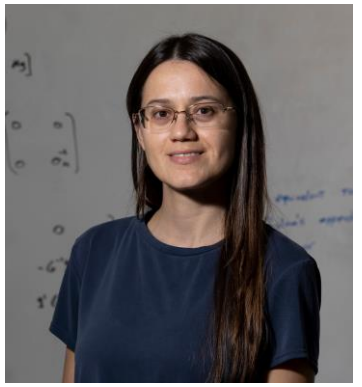
UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences



Increasing accuracy of genomic predictions: from SNP chips to sequence data

Daniela Lourenco

Shogo Tsuruta and Ignacy Misztal

6/12/2020

52nd Beef Improvement Federation Research Symposium and Convention

Genomics in animal and plant breeding

Theor Appl Genet (1983) 67:25–33

Genetic polymorphism in varietal identification and genetic improvement *

M. Soller [1] and J. S. Beckmann [2]

[1] Department of Genetics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
[2] Institute of Field and Garden Crops, Agricultural Research Organization, The Volcani Center 50250 Bet Dagan, Israel

Received July 14, 1982; Accepted July 3, 1983
Communicated by A. Robertson

© Springer-Verlag 1983

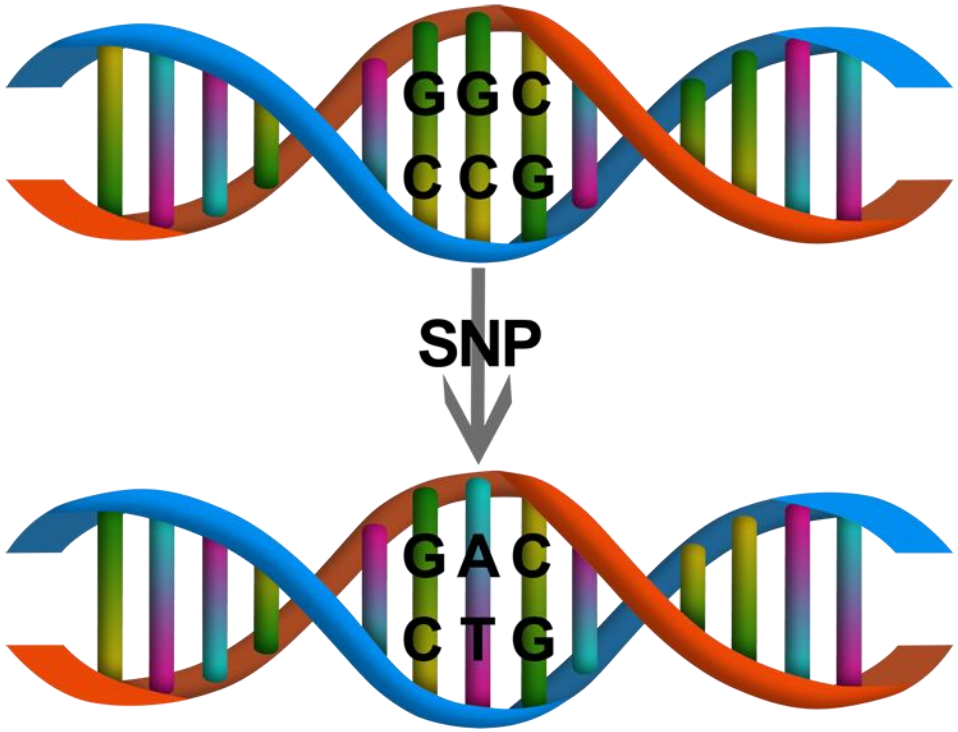
Summary. New sources of genetic polymorphisms promise significant additions to the number of useful genetic markers in agricultural plants and animals, and prompt this review of potential applications of polymorphic genetic markers in plant and animal breeding. Two major areas of application can be distinguished. The first is based on the utilization of genetic markers to determine genetic relationships. These applications include varietal identification, protection of breeder's rights, and parentage determination. The second area of application is based on the use of genetic markers to identify and map loci affecting quantitative traits and to monitor these loci during introgression or selection programs. A variety of breeding applications based on

Use of DNA markers

- Construct genetic relationships
- Parentage determination
- Identification of genes/QTL

SNP - Single nucleotide polymorphisms

Individual 1



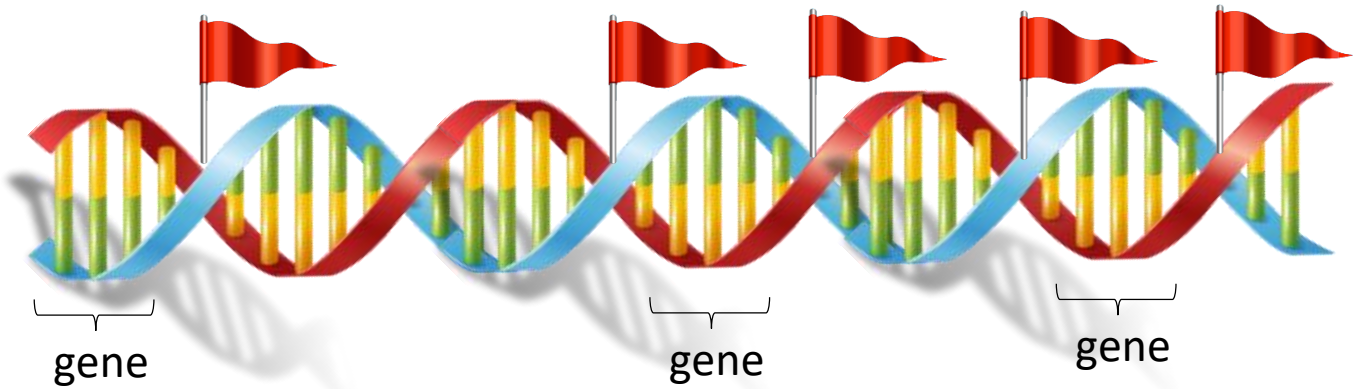
Individual 2

http://www.thinnergene.com/about-thinnergene/genetics-101/

"SNPs have become the bread-and-butter of DNA sequence variation" (Stonecking, 2001)

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

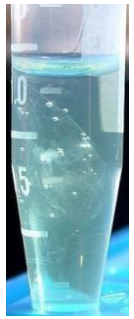
What are SNPs doing?



- Identify genes
- Track relationships
- Parentage determination

How many SNP in the genomic test?

- Livestock, poultry, aquaculture: 3,000 - 800,000 SNP

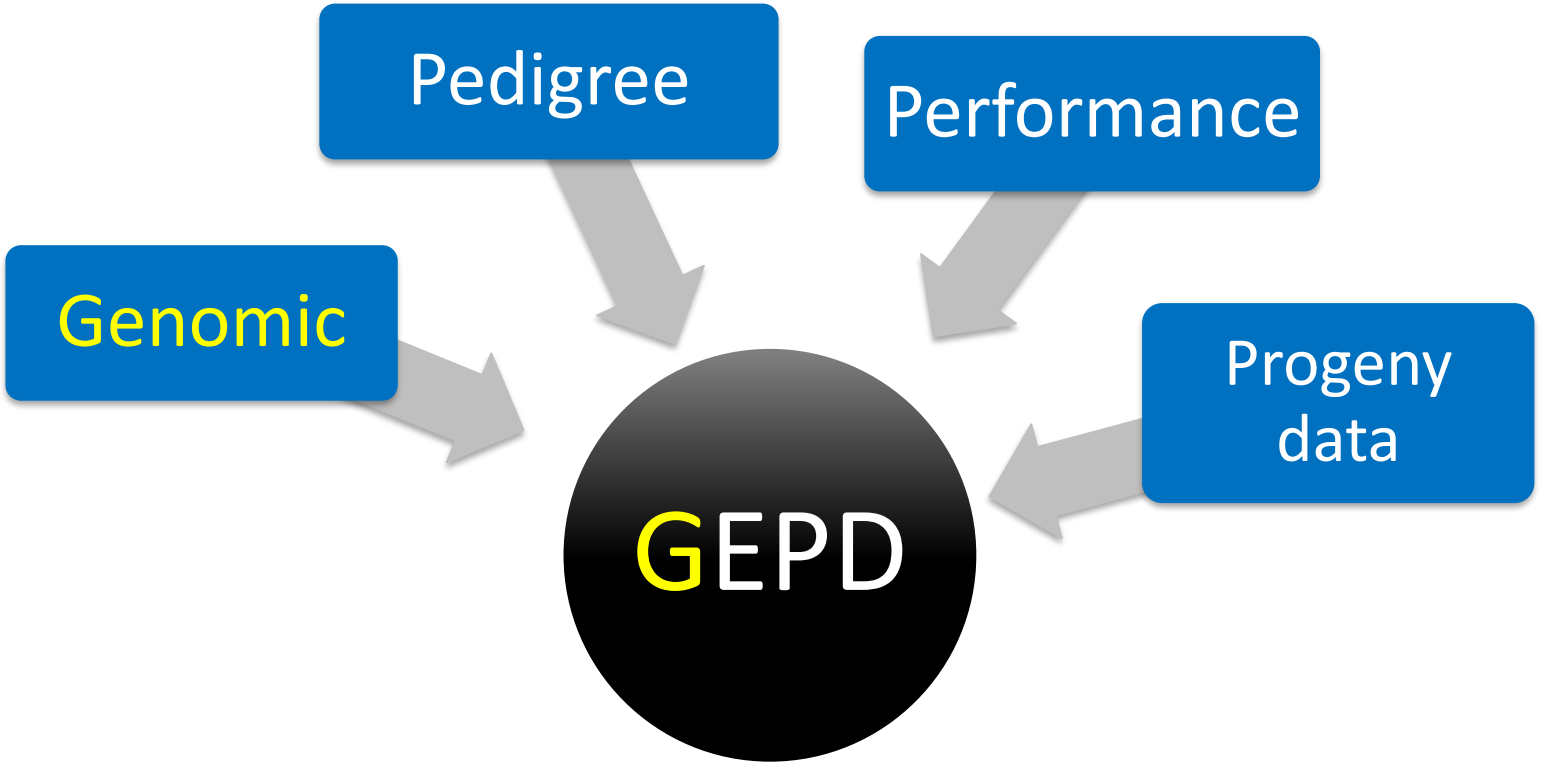


https://www.sciencelearn.org.nz/images/2543-dna-precipitate

https://www.lihs.cuhk.edu.hk/en-us/corefacilitiesandresources/corefacilities/geneexpressionprofiling,analysisandgenotyping.aspx

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Genomic Selection



UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Adoption of genomic testing

3,400,000

http://www.holsteinusa.net/programs_services/backgrounds.html

780,000

http://www.angus.org/AGI/default.aspx

460,000

https://www.usjersey.com/AJCA-NAJ-JMS/AJCA/AnimalIdentificationServices/HerdRegister.aspx

50,000 - 70,000

www.sheepcrc.org.au/

~50,000/line

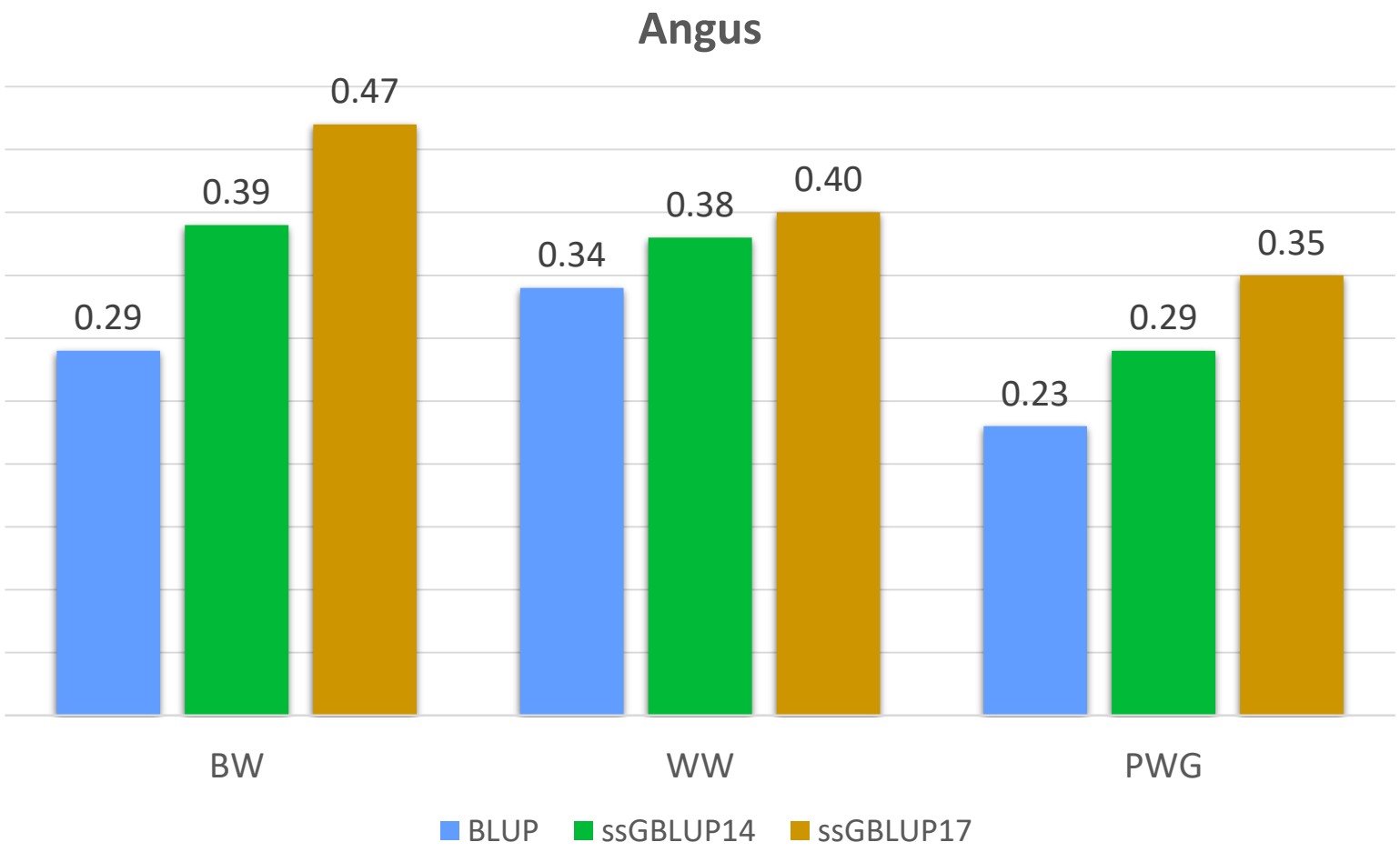
http://sesenfarm.com/raising-pigs/

~50,000/line

https://www.cobb-vantress.com/

2,000 - 10,000

Benefits of genomic testing in beef cattle



Average gain in accuracy

52k animals = 25%

335k animals = 36%

Genotyped for 50k SNP

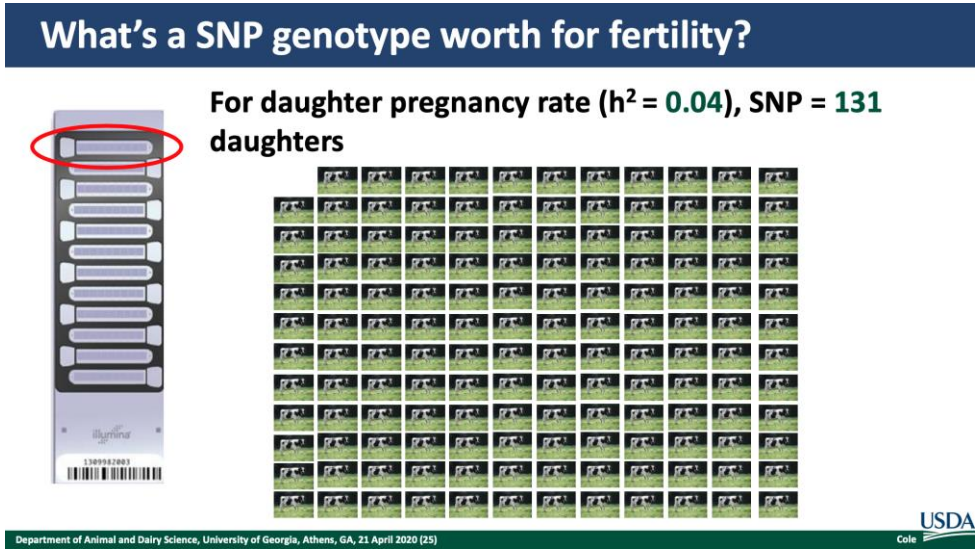
The gain in accuracy is because genomics is a non-redundant piece of information!

Lourenco et al., 2018

Benefits of genomic testing

Trait	PE	Trait	PE
Calving Ease Direct	26	Heifer Pregnancy	17
Birth Weight	23	Calving Ease Maternal	20
Weaning Weight	27	Milk	36
Yearling Weight	23	Mature Weight	15
Dry Matter Intake	12	Mature Height	9
Yearling Height	17	Carcass Weight	15
Scrotal Circumference	15	Carcass Marbling	11
Docility	12	Carcass Ribeye	17
Claw Angle	10	Carcass Fat	14
Foot Angle	10		

https://www.angus.org/AGI/GenomicEnhancedEPDs.pdf



Cole, 2020

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Benefits of genomic testing

$$\Delta G = \frac{i\, r\, \sigma_a}{L}$$

Genetic gain = (selection intensity * accuracy * genetic SD) / generation interval

Benefits of genomic testing in dairy cattle

PNAS

Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection

Adriana García-Ruiz[a,b], John B. Cole[b], Paul M. VanRaden[b], George R. Wiggans[b], Felipe J. Ruiz-López[a], and Curtis P. Van Tassell[b,1]

[a]Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Ajuchitlán, Querétaro, 76280, México; and [b]Animal Genomics and Improvement Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705

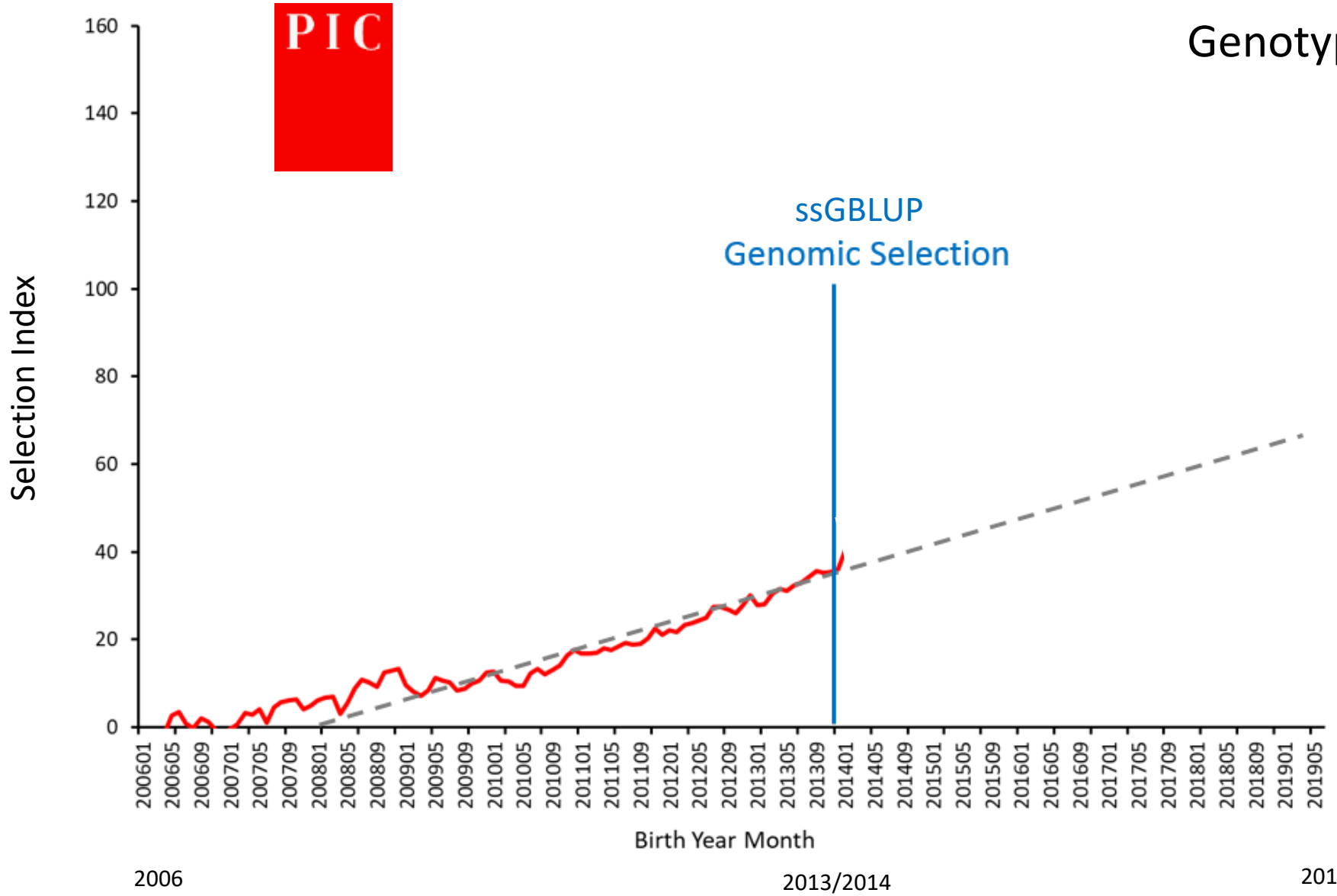
Genotyped for ~ 60k SNP

- Increase in genetic gain
  - ~50 to 100% for yield traits
  - 3- to 4-fold for traits with low heritability (fertility)

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Benefits of genomic testing in pigs

Genotyped for ~ 60k SNP



UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Benefits of genomic testing

- Increase in accuracy
  - Number of genotyped animals
  - Amount of performance records
  - Heritability of the trait
  - How informative the genomic test is
    - 50k SNP in beef cattle

Is the genomic test informative enough?

Do we have enough SNPs in the test?

Are we looking at the right SNPs?

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

About the right SNP

- Is the genomic testing informative enough?
- Do we have enough SNPs in the test?
- Are we looking at the right SNPs?

54k SNP cattle → 38k SNP

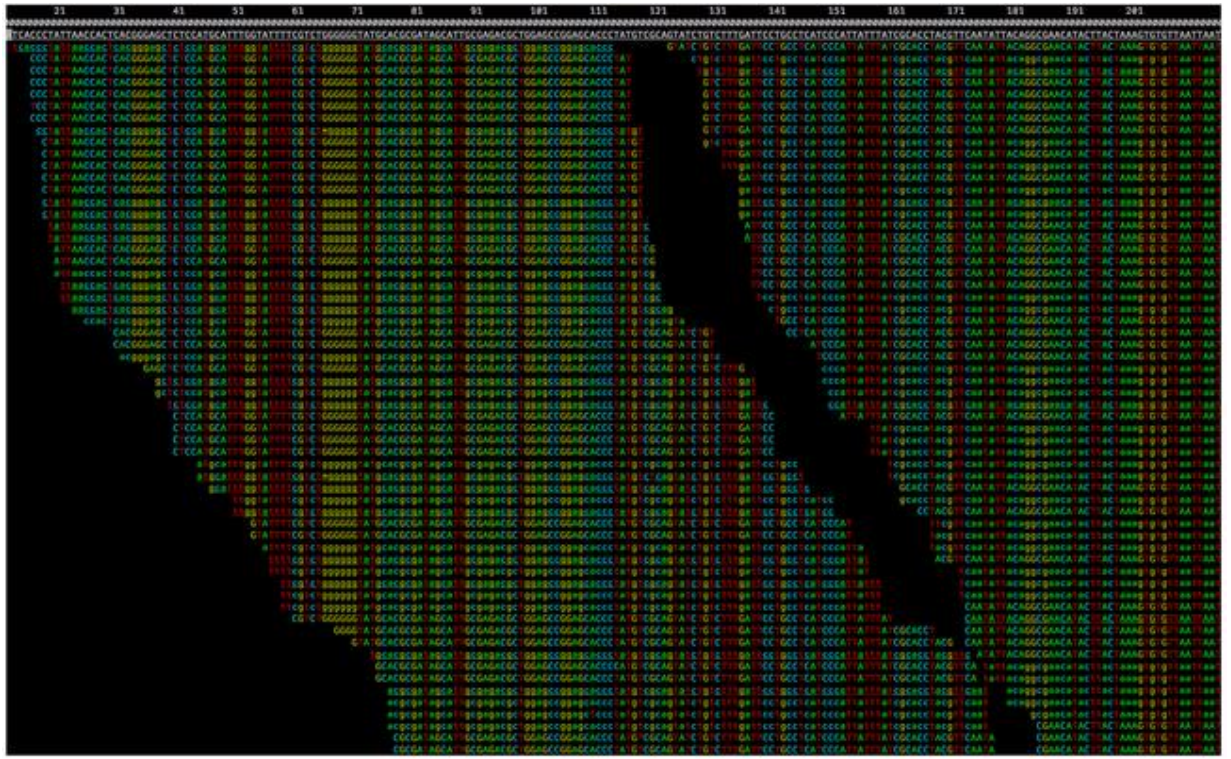
54k SNP pigs → 44k SNP

54k SNP chicken → 39k SNP

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Sequence information

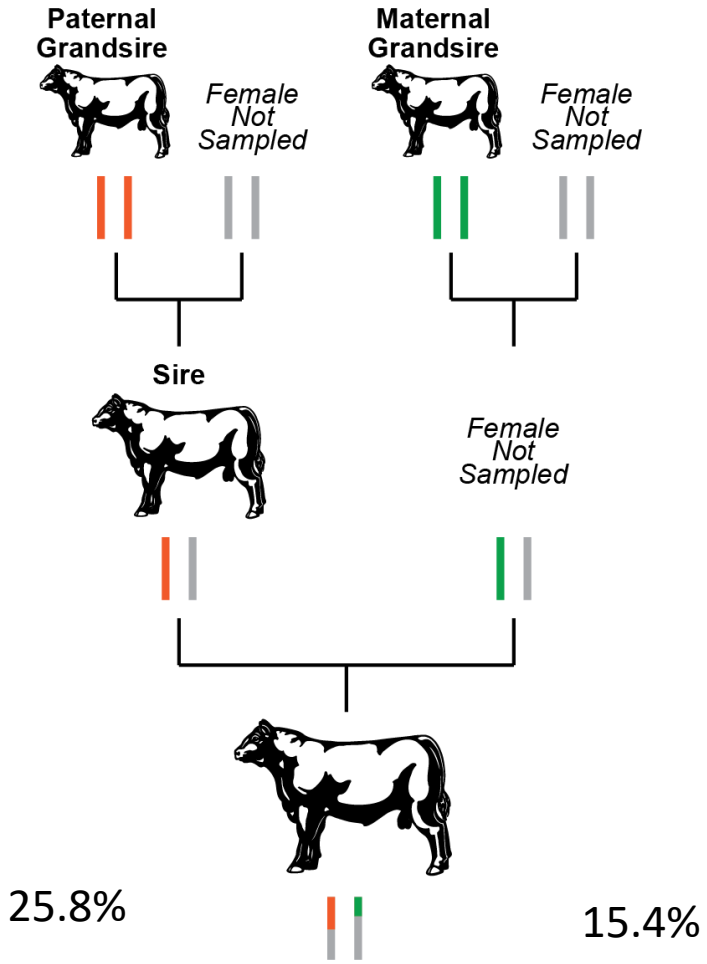
50k SNP



http://mtc.science/playing-around-with-ngs-step-by-step

30M SNP

SNPs capture relationships at genes

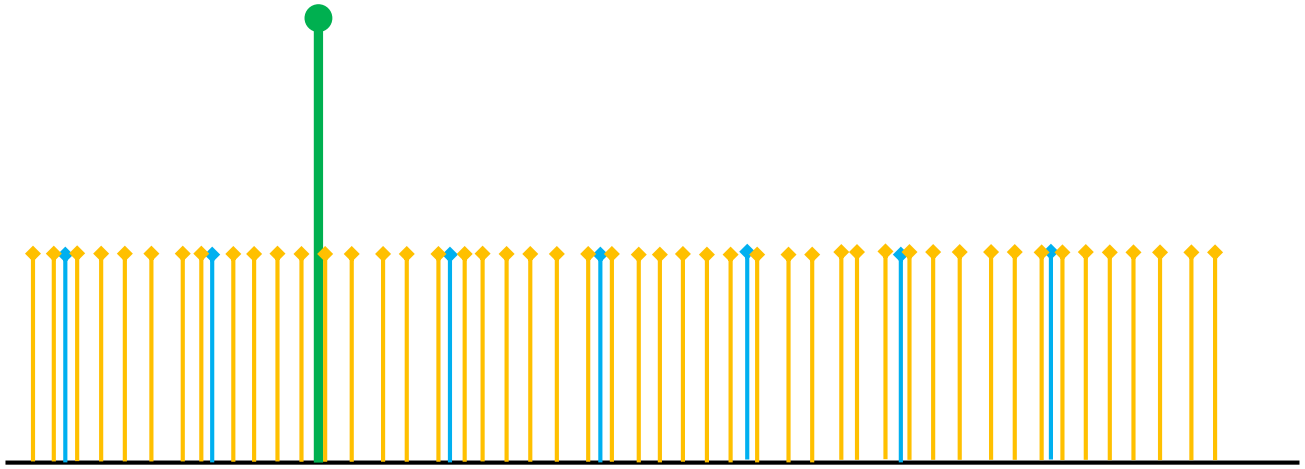


- Expected relationship between a bull and its grandparents?
- Observed relationships based on proportion of alleles shared between a bull and its paternal grandsire and maternal grandsire?

With sequence data, we may find SNPs that give more precise information about genes because they may be closer to genes

Jared Decker, University of Missouri

Sequence information



~ 50k SNP

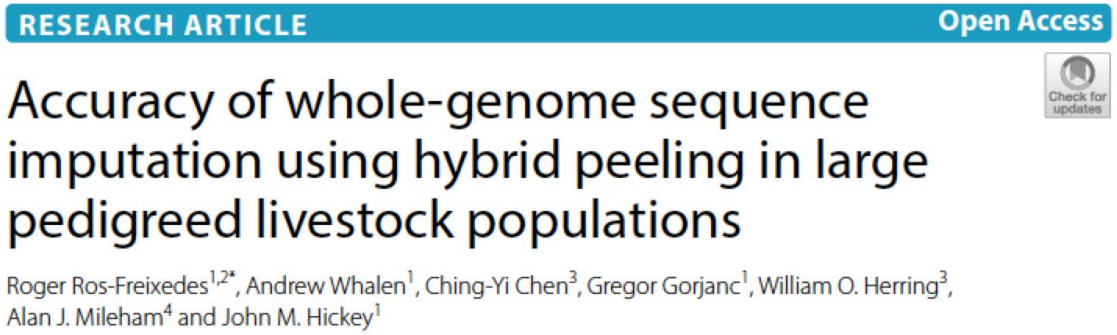
~ 30M SNP

- Trying to get a more precise information about the genes
  - More SNPs to cover larger areas of the DNA
  - Select SNP that are closer to genes for traits of interest

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Sequencing animals

- Becoming cheap
  - Sequence a small portion of animals
  - Imputation from 50k to sequence (filling the gaps)



RESEARCH ARTICLE Open Access

Accuracy of whole-genome sequence imputation using hybrid peeling in large pedigreed livestock populations

Roger Ros-Freixedes[1,2*], Andrew Whalen[1], Ching-Yi Chen[3], Gregor Gorjanc[1], William O. Herring[3], Alan J. Mileham[4] and John M. Hickey[1]

- 18,000 to 107,000 pigs
- Sequenced 2% of the animals
- Imputation to 20M to 30M
- Imputation accuracy of 94% - 98%

Largest sequence data in pigs

RESEARCH ARTICLE Open Access

Accuracy of whole-genome sequence imputation using hybrid peeling in large pedigreed livestock populations

Roger Ros-Freixedes[1,2*], Andrew Whalen[1], Ching-Yi Chen[3], Gregor Gorjanc[1], William O. Herring[3], Alan J. Mileham[4] and John M. Hickey[1]

- Gain in accuracy
- Multibreed
- Persistency of accuracy
- Up to 100,000 sequenced/imputed pigs
- 20M to 30M SNP

Do we work with all ~30M SNP?

Mining sequence data

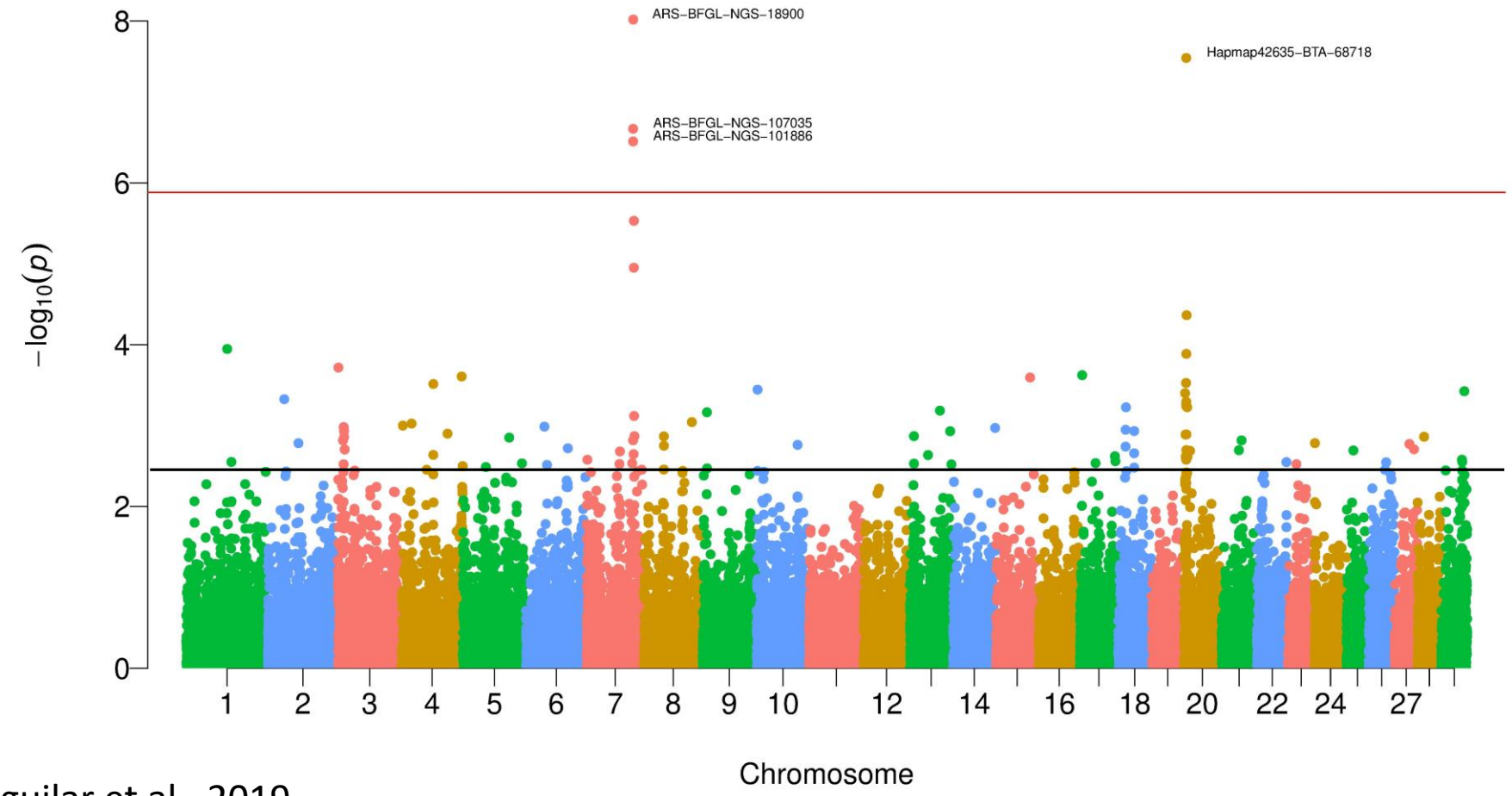
YOU COULD THINK OF IT AS IF SOME SNPs ARE CARRYING A TINY CAMPAIGN SIGN SUGGESTING WHICH GENE THEY'RE ASSOCIATED WITH.

Adapted from:
https://www.broadinstitute.org/visuals/explainer-genome-wide-association-studies

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Mining sequence data

Birth weight in Angus



- Causative SNP
- SNP give information about birth weight

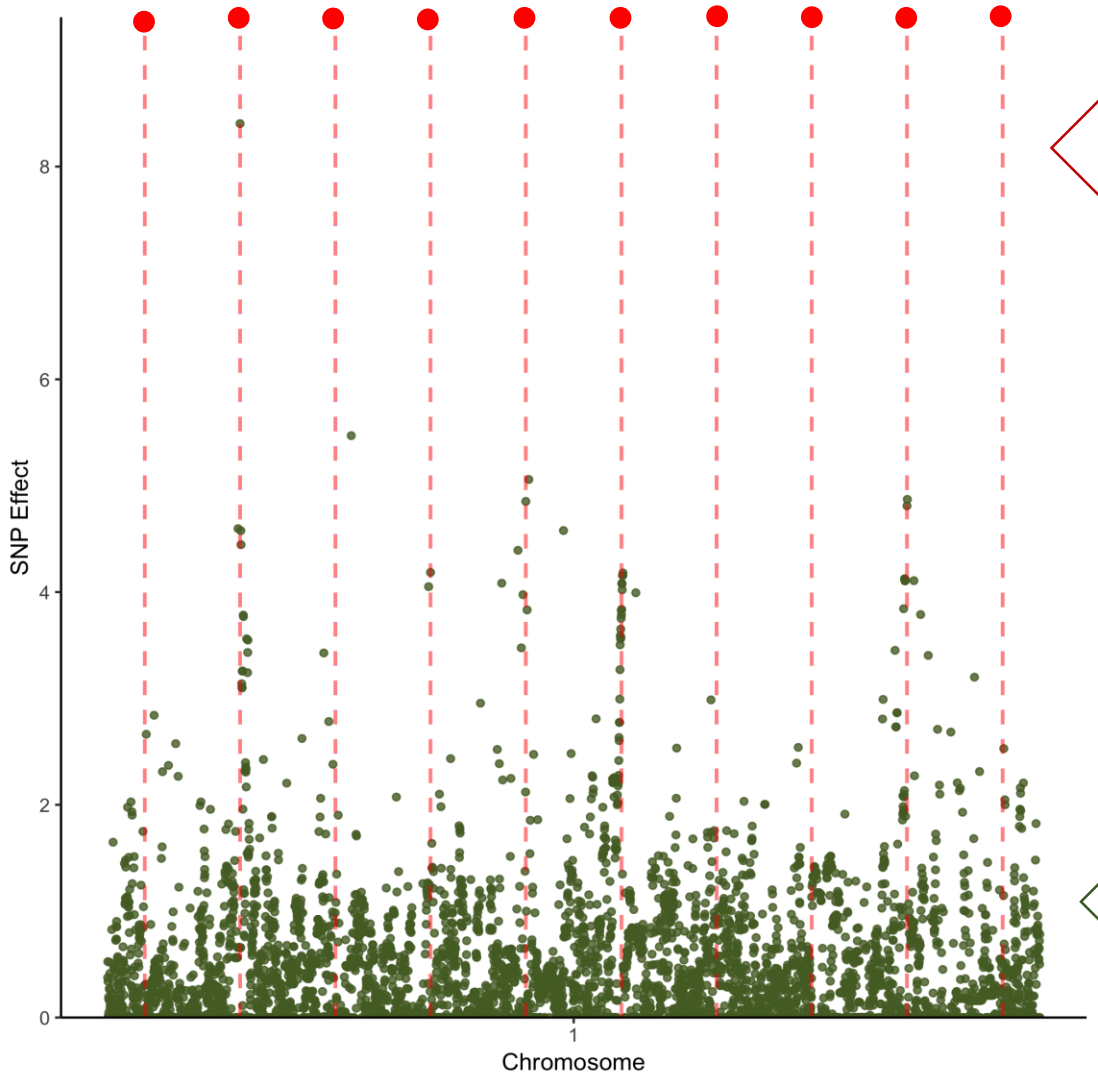
Aguilar et al., 2019

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Selecting causative SNP

- Issues
  - Effective population size
  - Number of animals with sequence and phenotypes
  - Large data – more confidence in finding causative SNP
  - Small data – more difficult to identify
    – more susceptible to errors

Selecting causative SNP



10 genes of equal effect

500 SNP to trace 10 genes

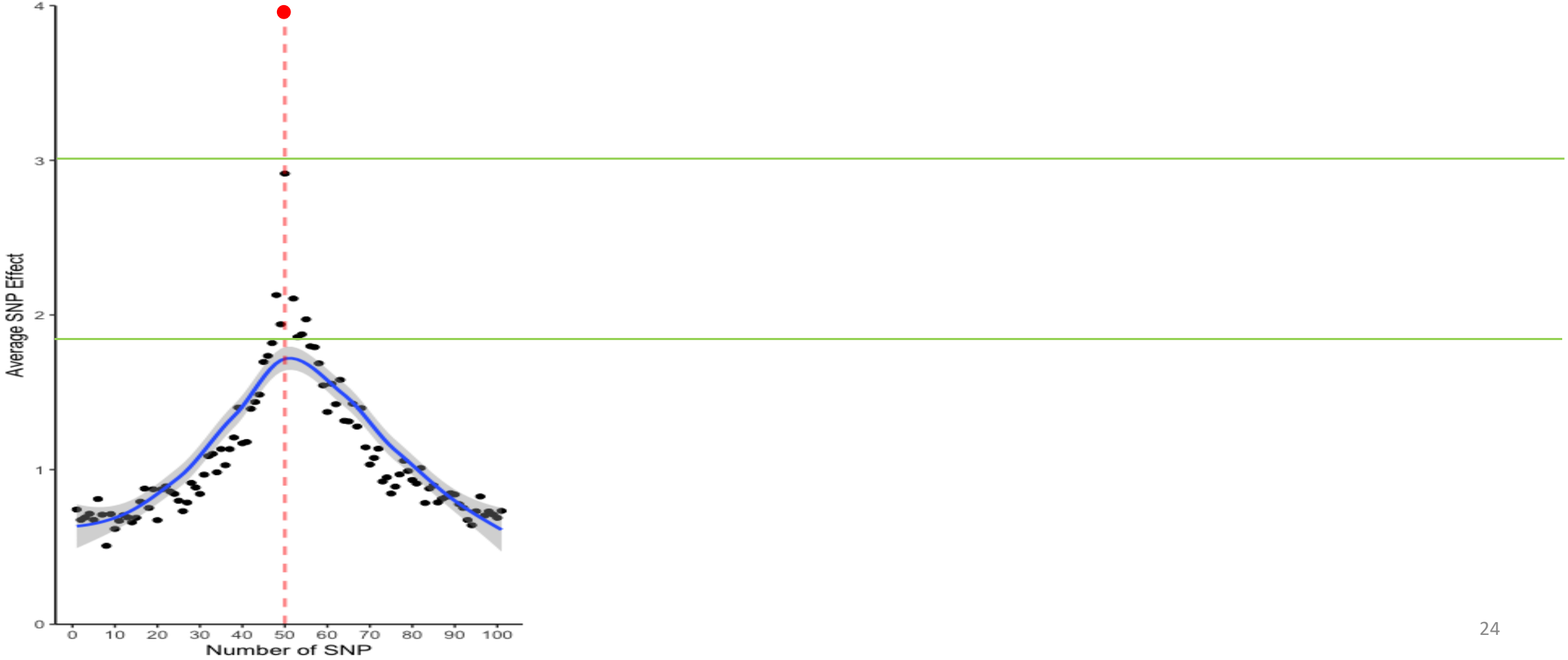
Pocrnic et al., 2020

Small population

Selecting causative SNP

Pocrnic et al., 2020

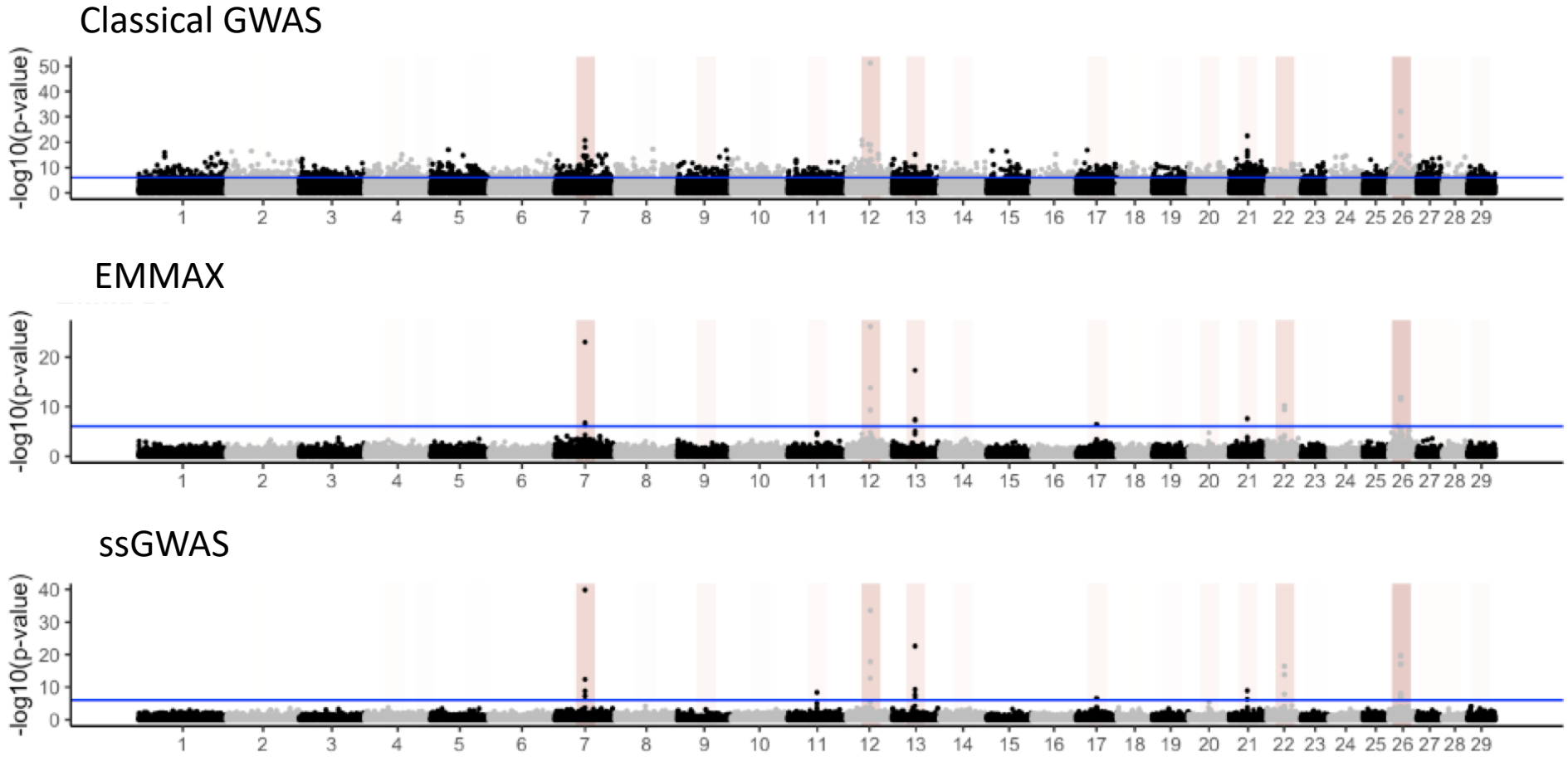
Ne = 60
Animals = 6000



Selecting causative SNP

- Small data: not enough power to discover which SNP are truly causative
- Methods are also important

Mancin et al., 2020



Using selected SNP to compute GEPD

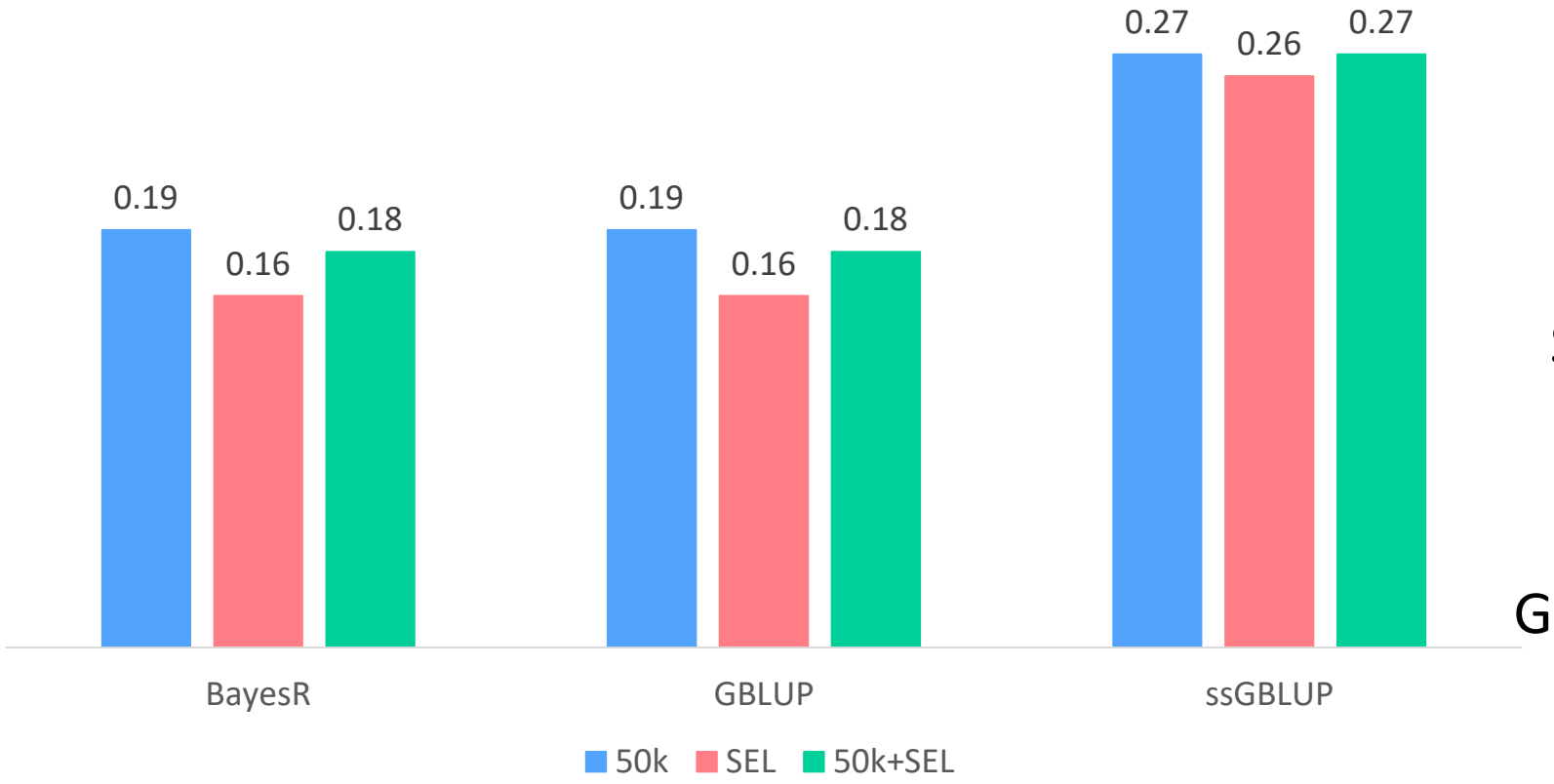
Jang et al., 2020

- Hanwoo cattle from South Korea
- 545k Marbling score records
- 1.3M animals in pedigree
- 1160 genotyped animals
- Imputation: 50k to 777k to 11.1M SNP
- 321k SNP selected out of 11.1M sequence SNP

en.prnasia.com

Gain in accuracy with selected SNP

Jang et al., 2020



No gain in accuracy with sequence SNP

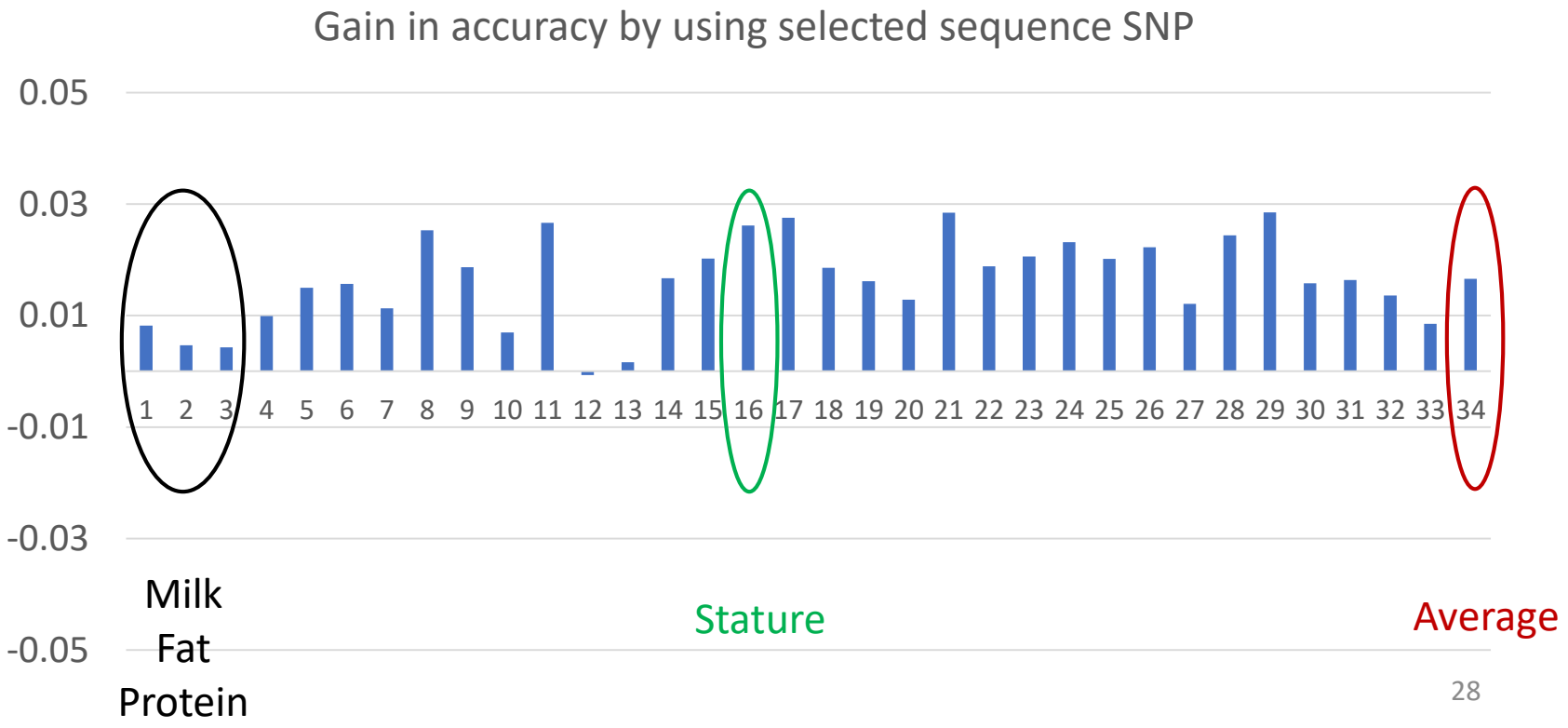
Small data: animals with SNP and records

GBLUP/BayesR = ~ 1200 records

ssGBLUP = 545k records

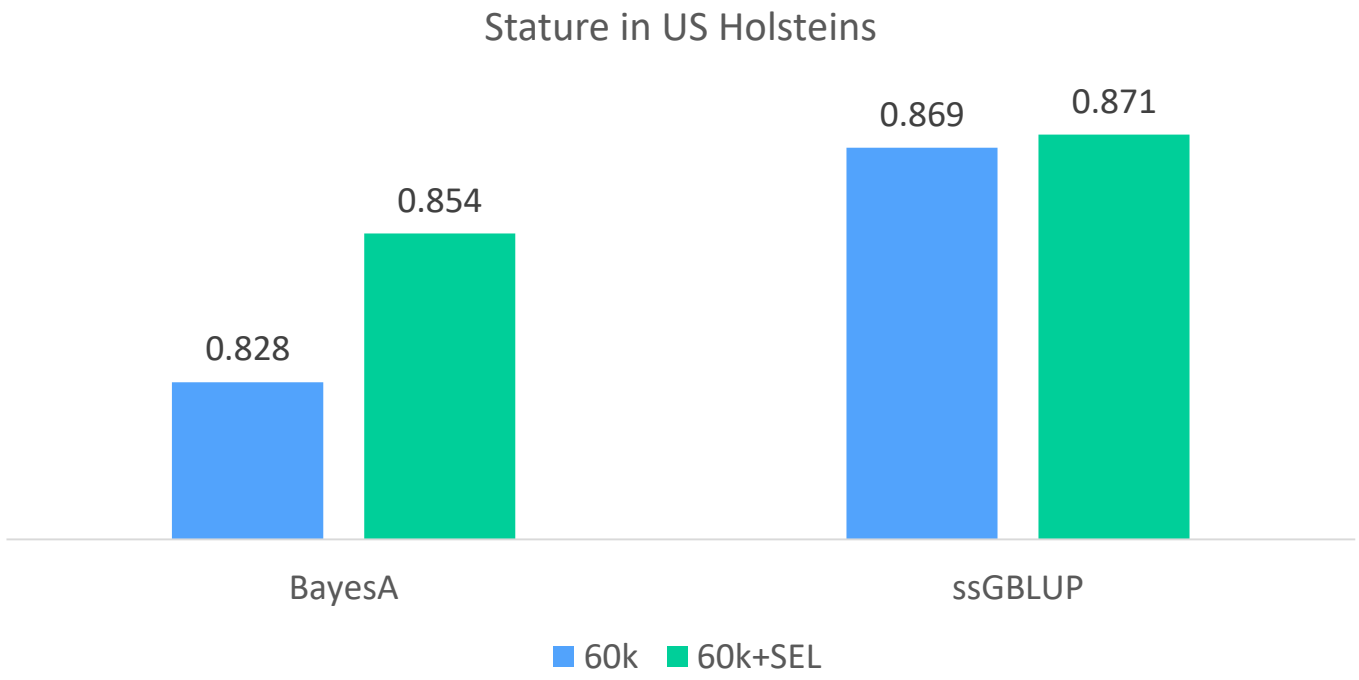
Gain in accuracy with selected SNP

- VanRaden et al. (2017)
  - 27k US Holstein bulls
  - Selected 16.6k SNP based on effect size on 33 traits
  - Added to the 60k SNP chip



Gain in accuracy with selected SNP

- Data from VanRaden et al. (2017) – Bayes A
  - 27k US Holstein bulls
  - Selected 16.6k SNP based on effect size on 33 traits
  - Added to the 60k SNP chip
- Fragomeni et al. (2019) – ssGBLUP



UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Simulated data – real genes

- Simulated population
  - 180,000 animals
  - 30,000 genotyped
  - 60,000 SNP
  - 100 genes (QTN)

- Single-step genomic predictions based on
  - SNP
  - SNP + QTN
  - QTN

Fragomeni *et al. Genet Sel Evol (2017) 49:59*
DOI 10.1186/s12711-017-0335-0

GSE Genetics Selection Evolution

RESEARCH ARTICLE Open Access

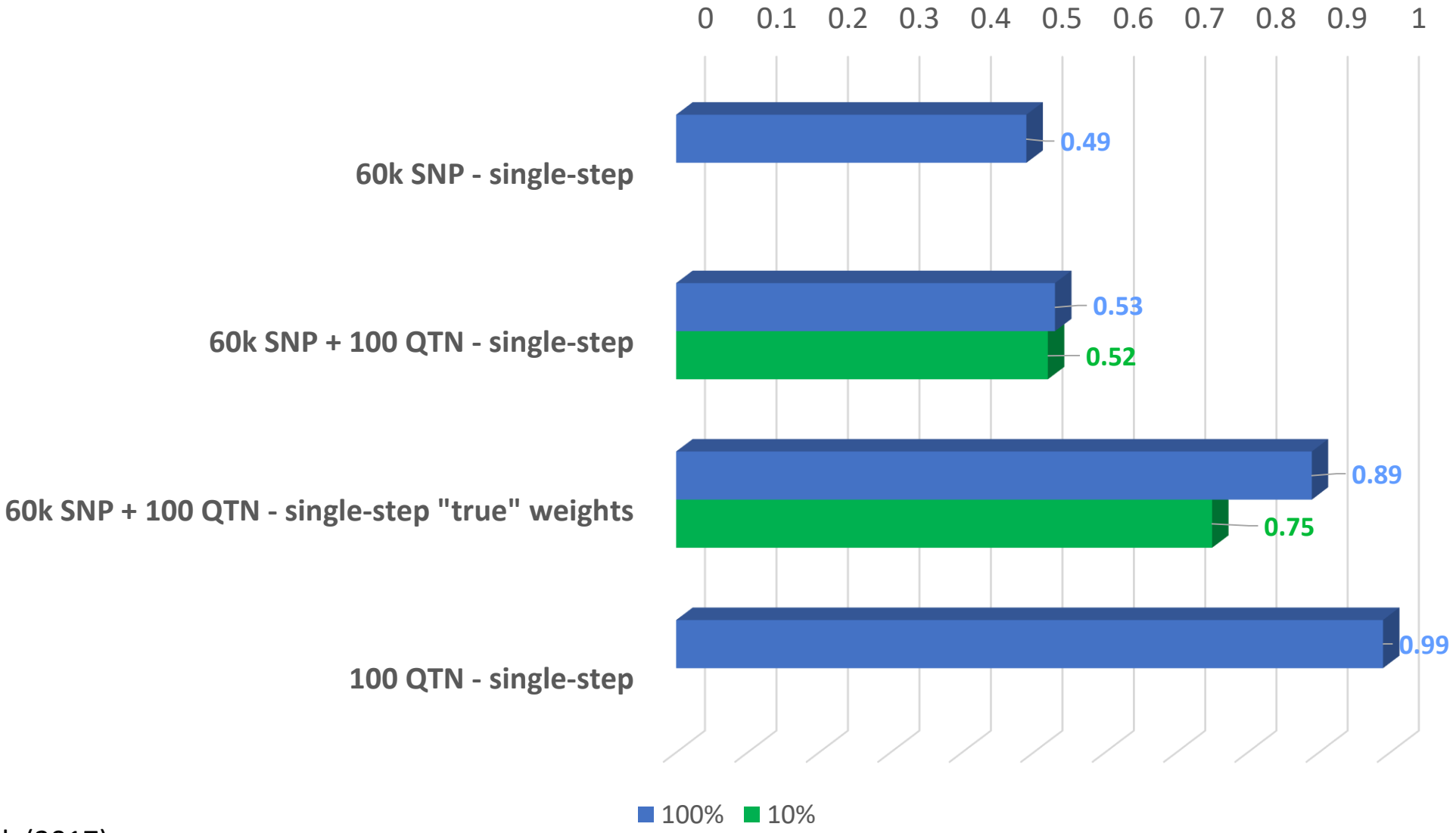
CrossMark

Incorporation of causative quantitative trait nucleotides in single-step GBLUP

Breno O. Fragomeni[1*], Daniela A. L. Lourenco[1], Yukata Masuda[1], Andres Legarra[2] and Ignacy Misztal[1]

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

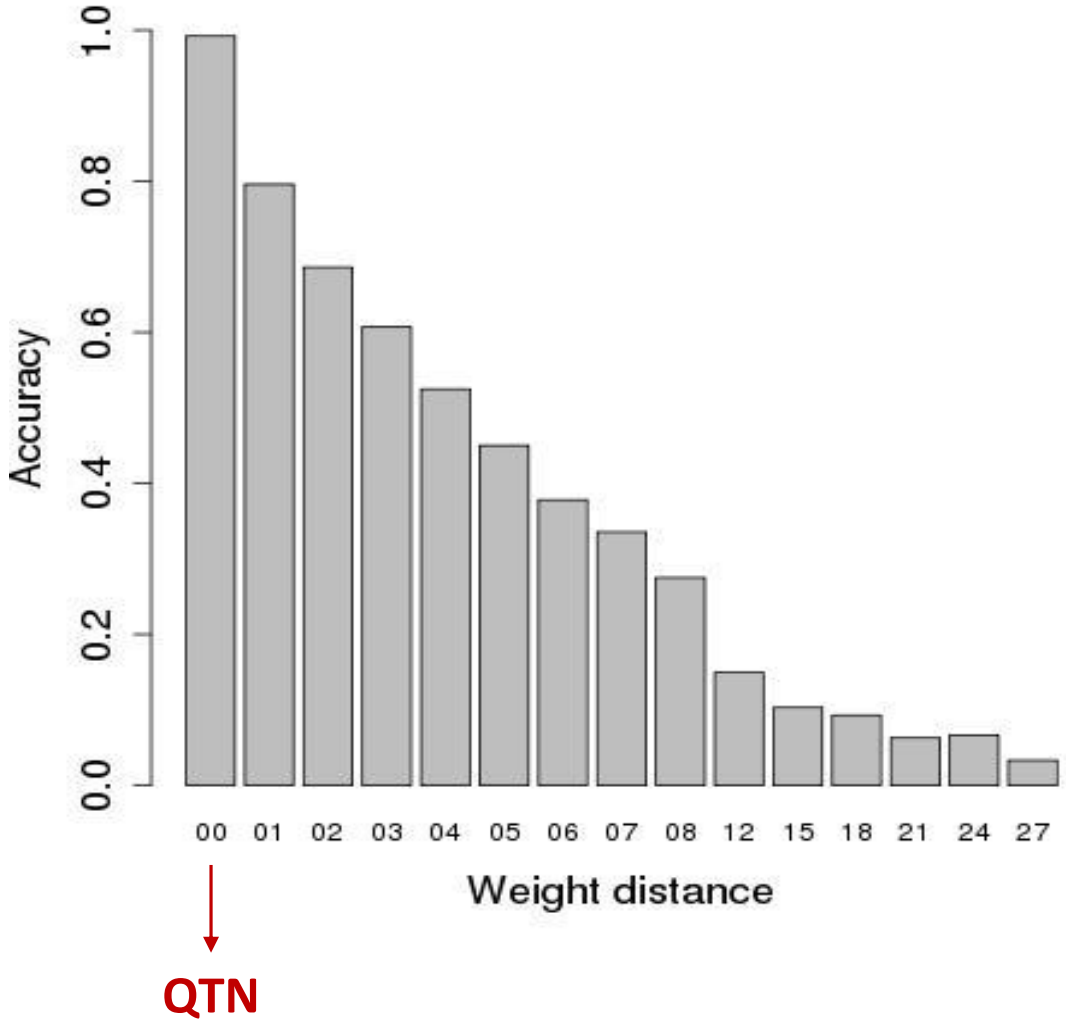
Accuracy with simulated genes



Fragomeni et al. (2017)

Accuracy with simulated genes

60k + 100 QTN – ssGBLUP APY – True weights



- Greatest accuracy if
  - all genes
  - locations are known
  - weights are optimal

Fragomeni et al. (2017)

SNP closely linked to genes

SNP that turn genes on and off

Amount of information they give

Is it possible to increase accuracy of genomic EPD with sequence data in real cattle populations?

We don't know all the genes

Knowing a bit may help a bit

Lots of work in progress

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Acknowledgements

USDA NIFA
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
zoetis
WAGYU
AMERICAN WAGYU ASSOCIATION
PIC
U.S. Registered Holsteins
FOR MAXIMUM PROFIT
HOLSTEIN ASSOCIATION USA, INC.
Cobb
Smithfield
THE MASCHHOFFS
Progressive Farming. Family Style.
DNA GENETICS
Greater gains. Improved performance.
methodgenetics
USDA United States Department of Agriculture
Agricultural Research Service
Warmwater Aquaculture Research Unit
PLEMDAT
ABCZ
SEMEX
Genetics for Life
PFHBiPM