

Single and multi-breed ssGBLUP using preselected variants from whole-genome sequence data in pigs

Sungbong Jang¹, R. Ros-Freixedes^{2,3}, J. Hickey³, C.Y. Chen⁴,
W.O. Herring⁴, I. Misztal¹, D. Lourenco¹

¹University of Georgia, USA

²Universitat de Lleida, Spain

³University of Edinburgh, Scotland

⁴Genus PIC, USA

7/7/2022



Animal Breeding and Genetics Group
College of Agricultural & Environmental Sciences
UNIVERSITY OF GEORGIA

WCGALP, Rotterdam



Motivation

- Genomic evaluation in farm animals is performed with SNP chip data
 - Number of SNP \approx 50k
- Number of QTL
 - Cattle: 192k for 701 traits
 - Pig: 35k for 716 traits
- Whole-genome
 - 20k ~ 25k genes
 - 30M ~ 40M SNP
- **Can 50k SNP include all information about genes and QTL?**

<https://www.animalgenome.org/cgi-bin/QTLdb/index>



Motivation

- WGS data
 - All genomic information
 - Becoming cheaper

So, how much?



In Humans: \$100 (UG 100)

www.ultimagenomics.com

- Should we use all SNP?
 - Genome has limited dimensionality
 - SNP within segments are likely inherited together due to strong LD
 - Redundant
 - Providing same information



Motivation

- Small to NO benefit with WGS in cattle

VanRaden et al., 2017
Fragomeni et al., 2019
Moghaddar et al., 2019
de Las Heras-Saldana et al., 2020
Lopez et al., 2021

- In pigs?

- Only few studies

Zhang et al., 2018
Song et al., 2019

- **Challenge**

- Small number of WGS pigs (< 300) and IWGS pigs (< 7k)
- No to limited benefits were observed
- Small effective population size (N_e) required large data to capture QTN

Jang et al., 2022



Objective

- Investigate the impact of using preselected variants from WGS for large-scale genomic predictions in single-breed and multi-breed scenarios



Data

- 3 terminal breeds
 - B1, B2, B3
 - MB (multi-breed)
- Traits
 - ADFI, ADG, BF
 - ADGX, BFX (recorded in crossbred animals)

- Same data used in previous presentation
- Ros-Freixedes et al. (2022)



Data

Breed	ADFI	ADG	BF	ADGX	BFX	Number of Animals in pedigree
B1	35k	0.36M	0.34M	150k	149k	1.13M
B2	40k	0.30M	0.30M	158k	156k	0.84M
B3	64k	0.94M	0.86M	299k	247k	3.14M
MB	140k	1.60M	1.50M	578k	525k	5.28M



Data

- Genotype
 - GGP-Porcine HD BeadChip (Chip1 – SBE)
 - Jointly imputed for MBE (Chip2 – MBE)
 - Chip2SB
- WGS and imputation
 - Imputation using AlphaPeel software
 - IWGS

Whalen et al., 2018

Ros-Freixedes et al., 2020



Data

Breed	N of genotyped individuals	N of SNP (Chip1)	N of sequenced individuals
B1	60,467	35,786	731
B2	41,572	40,311	760
B3	104,644	43,032	1,856
MB	206,634	41,303	3,347

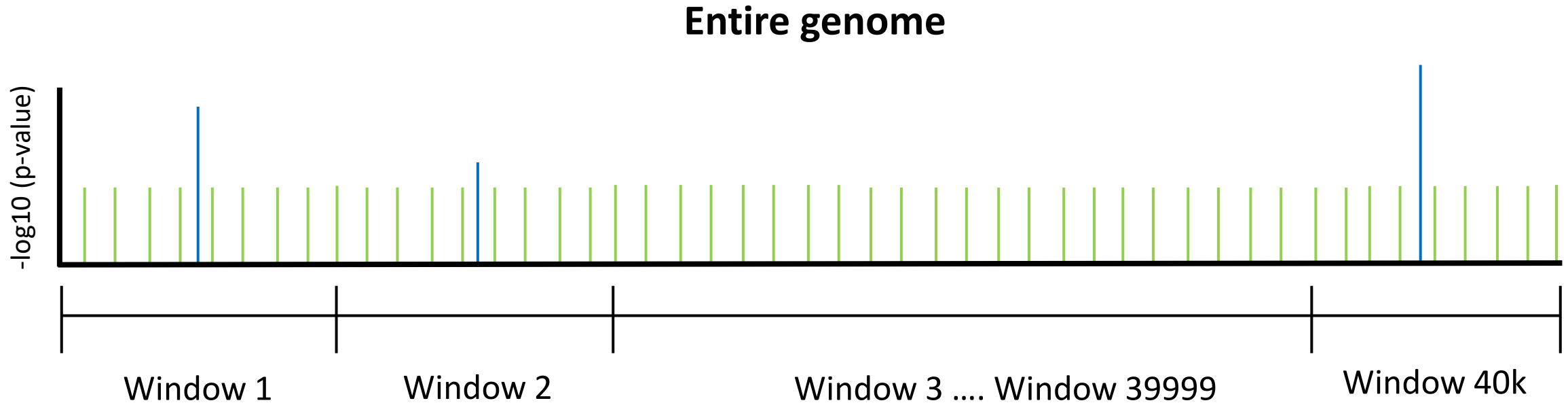


Data

- Preselected genotype panels Ros-Freixedes et al., 2022
- GWAS: Fast-LMM software Lippert et al., 2011
 1. Top40k
 - SNP with lowest p-value in each 55-kb windows
 2. ChipPlusSign
 - Combining **TopSign** to regular SNP chip data
 - **TopSign**: Only significant ones



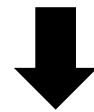
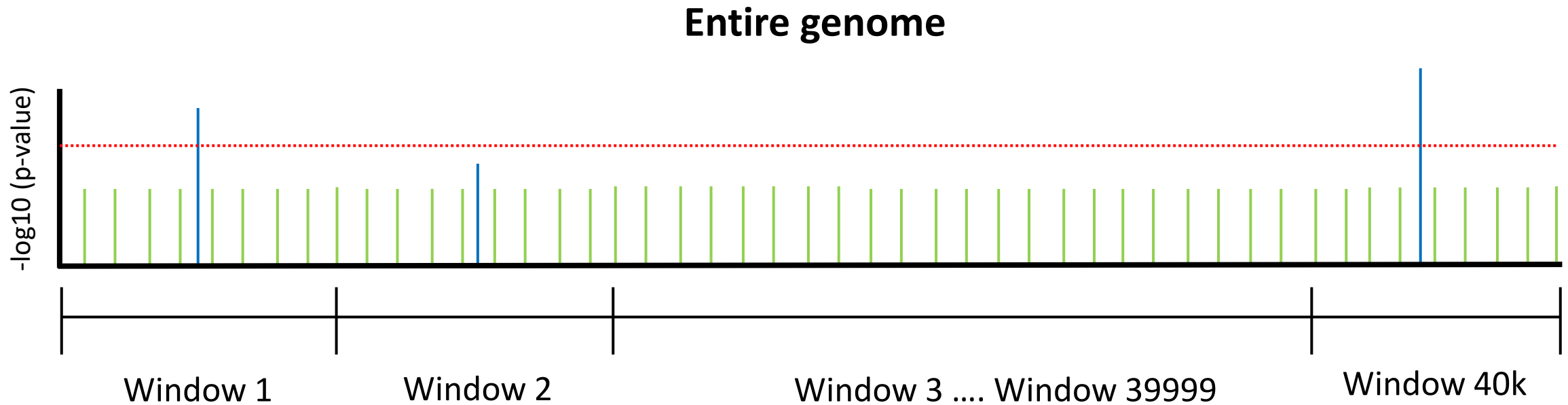
- **Top40k**



Extracting only 40k SNPs: Similar number as regular chip data (~40k)



- **ChipPlusSign**



Extracting only significant ones + SNP chip data



Genomic prediction

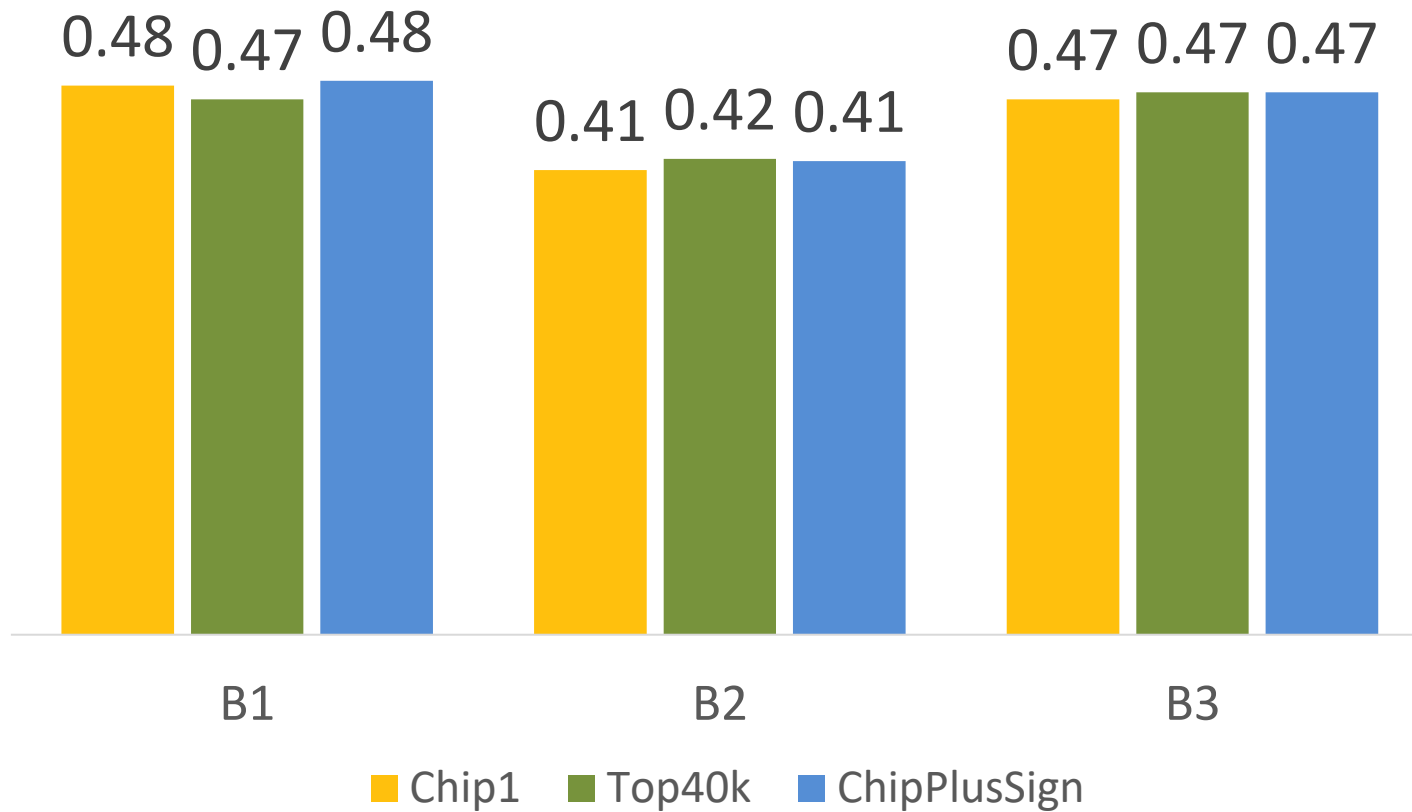
- ssGBLUP with APY
- UPG

- Validation
 - Prediction accuracy = $\text{cor}(\text{DEBV}, \text{GEBV})$

- Test sets
 - Entire litters from the last generation (youngest genotyped pigs)



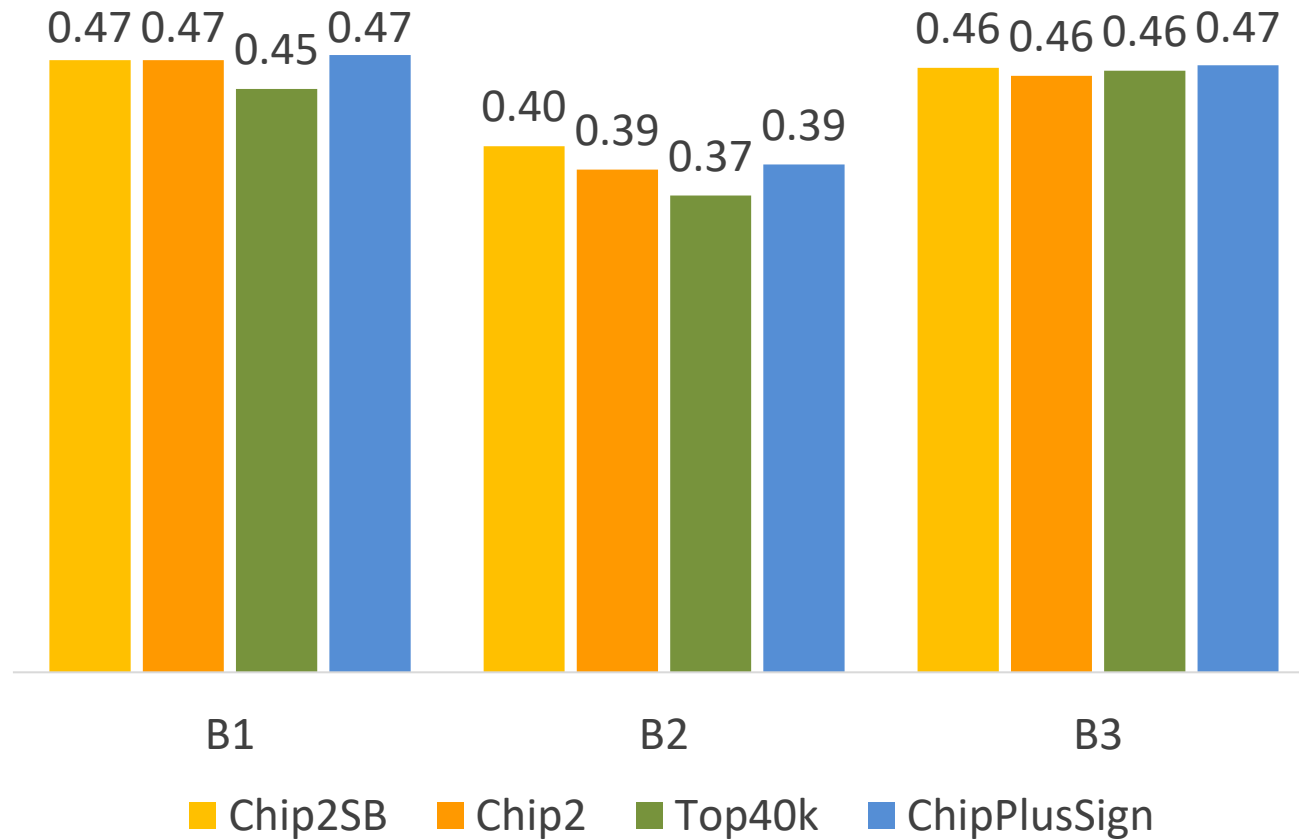
Single-breed evaluation



- Average acc
- Maximum gain
 - Compared to Chip1
 - ~0.03: Top40k
 - ~0.01: ChipPlusSign



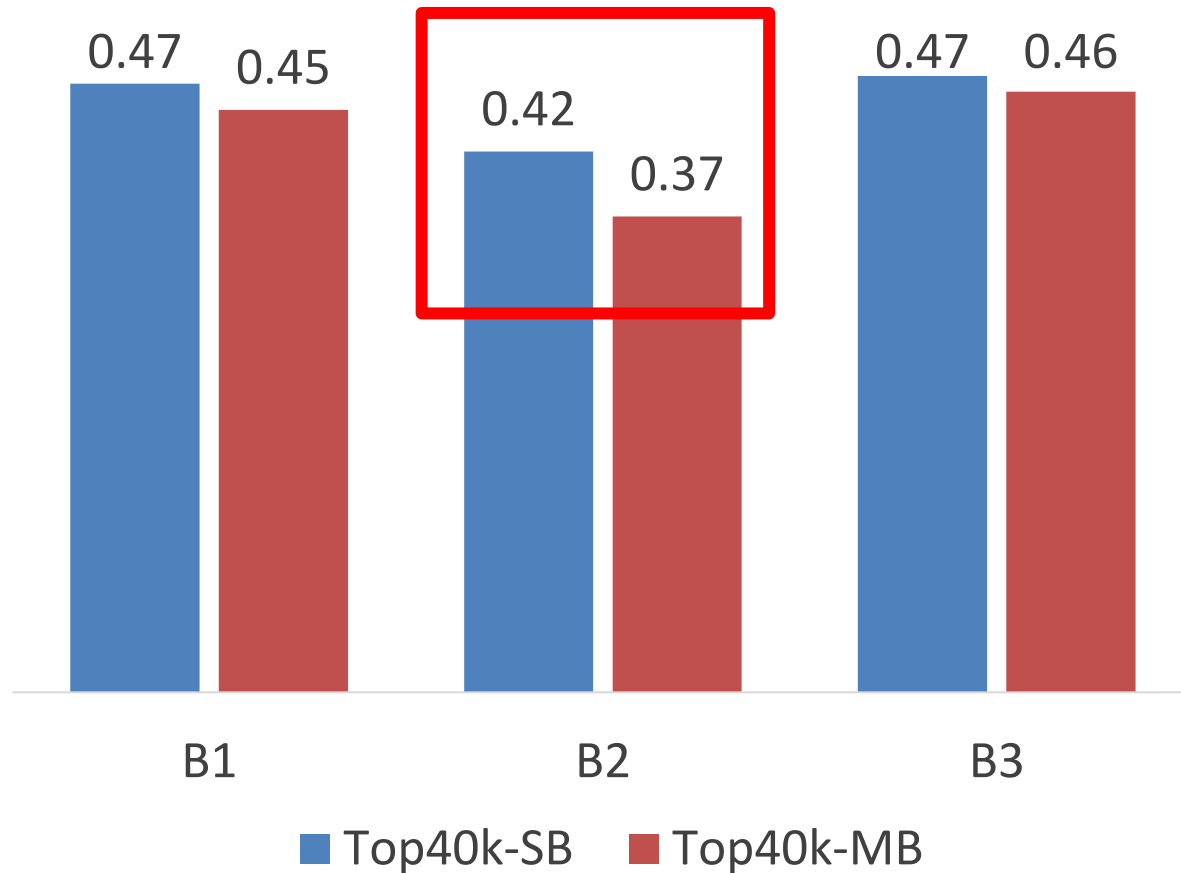
Multi-breed evaluation



- Average acc
- No benefit in MBE
 - Chip2SB vs Chip2
- Maximum gain
 - ~0.01: Top40k
 - ~0.01: ChipPlusSign



Top40k (SB GWAS vs MB GWAS)



- Average acc
- Top40k-SB > Top40k-MB
- More decrease in B1 and B2



Why limited to no benefits with WGS?

- Perfect prediction ($\text{acc} \approx 1$) is feasible
 - Real QTN
 - True effect as weight
 - Position in the genome

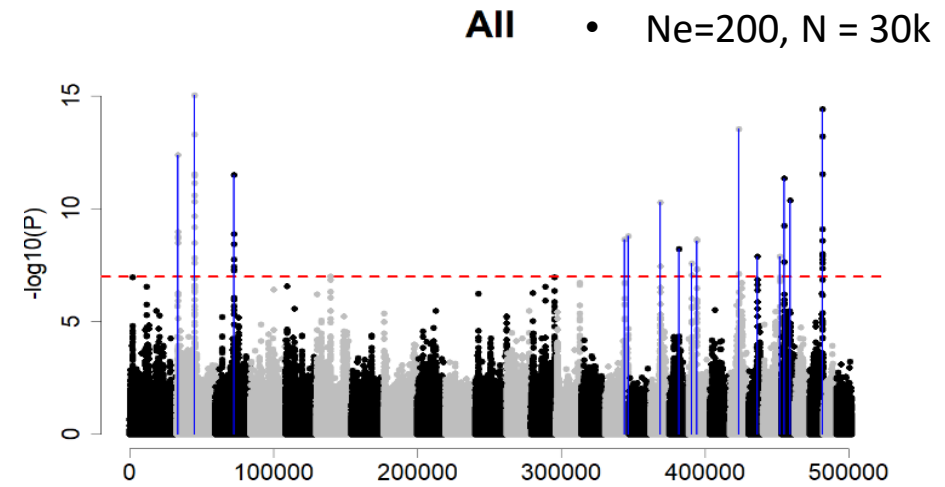
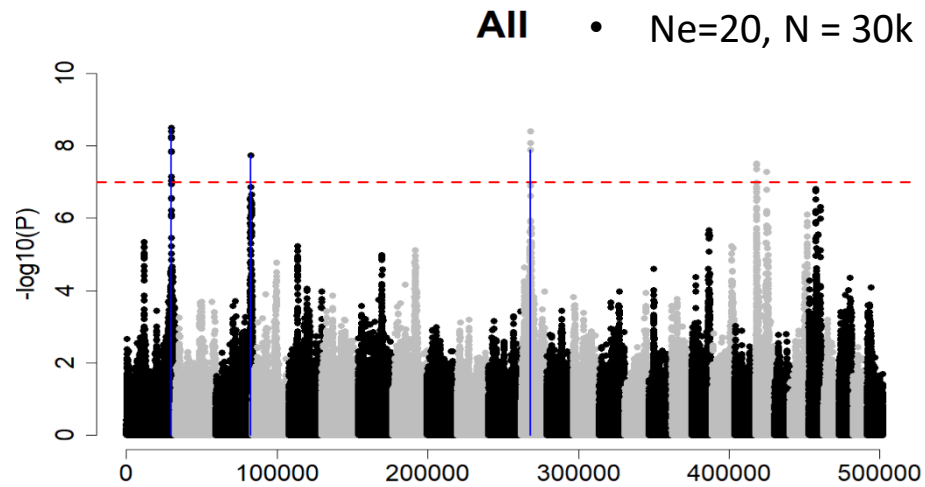
- *Ne*
 - Pigs and chickens have small *Ne*
 - Strong LD
 - Difficulty to pinpoint only causative variants

Fragomeni et al., 2017



Why limited to no benefits with WGS?

- GWAS resolution
 - QTN
 - Relationships
 - Noise (small sample size, imputation, etc)



Jang et al., 2022



Why limited to no benefits with WGS?

- Genomic prediction with preselected variants + SNP chip
 - Limited gain in small N_e (~2%)
 - Larger gain in large N_e (~9%), still limited
- Amount of information
 - Animals have a lot of information, high acc of GEBV
 - GEBV can be back-solved to SNP effects
 - $\hat{a} = \mathbf{DZ}'\mathbf{G}^{-1}\hat{u}$
 - $pval_i = 2(1 - \Phi(\left|\frac{\hat{a}_i}{sd(\hat{a}_i)}\right|))$

Jang et al., 2022



Conclusions

1. Preselected genotype panels from WGS improved the prediction accuracy depending on the breed, trait, size of the genotyped population, and genotype panels, but limited
2. Multi-breed scenario did not benefit but showed similar accuracy to the single-breed scenario



Acknowledgments



UNIVERSITY OF
GEORGIA



Thank you! Questions?

