Supporting Information for

DeepAnnotation: A novel interpretable deep learning-based genomic selection model that integrates comprehensive functional annotations

Wenlong Ma[1,2], Shenghua Qin[1,2], Weigang Zheng[1,2,4], Chao Wang[1,2], Bowen Lei[1,2], and Yuwen Liu[1,2,3,*]

[1]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-Omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China.

[2]Innovation Group of Pig Genome Design and Breeding, Research Centre for Animal Genome, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China.

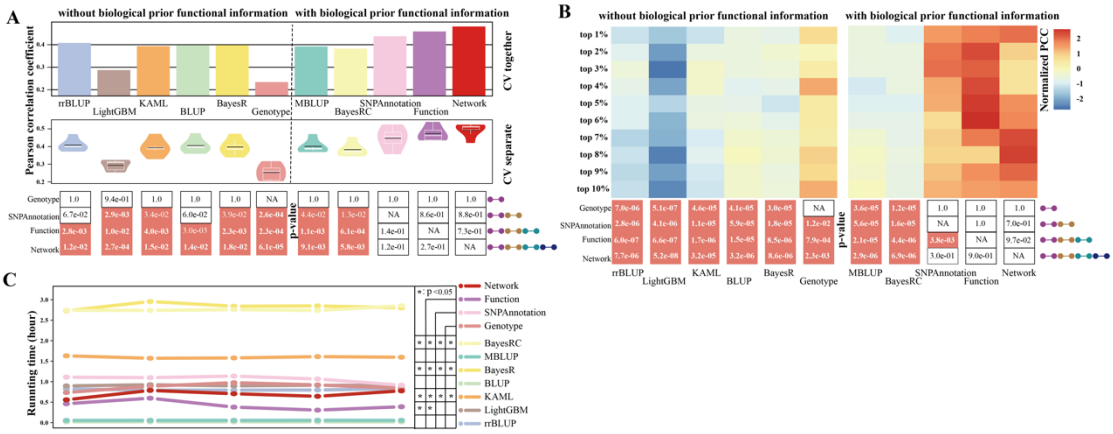
[3]Kunpeng Institute of Modern Agriculture at Foshan, Chinese Academy of Agricultural Sciences, Foshan 528226, China.

[4]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education & Key Lab of Swine Genetics and Breeding of Ministry of Agriculture and Rural Affairs, Huazhong Agricultural University, Wuhan 430070, People's Republic of China

[*]Correspondence address: Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-Omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China; E-mail: liuyuwen@caas.cn

Analyses

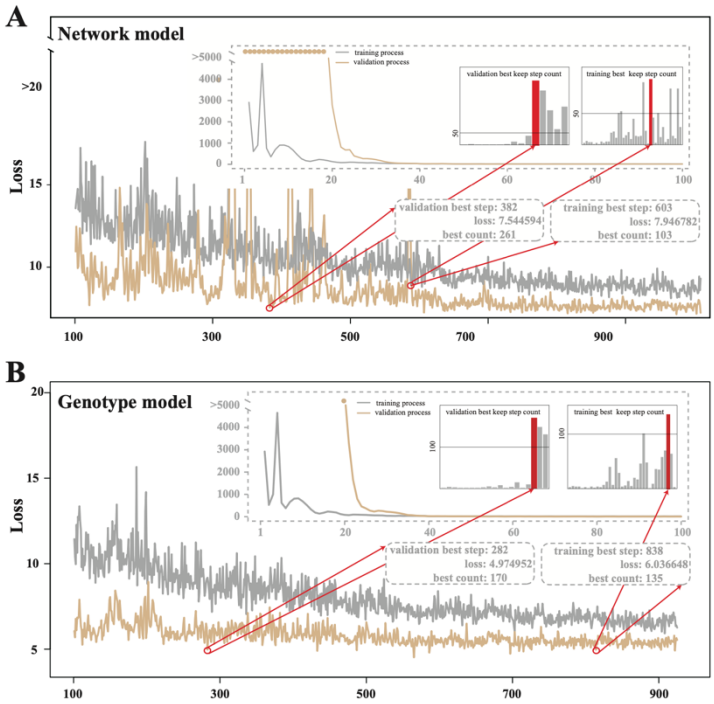
Dissecting the genetic basis of complex traits with DeepAnnotation



Support Figure 1. Prediction performance of DeepAnnotation through cross-validation compared with rrBLUP, LightGBM, KAML, BLUP, BayesR, MultiBLUP, BayesRC, and DeepAnnotation on LMP trait. (A) PCC evaluation of different models based on 5-fold cross-validation experiment. In the "CV together" approach, each testing fold from the 5-fold cross-validation were consolidated before metric computation. On the other hand, in the "CV separate" approach, metrics were computed separately for each testing fraction of the 5-fold cross-validation. The paired t-test *P*-values of DeepAnnotation compared with other models were displayed on the black boxes with pink represents significance with *P*-value < 0.05. (B) Spatial PCC values between the predicted and observed phenotypic values of top-ranked samples from top 1% to top 30%. The paired t-test *P*-values were displayed on the black boxes with pink represents significance with *P*-value < 0.05. (C) Elapsed training time of different models.

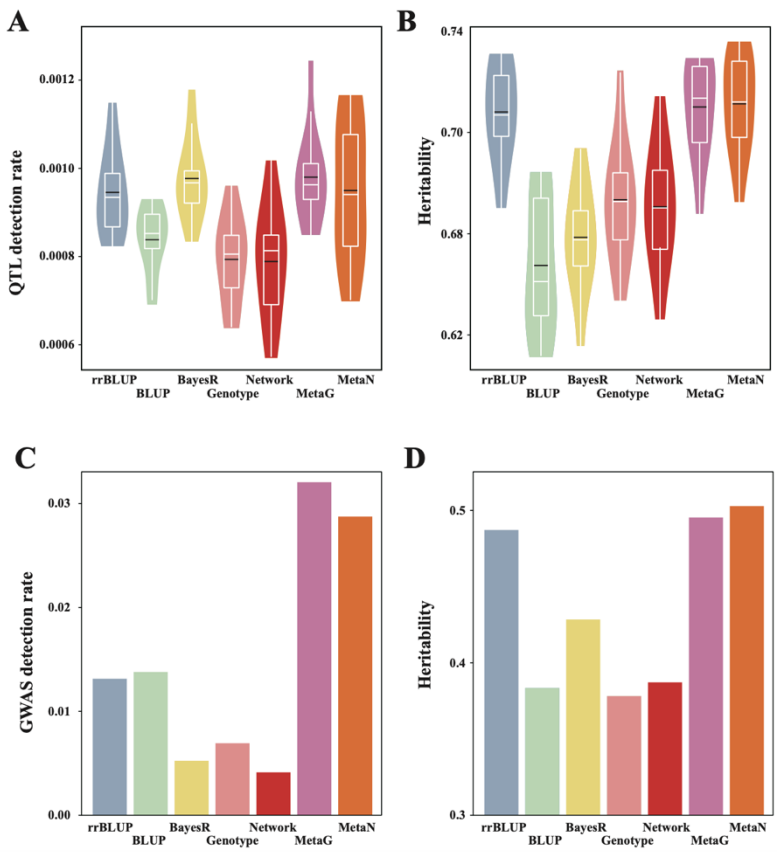
The prediction performance evaluations showed that as the number of functional annotations incorporated into the DeepAnnotation models increased, the accuracy gradually improved (Support Fig. 1). Since DeepAnnotation consistently outperformed other models in identifying individuals

with top-ranked phenotype values, we were interested in whether DeepAnnotation could correctly dissect the causal variants which potentially dominate the genetic variance.



Support Figure 2. Training and validation status of Network (A) and Genotype (B) models of each epoch based on the pre-trained result from 5-fold cross-validation.

Identifying causal variants is typically challenging. One way to do this is to simulate traits with pre-defined causal variants and check the detection rate. Therefore, we conducted this evaluation through 10 simulated traits with 10,000 pre-defined QTLs under 0.7 heritability (Methods). We used the Network and Genotype models to represent DeepAnnotation with and without functional annotations, respectively. For rigorous and consistent evaluation, for those models trained using 1,700 samples during 5-fold cross-validation, we tracked validation losses at each step and determined the optimal training epoch as the step after which the loss ceased to decrease further. The optimal training steps were 382 (Support Fig. 2A) and 282 (Support Fig. 2B) for the Network and Genotype models, respectively.



Support Figure 3. Dissecting the genetic basis of complex traits. (A) QTL detection rates distribution of different models on 10 simulated traits. (B) Estimated heritabilities distribution of significant SNPs (P-value < 0.05) identified by different models on 10 simulated traits. (C) GWAS significant SNPs (with adjusted P-value < 0.05) detection rates of different models on LMP. (D) Estimated heritabilities of significant SNPs (P-value < 0.05) identified by different models on LMP.

According to the performances and reasonable running time (Support Fig. 1B and 1D), we selected rrBLUP (an adjusted BLUP-based model), the original BLUP (no significantly different from MBLUP, paired t-test with *P*-value = 0.25) and BayesR (an adjusted Bayes-based model, not significantly different from BayesRC, paired t-test with *P*-value = 0.08) as baseline models. Across 10 simulated traits, the average QTL detection rates were 0.000793, 0.000838, 0.000946, and 0.000977 for Genotype, BLUP, rrBLUP, and BayesR, respectively (Support Fig. 3A). However, the

heritability explained by significant SNPs (*P*-value < 0.05) was 0.647, 0.658, 0.673, and 0.707 for BLUP, BayesR, Genotype, and rrBLUP, respectively (Support Fig. 3B). Notably, while the QTL detection rate for Genotype was the lowest, the heritability explained by its significant SNPs surpassed that of BayesR. This indicates that DeepAnnotation captures different aspects of genetic variance compared to other baseline models

To further evaluate the complementarity of DeepAnnotation with existing models, we employed a meta-analysis strategy to combine results. The metaG model (combining rrBLUP, BLUP, BayesR, and Genotype) achieved the highest QTL detection rate (0.000980, Support Fig. 3A) and the highest estimated heritability (0.709, Support Fig. 3B). Furthermore, we calculated these two indices for Network to test whether inveracious functional annotations still helpful. The result showed that the overall QTL detection rate was 0.000790 and heritability was 0.671, which were slightly lower than Genotype (Support Fig. 3A-3B). Besides, metaN (combination of rrBLUP, BLUP, BayesR, and Network) showed no significant improvement over metaG (QTL detection rate: 0.000950 vs 0.000980, heritability: 0.711 vs 0.709), highlighting the importance of comprehensive and authentic functional annotations.

For the real LMP trait, we further compared predicted causal variants with GWAS results, where SNPs with Bonferroni-adjusted *P*-value < 0.05 were assumed to be causal. The results aligned with those from simulated traits: for detection rates, they were 0.00413, 0.00524, 0.00693, 0.0131, 0.0138, 0.0288, and 0.0321 for Network, BayesR, Genotype, rrBLUP, BLUP, metaN, and metaG, respectively (Support Fig. 3C); for heritabilities, they were 0.378, 0.383, 0.387, 0.428, 0.487, 0.495,

and 0.503 for Genotype, BLUP, Network, BayesR, rrBLUP, metaG, and metaN, respectively (Support Fig. 3D). These findings demonstrate that DeepAnnotation complements commonly used BLUP- and Bayes-based models by improving GWAS detection rates and estimated heritability.

In summary, benefiting from the flexible framework, which incorporates comprehensive functional annotations and deep learning, DeepAnnotation suggests to explain different aspects of genetic variance, and could serve as a complement to existing models for dissecting the genetic basis of complex traits.

Methods

Simulated data

In this study, we used the simulated data to detect the genetic basis of complex traits by pre-defining causal variants. For each individual i, the simulated phenotype y_i was done by the following formula:

$$y_i = \boldsymbol{\mu} + Z_i g_i + \varepsilon_i$$

Where, $\boldsymbol{\mu}$ represents the overall mean, Z_i represents the genotype vector of individual i, $g_i \sim gamma(0.4, 1.66)$ represents the simulated effects vector of pre-defined QTLs, $\varepsilon_i \sim N(0, \sigma_e^2)$ represents the residual effect. Here, σ_e^2 represents the residual variance and the genetic variance $\sigma_g^2 = \sigma_e^2 \times h^2/(1-h^2)$, with h^2 represents the heritability. In our simulations, 10 simulated traits were done based on the original 11,633,164 genotypes by setting the heritability to 0.7 and the number of pre-defined QTLs to 10,000.

Detection rate and estimated heritability

To calculate the detection rate and estimated heritability, we first extracted the weights of all SNPs trained by each model through 5-fold cross-validation and calculated their significance levels using a meta-strategy with multiple testing correction via the 'RobustRankAggreg' R package. SNPs with an adjusted *P*-value < 0.05 were considered potential causal variants. For simulated traits, real causal variants corresponded to the predefined QTLs, whereas for the real LMP trait, causal variants were defined as SNPs with Bonferroni-adjusted *P*-values < 0.05 from GWAS analysis (plink using the parameters '--linear –adjust'). The detection rate was defined as the proportion of real causal variants among predicted causal variants. Heritability was defined as the genetic variance explained by these predicted causal variants and was estimated using the BLUP model with the GCTA software ('--reml-pred-rand --reml-est-fix --blup-snp').