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

Fragomeni BO[1], Lourenco DAL[1], Legarra A[2], Tooker ME [3], VanRaden PM[3], Misztal I[1]

[1]University of Georgia, Athens, USA

[2]INRA, Castanet-Tolosan, France

[3]AGIL ARS-USDA, Beltsville, USA

UNIVERSITY OF GEORGIA

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

- ssGBLUP was able to reach accuracies close to 1 with simulated causative variants
  - When priori used for weights were the simulated QTN effects
- GWA estimated weights had limited impact
  - GWA Methodology – no limitation in minimum and maximum weights

Objective

- Test different SNP weighting methods in GBLUP and ssGBLUP in field data which includes 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
- ssGBLUP
  - Same model as national evaluation for type traits
  - No deregressions
  - Matrix combining pedigree and genomic information (**H**)

Weighted Genomic relationship matrix

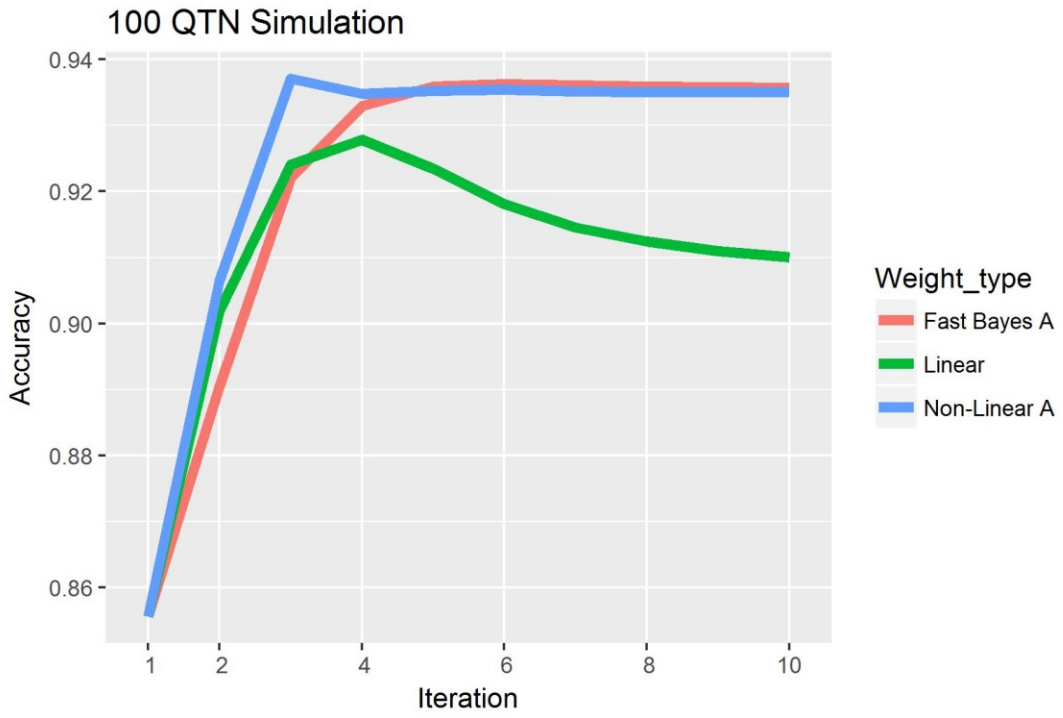
$$\mathbf{G} = \mathbf{ZDZ}' \frac{\sigma_s^2}{\sigma_a^2} = \frac{\mathbf{ZDZ}'}{\sum_i 2p_i\, q_i}$$

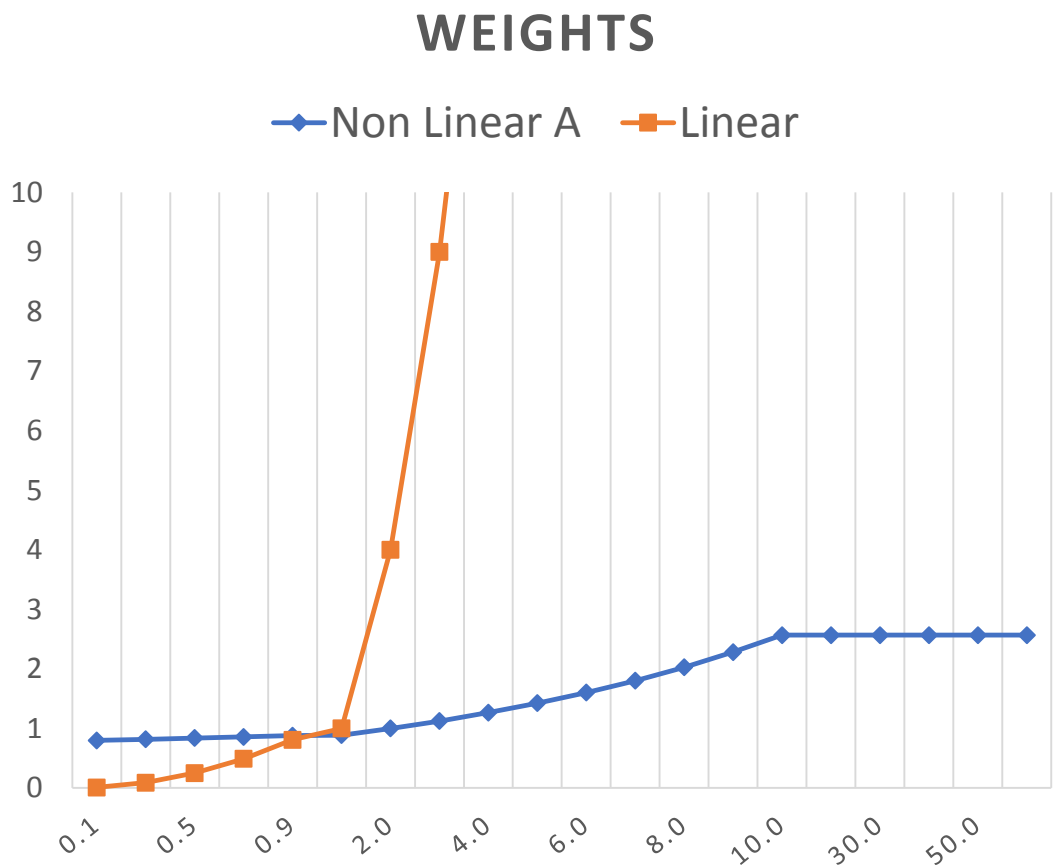
- Default
  - $\hat{d}_i = 1$
- Linear weights
  - $\hat{d}_i = \hat{u}_i^2$

- Non-linear A weights
  - $\hat{d}_i = 1.125^{\frac{|\widehat{u_i}|}{sd(u)}-2}$
    - Value capped at 10

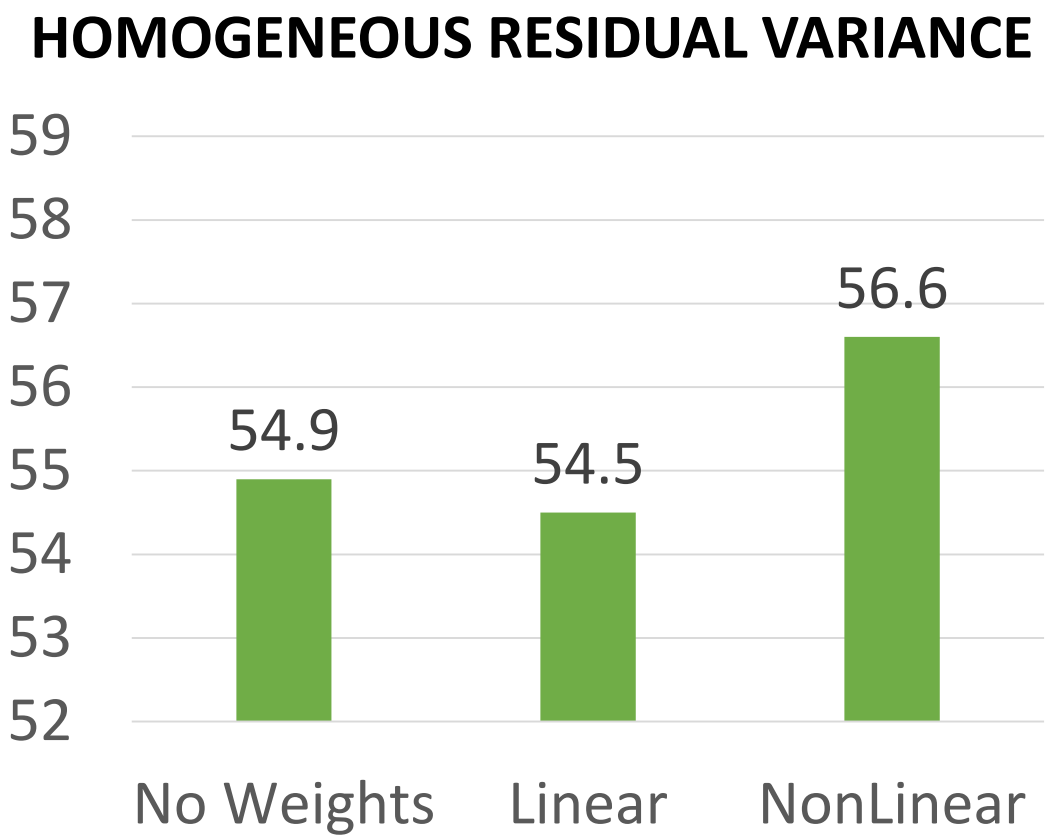
- Fast-Bayes A
  - $\hat{d}_i = \frac{SNP_{eff_i}^2 + df * S^2}{df+1}$

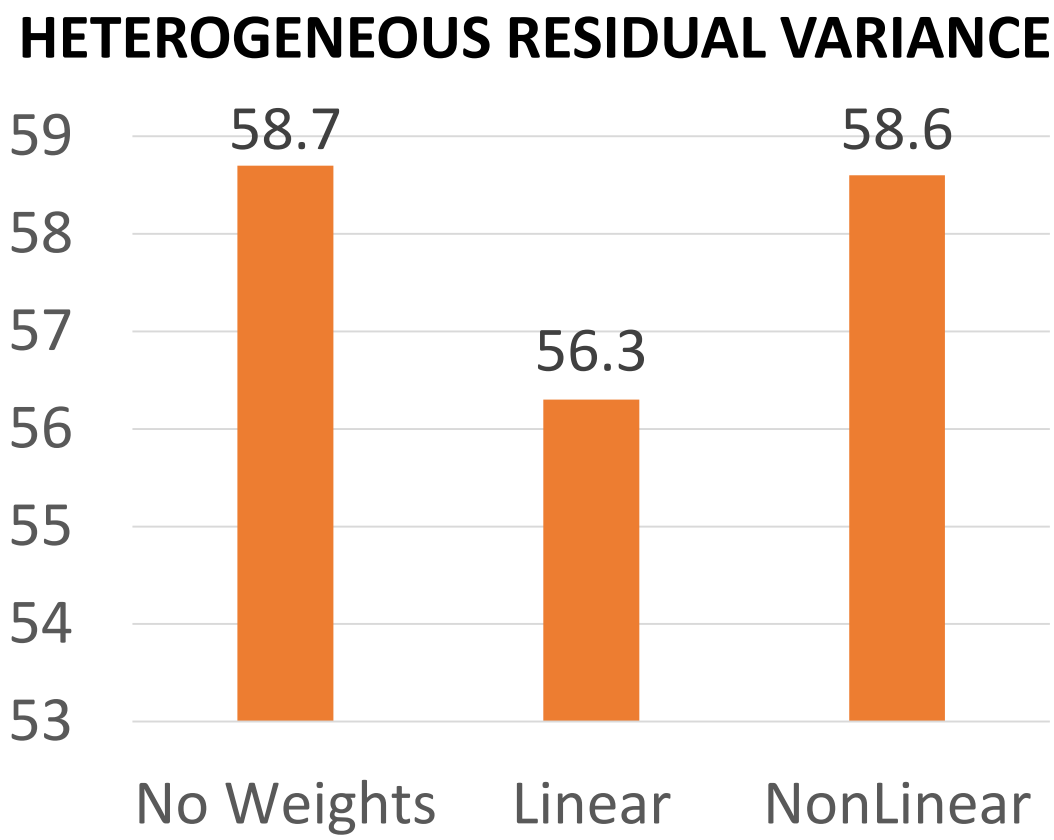
Weighted


100 QTN Simulation
Accuracy
0.94
0.92
0.90
0.88
0.86
1
2
4
6
8
10
Iteration
Weight_type
Fast Bayes A
Linear
Non-Linear A

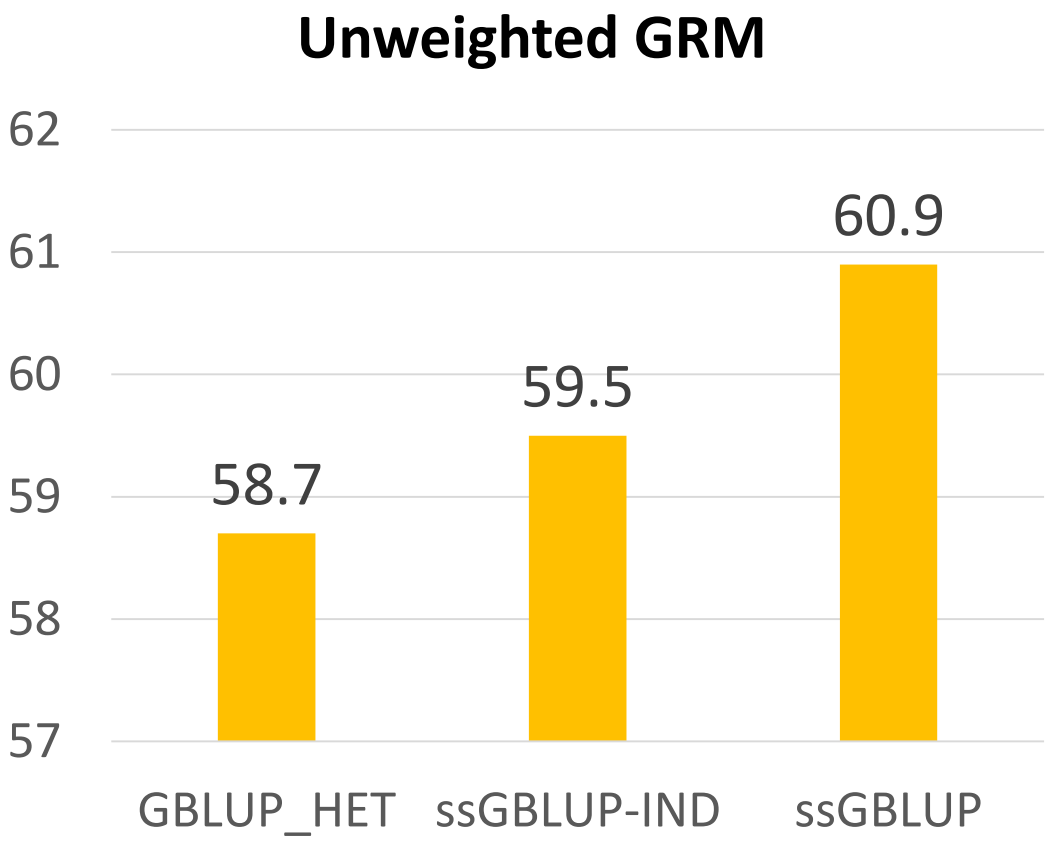

WEIGHTS
Non Linear A
Linear
10
9
8
7
6
5
4
3
2
1
0
0.1
0.5
0.9
2.0
4.0
6.0
8.0
10.0
30.0
50.0

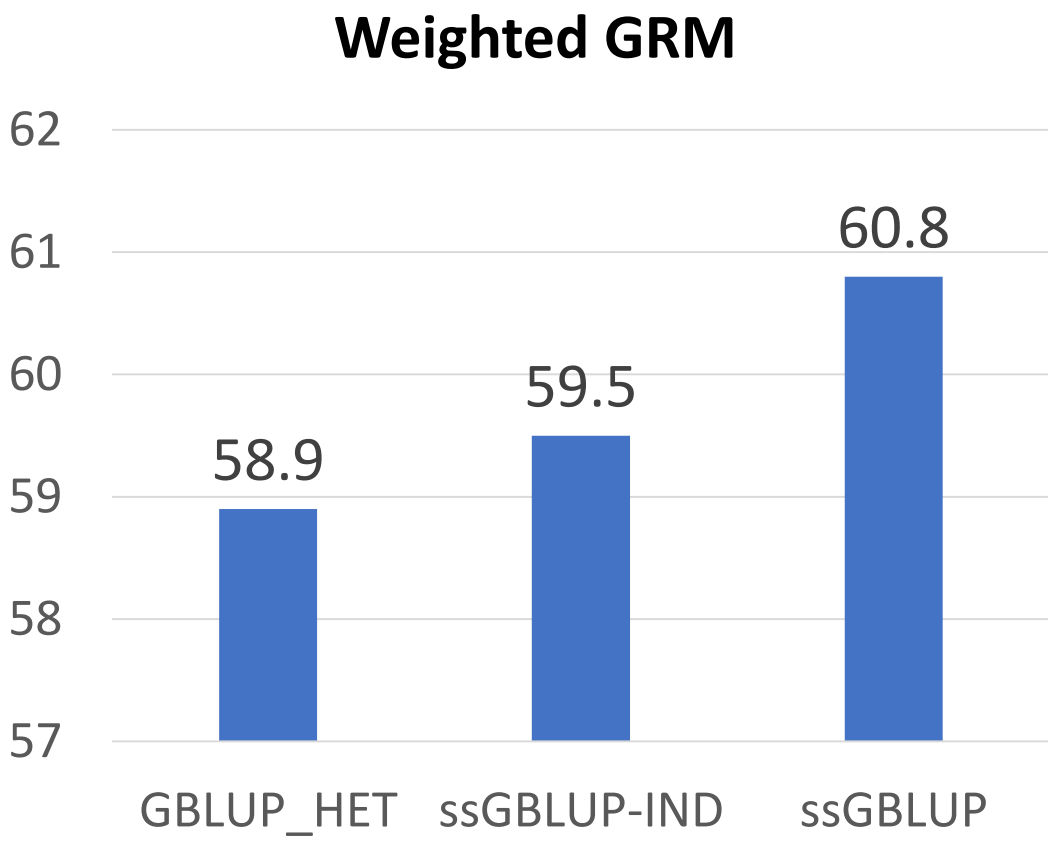
GBLUP – 54K SNP



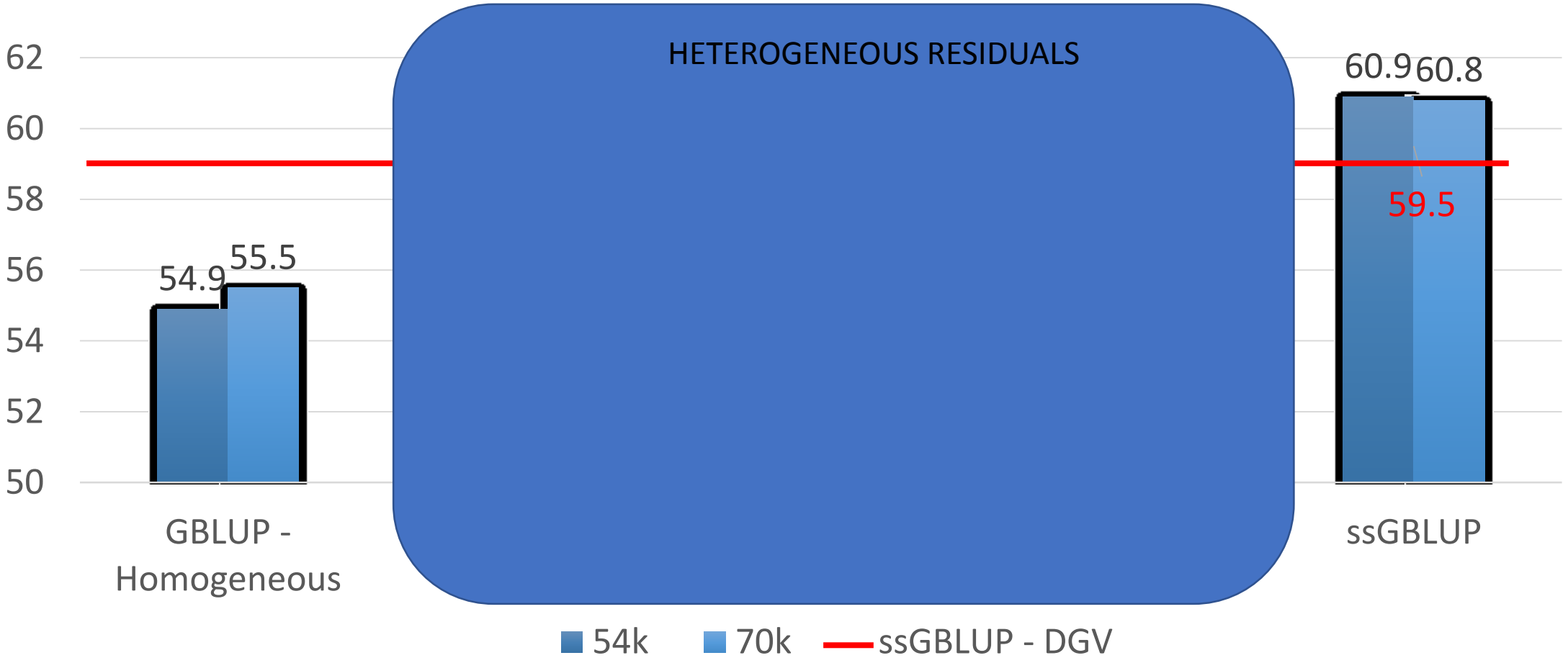


Weighted and unweighted reliabilities





Including causative variants



Conclusion

- Gains with inclusion of causative variants is limited when trait is polygenic
- Gains with causative variants has more impact in GBLUP than in ssGBLUP
  - More data is used in single-step methodology, therefore impact of prior is less important
- Non-linear methodology is better for weighting marker effects than linear weights