

# Use of causative variants and SNP weighting in a single-step GBLUP context

Breno Fragomeni<sup>1</sup>, Lourenco DAL<sup>1</sup>, Legarra A<sup>2</sup>, Tooker ME<sup>3</sup>, VanRaden PM<sup>3</sup>, Lawlor, T.<sup>4</sup>, Misztal I<sup>1</sup>

<sup>1</sup>University of Georgia, Athens, USA

<sup>2</sup>INRA, Castanet-Tolosan, France

<sup>3</sup>AGIL ARS-USDA, Beltsville, USA

<sup>4</sup>American Holstein Association, Brattleboro, USA



UNIVERSITY OF  
**GEORGIA**



7 – 11 February 2018

THE GLOBAL STANDARD  
FOR LIVESTOCK DATA

Annual Conference

icar2018.nz



11-16 February 2018

11<sup>th</sup> WORLD CONGRESS  
ON GENETICS  
APPLIED TO  
LIVESTOCK PRODUCTION  
wcalp.com



A O T E A C E N T R E , A U C K L A N D , N E W Z E A L A N D

# Motivation

- Decreasing costs of whole genome sequence
- Revived interest in causative variants for prediction
- Several authors are finding and using causative variants
  - No improvement :
    - Binsbergen et al., 2015 and Erbe et al., 2016
  - Up to 5% improvement:
    - Brondum et al. 2015 and Vanraden et al., 2017

# Motivation

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



RESEARCH ARTICLE

Open Access



## Incorporation of causative quantitative trait nucleotides in single-step GBLUP

Breno O. Fragomeni<sup>1\*</sup>, Daniela A. L. Lourenco<sup>1</sup>, Yukata Masuda<sup>1</sup>, Andres Legarra<sup>2</sup> and Ignacy Misztal<sup>1</sup>

- ssGBLUP was able to reach accuracies close to 1 in simulation
  - Simulated QTN position and effects known
  - GWA estimated weights had limited impact
- GWA Methodology – no limitation in minimum and maximum weights (Zhang et al., 2016)

# Objective

- Test different SNP weighting methods in GBLUP and ssGBLUP in field data with the inclusion of causative variants

# Field Data

- 4M Records for Stature
- 3M Cows
- 4.6M Animals in pedigree
- $h^2=0.44$
- 27k Genotyped Sires
  - 54k SNP
  - 54k SNP + 17k Causative Variants (VanRaden et al., 2017)

# Analysis

- **GBLUP**

- Multi-step approach
- Daughter deviation as phenotypes
- Genomic Relationship Matrix
- Homogeneous or heterogeneous residual variance – different reliabilities

- **ssGBLUP**

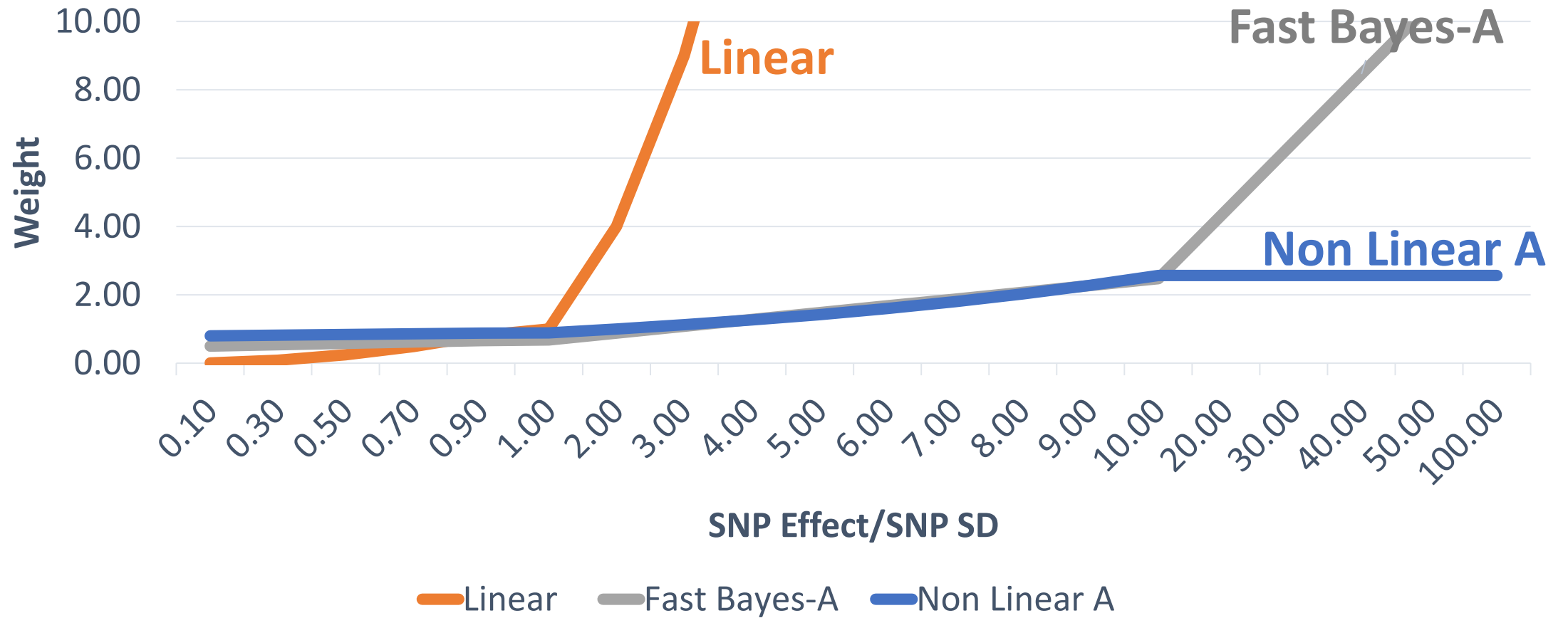
- Same model as national evaluation for type traits
- No deregressions
- Matrix combining pedigree and genomic information (**H**)

# Weighted genomic relationship matrix

$$\mathbf{G} = \mathbf{ZDZ}' \frac{\sigma_s^2}{\sigma_a^2} = \frac{\mathbf{ZDZ}'}{\sum_i 2p_i q_i}$$

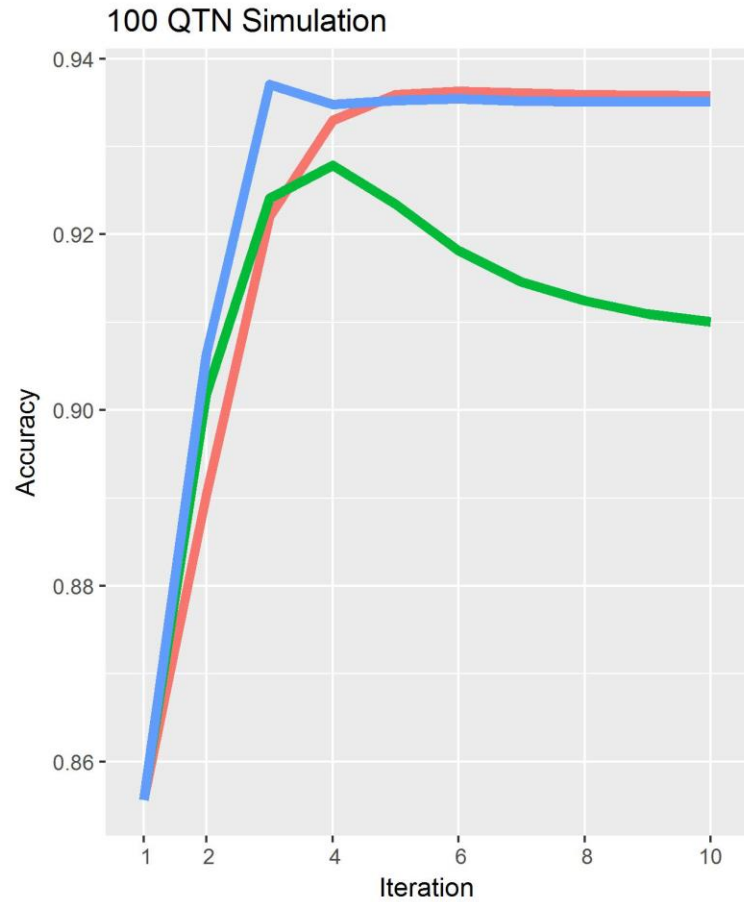
- Default
  - $\hat{d}_i = 1$ 
    - (VanRaden et al., 2008)
- Linear weights
  - $\hat{d}_i \sim \sigma_{SNP_i}^2$ 
    - (Zhang et al., 2010)
- Non-linear A weights
  - $\hat{d}_i = 1.125 \frac{|\hat{u}_i|}{sd(u)}^{-2}$ 
    - Value capped at 10
    - (VanRaden et al., 2008)
- Fast-Bayes A
  - $\hat{d}_i = \frac{\hat{u}_i^2 + df * S^2}{df + 1}$ 
    - (Sun et al., 2012)

# Weight matrix elements



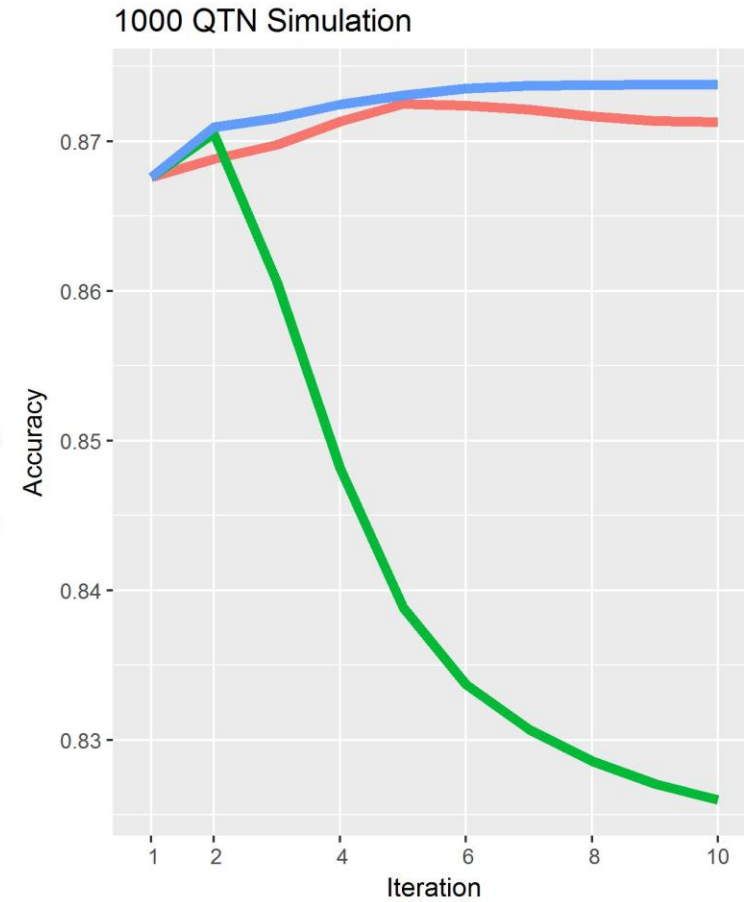


# Simulation results



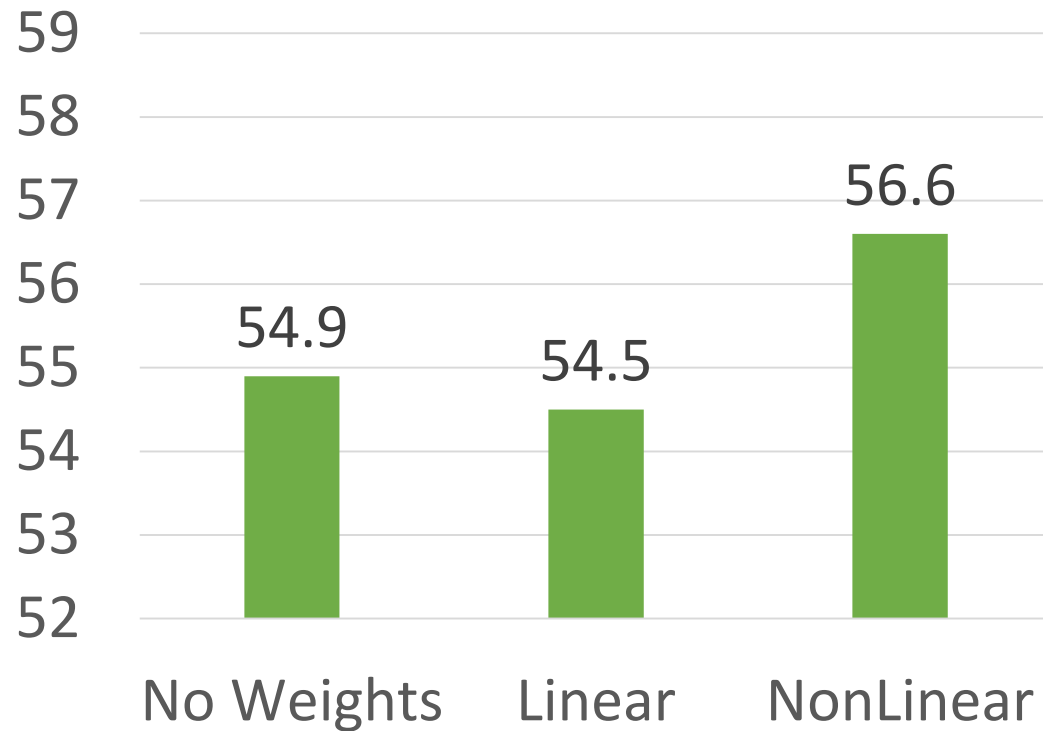
Weight\_type

- Fast Bayes A
- Linear
- Non-Linear A

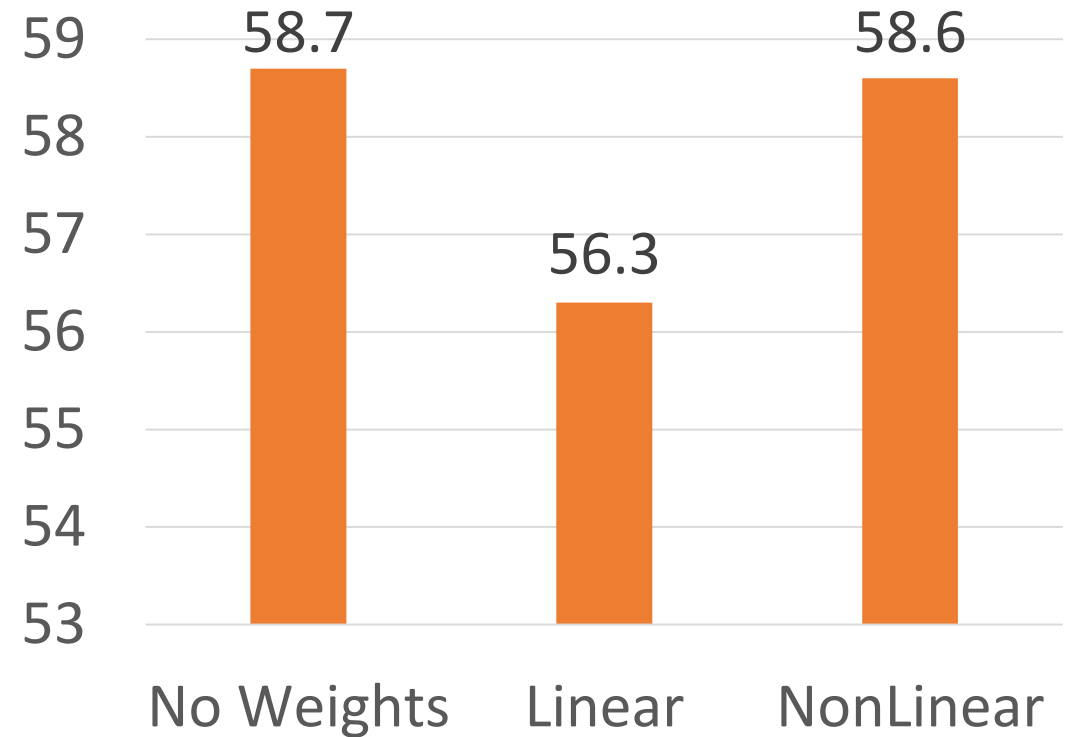


# GBLUP – 54K SNP - Reliabilities

## HOMOGENEOUS RESIDUAL VARIANCE

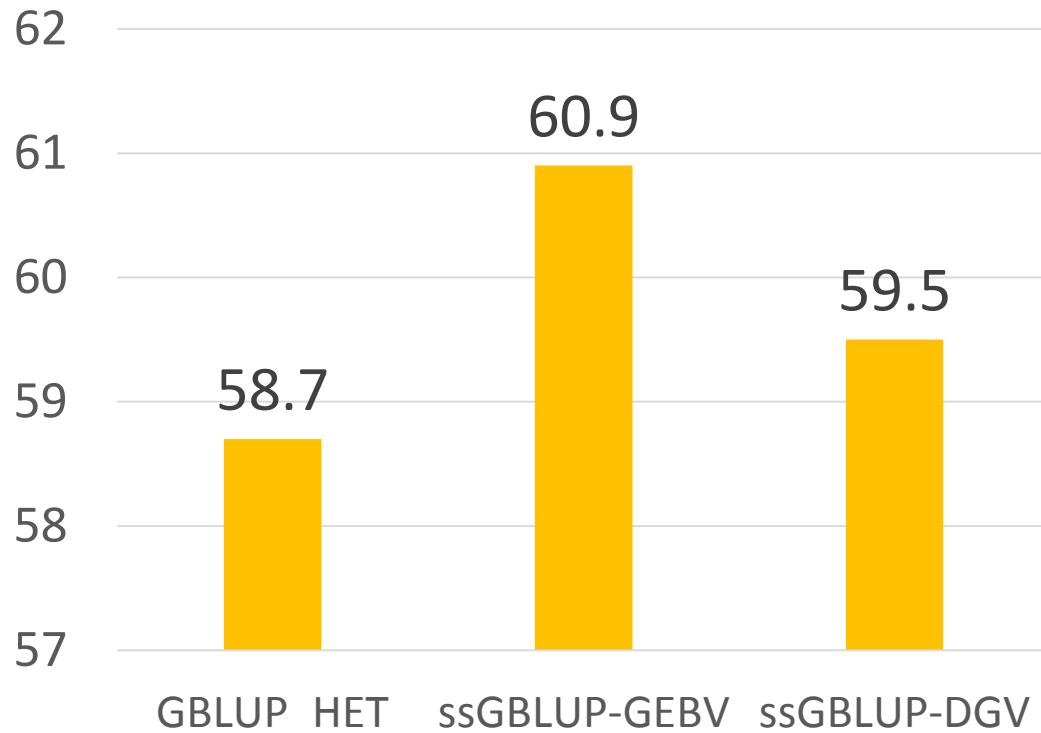


## HETEROGENEOUS RESIDUAL VARIANCE

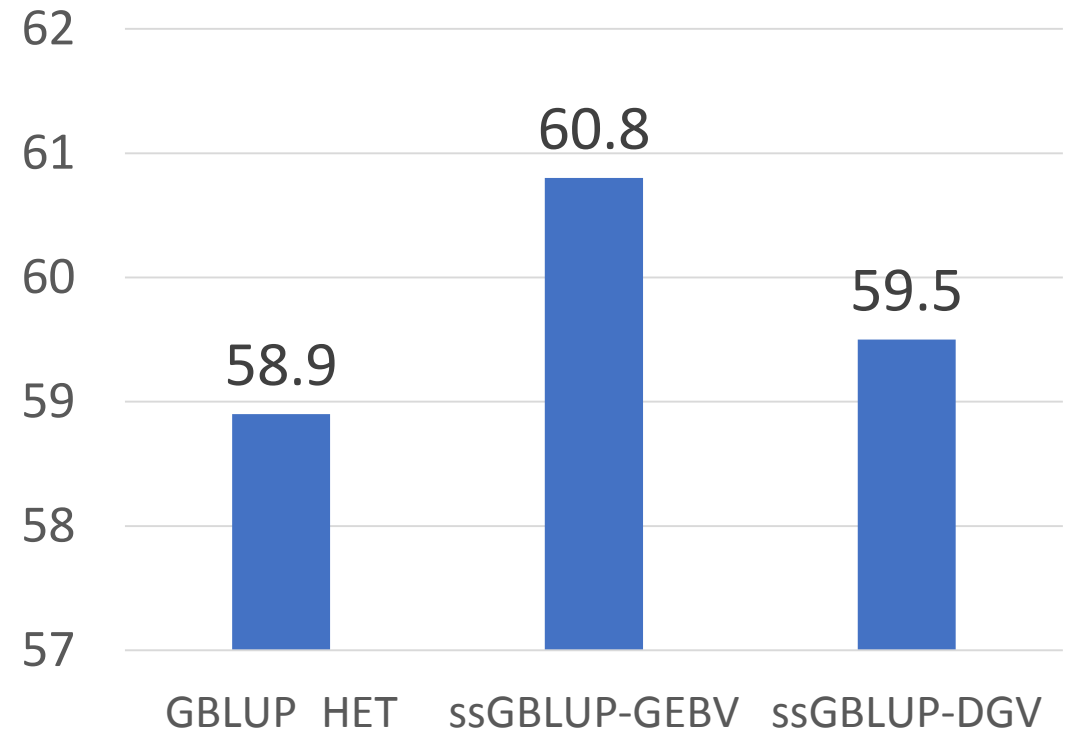


# GBLUP and ssGBLUP – 54K SNP - Reliabilities

## Unweighted GRM

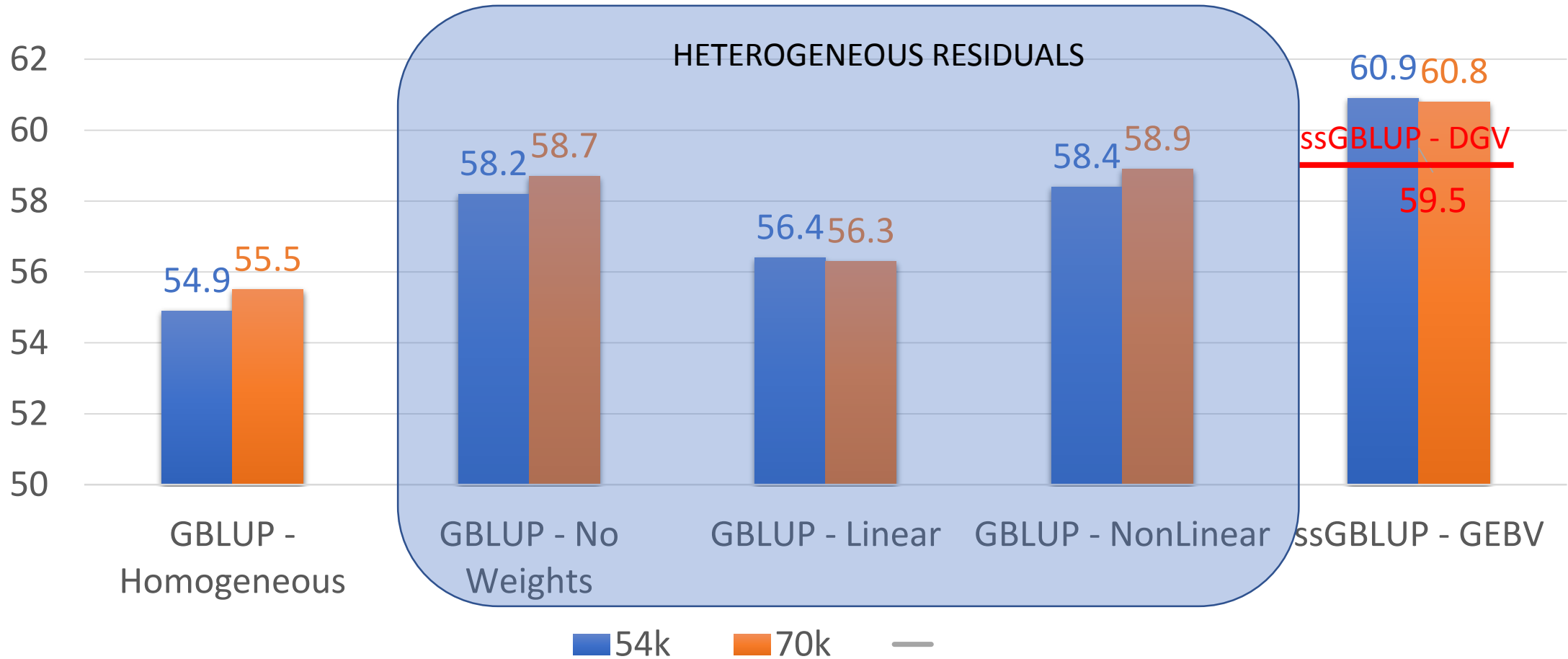


## Weighted GRM

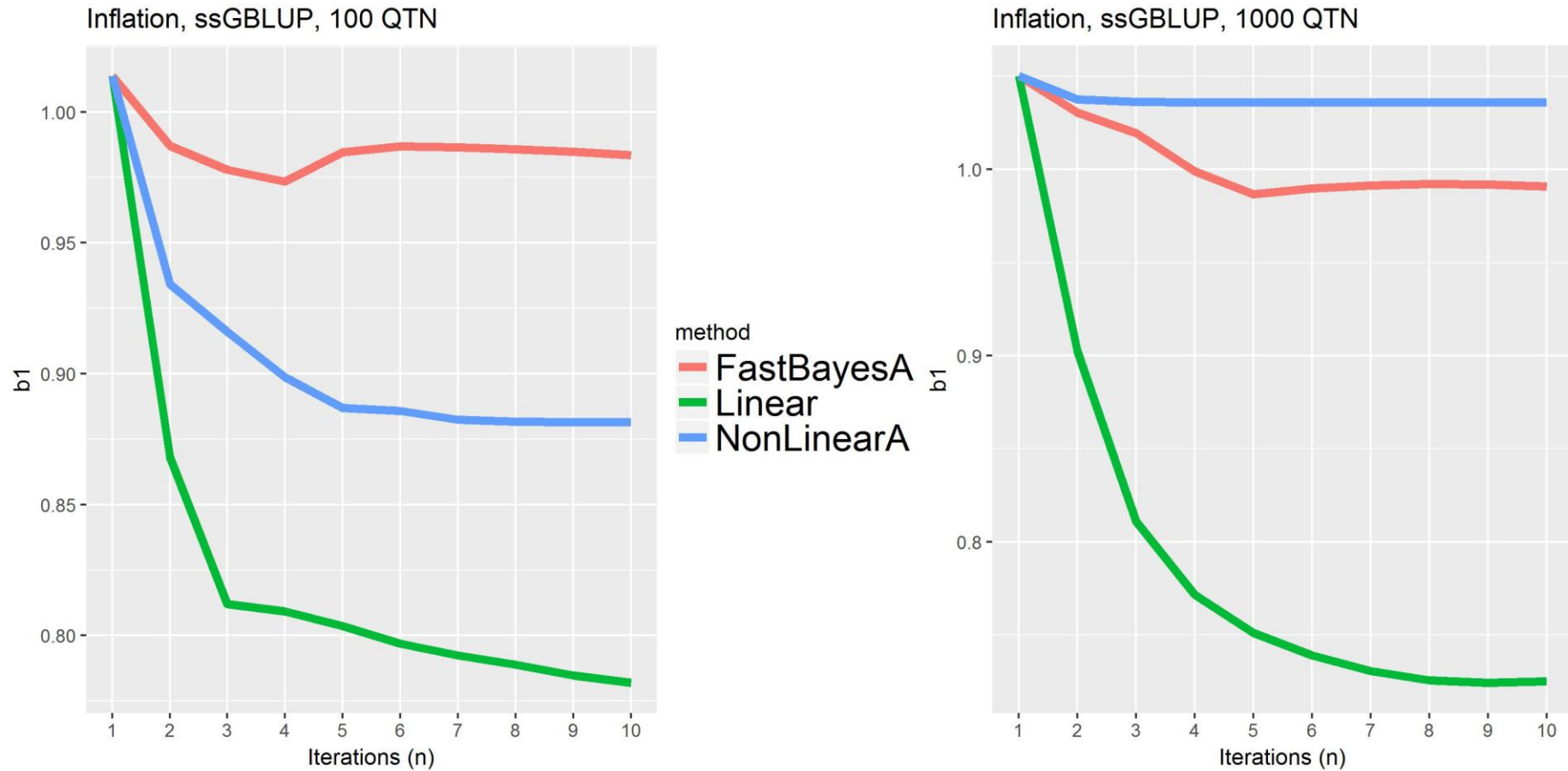


$$GEBV = w_1 PA + w_2 DGV$$

# Including causative variants



# Inflation coefficient: $b_1$



# Conclusions

- Gains with causative variants have more impact in GBLUP than in ssGBLUP
  - More data is used in single-step methodology, therefore impact of prior is less important
  - Sequence data might mask or fix methodology problems
- Non-linear methodology is better for weighting marker effects than linear weights
- Estimating weights in single-step GBLUP is still a research topic