional
premature STOP
codons or if only one
gene is knocked out

Signal in cis = expected
(5569/5834 for over
and
209/276 for under)



GWAS on the Co-occurrence

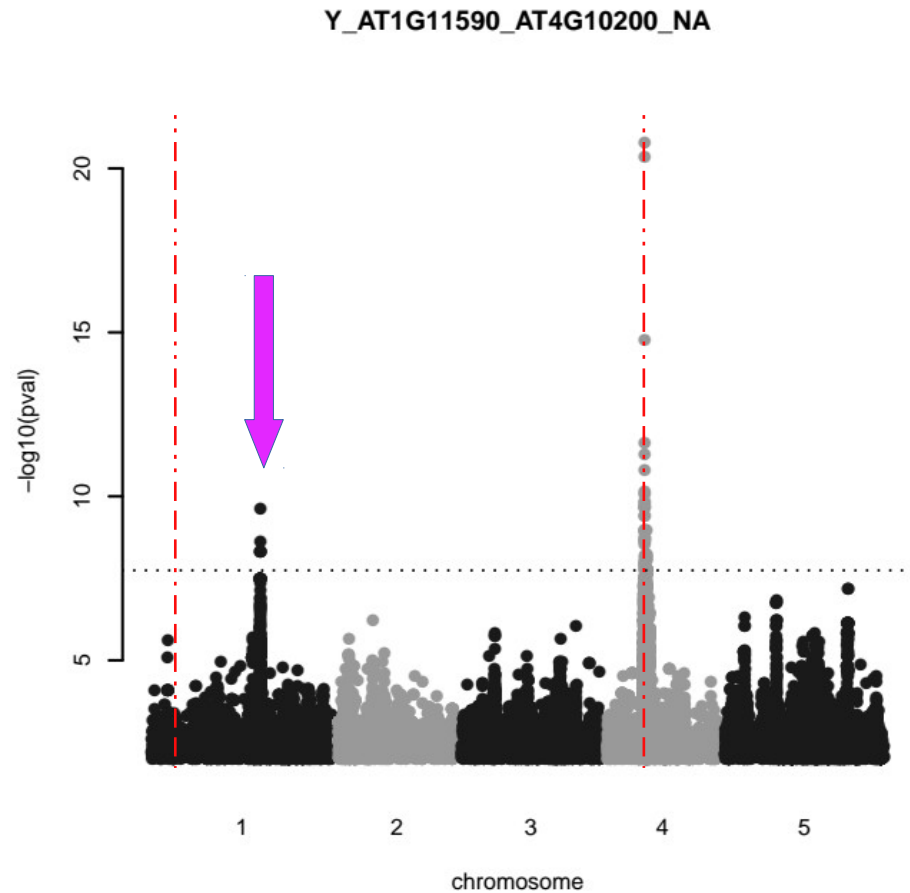
Create a binary trait (0|1)
which indicates if in a
given accessions both
genes contain functional
premature STOP
codons or if only one
gene is knocked out



GWAS on the Co-occurrence

Create a binary trait (0|1)
which indicates if in a
given accessions both
genes contain functional
premature STOP
codons or if only one
gene is knocked out

Additional signal in trans
(212/5834 for over
and
8/276 for under)



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



**In collaboration with
Pascal Braun (INET, HMGU)**

Summary III

**Tremendous amount of Natural Variation in
signaling pathways**

**If we want to understand G-P, we need to
understand the underlying gene regulatory
networks**



European Maize

SPONSORED BY THE



Federal Ministry
of Education
and Research

JMU Würzburg

Jan Freudenthal

Ammarah Anwar

William López Arboleda

Stephan Reinert

Laura Steinmann

