Improvement of genomic prediction accuracy for residual feed intake by prioritizing genetic markers identified by genome-wide association and gene expression

Sara de las Heras-Saldana*, Nasir Moghaddar, Samuel A. Clark, Julius van der Werf

School of Environmental and Rural Science, University of New England, Armidale, NSW Australia

*sdelash2@une.edu.au

Introduction

Genomic selection strategies applied to complex traits like residual feed intake (RFI) could improve selection for feed efficiency in livestock. The aim of this study was to test whether combining information from genome-wide association studies (GWAS) and expression of genes significantly associated (GSA) with RFI results could improve accuracy of genomic prediction.

Conclusion

The use of top SNPs in combination with selected SNPs located inside GSA reduced the bias in prediction compared with using only top SNPs from the independent GWAS set and slightly increased the accuracy of prediction by 2%. Genomic prediction accuracy can be improved when using selected SNPs from GWAS and GSA in larger datasets.

Methodology

Data

A total of 2,190 Angus steers with estimated RFI and imputed genotypes (50k and 770k) were uses. Further information about the dataset, calculation of RFI and imputation is found in [1]. The RFI was adjusted for contemporary groups (CG) as $y_{ij} = CG_i + e_j$

Sub-setting
Comparing two ways of cross-validation (CV)

4x4CV
training: 1,268
validation: 424
Independent GWAS: 500

4CV
GWAS-training: 1,644
validation: 546

Selection of SNPs
1) GWAS (700k)
Top-SNPs: $-\log_{10}(p\text{-value}) > 3.5$
Remove SNPs in LD ($r^2 > 0.95$)
2) SNP located on QTLs regions from genes which expression associated with RFI (GSA)

Prediction
GBLUP
$$y = \mu + Zg_1 + e$$
$$y = \mu + Zg_1 + Zg_2 + e$$
g_1: 50k;
g_2: top-SNPs or GSA-SNPs

Accuracy of prediction
Correlation:
r(GBV,adj.RFI)

Bias:
Bias: 1-b
$b \rightarrow y \sim a + bx$

Results

- Moderate heritability was observed for 50k and 770k (0.26 and 0.27).

- The number of top SNPs ($-\log 10(p\text{-value})$) was larger for the 4CV strategy than the 4x4CV (Table 1). The number of SNPs selected based on the gene expression information were from 2,439 to 23.

Table 1. Average number of SNP from the GWAS results and after removing SNPs in high LD ($r^2 \geq 0.95$) (Top) & the combination of Top-SNPs

TOP $_{-\log10(p\text{-value})}$	4CV		4x4CV	
	GWAS	Top	GWAS	Top
Top$_{1.3}$	29,384	2,475	29,179	846
Top$_2$	5,856	266	5,888	252
Top$_3$	652	102	611	94
Top$_4$	111	37	65	39
Top$_5$	27	10	8	5

Accuracy of prediction

- Higher accuracy was obtained for 4CV than the 4x4CV for 50k, 770k and Top$_S$ SNPs (Figure 1). The maximum benefit of using top SNPs was of 0.01 and 0.02 for 4CV and 4x4CV, respectively, compare to the accuracy obtained using the 50k.
  - Larger bias was observed by using the 4CV strategy when 50k+Top SNPs were used compare with using only one GRM. Smaller bias (0.14) was obtained when GRM from SNP pre-selected from GSA and 50k were used with an increase of 0.01 and 0.02 in accuracy of prediction than 50k or 770k respectively.
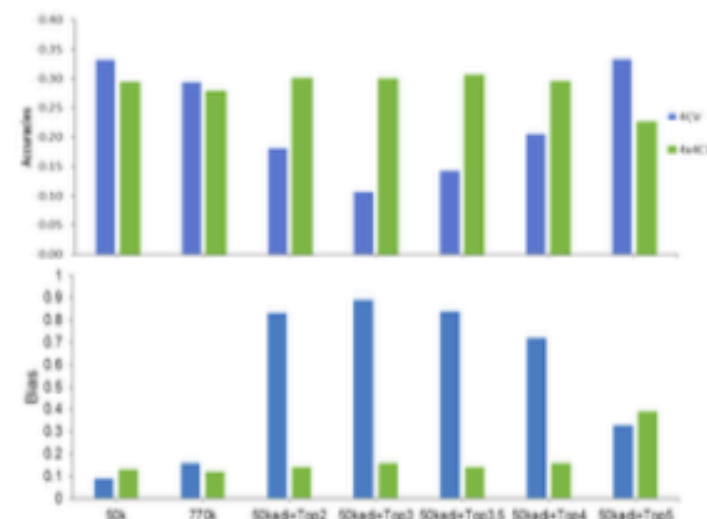

Figure 1. Average accuracies of prediction and bias for RFI breeding values using the GRM from the 50K, 770k or top SNPs. Strategies 4CV & 4x4CV. In x axis, the Top2 refers to SNPs with $-\log_{10}(p\text{-value}) > 2$.

Reference:
[1] de las Heras-Saldana, S., Clark, S. A., Duijvesteijn, N., Gondro, C., van der Werf, J. H. J., Chen, Y (2019). Combining information from genome-wide association and multi-tissue gene expression studies to elucidate factors underlying genetic variation for residual feed intake in Australian Angus cattle. BMC Genomics 20, 939.