

# Selected projects at University of Georgia

Ignacy Misztal and Daniela Lourenco

University of Georgia



# Recent projects

- Blupf90 software
  - Convergence improvements
    - Case of #phenotypes  $\ll$  # animals
  - Multiple categorical traits with large data
  - P-values in GWAS with national data sets
- Applications
  - Implementation in dairy
  - Canalization for disease resistance
- Potential negative effects of genomic selection
- Parameter estimation with large data
- Explaining peculiarities of GWAS

# What Manhattan plots show?

## QTLs and sequence data

Ignacy Misztal, Ivan Pocrnic\*, Daniela Lourenco

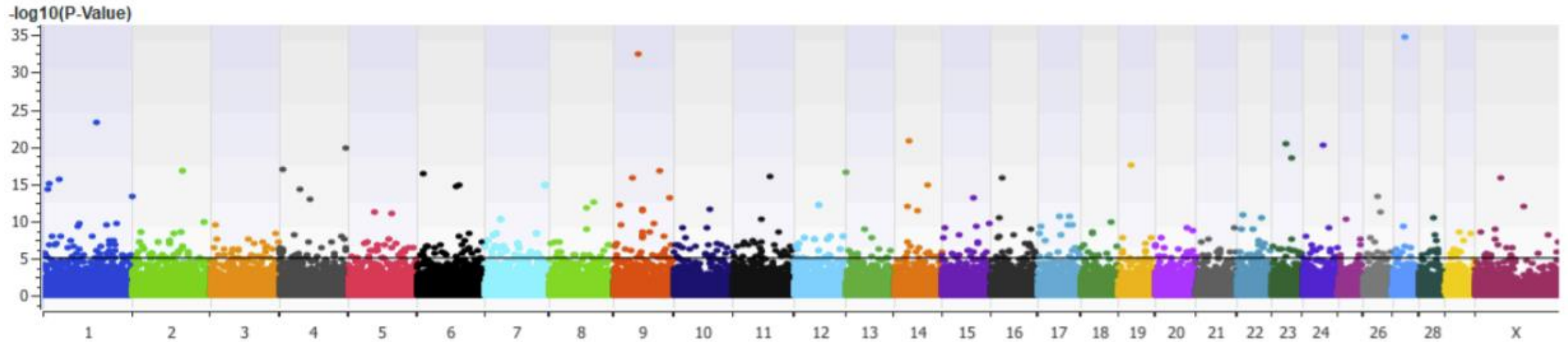
University of Georgia

\*now Roslin Institute

# Peculiarities of QTL detection - GWAS

- Lots of QTLs detected with small data sets
- Fewer QTLs detected with large data
- GWAS by
  - % of variance explained usually per 1Mb
  - p-values
- Few regions explain  $> 1\%$  additive variance
- Little or no gain with sequence data for ssGBLUP with commercial data

# First conception rate on 2k Holstein heifers



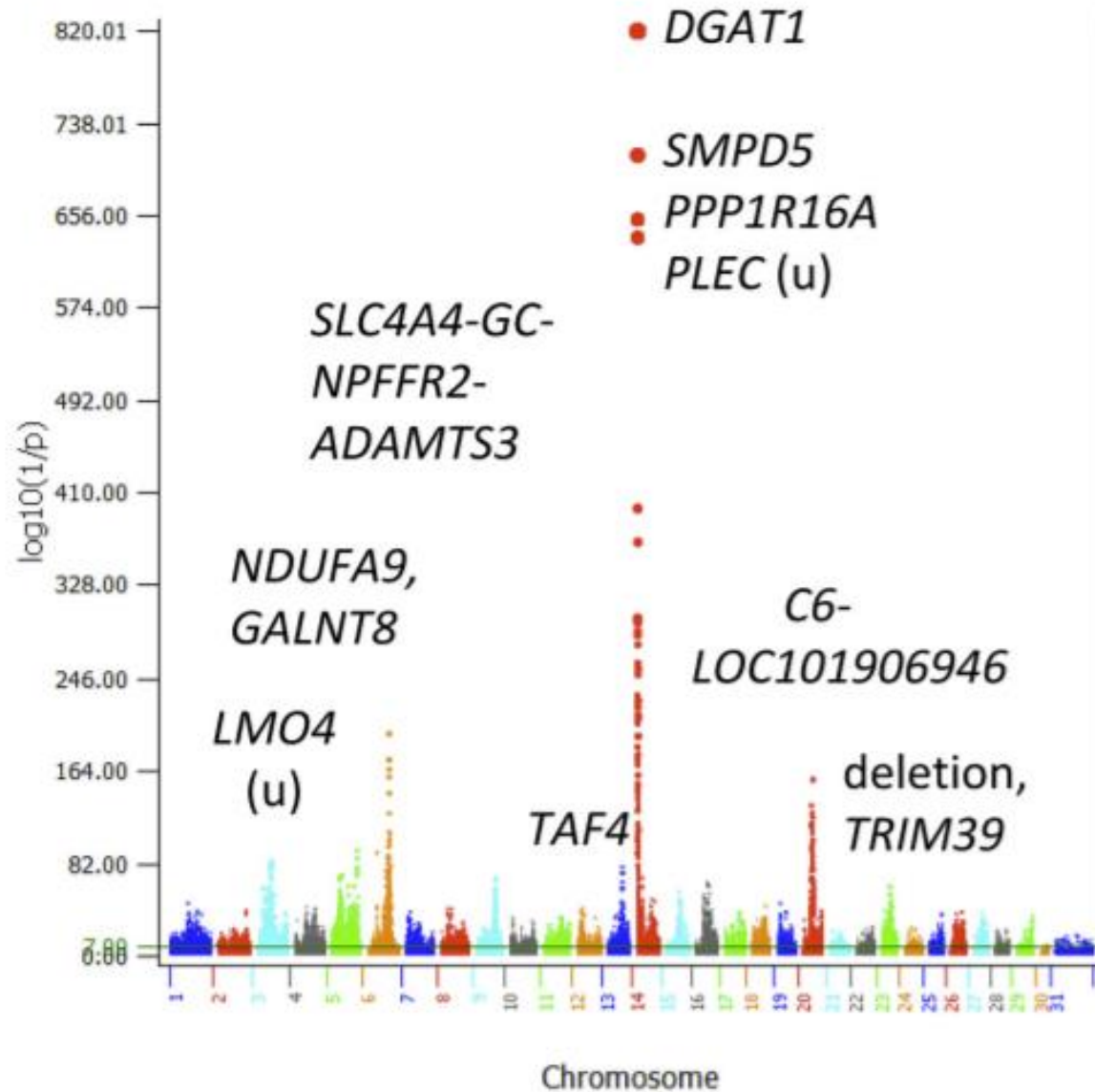
Estimated heritability 36% (normally 1%)

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

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

# GWAS on 294k Holstein cows

## A Milk yield: additive effect



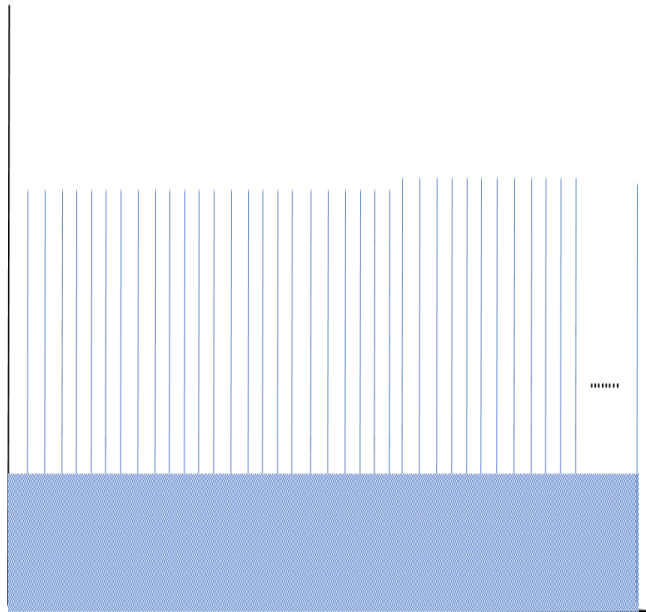
Jiang et al., 2019

doi: 10.3389/fgene.2019.00412

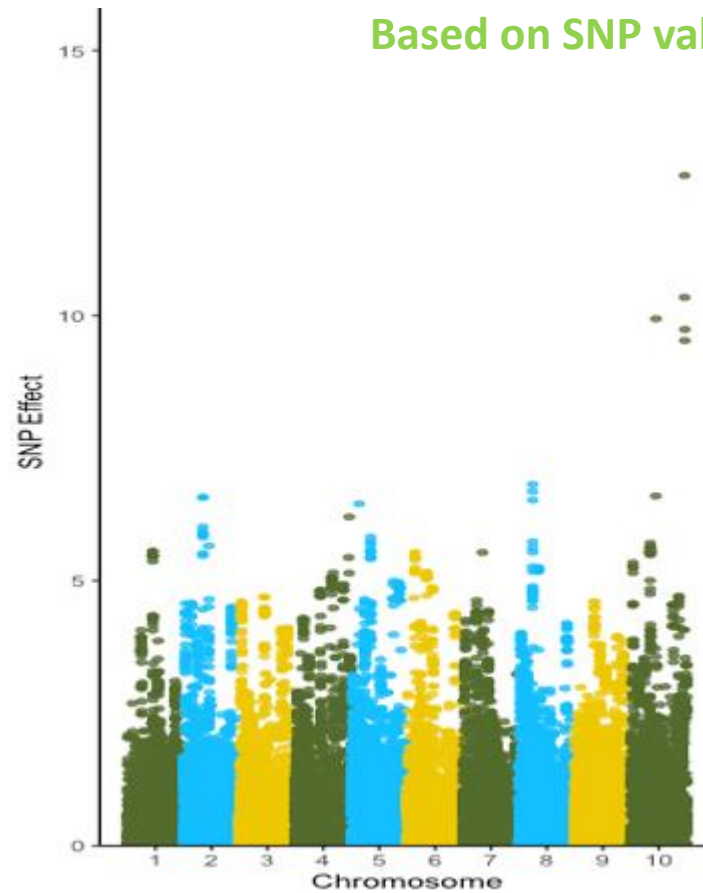
# Manhattan plots for simulated population with 100 identical equidistant QTNs



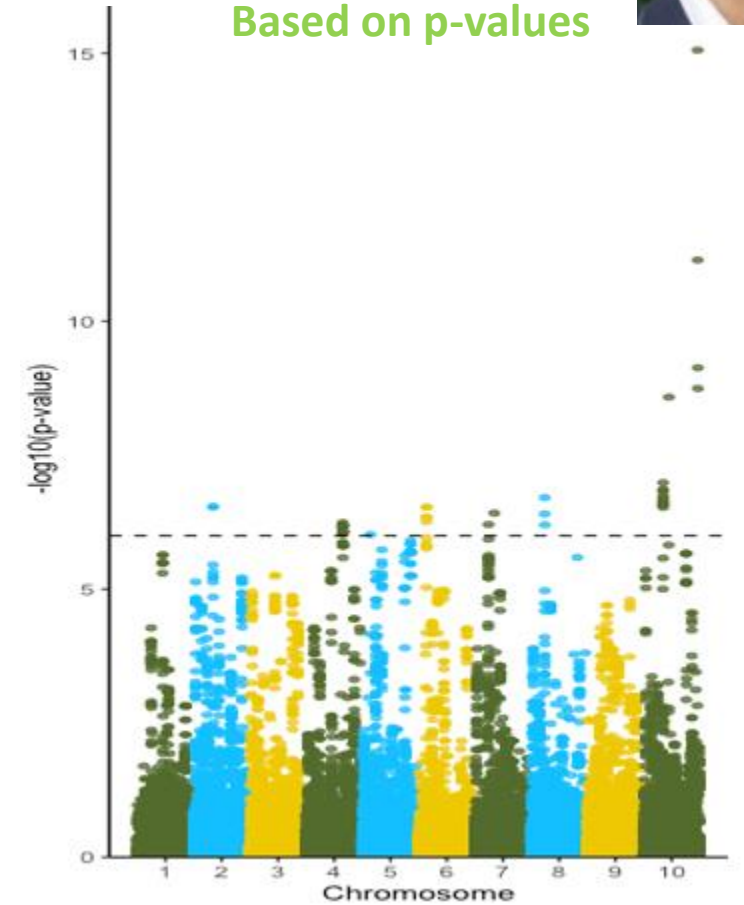
Expectation



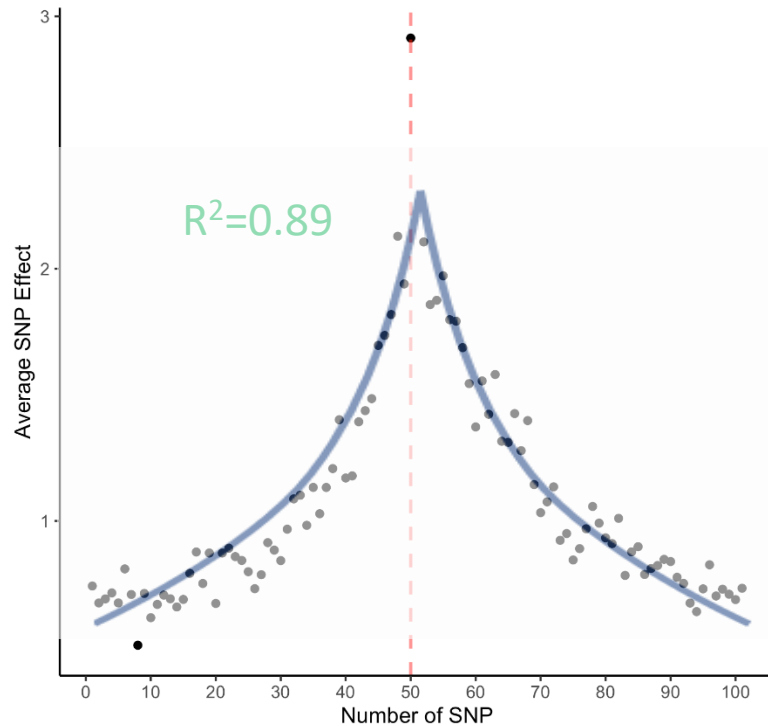
Based on SNP values



Based on p-values



# Plots averaged for 100 QTN



Pairwise linkage disequilibrium curve

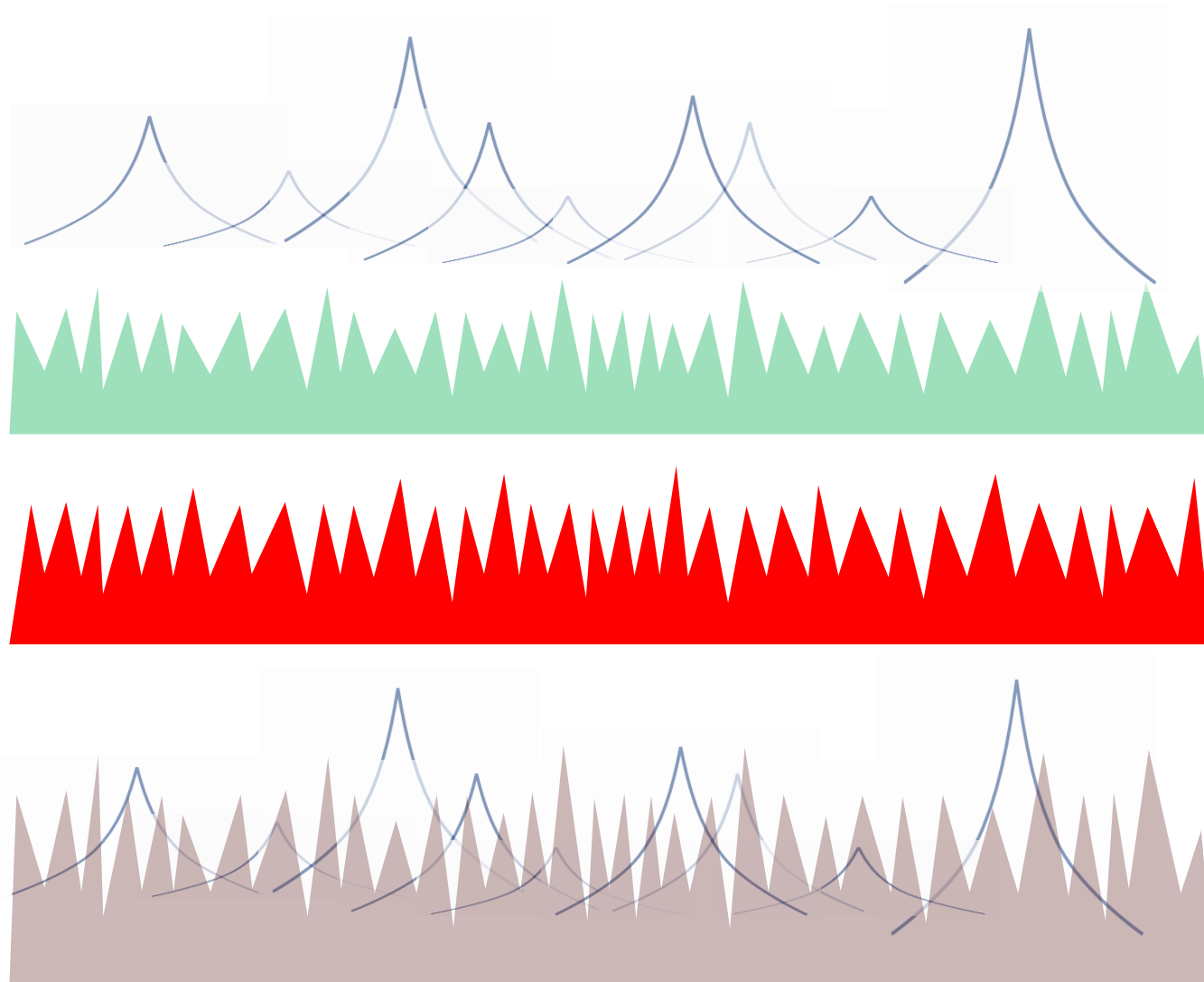


~ 2 Mb for cattle  
~ 5 Mb for pigs/chickens  
~ 15 kb for humans

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



# What is Manhattan plot composed of?



QTNs

**Bigger with larger QTN  
and larger data**

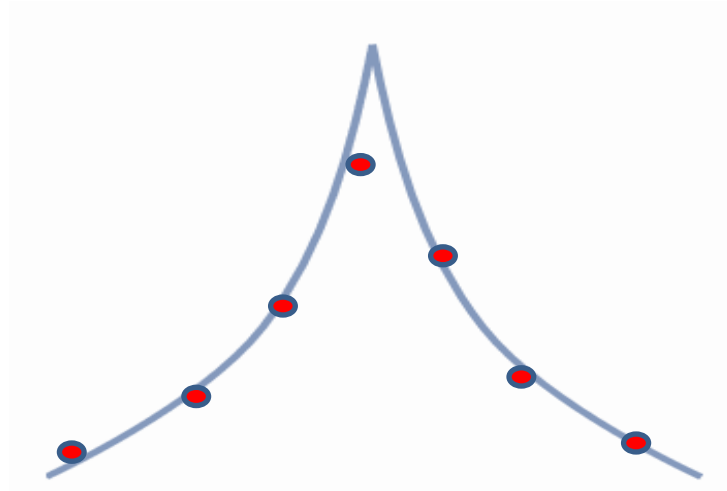
Relationships

Noise

**Smaller with more data**

Combined

# Why ssGBLUP accounts for QTN?



SNPs cover QTN LD curve

# P-values for GWAS in (ss)GBLUP

$$pval_i = 2 \left( 1 - \Phi \left( \left| \frac{s\hat{n}p_i}{sd(s\hat{n}p_i)} \right| \right) \right) \quad (\text{Chen et al., 2017})$$

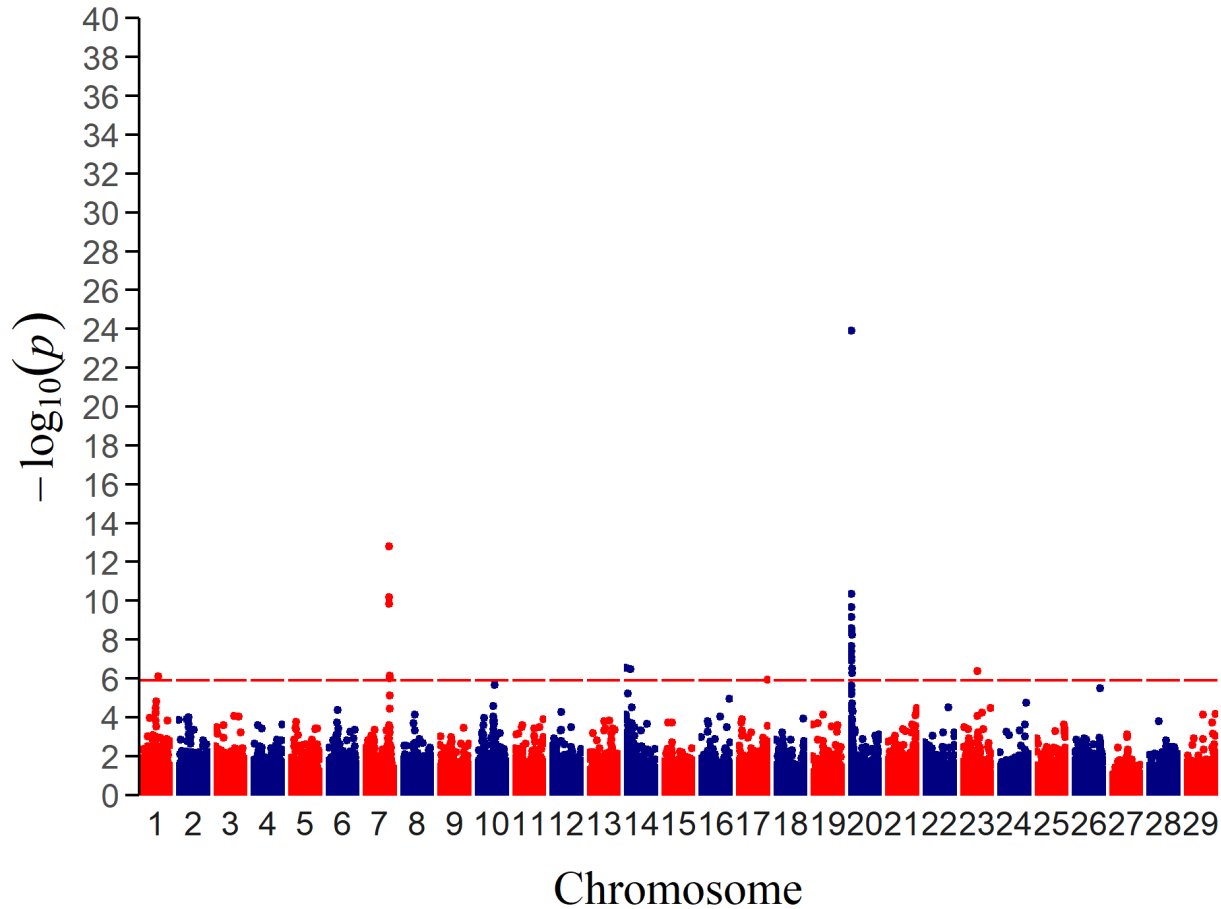
If  $sd(s\hat{n}p_i)$  approximately constant, Manhattan plots based on  $|s\hat{n}p_i|$  and  $pval_i$  similar

Large data – PEV from accuracy approximations (Bermann et al., 2021)

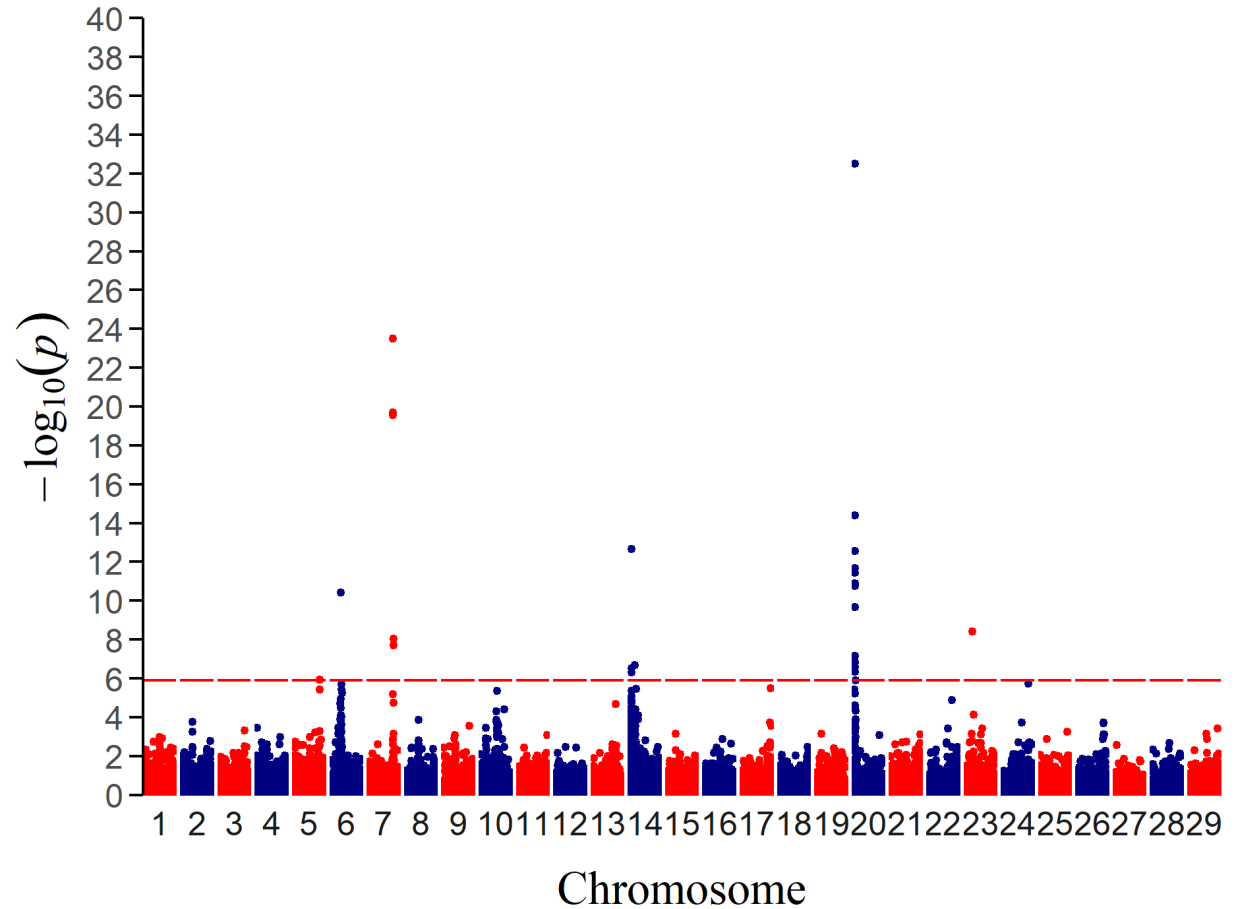
# Post-weaning gain in American Angus



## 50k genotyped animals



## 500k genotyped animals

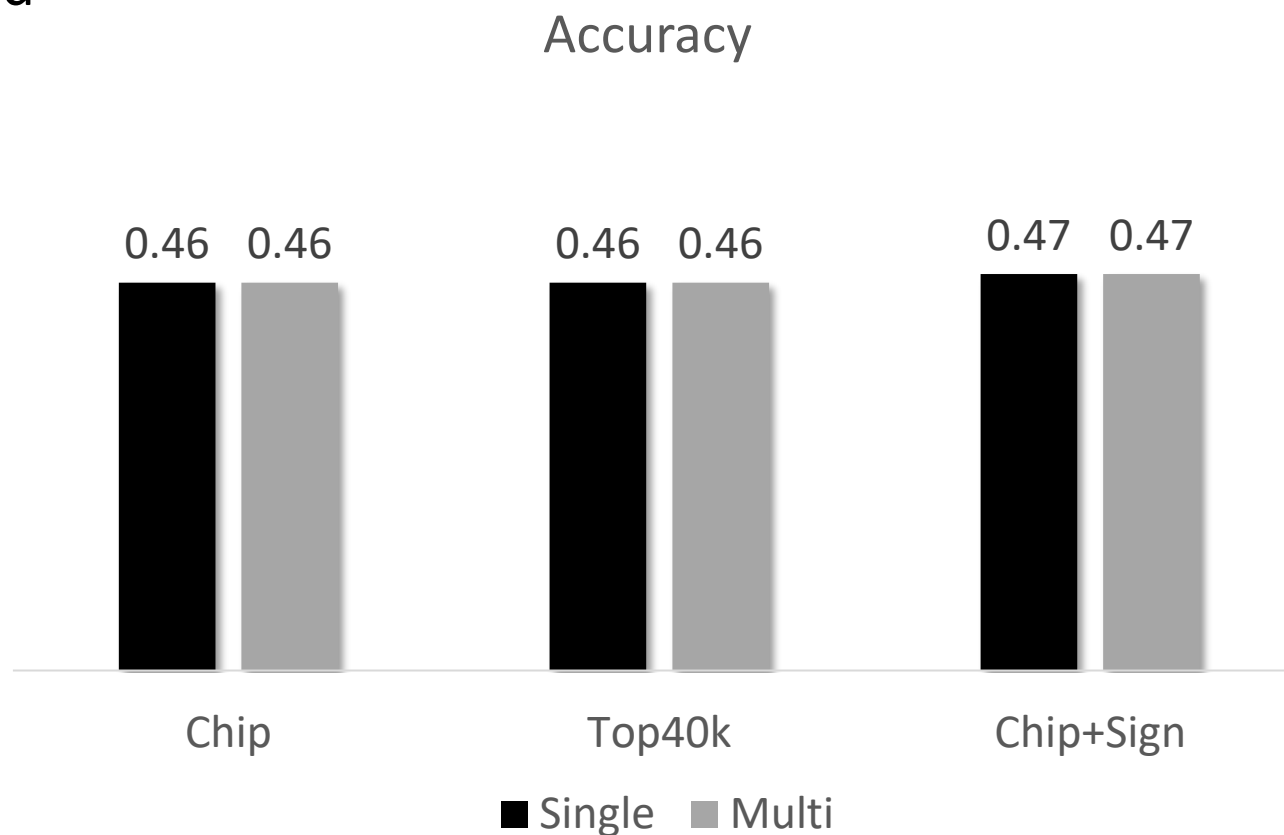


# Sequence data

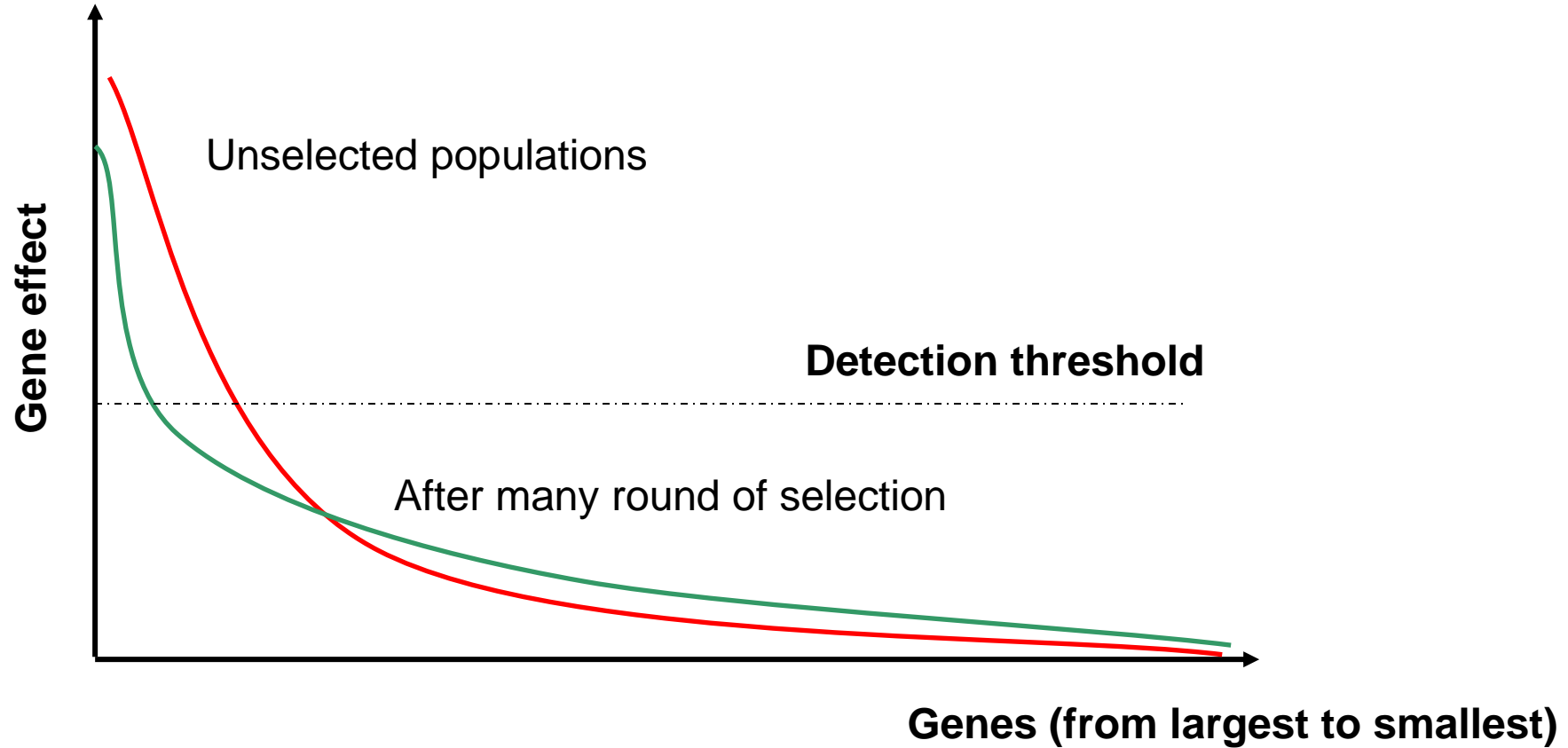
- 207k pigs with sequence
- 5M pedigree
- 1.5M records
- Single and multi-breed



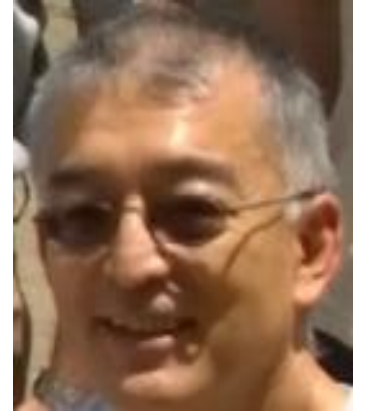
Jang et al., 2023



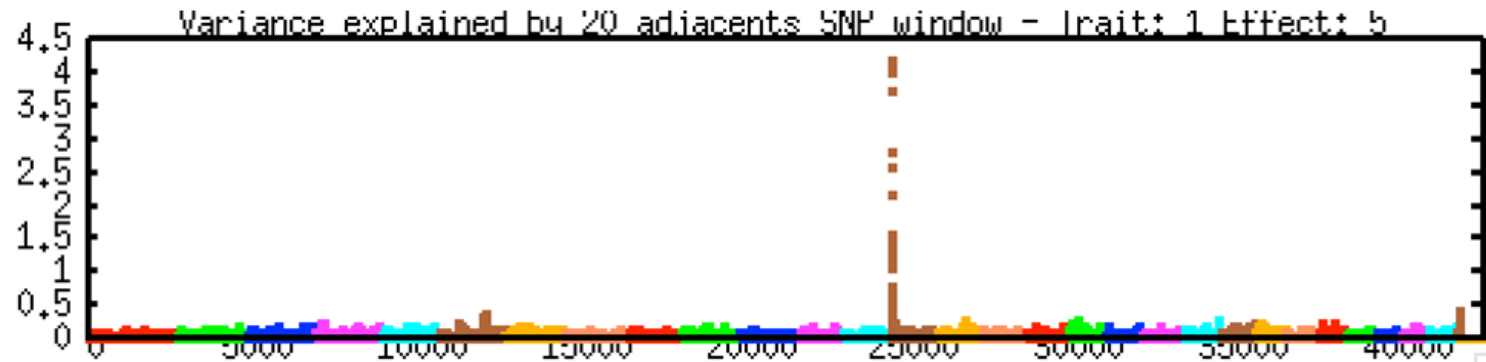
# Distribution of QTL effects



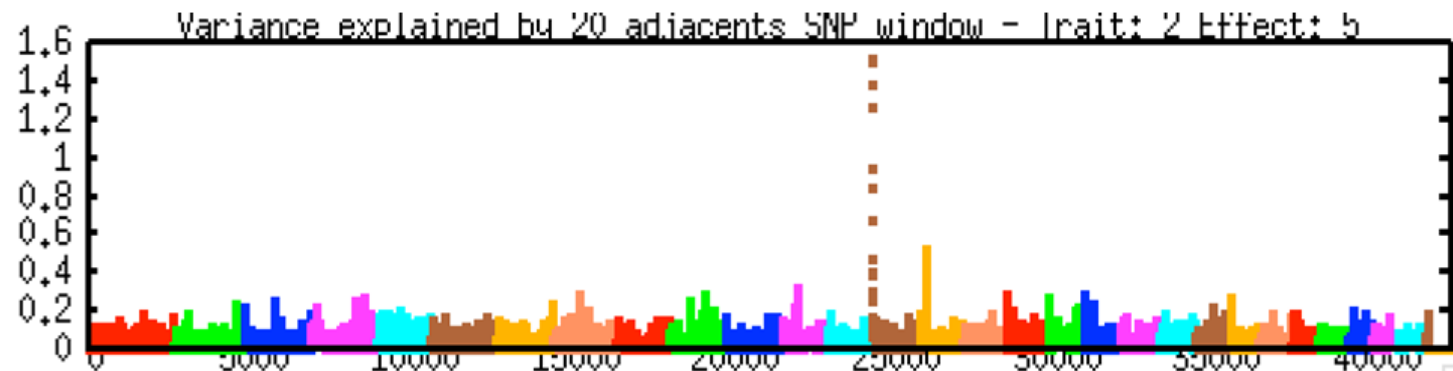
# GWAS using 35k Holstein bulls



## Milk – first parity



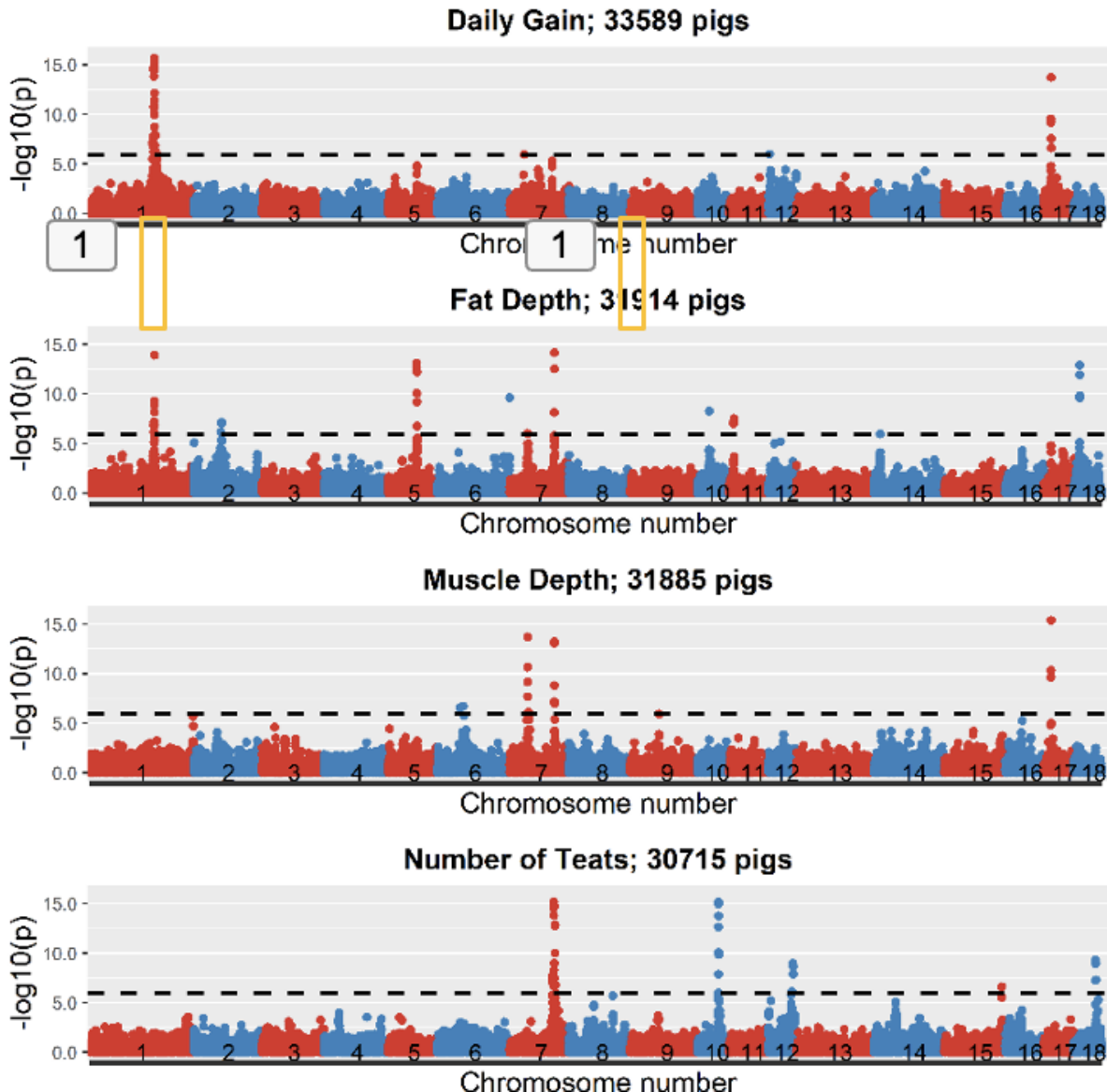
## Mortality – first parity



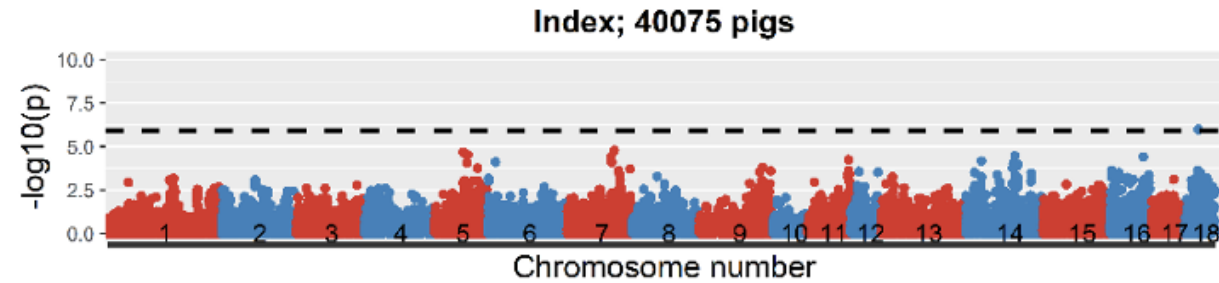
(Tokuhsa et al, 2014;  
Tsuruta et al., 2014)

# GWAS for various traits and index in pigs

Bijma, EAAP 23



## Index



- Different peaks in different lines
- Antagonistic pleiotropy



# Conclusions for GWAS

- QTN profile wide with small effective population size
- Large signals in GWAS due to QTN, relationships and noise (incl. Imputation)
- Most QTNs probably below detection limit
- ssGBLUP accounts for QTL with large data

# Positive and Negative Impacts of Genomic Selection

**Ignacy Misztal and Daniela Lourenco**

University of Georgia

# Negative effects of genomic selection

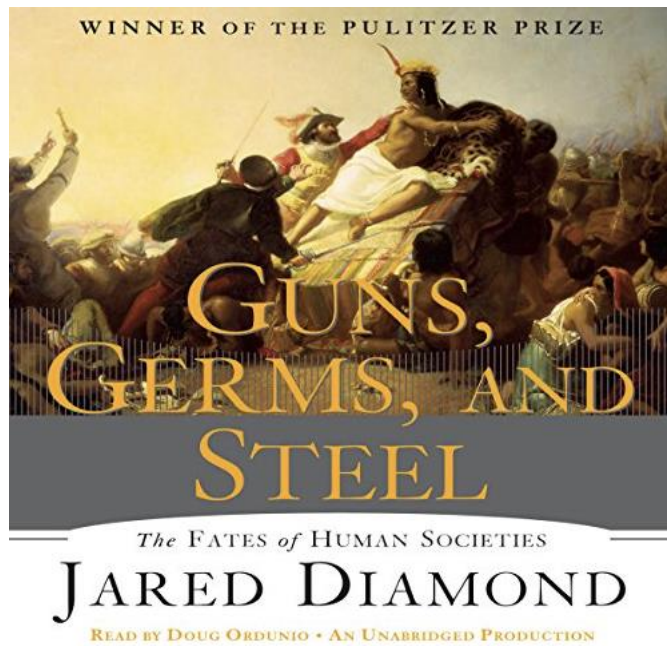
- Informal industry reports:
  - Deteriorating sow survival and pig mortality in pigs
  - Deteriorating feet & legs in beef
  - Short teats and increased calf mortality in dairy
  - Increased sensitivity to heat stress in dairy
  - Deteriorating disease resistance across species
- Recessive genes or pleiotropy?

# Genetic selection as optimization

- Selection for one trait or an index
- Gains on selected traits
- Losses on correlated antagonistic traits
- Losses compensated by improved environment/management

# History of selection strategies

- Domestication
- Unformal
- Large-scale single-trait for production traits
- Multi-trait with fitness traits
- Genomic





Livestock Production Science

Volume 93, Issue 1, 1 April 2005, Pages 3-14



Go to Livestock Production Science on ScienceDirect

## Genetics of adaptation and domestication in livestock ☆

[Sandrine Mignon-Grasteau](#)<sup>a</sup>  , [Alain Boissy](#)<sup>b</sup>, [Jacques Bouix](#)<sup>c</sup>,  
[Jean-Michel Faure](#)<sup>a</sup>, [Andrew D. Fisher](#)<sup>d</sup>, [Geoffrey N. Hinch](#)<sup>e</sup>, [Per Jensen](#)<sup>f</sup>,  
[Pierre Le Neindre](#)<sup>b</sup>, [Pierre Mormède](#)<sup>g</sup>, [Patrick Prunet](#)<sup>h</sup>, [Marc Vandeputte](#)<sup>i</sup>,  
[Catherine Beaumont](#)<sup>a</sup>

# Domestication

## Winners

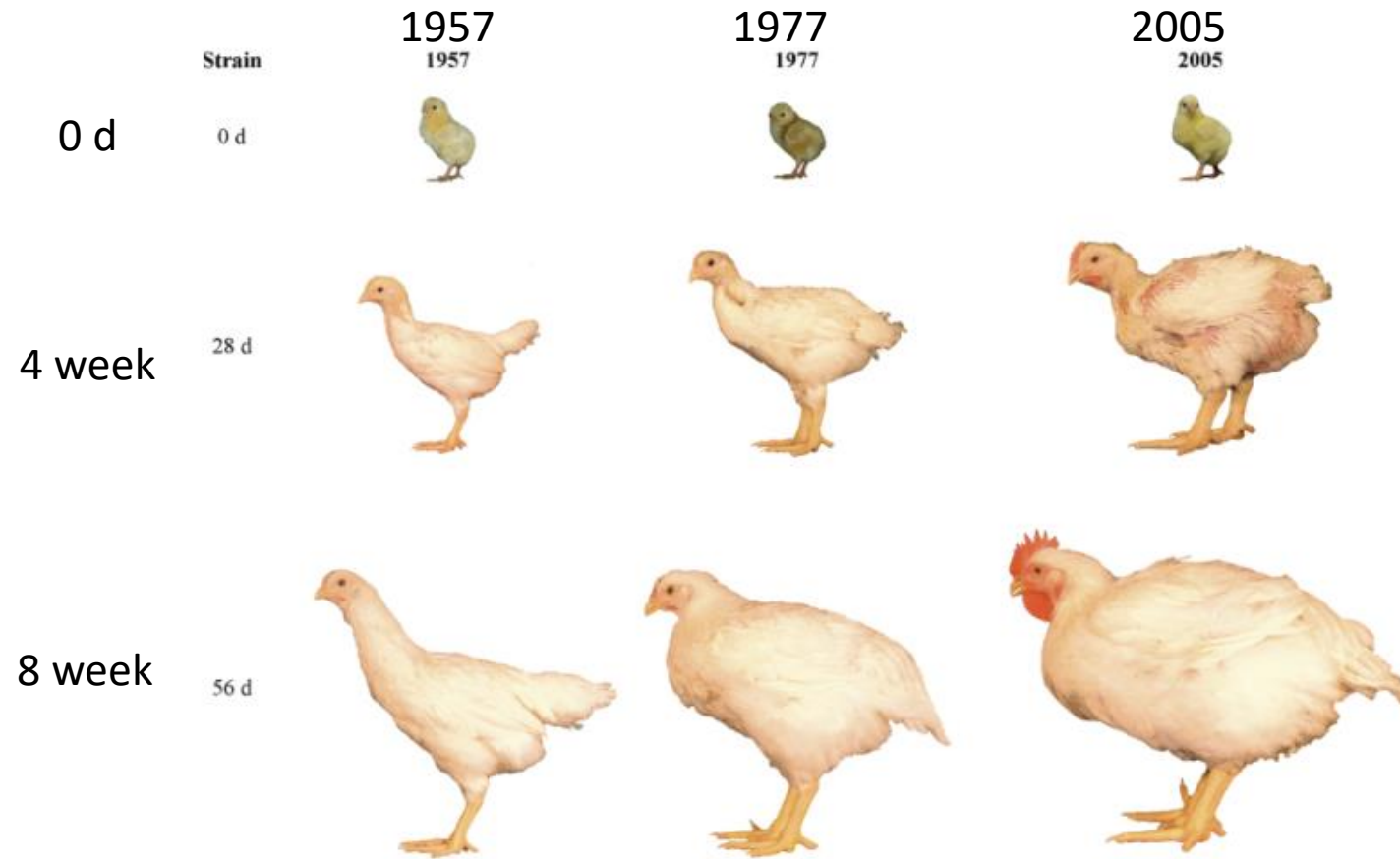
Growth  
Milk  
Mating procedures

## Losers

Food finding  
Seasonal reproduction  
Predator avoidance  
Brain size

...

# Example of effects of mostly single-trait selection



# Side effects of intensive selection for growth in broiler chicken

- Unlimited appetite / obesity → artificial lighting
- Poor survival of males → male supplementation
- Increased susceptibility to diseases → antibiotics
- Low hatchability → alternate heating/cooling of incubators
- ...

All companies – similar problems at same time

Initially problems kept confidential



# Undesirable side effects of selection for high production efficiency in farm animals: a review

W.M. Rauw<sup>a,\*</sup>, E. Kanis<sup>b</sup>, E.N. Noordhuizen-Stassen<sup>c</sup>, F.J. Grommers<sup>c</sup>

<sup>a</sup>*Department of Animal Science, Agricultural University of Norway, P.O. Box 5025, 1432 Ås, Norway*

<sup>b</sup>*Animal Breeding and Genetics Group, Wageningen Institute of Animal Science, Wageningen Agricultural University, P.O. Box 338, 6700 AH Wageningen, The Netherlands*

<sup>c</sup>*Department of Herd Health and Reproduction and Interdepartmental Section Veterinary Medicine and Society, University of Utrecht, P.O. Box 80151, 3508 TD Utrecht, The Netherlands*

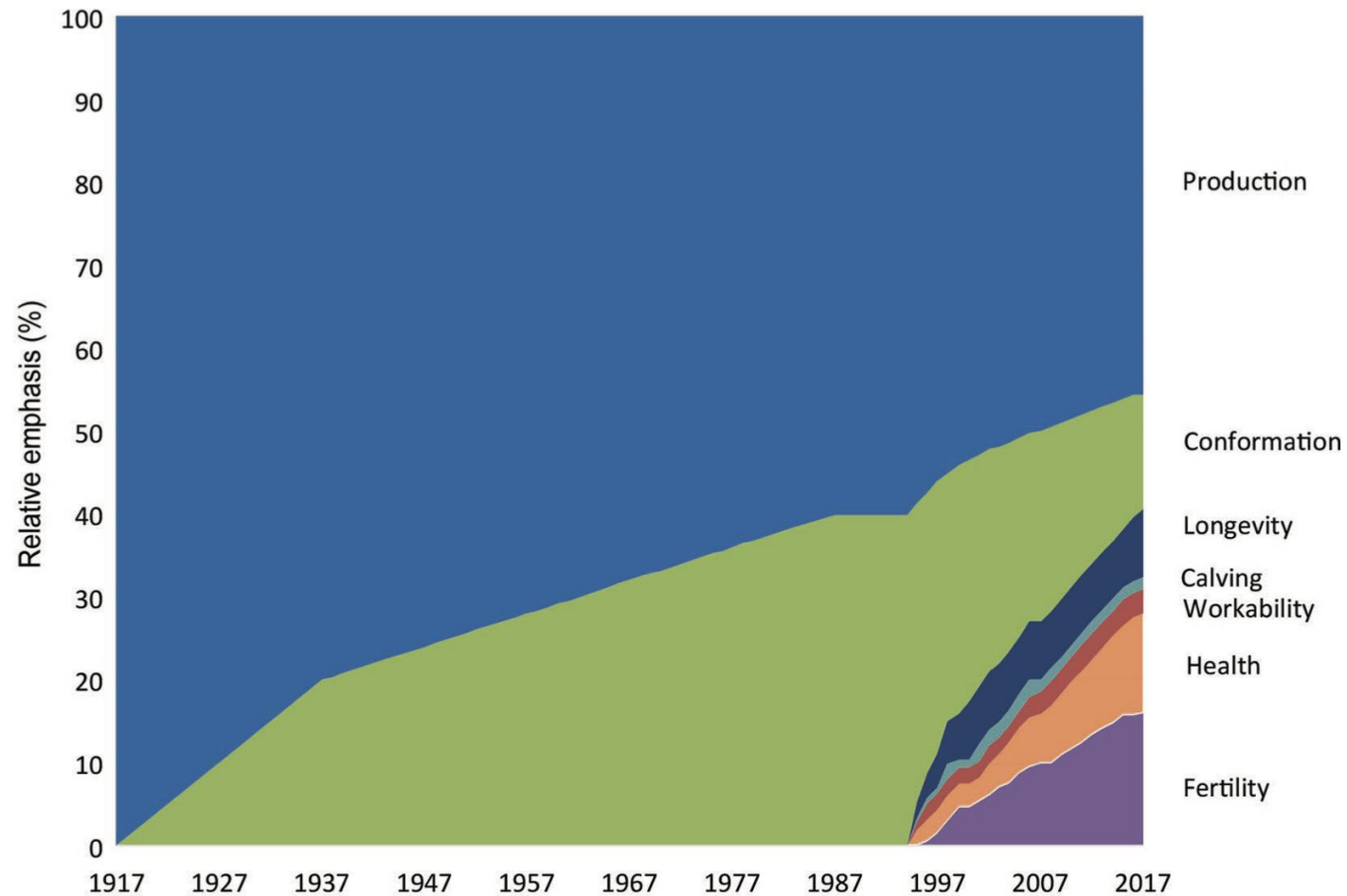
Received 4 July 1997; accepted 29 April 1998

...over 100 references on undesirable(cor)related effects of selection ... in broilers, pigs and dairy cattle....

Future application ... DNA-techniques .. ....more dramatic consequences....

Selection for more than production traits alone may prevent such.

# Changes in US dairy index

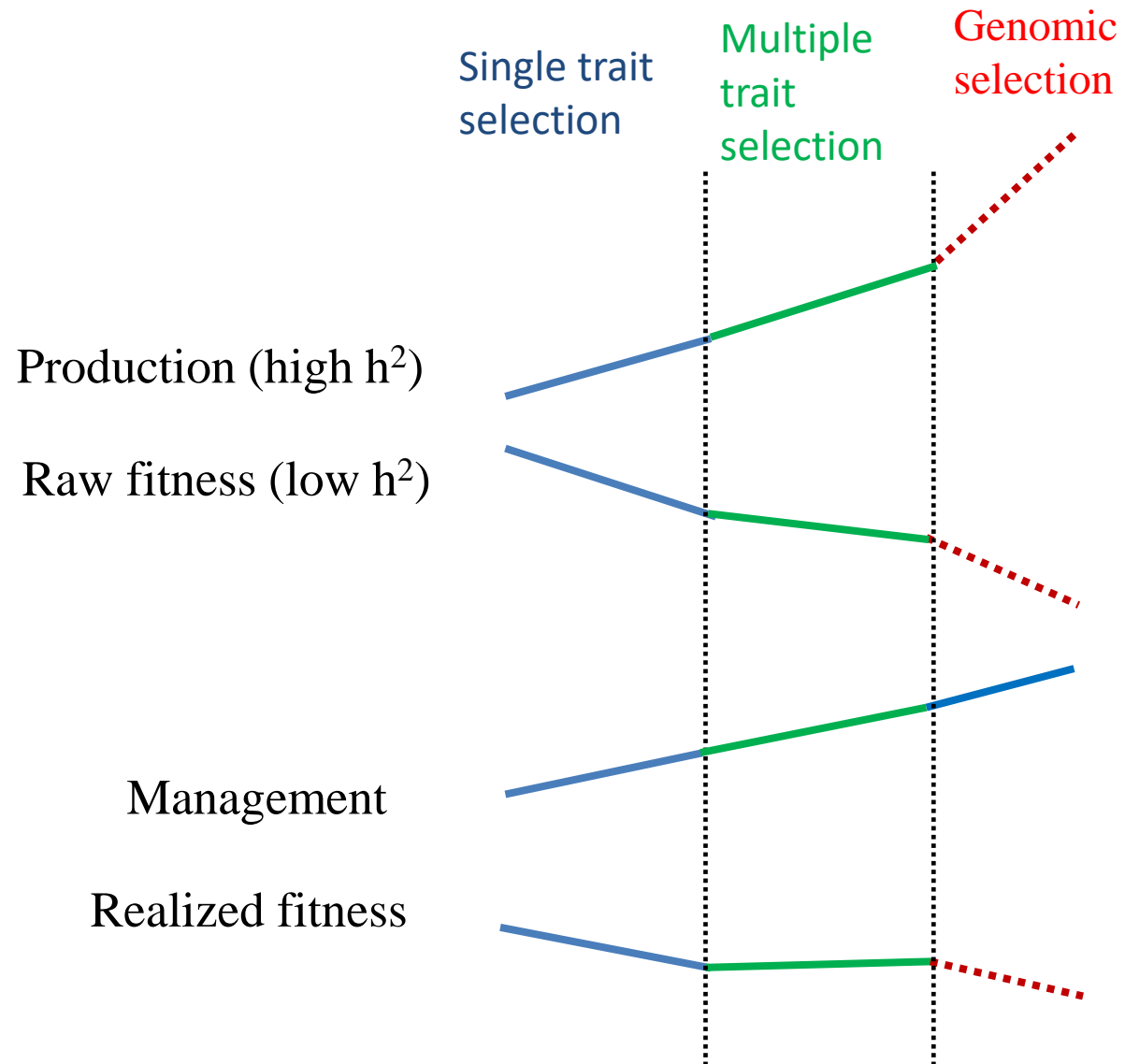


Miglior et al., 2017

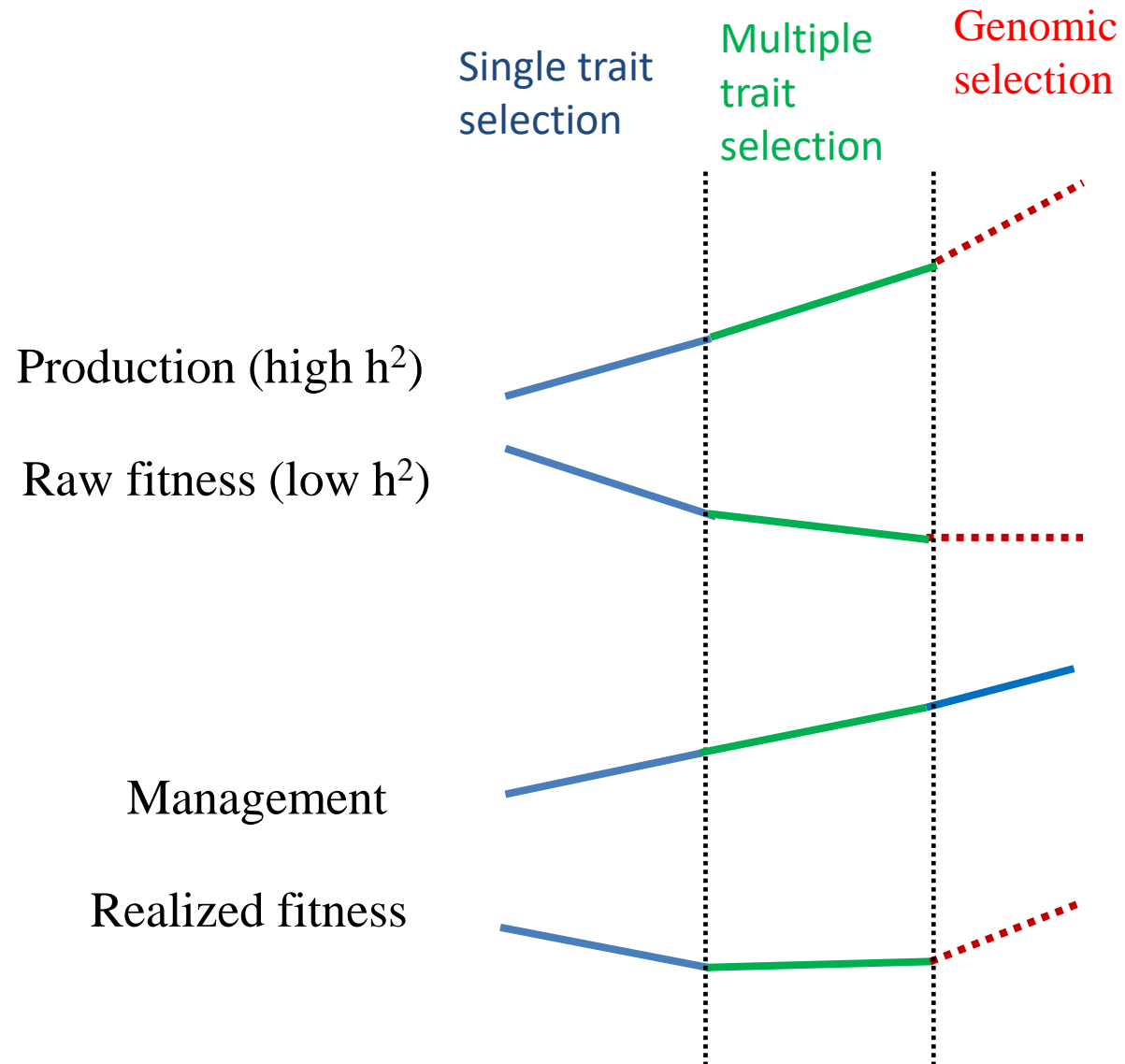
# Effects of genomic selection

- High accuracy for well recorded traits
- Low accuracy for low  $h^2$  traits with little information
- Lowered generation interval
  
- Acceleration of trends for selected traits
- Acceleration of correlated responses
  
- Changes in genetic parameters

# Hypothetical trend changes in 3 stages of genetic selection



# ... large data for fitness traits ...



# Selection and resource allocation theory

- Thresholds of minimum energy required for fitness, environment dependent (Van der Waaij, 2004; Rauw, 2012; Knapp, 2014)
- More energy for production → fitness more antagonistic
  - Genetic correlations → -1
  - $h^2$  of selection index decline
- Fast selection → unbalanced animals (Huber, 2015)

# Problems and species

- Need large data for accurate evaluations
- Small fraction of fertility to performance records in some species
  - ~ 1 in cattle
  - 1/15 in pigs
  - 1/200 in layers
- Problems with early mortality/morbidity when affected animals not genotyped

# Genomic gain for production and fitness traits – example in pigs

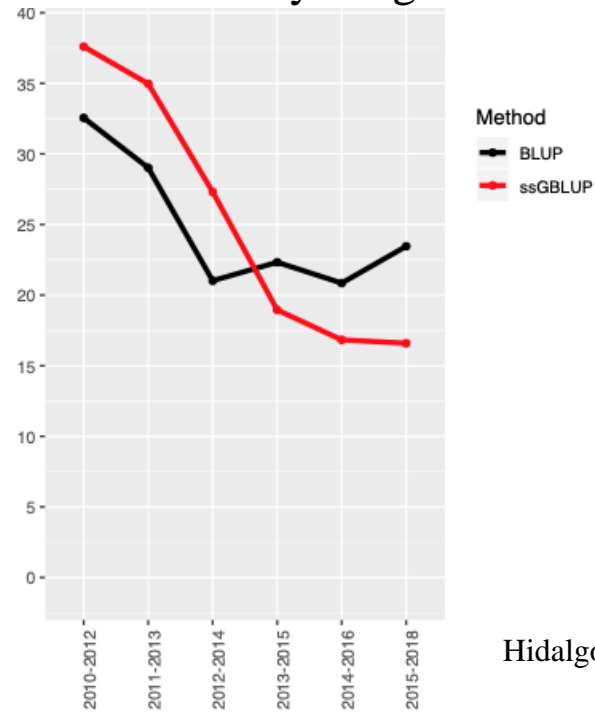
- 1000 sows per generation
- 15 piglets per sow
- 4 generations
  
- Gain per generation:
  - 0.65 phenotypic SD for growth
  - 0.02 phenotypic SD for number of born dead
  
- Genomic favors bigger populations with better recording



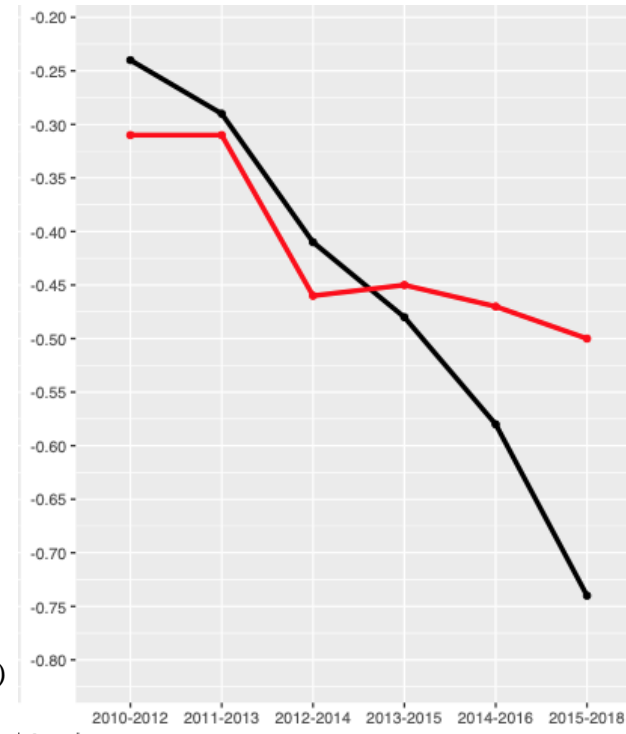
# Changes in (co)variances in pigs due to genomic selection



### Heritability for growth



### Genetic correlation with reproduction



Hidalgo et al. (2019)

Heritability halved, antagonistic correlations  $-0.3 \rightarrow -0.5$

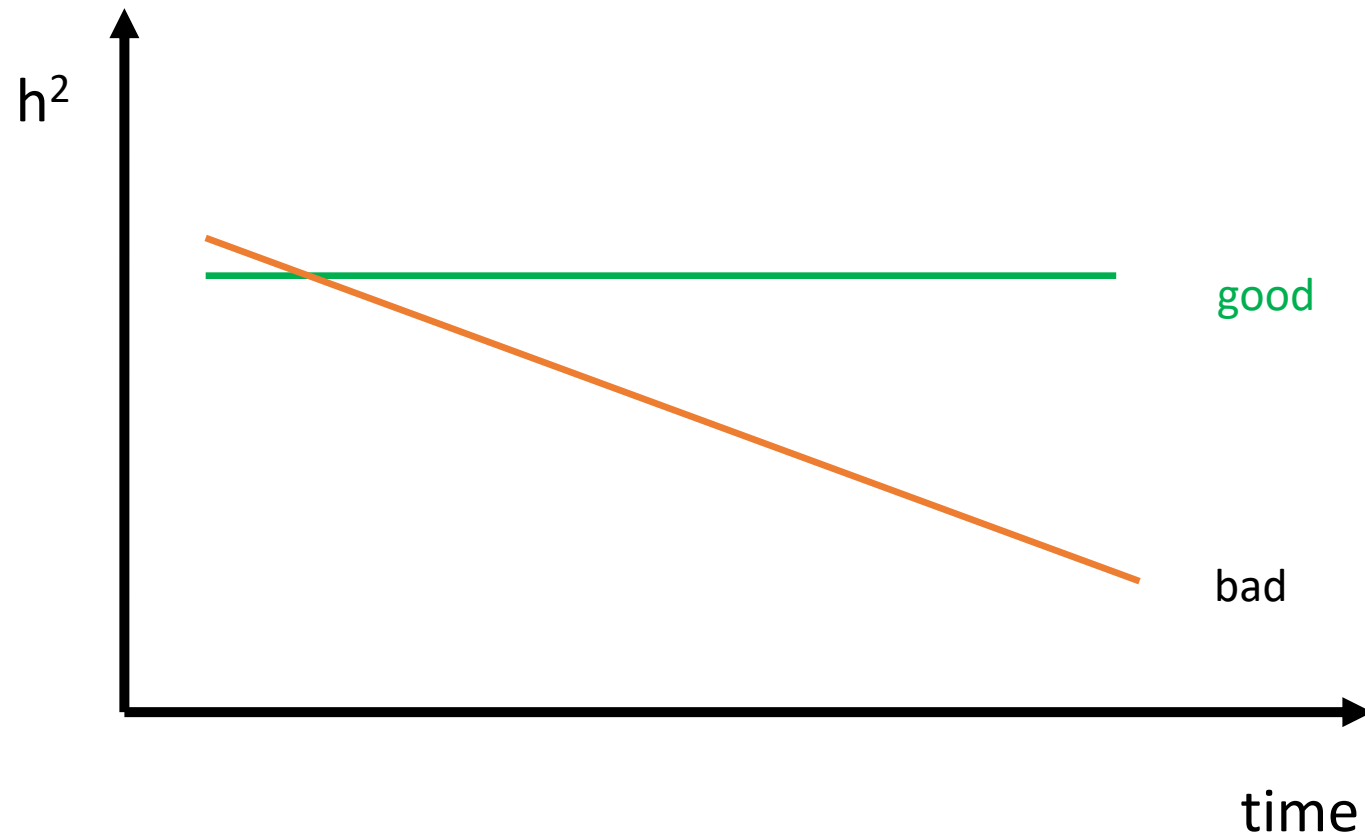
# Why changes in genetic parameters?

- Bulmer effect
- Changing resource allocation
  
- Changes in gene frequencies
- Changes in trait definitions
- G x E
- Recessives
- ...

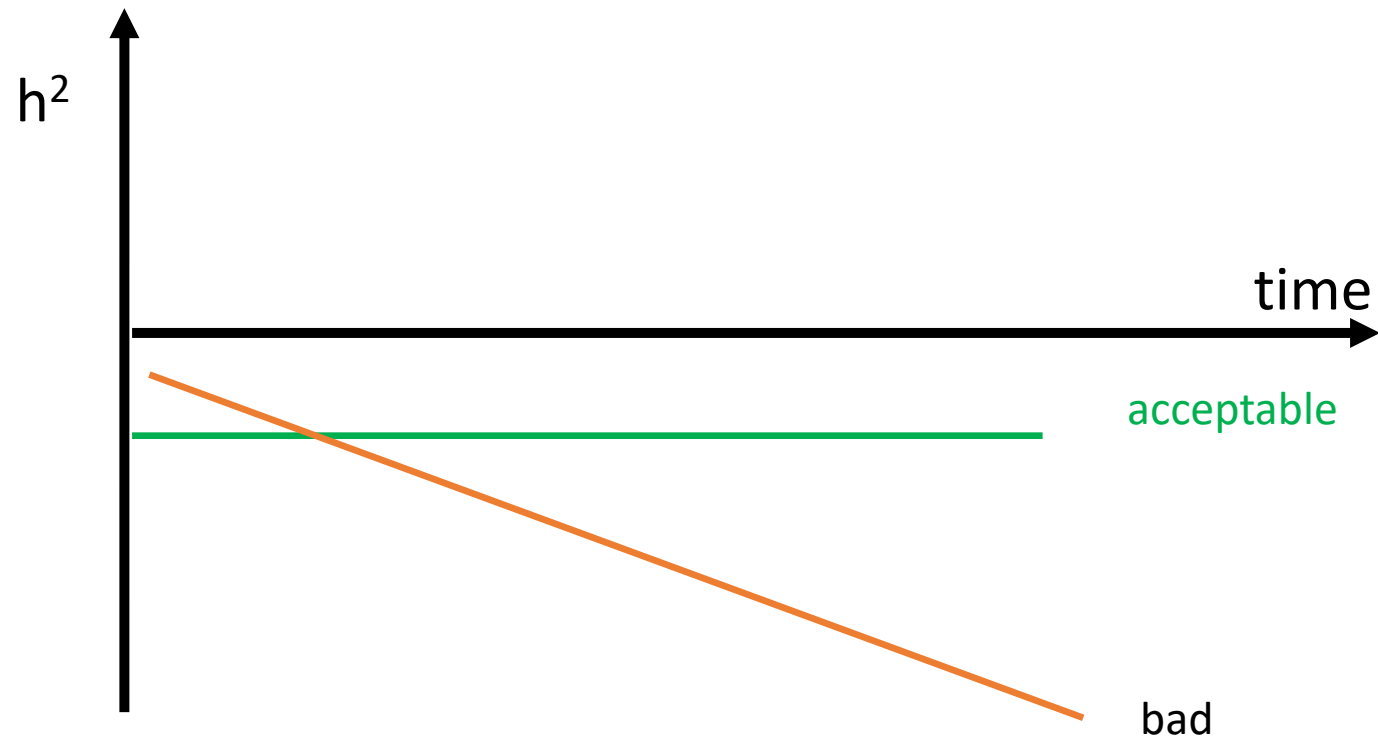
# How to circumvent negative effects?

- Start or expand recording for problematic traits
- Have a realistic index
  - Needs estimates for last generation
- Focus on traits where the parameters are changing rapidly
  - Needs estimates for last generation

# Possible changes in heritability



# Possible changes in genetic correlations



# Using theoretical and realized accuracies to estimate changes in heritabilities and genetic correlations

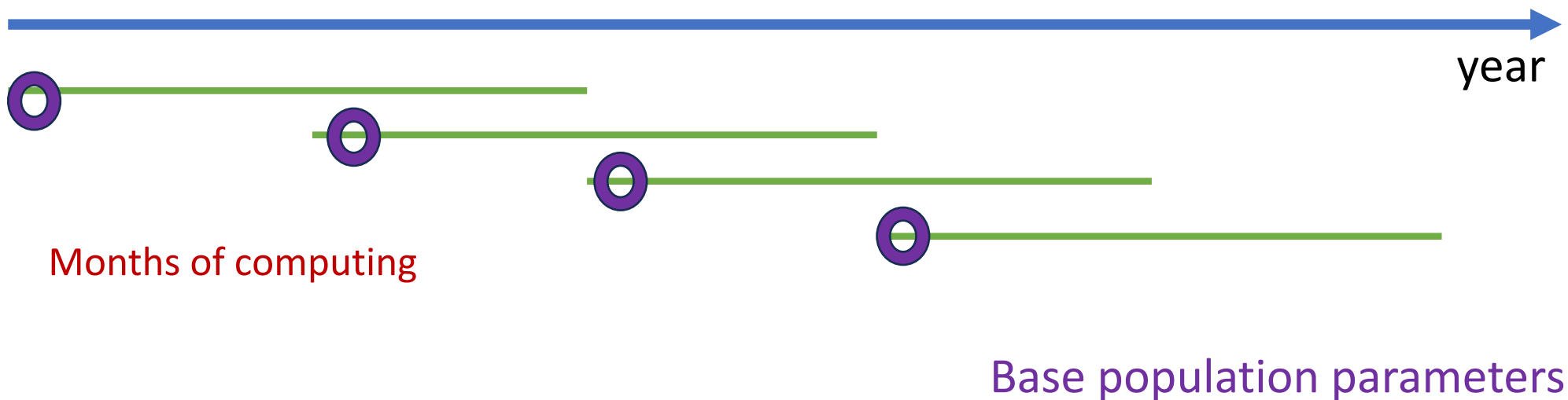
**Ignacy Misztal**

**University of Georgia**

# How to estimate parameter change by generation?

- If REML/Gibbs sampling
  - Base population parameters
  - Impossible computations with too many genotypes
  - Biases with too few generations (Cesarani et al., 2019)

Data sampling in Hidalgo et al.(2019)



# Requirements for new estimation

- All data including genomic
- Parameters by last generation/ year
- Reasonable computing cost



Hours of computing

Last generation parameters



# Realized and theoretical accuracies

Realized accuracy  $acc = corr(y - Xb, \hat{u})/h$

Legarra et al. (2008)

$y - Xb$  - adjusted phenotype

$\hat{u}$  - breeding value obtained without  
that phenotype

$h^2$  - heritability

Theoretical accuracy  $acc = \sqrt{\frac{Nh^2}{Nh^2 + M_e}}$

Daetwyler et al. (2008)

$N$  – number of genotyped animals with phenotypes  
 $M_e$  – number of independent chromosome segments

$M_e \approx 5k$  (chickens, pigs),  $10k$  (beef),  $15k$  (Holsteins)  
Pocrnic et al. (2017)

# Pig data set

150k records on growth ( $h^2=0.21$ )

25k records on fitness ( $h^2=0.05$ )

53k genotyped animals



Hollifield et al., 2021

Theoretical accuracy

$$\sqrt{\frac{53k * \text{growth}}{53k * 0.21 + 5k}} = 0.83$$

$$\sqrt{\frac{25k * \text{fitness}}{25k * 0.05 + 5k}} = 0.44$$

Realized accuracy

0.82

0.41

# Broiler chicken data set

820k phenotypes for growth

$$h^2=0.3$$

150k genotyped

Theoretical accuracy  $\sqrt{\frac{150k * 0.3}{150k * 0.3 + 5k}} = 0.94$

Realized accuracy 0.58

Both accuracy same 0.89 if  $h^2 = 0.13$

Company was using 0.14! (Breen, 2022)



Hidalgo et al., 2021

# Formulas for estimating heritability

$$\widehat{h^2}: \sqrt{\frac{Nh^2}{Nh^2 + M_e}} = \text{corr}(y - Xb, \hat{u})/h$$

$$\widehat{h^2} = \frac{c^2 + \sqrt{c^4 + 4c^2 M_e/N}}{2}, c = \text{corr}(y - Xb, \hat{u})$$

$$SE(\widehat{h^2}) \approx \frac{1}{\sqrt{N_{val}}} \left[ c + \frac{2c^2 + \frac{4M_e}{N}}{\sqrt{c^2 + \frac{4M_e}{N}}} \right] \widehat{h^2} \approx \frac{3c}{\sqrt{N_{val}}}$$

$N$  – # animals in reference    $N_{val}$  – number of animals in validation

# Heritability for milk in Holsteins



J. Dairy Sci. 104:5843–5853

<https://doi.org/10.3168/jds.2020-19789>

© 2021, The Authors. Published by Elsevier Inc. and Fass Inc. on behalf of the American Dairy Science Association®.  
This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Genomic predictions for yield traits in US Holsteins with unknown parent groups

A. Cesarani,<sup>1\*</sup>  Y. Masuda,<sup>1</sup>  S. Tsuruta,<sup>1</sup>  E. L. Nicolazzi,<sup>2</sup> P. M. VanRaden,<sup>3</sup>  D. Lourenco,<sup>1</sup>   
and I. Misztal<sup>1</sup> 

<sup>1</sup>Department of Animal and Dairy Science, University of Georgia, Athens 30602

<sup>2</sup>Council on Dairy Cattle Breeding, Bowie, MD 20716

<sup>3</sup>Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

# animals with phenotypes and genotypes	580k
# animals with validation	381k
Assumed # chromosome segments $M_e$	15k
Initial $h^2$	0.35
Calculated $h^2$	0.33

# How to estimate genetic correlations?

Predictivity for trait i

$$\text{corr}(y_i - Xb_i, \hat{u}_i) = \text{acc}_i h_i$$

What is predictivity from trait i to trait j?

$$\text{corr}(y_i - Xb_i, \hat{u}_j) = ?$$

.....  
.....

$$\text{corr}(y_i - Xb_i, \hat{u}_j) = \text{acc}_j \text{corr}_{ij} h_i$$

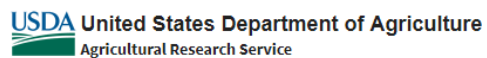
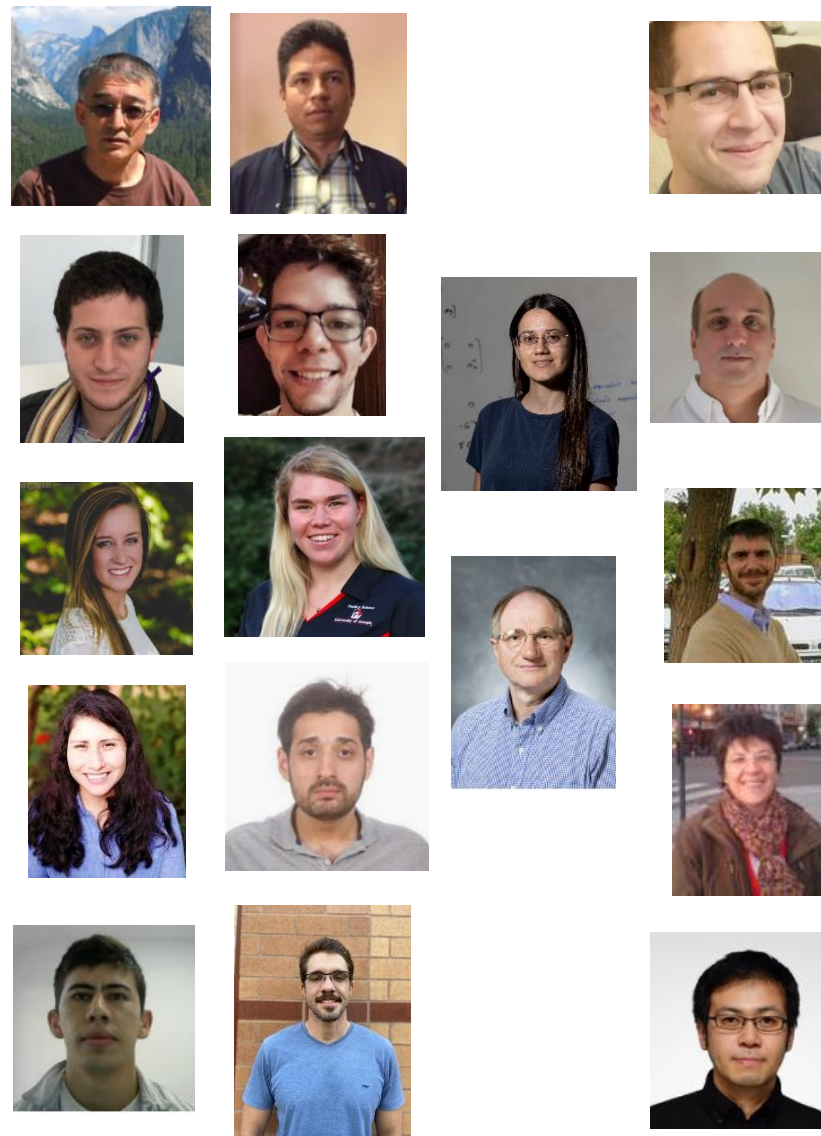
$$\text{corr}_{ij} = \frac{\text{corr}(y_i - Xb_i, \hat{u}_j)}{h_i \text{acc}_j}$$

$$SD(\text{corr}_{ij}) \approx \frac{1}{h_i \text{acc}_j \sqrt{N_{val}}}$$

# Conclusions

- Potential negative effects of genomic selection on fitness traits
  - faster correlated responses
  - Potentially increased antagonism
- Need new methods to estimate genetic parameters – use of predictivity promising
- Response to QTL wide for pigs & chickens– several Mb
  - Probably false QTL if no LD trail
  - ssGBLUP accounts for QTL with large data
- “Good” large QTLs probably fixed, remaining show pleiotropy

# UGA AB&G team



Warmwater Aquaculture Research Unit



Cool and Cold Water Aquaculture Research

