



# Genomic predictions for marbling score in Hanwoo cattle using sequence data

Sungbong Jang<sup>1</sup>

A. Garcia<sup>1</sup>, S.H. Lee<sup>2</sup>, S. Tsuruta<sup>1</sup>, D. Lourenco<sup>1</sup>

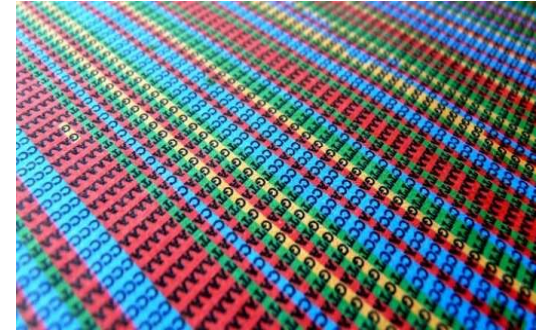
<sup>1</sup> University of Georgia, <sup>2</sup> Chungnam National University

# Hanwoo cattle

- Korean native breed
- Small size & brown coat color
- High marbling



# Why sequence data?



phys.org

- More information
  - SNPs chip cannot capture all variants in genome
  - Need more SNPs to cover whole genome
- Sequencing price is getting cheaper
  - Imputation



# Objectives

- Assess gain in accuracy using
  - Regular SNP chip
  - Sequence data
  - Regular SNP chip + sequence data
- Test different methods for genomic prediction



# Animal data

- Hanwoo Research Institute in South Korea
- 1.3M pedigree
- 545K marbling score phenotype
  - 1 to 9 score (9: High-marbling, 1: Low-marbling)
- 1,160 genotyped animals
- Pedigree-based  $h^2 = 0.51$



# Genomic data

- 50K, DEG, 50k + DEG, Imputed\_WGS
  - DEG: Selected SNPs from Imputed\_WGS

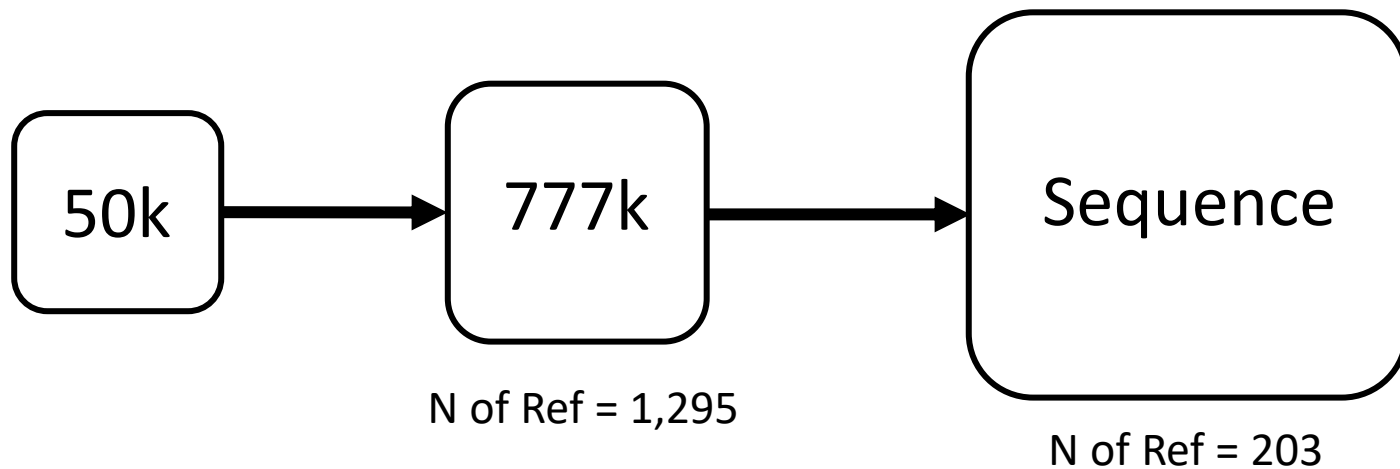
SNP panel	Number of SNPs after QC
50k	39,822
DEG	321,614
50k+DEG	360,407
Imputed_WGS	11,146,536



# Imputation

- Eagle 2.4v (Phasing)
- Minimac3 (Imputation, Filtered:  $R^2 > 0.6$ )
- Two-step

Das et al., (2016)  
Howie et al., (2012)



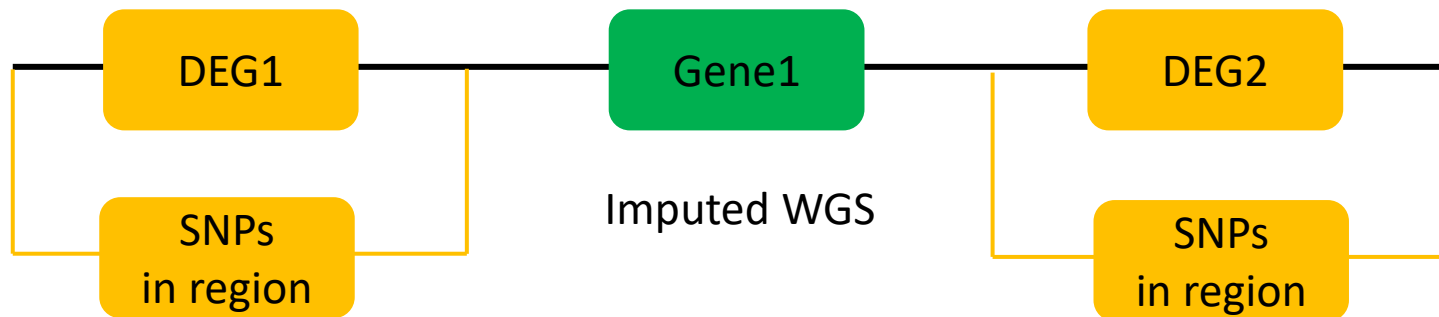
# DEG identification

High marbling vs Low marbling

High fat (%) vs Low fat (%)

- Differentially Expressed Genes
- RNA-seq analysis I (n = 24, Limma package)
- RNA-seq analysis II (n = 40, Linear regression)
- Total 336 DEGs were identified

Ritchie et al., (2015)





# Methods

- BayesR

$$p(\beta_j | \pi, \sigma_g^2) = \pi_1 \times N(0, 0 \times \sigma_g^2) + \pi_2 \times N(0, 10^{-4} \times \sigma_g^2) + \pi_3 \times N(0, 10^{-3} \times \sigma_g^2) \\ + \pi_4 \times N(0, 10^{-2} \times \sigma_g^2)$$

Erbe et al., (2012)



# Methods

- GBLUP (BLUPF90 programs)

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\lambda \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

- ssGBLUP (BLUPF90 programs)

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + H^{-1}\lambda \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

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

- $A_{22}$  = pedigree relationship matrix for genotyped animals
- $G$  = genomic relationship matrix



# Methods

- SNP weighting (BLUPF90 programs)
  - SNPs are assumed to have different effects on the traits

1.  $D = I, G = \frac{ZDZ'}{2\sum p_i(1-p_i)}$

2. GEBV from ssGBLUP

3.  $\hat{a} = \lambda DZ'G^{-1}\hat{u}$

4.  $d_i = CT \frac{|\hat{u}_i|}{sd(\hat{u})}^{-2}$

5. Iteration from 2

VanRaden (2008)

Wang et al., (2012)

Fragomeni et al., (2019)



# Methods

- Linear weights:  $d_i = \hat{u}_i^2 2p_i(1 - p_i)$   
(Falconer & Mackay, 1996)
- Non-linear weights:  $d_i = CT \frac{|\hat{u}_i|}{sd(\hat{u})} - 2$   
(VanRaden, 2008)

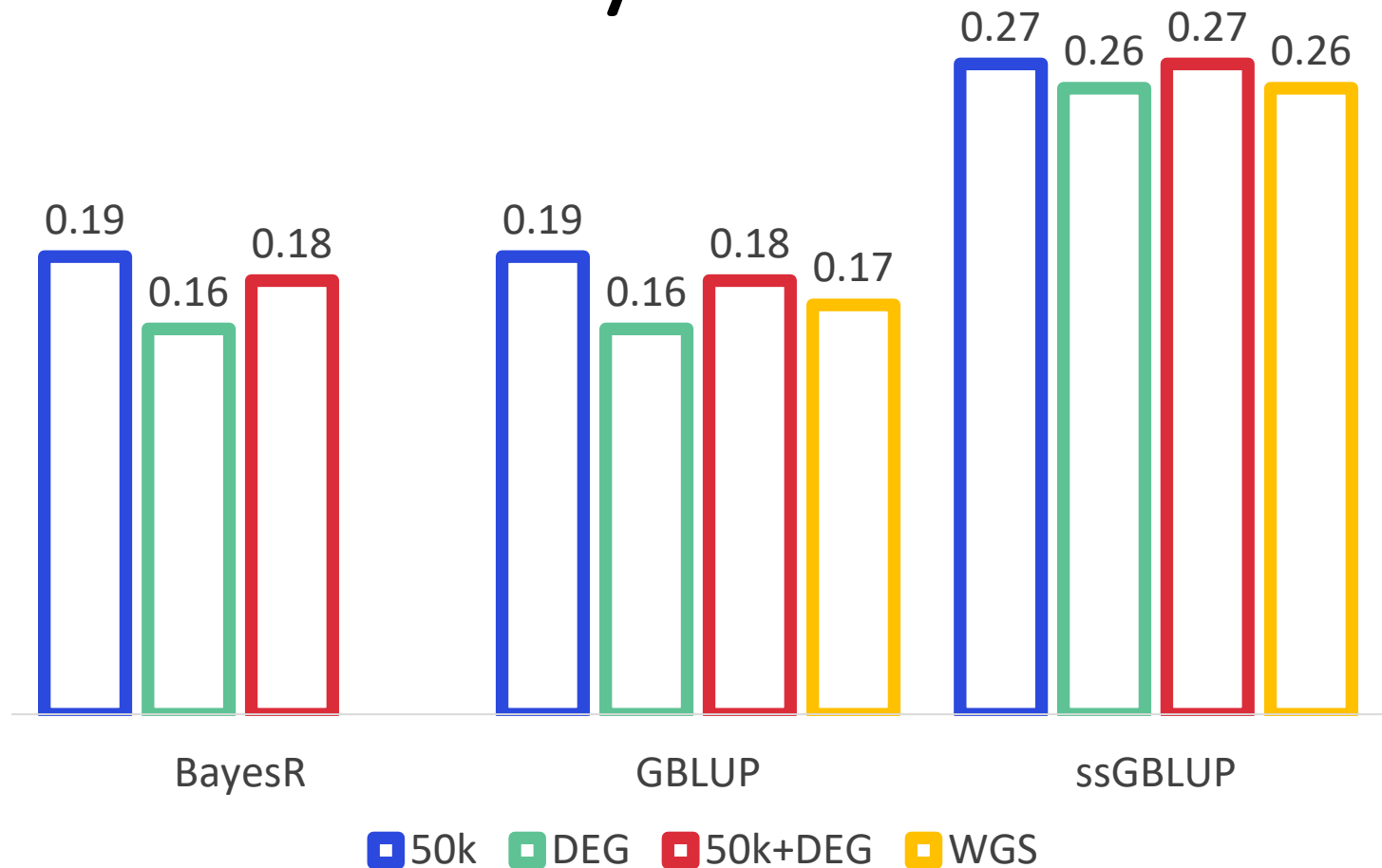


# Validation

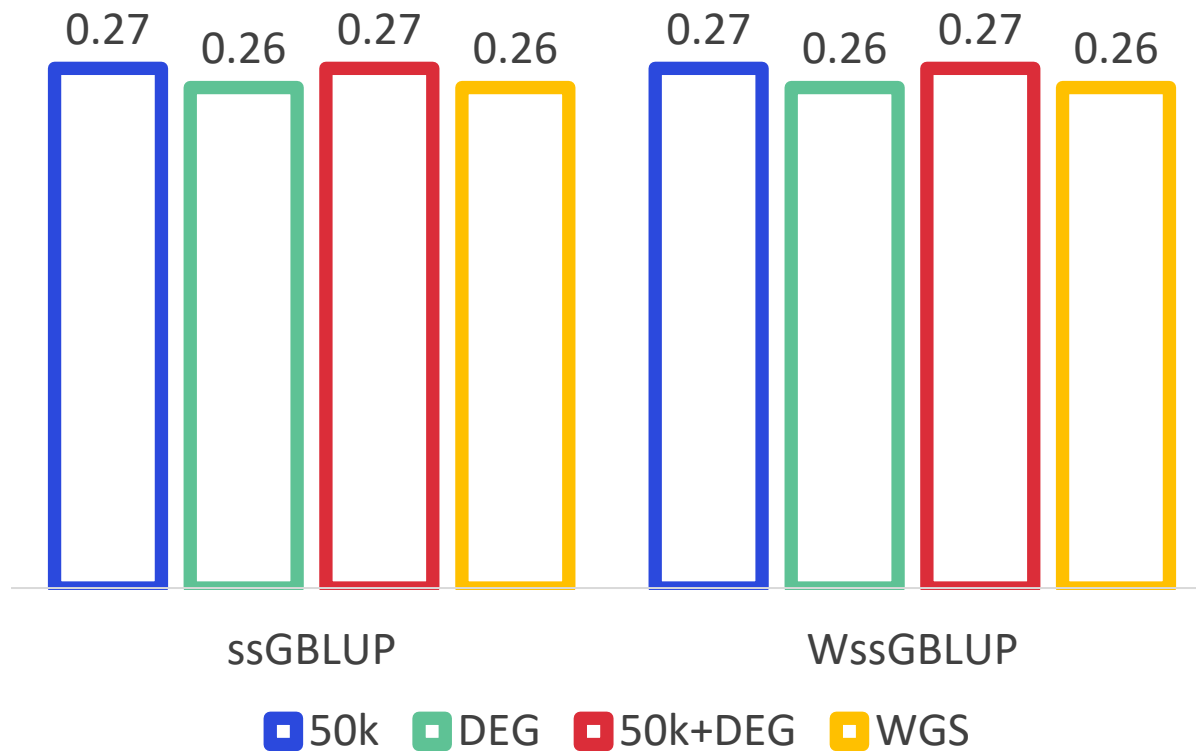
- Youngest genotyped animals ( $n = 169$ )
- Predictive ability:  $\text{cor}(\text{GEBV}, y_{adj})$
- Inflation ( $b_1$ ):  $y_{adj} = b_0 + b_1 \times \text{GEBV}$



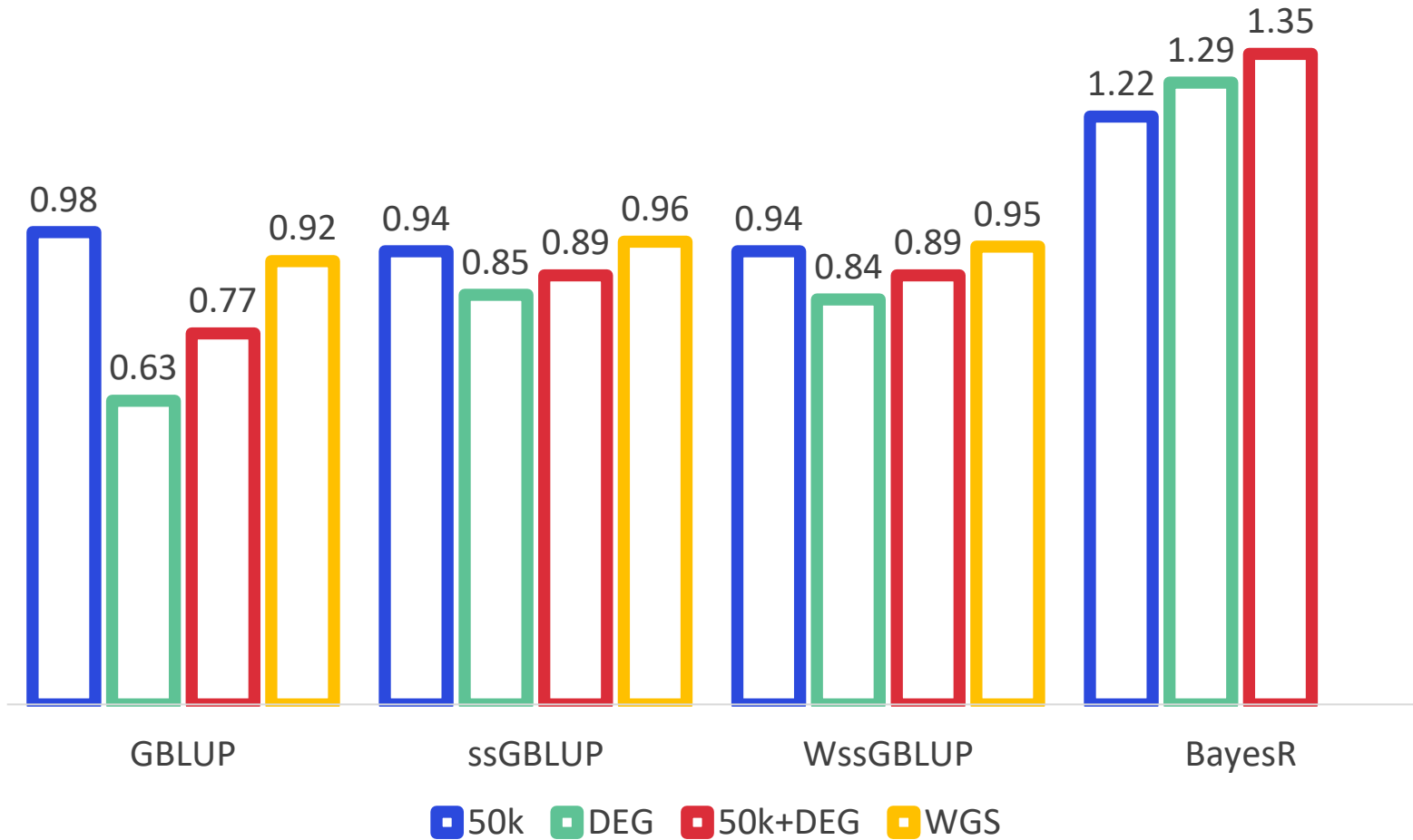
# Predictive ability



# Predictive ability



# Inflation





# Conclusion

- No gain using sequence data
- ssGBLUP outperforms GBLUP and BayesR
  - More data is used in ssGBLUP
- No increase in predictive ability using SNP weighting




# Limitations

- Small number of genotyped animals with records
- Non-stringent way to select SNPs (DEG)
- Are those selected SNPs (DEG) really causative?



# Acknowledgements





# Thank you!

[jsbng@uga.edu](mailto:jsbng@uga.edu)

