Progress in GWAS for large datasets with GBLUP and single-step GBLUP

Ignacy Misztal, Daniela Lourenco and Matias Bermann

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

Single-step GBLUP –pedigree and genomic relationships combined

Matrix H (Legarra ,2009)

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

Inverse of H (Aguilar et al., 2010)

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

G –genomic relationship matrix

1 –ungenotyped animals

2-genotyped animals

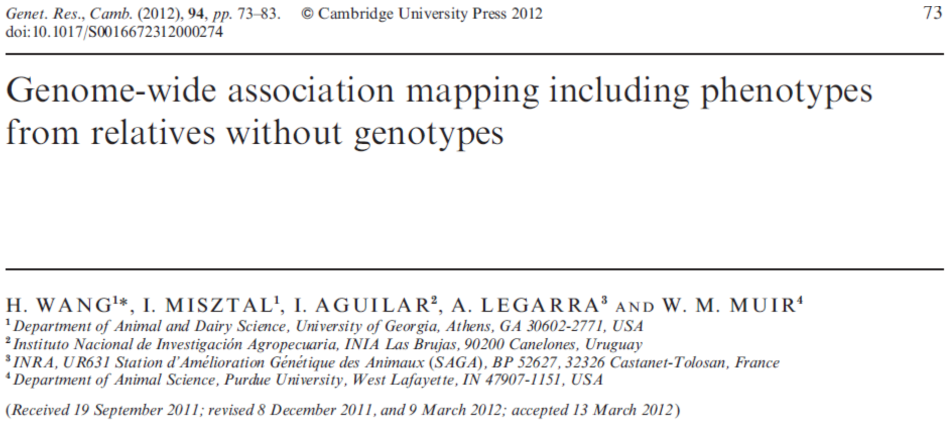
Christensen and Lund, 2010

Boemcke et al., 2011

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?



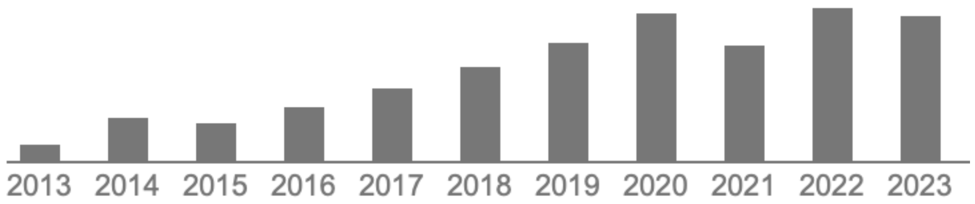
Genet. Res., Camb. (2012), **94**, *pp.* 73–83. © Cambridge University Press 2012 73
doi:10.1017/S0016672312000274

Genome-wide association mapping including phenotypes from relatives without genotypes

H. WANG[1]*, I. MISZTAL[1], I. AGUILAR[2], A. LEGARRA[3] AND W. M. MUIR[4]
[1] *Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602-2771, USA*
[2] *Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90200 Canelones, Uruguay*
[3] *INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), BP 52627, 32326 Castanet-Tolosan, France*
[4] *Department of Animal Science, Purdue University, West Lafayette, IN 47907-1151, USA*

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

Cited by 537



GWAS with ssGBLUP (Wang et al., 2012)

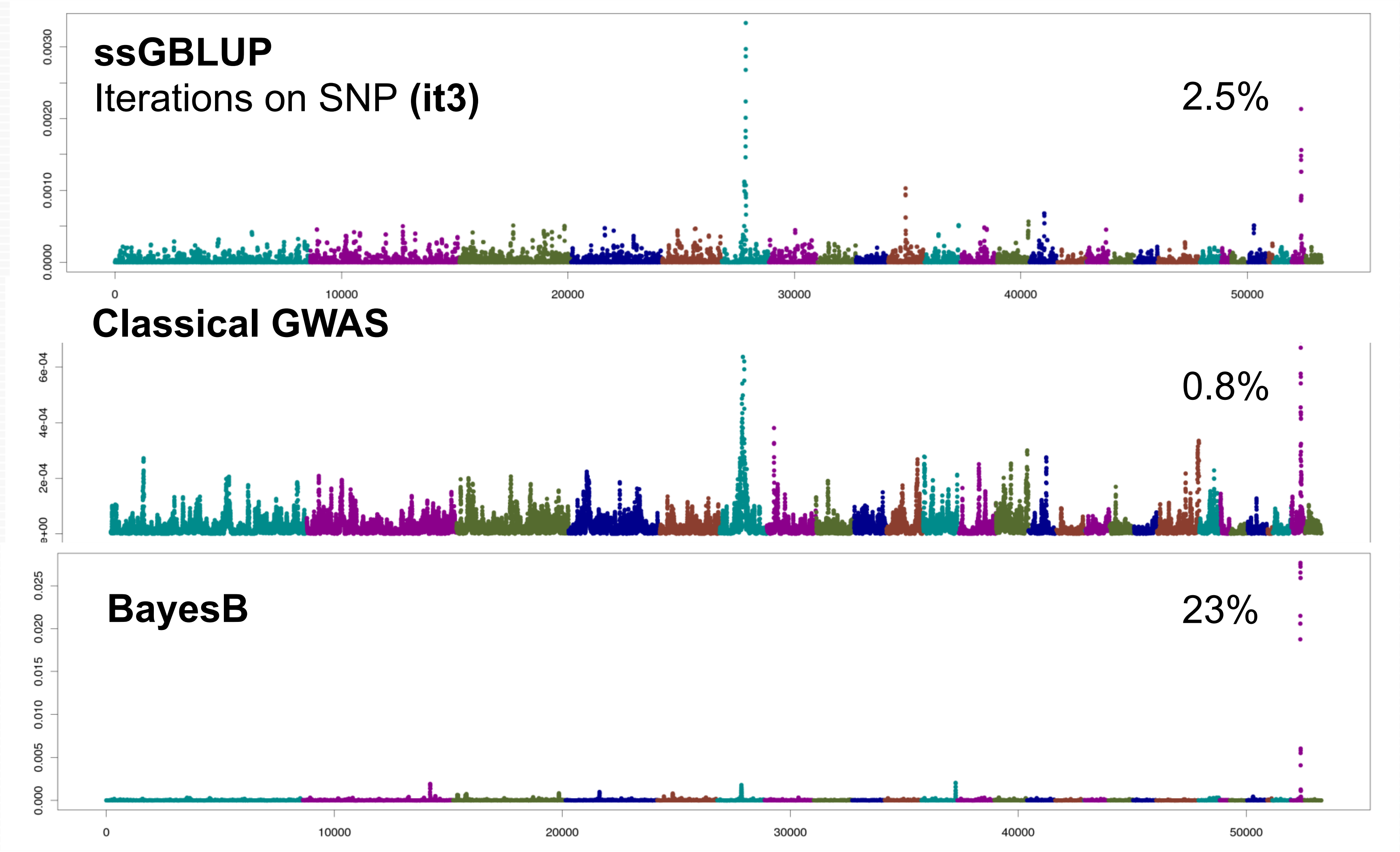
- Convert GEBV to SNP effects
- Estimate individual SNP variances
- Incorporate variances in G
- Possibly recompute GEBV and iterate

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. $D=n\ D/tr(D)$
7. Loop to 2

Output as % of variance explained in a window

Discrepancies in GWAS methods

Chicken weight



P-values for GWAS in (ss)GBLUP

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

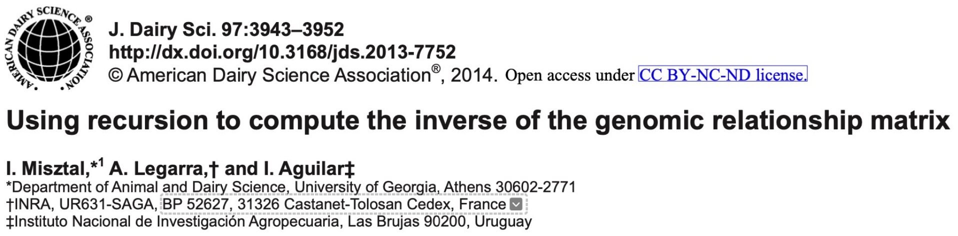
If $sd(\widehat{snp}_i)$ approximately constant, Manhattan plots based on $|\widehat{snp}_i|$ and $pval_i$ similar

Large data – APY algorithm

- Due to LD, genomic information compresses well: about 15k for cattle and about 5k for pigs and chicken

- APY algorithm: $u_{noncore}$ = P u_{core},+ ε

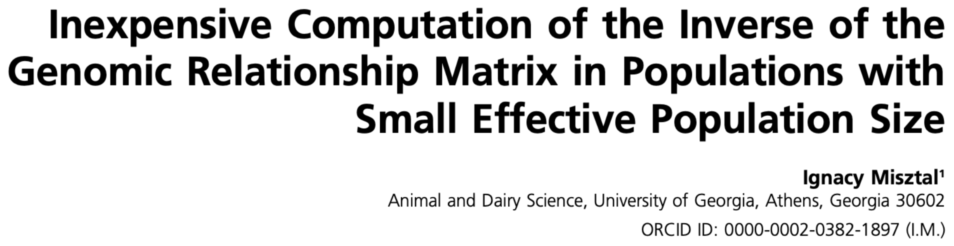
- Number of core animals ~ equal to dimensionality



J. Dairy Sci. 97:3943–3952
http://dx.doi.org/10.3168/jds.2013-7752
© American Dairy Science Association®, 2014. Open access under CC BY-NC-ND license.

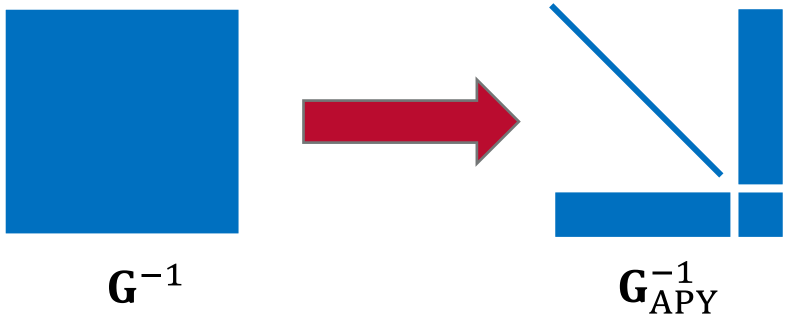
Using recursion to compute the inverse of the genomic relationship matrix

I. Misztal,*[1] A. Legarra,† and I. Aguilar‡
*Department of Animal and Dairy Science, University of Georgia, Athens 30602-2771
†INRA, UR631-SAGA, BP 52627, 31326 Castanet-Tolosan Cedex, France
‡Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay



Inexpensive Computation of the Inverse of the Genomic Relationship Matrix in Populations with Small Effective Population Size

Ignacy Misztal[1]
Animal and Dairy Science, University of Georgia, Athens, Georgia 30602
ORCID ID: 0000-0002-0382-1897 (I.M.)



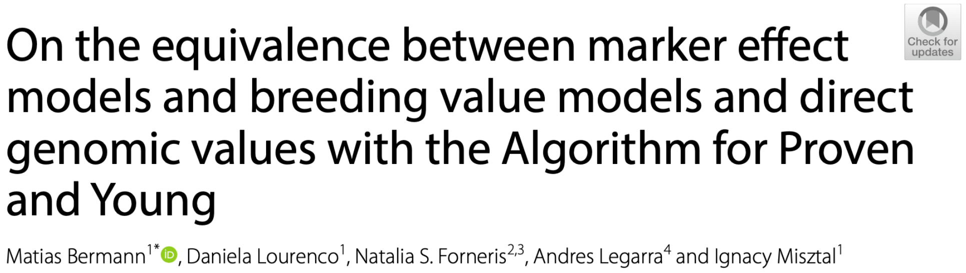
$$\mathbf{A}^{-1}+\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\end{pmatrix} \longrightarrow \mathbf{A}^{-1}+\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}_{\mathrm{APY}}^{-1}-\mathbf{A}_{22}^{-1}\end{pmatrix}$$

APY Single-step GWAS

- **Model**

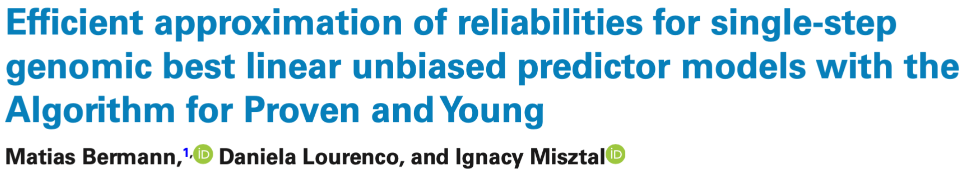
$$\boldsymbol{y} = \boldsymbol{W\alpha} + \boldsymbol{Zu} + \boldsymbol{\eta}$$

- **Procedure**
  1. Calculate $Var(\boldsymbol{u})^{-1} = \boldsymbol{H}_{APY}^{-1}$
  2. Estimate variance components
  3. Calculate $\hat{\boldsymbol{u}}_{2_c}$ and approximate $Var\left(\hat{\boldsymbol{u}}_{2_c}\right) = \boldsymbol{G}_{cc} - \boldsymbol{C}^{\boldsymbol{u}_{2c}\boldsymbol{u}_{2c}}$
  4. For each marker:
     1. Calculate $\hat{b}_i = \boldsymbol{x}'_{\boldsymbol{c}_i}\boldsymbol{G}_{cc}^{-1}\,\hat{\boldsymbol{u}}_{\boldsymbol{2}}$
     2. Calculate $sd\left(\hat{b}_i\right) = \sqrt{\boldsymbol{x}'_{c_i}\boldsymbol{G}_{cc}^{-1}(\boldsymbol{G}_{cc} - \boldsymbol{C}^{\boldsymbol{u}_{2c}\boldsymbol{u}_{2c}})\boldsymbol{G}_{cc}^{-1}\boldsymbol{x}_{c_i}}$
     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 Lourenco[1], Natalia S. Forneris[2,3], Andres Legarra[4] and Ignacy Misztal[1]



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

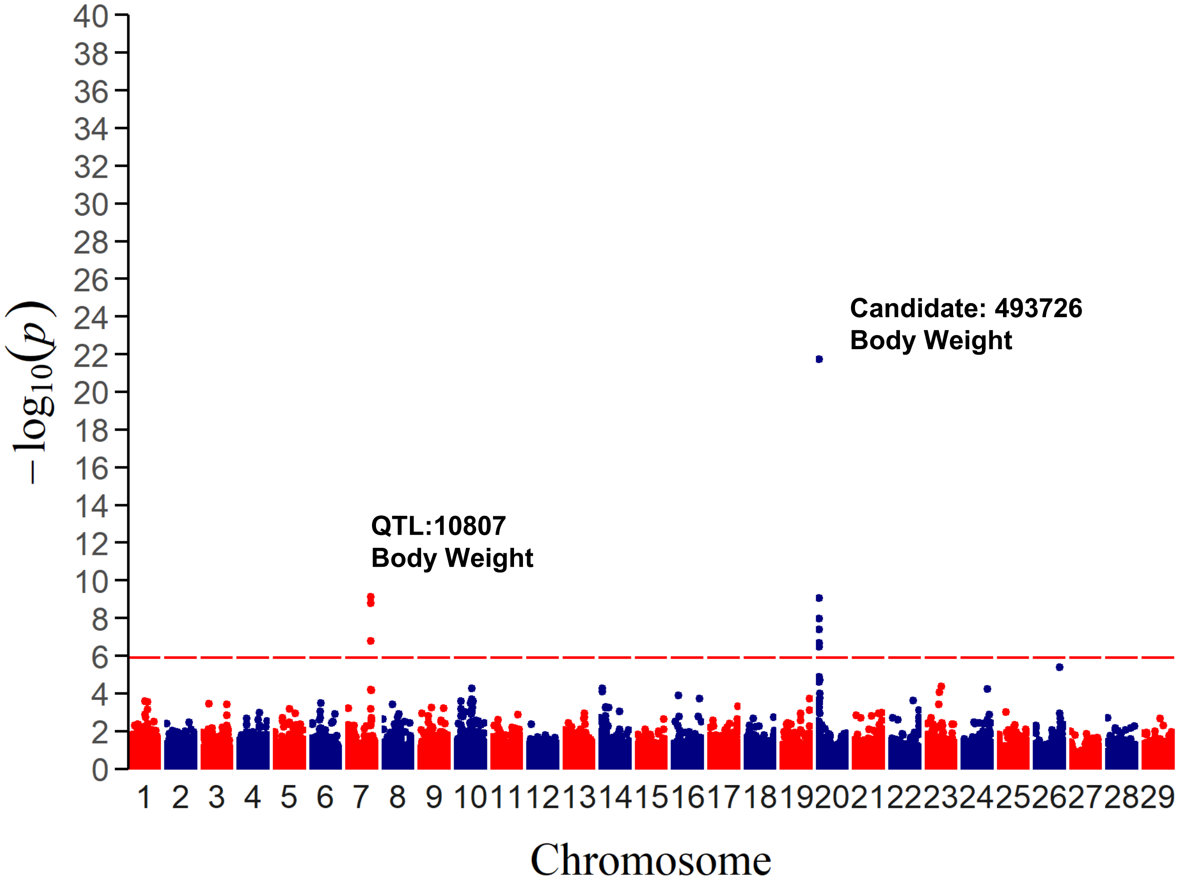
Matias Bermann,[1] Daniela Lourenco, and Ignacy Misztal

Application example

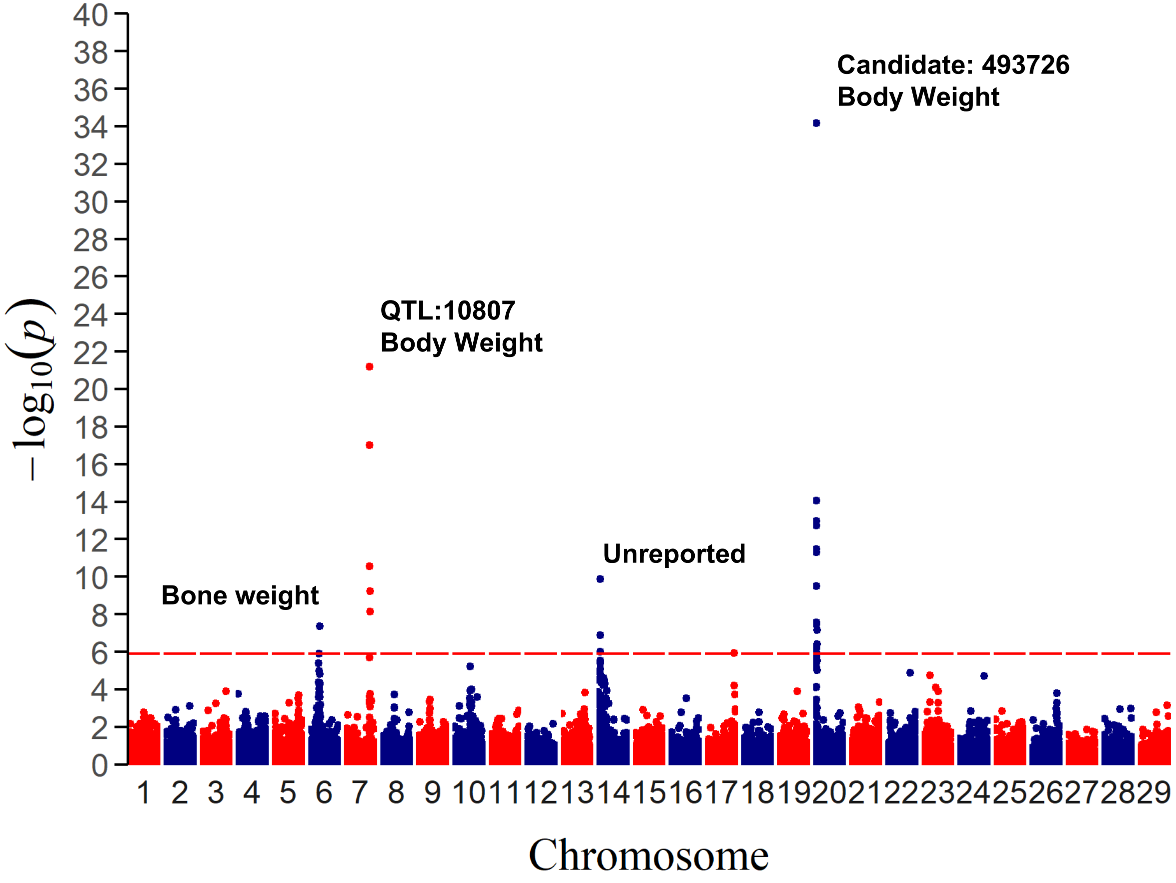
- Post-weaning gain in American Angus
- 845,000 phenotypes
- 450,000 genotypes
- 1,570,000 animals in the pedigree
- ssGWAS (50k genotyped animals) vs. APY-ssGWAS (450k genotyped animals)
- We expect:
  - Higher power
  - Less noise
  - Less false-positives

Leite et al.
(in progress)

50k genotyped animals



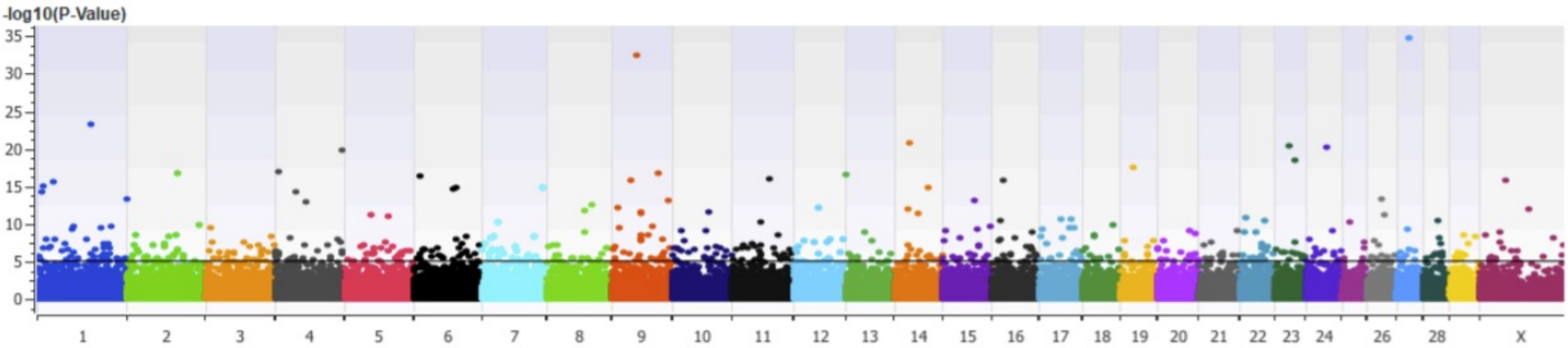
500k genotyped animals



Questions with GWAS and predictions

- GWAS by
  - % of variance explained usually per 1Mb
  - p-values
- Few regions explain > 1% additive variance
- Lots of QTLs detected with small data sets
- Fewer QTLs detected with large data

First conception rate on 2k Holstein heifers



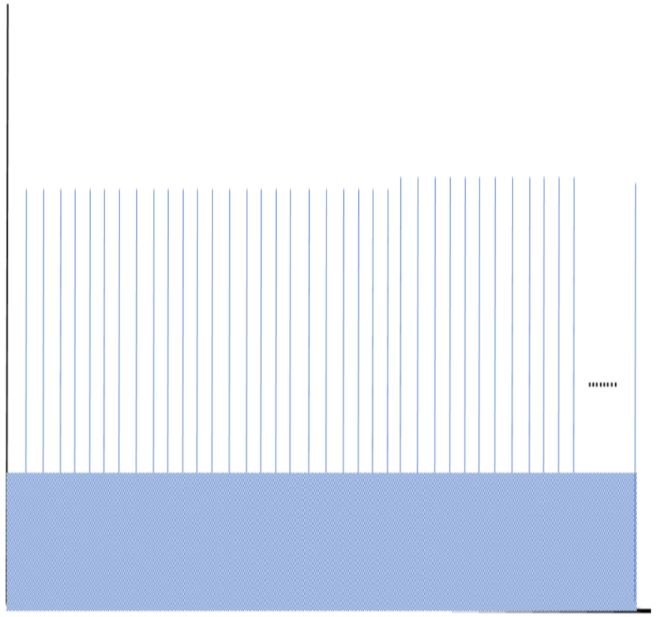
Estimated heritability 36% (normally 1%)

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

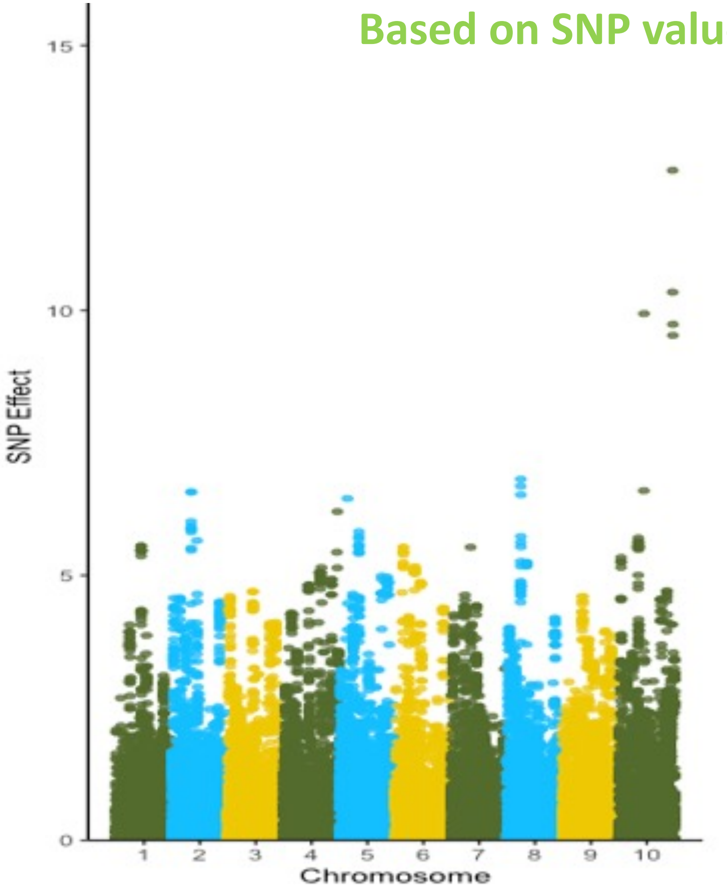
Galliou et al., 2020, https://doi.org/10.3390/genes11070767

Manhattan plots for simulated population with 100 identical equidistant QTNs

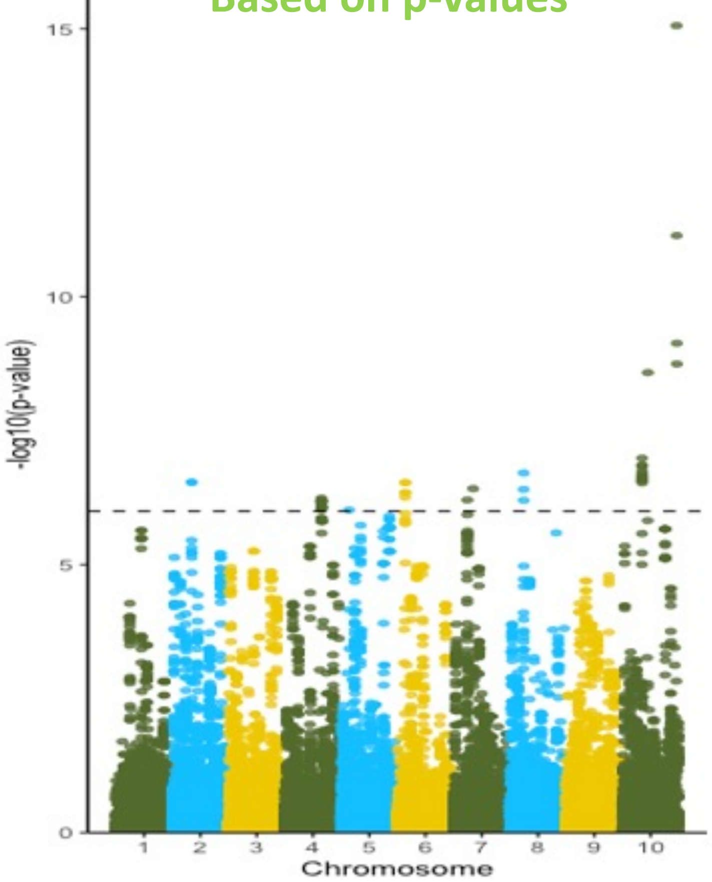
Expectation



Based on SNP values

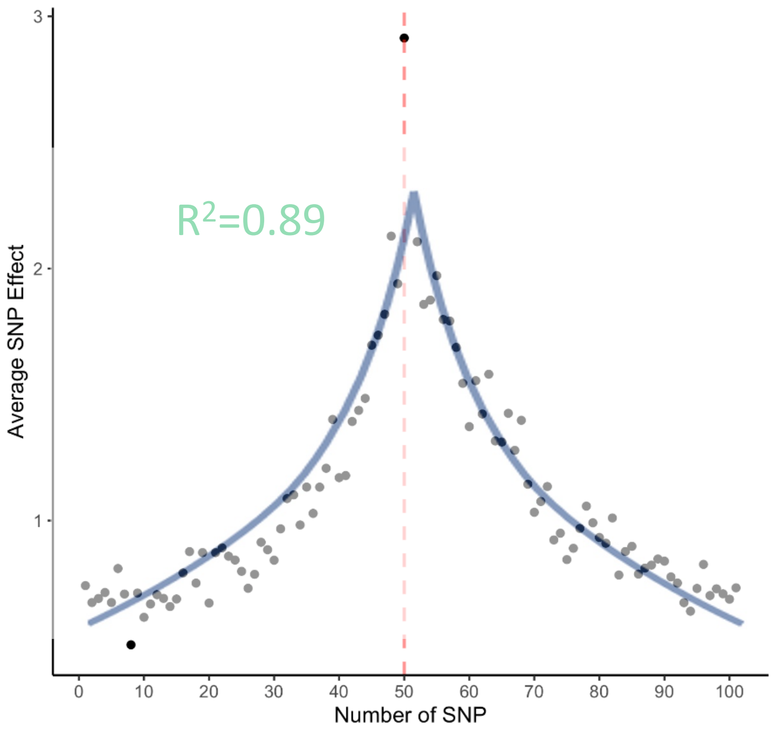


Based on p-values



Work started by Pocrnic et al. (2018)

Plots averaged for 100 QTN



Pairwise linkage disequilibrium curve

~ 2 Mb for cattle

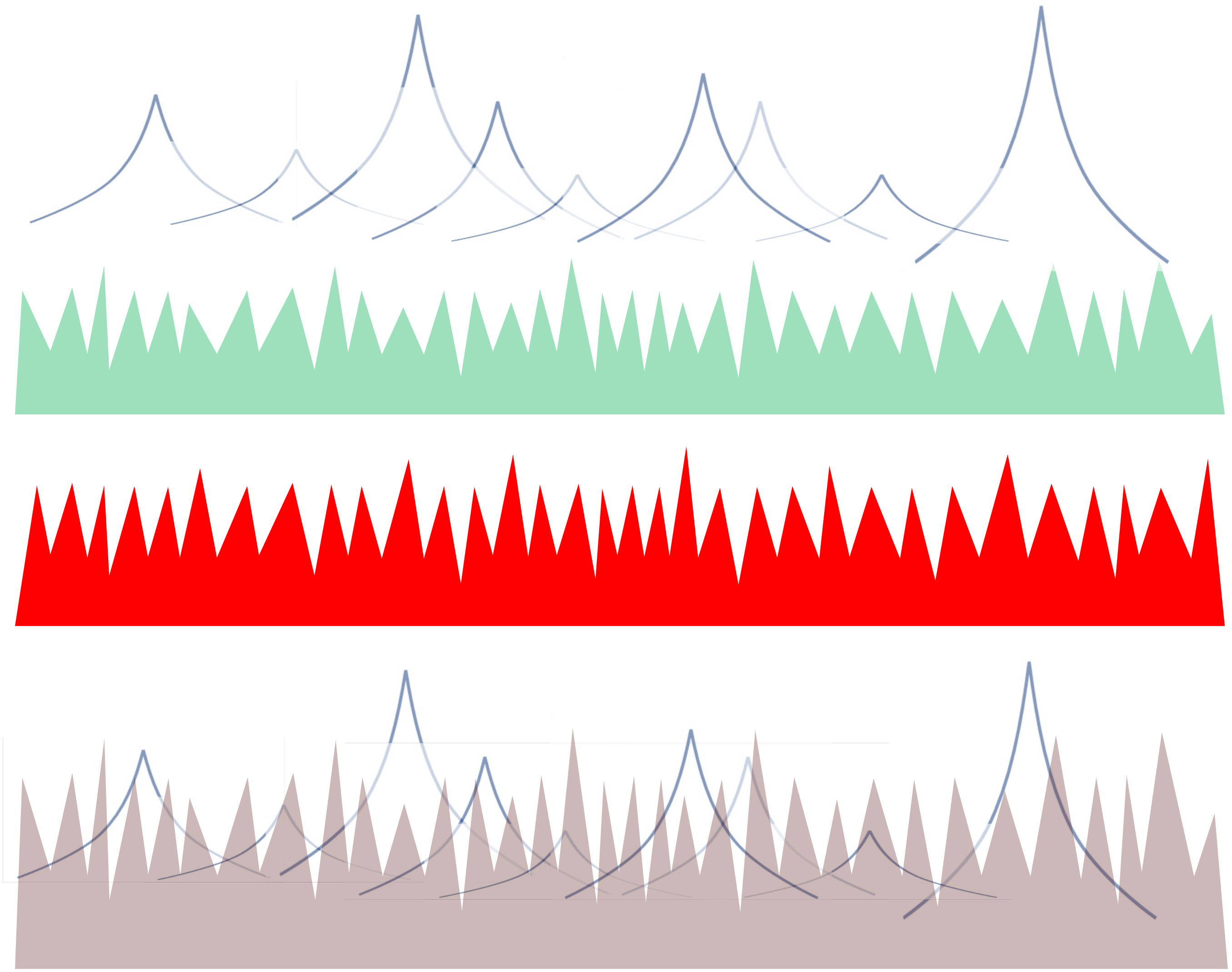
~ 5 Mb for pigs/chickens

~ 15 kb for humans

1/Ne Morgans for 80% QTN variance

Ne - effective population size

What is Manhattan plot composed of?



QTNs

Bigger with larger QTN and larger data

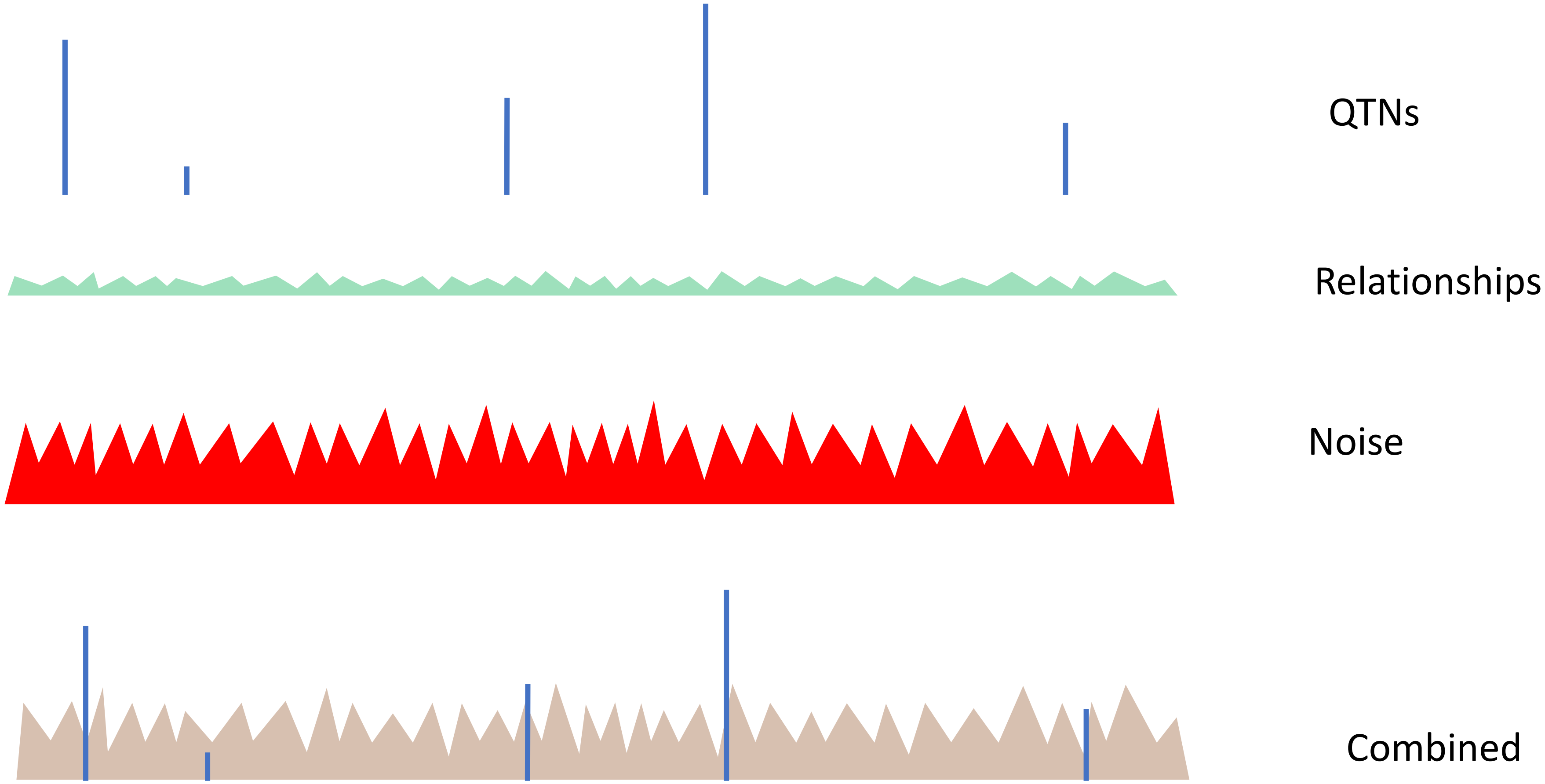
Relationships

Noise

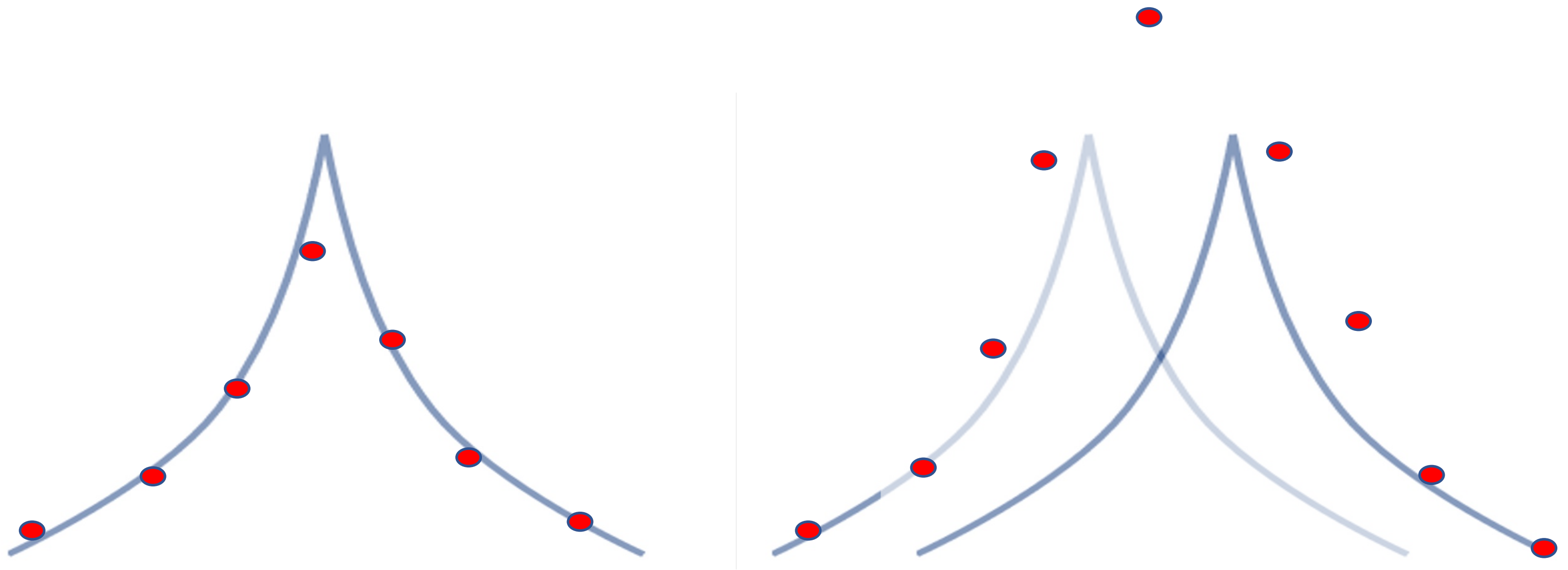
Smaller with more data

Combined

Large effective population size



Why GBLUP accounts for QTN?

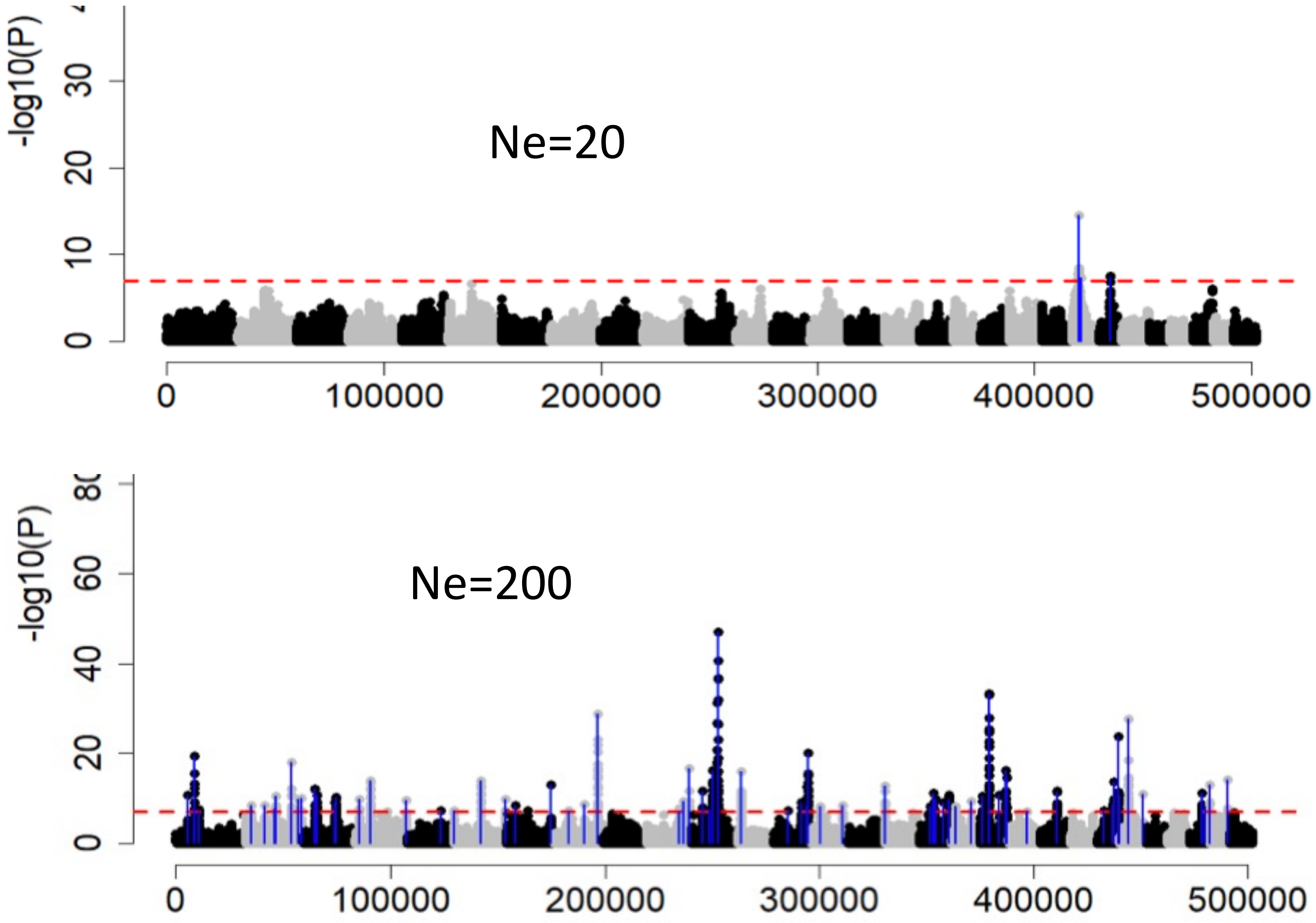


If 4 SNP per segment, 32 SNP account for 80% of QTN variance

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

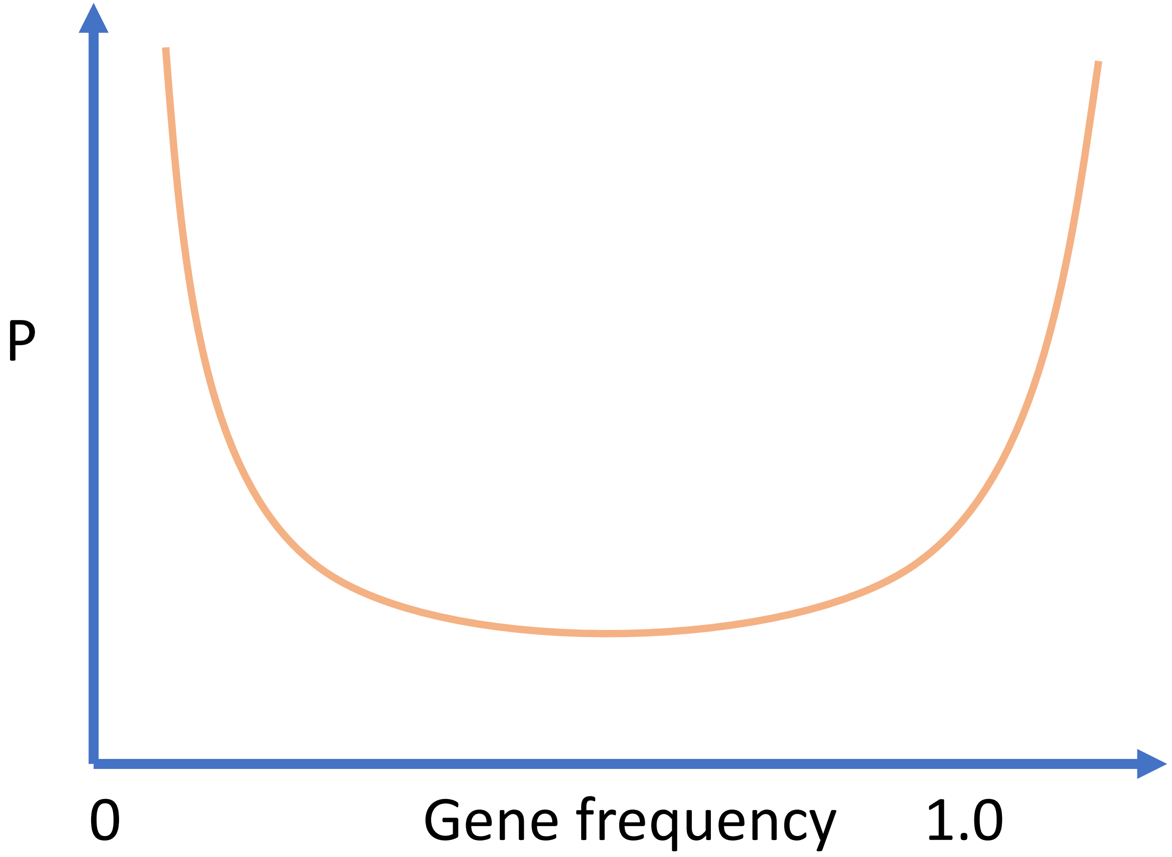
Why few QTN detected?

JOURNAL ARTICLE

AlphaSimR: an R package for breeding program simulations

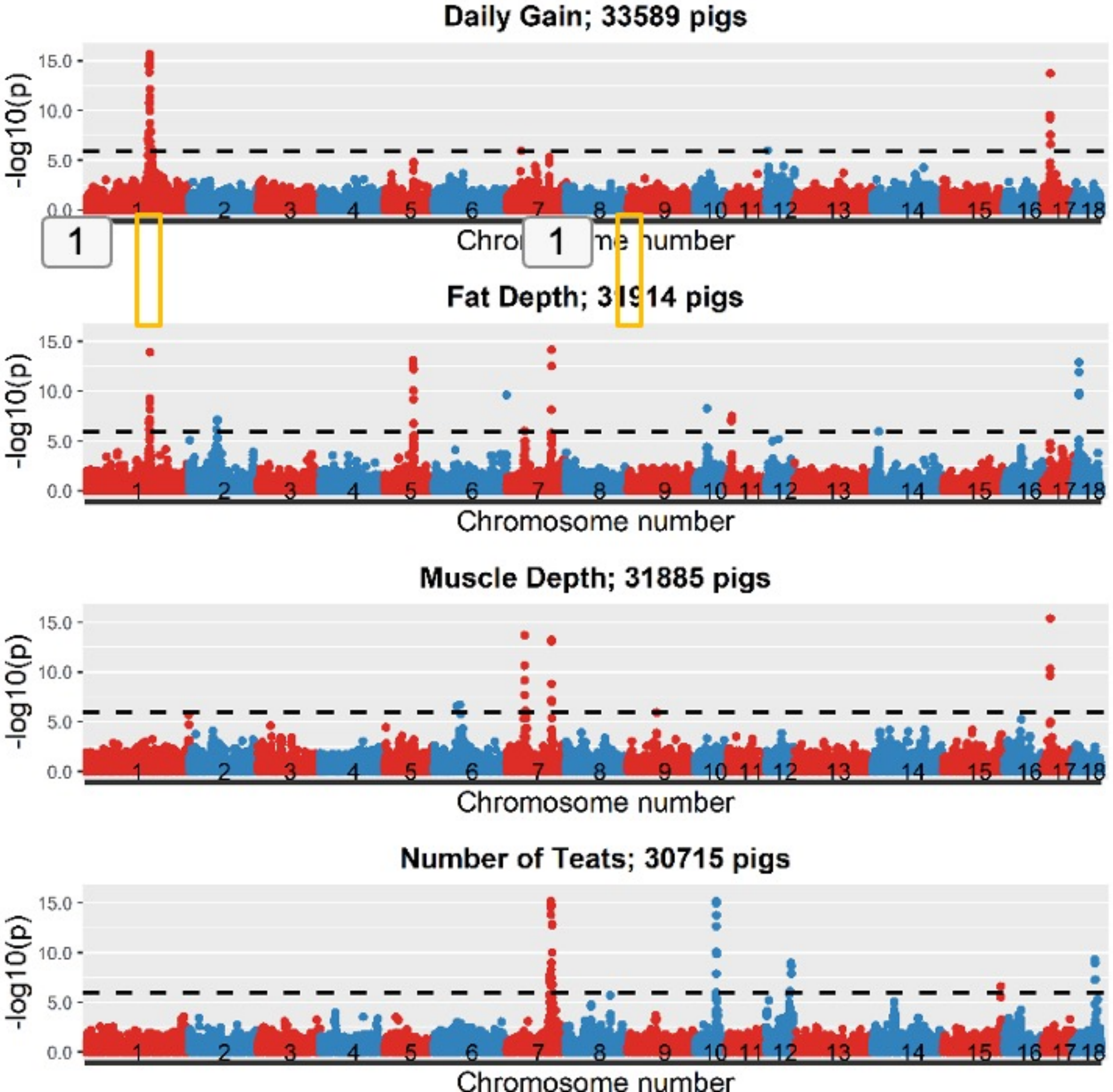
R Chris Gaynor ✉, Gregor Gorjanc, John M Hickey

Only 20-30% QTN with p > 0.3

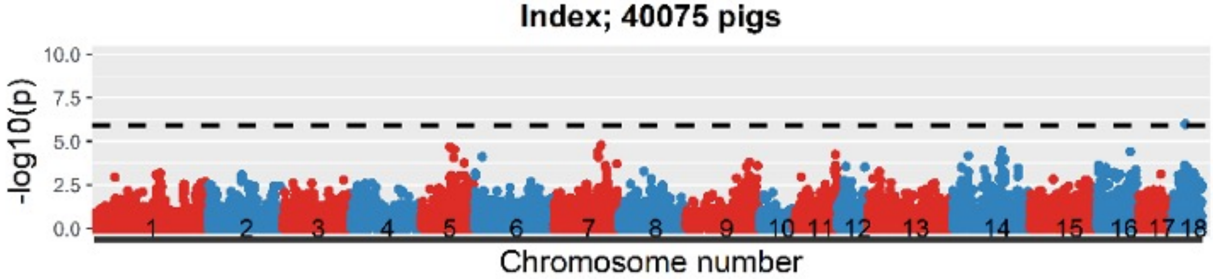


GWAS for various traits and index in pigs

Bijma, EAAP 23



Index



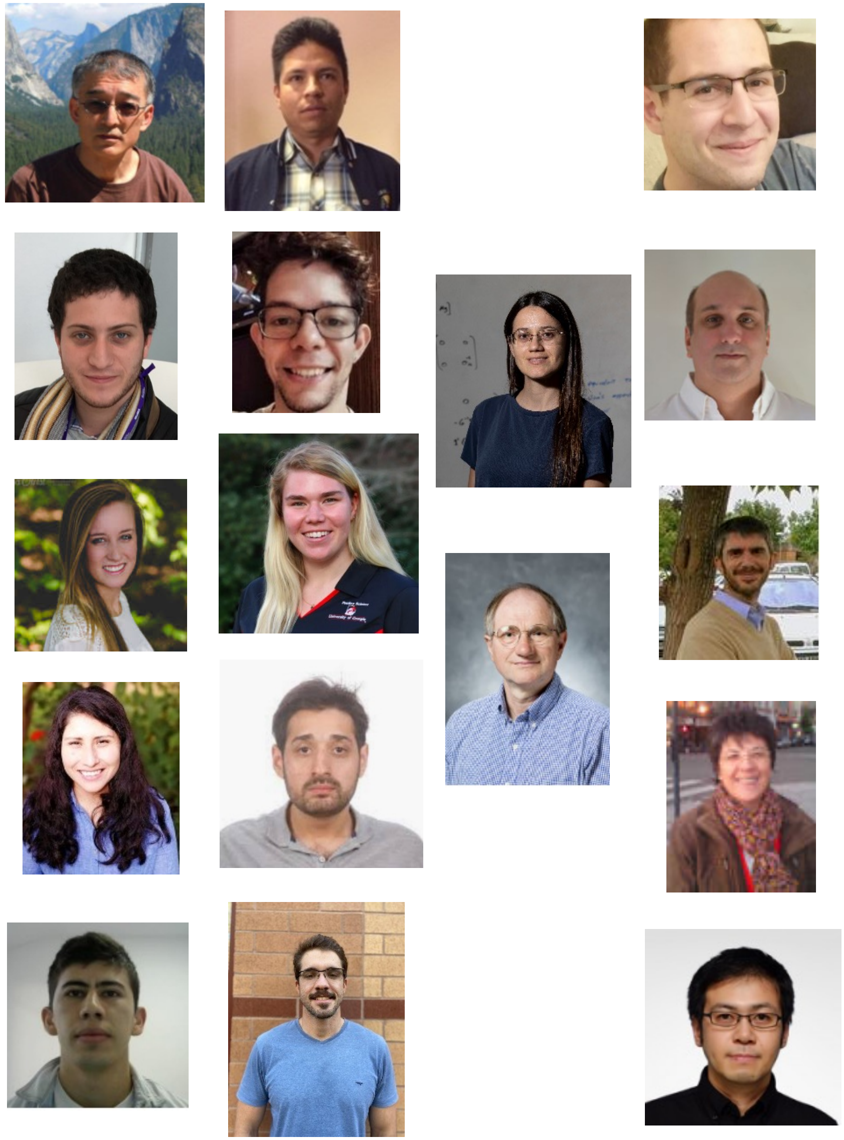
- Different peaks in different lines
- Antagonistic pleiotropy

Conclusions

- GWAS in farm animals affected by small effective population size
- Optimal window size 1-2 Mb for Ne=100
- Large signals in GWAS due to QTN, relationships and noise (incl. Imputation)
- Large QTL show pleiotropy – QTL not visible in index
- GWAS by single-step GBLUP for any data size with option for p-values

UNIVERSITY OF GEORGIA

UGA AB&G team



USDA

NIFA

PIC

ANGUS THE BUSINESS BREED
AMERICAN ANGUS ASSOCIATION
3201 Frederick Avenue • St. Joseph, MO 64506 • (816) 383-5100 • Fax (816) 233-9703 • E-mail: angus@angus.org

AGI

USDA United States Department of Agriculture
Agricultural Research Service
Warmwater Aquaculture Research Unit

zoetis

Smithfield®

USDA Agricultural Research Service
U.S. DEPARTMENT OF AGRICULTURE
Cool and Cold Water Aquaculture Research

Cobb®

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

DNA GENETICS
Greater gains. Improved performance.

ABCZ

Rural Development Administration
National Institute of Animal Science

WAGYU
AMERICAN WAGYU ASSOCIATION

BAYER Crop Science

THE MASCHHOFFS
Progressive Farming. Family Style.

ABS®

SEMEX®
Genetics for Life®

U.S. Registered Holsteins
FOR MAXIMUM PROFIT
HOLSTEIN ASSOCIATION USA, INC.

PEAK

PFHBiPM

PLEMDAT

Topigs Norsvin

methodgenetics

ANCP