GWAS Analysis for Behavioral Traits in Domestic Sheep

Training Material

Valentina Tsartsianidou / Postdoctoral Researcher

AUTH

07.05.2025

Introduction to GWAS
(Genome-Wide Association Studies)

What is GWAS?

A statistical approach based on genome variation testing associations between multiple genetic variants (across many genomes) and phenotypes of interest

Detection of differential allele frequencies of variants to reveal complex trait variation (Uffelmann, 2021)
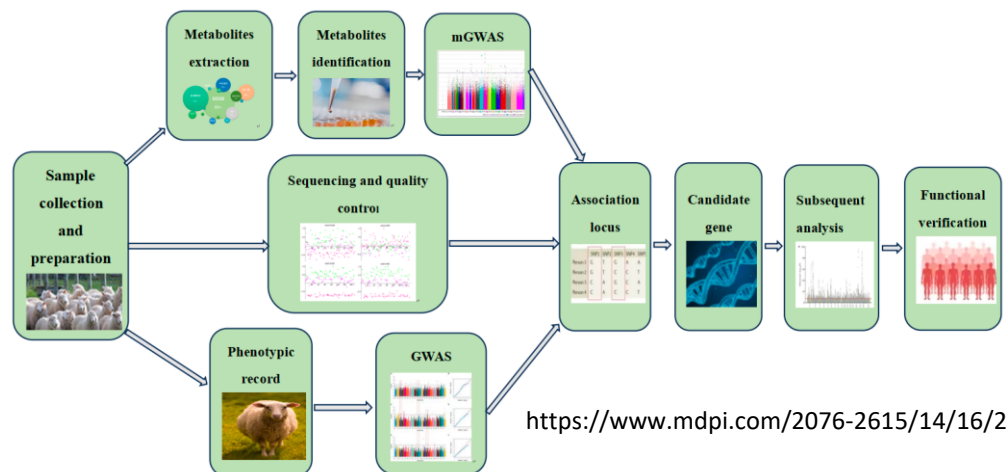
These variants are called causal (non-randomly linked) and partially control phenotypic variation

https://www.mdpi.com/2076-2615/14/16/2382

Applications in Livestock Genomics
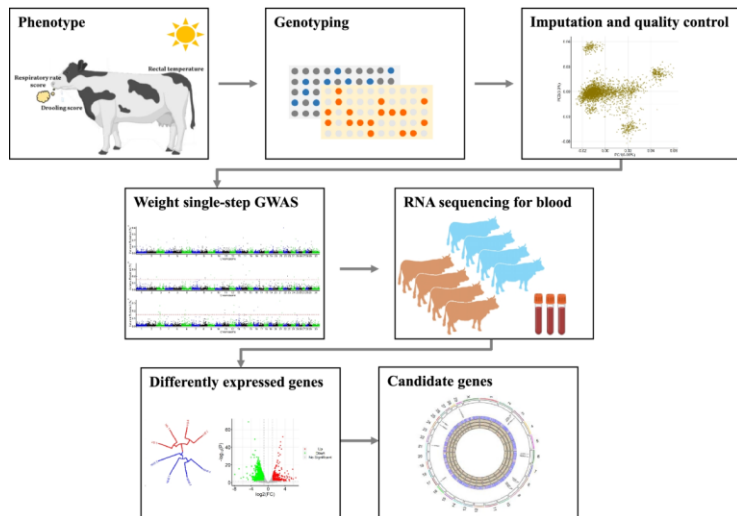
Traits of economic interest (production, reproduction, health, longevity)

↓

Quantitative characters - continuous variation within/between populations

↓

Influenced by multiple genes with relative effects - polygenic inheritance

Increasing interest on adaptive and behavioral traits

↓

Understanding of local adaptation, animal resilience/behavior to environmental changes

Implementation of genomic selection for genetic improvement

**GWAS enable the detection of specific SNP markers**

WHAT IS PHENOTYPE

Phenotype (P) = Genotype (G) + Environment (E)

- **Phenotype** : Trait, observable characteristic of an animal

- **Genotype** : Genetic profile of an individual. We can act on it. Inherited

- **Environment** : Events around the individual. We can partially act on it. Not inherited

WHAT IS GENOTYPE

- **Genotype:** Determined by alleles/DNA segments across genome of an individual coding for functions related to the trait (also used to describe the alleles on a particular locus)

- **Haplotype:** A group of tightly linked alleles/DNA segments from a single parent inherited together

- **Haplotype Blocks:** haplotypes of different length are sometimes co-inherited (Linkage Disequilibrium)

Animal 1
Animal 2
X
Animal 3
Animal 4
X
Animal 5
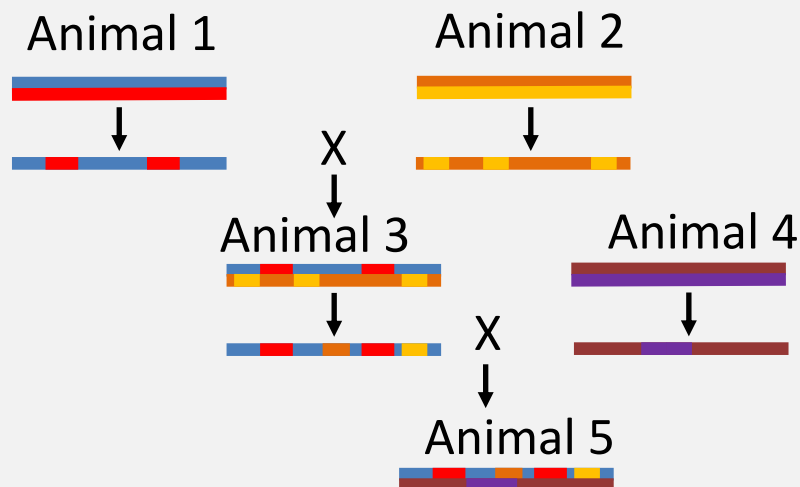
- **Recombination reduces LD (shorter haplotypes)**

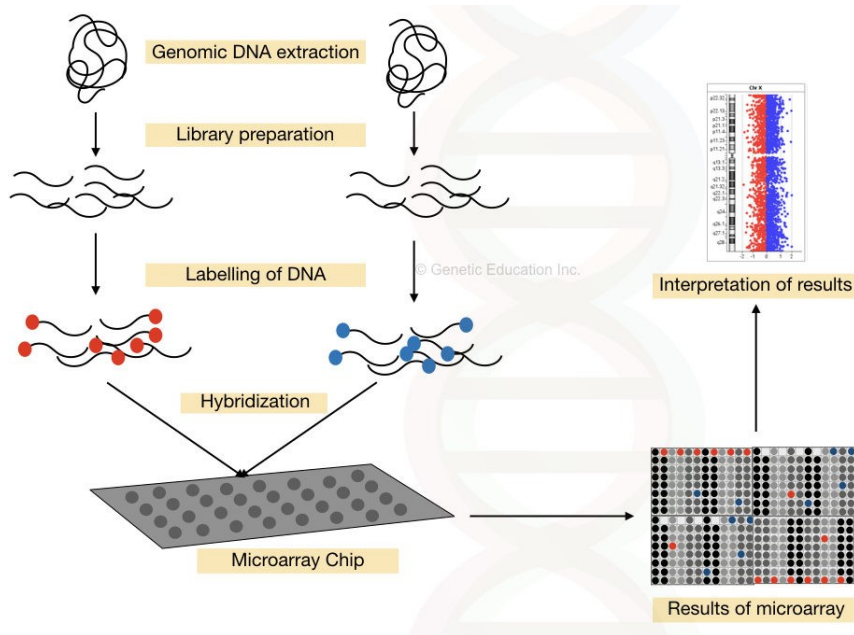- **Selection increase LD (longer haplotypes)**

WHAT IS GENETIC MARKER

- **Markers: Variable sites/sequences with known position on the genome**
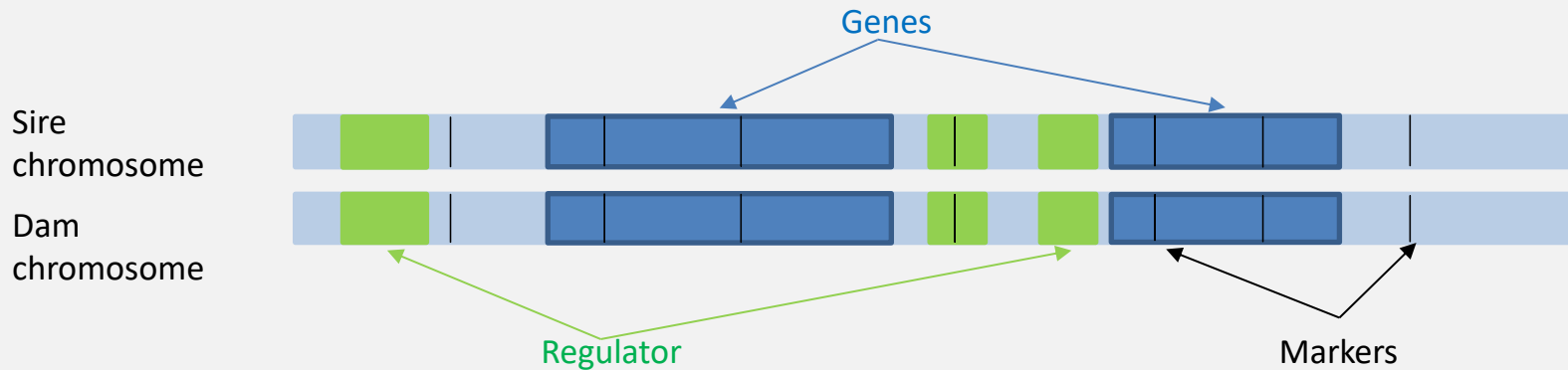
- **They can be linked with genes related to a trait of interest**
- **If they are linked with relevant genes, we can use them for selection purposes**
- **We can also use them to estimate the relationships between individuals**

High-throughput genotyping and GWAS

Genomic DNA extraction

Library preparation

Labelling of DNA

Hybridization

Microarray Chip

Interpretation of results

Results of microarray

© Genetic Education Inc.

https://geneticeducation.co.in/snp-array-high-throughput-snp-genotyping-technique/

SNP arrays: Genotyping of Single Nucleotide Polymorphisms (SNPs) across the genome

Development of species-specific arrays of thousands or million markers

In livestock, multiple global breeds are used for array designation and validation

GWAS studies testing association between phenotypes and the total
number of available SNPs

Classic Animal Model

Statistical methods first developed by Fisher (1918) and Wright (1921) to partition the observed phenotypic variation into variance components (genetic, environmental)

Phenotype (P) = Genotype (G) + Environment (E)

$$Y_i = X + u_i + \varepsilon$$

- **Phenotype (Y)** : Trait, observable characteristic of an individual

- **Fixed effect (X)** : Part of the environment clearly identified, eg: Farm, sex, year, …

- **Random effect (u)** : Animal genetic profile – estimated with pedigree information

- **Error (ε):** Residual due to imperfection of the model

GWAS Model

Test the association of each SNP marker with the phenotype

GWAS model ➜ **Y = X + *Marker effect* + u + ε**

- **Phenotype (Y)** : Trait, observable characteristic of an animal

- **Fixed effect (X)** : Environmental factors clearly identified, eg: Farm, sex, year, …

- **Marker (SNP) effect**: assessed for each SNP

- **Random effect (u)** : random (additive genetic) polygenic effect ⟶ Genomic Relatedness Matrix (GRM)

- **Error (ε):** Residual due to imperfection of the model

Control for genetic relatedness between individuals that may lead to false positive associations

GWAS Model

Manhattan plot

SNPs with different allele frequencies among subgroups of the study population

https://link.springer.com/article/10.1186/s12711-021-00682-7
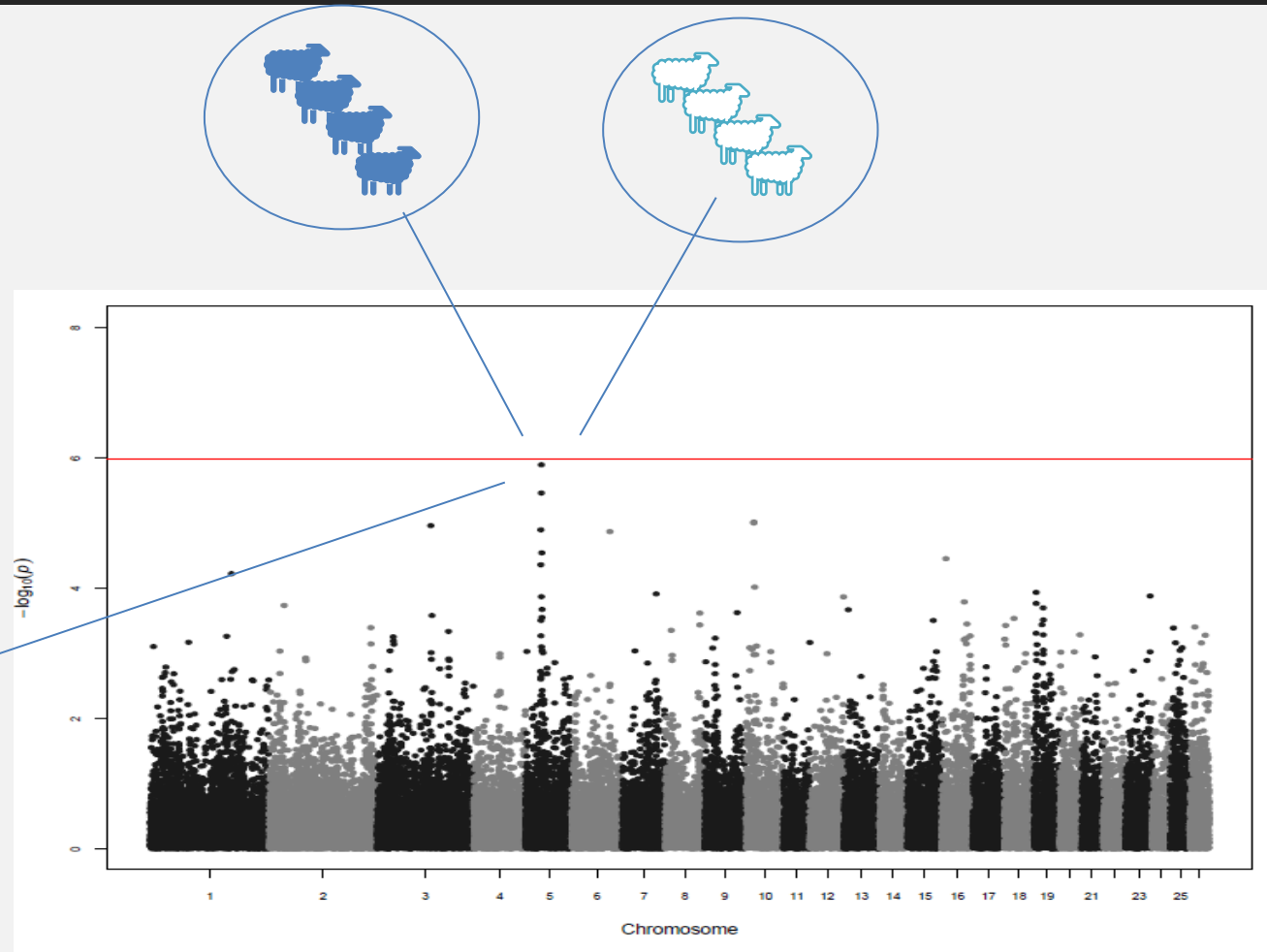
Case Study

Grazing behavior
of mountainous
Greek Boutsko
sheep

Study Design – Overview

- Phenotype collection
- Blood sampling
- DNA extraction and SNP Genotyping
- GWAS Workflow

Sampling locations

Sampling Locations in Greece

Smarter
SMAll RuminanTs breeding for Efficiency and Resilience
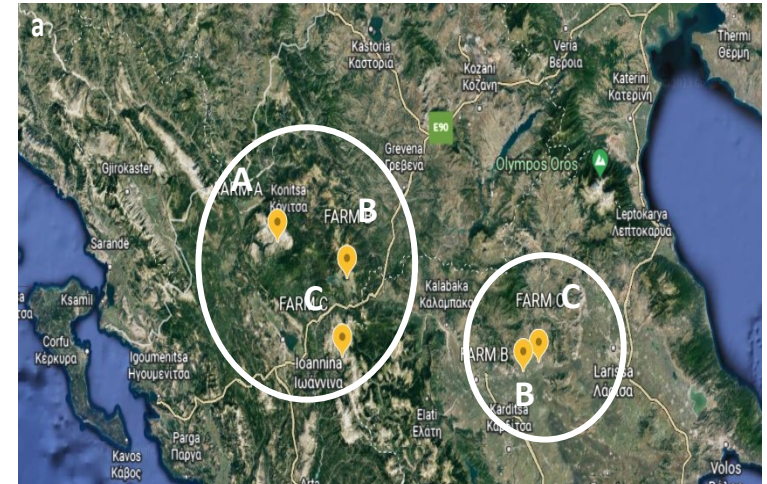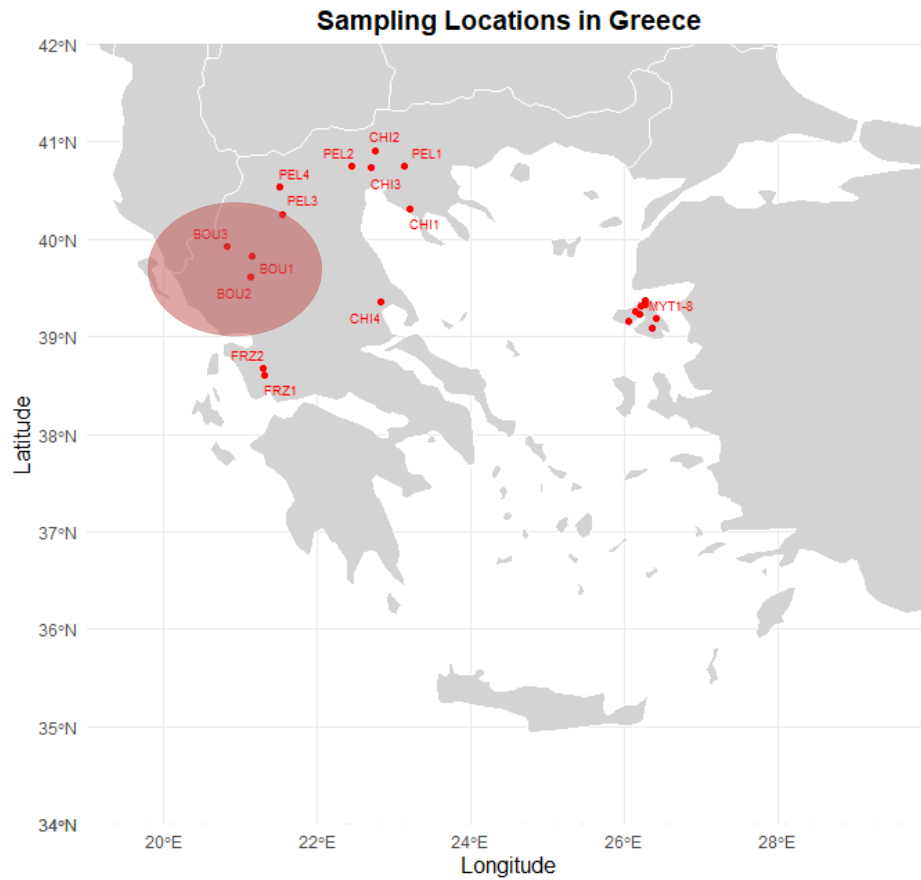
SNP Genotyping across genome

- Blood sample collection

- DNA extraction with Qiagen MiniBlood Kit
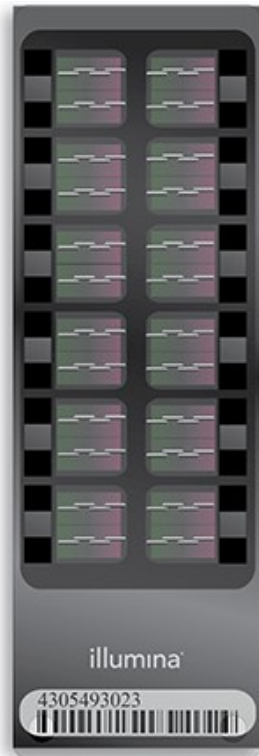
🐑 Illumina Ovine SNP50 BeadChip – 53,516 SNPs
(https://www.illumina.com/documents/products/datasheets/datas
heet_ovinesnp50.pdf)

Ped file
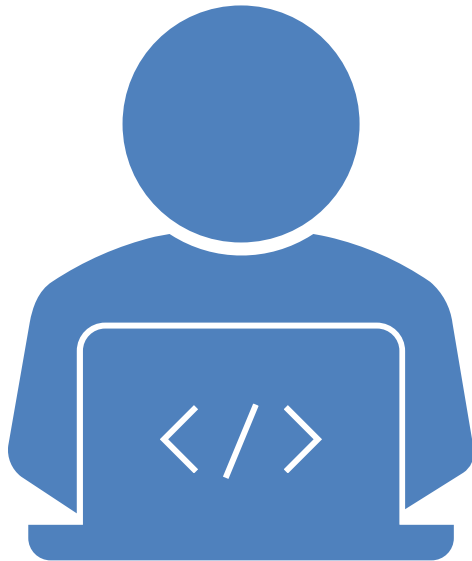
map file

binary format

.fam/.bed/.bim

GWAS Workflow

1. Data filtering

2. Assessment of population stratification

3. Association testing

4. Significance testing at genome-wide and chromosome level

Software and Tools

- PLINK v1.9

- GEMMA

- RStudio

Quality Control

Standard filtering steps using plink

Exclude non-autosomal and unmapped SNPs

SNP call rate

Minor allele frequency

Individual call rate

Genome-wide Association Study (GWAS)

GWAS with GEMMA

Farm
Distance
Speed
elevation gain

Linear mixed model

Continuous trait

Construction of GRM

File with fixed effects and covariates

Filtered genotypes with phenotype information
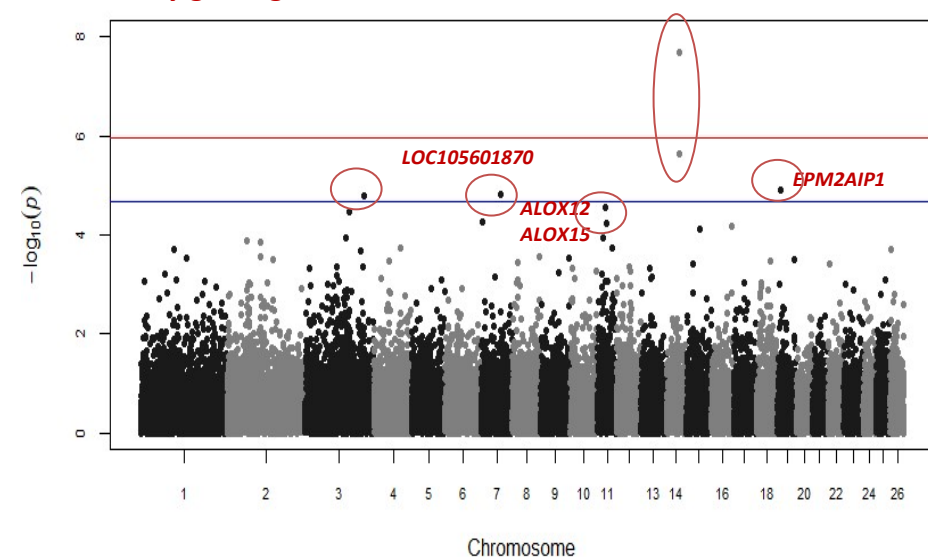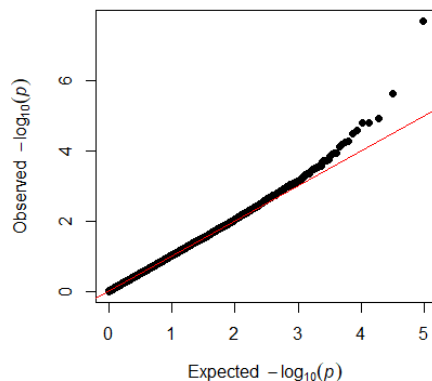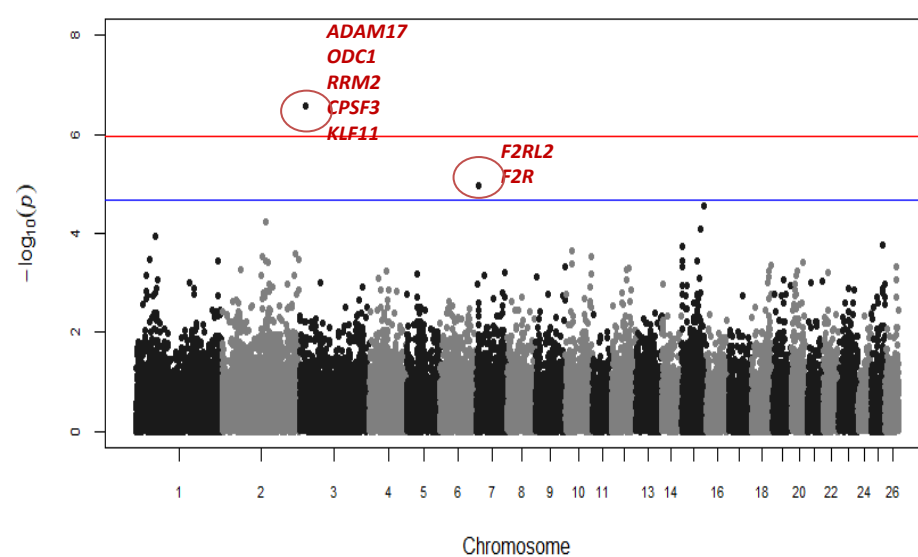
GWAS results – genome-wide Visualization in RStudio

Manhattan plots
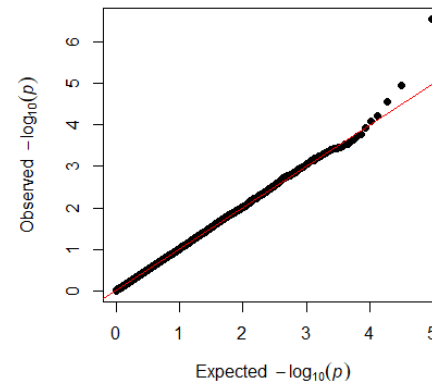Association between SNPs and phenotype under study
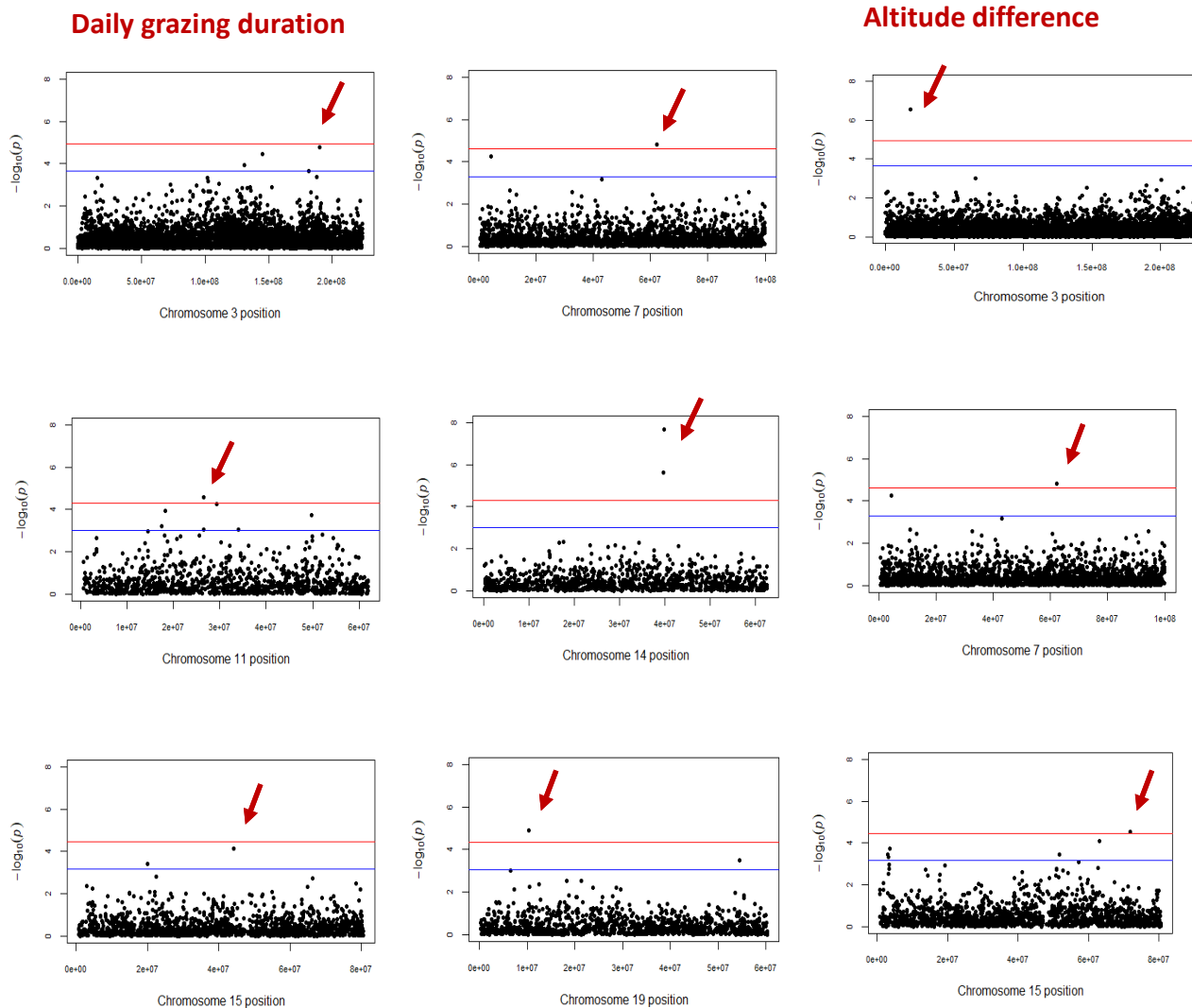
Daily grazing duration

Altitude difference

QQ plots
Expected vs Observed p-values

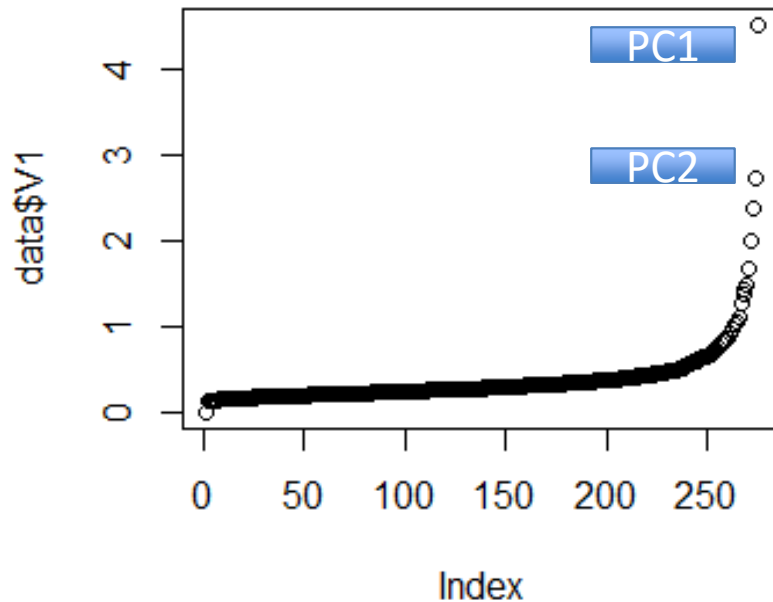Early deviations from diagonal line suggests hidden population structure

GWAS results – chromosome-wide Visualization in RStudio

Daily grazing duration
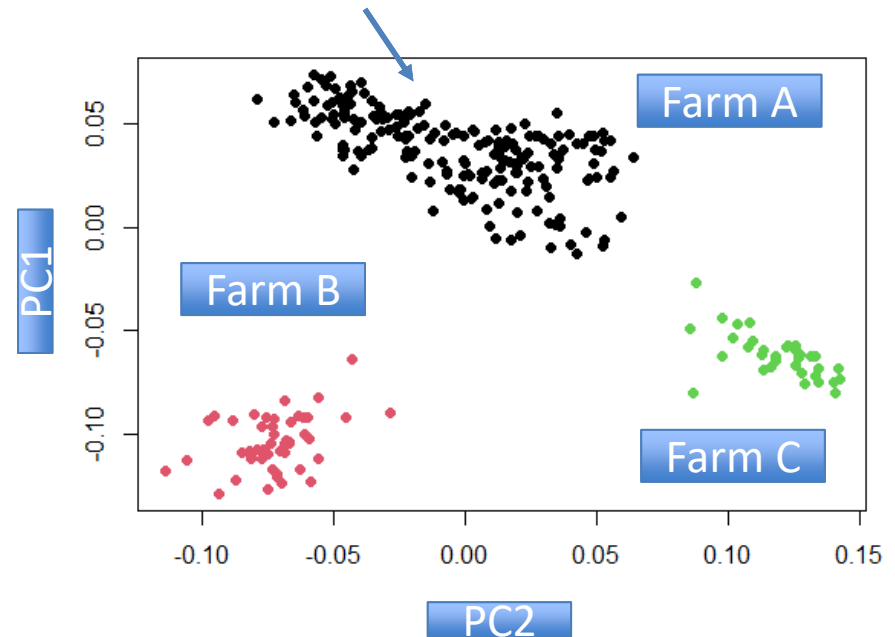
Altitude difference

PCA results – Visualization in RStudio

principal components are usually included in GWAS as covariates to correct for population structure
our case: farm=population structure

eigen values = variance explained by each PC

eigen vectors = principal components = directions of data variation

https://github.com/valentinatsar/GWAS/tree/main

Thank you for your attention