



AraPheno and the AraGWAS Catalog 2020: A major database update including RNA-seq and knockout mutation data for *Arabidopsis thaliana*



@arthurkorte



arthurkorte

Arthur Korte

San Diego – PAG 2020

Acknowledgment

Published online 23 October 2019

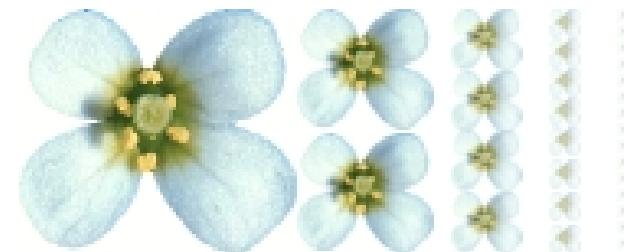
Nucleic Acids Research, 2020, Vol. 48, Database issue D1063–D1068
doi: 10.1093/nar/gkz925

AraPheno and the AraGWAS Catalog 2020: a major database update including RNA-Seq and knockout mutation data for *Arabidopsis thaliana*

Matteo Togninalli^{1,2,†}, Ümit Seren^{3,†}, Jan A. Freudenthal⁴, J. Grey Monroe⁵, Dazhe Meng^{3,6}, Magnus Nordborg³, Detlef Weigel⁵, Karsten Borgwardt^{1,2}, Arthur Korte^{④,*} and Dominik G. Grimm^{⑦,8,*}

¹Machine Learning and Computational Biology Lab, Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland, ²Swiss Institute of Bioinformatics, Basel, Switzerland, ³Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria, ⁴Center for Computational and Theoretical Biology, University Würzburg, Würzburg, Germany, ⁵Max Planck Institute for Developmental Biology, Tübingen, Germany, ⁶Google, Mountain View, USA, ⁷Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Straubing, Germany and ⁸Weihenstephan-Triesdorf University of Applied Sciences, Bioinformatics, Straubing, Germany

Received September 06, 2019; Revised September 26, 2019; Editorial Decision October 04, 2019; Accepted October 08, 2019

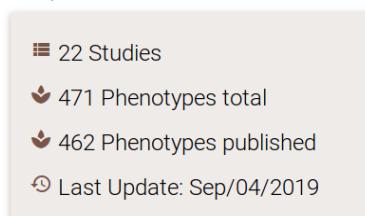


www.1001genomes.org

AraPHENO and AraGWAS



The screenshot shows the AraPheno homepage. At the top, there are two tabs: "AraPheno" (selected) and "AraRNASeq". On the right, there are links for "Data", "Tools", and "About". Below the tabs is a large green background image of Arabidopsis thaliana plants. In the center, the word "AraPheno" is displayed in white. Below it, a text block says: "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." At the bottom of the main area is a search bar with the placeholder "Search a phenotype, study, trait ontology (e.g. type FRI for phenotype, Atwell for study, or concentration for ontology) or accession name".



A modal window titled "Quick Stats" is open, showing the following data:

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



The screenshot shows the AraGWAS Catalog homepage. At the top, there is a navigation bar with links for "TAKE A TOUR?", "FAQ", and "ABOUT". Below the navigation is a large green background image of Arabidopsis thaliana flowers. In the center, the words "AraGWAS Catalog" are displayed in white. Below the title, a text block says: "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."

<https://arapheno.1001genomes.org/>

<https://aragwas.1001genomes.org>

AraPHENO

https://arapheno.1001genomes.org

AraPheno AraRNaseq

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

Public Studies
Browse through all available public *Arabidopsis thaliana* studies and retrieve additional information.
[VIEW STUDIES](#)

Public Phenotypes
Browse through all available public *Arabidopsis thaliana* phenotypes and studies and retrieve additional information.
[VIEW PHENOTYPES](#)

Master Accession Table
Browse through all *Arabidopsis thaliana* accessions and find out how many phenotypes are available for which accessions.
[VIEW ACCESSIONS](#)

Submit Phenotypes
Submit and publish your own studies and phenotypes in order to share it with the community.
[SUBMIT PHENOTYPES](#)

Download Database
Here you can download the full database, including all phenotypes and meta-information.
[DOWNLOAD DATABASE](#)

REST API
The REST API can be used to retrieve phenotype data and meta-information from AraPheno via URLs.
[REST API](#)

Take a Tour
Take a guided tour through AraPheno and familiarise yourself with all available public functions.
[TAKE A TOUR](#)

FAQ & Tutorials
Here, you can find detailed help about the functions of AraPheno, its integrated data and frequently asked questions.
[FAQ](#)

Feedback & Issues
Here, you can submit issues, feedback or any related questions regarding AraPheno.
[FEEDBACK](#)

462 Phenotypes

AraPheno



AraRNASeq

Phenotype: FT10

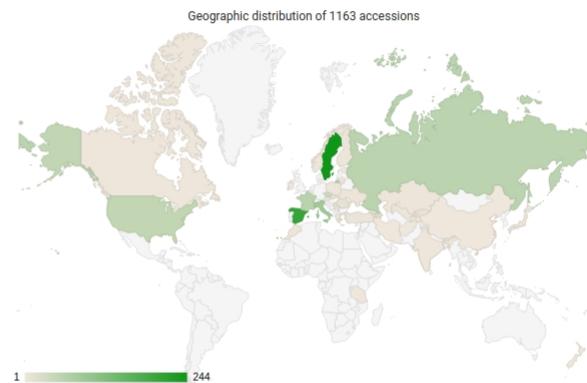


ID:	261
Study:	1001genomes flowering time phenotypes
DOI:	10.21958/phenotype:261 (# Cite)
AraGWAS	AraGWAS:261
Associations:	
Scoring:	Flowering time was scored as days until first open flower
Growth conditions:	Seeds for all 1135 accessions were surface-sterilized with chlorine gas. Seeds were distributed in pots with four replicates in a randomized block design, each replicate corresponding to one block. Plants were grown in growth chambers with the following settings: after 6 days of stratification in the dark at 4°C, constant temperature of 10°C with 16 hours light / 8 hours darkness, 65% humidity. All trays within a block were moved to a new shelf and rotated 180° every other day to minimize position effects.
Trait-Ontology:	days to flowering trait (TO:0000344)
Environment ontology:	growth chamber study (EO:0007269)
Unit:	day (UO:0000033)
Shapiro-Wilk*:	'4.62e-18' (p-value)
Submission date:	Aug. 14, 2016, 2:08 p.m.

* The Shapiro-Wilk test tests the null hypothesis that the data was drawn from a normal distribution.

Phenotype Publications

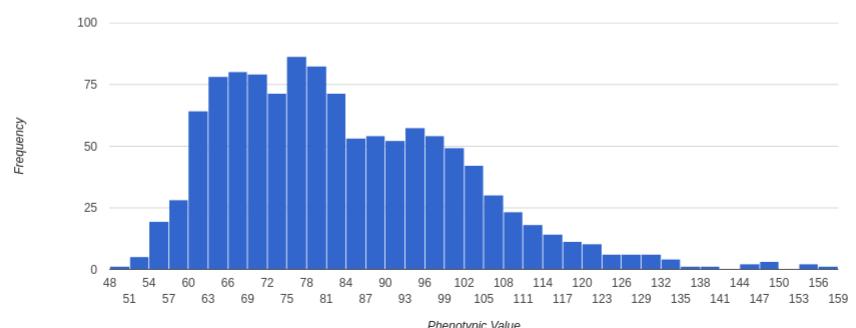
[1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*](#)
The 1001 Genomes Consortium
Cell. 2016 166(2): 481-491. doi: 10.1016/j.cell.2016.05.063



HISTOGRAM

EXPLORER

TABLE

[Show transformations](#)

AraRNASeq

The screenshot shows the AraPheno website homepage. At the top, there is a header bar with a URL bar showing <https://arapheno.1001genomes.org>, a battery icon at 67%, and a star icon. Below the header, there are two tabs: "AraPheno" (selected) and "AraRNASeq". The main content area has a green background image of a plant. The title "AraPheno" is centered in large white font. Below it, a subtitle says "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." A search bar is present. Below the search bar, there are several sections with icons and buttons:

- Public Studies**: Browse through all available public *Arabidopsis thaliana* studies and retrieve additional information. Button: [VIEW STUDIES](#).
- Public Phenotypes**: Browse through all available public *Arabidopsis thaliana* phenotypes and studies and retrieve additional information. Button: [VIEW PHENOTYPES](#).
- Master Accession Table**: Browse through all *Arabidopsis thaliana* accessions and find out how many phenotypes are available for which accessions. Button: [VIEW ACCESSIONS](#).
- Submit Phenotypes**: Submit and publish your own studies and phenotypes in order to share it with the community. Button: [SUBMIT PHENOTYPES](#).
- Download Database**: Here you can download the full database, including all phenotypes and meta-information. Button: [DOWNLOAD DATABASE](#).
- REST API**: The REST API can be used to retrieve phenotype data and meta-information from AraPheno via URLs. Button: [REST API](#).
- Take a Tour**: Take a guided tour through AraPheno and familiarise yourself with all available public functions. Button: [TAKE A TOUR](#).
- FAQ & Tutorials**: Here, you can find detailed help about the functions of AraPheno, its integrated data and frequently asked questions. Button: [FAQ](#).
- Feedback & Issues**: Here, you can submit issues, feedback or any related questions regarding AraPheno. Button: [FEEDBACK](#).

RNAseq data

AraPheno



AraRNASEq

D

AraRNASEq

AraPheno also contains *Arabidopsis thaliana* RNASeq measurements.

This database allows to look for them.



Search a gene, study, or accession name



Public RNASeq Studies

Browse through all available public *Arabidopsis thaliana* RNASeq studies and retrieve additional information.

[VIEW STUDIES](#)



Public RNASeq data

Browse through all available public *Arabidopsis thaliana* RNASeq values and retrieve additional information.

[VIEW RNASEQ DATA](#)



FAQ & Tutorials

Here, you can find detailed help about the functions of AraPheno, its integrated data and frequently asked questions.

[FAQ](#)



Submit RNASeq

Submit your own RNASeq data in order to share it with the community

[SUBMIT RNASEQ](#)

News & Updates

Quick Stats

AraPheno and the AraGWAS Catalog 2020: New paper

2 RNASeq studies

RNAseq data

AraPheno



AraRNASEq

RNAseq: AT1G01010

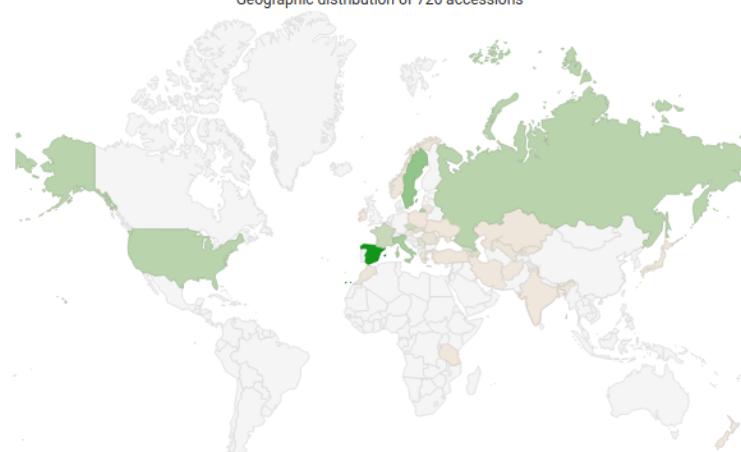


Name: AT1G01010

Geographic distribution of 726 accessions

ID: 40232

Growth Condition: None

Study: Epigenomic Diversity in *A. thaliana* (Kawakatsu et al. 2016)

DOI: 10.21958/maseq:40232 (Cite)

Scoring: TPM

Shapiro-Wilk*: '1.17e-26' (p-value)

Submission date: Sept. 3, 2019, 4:54 p.m.

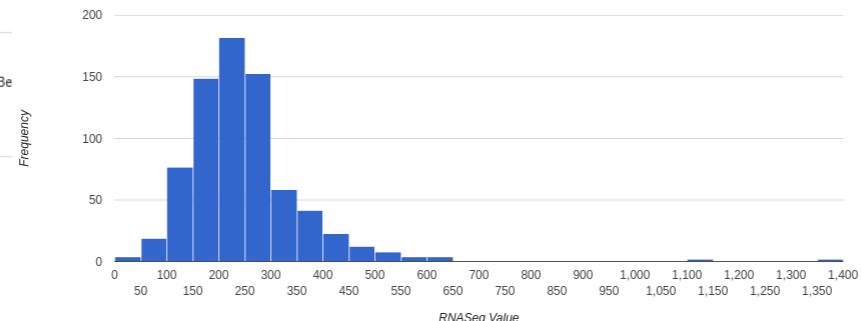
* The Shapiro-Wilk test tests the null hypothesis that the data was drawn from a normal distribution.

RNASeq Publications

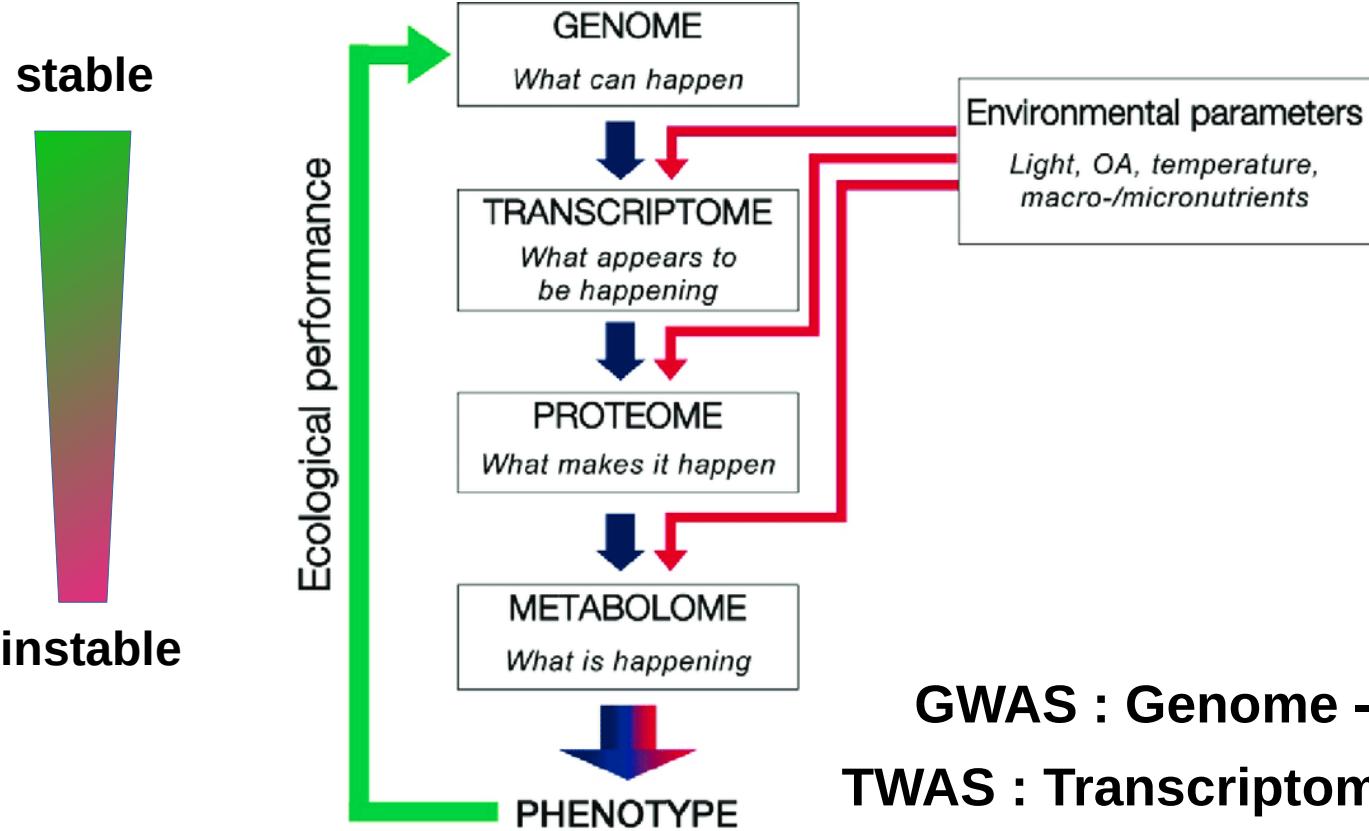
Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions.

Kawakatsu T, Huang SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, Chen H, Dubin M, Lee CR, Wang C, Bemm F, Be Consortium, Schork NJ, Weigel D, Nordborg M, Ecker JR

Cell. 2016;166:492-505. doi: 10.1016/j.cell.2016.06.044



The Omnics cascade



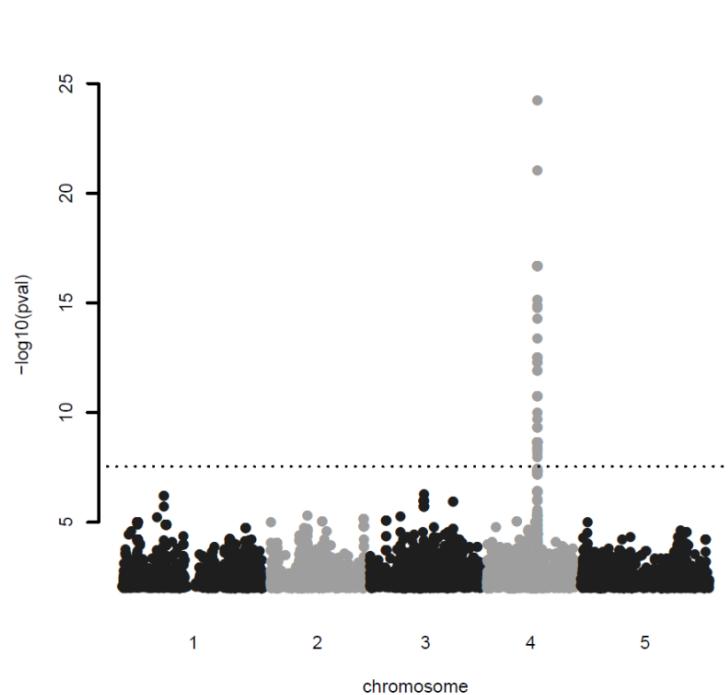
GWAS : Genome - Phenotype

TWAS : Transcriptome - Phenotype

eGWAS : Genome - Transcriptome

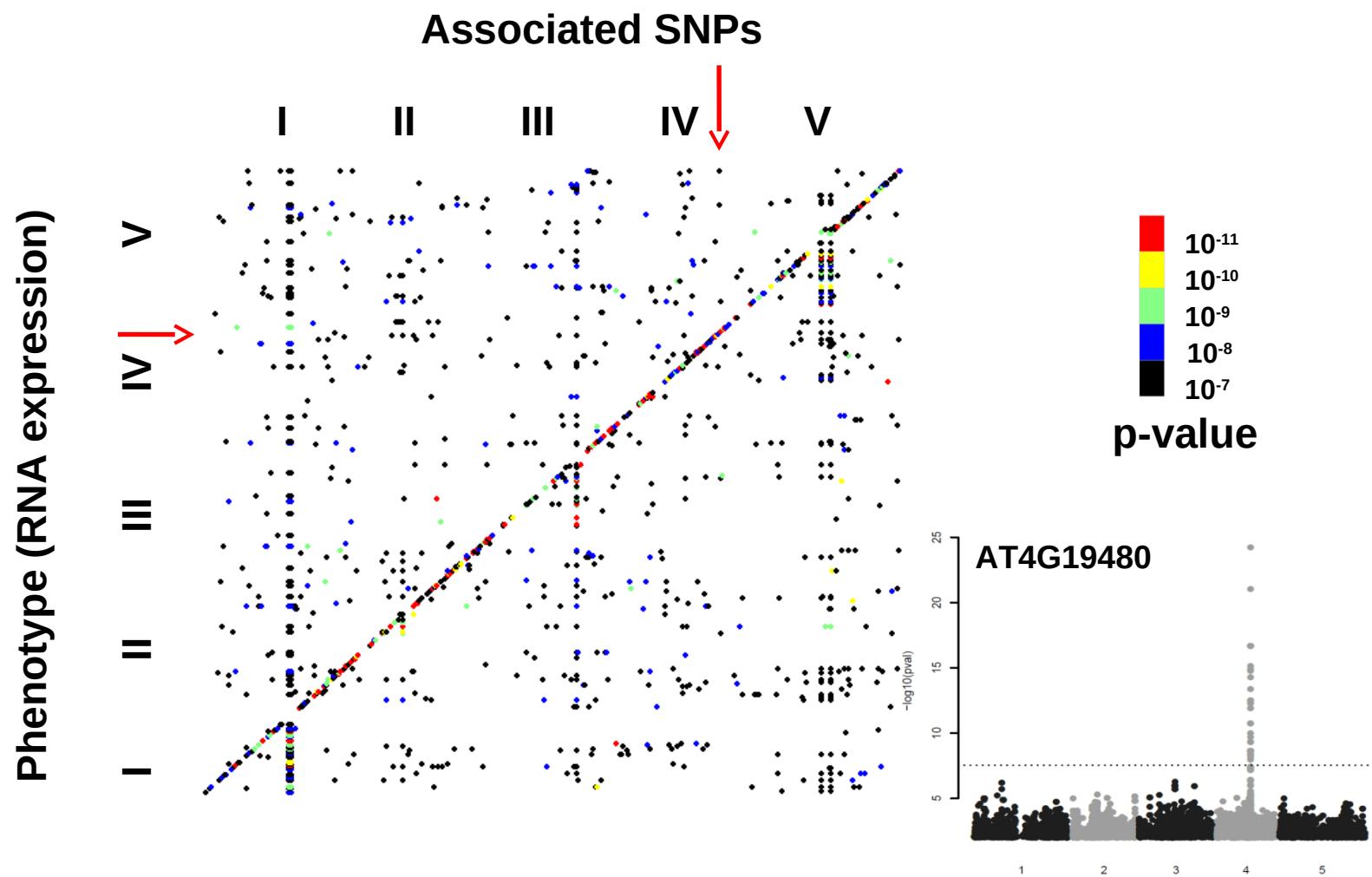
eGWAS

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



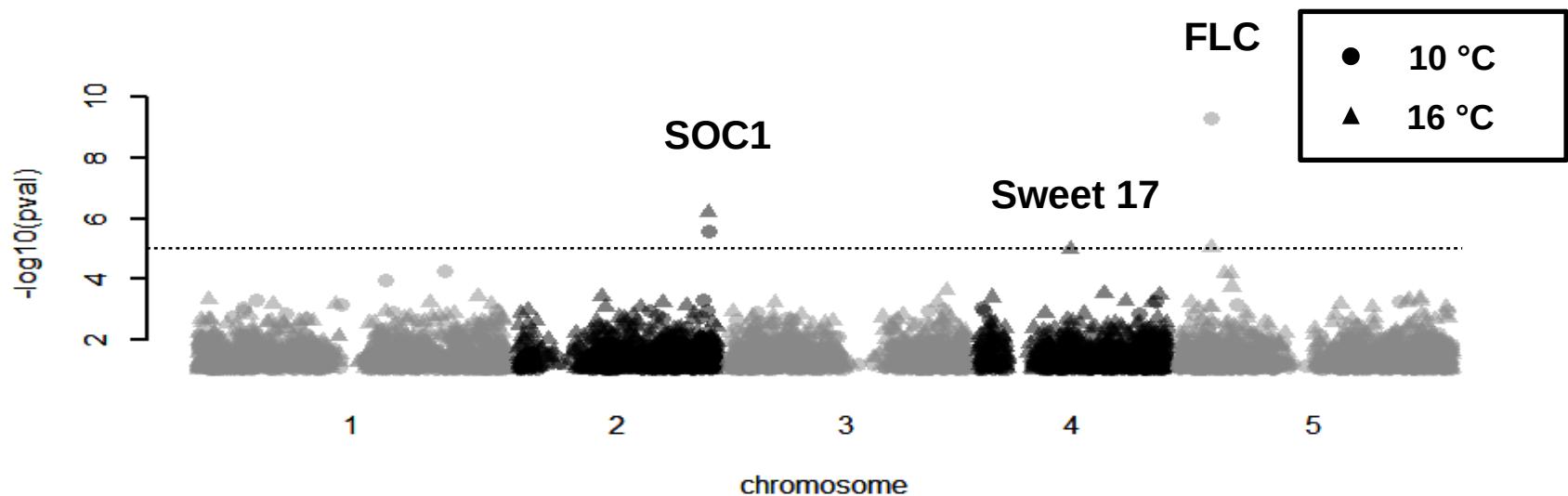
**A SNP on Chromosome 4 affects the expression of the gene AT4G19480
How does this affect the phenotype ?**

eGWAS



TWAS on Flowering Time

Using RNA expression to explain the phenotype



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

AraGWAS Catalog

AraGWAS Catalog

[TAKE A TOUR?](#) [FAQ](#) [ABOUT](#) [LINKS](#) [REST API DOCUMENTATION](#)

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.



Search the catalog (e.g. FLC, AT2G27035, chr2:1153551, flowering, PMC3002...)



Take a tour

Take a quick tour of AraGWAS to learn about the important functionalities.

TAKE A TOUR



Top Associations

Check out the top hits across the *Arabidopsis thaliana* genome.

TOP ASSOCIATIONS



Top Genes

Check out the genes with most hits across the *Arabidopsis thaliana* genome.

TOP GENES



Public GWA Studies

Browse through all available public *Arabidopsis thaliana* GWA studies.

GWA STUDIES



GWAS HitMap

Check out the hitmap of hits across the *Arabidopsis thaliana* genome for all public GWA studies.

GWAS HITMAP

standardized GWAS pipeline for all phenotypes in AraPheno

Using TensorFlow for GWAS



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

New Results

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?].

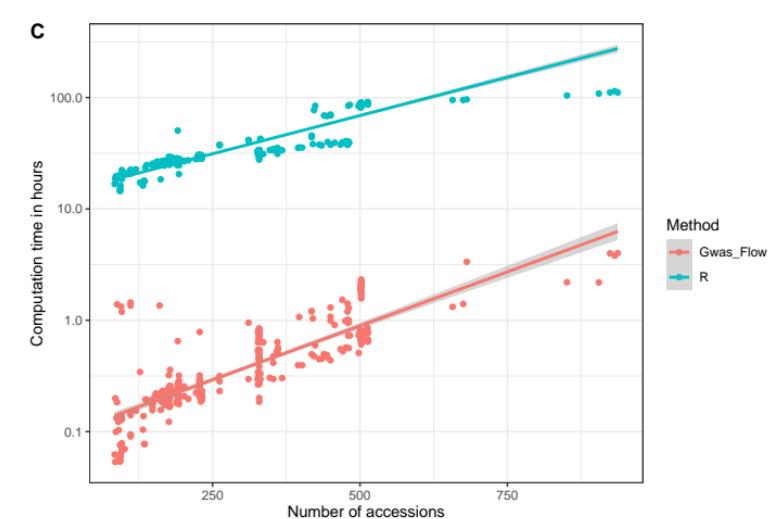
Abstract

Full Text

Info/History

Metrics

Preview



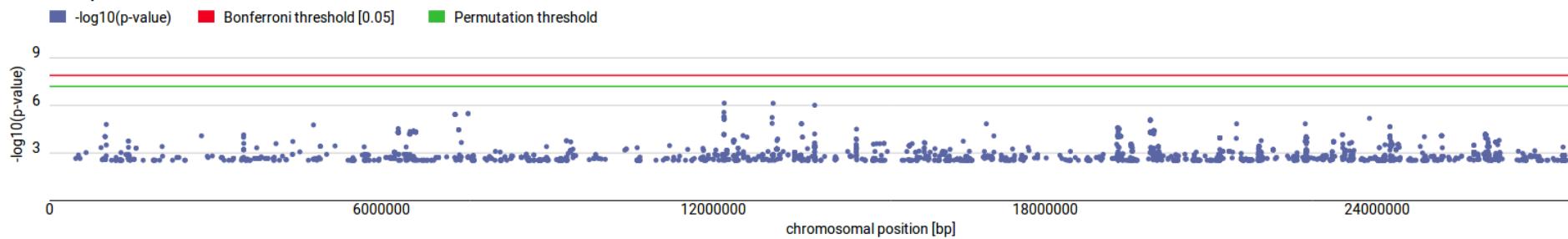
GWAS-Flow enables the calculation of permutation-based significance thresholds



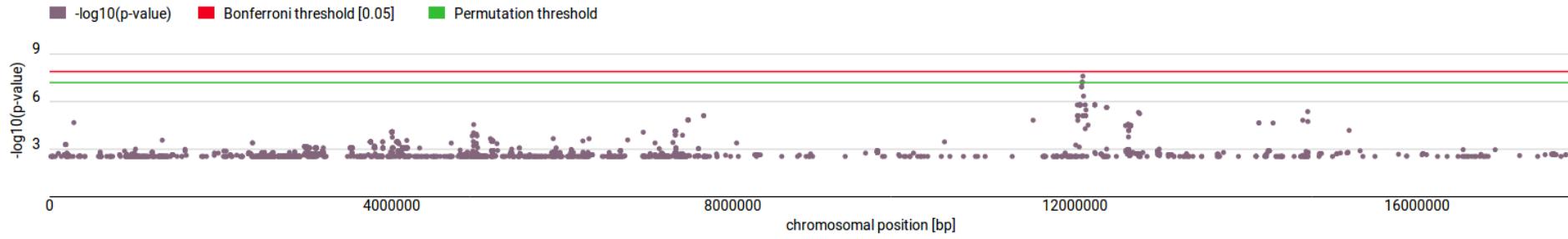
Calcium concentration in leaves

Manhattan Plots

Manhattan-plot for chromosome 1

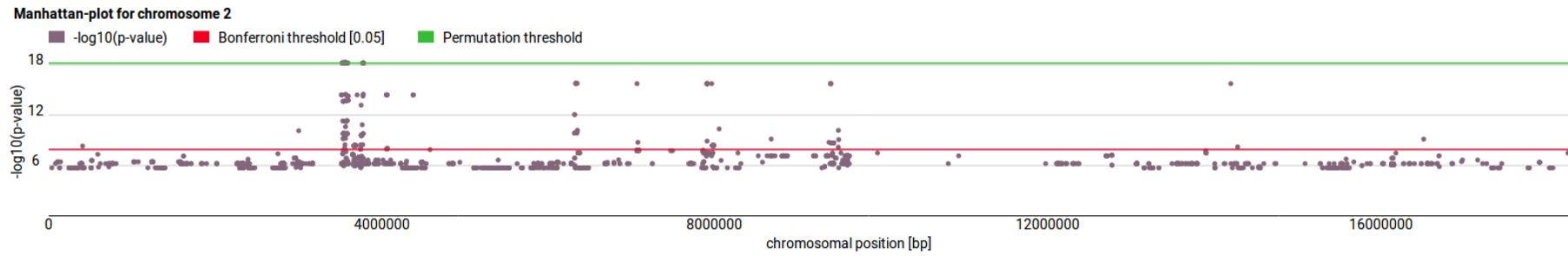
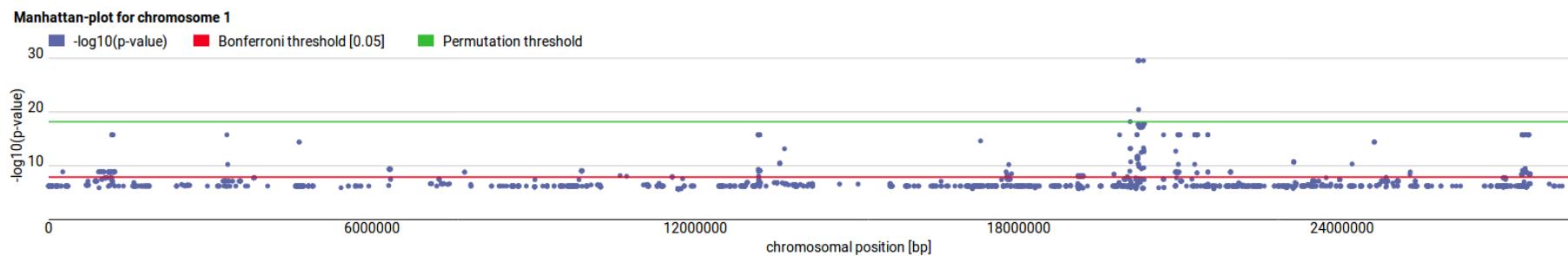


Manhattan-plot for chromosome 2

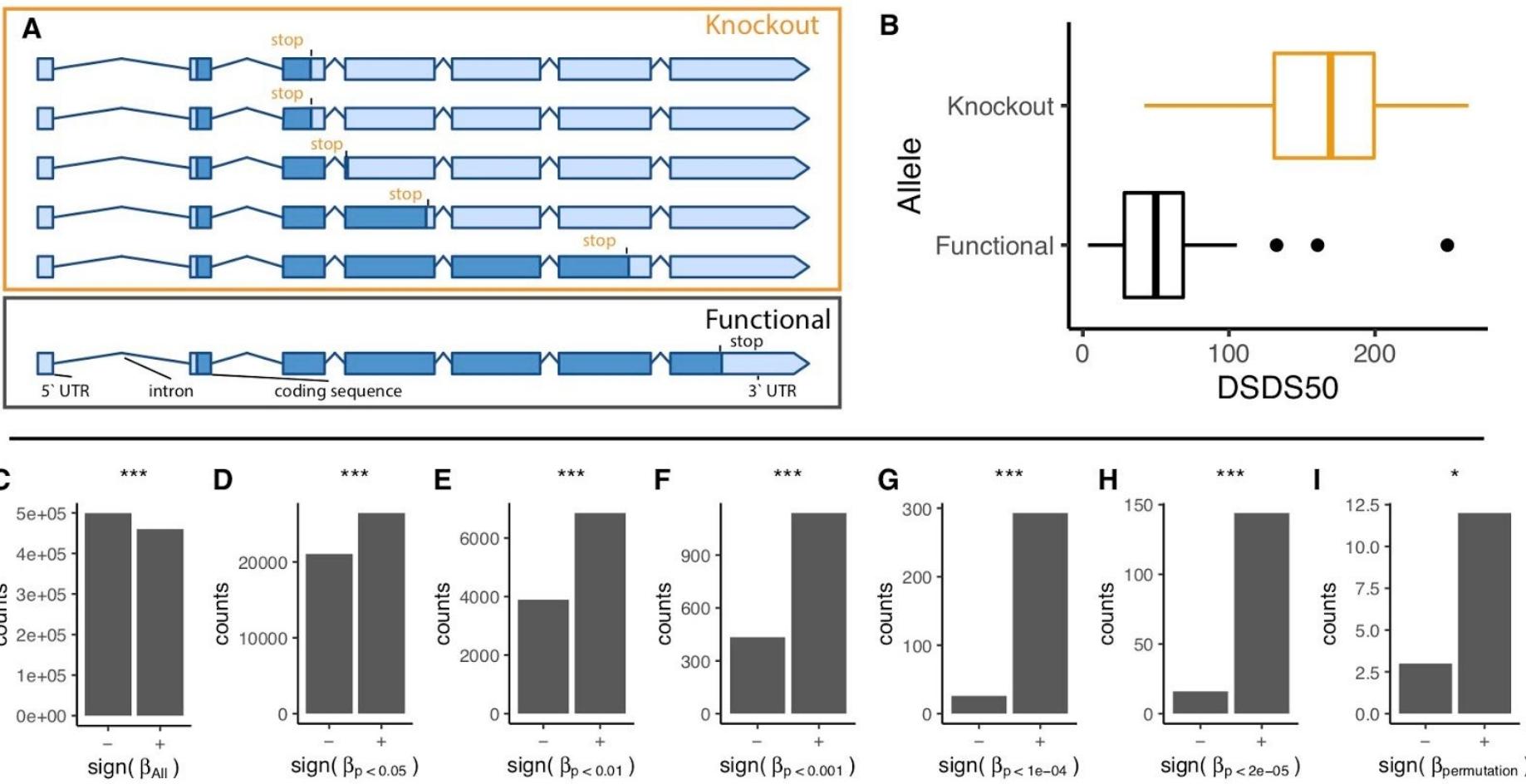


Chlorosis

Manhattan Plots

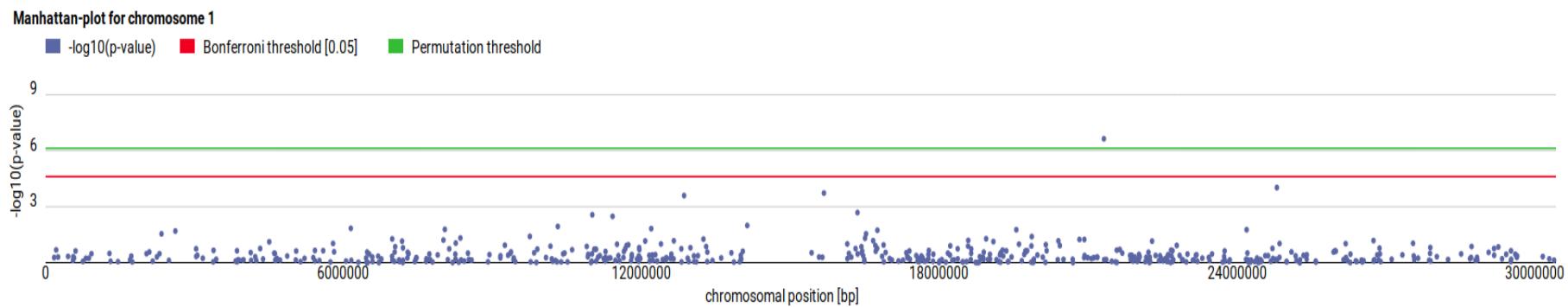


Association with ko-mutations



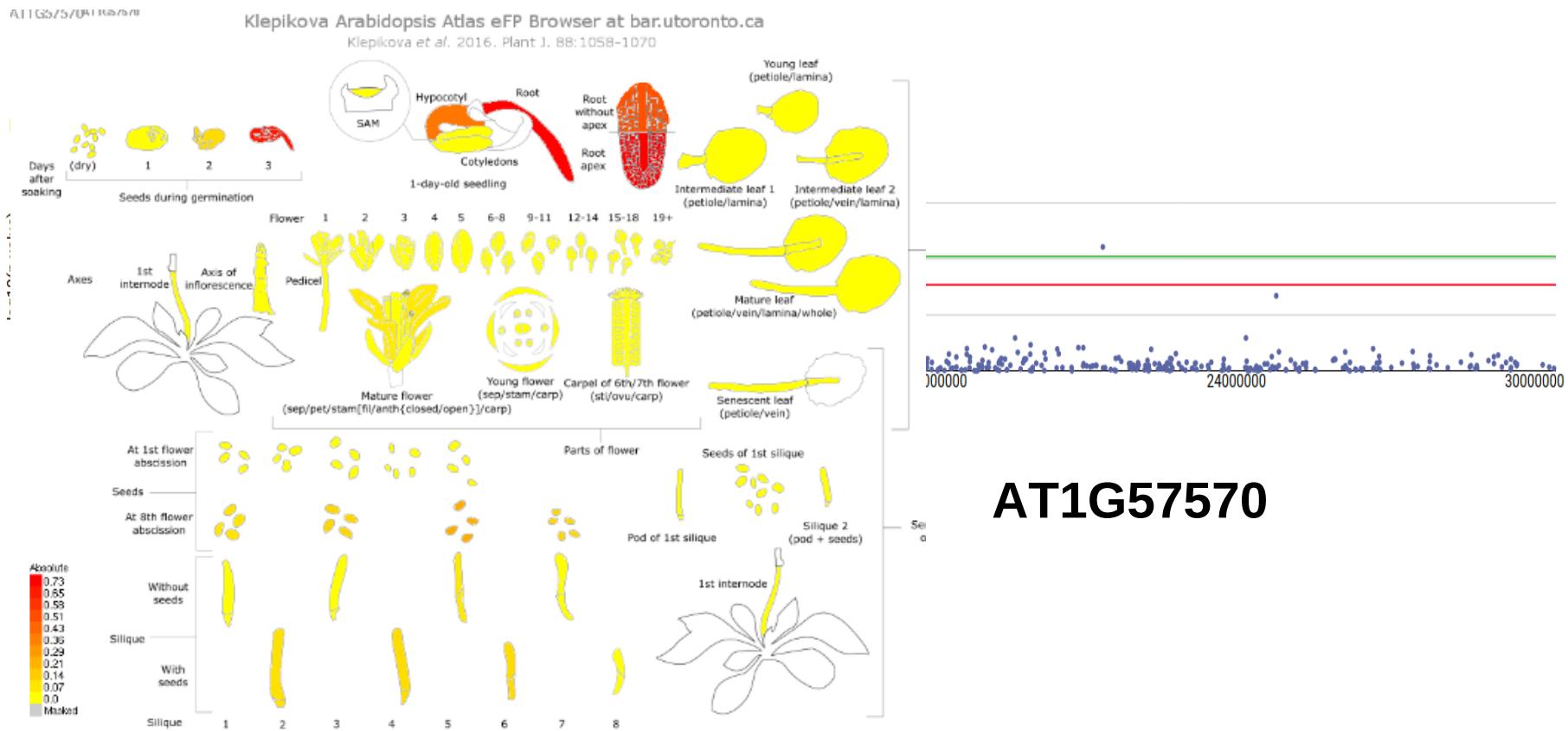
Association with ko-mutations

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)



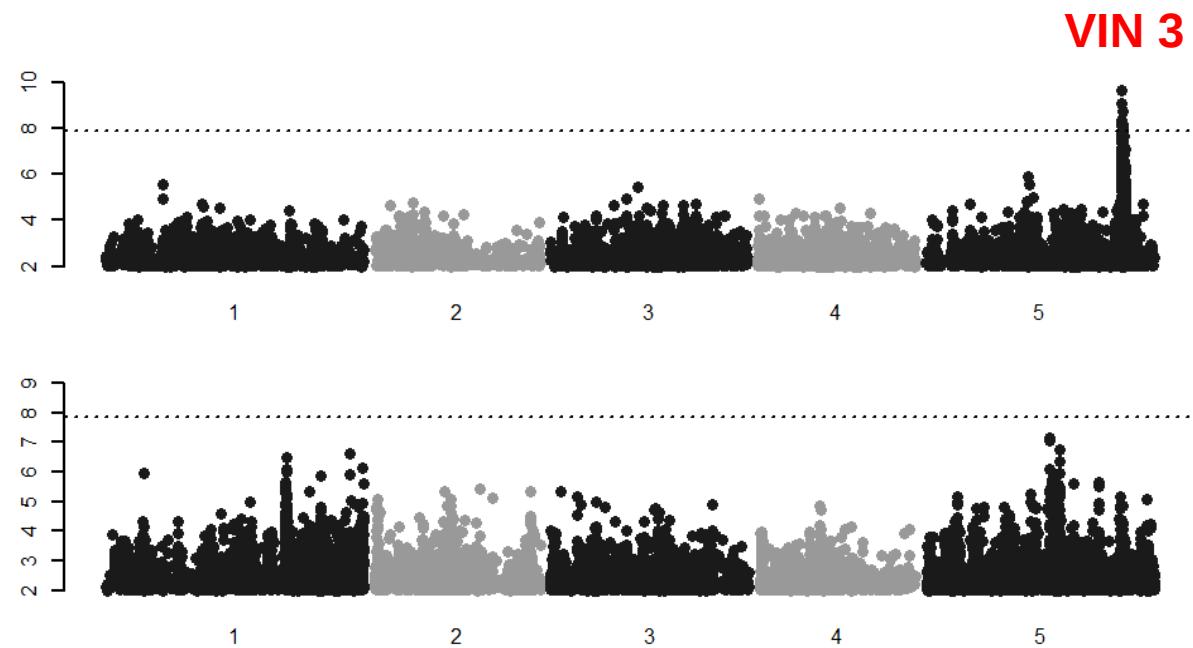
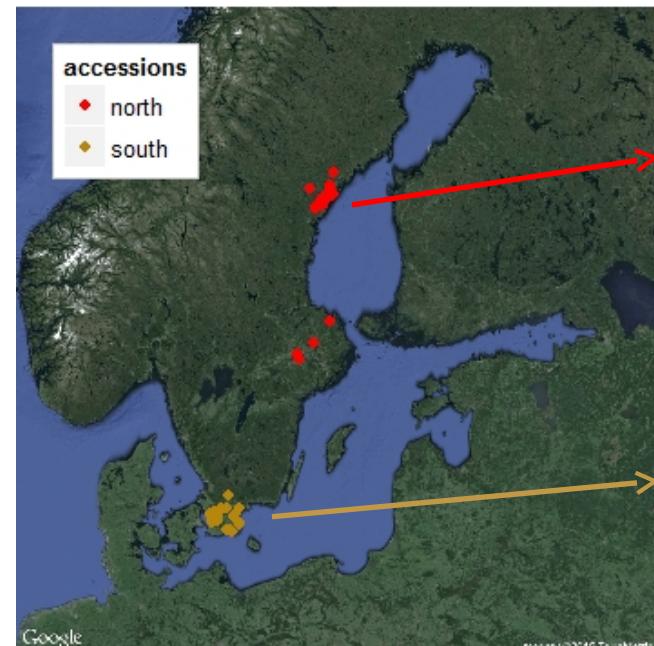
Horizontal GWAS

Genes with pleiotropic effects



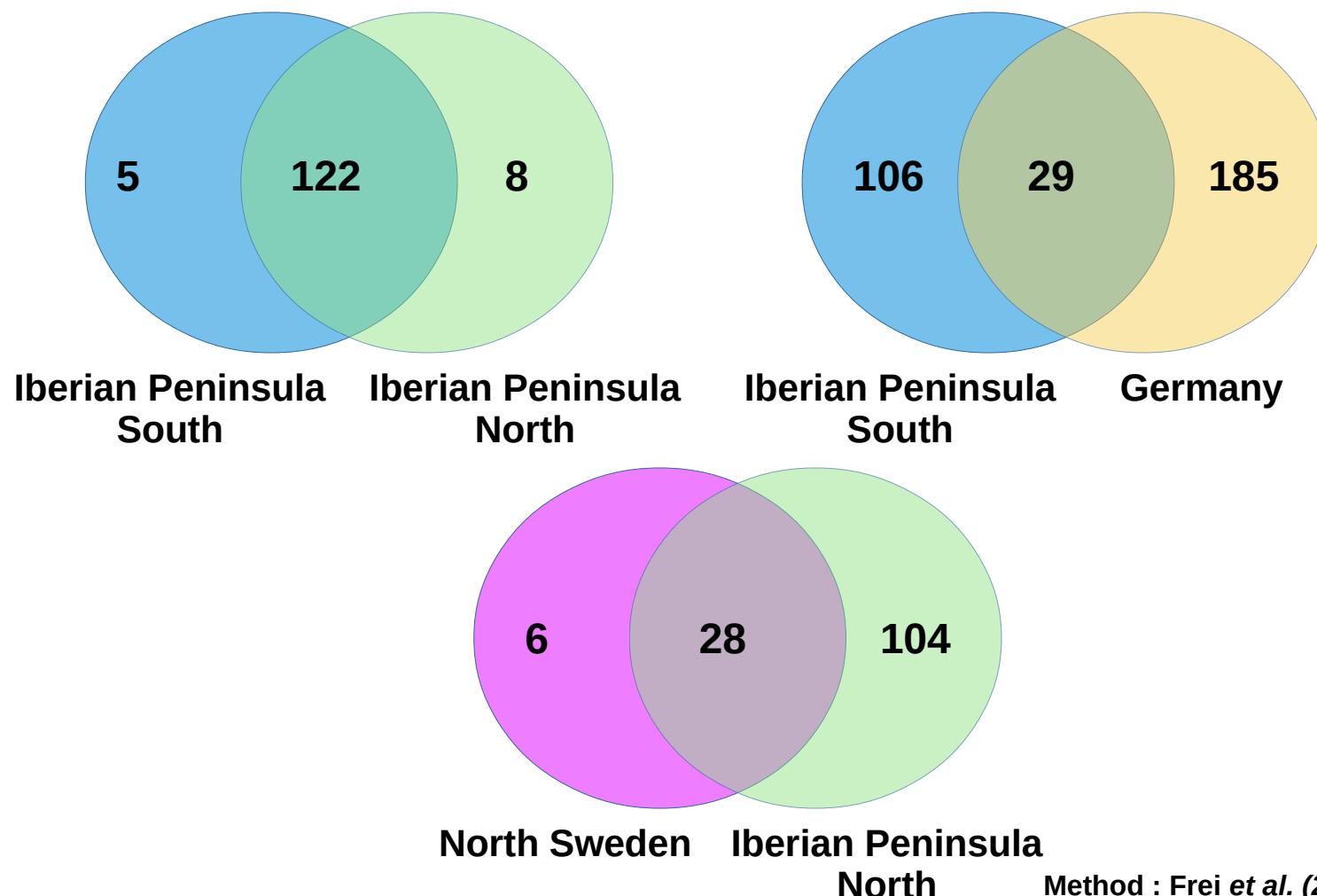
Polygenic overlap of traits / Natural Variation in Gene Regulatory Networks

GWAS on Flowering Time local subsets



Different results in different subsets

Overlap of causative genes in different subsets



Flowering time pathway in *A.thaliana*

