ORIGINAL RESEARCH

The Plant Genome 🛛 🕮 🛛

Relative importance of genotype, gene expression, and DNA methylation on complex traits in perennial ryegrass

Marta Malinowska¹ Anja Karine Ruud¹ | Just Jensen¹ Kimon Fiil Svane² | Abraham George Smith³ | Andrea Bellucci¹ | Ingo Lenk⁴ | Istvan Nagy¹ | Mattia Fois¹ | Thomas Didion⁴ | Kristian Thorup-Kristensen² | Christian Sig Jensen⁴ Torben Asp¹

¹Center for Quantitative Genetics and Genomics, Aarhus Univ., Slagelse, Denmark

²Dep. of Plant and Environmental Sciences, Univ. of Copenhagen, Taastrup, Denmark

³Dep. of Computer Science, Univ. of Copenhagen, Denmark

⁴Research Division, DLF Seeds A/S, Store Heddinge, Denmark

Correspondence

Torben Asp, Center for Quantitative Genetics and Genomics, Aarhus Univ., Slagelse, Denmark. Email: torben.asp@qgg.au.dk

Assigned to Associate Editor Fanjiang Kong.

Funding information Innovation Fund Denmark, Grant/Award Number: 5184-00035B

Marta Malinowska and Anja Karine Ruud contributed equally.

Abstract

The growing demand for food and feed crops in the world because of growing population and more extreme weather events requires high-yielding and resilient crops. Many agriculturally important traits are polygenic, controlled by multiple regulatory layers, and with a strong interaction with the environment. In this study, 120 F₂ families of perennial ryegrass (Lolium perenne L.) were grown across a water gradient in a semifield facility with subsoil irrigation. Genomic (single-nucleotide polymorphism [SNP]), transcriptomic (gene expression [GE]), and DNA methylomic (MET) data were integrated with feed quality trait data collected from control and drought sections in the semifield facility, providing a treatment effect. Deep root length (DRL) below 110 cm was assessed with convolutional neural network image analysis. Bayesian prediction models were used to partition phenotypic variance into its components and evaluated the proportion of phenotypic variance in all traits captured by different regulatory layers (SNP, GE, and MET). The spatial effects and effects of SNP, GE, MET, the interaction between GE and MET ($GE \times MET$) and GE \times treatment (GE_{Control} and GE_{Drought}) interaction were investigated. Gene expression explained a substantial part of the genetic and spatial variance for all the investigated phenotypes, whereas MET explained residual variance not accounted for by SNPs or GE. For DRL, MET also contributed to explaining spatial variance. The study provides a statistically elegant analytical paradigm that integrates genomic, transcriptomic, and MET information to understand the regulatory mechanisms of polygenic effects for complex traits.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. The Plant Genome published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

Abbreviations: CEL, cellulose content; DIC, deviance information criterion; DMDig, dry matter digestibility; DRL, deep root length; DW, dry matter weight; GBS, genotyping-by-sequencing; GE, gene expression; MET, DNA methylomic; QTL, quantitative trait loci; RNA-seq, RNA sequencing; SNP, single-nucleotide polymorphism.

1 | INTRODUCTION

Events of extreme weather fueled by climate change already have adverse effects on crop production (Lesk et al., 2016). In Denmark, fewer but heavier rainfalls during summer and more frequent heat waves of longer duration are predicted to challenge agricultural production (The Danish Environmental Protection Agency, 2020). Perennial ryegrass (*Lolium perenne* L.) is the predominant forage grass in temperate climates, including Denmark, because of its rapid growth and high forage quality, but growth and yield are challenged by such weather extremes (Buttler et al., 2019; Cheplick et al., 2000; Kemesyte et al., 2017). Hence, we need a deeper understanding of the biological processes and molecular mechanisms behind the responses to different unfavourable events like excess or limited water.

According to The Central Dogma (Crick, 1958, 1970), the genome is the basal information layer of the cell, static by nature, and does not directly provide information about the biological mechanisms it encodes. Downstream from the genomic layer, dynamic functional omics layers provide high-dimensional endophenotypes (intermediate phenotypes) and potential sources of information. The transcriptome and proteome are directly encoded in the genome and modulated by environmental influences, whereas the shape of the metabolome, and largely the epigenome, are products of proteome functionality and environmental signals, respectively (Haas et al., 2017). Many biological processes are regulated by interactions between the genome and downstream regulatory layers (Haas et al., 2017), omic \times omic and omic \times environment interactions, thus serving as modulators of genetic and environmental effects on the final phenotype, that is, they can be seen as intermediate phenotypes.

Because of increased availability and decreasing prices of next-generation sequencing, combined with fewer constraints because of computational power, there is a growing interest in the potential of downstream omics data to bridge the gap between genotype and phenotype. Several studies have used omics data as intermediate phenotypes, most notably in gene expression (GE) quantitative trait loci (QTL) and metabolic QTL mapping. Such studies aim to identify candidate genes, diagnostic markers, or genetic elements underlying complex phenotypes in crop plants as reviewed in Druka et al. (2010). The DNA methylation (MET) patterns could explain up to 65% of the phenotypic variance for plant height in an Arabidopsis thaliana (L.) Heynh. isogenic epirecombinant inbred line population, only segregating for wild type and mutant DNA methyltransferase ddm1-2 (Hu et al., 2015).

Particularly, downstream omics data has the potential to capture nonadditive genetic effects including epistasis. The

Core Ideas

- Yield, quality, and deep root length were screened in a semifield, state-of-the-art root phenotyping facility.
- This study integrates genomic, transcriptomic and MET information to understand complex traits better.
- Bayesian regression models fit high-dimensional data and account for interactions between and within omics layers.
- By combining multiple omics layers, it is possible to capture a larger part of phenotypic variation.

statistical meaning of epistasis is defined as deviations from additivity in a statistical model, whereas the biological meaning is the variations resulting from gene actions (Fisher, 1918). Although epistatic models based on low marker density have been shown to increase prediction accuracy compared with additive models (Schrauf et al., 2020), other authors have found no (Jarquín, Kocak, et al., 2014) or a negative (Lorenzana & Bernardo, 2009) effect of epistasis. To detect significant gain in predictive ability by including marker-based epistasis requires a large population size and depends on the genetic relationships within the population. For example, Raffo et al. (2022) observed a significant increase in predictive ability compared with an additive genetic model using leave-one-line-out cross-validation in >2,000 winter wheat (Triticum aestivum L.) breeding lines. Using a leave-one-breeding-cycle-out scheme, where close relatives like full-sibs were left out, no improvement was detected, which the authors attribute mainly to the influence of relationships among the individuals in the reference and validation populations (Raffo et al., 2022). Meanwhile, integrating endophenotypes in the prediction models can contribute to biological explanations of epistatic and other interactions (Civelek & Lusis, 2014).

The phenotype is the result of the interactions between an organism's genotype and the environment. Because of such interactions, predicting genetic breeding values across environments can be challenging. Nevertheless, predicting the phenotype from genotypic and environmental information is of utmost interest in medicine and crop breeding (Grinberg et al., 2020), for example, by accounting (Svane, Dam, et al., 2019; Svane, Jensen, et al., 2019) for environmental risk factors in the prediction of disease outcome or development of obesity (Huang & Hu, 2015) or the performance of a crop cultivar in different environments (van Eeuwijk et al., 2019). Adding secondary data to explain genotype × environment (G × E) interactions has become a focus in plant breeding

in recent years, increasing prediction accuracy and predicting performance in new environments (Jarquín, Crossa, et al., 2014; Malosetti et al., 2016; Robert et al., 2020). Downstream regulatory omic layers, which are intermediate phenotypes influenced by both genetic effects and the environment, are examples of such secondary data.

Clustering methods based on single omics have been shown to successfully group, for example, subtypes of breast cancer in humans (Sørlie et al., 2001). By adding whole-genome transcriptome profiles in Bayesian prediction models, Vazquez et al. (2016) demonstrated that increases in predictive ability of breast cancer outcome were not due to such subtype clustering but to other patterns in the gene expression data. Whole-genome methylation profiles offered a higher prediction power and accuracy than any commonly used covariates for predicting breast cancer survival (Vazquez et al., 2016).

Predicting performance and combining abilities in hybrids usually achieves relatively low predictability from parental genotype and phenotype, and heterosis is a product of complex gene × gene interactions and epigenetics (Zenke-Philippi et al., 2016). Thus, there have been many efforts to include soluble RNA, messenger RNA, and metabolites in hybrid performance prediction models in maize (*Zea mays* L.) (Fu et al., 2012; Schrag et al., 2018; Westhues et al., 2017; Zenke-Philippi et al., 2016) and rice (*Oryza sativa* L.) (Wang et al., 2019). For example, Dan et al. (2016) showed that metabolites provided high predictive abilities for predicting hybrid performance in rice using partial least square regression.

Azodi et al. (2020) developed prediction models trained on transcriptome data from the maize pangenome population (diverse inbred lines) to understand genetic mechanisms for flowering time, plant height, and yield. They found that genetic markers and transcriptome data captured different aspects of the phenotypic variation and that transcriptomes from the seedling stage (Hirsch et al., 2014) could predict phenotypes at the adult stage better than a baseline model using population structure with the first five principal components of genetic marker basis. However, the prediction of additive genotype based on genetic marker data and models, including transcript level data for phenotype prediction, were not compared appropriately.

Under field and semifield conditions, plants can experience heterogeneous conditions with variable spatial and temporal patterns. The plants' responses to combinations of stresses often have nonadditive effects on the phenotypes (Atkinson & Urwin, 2012; Cruz et al., 2020; Rasmussen et al., 2013; Thoen et al., 2017). Omics data generated under field conditions is noisier than data collected in controlled environments (Schrag et al., 2018). Despite the challenges, data from fieldgrown plants is essential for closing the gap between field and lab (Alexandersson et al., 2014; Cruz et al., 2020; Nelissen et al., 2020). Cruz et al. (2020) investigated the spatial autocorrelation for phenotype, metabolite, and GE data sampled across a field of the inbred maize line B104. They suggested that patterns of spatial autocorrelation among metabolites and transcripts, and part of the variability observed among individual plants, are due to microenvironmental factors with a spatial structure. In prediction models, such spatial effects should be disentangled from the true genetic effects in the omics data. Ignoring interactions with the genotype, including complex environmental–microenvironmental factors, can lead to biased estimates of residual variance and single-nucleotide polymorphism (SNP) heritability (Ni et al., 2019). González-Reymúndez et al. (2017) found that the effects of omic × treatment interactions explained variance in human breast cancer survival time not captured by clinical covariates.

In this study, we investigated the role and relative importance of genomic (SNP), transcriptomic (GE), and MET data to predict the yield, quality, and deep root length (DRL) of 120 F₂ families of perennial ryegrass. The plants were grown across a water gradient in a semifield facility with controlled subsoil irrigation (Svane, Jensen, et al., 2019) and phenotyped for three feed quality traits that can be affected by water availability. Root growth traits are important for drought tolerance, and DRL measured with multispectral imaging was included. We computed a K-nearest weighted spatial connectivity matrix to capture spatial effects. We then ran a sequence of Bayesian models, implemented in the BGLR package (Pérez & de los Campos, 2014), to partition the variance explained by spatial effects, omics, omic × omic interaction, and $GE \times$ treatment interactions. We further investigated the effects of the individual and different combinations of sets of predictors on prediction accuracy. By this effort, we were able to partition the genetic, spatial, and environmental components of the GE and methylation layers. To the best of our knowledge, neither treatment \times omic interactions, nor the influence of spatial effects in omics-augmented prediction models with different genotypes, have been investigated in plants.

2 | MATERIALS AND METHODS

2.1 | Plant material and experimental design

The experiment was conducted in RadiMax (Svane, Jensen, et al., 2019), a semifield root phenotyping facility located at Copenhagen University's experimental farm (Taastrup, Denmark) in 2017. The trial was conducted in two separate beds (hereafter referred to as Bed 3 and Bed 4), each divided into half-beds (one half-bed is an experimental unit, see Figure 1). Each bed is operated with independent subsurface water management systems (Figure 1) and has a V-shaped concrete bottom with soil depth increasing from 0.8 to 2.1 m in each unit (maximal depth: 2.1 m; slope: 15.8°; 18 subirrigation



FIGURE 1 Cross-section of a bed in RadiMax, with division in experimental units and specification of drought (Dry) and control (Wet) parts of the unit. The red 'X' represents approximately where, within each family or row, the sampling for omics data were collected. Each F_2 family was grown in rows across two units, that is, from the center to the edge of the facility. Sampling for transcriptomic and methylomic data was done from the unit equipped with minirhizotron tubes. Water was controlled by the subsurface irrigation lines, where plants toward the center of the facility had to grow deeper roots to reach water. The facility is described in detail in Svane, Jensen, et al. (2019) and the root imaging set up in Svane, Dam, et al. (2019)

tubes). The subsurface irrigation system was described in detail by Svane, Jensen, et al. (2019). In brief, driplines were placed along the bottom of each unit at the 0.20-m interval. Perforated pipes leaked water flowing by gravity, whereas the capillary soil movement provided irrigation to the plants. Subirrigation was controlled by the sensor system monitoring the water uptake (Supplemental Figure S1). Because of the water availability gradient created by a depth to the subsoil irrigation system and movable rain-out shelters, plants growing toward the deeper part of the unit had access to limited amounts of water and were expected to exhibit drought stress responses (called 'drought' treatment for simplicity). Units are fitted with minirhizotrons along one sloping bottom, allowing for root observation at various soil depths along half of each row (or an experimental unit). One-hundred twenty diploid F₂ families of forage type perennial ryegrass were randomly assigned to each of Beds 3 and 4 sown in October 2016 with a planting density of 2 gm^{-2} (0.5 gm^{-1} with 25 cm between the rows) and grown for 1 year in the facility. The plants were grown per F_2 family in rows (9.7 m long) across two units directly above the minirhizotrons tubes with a 25-cm distance between each row. Thus, a unique row number also refers to a unique F₂ family grown in that row. In this experiment, only half of the row that was located above the

minirhizotron tubes were considered (Figure 1), that is, one entire unit per bed.

2.2 | Phenotypic records

The grass was cut three times during the season. All cuts were carried out at 6 cm aboveground level with a Nibbi Brik 3 walk-behind tractor with cutter bar ESM 1.22 mt (NIBBI). Each row was divided into four subsamples at harvest-two from the deep midsections ('drought' treatment) and two from the border sections ('control')-and analyzed separately for yield- and quality-related phenotypes. As samples for RNA and methylomic data were collected from one unit per F₂ family, we have only included phenotypic data from the same unit (Figure 1), that is, two data points per F_2 family. The quality measurements were done by DLF Seeds (Store Heddinge, Denmark) using protocols for the ANKOM 2000 fiber analyzer (ANKOM Technology). The third cut was chosen for data collection, including sampling of the omics data, because of the timing and expected effect of the drought treatment. Three aboveground phenotypes were included: dry matter weight (DW) of the third cut, cellulose content (CEL), and dry matter digestibility (DMDig) in percentage of total DW. For DW, grass samples were stored for a maximum of 1 h in a cool environment and dried at 60 °C in a hot-air circulating oven until no further weight loss could be measured (24-40 h of drying). Then, the dried samples were milled in a Retsch SM 300 cutting mill to 1-mm particle size, and, after mixing, samples were subjected to chemical analysis.

Cellulose content (in percentage of dry matter) was estimated based on the acid detergent fiber (ADF) (ANKOM Technology, 2017a) and acid detergent lignin (ADL) (ANKOM Technology, 2022) values in percentage of dry matter. In principle, the cellulose content was described as the ADF content minus ADL (CEL = ADF – ADL), all based on percentage dry matter.

Dry matter digestibility was based on the neutral detergent fiber (ANKOM Technology, 2017b), which represents the cellulose, hemicellulose, and lignin part of the cell wall. After neutral detergent fiber analysis, the cell wall samples were treated with a cellulase mix (CELLULASE R-10 from Duchefa Biochemistry), and the indigestible neutral detergent fiber, neutral cellulase fiber (NCF), was measured by weighing (percentage of dry matter) (ANKOM Technology, 1998) using a modified protocol without gamanase. The DMDig is the fraction of the cell wall not digestible by the cellulase enzyme mix after 40 h of incubation at 39 °C. It is assumed that everything else is 100% digestible so that DMDig% = 100% - NCF%.

A total of 78,200 root images were taken with the multispectral camera system as described in Svane, Dam, et al. (2019) on four dates in 2017. These dates were the 10 May, 29 May, 6 July, and 22 August. The first imaging coincided with the beginning of the growing season, and the remaining three dates corresponded to a regular cutting regime under Danish growth conditions and the increasing drought stress, as assessed by a sensor network consisting of volumetric water content and temperature sensors (Acclima, Inc.) (Svane, Jensen, et al., 2019). The images were taken of roots through minirhizotron tubes {acrylic glass [poly(methyl methacrylate)]} installed beneath each row at a 15° angle to the horizontal, allowing imaging from 0.4 to 1.8 m depth. Pseudo-RGB images, generated using the red (660 nm), green (590 nm) and blue (450 nm) waveband from all four dates, were used for segmentation model training with RootPainter (Smith, Han, et al., 2022). From these 78,200 images, 4,000 were selected at random using the 'create training dataset' functionality from RootPainter with a target size specified as 900 \times 900. In short, a U-Net based convoluted neural network system in Python (v3.6.4.) using PyTorch (Paszke et al., 2017) was used to segment the root images. The first 450 images were annotated correctively with attention to detail as described in Smith, Han, et al. (2022). The subsequent 1,350 images were inspected at a faster pace to identify larger error regions or outliers, which would then be annotated correctively. Model training was then stopped, and all 78,200 images were segmented. Then, root length was extracted from the generated segmentations via skeletonization and pixel counting (Smith, Petersen, et al., 2020). The interactive model training, including annotation, model training, and reviewing model predictions, took 5.5 h using two NVIDIA RTX 2080 Ti GPUs. Segmenting all 78,200 images took ~1 d. In total, 22,175 images from 6 July were used for this study. Of the 300 tubes, image data from eight tubes were missing because of overheating failure of the camera computer and operator error during the day of imaging. The extracted root length from the segmentation was correlated to a manually curated grid count of 200 randomly chosen and not annotated images from the data set (Supplemental Figure S2), giving a R^2 value of .85.

Furthermore, six tubes were excluded because of bent minirhizotrons preventing the camera from reaching the deepest sections of the tubes. The DRL root profile trait was then computed from individual root image length measurements from 6 July using 1/470 pixels cm^{-1} as a calibration factor. The DRL was computed as the sum of root length (cm) below a cut-off soil depth of 110 cm. We based the length of 110 cm as the depth the average intensity in the dataset started to decline, thus we are only interested in the difference in length in the deeper part of the root system. For aboveground measurements, two phenotype data points were obtained per family, one from 'drought' and one from 'control' plants (Figure 1), whereas for root length, only one measurement per family was considered for further analysis, that is, based on the sum of roots 110 cm below the soil surface. In other words, we looked

at the total amount of roots below 110 cm as a phenotype not the root length in a specific location in the facility.

2.3 | Omics data

Omics data included SNPs from combined genotyping-bysequencing (GBS) and transcript-based SNPs from the RNA sequencing (RNA-seq) data, GE profiles from RNA-seq, and reduced-representation MET. Leaf samples for total GE and MET analyses were collected in RadiMax, coinciding with the third cut and sampling of material from which the aboveground phenotypes were also derived. For each F_2 family, samples consisting of 15–20 leaves were taken from two areas ('drought' and 'control'), generating two records per family. The samples were snap frozen in liquid N, ground in a Geno/Grinder 2000 (SPEX CertiPrep), and the ground tissue distributed for MET and RNA extractions.

To obtain methylomic data, epiGBS libraries were prepared as previously described (Malinowska et al., 2020) and sequenced on the Illumina HiSeq4000 platform $(2 \times 150$ bp) by Beijing Genomics Institute (Shenzhen, China). The sequencing data were processed using the WellMeth pipeline (Malinowska et al., 2020). Demultiplexed and trimmed reads were mapped to the perennial ryegrass pseudo-chromosome assembly (Nagy et al., 2022). The identification of MET sites was followed by polymerase chain reaction duplicate removal. The methylation level of each cytosine covered by the analysis was calculated as a proportion of methylated cytosine (#C) to sequencing depth of a position (methylated and unmethylated #T) [#C/(#C + #T)]. Single-base DNA methylation values ranged between 0 (not methylated) and 1 (completely methylated).

The RNA was extracted with the Sigma Aldrich Total Plant RNA kit, following their standard protocol for normal leaf tissues. Gene expression profiles were obtained using RNA-seq technology sequenced on the Illumina HiSeq4000 platform (2×100 bp, ~20 M reads per sample) by the Beijing Genomics Institute (Shenzhen, China). Transcript abundance was calculated with KALLISTO (Bray et al., 2016). The HISAT2 software (Kim et al., 2019) was used to align the transcripts to the perennial ryegrass reference genome (Nagy et al., 2022). Samtools (Li et al., 2009) was used to add read groups and mark and delete duplicates. Finally, SNPs were called with freebayes (Garrison & Marth, 2012) with settings set to minimum mapping quality of 30 and minimum coverage of eight.

The GBS sampling and library preparation was performed according to Byrne et al. (2013) and Elshire et al. (2011). Briefly, leaf material for GBS data was sampled from \sim 100 seedlings per F₂ family grown on Rockwool blocks. Genomic DNA extracted from these samples was digested

using the ApeKI (5-bp recognition site) and PstI (6-bp recognition site) restriction enzymes and sequenced on an Illumina HiSeq4000 platform. Genetic variants were called using GATK v3.7 and filtered on a mapping quality of 30.

The SNP datasets from RNA-seq and GBS sequencing were merged based on physical position and family. The 10,432 SNPs that were found to be in common between the two datasets were merged based on their allele frequencies, that is, DNA fragments carrying alternative and reference alleles were summed prior to allele frequency (AF) calculation. Afterward, SNPs were filtered based on sequencing depth >10, SNP missingness <50%, and minor allele frequency >0.05). Finally, family AF for each biallelic SNP was extracted as the sequencing depth for DNA fragments carrying the alternative base divided by total sequencing depth. Thus, numeric values ranged between 0 and 1. Missing data points were replaced by mean AF per SNP.

Initially, variant datasets included 192,814 and 1,787,068 SNPs for GBS and RNA-seq experiments, respectively. These were merged and filtered as described above to obtain 357,217 SNPs that were finally converted into AFs. A total of 139,003 transcripts were identified by KALLISTO (Bray et al., 2016). After filtering out transcripts with zero expression and median transcript per million (TPM) <1, $\log_2(TPM + 1)$ values for 22,602 transcripts were included in the analysis. The MET data were filtered based on a minimum read coverage of three and a maximum of 50% missing values per sample. If a family was removed in one dataset (control or drought), it was also removed in all the other available data. A total of 276.962 unique MET positions were kept for further analysis. The remaining missing methylation values (31% of all values) were replaced with average methylation per site over all samples.

2.4 | Spatial distribution characteristics

Moran's eigenvector maps were derived from K-nearest weighted spatial connectivity matrix with R packages adespatial (Dray et al., 2018) and ade4 (Dray & Dufour, 2007). The assumed distance between two neighboring rows equals one. The spatial information was used to compute the spatial autocorrelation of analyzed traits. Moran's coefficient estimates the strength of interdependence between each phenotypic score by comparing the value of x_i at location *i* with the value x_i at all other locations (Zhou and Lin, 2016). In other words, spatial autocorrelation expresses a correlation in a signal among neighboring locations in multidirectional space and evaluates whether the pattern is clustered, dispersed, or random. The spatial effects were evaluated on a by-row basis, that is, covering both drought and control treatment for each sample. Thus, the spatial effects are independent of treatment and within-F2 family genetic effects.

2.5 | Statistical models

A genomic relationship matrix for additive effects was calculated based on VanRaden (2008) modified to include family instead of individual AF (Fè et al., 2015). The AF matrix \mathbf{M} was centered by mean AF per SNP and multiplied by its transposed \mathbf{M}' and scaled by a constant.

The formula applied was as follows:

$$\mathbf{G} = \frac{\mathbf{M}\mathbf{M}'}{0.25\sum \bar{F}_j \left(1 - \bar{F}\right)_j},\tag{1}$$

where the scale parameter 0.25 is due to the presence of four alleles in F_2 families. \overline{F}_j refers to average AF for the marker *j*. Furthermore, **G** matrix diagonal elements were corrected for extra binomial variance because of reduced sequencing depth as derived by Cericola et al. (2018) for Equation 2:

$$D_{\mathrm{C}i} = D_{\mathrm{O}i} \left(1 - \frac{1}{\overline{\mathrm{SD}}_i + 1} \right),\tag{2}$$

where a corrected diagonal element D_{Ci} for every *i*th individual is obtained from the initial value D_{Oi} transformed based on individual mean sequencing depth \overline{SD}_i . Only diagonal elements are corrected because off-diagonal elements are not biased by sequencing depth because of the independence of sequencing depth of different loci.

Similarity matrices for gene expression and DNA methylation data were computed following the basic formula:

$$K = \frac{\mathbf{X}\mathbf{X}'}{p} \tag{3}$$

where \mathbf{X} is a matrix (centered and scaled) of considered features, p is the number of features (transcripts or MET sites).

A total of seven prediction models (Table 1) were evaluated based on data from all families grown under two treatments. Aside from three omics effects, the effect of the water availability (treatment), the spatial effect was included to account for micro- and macroenvironmental differences (like soil compactness, nutrient availability, shading. or location in the field) observed across units. The baseline model (Equation 4) defined the response of observed phenotypes (y; N = 240) and included treatment effect (T_i) assigned a flat Gaussian prior with mean zero and variance equal to 1×10^{10} , plus the normal, independent, and identically distributed random effects such as $l_i \sim \text{NIID}(0, \sigma_1^2)$, $\varepsilon_{ij} \sim \text{NIID}(0, \sigma_{\varepsilon}^2)$, and multivariate normal prior distribution $s_i \sim \text{MVN}(0, K_8 \sigma_8^2)$, where σ_1^2, σ_8^2 , σ_{ε}^2 are family, spatial neighborhood, and residual variances, respectively.

Baseline:
$$y_{ii} = \mu + T_i + l_i + s_i + \varepsilon_{ii}$$
 (4)

TABLE 1 Main effect and interaction of the seven models used to fit the data set

7 of 17

Model	SPAT	FAMILY	SNP	MET	GE	MET × GE	GE _C	GED	ε
M1	Х	Х							Х
M2	Х	Х	Х						Х
M3	Х	Х	Х	Х					Х
M4	Х	Х	Х		Х				Х
M5	Х	Х	Х	Х	Х				Х
M6	Х	Х	Х	Х	Х	Х			Х
M7	Х	Х	Х	Х	Х	Х	Х	Х	Х

Note. GE, transcriptomic effect; GE_c , transcriptomic × control effect; GE_D , transcriptomic-by-drought treatment effect; MET, methylomic effect; MET × GE, methylationby-expression effect; SNP, genetic additive effect; SPAT, spatial effect.

Subsequently, we extended the baseline model by sequentially adding omic information (Table 1). The full model, including all omics and interaction terms is defined in Equation 5: model is applied for DRL in Equation 6:

$$y_{ij} = s_i + g_i + u_{ij} + m_{ij} + u_{ij} + \varepsilon_{ij}$$
(6)

where each element was specified as above.

$$y_{ij} = \mu + T_i + l_i + s_i + g_i + u_{ij} + m_{ij} + u_{Ci} + u_{Di} + \varepsilon_{ij}$$
(5)

Where baseline model was extended by random effects of markers (SNP) with a multivariate normal prior distribution $g_i \sim \text{MVN}(0, \mathbf{G}\sigma_{\sigma}^2)$ followed by genomic main effects associated with GE, $u_{ij} \sim \text{MVN}(0, \mathbf{K}_{\mathbf{u}} \sigma_{\mathbf{u}}^2);$ MET, $m_{ij} \sim \text{MVN}(0, \mathbf{K}\sigma_{m_m}^2)$, as well as GE under both levels of water availability either when drought (GE_D) was applied $u_{Di} \sim \text{MVN}(0, \mathbf{K}_{ud}\sigma_{ud}^2)$ or not (GE_C) $u_{Ci} \sim \text{MVN}(0, \mathbf{K}_{uc} \sigma_{uc}^2)$. Therefore, **G** was a marker-derived genomic relationship matrix as calculated in Equation 1 and σ_{α}^2 was genomic variance. Similarity matrices for each layer of omic data (K_u , K_m , K_{ucd} , and K_{ud}) were calculated as specified in Equation 3 where σ_u^2 , σ_m^2 , σ_{uc}^2 , σ_{ud}^2 were variance parameters associated with GE, MET, and expression profiles under two levels of water availability. Additionally, the model was expanded by the interaction between gene expression and DNA methylation (MET × GE) $um_{ij} \sim MVN(0, \mathbf{K}_{um}\sigma_{um}^2)$, with $\mathbf{K}_{um} = \mathbf{K}_{u} \# \mathbf{K}_{m}$ (# represents cell-by-cell multiplication operator, known as the Hadamard product) and σ^2_{um} was the variance parameter associated with the interaction term. All variance parameters of the random effects were assigned priors with scaled-inverted χ^2 distribution and degrees of freedom close to zero (df0 = 0.0001) and R^2 parameter set equal to .8, with the default choices implemented using the Bayesian ridge regression specification in the Bayesian generalized linear regression (BGLR) R package.

Apart from aboveground phenotypes, a root trait, DRL, was analyzed in this study. Because there was one root record for each minirhizotron tube (therefore also family), a different model needed to be applied for this trait. The following full

2.6 | Assessment of predictive ability

The performance of models was evaluated using the leaveone-family-out cross-validation approach, dividing the data set into n training-testing subsets (n = number of families = 120), which led to 120 combinations. This procedure was repeated 20 times, and a total of 2,400 runs were performed for each training-testing combination. Predictive abilities were calculated as the Pearson's correlation between the families' phenotypes, corrected for the treatment ('fixed') effects estimates, and the predicted values for each treatment separately. Family and spatial effects were not included in the calculation of predictive abilities because of the assumption of independence between lines and location within the semifield. Hence, the baseline model was not included in the cross-validation scheme. Models were compared based on the average Pearson's correlation and standard deviation from each run. The predictive abilities of the models were assessed based on the ability of each model to predict phenotypic traits (above- and belowground) after accounting for the general effects of the water availability, that is, the treatment effects. The prediction's degree of inflation or deflation (bias) was measured by estimating the regression coefficient between observed and predicted values. Models with no inflation are expected to have a regression coefficient equal to one. The model fit was also evaluated by the deviance information criterion (DIC) (Spiegelhalter et al., 2002), where a smaller DIC value is considered better.

To evaluate the difference in predictive ability between models for each treatment, the Hotelling–Williams t test was

TABLE 2 Descriptive statistics for dry weight, cellulose content, and dry matter digestibility

Trait	Treatment	No.	Average (SD)	Range	Coefficient of variation
					%
DW (g)	control	120	78.0 (15.8)	40.4–114.7	20.2
	drought	120	100.8 (23.3)	49.7–165.0	23.1
DMDig (percentage of DM)	control	120	89.6 (1.6)	82.1–93.9	1.8
	drought	120	89.0 (1.3)	84.8–92.4	1.5
CEL (percentage of DM)	control	120	21.0 (1.2)	18.3–25.3	5.8
	drought	120	21.0 (1.2)	18.4–25.2	5.8
DRL (cm)	N/A	117	96.4 (43.5)	10.6-225.0	45.2

Note. CEL, cellulose content as percentage of DM; DMDig, dry matter digestibility as percentage of DM; DRL, deep root length (one measurement per F₂ family); DW, herbage dry matter (DM) weight.

applied (Dunn OJ & V, 1971). Differences were considered significant if the p value of the test was <.01.

2.7 | Software

Models were fitted using the R package BGLR (Pérez & de los Campos, 2014) in the R v3.5.1 environment (R Core Team, 2018). For each model, a single chain of 150,000 iterations was run with a burn-in setting of 20,000, and the remaining samples were thinned at the interval of 20 (Supplemental Figure S3). The convergence of the posterior chains to a stationary state was evaluated following Gelman's and Rubin's approach using chains in the R package CODA (Plummer et al., 2006). Scripts showing how to perform all the analyses are contained in Supplemental File S1.

3 | RESULTS

3.1 | Phenotypic data

Descriptive statistics of the four traits analyzed in this study are presented in Table 2. Mean values of traits under drought and control conditions were compared using ANOVA test. The phenotypic range was widest for DRL and narrowest for CEL. There was a statistically significant difference between the two treatments for DW and DMDig but not for CEL on a population level.

3.2 | Spatial distribution characteristics

Moran's spatial autocorrelation values were statistically significant (p = .001 for all traits), indicating that phenotypic values in the dataset were more spatially clustered than would be expected if underlying spatial processes were random, confirming the need for spatial correction in the predictive models. The variation in phenotypic traits is shown in Figure 2. Spatial patterns across both units of each trait are not a consequence of water availability but nonhomogeneous variation under semifield conditions, for example, uneven soil packing independent of treatment.

3.3 | Estimates of variance components

Three aboveground traits, DW, CEL, DMDig, and one belowground trait, DRL, were used as response variables in this study. We fitted a sequence of models designed to partition phenotypic variance into its components and evaluate how much more of it can be captured by introducing different layers of omics data. Estimates of variance components derived from the analysis of the aboveground traits are shown in Table 3.

The baseline model (M1) included treatment (flat prior), spatial variation, and family capturing between 60 (DMDig) and 73% (CEL) of phenotypic variance $(\sigma_s^2 + \sigma_1^2)/(\sigma_s^2 + \sigma_1^2 + \sigma_\epsilon^2)$ (Table 3). Expanding the baseline model by adding genomic effect (M2) had a negligible effect on residual or variance explained by spatial neighborhood. Addition of an additive genomic effect allocated over half of the variance from family to additive genetic effects (M2) for all traits. In both models (M1 and M2), variance captured by spatial effect is moderate for DW (~20%). However, for two other aboveground traits, variance corresponding to spatial effect is comparable or even bigger than summed family and SNP effect, thus highlighting the importance of accounting for microenvironmental variations under semifield conditions.

When the MET effect was added (M3), the total proportion of variance explained increased by $\sim 11\%$ for DW and CEL and 22% for DMDig. For all three aboveground traits, the addition of MET contributed almost exclusively to explaining residual variance, which indicates a potential for epigenetic variation to account for a considerable proportion of phenotypic variance of complex traits.

Model ^a	$\sigma_{\rm s}^2$	ճ ²	<mark>6</mark> 2 ც	$\sigma_{\rm m}^2$	σ_u^2	σ^2_{um}	σ_{uc}^2	σ^2_{ud}	σ_{e}^{2}	DIC
Dry ma	tter weight		I					1		
MI	0.17 (0.08–0.26)	0.33 (0.22–0.44)	1	1	1	1	I	1	0.28 (0.21-0.35)	454.8
M2	0.16 (0.07–0.25)	0.17 (0.07–0.27)	0.25 (0.11–0.39)	1	I	I	I	I	0.27 (0.21-0.34)	449.9
M3	0.16 (0.07–0.25)	0.17 (0.07–0.28)	0.24 (0.10-0.39)	0.11 (0.06–0.18)	I	I	I	I	0.17 (0.10-0.25)	384.4
M4	0.13 (0.06-0.21)	0.14 (0.06-0.24)	0.20 (0.07-0.33)		0.21 (0.08-0.35)	I	I	I	0.16 (0.08-0.23)	360.4
M5	0.13 (0.05-0.22)	0.14 (0.05–0.25)	0.20 (0.07-0.34)	0.09 (0.04-0.14)	0.17 (0.06-0.30)	I	I	I	0.11 (0.05-0.18)	300.4
M6	0.13 (0.05–0.22)	0.14 (0.05–0.25)	0.20 (0.06-0.34)	0.07 (0.04-0.12)	0.14 (0.05–0.26)	0.07 (0.03-0.11)	I	I	$0.09\ (0.03-0.16)$	263.3
M7	0.10 (0.03-0.17)	0.09 (0.03-0.16)	0.12 (0.04-0.23)	0.07 (0.03-0.11)	0.12 (0.04-0.22)	0.06 (0.03-0.10)	0.13 (0.04–0.25)	0.09 (0.03-0.18)	0.10(0.04-0.16)	271.0
Dry ma	tter digestibility									
M1	0.28 (0.17-0.41)	0.23 (0.14-0.33)	I	I	I	I	I	I	0.32 (0.24-0.40)	477.6
M2	0.27 (0.15–0.40)	0.13 (0.05–0.21)	$0.19\ (0.08-0.30)$	I	I	I	I	Ι	0.30 (0.23-0.38)	472.0
M3	0.28 (0.16-0.42)	0.13 (0.05-0.22)	0.20 (0.08-0.33)	0.18 (0.10-0.25)	1	1	I	1	0.11 (0.04-0.19)	277.6
M4	0.21 (0.09-0.34)	0.11 (0.04-0.19)	0.15 (0.06-0.26)	1	0.31 (0.17–0.48)	I	I	Ι	0.11 (0.05–0.18)	289.0
M5	0.21 (0.09-0.34)	0.11 (0.04-0.19)	0.15 (0.06-0.27)	0.10 (0.04-0.16)	0.20 (0.08-0.34)	1	I	1	0.08 (0.03-0.14)	220.7
M6	0.21 (0.09-0.34)	0.10 (0.03-0.18)	0.15 (0.05-0.26)	0.08 (0.03-0.13)	0.17 (0.06-0.29)	0.08 (0.03-0.12)	I	Ι	0.06 (0.02-0.11)	174.4
M7	0.12 (0.03-0.22)	0.07 (0.02-0.12)	0.09 (0.03-0.17)	0.07 (0.03-0.13)	0.13 (0.04-0.24)	0.07 (0.03-0.12)	0.12 (0.03-0.23)	0.10 (0.03-0.20)	0.06 (0.02–0.11)	179.4
Cellulos	ie content									
MI	0.32 (0.17-0.46)	0.36 (0.24-0.48)	I	1	1	1	1	I	0.25 (0.19-0.32)	437.5
M2	0.31 (0.17–0.46)	0.17 (0.07–0.28)	0.27 (0.13–0.44)	I	I	I	I	I	$0.25\ (0.19-0.31)$	432.1
M3	0.32 (0.17-0.47)	0.17 (0.06-0.30)	0.27 (0.10-0.43)	0.12 (0.06–0.18)	I	1	I	I	0.14 (0.06-0.20)	333.1
M4	0.20 (0.09-0.31)	0.13 (0.06-0.22)	0.19 (0.07–0.34)	I	0.34 (0.22–0.47)	1	I	I	0.05 (0.02-0.09)	141.4
M5	0.21 (0.10-0.32)	0.13 (0.05-0.22)	0.19 (0.06-0.34)	0.05 (0.03-0.08)	0.29 (0.17–0.42)	I	I	I	0.05 (0.02-0.08)	123.7
M6	0.21 (0.09-0.33)	0.12 (0.04-0.22)	0.18 (0.06-0.33)	0.05 (0.02-0.08)	0.26 (0.14–0.40)	0.05 (0.03-0.08)	I	I	0.05 (0.02-0.08)	117.1
M7	0.16 (0.06–0.28)	0.08 (0.03–0.15)	0.11 (0.03–0.21)	0.04 (0.02–0.07)	0.22 (0.11–0.33)	0.04 (0.02–0.07)	0.10 (0.03-0.20)	0.11 (0.03–0.22)	0.05 (0.02–0.08)	123.4
Note. DIC,	deviance information cr	iterion; σ_s^2 , spatial neigh	hborhood variance, $\sigma_{\rm I}^2$, f	family or residual genet	tic variance; σ_g^2 , genom	ic variance; and σ_{m}^{2} , σ_{u}^{2} ,	$\sigma^2_{um},\sigma^2_{uc},and\sigma^2_{ud}$ were	variance parameters a	ssociated with DNA me	ethylation,

Estimated variance components (95% highest density interval of posterior distributions) derived from Gaussian Bavesian ridge regression model for aboveground traits TABLE 3

gene expression, interactions between DNA methylation and gene expression, and expression profiles under two levels of water availability; σ_{e}^{2} , residual variance ^aMI-M7 are model numbers of the aboveground traits. The effects included in each of them are described in columns.

MALINOWSKA ET AL.



FIGURE 2 Variation of mean phenotypic values of four analyzed traits across two experimental units (in Beds 3 and 4) in the RadiMax facility. Light orange and light purple backgrounds indicate Bed 3 and Bed 4, respectively. Each row number indicates the position of a unique F_2 family in the facility

Incorporating the expression profile into the model (M4) further increased the proportion of the variance explained and captured a sizeable proportion of variance resulting from the microenvironment (σ_s^2), family, and SNPs. Estimated variance associated with GE varied from 0.21 for DW to 0.31 and 0.34 for DMDig and CEL, respectively. In other words, GE had both genetic and microenvironmental (spatial) components, whereas methylation seemed independent of spatial and genetic structures in the (aboveground) data.

Combining all omics data (M5) and adding the interaction term between MET and GE (M6) did not lead to a substantial reduction in the estimated residual variance (CEL) or did it by only a small margin (from 0.11 to 0.09 for DW and from 0.08 to 0.06 for DMDig) relative to model M4. However, the inclusion of both omics and interaction between the two (M5 and M6) reduced the variance explained by each omics component compared with single omic models (M3 and M4).

Separating the variance explained by the interaction between omics is of interest for understanding how one omic layer is associated with another. Variance estimates from the most comprehensive model (M7) suggest that by the inclusion of various omic layers, as well as interactions between them and treatment, the residual variance may be reduced from 64 (DW) and 80% (CEL and DMDig). A full model (M7) also had the best fit with the lowest DIC for DW and DMDig. Furthermore, the proportion of within-treatment variation explained by GE × treatment interaction (23–26%) shows the importance of considering such interactions, especially in trials with insufficient randomization or few replicates and significant micro- and macroenvironmental variations. However, it is worth noting that parts of the omic and treatment effects could be explained by their interactions in models M6 and M7. Nonetheless, for all three traits, in comparison to M2, incorporation of either MET, GE, or both, led to a substantial increase of phenotypic variance explained by the models, indicating that differences between lines across two treatments cannot be captured alone by marker effect, treatment, and spatial correction. Thus, to some extent, MET and GE contribute to explained phenotypic variance beyond additive genetic profile or microenvironmental condition.

Estimates of variance components for DRL are presented in Table 4. For each minirhizotron tube, there was one record per family available. Hence, models did not include the effect of treatment or family.

The spatial effect seemed to contribute substantially to variance the amount of variance explained for all five models for DRL to a greater extent than for aboveground traits. In the baseline model (M1-R), for root phenotype, additive genetic variance (σ_g^2) accounted for 0.60 of the total variance, over double of the spatial effect (Table 4). The introduction of two single omics (M2-R and M3-R) contributed to the total variance explained and captured a significant proportion of spatial and genotypic variation. In contrast to aboveground traits, the variance of DRL captured by MET (M3-R) was similar to SNP and GE, contributing equally to reducing the residual variance. Finally, combining all three omics data and

Model ^a	ھ ²	6 ² 8	$\sigma_{\rm m}^2$	6 ² u	σ^2_{um}	$\sigma_{\rm e}^2$	DIC
M1-R	0.27 (0.13–0.42)	0.60 (0.30-0.93)	I	I	I	0.31 (0.07–0.56)	248.2
M2-R	0.21 (0.09–0.35)	0.47 (0.19–0.75)	0.26 (0.11-0.44)	1	I	0.23 (0.04–0.45)	213.1
M3-R	0.21 (0.08-0.34)	0.33 (0.13-0.59)	I	0.37 (0.14–0.63)	I	0.21 (0.05–0.41)	209.7
M4-R	0.18 (0.06-0.31)	0.28 (0.09–0.51)	0.20 (0.07-0.35)	0.28 (0.10-0.51)	1	0.18 (0.04–0.36)	192.5
M5-R	0.16 (0.05–0.29)	0.24 (0.07–0.45)	0.17 (0.05–0.33)	0.24 (0.07–0.44)	0.15 (0.05–0.27)	0.17 (0.04–0.35)	188.0

nethylation and gene expression; σ_{e}^{2} , residual variance 'M1-R-M7-R are model numbers The Plant Genome 2000 0 11 of 17

interaction of MET and GE in M4-R and M5-R incrementally reduced residual variance down to 0.17 and affected variance components values for spatial and main effects of markers. A full model (M5-R) also showed the best fit with the lowest DIC value (188.0) out of all five models.

3.4 | Assessment of predictive ability

Leave-one-family-out cross-validation scheme was used to assess each model's ability to predict phenotypic traits of families (under both treatments) not used to train the model. For each family, models were fitted using data from the rest of the analyzed population, and the scheme was repeated 25 times, resulting in 25 predicted values for each family. Figure 3 shows the average Pearson's correlation between predicted and observed values of aboveground traits for each model analyzed and obtained for each treatment separately. Model M2 yielded the lowest prediction correlations: between 0.13 and 0.19 for DMDig and 0.20 and 0.29 for DW and CEL. On the other hand, the model extended by MET (M3) achieved similarly moderate prediction correlations with the highest value for DMDig under control at 0.39. Including GE (M4-M6) doubled predictive abilities in comparison to M2 and M3 for CEL (0.53-0.63) and substantially increased for DMDig (up to 0.59 for control and 0.25 for drought) and DW under drought (0.43-0.46). The full model (M7) achieved the highest correlation values between predicted and observed aboveground phenotypes. However, for CEL under drought (M4-M7) and DMDig under control (M4-M7), the overestimation of prediction seemed a potential issue for models as presented in Supplemental Figure S4. There was no significant inflation in the predicted values for the rest models and traits (including DRL and DW).

The predictive abilities of DRL are presented in Figure 4. The addition of MET and GE significantly increased predictive ability for the root length phenotype compared with M1-R. Compared with the aboveground phenotypes (Figure 3), MET also contributed significantly to the increase of models' predictive ability. In general, there was little difference between the full model and M3 and M4, with prediction oscillating around 0.45. The test for variance inflation did not indicate a significant under- or overdispersion in DRL predictions (Supplemental Figure S5).

4 | DISCUSSION

In recent years, the availability of different types of omic data has increased because of reduced sequencing costs, the development of public databases, and increased computing power. These data sets are large, multilayered, and highly dimensional. Although the methodology is available for analyzing



FIGURE 3 Leave-one-family-out cross-validations for all models and aboveground traits. Predictive ability is expressed as Pearson's correlation between predicted and observed values for control and drought. Different letters indicate significant ($p \le .01$) differences between models within each treatment and trait. Baseline model (M1) is not included in the figure as family, and spatial effects were not included in the calculation of predictive abilities



FIGURE 4 Leave-one-family-out cross-validations for deep root length. Predictive ability is expressed as Pearson's correlation between predicted and observed values. Different letters indicate significant ($p \leq$.01) differences between models

individual data sets, there is still a lack of consensus on integrating multiple layers of such data to better understand the basis and relative effects of the omic layers on complex traits in crops.

In this study, we investigated the relative contributions of different omic predictors to partition variance components for a better understanding of the different omics layers' effects on four traits related to yield, quality, and DRL in a semifield state-of-the-art root phenotyping facility. The plants were exposed to two different water treatments: drought and control (Svane, Dam, et al., 2019) (Figure 1).

We implemented a series of Bayesian regression models (Table 1) with a Gaussian prior to accommodate heterogeneous, high-dimensional omics inputs, account for nonlinear interactions between and within layers, omic x treatment (drought) interactions, and handling quantitative traits. By analyzing the before-mentioned phenotypes (Table 2) with different omics layers and accounting for interactions and spatial structures, we could quantify the effects of genomic, MET, and transcriptomic data on the traits analyzed (Tables 3 and 4). Additionally, extending the model beyond SNPs showed a decrease in residual variance. Although genetic markers can also model nonadditive genetic effects (epistasis, dominance), the power to explain variance is limited and dependent on the relationship between individuals. Considering the population of 120 F₂ families (genotyped on a family basis), a comparison of marker- and omics-based nonadditive genetic effects was not feasible.

Gene expression accounted for considerable parts of spatial and additive genetic variance for all phenotypes, but after partitioning out spatial, additive genetic, and GE × treatment effects, a large portion of variance was still explained by GE. This probably could be explained by the fact that GE provides a link between the analyzed trait and variation that cannot be explained by DNA-based genetic markers. In contrast, MET mostly accounted for a sizeable proportion of the residual variance for the quality traits but, interestingly, captured larger proportions of spatial and additive genetic variance for the deep root length.

Experimental and agricultural fields are rarely homogeneous, exhibiting variability in landscape attributes, soil properties, water, or nutrient availability. Effects on analyzed phenotypic traits can stem from soil compactness, nutrient availability, or even thermal distribution in the semifield. Spatial variation of the RadiMax facility was shown based on Beds 1 and 2 (Guo et al., 2020). In our data, we identified that phenotypic traits of perennial ryegrass grown in Beds 3 and 4 also were spatially clustered (Figure 2), confirming the need to account for a nonrandom distribution of spatial effects in the semifield facility. The spatial effect was relatively high for all models (Tables 3 and 4). Together with the treatment effect in the model, the introduction of the spatial effect explained micro- and macroenvironment in the semifield in a more comprehensive way. Furthermore, accounting for spatial effects allowed for better variance partitioning, highlighting nonadditive genetic effects and regulatory effects independent of the additive genetic architecture of the omics components in the models. Accounting for the spatial effects can be an effective means to reduce the noise level in nonuniform data, as shown here and by Guo et al. (2020), and can be particularly useful for datasets with limited numbers of trials. Spatial effects can also account for otherwise unexplained noise in randomized experiments with many replicates.

Because the genetic effect on the phenotype is passed from the genome through multiple regulatory layers like MET patterns and GE, a combination of these downstream regulatory layers can compensate for incomplete information in any single regulatory layer and provide a more thorough investigation of genotype-phenotype associations (Ritchie et al., 2015). However, it is important to remember that, unlike DNA sequence information, methylation patterns and transcriptome information are not stably inherited in the population. Their levels are affected by environmental conditions and factors like plant tissue, conditions at sampling, and environmental effects on the plant until sampling. Nevertheless, GE (along with GE_{C} and GE_{D}) consistently explained the highest proportion of variance captured by models compared with MET and SNP for the aboveground traits and showed excellent predictive ability in models M4–M7 (Figure 3). Similar results regarding improved phenotype prediction of complex traits were shown before (Azodi et al., 2020; Fu et al., 2012; Westhues et al., 2017). Gene expression explained a substantial part of the genetic variance for all the investigated phenotypes, and models with GE had higher prediction accuracy than models without (Table 3). The model with MET as the single omic was better for predicting root intensity than the model with GE or a combination of GE and MET (Figure 4). DNA methylomic explained residual variance not accounted for by SNPs or GE (Table 3) for aboveground traits but only for DMDig was this reflected in increased predictive ability.

Changes in epigenetic (MET) patterns likely act slower than fluctuations in GE. Variability in MET profiles between plants can be caused by technical noise, developmental stochasticity, and microenvironmental variance, as well as being coupled with genetic variation (Banta & Richards, 2018; Richards, 2006). However, epigenetic diversity is also expected to be a consequence of transgenerational epigenetic variation, which is primarily a nongenetic phenomenon (Richards, 2006) that can influence qualitative and quantitative traits (Banta & Richards, 2018; Noshay & Springer, 2021; Varona et al., 2015). One aim of our multiomic prediction models was to partition transcriptomic and epigenetic variance from the total genetic variance. Thus, there is a need to identify and account for nongenetic variance in the model evaluation and prediction.

In contrast to GE, MET was not confounded by spatial effects for the aboveground traits (Table 3), whereas the opposite was true for DRL (Table 4). This may be because the aboveground phenotypes were obtained from the third cut, meaning the regrowth after the second cut. Although cutting also affects root growth and biomass (Vinther, 2006), that is, frequent cutting reduces both traits, it is reasonable to assume that the aboveground phenotypes represented a shorter growth window than the root trait.

Additionally, changes in the microenvironment may happen faster aboveground because of within-field differences in shading during the day, exposure to wind, or relative humidity on top of soil heterogeneity, which is more persistent. These rapid changes might affect the aboveground phenotypes more than the roots. Gene expression effects captured a spatial component for both aboveground and root traits. Longer lasting spatial effects are more likely to induce changes in MET patterns. Because the spatial component of MET for DRL was considerable, we speculate that factors contributing to soil heterogeneity, including water and nutrient availability, triggered changes in MET. This also indicates a potentially important role of MET in shaping perennial ryegrass phenotypes during a growing season as a response to the microenvironment. Cortijo et al. (2014) suggested an important role of methylation on primary root length in Arabidopsis. Further studies are needed to identify to what extent heritable, differentially methylated sites or regions affect phenotypes and can be used in plant breeding.

Adding both GE and MET to the models for aboveground traits did not increase predictive ability compared with models with only GE (Figure 3). However, these models usually had a better DIC fit, perhaps because of more precise variance allocation, which reduces bias (Ni et al., 2019). The GE \times MET interaction term was usually minor compared with the variance explained by each of the omics in the model but relatively larger for DRL, indicating a sizeable modifying effect of one omic on the other. González-Reymúndez et al. (2017) suggested that larger sample sizes may be needed to get the full benefits of including omic \times omic interactions in prediction models.

Although the spatial effects explained the by-row heterogeneity, the different ryegrass families were also expected to have a heterogeneous response to the different water treatments. Therefore, including the GE \times treatment interaction term in the model (M7) for the aboveground traits could identify the magnitude of the phenotypic variance depending on the treatment (Table 3). Thus, the GE \times treatment interaction explained more residual variance than the simpler models and increased predictive ability (Figure 3). Likewise, the interaction terms separated variance caused by treatment from the remaining genetic component of the GE matrix, resulting in a more precisely allocated variance.

Our analysis allowed us to consider multiple variables simultaneously (from environmental to multiomics components), which resulted in enhanced predictive abilities of more complex models. Overall, integration of additional predictors can help dissect the complex traits and further strengthen models' predictive abilities. However, in our cross-validation scheme, we observed heterogeneity in predictive abilities under different treatments and a degree of inflation of some of the predicted values for DMDig and CEL. Depending on the treatment for which it was explored, differences between predictive ability values could be due to the sample size of the analyzed population, which did not allow for the exploration of enough phenotypic heterogeneity to achieve a desired degree of generalizability. It could also be speculated that differential treatment effect, which was mainly pronounced on the level of GE and MET patterns compared with phenotypic traits (Table 2), lies at the source of observed discrepancies in predictive abilities under control and drought.

We have partitioned variance components to better understand the effect of three omic layers and showed the importance of accounting for the heterogeneity of experimental semifields. Both MET and GE play important roles in shaping phenotypes, which was illustrated by the variance explained by these predictors (Tables 3 and 4). An alternative hypothesis of complex trait causation states that phenotypic variation is not hierarchical but an interconnection between variations across omic levels. (Ritchie et al., 2015) By combining multiple omics layers, we assume that we capture a larger part of these interconnections and bridge the genotype–phenotype gap. In our study, we partitioned the variance to identify the relative effects of each layer on the phenotypes and how the different regulatory layers were connected in terms of unique and overlapping variance.

The partitioning also quantified additive and more complex components of the variance explained by MET and GE. Recently, important steps have been taken to obtain breeding values using an intermediate omic feature in simulated data sets (Christensen et al., 2021; Weishaar et al., 2020). Importantly, any increase in prediction accuracy of breeding values would depend on the relationship between the omics features and the phenotype and heritability of omics features because only the genetic part of each omics feature is passed on to the next generation. However, validating the methods on real data is not straightforward, as naïve cross-validation may favor methods that predict accurate phenotypes and not accurate breeding values (Christensen et al., 2021). Although much work is needed to develop the methods and validate their suitability on real data, advances like this show that multiomic data can successfully improve future breeding value estimation. This is supported by our findings that there are unique and sizeable genetic components of both MET and GE data, which is a prerequisite for improving the accuracy.

The data set comprised a unique combination of transcriptomic, methylomic, and genetic data, as well as relevant phenotypes, which made it suitable for the development and testing of the models and assessment of variance components. The model framework can be applied to any multiomic dataset to reveal the relative importance of the variance components. Compared with the data set dimensions, the study's main limitation is a relatively small sample size. Further research with a larger sample size would allow extending the scope of analysis to include epistatic and dominance effects as well as to improve the prediction accuracy, eliminate possible bias, and strengthen the conclusion. Nevertheless, this study presents a partitioning of the phenotypic variance into different regulatory layers for the first time. This will pave the way for future studies as the cost of obtaining multiomics data will continue to decrease.

DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its Supplemental Materials.

ACKNOWLEDGMENTS

The authors would like to thank Qiongxian Lu and Maria De la O Leyva Pérez for their help with sampling and the lab technicians at DLF Seeds for obtaining the phenotypic data. This work was supported by Innovation Fund Denmark (grant number 5184-00035B) and Promilleafgiftsfonden for Landbrug.

AUTHOR CONTRIBUTIONS

Marta Malinowska: Conceptualization; Formal analysis; Investigation; Methodology; Writing - original draft. Anja Karine Ruud: Conceptualization; Formal analysis; Investigation; Writing - original draft. Just Jensen: Methodology; Supervision; Writing – review & editing. Simon Fiil Svane: Data curation; Investigation; Writing - review & editing. Abraham George Smith: Data curation; Investigation; Writing - review & editing. Andrea Bellucci: Data curation; Writing - review & editing. Ingo Lenk: Data curation; Writing - review & editing. Istvan Nagy: Data curation; Writing - review & editing. Mattia Fois: Investigation; Writing review & editing. Thomas Didion: Data curation; Writing - review & editing. Kristian Thorup-Kristensen: Funding acquisition; Project administration; Resources; Supervision. Christian Sig Jensen: Funding acquisition; Project administration; Resources. Torben Asp: Funding acquisition; Project administration; Resources; Supervision; Writing - review & editing.

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study nor in the collection, analysis, or interpretation of data and the writing of the manuscript.

ORCID

Marta Malinowska https://orcid.org/0000-0002-2152-4183

Just Jensen D https://orcid.org/0000-0003-3291-8468

REFERENCES

- Alexandersson, E., Jacobson, D., Vivier, M. A., Weckwerth, W., & Andreasson, E. (2014). Field-omics—Understanding large-scale molecular data from field crops. *Frontiers in Plant Science*, 5, 286. https://doi.org/10.3389/fpls.2014.00286
- ANKOM Technology. (1998). *Method for determining neutral cellulase plus gamanase (NCGD)*. ANKOM Technology. https://www.ankom. com/analytical-methods-support/daisy-incubators
- ANKOM Technology. (2022). Method 8—determining acid detergent lignin in beakers. ANKOM Technology. https://www.ankom.com/ analytical-methods-support/fiber-analyzer-delta
- ANKOM Technology. (2017a). ADF Method 12–Acid detergent fiber in feeds—Filter bag technique (for A2000 and A2000I). ANKOM Technology. https://www.ankom.com/sites/default/files/documentfiles/Method_12_ADF_A2000.pdf
- ANKOM Technology. (2017b). NDF Method 6—Neutral detergent fiber in feeds—Filter bag technique (for A2000 and A2000I). ANKOM Technology. https://www.ankom.com/sites/default/files/documentfiles/Method_6_NDF_A200.pdf
- Atkinson, N. J., & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of Experimental Botany*, 63, 3523–3543. https://doi.org/10.1093/jxb/ers100
- Azodi, C. B., Pardo, J., VanBuren, R., de los Campos, G., & Shiu, S.-H. (2020). Transcriptome-based prediction of complex traits in maize. *The Plant Cell*, 32, 139–151. https://doi.org/10.1105/tpc.19.00332
- Banta, J. A., & Richards, C. L. (2018). Quantitative epigenetics and evolution. *Heredity*, 121, 210–224. https://doi.org/10.1038/s41437-018-0114-x
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Nearoptimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34, 525–527. https://doi.org/10.1038/nbt.3519
- Buttler, A., Mariotte, P., Meisser, M., Guillaume, T., Signarbieux, C., Vitra, A., Preux, S., Mercier, G., Quezada, J., Bragazza, L., & Gavazov, K. (2019). Drought-induced decline of productivity in the dominant grassland species Lolium perenne L. depends on soil type and prevailing climatic conditions. *Soil Biology and Biochemistry*, *132*, 47–57. https://doi.org/10.1016/j.soilbio.2019.01.026
- Byrne, S., Czaban, A., Studer, B., Panitz, F., Bendixen, C., & Asp, T. (2013). Genome wide allele frequency fingerprints (GWAFFs) of populations via genotyping by sequencing. *PLoS One*, *8*, e57438. https://doi.org/10.1371/journal.pone.0057438
- Cericola, F., Lenk, I., Fè, D., Byrne, S., Jensen, C. S., Pedersen, M. G., Asp, T., Jensen, J., & Janss, L. (2018). Optimized use of low-depth genotyping-by-sequencing for genomic prediction among multi-parental family pools and single plants in perennial ryegrass (*Lolium perenne L.*). *Frontiers in Plant Science*, 9, 369. https://doi. org/10.3389/fpls.2018.00369

- Cheplick, G. P., Perera, A., & Koulouris, K. (2000). Effect of drought on the growth of *Lolium perenne* genotypes with and without fungal endophytes. *Functional Ecology*, 14, 657–667. https://doi.org/10. 1046/j.1365-2435.2000.00466.x
- Christensen, O. F., Börner, V., Varona, L., & Legarra, A. (2021). Genetic evaluation including intermediate omics features. *Genetics*, 219, iyab130. https://doi.org/10.1093/genetics/iyab130
- Civelek, M., & Lusis, A. J. (2014). Systems genetics approaches to understand complex traits. *Nature Reviews Genetics*, 15, 34–48. https://doi. org/10.1038/nrg3575
- Cortijo, S., Wardenaar, R., Colomé-Tatché, M., Gilly, A., Etcheverry, M., Labadie, K., Caillieux, E., Hospital, F., Aury, J.-M., Wincker, P., Roudier, F., Jansen, R. C., Colot, V., & Johannes, F. (2014). Mapping the epigenetic basis of complex traits. *Science*, 343, 1145–1148. https://doi.org/10.1126/science.1248127
- Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227, 561– 563. https://doi.org/10.1038/227561a0
- Crick, F. H. (1958). On protein synthesis. Symposia of the Society for Experimental Biology, 12, 138–163.
- Cruz, D. F., De Meyer, S., Ampe, J., Sprenger, H., Herman, D., Van Hautegem, T., De Block, J., Inzé, D., Nelissen, H., & Maere, S. (2020). Using single-plant-omics in the field to link maize genes to functions and phenotypes. *Molecular Systems Biology*, *16*, e9667. https://doi. org/10.15252/msb.20209667
- Dan, Z., Hu, J., Zhou, W., Yao, G., Zhu, R., Zhu, Y., & Huang, W. (2016). Metabolic prediction of important agronomic traits in hybrid rice (*Oryza sativa* L.). *Scientific Reports*, 6, 21732. https://doi.org/10. 1038/srep21732
- Dray, S., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., & Wagner, H. (2018). *adespatial: Multivariate multiscale spatial analysis*. R package version 0.2-0.
- Dray, S., & Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, 1– 20. https://doi.org/10.18637/jss.v022.i04
- Druka, A., Potokina, E., Luo, Z., Jiang, N., Chen, X., Kearsey, M., & Waugh, R. (2010). Expression quantitative trait loci analysis in plants. *Plant Biotechnology Journal*, 8, 10–27. https://doi.org/10. 1111/j.1467-7652.2009.00460.x
- Dunn, O. J., & Clark, V. (1971). Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statisti*cal Association, 66, 904–908. https://doi.org/10.2307/2284252
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotypingby-sequencing (GBS) approach for high diversity species. *PLoS One*, *6*, e19379. https://doi.org/10.1371/journal.pone.0019379
- Fè, D., Cericola, F., Byrne, S., Lenk, I., Ashraf, B. H., Pedersen, M. G., Roulund, N., Asp, T., Janss, L., Jensen, C. S., & Jensen, J. (2015). Genomic dissection and prediction of heading date in perennial ryegrass. *BMC Genomics*, *16*, 921. https://doi.org/10.1186/s12864-015-2163-3
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Philosophical Transactions of the Royal Society of Edinburgh*, 52, 399–433.
- Fu, J., Falke, K. C., Thiemann, A., Schrag, T. A., Melchinger, A. E., Scholten, S., & Frisch, M. (2012). Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene expression data. *Theoretical and Applied Genetics*, 124, 825–833. https://doi.org/10. 1007/s00122-011-1747-9

- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907. https://doi.org/10. 48550/arXiv.1207.3907
- González-Reymúndez, A., de los Campos, G., Gutiérrez, L., Lunt, S. Y., & Vazquez, A. I. (2017). Prediction of years of life after diagnosis of breast cancer using omics and omic-by-treatment interactions. *European Journal of Human Genetics*, 25, 538–544. https://doi.org/ 10.1038/ejhg.2017.12
- Grinberg, N. F., Orhobor, O. I., & King, R. D. (2020). An evaluation of machine-learning for predicting phenotype: studies in yeast, rice, and wheat. *Machine Learning*, 109, 251–277. https://doi.org/10.1007/ s10994-019-05848-5
- Guo, X., Svane, S. F., Füchtbauer, W. S., Andersen, J. R., Jensen, J., & Thorup-Kristensen, K. (2020). Genomic prediction of yield and root development in wheat under changing water availability. *Plant Methods*, 16, 90. https://doi.org/10.1186/s13007-020-00634-0
- Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., Peñagaricano, F., Lindquist, E., Pedraza, M. A., Barry, K., de Leon, N., Kaeppler, S. M., & Buell, C. R. (2014). Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell*, 26, 121–135. https://doi.org/10.1105/tpc.113.119982
- Hu, Y., Morota, G., Rosa, G. J. M., & Gianola, D. (2015). Prediction of plant height in *Arabidopsis thaliana* using DNA methylation data. *Genetics*, 201, 779–793. https://doi.org/10.1534/genetics.115.177204
- Huang, T., & Hu, F. B. (2015). Gene-environment interactions and obesity: recent developments and future directions. *BMC Medical Genomics*, 8, S2. https://doi.org/10.1186/1755-8794-8-S1-S2
- Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S., & Ralser, M. (2017). Designing and interpreting 'multi-omic' experiments that may change our understanding of biology. *Current Opinion* in Systems Biology, 6, 37–45. https://doi.org/10.1016/j.coisb.2017.08. 009
- Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Pérez, P., Calus, M., Burgueño, J., & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, 127, 595–607. https://doi.org/10. 1007/s00122-013-2243-1
- Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., & Lorenz, A. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics*, 15, 740. https:// doi.org/10.1186/1471-2164-15-740
- Kemesyte, V., Statkeviciute, G., & Brazauskas, G. (2017). Perennial ryegrass yield performance under abiotic stress. *Crop Science*, 57, 1935–1940. https://doi.org/10.2135/cropsci2016.10.0864
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*, 907–915. https://doi. org/10.1038/s41587-019-0201-4
- Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, 529, 84. https:// doi.org/10.1038/nature16467
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
- Lorenzana, R. E., & Bernardo, R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics*, 120, 151–161. https://doi.org/10. 1007/s00122-009-1166-3

- Malinowska, M., Nagy, I., Wagemaker, C. A. M., Ruud, A. K., Svane, S. F., Thorup-Kristensen, K., Jensen, C. S., Eriksen, B., Krusell, L., Jahoor, A., Jensen, J., Eriksen, L. B., & Asp, T. (2020). The cytosine methylation landscape of spring barley revealed by a new reduced representation bisulfite sequencing pipeline, WellMeth. *The Plant Genome*, 13, e20049. https://doi.org/10.1002/tpg2.20049
- Malosetti, M., Bustos-Korts, D., Boer, M. P., & van Eeuwijk, F. A. (2016). Predicting responses in multiple environments: Issues in relation to genotype × environment interactions. *Crop Science*, 56, 2210–2222. https://doi.org/10.2135/cropsci2015.05.0311
- The Danish Environmental Protection Agency. (2020). *Fremtidens klima*. (In Danish.) The Danish Environmental Protection Agency (Miljøministeriet/Miljøstyrelsen). https://www.klimatilpasning.dk/viden-om/fremtidens-klima/
- Nagy, I., Veeckman, E., Liu, C., Vandepoele, K., Jensen, C. S., Ruttink, T., & Asp, T. (2022). Chromosome-scale assembly and annotation of the perennial ryegrass genome. *BMC Genomics*, 23, 505. https://doi. org/10.1186/s12864-022-08697-0
- Nelissen, H., Sprenger, H., Demuynck, K., De Block, J., Van Hautegem, T., De Vliegher, A., & Inzé, D. (2020). From laboratory to field: yield stability and shade avoidance genes are massively differentially expressed in the field. *Plant Biotechnology Journal*, 18, 1112–1114. https://doi.org/10.1111/pbi.13269
- Ni, G., van der Werf, J., Zhou, X., Hyppönen, E., Wray, N. R., & Lee, S. H. (2019). Genotype–covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. *Nature Communications*, 10, 2239. https://doi.org/10.1038/s41467-019-10128-w
- Noshay, J. M., & Springer, N. M. (2021). Stories that can't be told by SNPs; DNA methylation variation in plant populations. *Current Opinion in Plant Biology*, *61*, 101989. https://doi.org/10.1016/j.pbi.2020. 101989
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & A, L. (2017). Automatic differentiation in pytorch. Paper presented at 31st Conference on Neural Information Processing Systems, 4–9 December 2017, Long Beach, CA. Neural Information Processing Systems. https://openreview.net/ pdf?id=BJJsrmfCZ
- Pérez, P., & de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198, 483–495. https://doi.org/10.1534/genetics.114.164442
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
- Raffo, M. A., Sarup, P., Guo, X., Liu, H., Andersen, J. R., Orabi, J., Jahoor, A., & Jensen, J. (2022). Improvement of genomic prediction in advanced wheat breeding lines by including additive-by-additive epistasis. *Theoretical and Applied Genetics*, 135, 965–978. https:// doi.org/10.1007/s00122-021-04009-4
- Rasmussen, S., Barah, P., Suarez-Rodriguez, M. C., Bressendorff, S., Friis, P., Costantino, P., Bones, A. M., Nielsen, H. B., & Mundy, J. (2013). Transcriptome responses to combinations of stresses in Arabidopsis. *Plant Physiology*, *161*, 1783–1794. https://doi.org/10.1104/ pp.112.210773
- Richards, E. J. (2006). Inherited epigenetic variation–revisiting soft inheritance. *Nature Reviews Genetics*, 7, 395–401. https://doi.org/10. 1038/nrg1834

- Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., & Kim, D. (2015). Methods of integrating data to uncover genotype–phenotype interactions. *Nature Reviews Genetics*, *16*, 85–97. https://doi.org/10. 1038/nrg3868
- Robert, P., Le Gouis, J., & Rincent, R., The BreedWheat Consortium. (2020). Combining crop growth modeling with trait-assisted prediction improved the prediction of genotype by environment interactions. *Frontiers in Plant Science*, 11, 827. https://doi.org/10.3389/fpls.2020. 00827
- Schrag, T. A., Westhues, M., Schipprack, W., Seifert, F., Thiemann, A., Scholten, S., & Melchinger, A. E. (2018). Beyond genomic prediction: Combining different types of *omics* data can improve prediction of hybrid performance in maize. *Genetics*, 208, 1373–1385. https://doi. org/10.1534/genetics.117.300374
- Schrauf, M. F., Martini, J. W. R., Simianer, H., de Los Campos, G., Cantet, R., Freudenthal, J., Korte, A., & Munilla, S. (2020). Phantom epistasis in genomic selection: On the predictive ability of epistatic models. *G3 Genes|Genomes|Genetics*, 10, 3137–3145. https://doi. org/10.1534/g3.120.401300
- Smith, A. G., Han, E., Petersen, J., Olsen, N. A. F., Giese, C., Athmann, M., Dresbøll, D. B., & Thorup-Kristensen, K. (2022). RootPainter: Deep learning segmentation of biological images with corrective annotation. *New Phytologist*, https://doi.org/10.1111/nph.18387
- Smith, A. G., Petersen, J., Selvan, R., & Rasmussen, C. R. (2020). Segmentation of roots in soil with U-Net. *Plant Methods*, 16, 13. https://doi.org/10.1186/s13007-020-0563-0
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639. https://doi.org/10.1111/1467-9868.00353
- Svane, S. F., Dam, E. B., Carstensen, J. M., & Thorup-Kristensen, K. (2019). A multispectral camera system for automated minirhizotron image analysis. *Plant and Soil*, 441, 657–672. https://doi.org/10.1007/ s11104-019-04132-8
- Svane, S. F., Jensen, C. S., & Thorup-Kristensen, K. (2019). Construction of a large-scale semi-field facility to study genotypic differences in deep root growth and resources acquisition. *Plant Methods*, 15, 26. https://doi.org/10.1186/s13007-019-0409-9
- Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lønning, P. E., & Børresen-Dale, A.-L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, 98, 10869–10874. https://doi.org/10.1073/pnas.191367098
- Thoen, M. P. M., Davila Olivas, N. H., Kloth, K. J., Coolen, S., Huang, P.-P., Aarts, M. G. M., Bac-Molenaar, J. A., Bakker, J., Bouwmeester, H. J., Broekgaarden, C., Bucher, J., Busscher-Lange, J., Cheng, X., Fradin, E. F., Jongsma, M. A., Julkowska, M. M., Keurentjes, J. J. B., Ligterink, W., Pieterse, C. M. J., ... Dicke, M. (2017). Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. *The New Phytologist*, 213, 1346–1362. https://doi.org/10. 1111/nph.14220
- van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., Malosetti, M., Iwata, H., Quiroz, R., Kuppe, C., Muller, O., Blazakis, K. N., Yu, K., Tardieu, F., & Chapman, S. C. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science*, 282, 23–39. https://doi.org/10.1016/j.plantsci.2018.06.018

- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91, 4414–4423. https://doi.org/ 10.3168/jds.2007-0980
- Varona, L., Munilla, S., Mouresan, E. F., González-Rodríguez, A., Moreno, C., & Altarriba, J. (2015). A Bayesian model for the analysis of transgenerational epigenetic variation. G3: Genes|Genomes|Genetics, 5, 477–485. https://doi.org/10.1534/g3. 115.016725
- Vazquez, A. I., Veturi, Y., Behring, M., Shrestha, S., Kirst, M., Resende, M. F. R., Jr., & de Los Campos, G. (2016). Increased proportion of variance explained and prediction accuracy of survival of breast cancer patients with use of whole-genome multiomic profiles. *Genetics*, 203, 1425–1438. https://doi.org/10.1534/genetics.115.185181
- Vinther, F. P. (2006). Effects of cutting frequency on plant production, N-uptake and N2 fixation in above- and below-ground plant biomass of perennial ryegrass–white clover swards. *Grass and Forage Science*, 61, 154–163. https://doi.org/10.1111/j.1365-2494.2006.00519.x
- Wang, S., Wei, J., Li, R., Qu, H., Chater, J. M., Ma, R., Li, Y., Xie, W., & Jia, Z. (2019). Identification of optimal prediction models using multiomic data for selecting hybrid rice. *Heredity*, 123, 395–406. https:// doi.org/10.1038/s41437-019-0210-6
- Weishaar, R., Wellmann, R., Camarinha-Silva, A., Rodehutscord, M., & Bennewitz, J. (2020). Selecting the hologenome to breed for an improved feed efficiency in pigs—A novel selection index. *Journal of Animal Breeding and Genetics*, 137, 14–22. https://doi.org/10.1111/ jbg.12447
- Westhues, M., Schrag, T. A., Heuer, C., Thaller, G., Utz, H. F., Schipprack, W., Thiemann, A., Seifert, F., Ehret, A., Schlereth, A., Stitt, M., Nikoloski, Z., Willmitzer, L., Schön, C. C., Scholten, S., & Melchinger, A. E. (2017). Omics-based hybrid prediction in maize. *Theoretical and Applied Genetics*, 130, 1927–1939. https://doi.org/ 10.1007/s00122-017-2934-0
- Zenke-Philippi, C., Thiemann, A., Seifert, F., Schrag, T., Melchinger, A. E., Scholten, S., & Frisch, M. (2016). Prediction of hybrid performance in maize with a ridge regression model employed to DNA markers and mRNA transcription profiles. *BMC Genomics*, 17, 262. https://doi.org/10.1186/s12864-016-2580-y
- Zhou, X., & Lin, H. (2016). Moran's *I*. In Shekhar, S., Xiong, H., Zhou, X. (Eds.), *Encyclopedia of GIS*. https://doi.org/10.1007/978-3-319-23519-6_817-2

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Malinowska, M., Ruud, A. K., Jensen, J., Svane, S. F., Smith, A. G., Bellucci, A., Lenk, I., Nagy, I., Fois, M., Didion, T., Thorup-Kristensen, K., Jensen, C. S., & Asp, T. (2022). Relative importance of genotype, gene expression, and DNA methylation on complex traits in perennial ryegrass. *The Plant Genome*, e20253. https://doi.org/10.1002/tpg2.20253