Integrating gene expression data into Genomic prediction

Zhengcao Li[1], Ning Gao[2], Johannes W. R. Martini[3], Henner Simianer[1]

[1]University of Goettingen, Department of Animal Sciences, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Goettingen, Germany

[2]National Engineering Research Center for Breeding Swine Industry, College of Animal Science, South China Agricultural University, Guangzhou, China

[3]KWS SAAT SE, Einbeck, Germany.

My project overview

(1) to integrate transcriptome data with whole-genome sequence data for genomic-transcriptomic prediction

(2) to assess whether our statistical models can capture substantial proportions of phenotypic variance explained by transcriptome data

(3) to test whether accounting for transcriptome data can improve phenotype prediction.

Materials

Drosophila Genetic Reference Panel (DGRP)

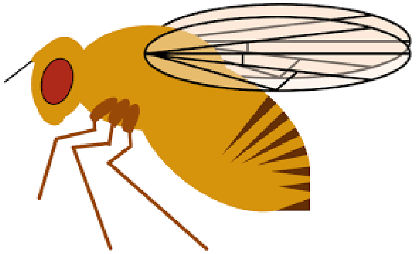
- **Availability of Supporting Data**: http://dgrp2.gnets.ncsu.edu
- *Drosophila* Whole-Genome Sequence Data
- *Drosophila* Gene Expression Data
- *Drosophila* Phenotype Data

Whole-Genome Sequence Data

- 2’863’909 SNPs from 185 fully sequenced inbred lines

Expression Data

- The abundances of RNA products of 18’140 genome-wide annotated genes and novel transcribed regions (NTRs) in 185 DGRP lines

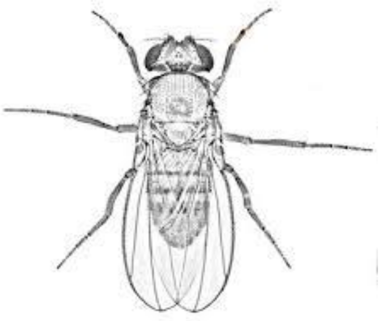


Phenotype data

(all phenotypes are based on the line mean of several observations)

9 traits

1. Startle response (SR)
2. Starvation resistance (SRE)
3. Food intake (FI)
4. Alcohol sensitivity and tolerance (AST)



Olfactory perceptions to 5 chemical odorants:

5. Olfactory perceptions to Ethyl Butyrate (OPEB)
6. Olfactory perceptions to 2-Heptanone (OP2H)
7. Olfactory perceptions to Methyl Salicylate (OPMS)
8. Olfactory perceptions to l-Carvone (OPlC)
9. Olfactory perceptions to 1-Hexanol (OP1H)

Statistical models

methods	data
GBLUP	Whole-Genome Sequence Data
TBLUP (linear kernel)	Gene Expression data
Reproducing Kernel Hilbert Space Regression (RKHSR) (Gaussian kernel)	Gene Expression data
GTBLUP (GBLUP + TBLUP)	Whole-Genome Sequence Data Gene Expression data
GRBLUP (GBLUP + RKHSR)	Whole-Genome Sequence Data Gene Expression data

Comparison of predictive abilities

- Pearson correlation coefficients between predicted phenotypes and observed phenotypes.

- The different approaches were assessed using 20 replicates of a 5-fold cross-validation (CV).

- The final predictive ability of each model was the mean of the prediction accuracies across 100 estimates.

Estimation of omics-augmented broad sense heritability

(based on the between line effects)

The omics-augmented broad sense heritability is defined as the proportion of phenotypic variance explained by whole genome SNP marker effects and other utilized omics data effects.

$$\widehat{H}_o^2 = \frac{\hat{\sigma}_g^2 + \hat{\sigma}_{omics}^2}{\hat{\sigma}_g^2 + \hat{\sigma}_{omics}^2 + \hat{\sigma}_e^2}$$

(1) SNP-based genomic narrow sense heritability

for GBLUP ($\hat{h}_G^2$)

$$\hat{h}_G^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2+\hat{\sigma}_e^2}$$

(2) SNP and gene expression data-augmented broad sense heritability

for GTBLUP ($\hat{H}_{GT}^2$)

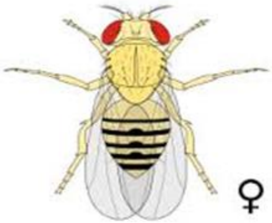
$$\hat{H}_{GT}^2 = \frac{\hat{\sigma}_g^2+\hat{\sigma}_t^2}{\hat{\sigma}_g^2+\hat{\sigma}_t^2+\hat{\sigma}_e^2}$$

for GRBLUP ($\hat{H}_{GR}^2$)

$$\hat{H}_{GR}^2 = \frac{\hat{\sigma}_g^2+\hat{\sigma}_v^2}{\hat{\sigma}_g^2+\hat{\sigma}_v^2+\hat{\sigma}_e^2}$$

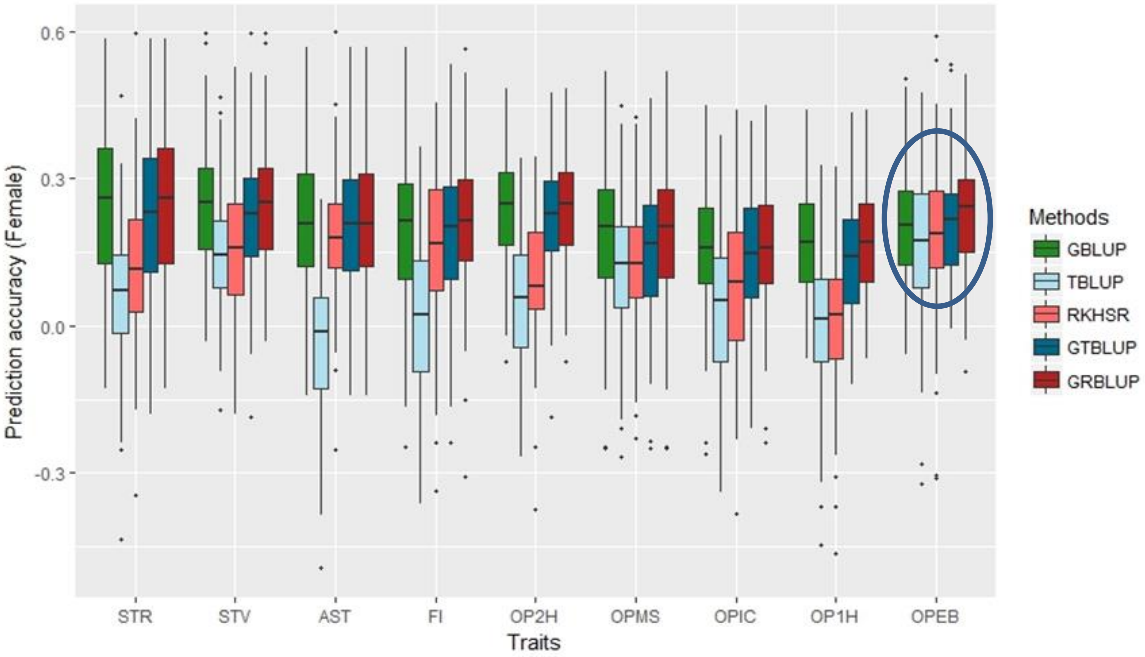
The variance components $\hat{\sigma}_g^2$, $\hat{\sigma}_t^2$, $\hat{\sigma}_v^2$, $\hat{\sigma}_e^2$ were estimated from the entire data sets.

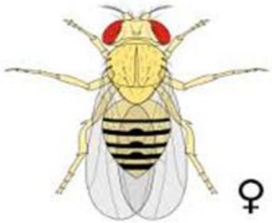
Results



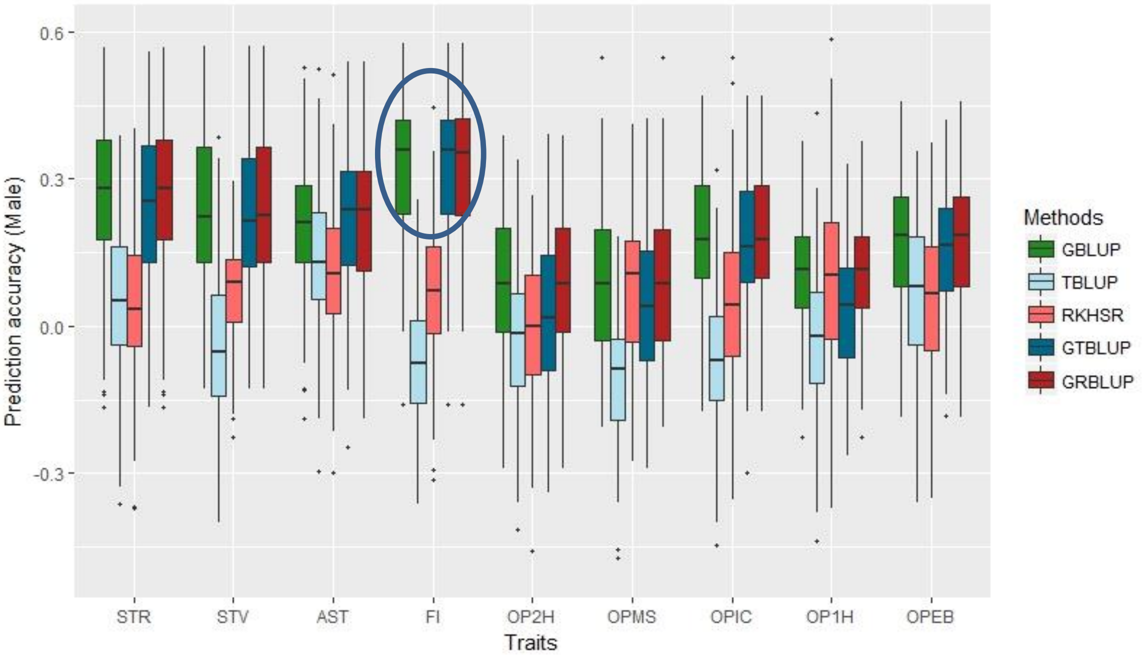
Predictive abilities for 9 traits with 5 statistical models in females

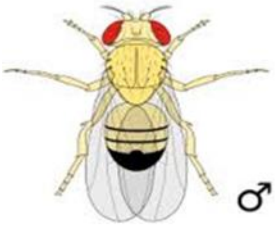
Results



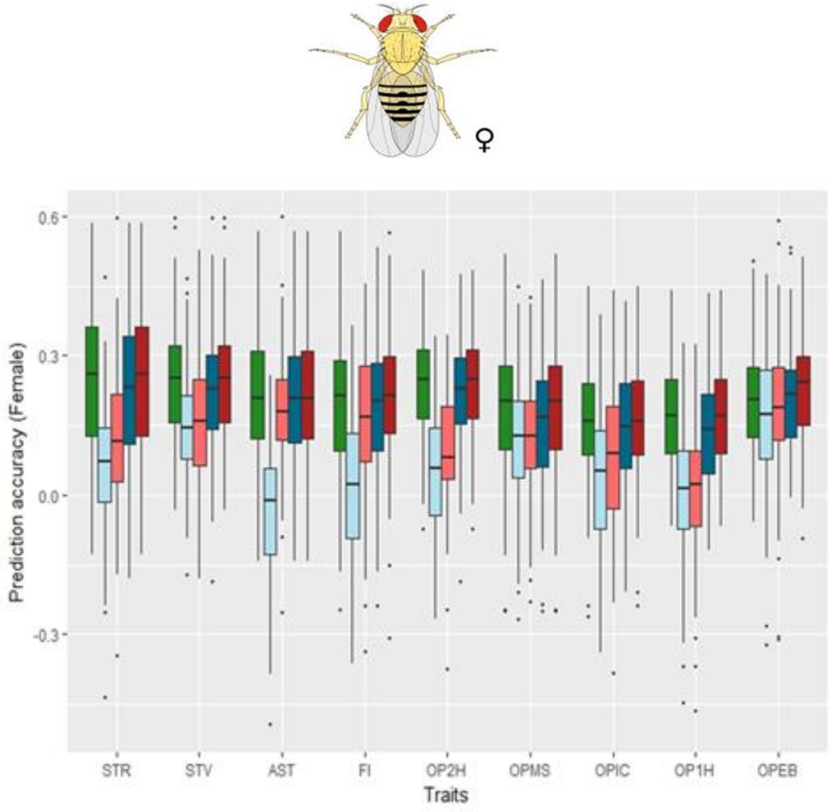


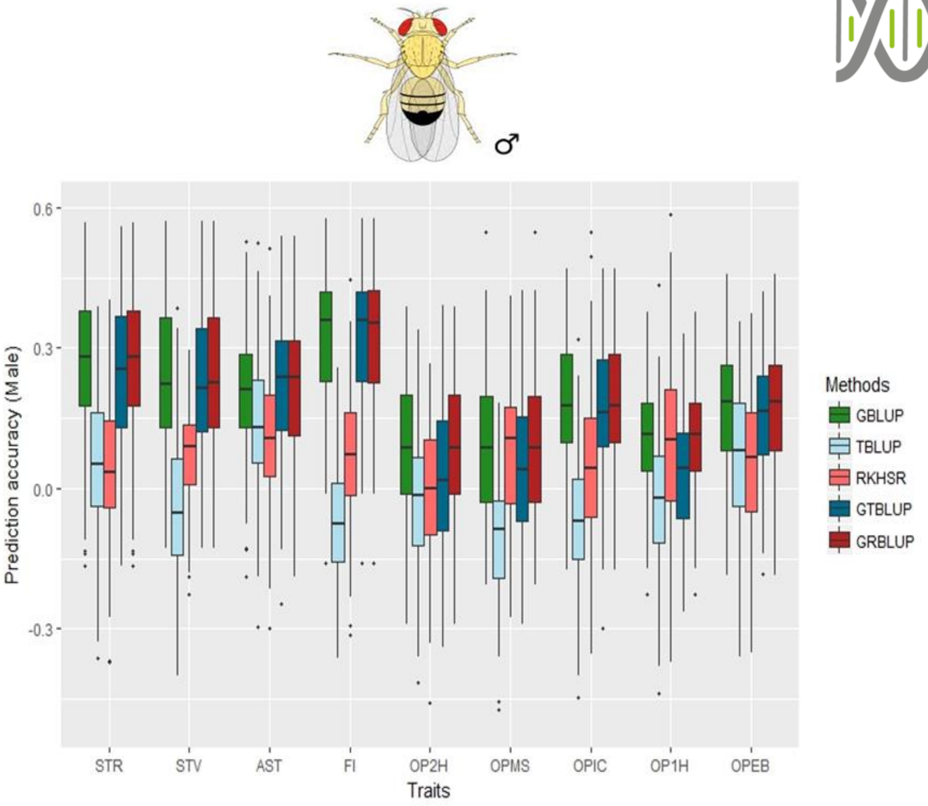
Predictive abilities for 9 traits with 5 statistical models in females





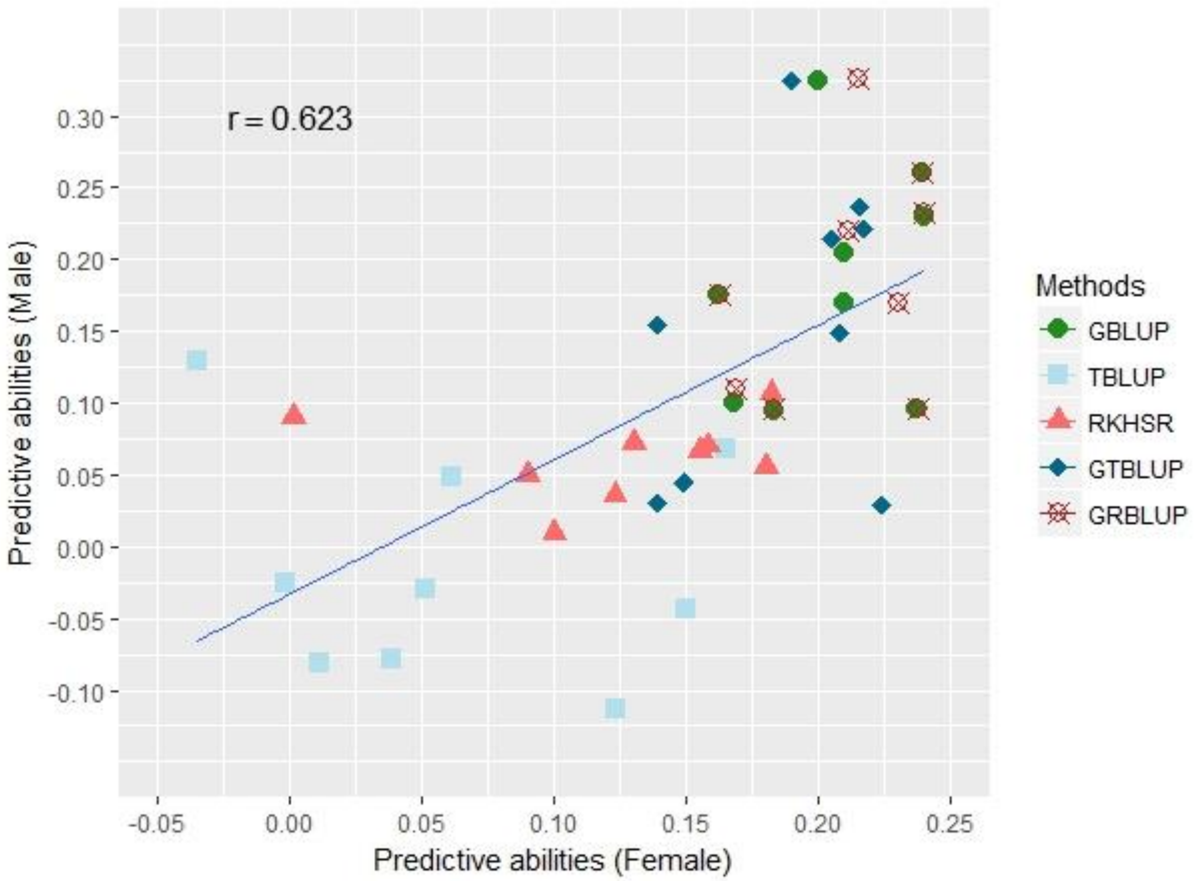
Predictive abilities for 9 traits with 5 statistical models in males



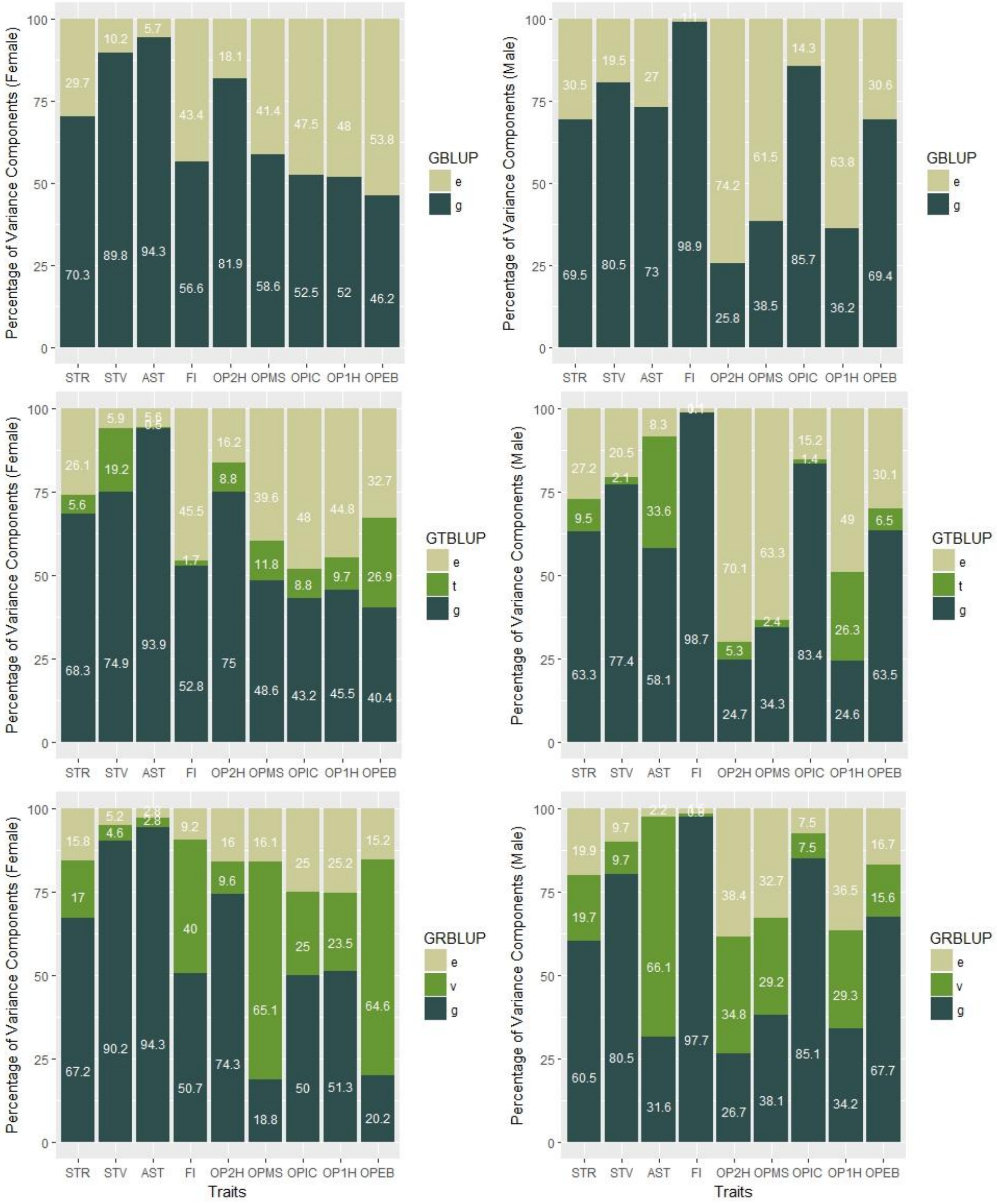


Comparison of predictive abilities between females and males for 9 traits with 5 statistical models

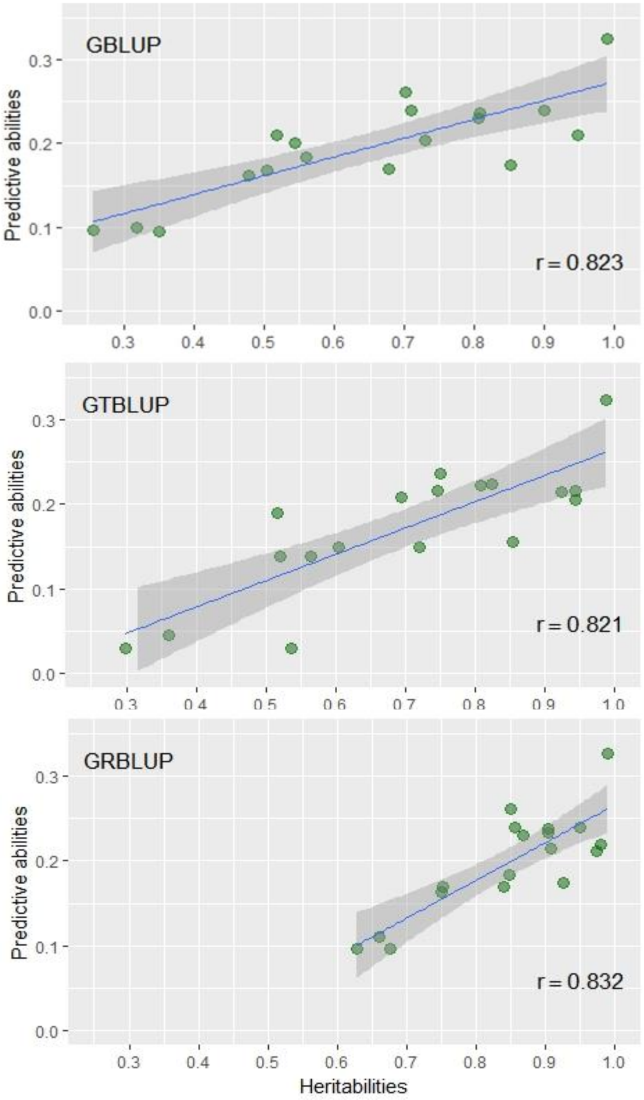
Correlation of predictive abilities between females and males across 9 traits and 5 statistical models



Percentage of variance components for GBLUP, GTBLUP and GRBLUP on 9 traits for females (left) and males (right)



The correlation between predictive abilities and heritabilities for GBLUP, GTBLUP, GRBLUP across all traits and both sex



Conclusion

- GRBLUP and GBLUP provided similar predictive ability, but GRBLUP could capture more phenotypic variance components explained by transcriptome data.
- GRBLUP has excellent goodness of fit with transcriptome data. The better goodness of fit of GRBLUP in general did not translate into a better predictive ability in most traits.
- The sample size was (too) small
- Gene expression was not measured at a time point and in a specific tissue which was functionally linked to the trait of interest.
- We suspect that adding more specifically collected transcriptome data has the potential to improve genomic predictions in larger scale applications.

Acknowledgements

Thanks China Scholarship Council and EAAP for financial support.

Thanks for listening