ICQG6 icqg6.org
SIXTH INTERNATIONAL CONFERENCE OF QUANTITATIVE GENETICS
First two weeks of November 2020

華南農業大學
South China Agricultural University

Improving accuracy of genomic prediction by fitting epistasis

Jinyan Teng[1], Ning Gao[2], Shaopan Ye[1], Jiaqi Li[1] & Zhe Zhang[1]*

[1] Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou, Guangdong 510642, China

[2] State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou, Guangdong 510006, China

1. INTRODUCTION

Genomic prediction (GP), first proposed in 2001[1], has been widely used for predicting the genetic or phenotypic value of complex trait. Though numbers of genomic prediction studies based on additive genetic effects (A) archived a considerable accuracy, alongside the availability of big data, it is potential to include genetically interactive effects (D and I) in genomic prediction model. Since the biological interaction of gene sets from KEGG pathway was publicly accessible, incorporating the KEGG as a carrier of interaction (i.e. epistasis[2]) into genomic prediction model may be a method potentially improving accuracy of genomic prediction.

2. MATERIALS AND METHODS

In the present study, genome-enable best linear unbiased prediction (GBLUP)[3] and a pathway model (PGBLUP) that polygenes effect combined a separate random component matrix from KEGG in model were validated in a German Holstein population. GBLUP model can be expressed as:

$$\mathbf{y} = \mu + \mathbf{Zg} + \boldsymbol{\epsilon},$$

and PGBLUP can be expressed as:

$$\mathbf{y} = \mu + \mathbf{Zg} + \mathbf{Zp} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is the observations, μ is the overall mean, $\mathbf{Z}$ is a design matrix allocating observations to genetic values, $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$ is the polygenic genetic values, $\mathbf{p} \sim N(0, \mathbf{G_p}\sigma_p^2)$ is the genetic values captured by variants in the KEGG pathway, in which $\mathbf{G}$ and $\mathbf{G_p}$ matrix were constructed referring to VanRaden (2008)[3], and $\boldsymbol{\epsilon} \sim N(0, \mathbf{I}\sigma_\epsilon^2)$ is the residual error and $\mathbf{I}$ is the identity matrix.

The cattle population used in this study includes 2,000 genotyped individuals with 54K SNPs and six traits with highly reliable estimated breeding values. Model assessment was performed by 10 times 10-fold cross-validation. Predictive accuracy was defined as the Pearson's correlation coefficient between the predicted genetic values ($\mathbf{Z}\hat{\mathbf{g}}$ for GBLUP and $\mathbf{Z}\hat{\mathbf{g}} + \mathbf{Z}\hat{\mathbf{p}}$ for PGBLUP) and phenotypic values ($\mathbf{y}$) in the validation group.

3. RESULTS

The predictive accuracy (Figure 1) of PGBLUP incorporating the best KEGG pathway outperformed GBLUP in six tested traits with the relative advantage of from 0.4% to 15%.

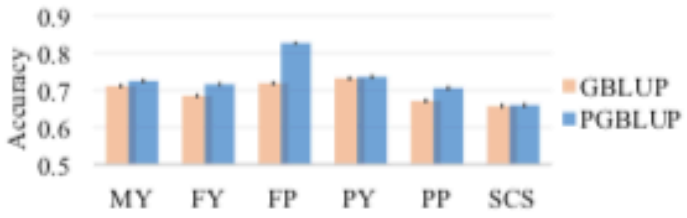

Figure 1. Predictive accuracies of all models for each trait in the cattle population. *MY, milk yield; FY, fat yield; FP, fat percentage; PY, protein yield; PP, protein percentage; SCS, Somatic cell score.*

In addition, the best pathway (Table 1) used in the PGBLUP model is Amyotrophic lateral sclerosis pathway (bta05014) for SCS and p53 signaling pathway (bta04115) for other five tested traits.

Table 1. Description of the best KEGG pathway in PGBLUP model.

KEGG	No. of genes (represented[1]/total[2])	No. of SNPs[3]
bta04115	34 / 69	59
bta05014	25 / 48	60

[1] The number of genes represented in the PGBLUP; [2] total number of genes in the KEGG pathway; [3] the number of SNPs mapped in the KEGG pathway.

4. CONCLUSIONS

This study concluded that KEGG pathway can act as a carrier of interaction in genomic prediction model, and there is always an optimal pathway can improve the predictive accuracy in the tested cattle population.

ACKNOWLEDGMENTS

We thank the Vereinigte Informationssysteme Tierhaltung w.V. for providing the German Holstein data.

REFERENCES

[1] Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1820.

[2] Bateson, W. 1909. Mendel's Principles of Heredity. Cambridge University Press, Cambridge.

[3] VanRaden P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423.

Email: Jinyan Teng – kingyan312@live.cn; Zhe Zhang* (corresponding author) – zhezhang@scau.edu.cn