Adding gene transcripts into genomic prediction improves accuracy and reveals sampling time dependence

B.C. Perez[1], M.C.A.M. Bink[1], K.L. Svenson[2], G.A. Churchill[2], M.P.L. Calus[3*]

[1] *Hendrix Genetics B.V., Research and Technology Center (RTC), P.O. Box 114, 5830 AC Boxmeer, the Netherlands.*

[2] *The Jackson Laboratory, Bar Harbor, Maine, United States of America.*

[3] *Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands.*

*Corresponding author: mario.calus@wur.nl *Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands*

Running title: Transcriptomic prediction accuracy

ABSTRACT

Recent developments allowed generating multiple high quality 'omics' data that could increase predictive performance of genomic prediction for phenotypes and genetic merit in animals and plants. Here we have assessed the performance of parametric and non-parametric models that leverage transcriptomics in genomic prediction for 13 complex traits recorded in 478 animals from an outbred mouse population. Parametric models were implemented using best linear unbiased prediction (BLUP), while non-parametric models were implemented using the gradient boosting machine algorithm (GBM). We also propose a new model named GTCBLUP that aims to remove between-omics-layer covariance from predictors, whereas its counterpart GTBLUP does not do that. While GBM models captured more phenotypic variation, their predictive performance did not exceed the BLUP models for most traits. Models leveraging gene transcripts captured higher proportions of the phenotypic variance for almost all traits when these were measured closer to the moment of measuring gene transcripts in the liver. In most cases, the combination of layers was not able to outperform the best single-omics models to predict phenotypes. Using only gene transcripts, the GBM model was able to outperform BLUP for most traits except body weight, but the same pattern was not observed when using both SNP genotypes and gene transcripts. Although the GTCBLUP model was not able to produce the most accurate phenotypic predictions, it showed highest accuracies for breeding values for 9 out of 13 traits. We recommend using the GTBLUP model for prediction of phenotypes and using the GTCBLUP for prediction of breeding values.

INTRODUCTION

Predicting complex traits is a fundamental aim of quantitative genetics. The use of whole genome single nucleotide polymorphisms (SNP) revolutionized the prediction of breeding values, resulting in the process widely known as genomic prediction (GP) (Meuwissen *et al.* 2001). A number of statistical approaches are now applied routinely in breeding programs, such as genomic best linear unbiased prediction (GBLUP) (VanRaden 2008), ridge regression (Whittaker *et al.* 2000), or methods from the "Bayesian Alphabet" (Gianola *et al.* 2009). More recently, machine learning algorithms have been tested in the context of genomic prediction (González-Recio *et al.* 2013; Pook *et al.* 2020; Zingaretti *et al.* 2020). These models may have several advantages when compared to traditional linear models, such as capturing interactions between predictors (non-additive effects), automatic variable selection and for making fewer assumptions regarding the underlying genetic architecture of phenotypes (Nayeri *et al.* 2019; Pérez-Enciso and Zingaretti 2019). However, compared to the linear models mentioned above, prediction performance from machine learning methods has shown mixed results (Azodi *et al.* 2019; Abdollahi-Arpanahi *et al.* 2020; Perez *et al.* 2022). There seems to be no "one-size-fits-all" model, as results are dependent on trait genetic architecture, size of the data, and on fine tuning of hyperparameters.

Recent development of low-cost high throughput molecular technologies allowed generating multiple high quality 'omics' data can be measured with high accuracy (Fernie and Schauer 2009; Tohge and Fernie 2015; Chawade *et al.* 2016). This has led to interest in utilizing these as new layers of information to improve the predictive performance of genomic prediction models, ultimately contributing to improve efficiency of breeding programs (Guo *et al.* 2016; Li *et al.* 2019). For example, gene expression levels measured in tissue samples by direct RNA sequencing (RNA-seq) is now readily available to animal breeders (Ozsolak and Milos 2011). To incorporate these new sources of data into genomic prediction models requires new strategies for integration with the already widely used genome-wide marker data. Although most of the literature focusing on the inclusion of gene-expression data into genomic

63 models to improve predictive performance aimed at predicting phenotypes (Takagi *et al.* 2014;
64 Guo *et al.* 2016; Schrag *et al.* 2018; Azodi *et al.* 2019; Li *et al.* 2019; Morgante *et al.* 2020),
65 fitting gene transcript levels as an additional layer of information into genomic models could
66 indirectly improve the prediction of breeding values. Christensen *et al*. (2021) presented a two-
67 step method to incorporate such intermediate omics into genomic evaluations considering
68 complete and incomplete omics-data scenarios. Results were validated using simulated data
69 and suggested superiority of the single-step method including both the intermediate omics and
70 genomics data, over the traditional genomic best linear unbiased prediction (GBLUP) using
71 only genomics data. Similar results were observed by Michel *et al*. (2021) when investigating
72 the integration of gene expression into genomic prediction for disease resistance in wheat by
73 using a hybrid relationship matrix for merging both layers of omics data. A pending issue that
74 remains, is the adequate handling of associations between layers of data that may lead to
75 inflated relative contributions of individual layers when ignored (Holm *et al.* 2010; Christensen
76 *et al.* 2021). Wade *et al*. (2021) have suggested that the benefits of multi-omics integration
77 models over single-omic models are achieved once redundancy of predictors is decreased.
78 Therefore, multi-omics models should either automatically or through adequate
79 parametrization be able to identify and manage information redundancy across multiple omics-
80 layers.

81 In the present study we used data from the Diversity Outbred (DO) mouse population
82 (Churchill *et al.* 2012; Svenson *et al.* 2012) to evaluate the utility of gene expression in addition
83 to genome-wide genetic markers for genomic prediction using different modeling strategies.
84 To this end, the objectives of this study were to: (1) assess the proportions of phenotypic
85 variance explained by genetic markers and gene transcripts in complex traits recorded in at
86 least two time points; (2) evaluate the predictive accuracy for phenotypes using transcripts
87 and/or marker information for the traits investigated using linear models and the gradient
88 boosting algorithm; and (3) evaluate how the inclusion of transcripts affects estimation of
89 genomic breeding values (GEBV) from BLUP models. The linear models proposed vary in

number of components, how interactions were modeled, and conditioning of one component on another. The gradient boosting machine algorithm was chosen for its ability to automatically control redundancy and implicitly account for non-linear effects in prediction, while the BLUP models tested comprise parametric approaches to incorporate genomics and transcriptomics, considering or ignoring the interactions between them.

MATERIAL AND METHODS

Data

Phenotypes

Data used for this study were obtained from The Jackson Laboratory (Bar Harbor, ME) and comprise a subset of the dataset used in Perez *et al*. (2022). The 478 DO mice originated from 4 non-overlapping generations (4, 5, 7 and 11) with males and females represented equally. The total number of animals per generation was 47, 47, 192 and 192 for generations 4, 5, 7 and 11, respectively, with slight variation in the numbers of missing records across traits (Table 1). The mice were maintained on either standard high fiber (chow, n=239) or high fat diet (n=239) from weaning until 23 weeks of age. The proportion of males and females within each diet category was close to 50-50 for all generations, as well as within each litter-generation combination (two litters per generation). This population is maintained under a systematic mating scheme, designed to limit population structure and relatedness. On average, the animals were related to each other at a level equivalent to first cousins, which is by design (Svenson *et al.* 2012). More elaborate descriptions of population structure, husbandry and phenotyping methods can be found in Svenson *et al*. (2012) and Tyler *et al*. (2021).

Table 1 gives for each trait a brief description, the numbers of observations and the estimated heritability. We considered six traits based on range of heritability and presumed genetic architectures. The chosen traits were measured at two or three times during the

animal's life, resulting in 13 distinct traits in total. The analyzed traits were bone mineral density at 12 (BMD12) and 21 (BMD21) weeks, body weight at 10, 15 and 20 weeks (BW10, BW15 and BW20); circulating cholesterol at 8 (CHOL8) and 19 (CHOL19) weeks, adjusted body fat percentage at 12 (FATP12) and 21 (FATP21), circulating glucose at 8 (GLUC8) and 19 (GLUC19) weeks, circulating triglycerides at 8 (TRGL8) and 19 weeks (TRGL19). These traits can be categorized into measurements of body composition (bone mineral density, body weights and fat percentage) and clinical plasma chemistries (circulating glucose and triglycerides). Phenotypic records were pre-corrected for fixed effects of diet, generation, litter, and sex (Perez *et al.* 2022). Therefore, the pre-corrected phenotypes (y^*) analyzed here comprise the sum of the additive genetic effect and residual terms.

TABLE 1

Genotypes

The genotype data used for the animals in this study, were obtained from their derived founder haplotypes (for details: see Perez *et al.* 2022). The complete genotype file used for the analyses included 64,000 markers on an evenly spaced grid, and the average distance between markers was 0.0238 cM. The full genotype dataset was cleaned based on the following criteria: variants with minor allele frequency < 0.05, call rates < 0.90 and linear correlation between subsequent SNPs > 0.80 were removed. After quality control, a total of 50,122 SNP markers were available for the mice with phenotypic, genotypic and transcriptomic records.

Transcript levels

Transcriptome-wide expression levels were measured from whole livers as previously described (Munger *et al.* 2014; Chick *et al.* 2016) for 478 animals at 26 weeks of age. The

RNA sample was sequenced using single-end RNA-Seq (Munger *et al.* 2014) and aligned transcripts to strain-specific genomes from the DO founders (Chick *et al.* 2016). Read counts were estimated using an expectation maximization algorithm (EMASE, https://github.com/churchill-lab/emase). Read counts were previously corrected for the effects of sex, diet, and batch by normalizing in each sample using upper quantile normalization and applying a rank Z transformation across samples. After quality control, quantification of transcripts was available for 11,770 genes (Tyler *et al.* 2017).

Statistical models

Below we introduce five best linear unbiased prediction (BLUP) models and three gradient boosting machine (GBM) models with their acronyms and key features summarized in Table 2.

Best linear unbiased prediction

GBLUP

The statistical model of GBLUP is:

$$\mathbf{y}^* = \mathbf{1}\mu + \mathbf{g} + \mathbf{e},$$

where $\mathbf{y}^*$ is the vector of pre-corrected phenotypes, $\mathbf{1}$ is a vector of ones, μ is the intercept, $\mathbf{g}$ is the vector of random additive genetic values, where $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$, $\mathbf{G}$ is the additive genomic relationship matrix between genotyped individuals, and σ_g^2 is the additive genomic variance. The matrix $\mathbf{G}$ is constructed following the second method described by VanRaden (2008) as $\frac{\mathbf{ZZ}'}{m}$ where $\mathbf{Z}$ is the matrix of centered and standardized genotypes for all individuals and m is the number of markers. Finally, $\mathbf{e}$ is the vector of random residual effects where $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$ with σ_e^2 being the residual variance, and $\mathbf{I}$ is an identity matrix.

166 ***TBLUP***

167 To evaluate the performance of transcriptomic data for predicting complex traits, we
168 used a Transcriptomic Best Linear Unbiased Predictor (TBLUP) model. This model is similar
169 to GBLUP, but using a transcriptomic relationship matrix, which evaluates the similarity among
170 animals based on gene expression levels (Guo *et al.* 2016).

171 The statistical model of TBLUP is:

172 $$\mathbf{y}^* = \mathbf{1}\mu + \mathbf{t} + \mathbf{e},$$

173 where $\mathbf{y}^*$, $\mathbf{1}$ and μ are defined as above, $\mathbf{t}$ is the vector of random transcript level effects,
174 where $\mathbf{t} \sim N(\mathbf{0}, \mathbf{T}\sigma_t^2)$ and $\mathbf{T}$ is the transcriptomic relationship matrix built according to the
175 formula $\frac{\mathbf{WW}'}{k}$ where $\mathbf{W}$ is the matrix of centered and standardized expression levels for all
176 animals and *k* is the number of genes, and σ_t^2 is the variance explained by gene transcripts.

177

178 ***GTBLUP and GTIBLUP***

179 The GTBLUP model fitted the $\mathbf{g}$ and $\mathbf{t}$ as independent random effects, each with their
180 own variance component (Guo *et al.* 2016; Li *et al.* 2019). The model is $\mathbf{y}^* = \mathbf{1}\mu + \mathbf{g} + \mathbf{t} + \mathbf{e}$,
181 where all the parameters are as defined above.

182 The GTIBLUP model fitted $\mathbf{g}$, $\mathbf{t}$, and the interaction between $\mathbf{g}$ and $\mathbf{t}$ with an additional
183 variance component (Morgante *et al.* 2020). This model is $\mathbf{y}^* = \mathbf{1}\mu + \mathbf{g} + \mathbf{t} + \mathbf{gt} + \mathbf{e}$,
184 where $\mathbf{y}^*$, $\mathbf{1}\mu$, $\mathbf{g}$, $\mathbf{t}$ and $\mathbf{e}$ are as defined above, and $\mathbf{gt}$ is the vector of interaction (between
185 genomic and transcriptomic) effects, where $\mathbf{gt} \sim N(\mathbf{0}, \mathbf{G}\#\mathbf{T}\sigma_{gt}^2)$ and # is the Hadamard
186 product.

187

188

189

GTCBLUP

The GTCBLUP model was similar to GTBLUP in that the $\mathbf{g}$ and $\mathbf{t}$ that were fitted as independent random effects, each with their own variance component. However, for this model the transcript levels were conditioned on SNP genotypes, yielding a matrix $\mathbf{W_c}$ computed as: $\mathbf{W_c} = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \mathbf{I}\lambda)^{-1}\mathbf{Z}')\mathbf{W}$, where $\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \mathbf{I}\lambda)^{-1}\mathbf{Z}'$ is the so-called "smoother matrix" (Hastie *et al.* 2009), $\mathbf{Z}$ is the matrix of centered and standardized genotypes as before, $\mathbf{I}$ is an identity matrix, and $\lambda = \frac{m*\sigma_e^2}{\sigma_g^2}$, σ_e^2 is the residual variance, and σ_g^2 is the additive genomic variance, both variances estimated with the GBLUP model (including only $\mathbf{g}$). Using the smoother matrix, i.e. including $\mathbf{I}\lambda$ rather than using $\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, reflects that the effects associated with the SNPs are estimated as random rather than fixed effects. The aim of this model is to remove any variance from transcripts that is correlated to variance in genotypes, such that any phenotypic variance both associated with variance in genotypes and transcripts automatically will be associated with the genotypes only. The model is $\mathbf{y}^* = \mathbf{1}\mu + \mathbf{g} + \mathbf{t_c} + \mathbf{e}$, where $\mathbf{t_c} \sim N(\mathbf{0}, \mathbf{T_c}\sigma_t^2)$ and $\mathbf{T_c}$ is computed as $\frac{\mathbf{W_c}\mathbf{W_c'}}{k}$, and all other parameters are as defined above.

Gradient boosting machine models

Gradient boosting machine (GBM) is an ensemble learning technique that applies an iterative process of assembling "weak learners" into a stronger learner, being largely used for both classification and regression problems (Friedman 2001). In the scope of this investigation, the GBM algorithm represents a non-parametric approach capable of implicitly fitting not only the additive effects of SNP and gene transcripts, but also the within- and between-omics layers interactions. The GBM is also capable of performing automatic feature selection, prioritization of important variables and discarding variables containing irrelevant or redundant information. A detailed description of the gradient boosting machine algorithm and

its application in genomic prediction can be found in Friedman (2002), González-Recio *et al*. (2010; 2013) and Perez *et al*., (2022).

To obtain the best possible results from the GBM algorithm, a grid search approach was used to determine the combination of hyperparameters that minimized the mean squared error of prediction within the inner training set for each trait. Details of the hyperparameter search method used are found in Perez *et al*. (2022). We implemented the GBM model using the "gbm" R package (Ridgeway 2020).

We tested three different GBM models. The first model considered only SNP genotypes as predictors (GGBM), the second model considered only (standardized) gene transcript levels as predictors (TGBM) and a third model that considered both genetic markers and transcript levels together as predictors (GTGBM). Our objective was to investigate if GBM models could capture within and between omics layers associations, while also reducing within and between omics layers redundancy by performing automatic variable selection. It is important to note here that although here we used "G" and "T" letters to refer to genomics and transcriptomics data in the GBM model's acronyms, predictors were fit directly in the model and not as relationship matrices

TABLE 2

Variance explained by genetic markers, transcript levels and combinations of both

To understand how much of the phenotypic variance can be explained by using SNP genotypes, gene transcript levels and the combinations of both sources of information, we estimated variance components using the GBLUP, TBLUP, GTBLUP, GTIBLUP and GTCBLUP models. Estimates of variance components along with the residual variance (σ_e^2) were obtained from a Bayesian approach analysis, using the BGLR R package (Pérez and de

los Campos 2014). For all models, the Gibbs sampler was run for 60,000 iterations, with a 20,000 burn-in period and a thinning interval of 10 iterations. Consequently, inference was based on 4,000 posterior samples.

For the GTIBLUP model, we calculated the portion of variance explained by SNP genotypes ($h^2 = \frac{\sigma_g^2}{\sigma_g^2+\sigma_t^2+\sigma_{gt}^2+\sigma_e^2}$), gene transcripts ($t^2 = \frac{\sigma_t^2}{\sigma_g^2+\sigma_t^2+\sigma_{gt}^2+\sigma_e^2}$) and from the interaction between effects from genetic markers and gene transcripts ($gt^2 = \frac{\sigma_{gt}^2}{\sigma_g^2+\sigma_t^2+\sigma_{gt}^2+\sigma_e^2}$). Consequently, the sum of h^2, t^2 and gt^2 represent the portion of phenotypic variance explained by two layers of omics data and by the between-omics-layer interactions. The parameters h^2, t^2 and gt^2 for the other models were calculated similarly but omitted any variance components associated with effects not included in the model.

Model Performance

Performance of predictions from the models was measured by the accuracy, computed as the Pearson correlation ($r_{y^*,\hat{y}}$), and the relative root-mean squared error of prediction (RRMSE) between predictions ($\hat{y}$) and pre-corrected phenotypes (y^*): RRMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y^*-\hat{y})^2}\,/\,\sigma_p$, where σ_p is the phenotypic standard deviation. In all analyses, we used a forward prediction validation scheme in which animals from older generations (4, 5, 7) were used as the reference and animals from the younger generation (11) as the validation subset. The standard error (SE) around the $r_{y^*,\hat{y}}$ estimates were obtained by calculating the standard deviations from 10,000 bootstrap samples (Davison and Hinkley 1997). The bootstrapping procedure was implemented using the “boot” R package (Canty and Ripley 2021). We have also assessed prediction bias by obtaining the regression coefficient from the linear regression of corrected phenotypes on model predictions. For these results, values above 1 indicate deflation, while values below 1 indicate inflation of predicted values.

To assess the proportion of variance explained by the models tested, we have calculated the coefficient of determination (R^2) from the regression of corrected phenotypes on model predictions for all traits. For the GBM models we have used results from the model using the previously obtained best hyperparameter set from the standard grid-search procedure to assess the model R^2 for prediction within the reference set.

For the BLUP models proposed to integrate SNP genotypes and gene transcripts (GTBLUP, GTIBLUP and GTCBLUP), in addition to $r_{y^*,\hat{y}}$ we have also calculated the correlations between the solutions for each random effect included in the model ($\mathbf{g}$, $\mathbf{t}$, $\mathbf{t_c}$ or $\mathbf{gt}$) and $\hat{\mathbf{y}}$, as well as pairwise comparisons between all components in the model. Here, we also focus on solutions from the additive genetic component from these models to assess if the prediction of genomic breeding values (GEBV) can be improved by using models capable of integrating SNP genotypes and gene transcripts for genomic prediction.

Data Availability

All data associated with this manuscript, and the code developed and used to perform analyzes described in this manuscript, can be obtained at https://doi.org/10.6084/m9.figshare.15081636.v1. All software used is publicly available.

RESULTS

Variance components estimation percentage of variance explained within the reference set

Genomic heritabilities (h^2) obtained with GBLUP ranged from 0.08 to 0.44, representing a wide range of magnitudes across traits (Figure 1 and Figure 2). When only fitting transcript levels as predictors (TBLUP), the percentage of variance explained (t^2) ranged from 0.22 to 0.75 and in general it was higher than h^2 when comparing within the same trait. The exceptions to that were observed for BMD12 (h^2 = 0.39 and t^2 = 0.35) and GLUC8 (h^2 = 0.30 and t^2 = 0.22). When comparing the same trait measured at different time points,

t^2 from TBLUP was higher for phenotypes collected closer to the 26 weeks of age (i.e. the age at mRNA data sampling).

In terms of the total phenotypic variance explained, GTBLUP and GTIBLUP showed similar results (Figure 1 and Figure 2). For body weights (BW10, BW15, BW20) and fat percentage (FATP12 and FATP21) traits, the variance explained by genetic markers (in GTBLUP and GTIBLUP) was drasticly lower when compared to GBLUP for the same traits. For the remaining traits the decrease in genetic variance captured by markers was much lower. For the interaction component in GTIBLUP (gt^2), results observed varied according to the trait analysed but in general, it was low compared to h^2 and t^2. The only exception to that was observed for TRGL8, in which gt^2 was higher than h^2 and t^2. For CHOL8, GLUC19 and TRGL19, gt^2 was either similar to h^2 or t^2.

For GTCBLUP, differently from GTBLUP and GTIBLUP, the additive genetic variance captured was always in line with results from GBLUP. On the other hand, the variance explained by transcripts (t^2) from GTCBLUP was always lower than observed by other models including transcripts as predictors (TBLUP, GTBLUP and GTIBLUP).

FIGURE1

FIGURE 2

The variance explained (represented by the R^2 parameter) within the reference data by parametric models was in general lower than by the non-parametric counterparts (Table 3). Independent of being a parametric or non-parametric model, the use of gene transcripts (TBLUP and TGBM) as predictors explained in most cases more of the variance than using exclusively SNP genotypes (GBLUP and GGBM). For GTBLUP, GTIBLUP and GTGBM, the variance explained was at least similar to observed for TBLUP and TGBM, but generally

higher. For GTCBLUP, variance explained by the model was slightly to moderately higher than observed for GBLUP model, but always smaller than observed for GTBLUP, GTIBLUP and GTGBM. The average R^2 when considering only traits recorded earlier (suffixes 8, 10 or 12) and later (suffixes 19, 20 or 21) moments were 76% and 83%, respectively, when using TBLUP, being the largest differece observed across models when considering these two groups of traits.

TABLE 3

Prediction performance – Phenotype prediction

In Table 4 accuracies are shown for predicted phenotypes for BLUP and GBM models using either SNP genotypes, transcript levels or both as predictors. Here we considered GBLUP to be the reference method. It showed prediction accuracies ranging from 0.01 to 0.29, these were highly positively correlated to the portion of variance explained by SNP genotypes by the same model, except for CHOL19. When comparing predictive performance between GBLUP and GGBM models, GBLUP yielded highest prediction accuracies for 7 traits, while GGBM had best predictive performance for 6 traits out of 13.

For models that include only gene transcripts (TBLUP and TGBM), the TBLUP approach showed predictive accuracies ranging from 0.03 to 0.61, having the best performance for only 4 out of 13 traits. The TGBM model was able to overcome TBLUP for 7 traits, with prediction accuracies ranging from 0.04 to 0.58. For BMD21 and BW15, predictive accuracy was identical between TBLUP and TGBM. The differences between accuracies from TBLUP and TGBM was higher than between GBLUP and GGBM.

For models that combined SNP genotypes and gene transcripts levels (GTBLUP, GTCBLUP, GTIBLUP and GTGBM), GTBLUP had the highest predictive accuracy for 5 traits out of 13. The second-best model overall was GTGBM, with the highest predictive accuracy for 4 traits. For every trait that GTIBLUP had the highest prediction accuracy, it was identical

to the result for GTBLUP, while the GTCBLUP never had the highest predictive accuracy (Table 4).

The prediction error (RRMSE) and bias (β) for model's predictions are presented in Supplementary Tables S1 and S2, respectively. Considering single-omics models, on average BLUP models (GBLUP and TBLUP) yielded less biased predictions than GBM models (GGBM and TGBM). For models integrating SNP genotypes and gene transcripts, GTBLUP and GTIBLUP showed similar bias across traits, while GTGBM had on average less bias than the BLUP models. For the GTCBLUP model, predictions were inflated ($\beta < 1$) for all traits but BMD21. In terms of prediction error, differences between models were smaller than observed for bias (Supplementary Table S2) or predictive accuracies (Table 4). The lowest RRMSE values were observed for FATP21, while the highest were observed for GLUC19. The RRMSE values for all traits analyzed were all around 1, indicating the average prediction errors were close to one phenotypic standard deviation.

TABLE 4

Predictive ability for GEBV and other model components, and the correlation between them in BLUP models

In Table 5 the Pearson's coefficient correlation between model components solutions ($\hat{g}$, $\hat{t}$, $\widehat{t_c}$ and $\hat{g}_t$) for the different BLUP models and corrected phenotypes (y^*) are shown. Overall, results for GTBLUP and GTIBLUP were similar across traits. These two models had the most accurate GEBV ($\rho_{\hat{g}\_y^*}$) exclusively for GLUC8, while for BMD12 results from these models were matched by GTCBLUP. For GLUC19, all four parametric multi-omics models yielded the same accuracy for GEBV, which was the lowest (0.01) across traits. In 8 out of 13 traits the GEBV estimated using GTCBLUP model was the most accurate across all models. The correlation between $\hat{t}$ and y^* ($\rho_{\hat{t}\_y^*}$) was also similar between GTBLUP and GTIBLUP, being always higher for these two models than observed for GTCBLUP. For

GTCBLUP exclusively, $\rho_{\hat{t}\_y^*}$ was low and negative for CHOL19 (-0.08), GLUC19 (-0.05) and TRGL8 (-0.07). For most traits, although a slight increase in the total variance explained was observed within the reference dataset (Figures 1 and 2) when comparing GTBLUP and GTIBLUP, there was not a proportional increase in $\rho_{\hat{g}\_y^*}$ in the validation (Table 5). For GTCBLUP on the other hand, for all traits there was an increase in the variance explained by SNP genotypes (g^2 in Figure 1 and Figure 2) when compared to GTBLUP and GTIBLUP, and the same pattern was observed for $\rho_{\hat{g}\_y^*}$. Results for $\rho_{\hat{g}\_\hat{t}}$ varied from +0.14 to +0.29 for GTBLUP and from +0.13 to +0.29 for GTIBLUP. For GTCBLUP, values for $\rho_{\hat{g}\_\hat{t_c}}$ were all negative and close to zero, ranging from -0.13 to -0.03 (Table 5). The values for $\rho_{\hat{g}\_\widehat{gt}}$, only calculated for GTIBLUP, were close to zero for most traits with an exception for CHOL8 and CHOL19, for which $\rho_{\hat{g}\_\widehat{gt}}$ was 0.18. A similar pattern was observed for $\rho_{\hat{t}\_\widehat{gt}}$, for which values varied from -0.12 to +0.06, with the largest differences from $\rho_{\hat{g}\_\widehat{gt}}$ observed for CHOL8 and CHOL19.

TABLE 5

DISCUSSION

Here, we investigated parametric and non-parametric approaches to leverage transcriptomic data for the prediction of complex phenotypes. To accomplish that, we used 478 animals from the DO Mouse population (Svenson *et al.* 2012), for which information on phenotypes (Churchill *et al.* 2012) for a wide range of quantitative traits, SNP genotypes and gene transcript levels from liver tissue (Tyler *et al.* 2017) were available on the same animals. We used the genomic (GBLUP) and transcriptomic (TBLUP) best linear unbiased prediction models to evaluate the value of these omics data to predict phenotypes. In addition, we evaluated models to integrate genome and transcriptome data by modelling both layers

393 independently (GTBLUP) or including an interaction component between the genome and
394 transcriptome (GTIBLUP). Finally, we proposed the GTCBLUP model that removes the
395 between-omics-layer information redundancy. The gradient boosting machine (GBM)
396 algorithm was investigated as a non-parametric approach potentially able to perform variable
397 selection and capture non-linear effects by fitting either SNP genotypes (GGBM), gene
398 transcript levels (TGBM) or to integrate both layers implicitly modeling interactions within and
399 between omics layers (GTGBM).

400 Using data from six distinct traits measured at two or three time points (resulting in 13
401 traits in total), we first assessed the proportion of phenotypic variance explained by each
402 variance component in the parametric models (Figure 1 and Figure 2). The variance explained
403 by SNP genotypes and gene transcript levels (and their interaction) varied by trait, time of
404 measurement and the model used. When using transcripts as predictors, two main patterns
405 were observed. For 5 out 13 traits (Figure 1), the TBLUP model explained much more of the
406 phenotypic variance than GBLUP. For the other 8 out of 13 traits (Figure 2), TBLUP explained
407 less variance than GBLUP. The observation that the portion of variance explained by gene
408 transcripts is strongly trait-specific is in line with results observed when assessing the
409 proportion of variance from gene transcripts for complex traits in *Drosophila* (Morgante *et al.*
410 2020). Ehsani *et al*. (2012) have analyzed data from an F2 mice population using models
411 integrating genotype markers and liver transcriptomics data. The authors reported that
412 transcripts explained 79%, while genotypes explained 36% of the phenotypic variance for body
413 weight at 8 weeks of age. It is important to emphasize that in Ehsani *et al*., (2012), RNA
414 samples were measured at the same time point as phenotypes were collected. Other authors
415 have observed that the genetic markers always explained a bigger portion of variance than
416 transcripts in maize (Guo *et al.* 2016; Azodi *et al.* 2019) and *Drosophila* (Li *et al.* 2019).

417 There are several possible reasons for the conflicting results found in literature when
418 assessing the value of transcripts to explain variation in complex phenotypes. Differently from
419 genotypes, transcripts are affected by many factors such as the tissue from where samples

420 are collected, the moment in life of sampling and the environmental conditions that the animal
421 was exposed to. These variables most likely impact the variance explained (and concomitantly
422 prediction performance) by transcripts. In Azodi *et al*. (2019) for example, transcripts were
423 quantified from whole seedlings while phenotypes were recorded at a much older age, which
424 could explain the limited predictive ability of transcriptomics data. In the present study
425 transcripts were measured when the mice were 26 weeks of age, while all phenotypes were
426 recorded at younger ages (from 8 to 21 weeks of age). In our results, phenotypes recorded
427 closer to 26 weeks of age had a larger proportion of phenotypic variance explained by
428 transcripts than measurements made earlier in the animal's life for the same phenotype
429 (Figure 1 and Figure 2). For BW and FATP the transcripts explained a larger proportion of
430 phenotypic variance at all time points. For BMD, CHOL, GLUC and TRGL this was the case
431 when there was 4 (BMD) to 6 weeks (CHOL, GLUC and TRGL) in-between measuring
432 phenotypes and transcripts, while genomics explained more phenotypic variance when this
433 time frame increased to 14 (BMD) or 18 weeks (CHOL, GLUC and TRGL) in-between. It is
434 important to emphasize here that although this outcome may have been expected beforehand,
435 to our knowledge it is the first time that this link between amount of variance explained by
436 transcript versus the time difference between measuring transcripts and phenotypes has been
437 shown empirically. One other aspect that must be considered here is that the gene expression
438 from whole maize seedlings (Azodi *et al.* 2019) is probably much less related to traits collected
439 later in life than the gene expression from liver tissue available for the DO mouse dataset. It
440 is widely known that the liver is strongly linked to many metabolic pathways (Ponsuksili *et al.*
441 2019), and therefore likely also especially to the BW and FATP traits used here, while the
442 variation contained in a sample collected from whole seedlings do not reflect a specific tissue
443 but a pool of all tissues in this organism.

444 When fitting both SNP genotypes and gene transcripts as predictors the portion of
445 variance explained by SNP genotypes varied drastically from the GBLUP model. For all BW
446 and FATP traits, the proportion of variance explained by genotypes using GTBLUP and

GTIBLUP was much lower than for GBLUP. Ehsani *et al*. (2012) and Takagi *et al*. (2014) observed a reduction in captured genetic variance by SNP genotypes of around 50% when fitting genotypes together with transcripts compared to models using fitting only genotypes as predictors for complex traits in other mice populations. This seems to confirm the hypothesis that there is redundant information between the genome and transcriptome layers (Wade *et al.* 2021), as also shown to be the case in *Drosophila* (Morgante *et al.* 2020). In our experience, it seems that the closer the phenotype analyzed is to the moment of RNA sampling, the higher the decrease in genetic variance captured by SNP genotypes in GTBLUP and GTIBLUP. This was observed for almost all traits we analyzed in different magnitudes. Takagi *et al*. (2014) analyzed circulating cholesterol at 10 weeks of age in mice and reported a large decrease in the genetic variance captured from SNP genotypes from models including only SNP genotypes (g^2= 46%) and together with liver transcripts (g^2= 19%) also measured at 10 weeks of age. In the present study, we observed only a slight decrease in genetic variance estimated when comparing GTBLUP (g^2= 28%) and GBLUP (g^2= 38%) for CHOL8. This seems to confirm that for the same phenotype, measurements made closer to the RNA sampling are prone to exhibit this pattern in a higher magnitude than measurements. This was further substantiated by the results observed for GTCBLUP. By conditioning the transcripts on the genotypes, the portion of variance explained by SNP genotypes was similar to the GBLUP model, while the variance explained by gene transcripts was much lower than estimated with TBLUP, GTBLUP and GTIBLUP.

The formal variance partitioning achieved with the BLUP models cannot be achieved with the non-parametric GBM models. To compare the performance of GBM and BLUP in terms of explained variance we investigated the model R^2 within the reference set. For the GBM models it was almost always higher than for the BLUP models (Table 3). From our results, this pattern is recognizable for almost all traits analyzed, in which the GBM algorithm is able to capture a higher portion of variance than the parametric counterpart within the reference dataset (Table 3) but fails to outperform these models when predicting in the

validation set (Table 4). The only exception to that is observed for GLUC19, for which in addition to the much larger portion of variance explained within the reference set, the GBM algorithm also outperformed BLUP models by a large margin for prediction purposes. The presence of noise in the data, limited size of the training set and the underlying complexity of the event being modelled are often cited as common causes of overfitting in machine learning models (Vabalas *et al.* 2019; Ying 2019). Here we used a training dataset of 286 animals and the high number of predictors in the models, coupled with the unavoidable presence of collinearity within and between omics layers may have caused GBM models to overfit. Takagi *et al*. (2014) have also analyzed datasets of similar size from another heterogeneous mice population using parametric models that integrated genomics and transcriptomics data, reporting large portions of phenotypic variance captured within training sets that weren't necessarily translated to high predictive accuracy in the validation set for circulating glucose and cholesterol. The authors argue that the high number of model's parameters to be estimated and the small number of animals with observations available could be the main cause of this pattern. We have observed a big impact of hyperparameters on the in predictive accuracies of the GBM models (results not shown). Having access to larger datasets could help to elucidate the magnitude of this impact for the models analyzed here since it would decrease the impact of hyperparameter definition in predictive performance, improving strength of evidence for any differences found between GBM and other models tested. The forward-prediction validation method adopted in the present study was previously described in Perez *et al*. (2022) and mimics prediction in animal and plant breeding, where a predictive model is trained based on data from individuals that have genomics (here also transcriptomics) and phenotype data available, while prediction is intended for un-phenotyped younger individuals. Here, training and validation generations were also 3 generations apart from each other, which erodes strong relationships (e.g., parent-progeny) and therefore should not be the cause of model overfitting suggested for the GBM models.

Prediction performance for phenotypes can be improved by combining genotypes and transcripts, however our results suggested that the magnitude of improvement is dependent on the trait analyzed (Table 4). In line with the observed differences in variance explained by model components, TBLUP showed a better predictive ability than GBLUP for most traits except CHOL19, GLUC8 and GLUC19. In contrast, Takagi *et al*. (2014) reported higher predictive accuracies for circulating glucose and cholesterol when using liver transcripts as predictors when compared to using genotypes in a different heterogeneous mice population (Valdar *et al.* 2006). Two aspects may explain the differences observed between studies. First, Takagi *et al*. (2014) performed a cross-validation scheme by randomly sampling individuals as reference and validations sets while we performed a forward-validation scheme, in which phenotype prediction for younger animals was based on estimates from older generations. Even though animals are randomly sampled, the overall similarity between reference and validation sets is higher in Takagi *et al*. (2014), which in general leads to higher magnitude of predictive accuracy (Pszczola *et al.* 2012; Werner *et al.* 2020). It is important to emphasize that although this is true when dealing exclusively with SNP genotypes, we cannot confirm that the affirmative holds when dealing with transcripts as predictors. A second relevant aspect is that in the present study phenotypes for CHOL and GLUC were collected at 8 and 19 weeks, while RNA samples were taken at 26 weeks of age. As previously mentioned, in Takagi *et al*. (2014) RNA samples and phenotypes were collected at the same age (10 weeks). In the present study, phenotypes recorded at a closer time point to transcript profiling result in higher predictive performance from transcripts (Table 4).

The TGBM model was able to overcome the TBLUP model for several traits, which was not the case when comparing GBLUP and GGBM. This result may indicate that interactions between transcripts were more easily captured or are more relevant than between SNPs. When fitting only genotype markers, the model is limited by incomplete linkage disequilibrium between the SNP and quantitative trait loci to perform an accurate detection of possible interactions, while there is no such limiting factor when using gene transcripts. One

other hypothesis is that as transcripts are more strongly linked to phenotypes than genetic markers, transcript-by-transcript interactions are also likely to affect the phenotype more strongly than SNP-by-SNP interactions (Green *et al.* 2019), hence the former is expected to have a clearer and more detectable signal. Morgante *et al*. (2020) have used the random forest model, a non-parametric ensemble machine learning method like GBM, to predict complex phenotypes in *Drosophila* using gene transcripts as predictors but did not observe a superior predictive ability when compared to the TBLUP model. While TGBM consistently outperformed TBLUP in our study, the GTGBM was only partly outperforming GTBLUP and GTIBLUP. This could mean that the inclusion of SNP genotypes together with gene transcripts as predictors in the GTGBM model may have impaired the ability of GBM to capture linear and non-linear signals from within- and between-omics layers. The exact cause remains unclear, but the size of dataset together with the substantial increase in number of predictors when going from TGBM to GTGBM may be in the roots of it. It is likely that the GBM algorithm may require more data to be able to accurately capture all patterns from the complex relationship between omics layers underlying quantitative traits. If this is indeed the case, testing these models using a larger dataset could help to confirm this hypothesis. In Azodi *et al*. (2019) machine learning models integrating genomics and transcriptomics data were also not able to outperform single-omics models in terms of predictive accuracy for three traits in maize using a dataset of similarly limited size as in the present study.

The linear association for solutions from BLUP models' components predicted for animals within the validation set were very similar between GTBLUP and GTIBLUP. For these two models $\rho_{\hat{g}\_y^*}$, $\rho_{\hat{t}\_y^*}$ and $\rho_{\hat{g}\_\hat{t}}$ were almost identical across all traits (Table 5). Although the inclusion of an interaction component in GTIBLUP captured between 9% and 26% of phenotypic variance (represented by gt^2 in Figure 1 and Figure 2) within the reference set, it did not seem to affect the relationship between other components. The low values observed for $\rho_{\hat{g}\_\widehat{gt}}$ and $\rho_{\hat{t}\_\widehat{gt}}$ in GTIBLUP also seem to suggest that the interaction component is capturing a portion of variance not directly shared with $\hat{g}$ or $\hat{t}$ components, and therefore it

554 does not affect the relationship between other components. The linear association between $\hat{g}$
555 and $\hat{t}$ in GTBLUP and GTIBLUP models was always higher than the association between $\hat{g}$
556 and $\hat{t_c}$ in GTCBLUP across all traits analyzed. Since in the GTCBLUP model the transcript
557 relationship matrix was conditioned on the variance of SNP genotypes (t_c), the linear
558 association between solutions for the two components was expected to closer to zero than
559 the observed in GTBLUP or GTIBLUP.

560 In this paper, we proposed the GTCBLUP model as an alternative to integrate genome
561 and transcriptome data for genomic prediction. There has been an increasing interest in the
562 use of intermediate omics data in animal and plant breeding (Guo *et al.* 2016; Yang *et al.*
563 2017; Azodi *et al.* 2019; Morgante *et al.* 2020; Christensen *et al.* 2021; Michel *et al.* 2021),
564 such as transcriptomics, metabolomics, or microbiome data. The inclusion of new layers of
565 omics data into genomic prediction models could arguably help in capturing additional portions
566 of variance not explained by genotype data, but at the same time, these layers most likely
567 contain overlapping information, increasing collinearity between predictors. Modelling the
568 relationship between G and T components could be an efficient way to realize the added value
569 of integrating such omics data into genomic prediction models (Wade *et al.* 2021), but this
570 could also be a challenge given the increase in number of parameters to be estimated. The
571 advantage of the GTCBLUP is that as pre-processing step it conditions the variance contained
572 in transcripts on the variance of genotypes to minimize the amount of redundant information
573 without having to increase model complexity. In general, the GTCBLUP model was able to
574 produce GEBV that were at least as accurate as or slightly more accurate than the GBLUP
575 model. The percentage of variance explained by SNP genotypes in GTCBLUP was similar to
576 that with the GBLUP model, while it was always lower when using GTBLUP and GTIBLUP
577 (Figure 1 and Figure 2). The observed reduction in additive genetic variance for GTBLUP and
578 GTIBLUP when compared to GBLUP indicates strong redundancy in information contained in
579 the genomic and transcriptomic layers. So, the conditioning of transcripts on SNP genotypes
580 in GTCBLUP allowed this model to perform a more accurate variance partitioning for the

581 additive genetic component, which consequently resulted in a more accurate estimation of
582 GEBV (Table 5). An interesting alternative way to consider the covariance between genomics
583 and transcriptomics layers, is by explicitly modelling it, as can be done using the CORE-
584 GREML method (Zhou *et al.* 2020), implemented in the MTG2 software (Lee and van der Werf
585 2016). We considered this method to evaluate its potential, as well as to try and assess the
586 magnitude of overlapping information between genotype and transcript data for the traits
587 analysed. Results (Supplementary Table S3) indicated correlations from -0.47 to +0.71, with
588 highest values observed for BW10, BW15, BW20, CHOL8 and CHOL19, FATP19 but the
589 correlation coefficient was never significantly different from 0. Coincidently for most of these
590 traits GTCBLUP obtained the most accurate GEBV when compared to GBLUP, GTBLUP and
591 GTIBLUP. These trends seem to further support that, for specific traits, genomics and
592 transcriptomics layers contain largely overlapping information, and although removing this
593 redundancy does not result in more accurate phenotype prediction, it may contribute to obtain
594 more accurate GEBV and consequently could improve breeding decisions. The lack of
595 significance of the estimated correlations between genomics and transcriptomics data may be
596 due to the limited size of the data used here, while based on the results obtained in the present
597 study this does not provide a limitation for the GTCBLUP model.

598 One limitation of the GTCBLUP model is that it does not accommodate missing omics
599 information, so all reference individuals must have genomics and transcriptomics data
600 available. In the context of breeding programs, a situation in which all reference animals have
601 multiple omics data available is unlikely to happen due to high costs involved in the collecting
602 this kind of information. However, based on the observed decrease in costs of genotyping
603 which has enabled large-scale genotyping, we may expect similar developments for the costs
604 of transcriptomics and other intermediate phenotypes in the near future (Uzbas *et al.* 2019).
605 At the same time, there have been some recent model developments that enable including
606 other omics data in genomic prediction, when these other omics data are not available for all
607 animals (Christensen *et al.* 2021; Zhao *et al.* 2022).

CONCLUSION

We have assessed prediction models that incorporate genetic markers and transcriptomics data in genomic prediction of complex phenotypes in mice. The proportion of phenotypic variance explained by transcripts was almost always higher when traits were measured closer to the time of measuring gene transcripts. While GBM models explained more variance in the reference data, their predictive performance did not exceed the GBLUP models. Models including SNP genotypes and gene transcripts did not consistently outperform the best single-omics models to predict phenotypes. While TGBM model was able to outperform TBLUP, this was not the case for GTGBM compared to GTBLUP and GTIBLUP. The newly developed GTCBLUP model was able to force all phenotypic variance associated with SNP genotypes into its additive genetic component, by conditioning gene transcripts on SNP genotypes. GTCBLUP generally yielded considerably lower accuracies of phenotypic predictions than the other models including SNP genotypes and gene transcripts, but it showed the best accuracies for breeding values for most traits. We recommend using the GTBLUP model for prediction of phenotypes, while the GTCBLUP should be preferred when the aim is to estimate breeding values.

FUNDING

This study is part of the GENE-SWitCH project that received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 817998. Gary Churchill acknowledges support by the National Institutes of Health (NIH) grant R01 GM070683.

CONFLICTS OF INTEREST

The authors report no conflicts of interest related to the present manuscript. Bruno Perez and Marco Bink are employees of Hendrix Genetics (Boxmeer, The Netherlands).

635 **REFERENCES**

636 Abdollahi-Arpanahi R, Gianola D, Peñagaricano F. 2020. Deep learning versus parametric
637 and ensemble methods for genomic prediction of complex phenotypes. Genet Sel Evol.
638 52:12.

639 Azodi CB, Bolger E, McCarren A, Roantree M, de los Campos G, Shiu S-H. 2019.
640 Benchmarking parametric and machine learning models for genomic prediction of
641 complex traits. G3 (Bethesda). 9:3691-3702.

642 Canty A, Ripley B. 2021. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-28.

643 Chawade A, Alexandersson E, Bengtsson T, Andreasson E, Levander F. 2016. Targeted
644 proteomics approach for precision plant breeding. J Prot Res. 15:638-646.

645 Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL,
646 Churchill GA, Gygi SP. 2016. Defining the consequences of genetic variation on a
647 proteome-wide scale. Nature. 534:500-505.

648 Christensen OF, Börner V, Varona L, Legarra A. 2021. Genetic evaluation including
649 intermediate omics features. Genetics. 219:iyab130.

650 Churchill GA, Gatti DM, Munger SC, Svenson KL. 2012. The diversity outbred mouse
651 population. Mamm Genome. 23:713-718.

652 Davison AC, Hinkley DV. 1997. Bootstrap methods and their application. Cambridge
653 University Press, New York.

654 Ehsani A, Sørensen P, Pomp D, Allan M, Janss L. 2012. Inferring genetic architecture of
655 complex traits using Bayesian integrative analysis of genome and transcriptome data.
656 BMC Genom. 13:456.

657 Fernie AR, Schauer N. 2009. Metabolomics-assisted breeding: a viable option for crop
658 improvement? Trends Genet. 25:39-48.

659 Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. Ann Stat.
660 29:1189-1232.

661 Friedman JH. 2002. Stochastic gradient boosting. Comp Stat Data Anal. 38:367-378.

662 Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R. 2009. Additive genetic
663 variability and the Bayesian alphabet. Genetics. 183:347-363.
664 González-Recio O, Jiménez-Montero JA, Alenda R. 2013. The gradient boosting algorithm
665 and random boosting for genome-assisted evaluation in large data sets. J Dairy Sci.
666 96:614-624.
667 González-Recio O, Weigel KA, Gianola D, Naya H, Rosa GJM. 2010. L-2-Boosting algorithm
668 applied to high-dimensional problems in genomic selection. Genet Res. 92:227-237.
669 Green RM, Leach CL, Diewert VM, Aponte JD, Schmidt EJ, Cheverud JM, Roseman CC,
670 Young NM, Marcucio RS, Hallgrimsson B. 2019. Nonlinear gene expression-phenotype
671 relationships contribute to variation and clefting in the A/WySn mouse. Dev Dyn.
672 248:1232-1242.
673 Guo Z, Magwire MM, Basten CJ, Xu Z, Wang D. 2016. Evaluation of the utility of gene
674 expression and metabolic information for genomic prediction in maize. Theor Appl Genet.
675 129:2413-2427.
676 Hastie T, Tibshirani R, Friedman J. 2009. The Elements of Statistical Learning. Springer, New
677 York.
678 Holm K, Melum E, Franke A, Karlsen TH. 2010. SNPexp - A web tool for calculating and
679 visualizing correlation between HapMap genotypes and gene expression levels. BMC
680 Bioinformatics. 11:600.
681 Lee SH, van der Werf JHJ. 2016. MTG2: an efficient algorithm for multivariate linear mixed
682 model analysis based on genomic information. Bioinformatics. 32:1420-1422.
683 Li Z, Gao N, Martini JWR, Simianer H. 2019. Integrating gene expression data into genomic
684 prediction. Front Genet. 10.
685 Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using
686 genome-wide dense marker maps. Genetics. 157:1819-1829.
687 Michel S, Wagner C, Nosenko T, Steiner B, Samad-Zamini M, Buerstmayr M, Mayer K,
688 Buerstmayr H. 2021. Merging genomics and transcriptomics for predicting Fusarium head
689 blight resistance in wheat. Genes. 12:114.

690 Morgante F, Huang W, Sørensen P, Maltecca C, Mackay TFC. 2020. Leveraging multiple
691 layers of data to predict drosophila complex traits. G3 (Bethesda). 10:4599-4613.
692 Munger SC, Raghupathy N, Choi K, Simons AK, Gatti DM, Hinerfeld DA, Svenson KL, Keller
693 MP, Attie AD, Hibbs MA*, et al.* 2014. RNA-Seq alignment to individualized genomes
694 improves transcript abundance estimates in multiparent populations. Genetics. 198:59-
695 73.
696 Nayeri S, Sargolzaei M, Tulpan D. 2019. A review of traditional and machine learning
697 methods applied to animal breeding. Ann Health Res Rev. 20:31-46.
698 Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. Nat
699 Rev Gen. 12:87-98.
700 Pérez-Enciso M, Zingaretti LM. 2019. A guide for using deep learning for complex trait
701 genomic prediction. Genes. 10:553.
702 Perez BC, Bink MCAM, Svenson KL, Churchill GA, Calus MPL. 2022. Prediction performance
703 of linear models and gradient boosting machine on complex phenotypes in outbred mice.
704 G3 (Bethesda). 12:jkac039.
705 Pérez P, de los Campos G. 2014. Genome-wide regression and prediction with the BGLR
706 statistical package. Genetics. 198:483-495.
707 Ponsuksili S, Trakooljul N, Hadlich F, Methling K, Lalk M, Murani E, Wimmers K. 2019.
708 Genetic regulation of liver metabolites and transcripts linking to biochemical-clinical
709 parameters. Front Genet. 10:348.
710 Pook T, Freudenthal J, Korte A, Simianer H. 2020. Using local convolutional neural networks
711 for genomic prediction. Front Genet. 11:561497.
712 Pszczola M, Strabel T, Mulder HA, Calus MPL. 2012. Reliability of direct genomic values for
713 animals with different relationships within and to the reference population. J Dairy Sci.
714 95:389-400.
715 Ridgeway G. 2020. Generalized Boosted Models: A guide to the gbm package. https://cran.r-
716 project.org/web/packages/gbm/vignettes/gbm.pdf. Accessed 5 September, 2021.

717 Schrag TA, Westhues M, Schipprack W, Seifert F, Thiemann A, Scholten S, Melchinger AE.
718 2018. Beyond genomic prediction: combining different types of omics data
719 can improve prediction of hybrid peformance in maize. Genetics. 208:1373-1385.
720 Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, Palmer AA, McMillan L,
721 Churchill GA. 2012. High-resolution genetic mapping using the mouse diversity outbred
722 population. Genetics. 190:437-447.
723 Takagi Y, Matsuda H, Taniguchi Y, Iwaisaki H. 2014. Predicting the phenotypic values of
724 physiological traits using SNP genotype and gene expression data in mice. PLOS ONE.
725 9:e115532.
726 Tohge T, Fernie AR. 2015. Metabolomics-inspired insight into developmental, environmental
727 and genetic aspects of tomato fruit chemical composition and quality. Plant Cell Physiol.
728 56:1681-1696.
729 Tyler AL, El Kassaby B, Kolishovski G, Emerson J, Wells AE, Mahoney JM, Carter GW. 2021.
730 Effects of kinship correction on inflation of genetic interaction statistics in commonly used
731 mouse populations. G3 (Bethesda). 11:jkab131.
732 Tyler AL, Ji B, Gatti DM, Munger SC, Churchill GA, Svenson KL, Carter GW. 2017. Epistatic
733 networks jointly influence phenotypes related to metabolic disease and gene expression
734 in diversity outbred mice. Genetics. 206:621-639.
735 Uzbas F, Opperer F, Sönmezer C, Shaposhnikov D, Sass S, Krendl C, Angerer P, Theis FJ,
736 Mueller NS, Drukker M. 2019. BART-Seq: cost-effective massively parallelized targeted
737 sequencing for genomics, transcriptomics, and single-cell analysis. Genome Biol. 20:155.
738 Vabalas A, Gowen E, Poliakoff E, Casson AJ. 2019. Machine learning algorithm validation
739 with a limited sample size. PLOS ONE. 14:e0224365.
740 Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS,
741 Rawlins JNP, Mott R, Flint J. 2006. Genome-wide genetic association of complex traits in
742 heterogeneous stock mice. Nat Genet. 38:879-887.
743 VanRaden PM. 2008. Efficient methods to compute genomic predictions. J Dairy Sci.
744 91:4414-4423.

745 Wade AR, Duruflé H, Sanchez L, Segura V. 2021. eQTLs are key players in the integration
746 of genomic and transcriptomic data for phenotype prediction.
747 bioRxiv.2021.2009.2007.459279.
748 Werner CR, Gaynor RC, Gorjanc G, Hickey JM, Kox T, Abbadi A, Leckband G, Snowdon RJ,
749 Stahl A. 2020. How population structure impacts genomic selection accuracy in cross-
750 validation: implications for practical breeding. Front Plant Sci. 11:592977.
751 Whittaker JC, Thompson R, Denham MC. 2000. Marker-assisted selection using ridge
752 regression. Genet Res. 75:249-252.
753 Yang Y-l, Zhou R, Li K. 2017. Future livestock breeding: Precision breeding based on multi-
754 omics information and population personalization. J Integr Agricult. 16:2784-2791.
755 Ying X. 2019. An overview of overfitting and its solutions. J Phys Conf Ser. 1168:022022.
756 Zhao T, Zeng J, Cheng H. 2022. Extend mixed models to multi-layer neural networks for
757 henomic prediction including intermediate omics data.
758 Genetics.https://doi.org/10.1093/genetics/iyac1034.
759 Zhou X, Im HK, Lee SH. 2020. CORE GREML for estimating covariance between random
760 effects in linear mixed models for complex trait analyses. Nat Comm. 11:4208.
761 Zingaretti LM, Gezan SA, Ferrão LFV, Osorio LF, Monfort A, Muñoz PR, Whitaker VM, Pérez-
762 Enciso M. 2020. Exploring deep learning for complex trait genomic prediction in polyploid
763 outcrossing species. Front Plant Sci. 11:25.

764

765

766 **Table 1.** Number of available observations (N), the extended description of traits, age of the
767 animals at phenotype measurement, and estimated heritability.

Trait	N	Trait description	Age at measurement	Estimated heritability[1]
BMD12	471	Bone mineral density	12 weeks	0.39
BMD21	471	Bone mineral density	21 weeks	0.41
BW10	478	Body weight	10 weeks	0.42
BW15	478	Body weight	15 weeks	0.35
BW20	478	Body weight	20 weeks	0.37
CHOL8	474	Circulating cholesterol	8 weeks	0.38
CHOL19	474	Circulating cholesterol	19 weeks	0.45
FATP12	471	Body fat percentage	12 weeks	0.35
FATP21	471	Body fat percentage	21 weeks	0.32
GLUC8	425	Circulating glucose	8 weeks	0.31
GLUC19	425	Circulating glucose	19 weeks	0.22
TRGL8	473	Circulating triglycerides	8 weeks	0.36
TRGL19	473	Circulating triglycerides	19 weeks	0.31

768 [1] *Standard errors for the heritability ranged from 0.07 to 0.09*

769

770 **Table 2.** Overview of models applied to SNP genotypes and/or individual levels of gene
771 transcripts.

Model acronym		Explanatory variables		
		SNP genotypes	Gene transcripts	Interaction modelled
GBLUP		Yes	No	No
	GGBM	Yes	No	Yes (Implicitly)
TBLUP		No	Yes	No
	TGBM	No	Yes	Yes (Implicitly)
GTBLUP		Yes	Yes	No
GTCBLUP		Yes	Yes	No
GTIBLUP		Yes	Yes	Yes (Explicitly)
	GTGBM	Yes	Yes	Yes (Implicitly)

772

773

774 **Table 3.** Model R^2 (x100) for the best linear unbiased prediction (GBLUP, GTBLUP,
775 GTCBLUP and GTIBLUP) and gradient boosting (GGBM, TGBM and GTGBM) approaches
776 within training data[1].

Trait[1]	Model[2]							
	Only SNP		Only gene transcripts		SNP + gene transcripts			
	GBLUP	GGBM	TBLUP	TGBM	GTBLUP	GTIBLUP	GTCBLUP	GTGBM
BMD12	75	90	79	96	85	92	88	95
BMD21	84	85	88	93	87	96	92	98
BW10	78	92	93	96	91	95	90	97
BW15	75	87	91	94	93	96	94	96
BW20	80	87	92	93	95	97	95	97
CHOL8	84	85	69	97	80	93	87	98
CHOL19	82	83	85	95	88	96	92	98
FATP12	75	76	89	97	92	96	95	98
FATP21	80	82	93	97	95	97	94	98
GLUC8	77	85	66	95	81	93	87	97
GLUC19	62	80	67	96	75	90	84	98
TRGL8	65	81	62	92	80	96	88	97
TRGL19	71	86	70	96	78	95	84	98
Mean (all)	76	85	80	95	86	95	90	97
Mean (T1)	75	85	76	96	85	94	89	97
Mean (T2)	76	84	83	95	86	95	90	98

777 T1 = average R^2 of the column considering only traits recorded earlier in life (suffixes 8, 10 and 12).

778 T2 = average R^2 of the column considering only traits recorded later in life (suffixes 19, 20 and 21).

779 [1]For a description of the traits, see Table 1.

780 [2]For a description of the models, see Table 2.

781 [3]BW15 trait was ignored when calculating average performance considering exclusively T1 and T2.

782

783 **Table 4.** Accuracies of predicted pre-corrected phenotypes for the validation subset with the
784 proposed models. For each group of models, the result with the highest accuracy is
785 indicated in bold, identical results between two or more models are indicated in underline.

Trait[1]	Model[2]							
	Only SNP		Only gene transcripts		SNP + gene transcripts			
	GBLUP	GGBM	TBLUP	TGBM	GTBLUP	GTIBLUP	GTCBLUP	GTGBM
BMD12	**0.19**	0.16	**0.27**	0.25	0.27	0.27	0.20	0.25
BMD21	**0.29**	0.28	0.38	0.38	**0.42**	0.40	0.29	0.38
BW10	0.18	**0.20**	**0.48**	0.42	0.47	0.47	0.19	0.42
BW15	**0.15**	0.12	0.52	0.52	0.51	0.51	0.22	0.49
BW20	**0.18**	0.16	**0.61**	0.58	0.60	0.60	0.30	0.54
CHOL8	0.14	**0.17**	0.14	**0.15**	**0.18**	0.17	0.16	0.16
CHOL19	**0.25**	0.20	**0.19**	0.16	**0.26**	0.23	0.22	0.23
FATP12	0.14	**0.15**	0.44	**0.45**	0.44	0.44	0.28	**0.46**
FATP21	**0.21**	0.20	0.54	**0.56**	**0.54**	0.53	0.35	0.52
GLUC8	0.10	**0.12**	0.03	**0.04**	0.08	0.09	0.11	**0.15**
GLUC19	0.01	**0.10**	0.03	**0.05**	0.04	0.05	-0.05	**0.11**
TRGL8	0.08	**0.11**	0.06	**0.08**	**0.07**	0.06	0.05	0.06
TRGL19	**0.15**	0.13	0.17	**0.19**	0.17	0.18	0.12	**0.19**

786 [1]For a description of the traits, see Table 1.

787 [2]For a description of the models, see Table 2.

788

Table 5. Pearson's coefficient correlation (ρ) between model's components (ĝ, t̂, $\widehat{t_c}$ and $\widehat{gt}$) solutions and corrected phenotypes (y^*) for BLUP models proposed[1]. Numbers in bold (per row) show the best values for the accuracy of GEBV ($\rho_{\hat{g}\_y^*}$) across models, identical results between two or more models are indicated in underline.

Trait[2]	Model[3]											
	GTBLUP			GTIBLUP					GTCBLUP			GBLUP
	$\rho_{\hat{g}\_y^*}$	$\rho_{\hat{t}\_y^*}$	$\rho_{\hat{g}\_\hat{t}}$	$\rho_{\hat{g}\_y^*}$	$\rho_{\hat{t}\_y^*}$	$\rho_{\hat{g}\_\hat{t}}$	$\rho_{\hat{g}\_\widehat{gt}}$	$\rho_{\hat{t}\_\widehat{gt}}$	$\rho_{\hat{g}\_y^*}$	$\rho_{\widehat{t_c}\_y^*}$	$\rho_{\hat{g}\_\widehat{t_c}}$	$\rho_{\hat{g}\_y^*}$
BMD12	<u>0.20</u>	0.24	0.22	<u>0.20</u>	0.22	0.21	-0.04	0.06	<u>0.20</u>	0.04	-0.10	0.19
BMD21	0.29	0.38	0.28	0.30	0.37	0.29	0.04	-0.03	**0.31**	0.08	-0.03	0.29
BW10	0.18	0.46	0.27	<u>0.19</u>	0.47	0.27	0.04	0.02	<u>0.19</u>	0.03	-0.13	0.18
BW15	0.15	0.52	0.29	0.15	0.51	0.28	0.02	-0.02	**0.18**	0.08	-0.06	0.15
BW20	0.17	0.61	0.25	0.17	0.60	0.26	-0.09	-0.02	**0.21**	0.16	-0.10	0.18
CHOL8	0.14	0.13	0.16	0.14	0.13	0.15	0.18	-0.01	**0.15**	0.04	-0.08	0.14
CHOL19	0.24	0.14	0.16	0.25	0.12	0.15	0.18	-0.02	**0.27**	-0.08	-0.09	0.25
FATP12	0.18	0.43	0.14	0.18	0.42	0.13	-0.03	-0.12	**0.20**	0.15	-0.05	0.16
FATP21	0.22	0.52	0.16	0.22	0.53	0.16	-0.09	-0.10	**0.26**	0.16	-0.10	0.21
GLUC8	<u>0.12</u>	0.02	0.22	<u>0.12</u>	0.01	0.20	0.02	0.06	0.11	0.04	-0.09	0.10
GLUC19	<u>0.01</u>	0.04	0.19	<u>0.01</u>	0.02	0.18	-0.04	0.05	<u>0.01</u>	-0.05	-0.11	<u>0.01</u>
TRGL8	0.08	0.05	0.16	<u>0.09</u>	0.05	0.15	0.04	0.03	<u>0.09</u>	-0.07	-0.07	0.08
TRGL19	0.14	0.15	0.15	0.14	0.15	0.14	0.03	-0.05	**0.17**	0.03	-0.09	0.15

[1]$\rho_{\hat{g}\_y^*}$ = correlation between additive genetic effect and corrected phenotypes; $\rho_{\hat{t}\_y^*}$ = correlation between gene transcripts effect and corrected phenotypes; $\rho_{\hat{g}\_\hat{t}}$ = correlation between the additive genetic and gene transcripts effects; $\rho_{\hat{g}\_\widehat{gt}}$ = correlation between the additive genetic effect and the interaction between genetic and gene transcript effects; $\rho_{\hat{t}\_\widehat{gt}}$ = = correlation between the additive genetic effect and the interaction between genetic and gene transcript effects; $\rho_{\widehat{t_c}\_y^*}$ = correlation between gene transcripts conditioned on SNP genotypes and corrected phenotypes; $\rho_{\hat{g}\_\widehat{t_c}}$ = correlation between the additive genetic effect and gene transcripts conditioned on SNP genotypes.

[2]For a description of the traits, see Table 1.
[3]For a description of the models, see Table 2.

799 **Figure 1.** Percentage of variance explained by SNP genotypes (g^2), gene transcripts (t^2), the
800 interaction between them (gt^2) and not explained (e^2) by GBLUP, TBLUP, GTBLUP,
801 GTIBLUP and GTCBLUP models tested for the traits BW and FATP.

802 [1]For a description of the traits, see Table 1.

803 [2]For a description of the models, see Table 2.

804

805

806 **Figure 2.** Percentage of variance explained by SNP genotypes (g^2), gene transcripts (t^2), the
807 interaction between them (gt^2) and not explained (e^2) by GBLUP, TBLUP, GTBLUP,
808 GTIBLUP and GTCBLUP models tested for the traits BMD, CHOL, GLUC and TRGL.

809

810 [1]For a description of the traits, see Table 1.

811 [2]For a description of the models, see Table 2.

812

SUPPLEMENTARY MATERIAL

Table S1. Regression coefficient of the corrected phenotypes on prediction for validation animals from all models tested. Values closer to 1 indicate less bias.

Trait[1]	Model[2]							
	Only SNP[2]		Only gene transcripts[2]		SNP + gene transcripts[2]			
	GBLUP	GGBM	TBLUP	TGBM	GTBLUP	GTIBLUP	GTCBLUP	GTGBM
BMD12	1.16	1.48	1.52	1.36	1.47	1.63	0.88	1.23
BMD21	1.47	1.82	1.67	1.34	1.65	1.95	1.13	1.21
BW10	1.07	1.46	1.25	1.46	1.28	1.44	0.85	1.16
BW15	0.80	1.82	1.08	1.44	1.06	1.17	0.70	1.12
BW20	1.01	1.82	1.09	1.36	1.09	1.16	0.82	1.07
CHOL8	1.17	1.78	0.92	1.31	1.05	1.06	0.76	0.79
CHOL19	0.99	0.79	0.79	1.43	0.92	1.00	0.83	0.69
FATP12	0.55	0.76	0.96	1.16	0.99	1.09	0.89	1.18
FATP21	1.10	0.83	1.02	1.25	1.04	1.12	0.99	1.14
GLUC8	0.38	0.44	0.29	1.27	0.31	0.31	0.68	0.52
GLUC19	0.35	1.82	0.33	1.28	0.36	0.37	0.26	0.45
TRGL8	0.54	0.61	0.36	1.83	0.37	0.36	0.58	0.86
TRGL19	0.96	1.12	0.75	1.15	0.74	0.85	0.72	1.23
Mean	0.89	1.27	0.93	1.36	0.95	1.04	0.79	0.97
Min	0.35	0.44	0.29	1.15	0.31	0.31	0.26	0.45
Max	1.47	1.82	1.67	1.83	1.65	1.95	1.13	1.23

[1]For a description of the traits, see Table 1.
[2]For a description of the models, see Table 2.

820 **Table S2.** Relative root-mean squared error (RRMSE) for predictions on the validation set.
821 Lower values indicate lower prediction error.

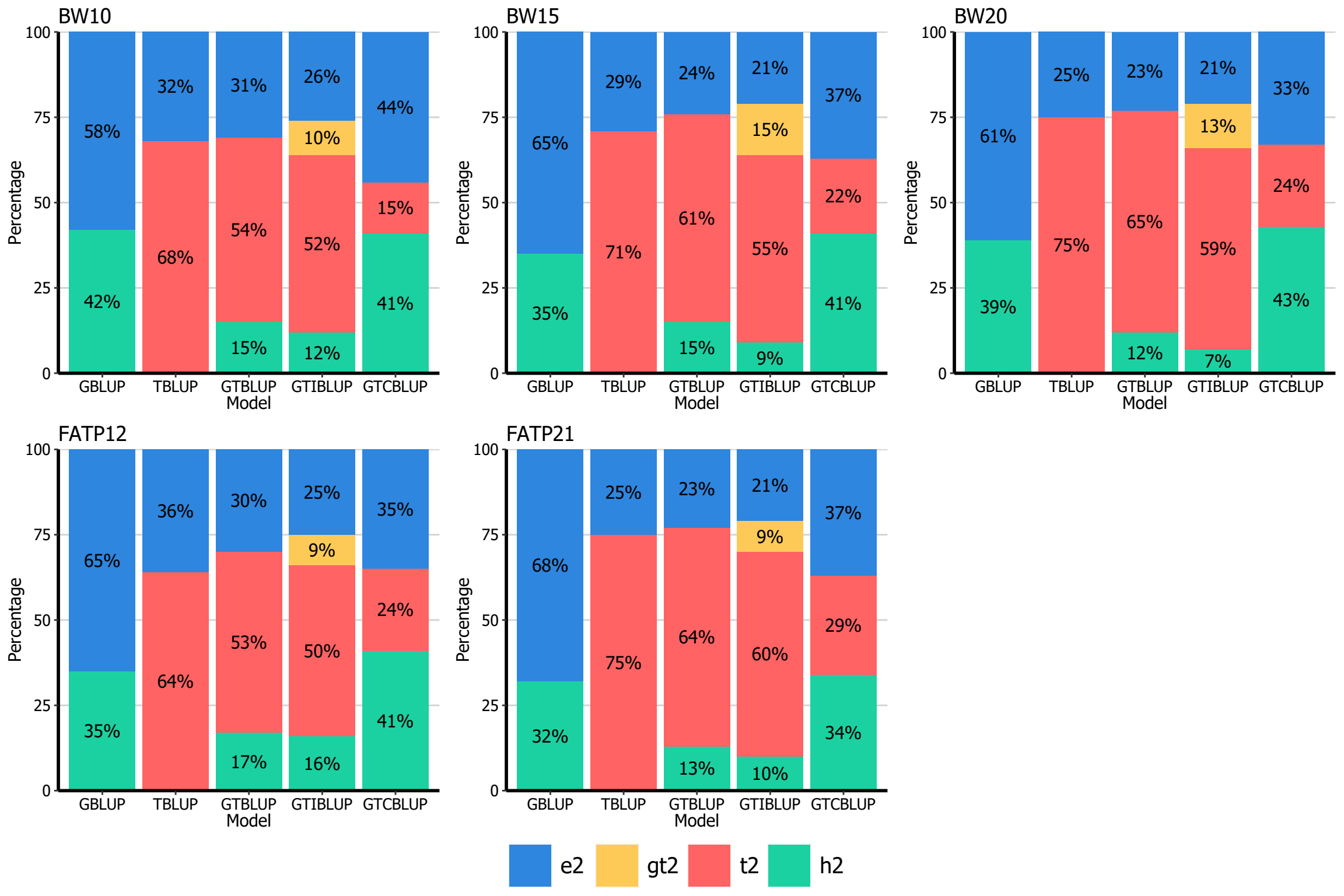
Trait[1]	Model[2]							
	Only SNP		Only gene transcripts		SNP + gene transcripts			
	GBLUP	GGBM	TBLUP	TGBM	GTBLUP	GTIBLUP	GTCBLUP	GTGBM
BMD12	1.00	1.02	0.99	1.02	0.98	0.99	1.01	0.99
BMD21	0.99	1.02	0.96	0.99	0.95	0.96	0.98	0.96
BW10	0.88	0.90	0.85	0.87	0.85	0.85	0.88	0.85
BW15	0.86	0.86	0.85	0.85	0.86	0.86	0.87	0.86
BW20	0.87	0.88	0.84	0.86	0.84	0.84	0.87	0.84
CHOL8	0.89	0.90	0.89	0.91	0.89	0.89	0.90	0.87
CHOL19	0.95	0.97	0.96	0.97	0.95	0.95	0.96	0.95
FATP12	0.91	0.91	0.86	0.87	0.86	0.86	0.89	0.87
FATP21	0.85	0.87	0.77	0.78	0.77	0.77	0.81	0.77
GLUC8	1.15	1.15	1.17	1.14	1.15	1.15	1.15	1.15
GLUC19	1.31	1.27	1.33	1.27	1.31	1.31	1.34	1.31
TRGL8	1.07	1.09	1.14	1.14	1.11	1.10	1.13	1.10
TRGL19	1.19	1.22	1.18	1.16	1.18	1.17	1.20	1.21
Mean	0.99	1.00	0.98	0.99	0.98	0.98	1.00	0.98
Min	0.85	0.86	0.77	0.78	0.77	0.77	0.81	0.77
Max	1.31	1.27	1.33	1.27	1.31	1.31	1.34	1.31

822 [1]For a description of the traits, see Table 1.
823 [2]For a description of the models, see Table 2.
824

Table S3. Estimates of model parameters for the CORE-GREML model: the percentage of variance explained by SNP genotypes (g^2) and gene transcripts (t^2), the estimated correlation between g and t components (ρ_{gt}), p-values for the estimated correlation ($p[\rho_{gt}]$) and p-values for the likelihood ratio tests to determine whether the model fit by CORE-GREML was better than that by the standard model not including a covariance component ($p[LKH]$).

Trait[1]	CORE-GREML parameters				GREML vs CORE-GREML		
	g^2	t^2	ρ_{gt}	$p[\rho_{gt}]$	GREML	CORE-GREML	$p[LKH]$
BMD12	0.24	0.20	+0.21	0.62	-181.4712	-181.4062	0.71
BMD21	0.28	0.26	+0.14	0.55	-177.2119	-177.1537	0.60
BW10	0.04	0.40	+0.58	0.29	-695.5398	-695.3162	0.50
BW15	0.03	0.45	+0.60	0.45	-696.1394	-696.7851	0.25
BW20	0.01	0.60	+0.63	0.32	-679.8870	-680.7453	0.19
CHOL8	0.27	0.13	+0.55	0.45	-483.6665	-483.9623	0.44
CHOL19	0.21	0.18	+0.71	0.07	-388.3944	-389.8146	0.09
FATP12	0.17	0.73	-0.23	0.50	-17.1618	-17.0426	0.62
FATP21	0.08	0.89	-0.47	0.34	-3.5900	-3.4460	0.59
GLUC8	0.21	0.09	-0.22	0.76	-510.4316	-510.4436	0.87
GLUC19	0.02	0.15	-0.26	0.72	-402.9910	-403.0818	0.67
TRGL8	0.08	0.14	+0.26	0.23	-226.2193	-226.9104	0.25
TRGL19	0.09	0.25	+0.07	0.90	-171.9591	-171.9595	0.97

[1]For a description of the traits, see Table 1.


BW10
BW15
BW20
FATP12
FATP21
Percentage
Model
GBLUP
TBLUP
GTBLUP
GTIBLUP
GTCBLUP
e2
gt2
t2
h2