

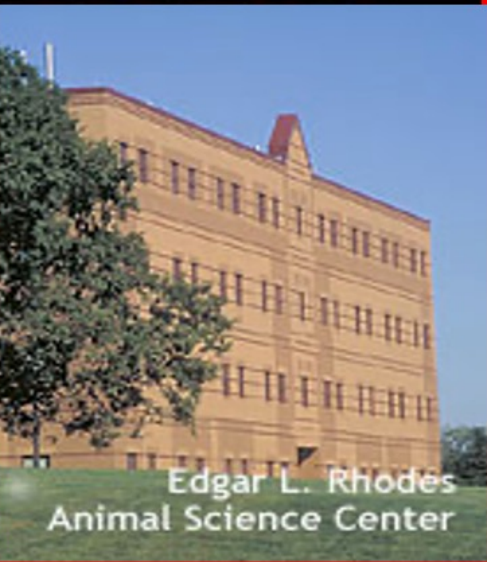
GWAS in large animal studies – why so few QTLs identified?

Ignacy Misztal
University of Georgia





THE UNIVERSITY OF GEORGIA



CIS Mini Conference, Center for Statistics and Human Genetics,
University of Edinburgh, March 26, 2024

Georgia Museum of Art

Edgar L. Rhodes
Animal Science Center

Research in Breeding & Genetics lab at UGA

- Focus on methods useful for genetic evaluation of animals
- New methods put into computing package– BLUPF90
- Methods used worldwide

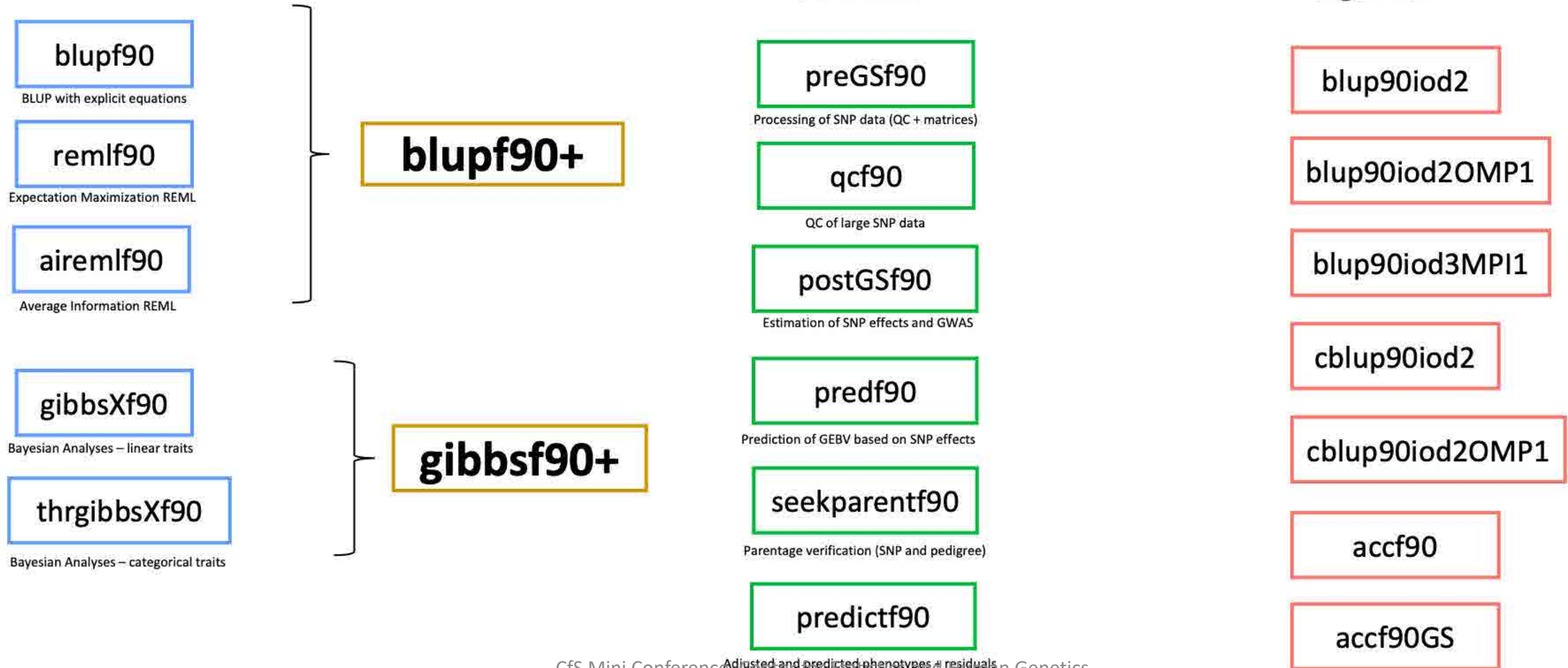
- Sponsors across species
 - Nearly all US animal breeding companies (dairy, beef, pigs, broilers, layers, fish)
 - Bayer (crops)

- Extra applications in bees, humans (schizophrenia) and trees

- 20+ papers/year

- Access to most comprehensive data sets anywhere

BLUPF90 software suite



Single-step GBLUP –BLUP with joint pedigree and genomic relationships

Idea (Misztal et al., 2009)

$$\begin{array}{ccccc} \text{Joint} & & \text{Pedigree} & & \text{Genomic} \\ \text{relationships} & = & \text{relationships} & + & \text{addition} \\ \mathbf{H} & = & \mathbf{A} & + & \mathbf{\Delta} \end{array}$$

Real (Legarra et al., 2009)



$$\mathbf{H} = \mathbf{A} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} [\mathbf{G} - \mathbf{A}_{22}] \begin{bmatrix} \mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

1 - ungenotyped
2 - genotyped

Inverse (Aguilar et al., 2010)



$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

Christensen and Lund, 2010
Boemcke et al., 2011

Specificity of plant and animal breeding

- Plants

- Find genes in wild species
- Introgress into inbred lines
- Genetic evaluation of inbred crosses across environments
 - All crosses genotyped

- Animals

- Selection usually within breeds and lines
- Commercial animals purebreds or crossbreds
- Many animals ungenotyped
- Single-step GBLUP dominant methodology

ssGBLUP for Genome Wide Association Studies

- Large research interest in GWAS
- Limitations for current methods
 - Simple models
 - Single trait
 - Complicated if not all animals genotyped



Can ssGBLUP be used for GWAS?

Genet. Res. Camb. (2012), **94**, pp. 73–83. © Cambridge University Press 2012.
doi:10.1017/S0016672312000274

73

Genome-wide association mapping including phenotypes
from relatives without genotypes

Cited by 537

H. WANG¹*, I. MISZTAL¹, I. AGUILAR², A. LEGARRA³ AND W. M. MUIR⁴

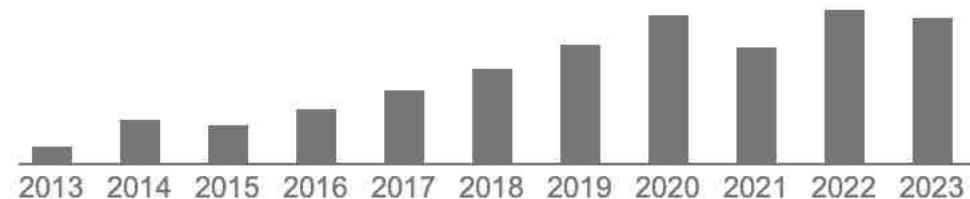
¹Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602-2771, USA

²Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90200 Canelones, Uruguay

³INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), BP 52627, 32526 Castelnau-Tolosan, France

⁴Department of Animal Science, Purdue University, West Lafayette, IN 47907-1351, USA

(Received 19 September 2011; revised 8 December 2011; and 9 March 2012; accepted 13 March 2012)



CfS Mini Conference, Center for Statistics and Human Genetics,
University of Edinburgh, March 26, 2024

GWAS with ssGBLUP (Wang et al., 2012)

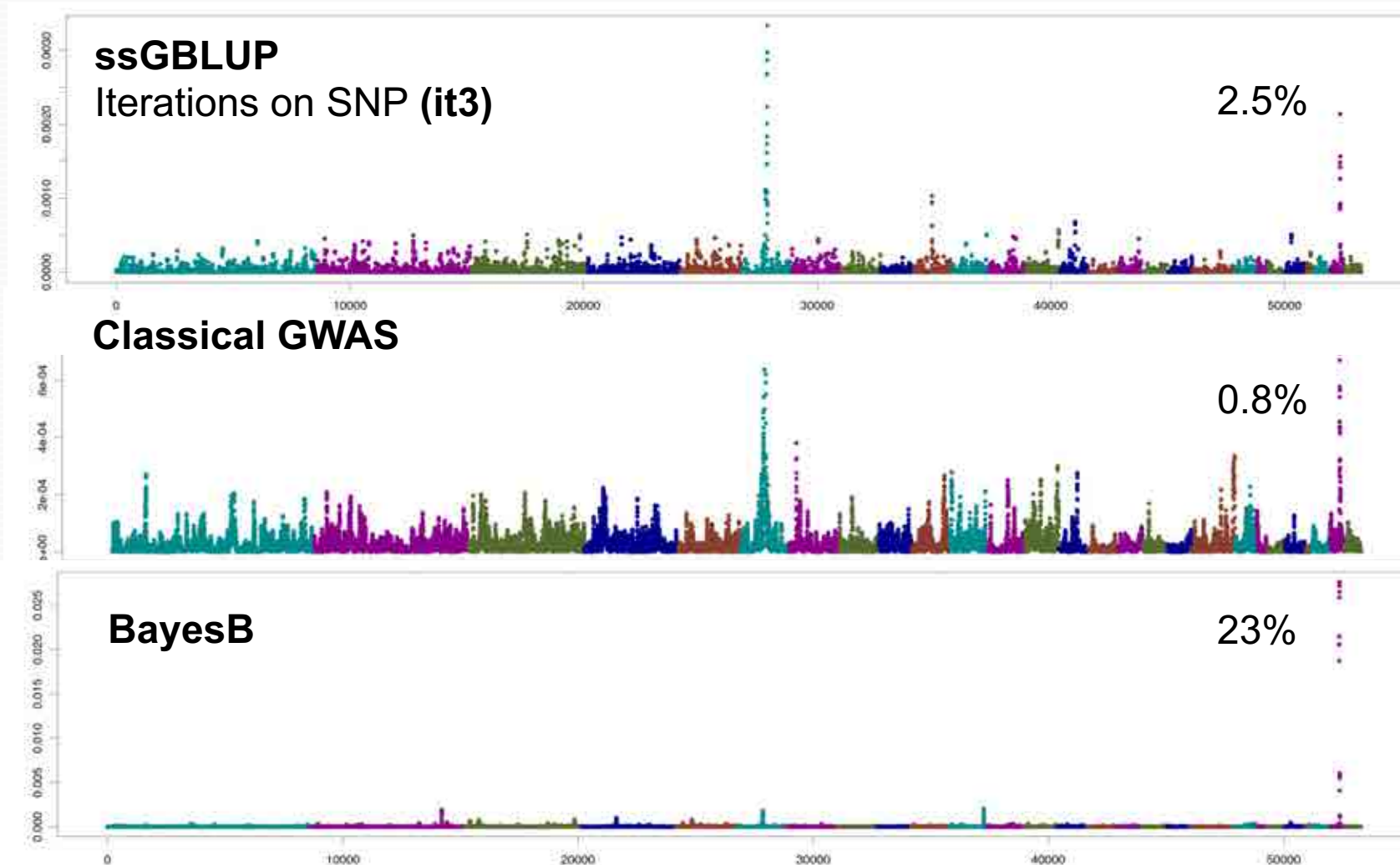
- Convert GEBV to SNP effects
- Estimate individual SNP variances
- Incorporate variances in G
- Possibly recompute GEBV and iterate

1. $D=I$
2. $G=ZDZ'/q$
3. Compute a
4. $u=DZ'/q G^{-1} a$
5. $d_i=2p_i(1-p_i)u_i^2$
6. $D=n D/\text{tr}(D)$
7. Loop to 2

Output as % of variance explained in a window

Discrepancies in GWAS methods

Chicken weight



Are p-values possibly in (ss)GBLUP?

$$pval_i = 2 \left(1 - \Phi \left(\left| \frac{\widehat{snp}_i}{sd(\widehat{snp}_i)} \right| \right) \right) \quad (\text{Chen et al., 2017})$$

In ssGLUP conversions:

GEBV to SNP effects

PEV(GEBV) to PEV(SNP)

Aguilar et al. (2021)

Large data - recursion as basis for genetic evaluation

Pedigree relationships (Henderson, 1976):

$$u_i = f(\textit{sire}, \textit{dam}) + \varphi$$

Genomic relationships:

$$u_i = f(\textit{thousand animals}) + \varphi$$

Misztal et al., 2014

- How many animals in recursion?
About 6 k in chicken
About 14k in Holsteins

APY algorithm



\mathbf{G}^{-1}



$\mathbf{G}_{\text{APY}}^{-1}$

Limited dimensionality of genomic information - chromosome segments

- Theory of junctions (Fisher, 1949):

- Heterogenetic and homogenic tracts in genome



- For randomly mating population of constant size the number junctions (Stam, 1980):

$$E(M_e) = 4 \text{ Effective population size } (N_e) * \text{ Genome size } (L)$$

- Independent chromosome segments M_e (Goddard, 2009; Daetwyler et al., 2010)
- Need $12 M_e$ SNPs to detect 90% of junctions (MacLeod et al., 2005)

Estimated dimensionality, effective population size and optimal number of SNP

Specie	Estimated dimensionality	Effective population size (L=30M)	Optimal number of SNP (12 x Me)
Holsteins	14k	149	180k
Angus	11k	113	130k
Pigs	6k	43 (L=20M)	72k
Chicken	6k	44	72k
Human	360k+	3,000+	5M+

Pocrnic et al. (2016)

APY Single-step GWAS

- **Model**

$$y = W\alpha + Zu + \eta$$

- **Procedure**

1. Calculate $Var(\mathbf{u})^{-1} = \mathbf{H}_{APY}^{-1}$
2. Estimate variance components
3. Calculate $\hat{\mathbf{u}}_{2c}$ and approximate $Var(\hat{\mathbf{u}}_{2c}) = \mathbf{G}_{cc} - \mathbf{C}^{u_{2c}u_{2c}}$
4. For each marker:
 1. Calculate $\hat{b}_i = \mathbf{x}'_{ci} \mathbf{G}_{cc}^{-1} \hat{\mathbf{u}}_2$
 2. Calculate $sd(\hat{b}_i) = \sqrt{\mathbf{x}'_{ci} \mathbf{G}_{cc}^{-1} (\mathbf{G}_{cc} - \mathbf{C}^{u_{2c}u_{2c}}) \mathbf{G}_{cc}^{-1} \mathbf{x}_{ci}}$
 3. Calculate p-value as $pvalue_i = 1 - \Phi\left(\frac{\hat{b}_i}{sd(\hat{b}_i)}\right)$

On the equivalence between marker effect models and breeding value models and direct genomic values with the Algorithm for Proven and Young

Matias Bermann^{1*}, Daniela Lourenco¹, Natalia S. Forneris^{2,3}, Andres Legarra⁴ and Ignacy Misztal¹

Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the Algorithm for Proven and Young

Matias Bermann^{1*}, Daniela Lourenco, and Ignacy Misztal

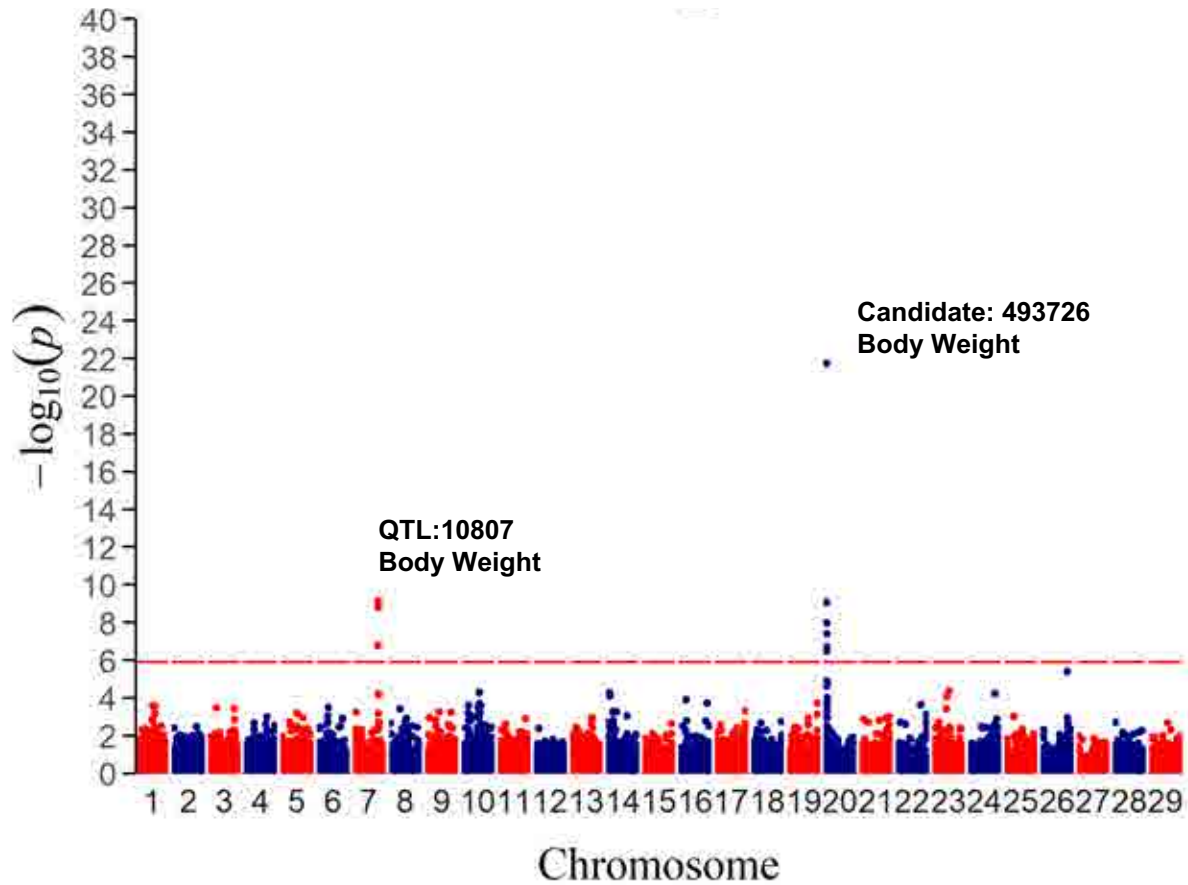
Application example

- Post-weaning gain in American Angus
- 845,000 phenotypes
- 450,000 genotypes
- 1,570,000 animals in the pedigree
- ssGWAS (50k genotyped animals) vs. APY-ssGWAS (450k genotyped animals)

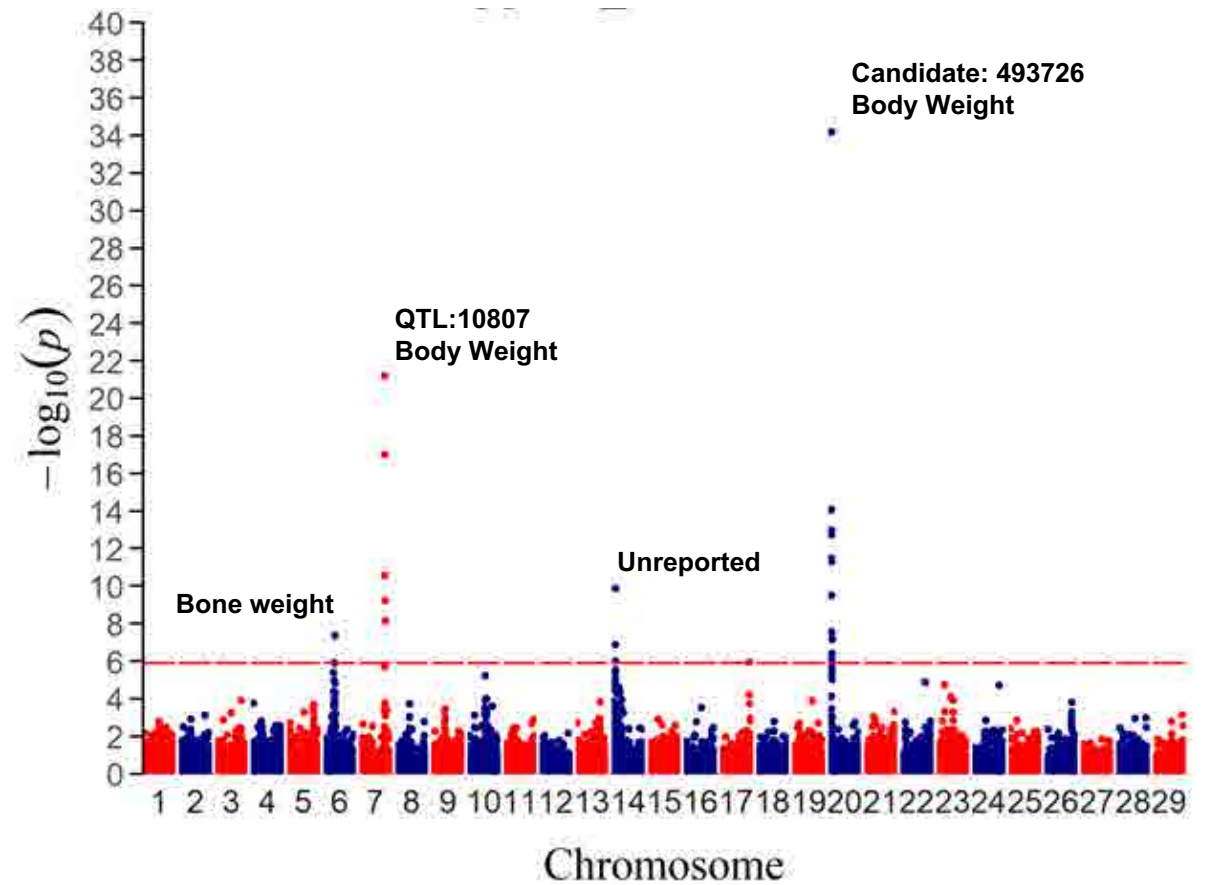


Leite et al.
(2024)

50k genotyped animals



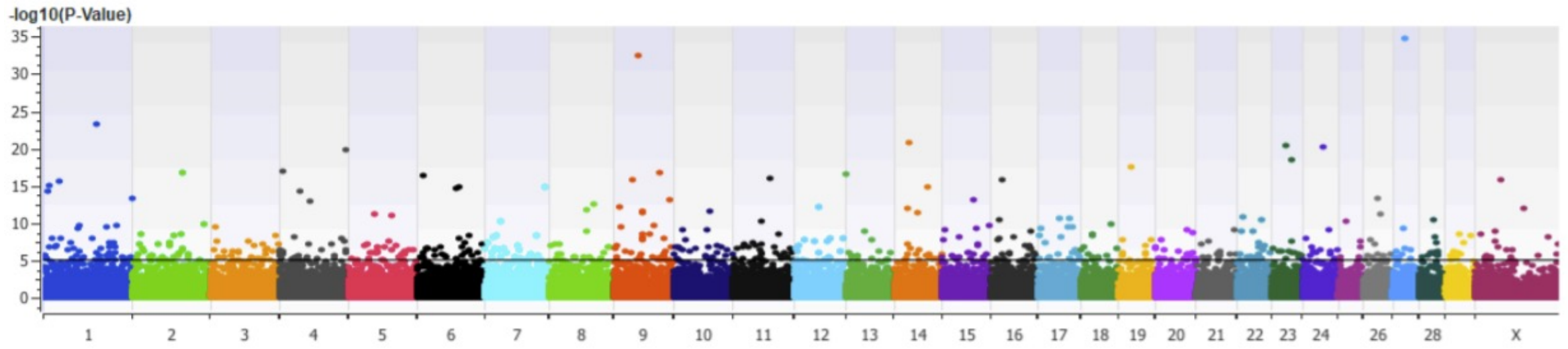
500k genotyped animals



Questions with GWAS and predictions in animal datasets

- GWAS by
 - p-values
 - % of variance explained usually per 1Mb, why 1 Mb?
 -
- Few regions explain $> 1\%$ additive variance
- Lots of QTLs “detected” with small data sets

First conception rate on 2k Holstein heifers



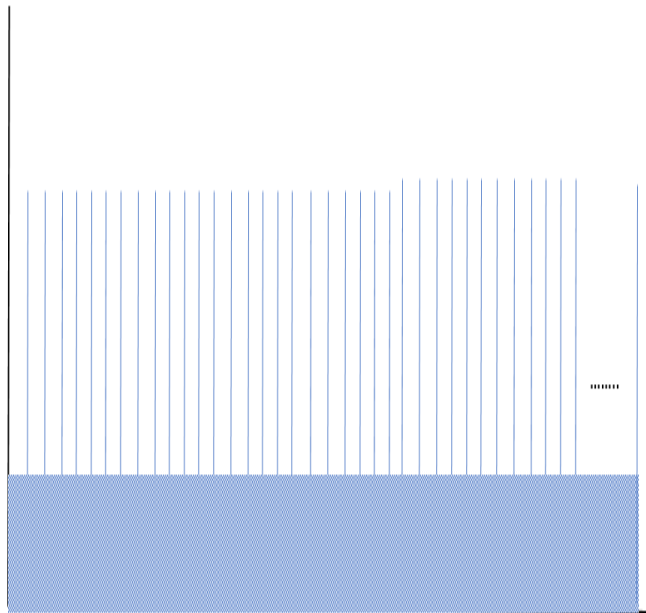
Estimated heritability 36% (normally 1%)

Identified 146 unique loci at $p < 5 \times 10^{-8}$ level

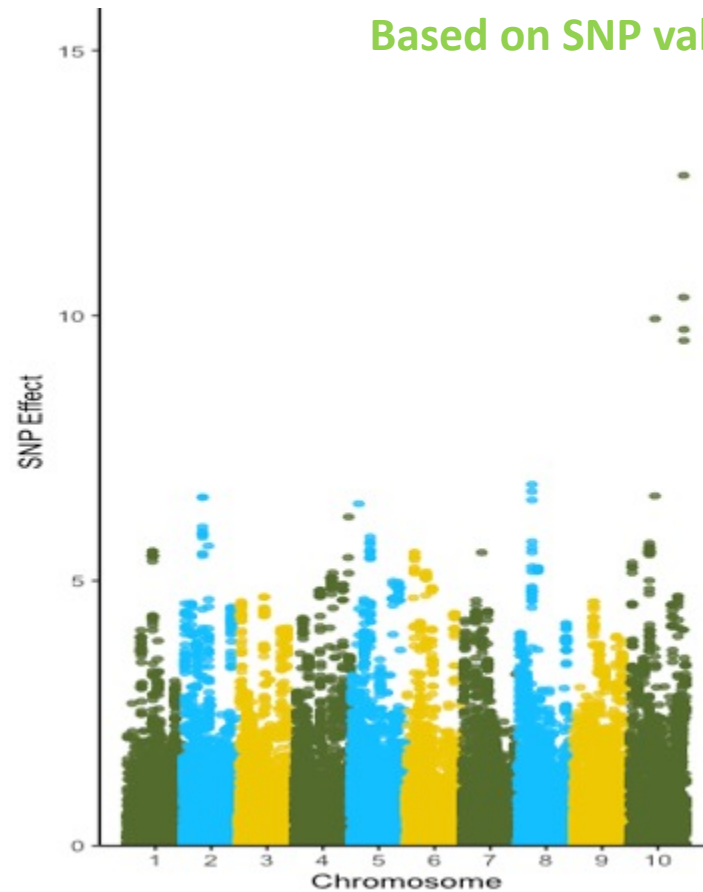
Manhattan plots for simulated population with 100 identical equidistant QTNs



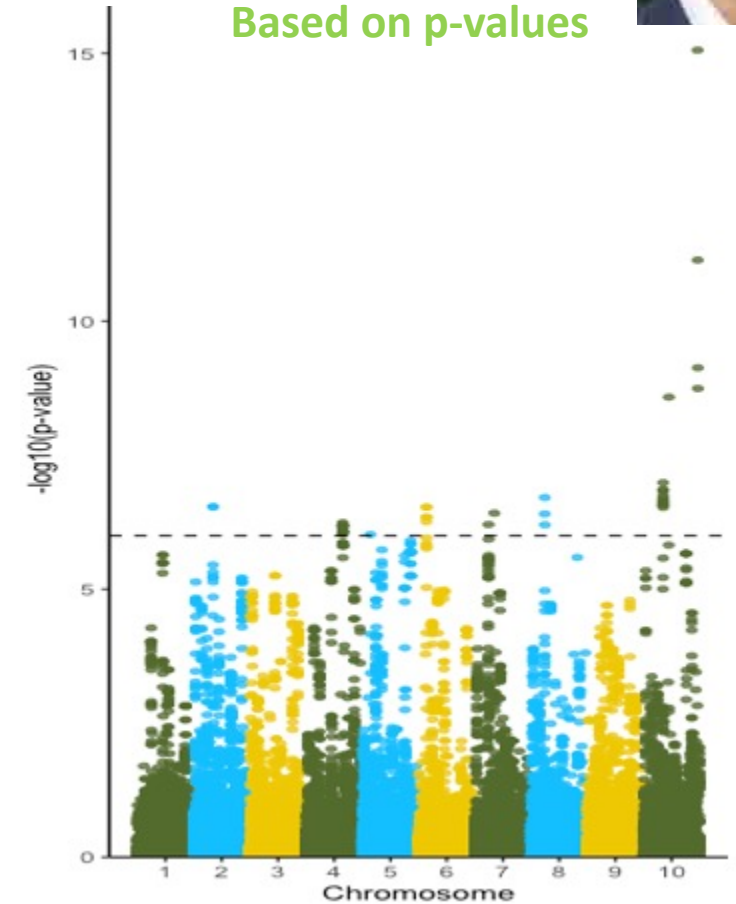
Expectation



Based on SNP values



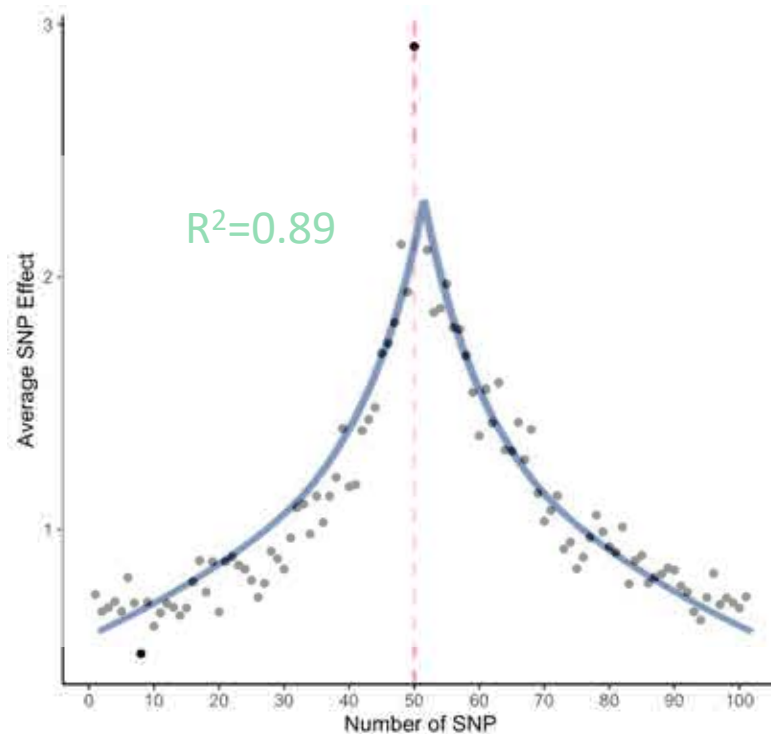
Based on p-values



Work started by Pocrnic et al. (2018)

CfS Mini Conference, Center for Statistics and Human Genetics, University of Edinburgh, March 26, 2024

Plots averaged for 100 QTN



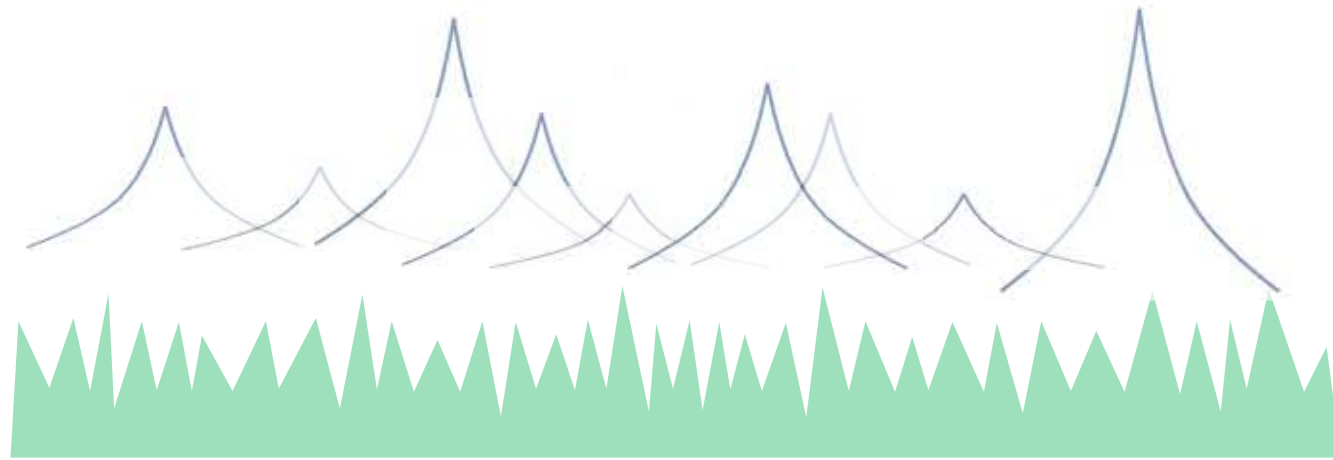
Pairwise linkage disequilibrium curve



~ 2 Mb for cattle
~ 5 Mb for pigs/chickens
~ 15 kb for humans

$1/N_e$ Morgans for 80% QTN variance
 N_e - effective population size

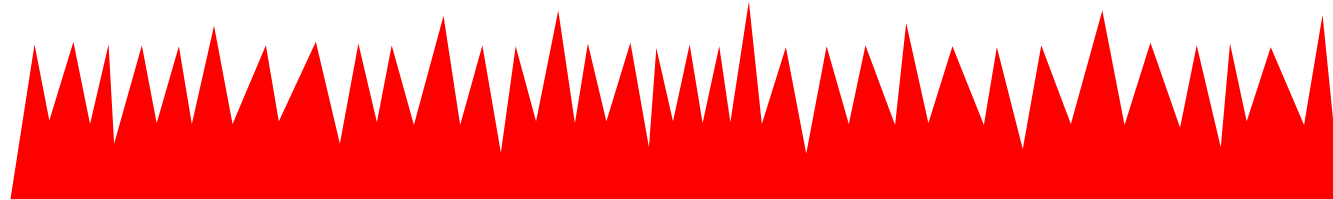
What is Manhattan plot composed of?



QTNs

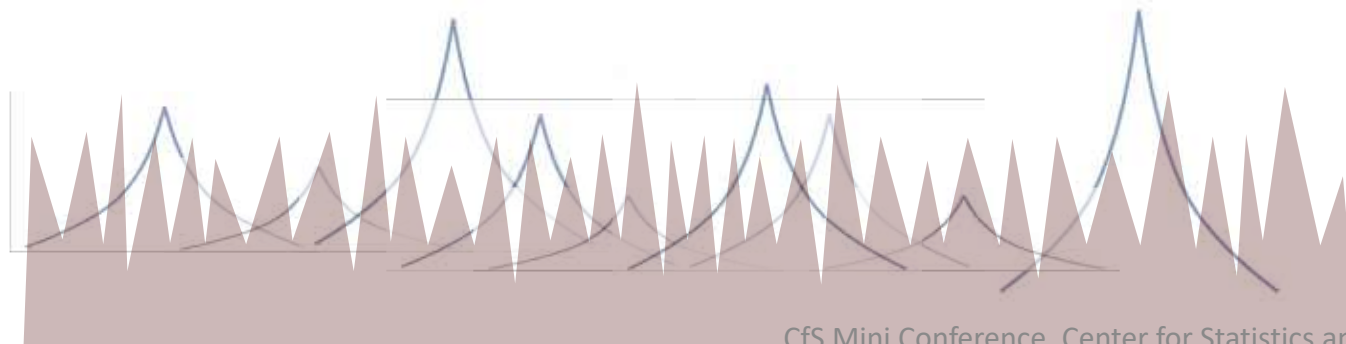
**Bigger with larger QTN
and larger data**

Relationships



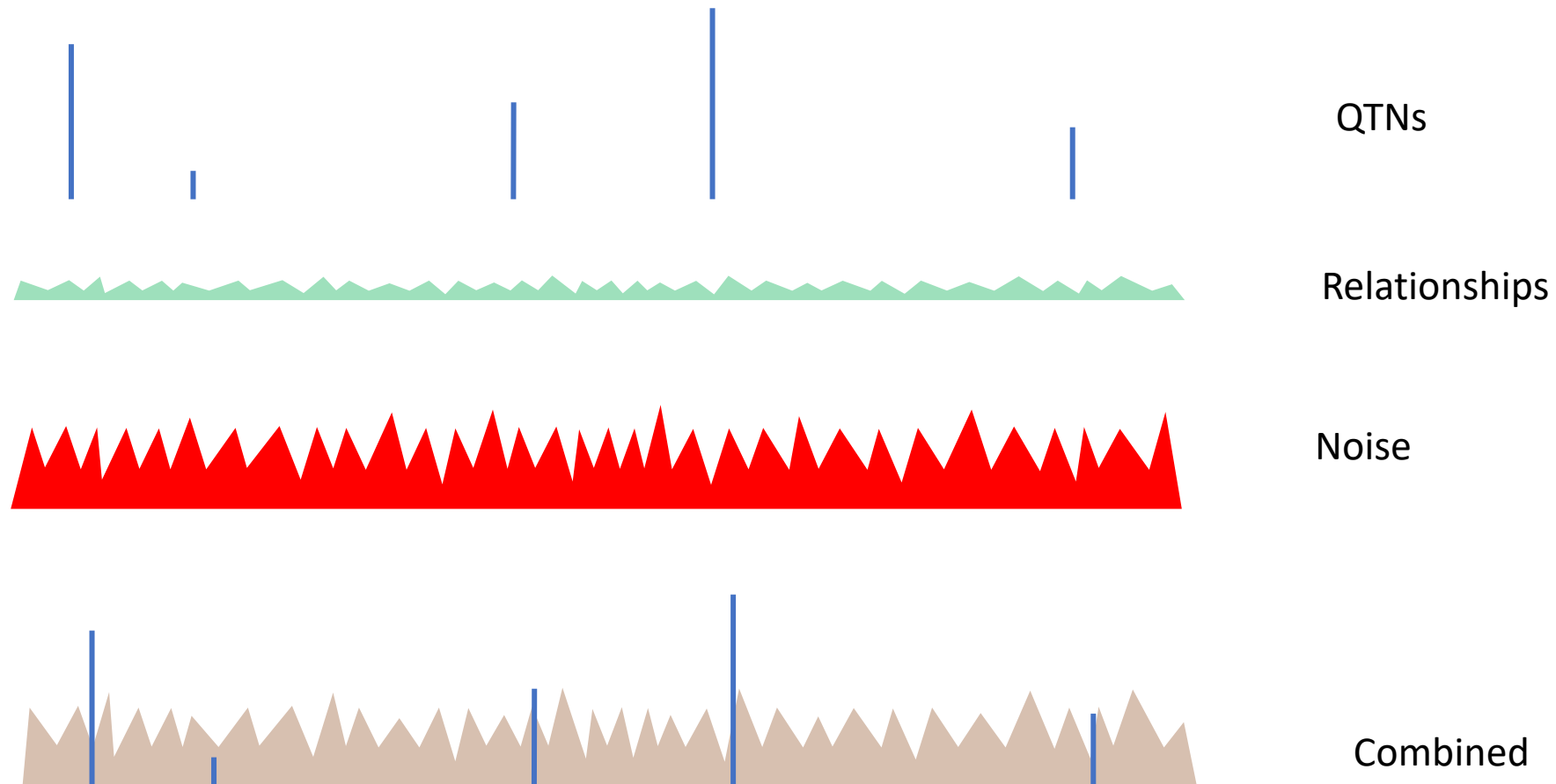
Noise

Smaller with more data

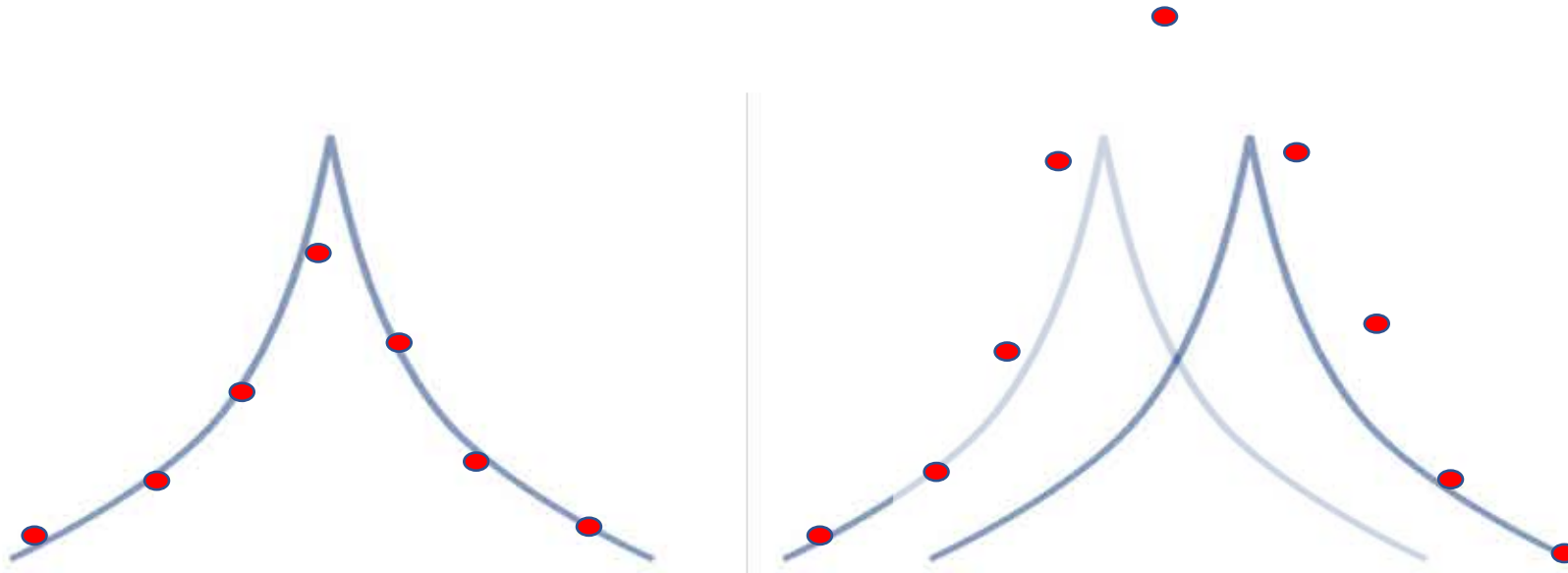


Combined

Large effective population size



Why GBLUP accounts for QTN?

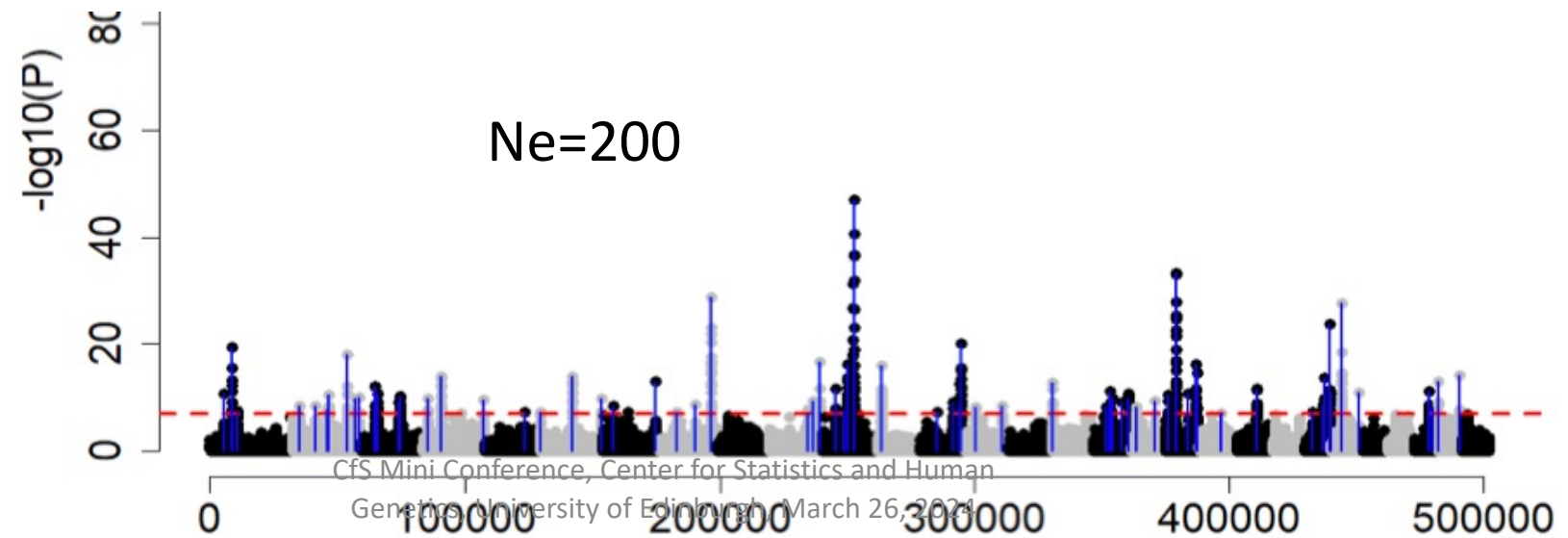
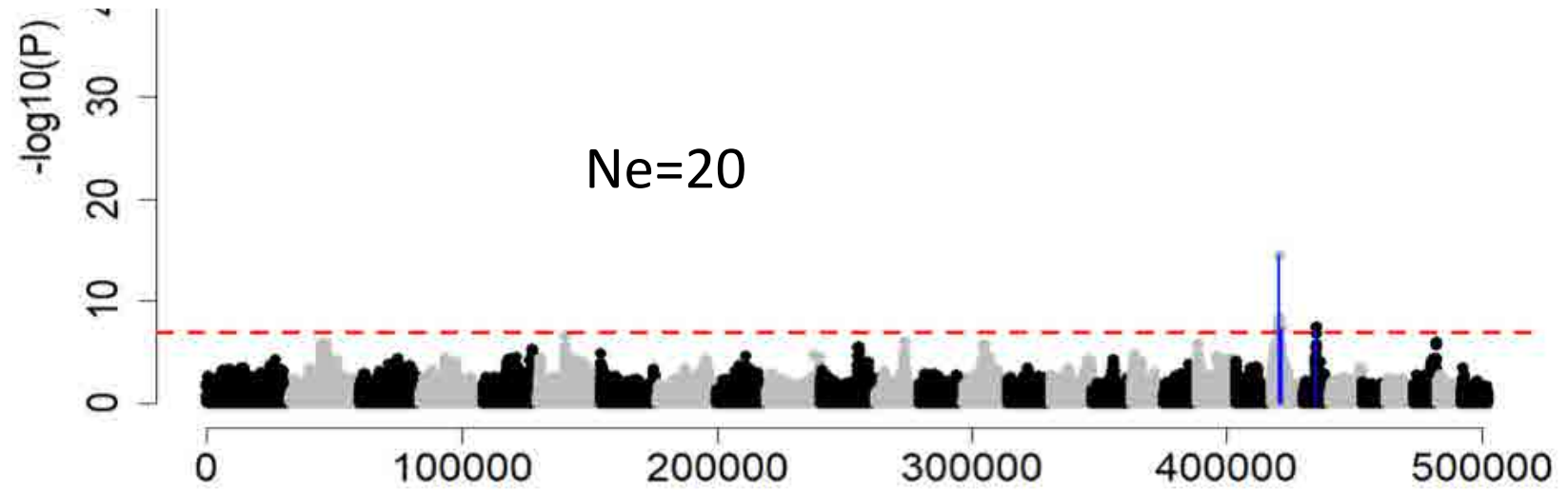


If 4 SNP per segment, 32 SNP account for 80% of QTN variance

Need chip with 16 NeL SNP to mostly account for QTN

About 20k for pigs/broilers, 60k for cattle, 5m for humans

Effective population size affects GWAS



Sungbong et al., 2021

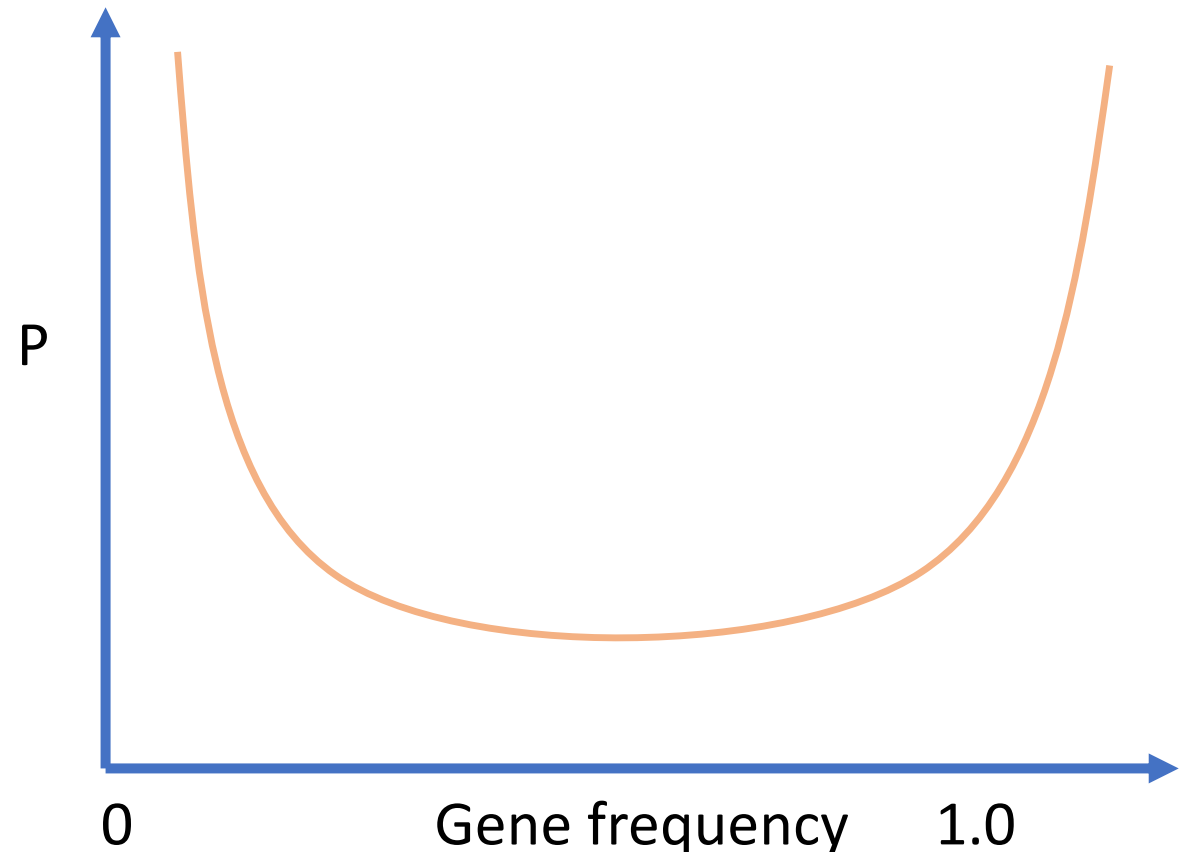
Why few QTN detected?

JOURNAL ARTICLE

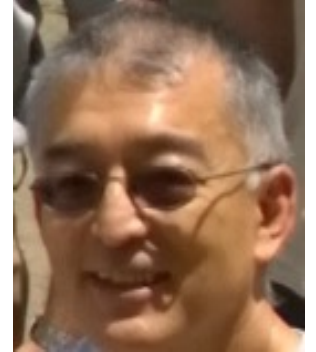
AlphaSimR: an R package for breeding program simulations

R Chris Gaynor , Gregor Gorjanc, John M Hickey

Only 20-30% QTN with $p > 0.3$

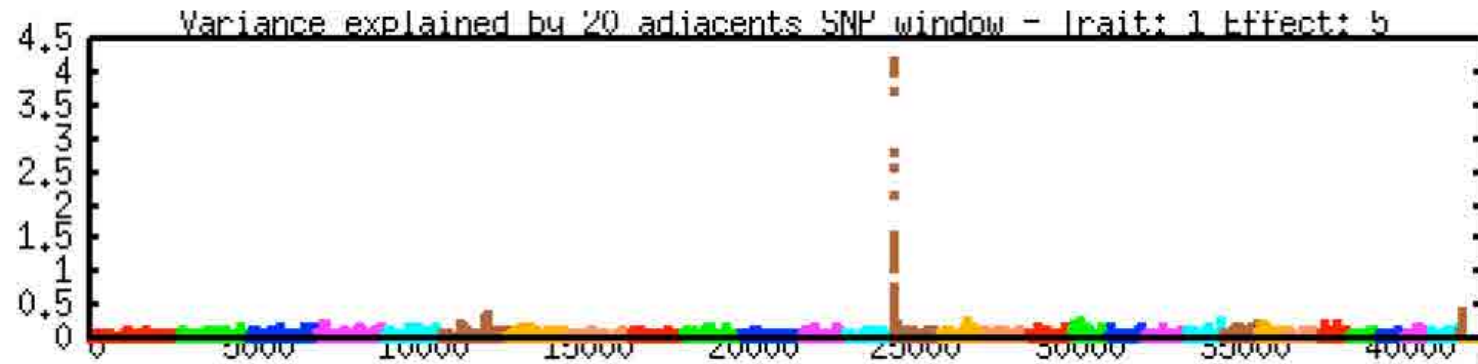


Can large QTL exist despite selection?

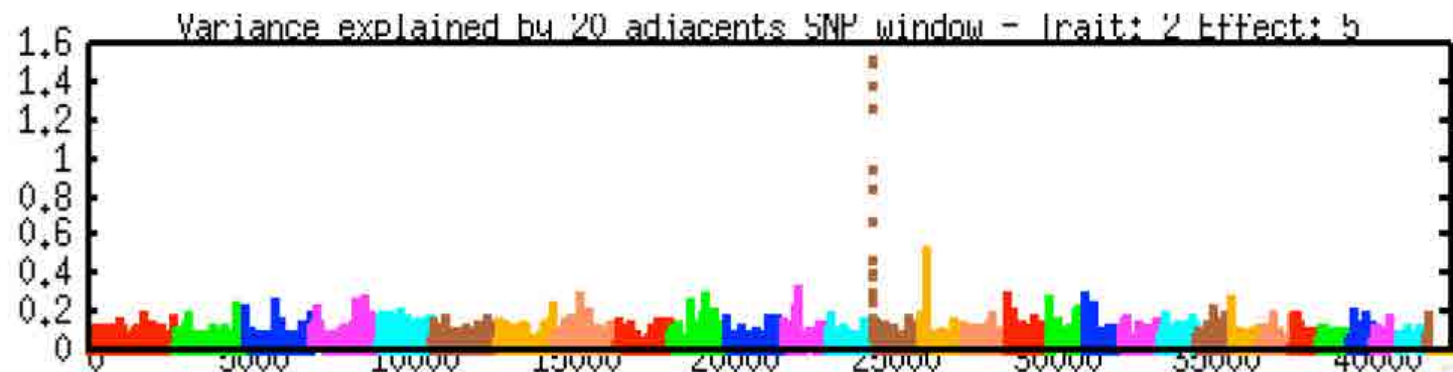


- Genetics and genomics of mortality in US Holsteins
- (Tokuhisa et al, 2014; Tsuruta et al., 2014)
- 6M records, SNP50k genotypes of 35k bulls

Milk – first parity



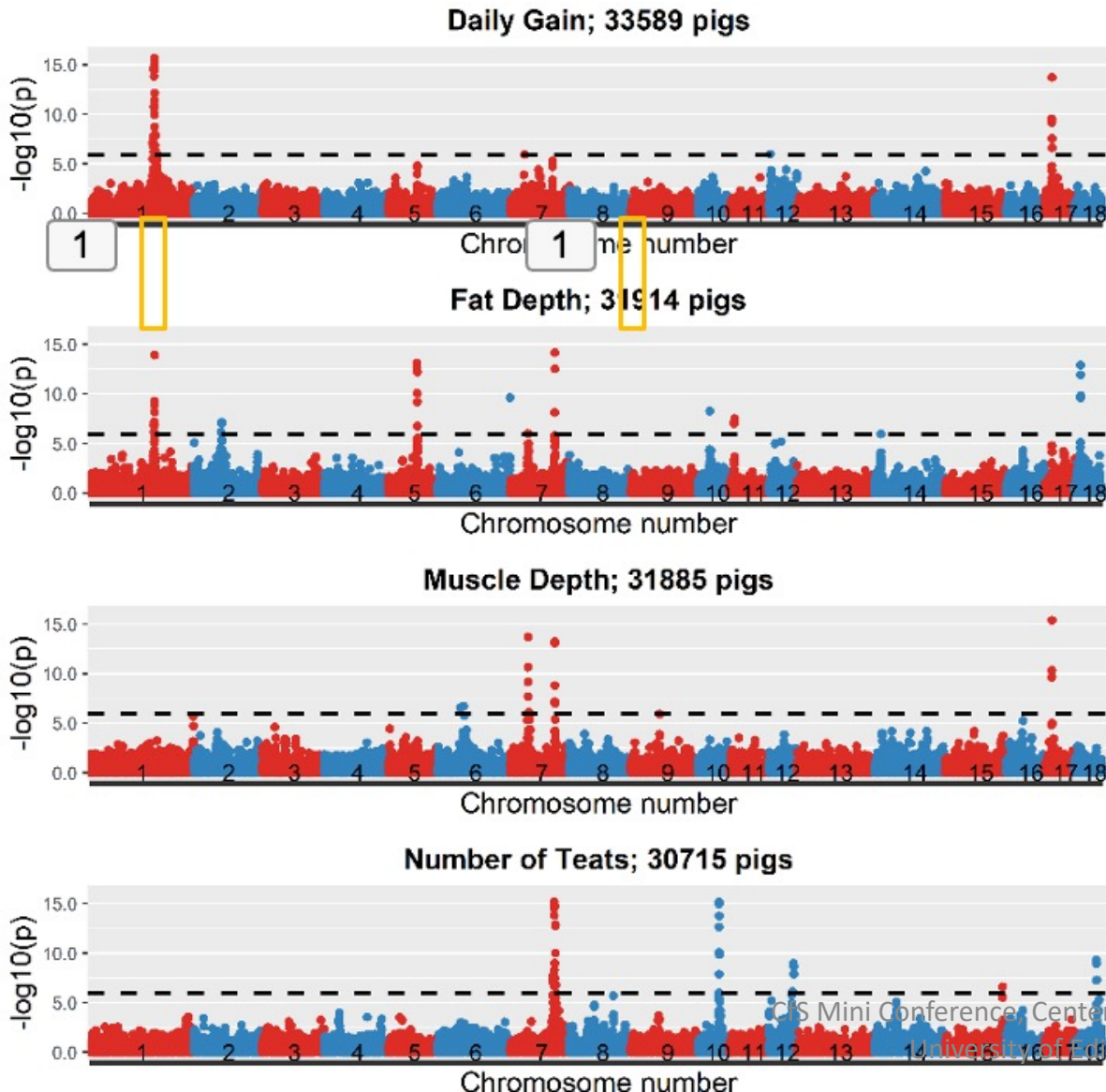
Mortality – first parity



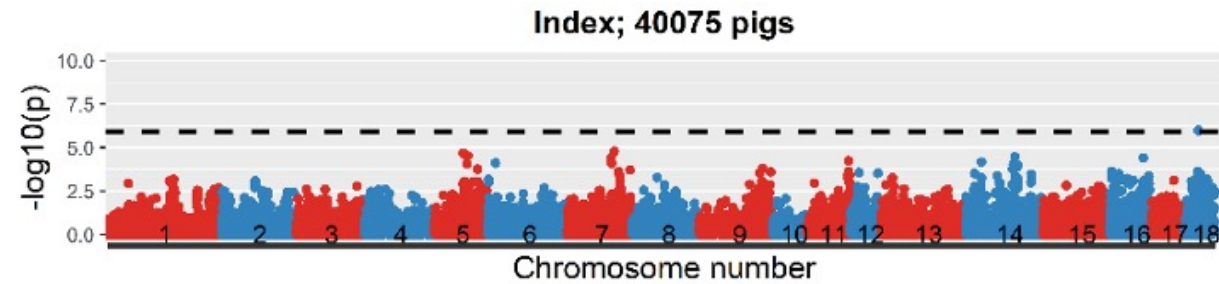
GWAS for various traits and index in pigs



Bijma, EAAP 23



Index



- Different peaks in different lines
- Antagonistic pleiotropy

Conclusions

- GWAS affected by effective population size
- Optimal window size for GWAS 1-2 Mb for $N_e=100$
- Large signals in GWAS due to QTN, relationships and noise (incl. Imputation)
- Large QTL in farm populations show pleiotropy – QTL not visible in index

UGA AB&G team

