Short course on

Methods and Tool for Genomic Predictions and GWAS in Breeding Programs

Mehdi Sargolzaei
Select Sires Inc.
University of Guelph

Daniela Lourenco
University of Georgia

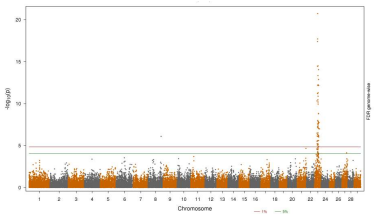
20 -24 February 2023, University of New England

1

GWAS Applications in Livestock

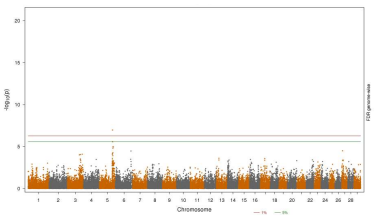
- Understanding the genetic architecture of traits
- Uncovering causative mutations affecting economically important traits
- Improving accuracy of Genomic Selection

h2 0.25



Major gene + polygene

h2 0.25



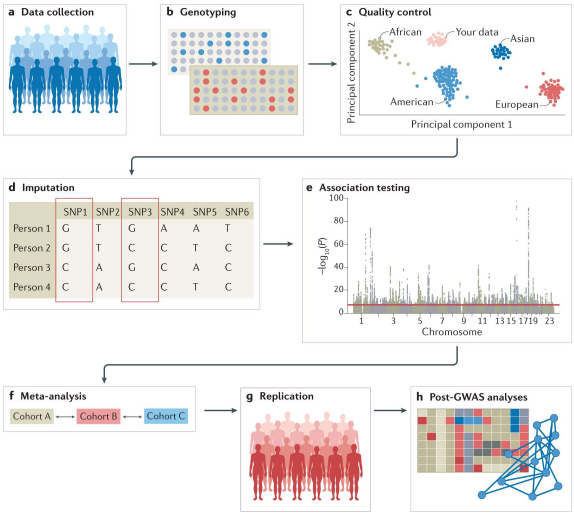
More polygenic

Results from the first 700 Ontario commercial Holsteins tested as part of the 5000 cow project – Dr. Mallard

2

2

Steps for Conducting GWAS



Uffelmann, E., Huang, Q.Q., Munung, N.S. et al. Genome-wide association studies. Nat Rev Methods Primers 1, 59 (2021).

Methods

- SNP by SNP GWAS
- SNP by SNP GWAS when fitting G matrix
- GBLUP/RBLUP/ssGBLUP
- Bayesian method

+Haplotype models

Methods

- Family-based GWAS
- Random samples (Mostly common variants)
- Selective genotyping (Mostly rare variants)
- Isolated population
  - Rare variants may be observed in higher frequencies

5

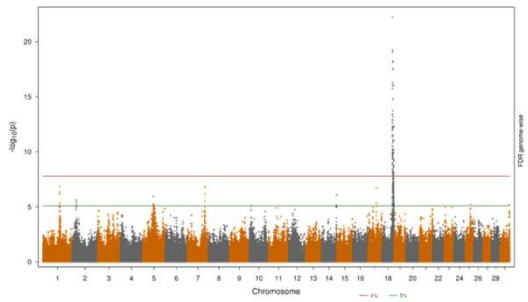
5

Points to Consider When Performing GWAS

6

6

Genome-Wide Association Studies (GWAS)



- GWAS identifies SNPs which are associated with a trait. Most of the time it cannot specify causal genes/mutations.
- Missing h^2: Genetic variation calculated from GWAS does not completely explain the heritability of quantitative traits.

7

7

Some Challenges

- Sample size (for mapping rare variants)
- Extent of LD and fine mapping
- Population stratification & structure
- Multiple testing (association by chance)
- Under-representation of rare variants due to ascertainment bias
- GWAS on EBV or de-regressed EBV

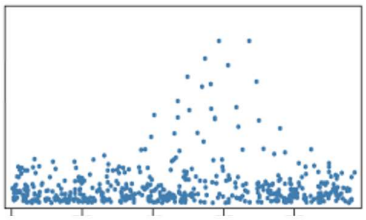
8

8

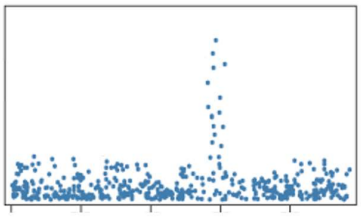
Extent of LD and Fine Mapping

When LD is extended over long distances, fine mapping is challenging

High extent of LD



Low extent of LD

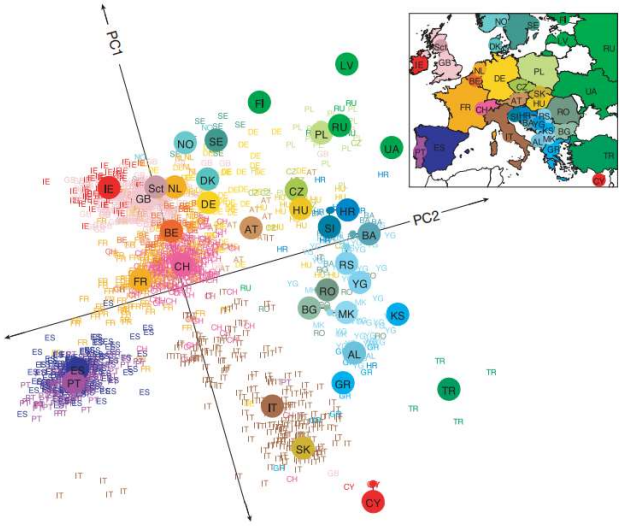


GWAS from different breeds may provide additional info to locate the causative mutation

9

9

Population Stratification & Structure



Ref: Genes mirror geography within Europe. John November et al. Nature 2008.

10

10

Population Stratification & Structure

- Unusual allele frequency differences between subpopulations
- Phenotypes correlated with locations cause spurious associations
- Family structure or cryptic relatedness also results in spurious associations
- Systematic ancestry differences between subgroups and also admixture

11

11

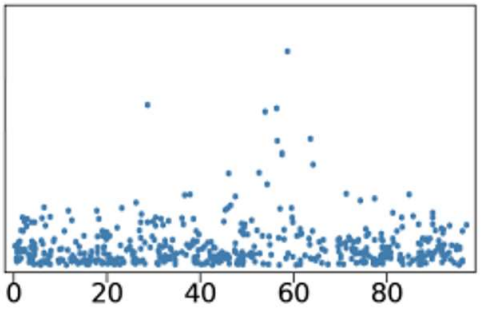
Population Stratification & Structure

Japan

Good in Baseball

Not as good in Baseball

Great Britain



12

12

Population Stratification & Structure

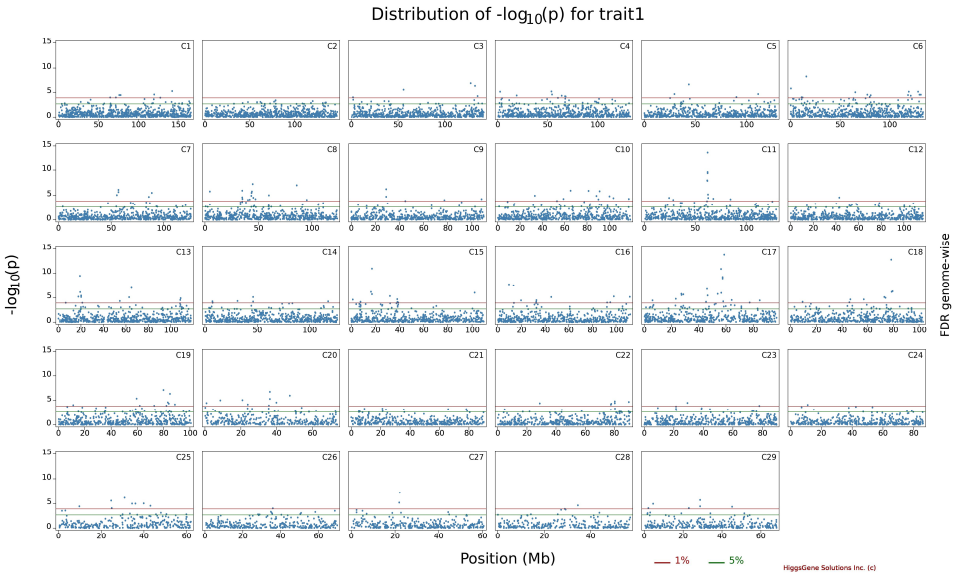
Solution:

- Structured association (clustering)
- Fitting principal components as covariates in the model
- Fitting genomic relationship matrix in the model (Genomic control)
- GBLUP/RBLUP/ssGBLUP

13

13

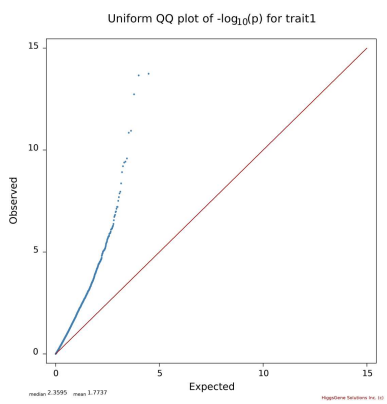
Population Stratification & Structure

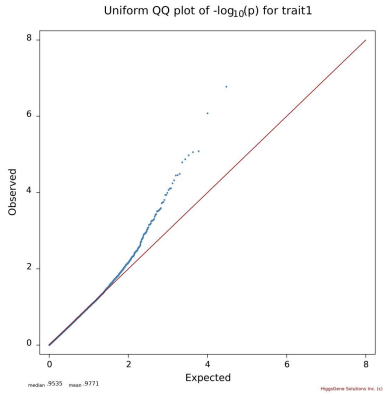


14

14

Population Stratification & Structure





15

15

Multiple Testing Issue

H0 = SNP is not associated with the phenotype (α = 5%)

Probability of type I error for single test is α

Type I error: Incorrect rejection of H0 (False positive)

If there are 1000 tests and tests are independent, then we expect 50 false positive associations!

16

16

Multiple Testing Issue

Controlling type I error rate:

- Permutation
- Bonferroni correction
- False discovery rate
- Positive false discovery rate

17

17

Multiple Testing Issue

Bonferroni correction:

$$\alpha/n$$

Simple but very conservative

$\alpha = 0.05$
$n = 1000$

Significant level = 0.00005

18

18

Multiple Testing Issue

False Discovery Rate (FDR):

For controlling FDR at 5% level

- Sort p-values from the smallest to the largest
- Find the first p-value that is larger than (j / n) * 0.05, where j is the rank of p-value
- Declare all p-values with rank less than j as significant

19

19

Multiple Testing Issue

Positive False Discovery Rate (pFDR)

More complicated than FDR but in some cases better than FDR

With this method p-values are transformed to q-values

https://www.bioconductor.org/packages/release/bioc/html/qvalue.html
https://github.com/StoreyLab/qvalue

20

20

Ascertainment Bias

- Ascertainment bias is introduced when SNPs are not a random sample of DNA polymorphism
- SNP chips are designed so that well segregating SNP (intermediate allele frequency) are selected (under-representation of rare variants)
- One solution is to impute to the sequence level (not practical yet!)

21

21

GWAS for Sex-Limited Traits

Dairy cattle case:

- Almost all males are genotyped
- Only fraction of females are genotyped

- Calculate EBV for males using daughters information
  - Double counting issue

- Calculate de-regressed EBV for males
  - There could be bias in de-regressed EBV

22

22

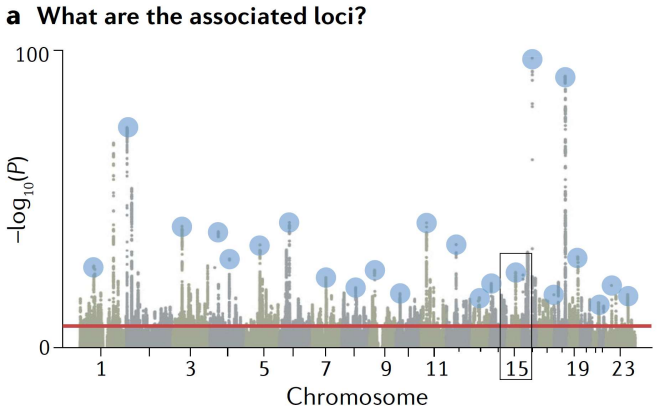
Functional Follow-up of GWAS

- In the last decade, a large number of GWAS on high quality phenotypes and big genomic data has resulted in uncovering of numerous SNP association for many traits
- Now, the big challenge is the interpreting the results in biological context

23

23

Functional Follow-up of GWAS



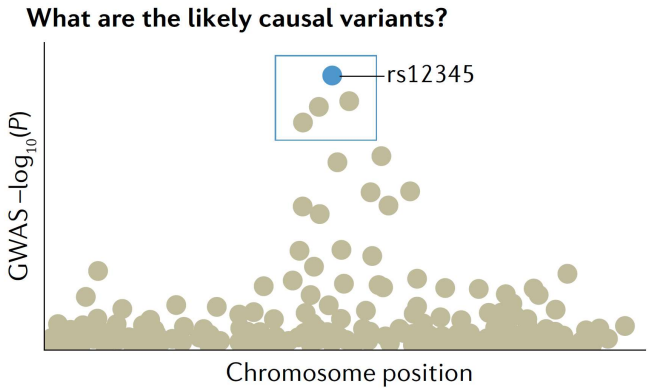
Uffelmann E., et al., Nature Review (2021) 1:59

24

24

Functional Follow-up of GWAS

Use statistical fine-mapping to identify the most credible SNP set

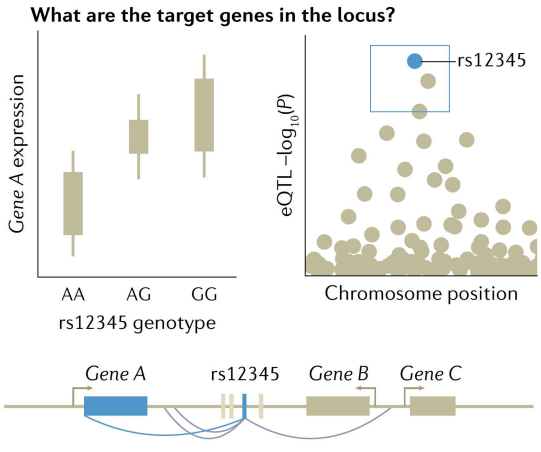


Uffelmann E., et al., Nature Review (2021) 1:59

25

Functional Follow-up of GWAS

Identify most likely target genes by mapping expression Quantitative Trait Loci (eQTLs)



Uffelmann E., et al., Nature Review (2021) 1:59

26

Functional Follow-up of GWAS

Identify pathways that may mediate the trait using enrichment analysis

What are the affected pathways?

Uffelmann E., et al., Nature Review (2021) 1:59

27

27

Confirmation

Replication in independent samples

28

28