Contents lists available at ScienceDirect

New Crops

journal homepage: www.keaipublishing.com/en/journals/new-crops/

Opinion Paper

ARTICLE INFO

Keywords: Artificial intelligence Crop breeding Deep learning Soybean Genomic prediction

AI breeder: Genomic predictions for crop breeding

Wanjie Feng^{a,1}, Pengfei Gao^{a,1}, Xutong Wang^{a,b,*}

^a National Key Laboratory of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Road, Hongshan District, Wuhan, Hubei 430070, China
^b Hubei Hongshan Laboratory, No. 1 Shizishan Road, Hongshan District, Wuhan, Hubei 430070, China

ABSTRACT

| The integration of Artificial Intelligence (AI) into crop breeding represents a paradigm shift toward data-driven |
|---|
| agricultural practices, aiming to enhance the efficiency and precision of crop improvement. In this perspective, |
| we critically evaluate the impact of genomic prediction models like SoyDNGP (Soybean Deep Neural Genomic |
| Prediction) on crop breeding. We discuss their current applications, challenges, and future potential. Addressing |
| existing obstacles such as optimizing parent selection, accurately predicting the combined effects of multiple |
| traits and genes, advancing explainable deep learning, and incorporating environmental factors, we propose |
| practical approaches to overcome these challenges. Our insights aim to unlock the full potential of AI in genomic |
| prediction, contributing to a comprehensive understanding of AI's role in agriculture. We advocate for future |
| research efforts that harness AI to cultivate sustainable and equitable food systems. |

Introduction

The rapid paradigm shifts brought about by artificial intelligence (AI) technology have substantially influenced diverse segments of our lives, leading to novel ways of living while changing how we work. AI development has also spurred a new phase in the life sciences. A clear example of this is *AlphaFold*, and its ability to successfully predict the shapes of proteins, marking a breakthrough in protein research and a historic moment in the life sciences (Jumper et al., 2021). Moreover, as the world's population grows and the climate changes, we need to develop quicker and more effective methods to accelerate the crop breeding process (Newman and Furbank, 2021; Xiong et al., 2022).

Using AI technologies has the potential to substantially accelerate the sluggish process of crop breeding. By predicting crop traits before planting with the assistance of AI models, breeders could cut years of work down to one season. AI will revolutionize the breeding process by making accurate predictions from comprehensive genetic data (Xu et al., 2022). The rise of "AI Breeders" marks a significant step, bringing together advanced computing and comprehension of biology. This integrated approach improves the skills of breeders with AI's ability to forecast the possibility of breeding results, enhancing the accuracy and speed of breeding processes.

AI is a key player in the evolution of crop enhancement, enabling researchers to chart a path toward overcoming the critical demand for sustainable agricultural practices. By harnessing the extensive potential of genetic data using predictive models, AI Breeders can push the limits of traditional selective breeding, leading us into an era in which crop improvement is driven by science and data. Building upon our preliminary foundation of predicting phenotypes according to genotypes, our focus has shifted to genomic prediction, sharing perspectives and possibilities, and expressing viewpoints and prospects, with the aim of promoting the application of genomic prediction in breeding, thus accelerating precision-designed breeding.

The evolution of genomic prediction

Genomic prediction, also known as Genomic Selection (GS), is a cornerstone of modern crop improvement strategies. It is the science of forecasting the performance of putative breeding candidates according to their genetic makeup. The evolution of GS has been a critical step in modern plant breeding. Originally identified by Meuwissen et al. (2001) for animal breeding, GS has become a central strategy for determining the phenotypes of polygenic traits in plants and crops using genome-wide markers. This method has been confirmed in diverse crop species, including wheat, maize, and rice, underscoring its usefulness in practical breeding programs (Montesinos-Lopez et al., 2021; Wang et al., 2023; Yan et al., 2023). However, challenges are still present, particularly in developing accurate

Received 6 November 2023; Received in revised form 27 November 2023; Accepted 3 December 2023 Available online 14 December 2023

2949-9526/© 2024 Henan University. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).







^{*} Corresponding author at: National Key Laboratory of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Road, Hongshan District, Wuhan, Hubei 430070, China.

E-mail address: xtwang@mail.hzau.edu.cn (X. Wang).

¹ These authors contributed equally to this work.

https://doi.org/10.1016/j.ncrops.2023.12.005

predictions within high-dimensional marker spaces in which genotypic markers outnumber the population size. Traditional GS models, dependent upon statistical algorithms like BLUP-based (e.g., ridge regression BLUP and genomic relationship BLUP)(BLUP: best linear unbiased prediction) and Bayesian-based algorithms, often have difficulty with the high dimensionality of marker data and the complexity of relationships within and between genotypes and phenotypes (Danilevicz et al., 2022).

Deep learning technologies have ushered in a new era in genomic selection. Methods such as DeepGS (an R package for predicting phenotypes from genotypes using deep learning techniques) and DNNGP (deep neural network for genomic prediction) employ deep neural networks and strategies, including convolution, sampling, dropout, and ensemble learning to manage high-dimensional genotypic data complexity. DeepGS, for example, complements traditional methods such as rrBLUP by providing more accurate phenotypic value selection (Ma et al., 2018). DNNGP distinguishes itself by combining multi-omics data throughout plants, using a deep neural network with a multilayered hierarchical structure for dynamic feature learning. This model manages the breeding data at diverse scales, offering improved prediction accuracy and quicker computation durations than typical methods, making it a valuable tool for GS platforms (Wang et al., 2023).

Our recently developed SoyDNGP model is another significant example of the use of deep learning for genomic prediction in crop breeding (Gao et al., 2023). A key step forward in the SoyDNGP model is its novel methodology for processing genetic variant information. Traditional models usually convert DNA sequences into "one-hot encoding" binary data, resulting in large, sparse matrices lacking information on the full biological significance of nucleotide sequences. In contrast, SoyDNGP transforms genetic variants into an image-like data format, allowing the convolutional neural network (CNN) to process this information similarly to image recognition tasks. This unique representation allows the SoyDNGP to characterize complex patterns and features within genomic data, leveraging multiple convolutional, and pooling layers to determine the intricate relationships and interactions between different genetic regions.

Parental selection with genomic prediction model

The SoyDNGP model structure employed a novel approach for converting genetic variant information, which is more resource-efficient than the traditional one-hot encoding approaches utilized in diverse deep learning applications. The genotype of each sample was transformed into a three-channel image-like format to capture information from both homozygous and heterozygous mutation sites (Fig. 1A). This structure improved the prediction model's functionality, enabling it to identify optimal parent mixtures, a feature that is useful for crops that benefit from cross-breeding approaches (Fig. 1B). Our method mimics the crossbreeding process by digitally merging the genotypes of potential parent plants to simulate both



Fig. 1. Overview of the SoyDNGP model architecture and functional applications.

(A) Schematic representation of the SoyDNGP model framework. Genotype files displayed in VCF format are transformed into a three-channel image-like input for each sample. The core architecture is composed of 12 convolutional layers followed by a single fully connected layer, specialized for regression analysis of quantitative traits and classification of qualitative traits. (B) Use in the selection of parental lines. Virtual F1 genotypes are developed by digitally simulating the combination of variable sites from the parental lines, such as the haplotype C12, representing a virtual hybrid between P1 and P2. Using combinatorial approaches, we can evaluate all possible parental line combinations. The SoyDNGP models are then utilized to forecast the phenotypes for specific traits such as T1, T2, and T3. For each trait, a ranking index is compiled to align with expectations. The rankings "1", "2", and "3" denote the sorted preferences for the given trait. (C) Application in multi-trait aggregate breeding. For the desired integration of traits, initial weights are given to each trait, which contributes to an Expected Index (EI) for the overall composite trait. For example, sample C12 has a score of 1.4, calculated by weighting its trait scores ($1 \times 0.6 + 3 \times 0.2 + 2 \times 0.1$). Comparatively, samples C23 and C13 have scores of 2.8 and 2.3, respectively. Ultimately, the EI ranks composite traits as "1", "3", and "2" for selection purposes, with P1 and P2 emerging as the optimal pairing for the desired composite trait. (D) Prospects for multi-gene aggregate breeding. By randomly disturbing each genetic site once or several times away from the reference (Ref.) to the alternate (Alt.) allele—accounting for three possible states: homozygous Alt., or heterozygous—we can develop 3" potential in silico haplotypes. Using our models, we can forecast outcomes for single or combined traits. This method allows for evaluating the most advantageous gene combinations for specified traits. (E) Application for elucidating trait variat

homozygous and heterozygous genetic sites throughout their offspring. While we primarily demonstrate this in F1 generation examples, the method is equally applicable to F2, F3, and F4 generations, allowing users to explore genetic combinations in these contexts. By inputting these virtual genotypes into our model, we are able to predict a range of possible phenotypes. For instance, using 100 potential parent lines, our model can quickly generate 4,950 sets of predicted phenotypes. We can then rank these predicted traits to determine the most promising parent combinations (Fig. 1B). This approach allows for a more efficient and less costly method for breeders to make informed decisions in the traditionally difficult and uncertain task of crossbreeding.

Genomic prediction using multi-trait aggregate breeding

The quest for improved crop varieties often requires simultaneous enhancement of multiple phenotypic traits, a process called multi-trait aggregate breeding (Akdemir et al., 2019). Multi-trait aggregate breeding depends upon a comprehension that certain traits can be genetically linked and that enhancements in one trait may result in changes, both positive and negative, in others. To properly work within this intricate landscape, breeders must predict how multiple genes interact and modulate diverse phenotypes (Jia and Jannink, 2012). This complex task necessitates the careful balance and selection of several traits to achieve the desired combination in the progeny. Traditional breeding approaches, while effective, are usually slow and complicated when dealing with polygenic traits with low heritability (Crossa et al., 2017).

AI-powered models excel in this area by employing vast datasets to learn complex patterns of genetic interactions. They are able to prioritize traits and examine the collective genetic contributions causing multiple phenotypes, allowing the prediction of breeding outcomes with increased precision (Yan and Wang, 2023). We can prioritize traits according to their importance and assign weights to the prediction outcomes (Fig. 1C). This weighted approach allows us to combine and rank the results for multiple target traits, allowing for unified evaluation. The highest-scored genotypes, whether individual germplasm resources or virtually combined parental genotypes, can be selected as preferred breeding material (Fig. 1C). In this way, the model can analyze the multivariate data to determine optimal combinations of alleles across diverse traits, informing the selection of parental genotypes likely to generate the optimal aggregate breeding results. This approach streamlines the selection process and opens up new possibilities for breeders to explore combinations of traits that might have previously been too challenging to consider. Therefore, AI serves as an accelerator for developing novel varieties to meet the growing demands for higher yield, improved quality, and enhanced resilience in crops, ensuring food security and sustainable agricultural practices.

To optimize breeding outcomes, we aim to characterize parent combinations with an ideal mix of beneficial haplotypes for various desired traits. Leveraging our trait prediction model, we can experiment with virtual genotypes by comprehensively swapping specific genetic loci, particularly those with high importance as indicated by genetic mapping or functional genomics. By analyzing the phenotypic data generated by the model, we can prioritize and select the most promising haplotype combinations (Fig. 1D). This approach hinges on available genotypes or hypothetical parental pairings and surpasses traditional GWAS and association studies, particularly with respect to complex traits with low heritability, like crop yield. By focusing on the cumulative influence of multiple key genes, we have improved our ability to produce meaningful improvements in plant breeding programs.

Genomic prediction using explainable AI models

For genomic prediction in agriculture, the interpretability of AI models, known as explainable AI (XAI), is not merely academic; it has practical implications (Novakovsky et al., 2023). Being able to explain predictions allows for rapid development of genetic markers for breeding, and can lead to more robust and resilient crop varieties. It also develops the confidence of breeders in AI-based decisions, which is critical for their adoption in breeding programs. The state-of-the-art approach for XAI includes various methodologies, such as model-agnostic approaches that approximate the predictions of complex models with more interpretable ones, or perturbation-based methods that assess the impact of input variations on predictions (Zhou and Troyanskaya, 2015; Xu and Jackson, 2019; Ivanovs et al., 2021). For instance, to progress toward explainable deep learning, we can employ sequence perturbation in our refined models to examine the effects of mutations at specific loci. This approach is unique compared to other genomic prediction models that predict chromatin accessibility or transcription factor binding via sequence perturbations; we focus on locus-specific mutations and phenotypic impacts. This enables us to evaluate the influence of each locus on the phenotype, extracting critical variant location information. While this technique has similarities with GWAS, it remains distinct (Fig. 1E). However, these approaches often provide only partial insights and can be computationally expensive or impractical for large-scale genomic data.

Within the SoyDNGP model architecture, we implemented an attention mechanism, directing the neurons to prioritize key features throughout the extraction, assigning increased weight to more pivotal locational attributes. This is intended to allow for the extraction of significant locational data from the genotype, similar to how GWAS quantifies the relevance of each genetic association. By focusing on the attention weights, we can pinpoint the most influential loci. To interpret the influence of genetic variation on specific traits, we constructed a decoder that mirrors the structure of the trained fully convolutional neural network (FCNN) model using transposed convolutions ((Fig. 2A). Each transposed convolution block in the decoder is shaped and weighted to align with its counterpart in the FCNN, guaranteeing that when a tensor filled with ones and shaped like the FCNN output is inputted, the resulting weight distribution matrix is aligned with the feature map dimensions (Fig. 2A). This matrix's weight distribution enables us to approximate the model's allocation of weights to different loci, gauging the impact of variant sites on traits. Our trials demonstrated that the attention mechanism could accurately characterize relevant loci for single-gene traits like soybean FC and POD (Sedivy et al., 2017) (Fig. 2B). However, it was less successful for complex quantitative traits. This could be due to the requirement of highly accurate phenotype models, and the influence of the model's activation functions on weight assignment. Removing these functions harms the prediction accuracy, resulting in unreliable information from an imprecise model.

Determining how to utilize the network to meaningfully interpret locational data remains unresolved and is an ongoing challenge we are determined to address in the future. The push for XAI in genomics is, therefore, not just about uncovering AI but is intricately linked with the desire to advance our fundamental understanding of complex genetic traits. This pursuit will require interdisciplinary collaboration and integration of computational modeling, bioinformatics, and experimental biology. Only through such a concerted effort can we hope to advance XAI in a meaningful way that serves the demands of genomic prediction and assists in securing the future of agriculture.

Genome prediction integrating environmental factors

The influence of environmental factors on crop growth and development is especially significant. In soybeans, the flowering period is notably influenced by environmental conditions. This significance is highlighted by the integral role of multi-environmental experiments in crop breeding, underscoring the necessity of incorporating environmental factors into genome prediction models. Research has indicated that models considering the interplay between environmental and genetic factors can substantially improve predictive ability (Crossa et al., 2016; Cuevas et al., 2017). These environmental factors are varied and assorted into two main categories: readily observable factors, including the location of crop growth, temperature, light, and humidity, and those that are more challenging to measure, including the internal cellular environment of plants (Xu et al., 2022).

W. Feng, P. Gao and X. Wang



Fig. 2. Depiction of the Decoder Structure of SoyDNGP and Preliminary Soybean Applications.

(A) Illustration of the encoder-decoder framework in SoyDNGP. The transposed convolution blocks of the decoder are constructed and weighted to match their corresponding elements in the fully convolutional neural network (FCNN). This guarantees that when a tensor, resembling the FCNN output in shape, but filled with ones, is fed into the system, the constructed weight distribution matrix accurately reflects the dimensions of the feature map. The distribution of weights in this matrix allows for the estimation of the model prioritizing different loci and assessing the influence of genetic variations on trait expression. (B) Analysis of the weight map for characterizing soybean traits such as flower and pod colors. For the SoyDNGP-Baseline version, we performed weight inversion on its associated decoder, normalizing the resulting weight distribution matrix using the z-score and using the 3 σ rule to remove outliers for Manhattan plot visualization. With a normal distribution, values within the range of (μ -3 σ , μ + 3 σ) are likely with a probability of 0.9974. Employing thresholds of 3 σ , 5 σ , and 7 σ , we evaluated the significance of loci weights. Our findings demonstrate that for the flower color (FC) and pod color (POD) traits, the SoyDNGP model has significant attention bias on chromosomes 13 and 18, respectively. After sorting the weight matrix by weight size and choosing variant sites with weights in the top 99.9%, we determined the following loci for the FC trait: 'Chr13_16869677,' 'Chr13_17001366,' 'Chr13_17125100,' and 'Chr13_17304314.' These are within 10 kb of the established major gene *Glyma.130772100*. Similarly, for the POD trait, the model's focus included loci 'Chr19_35024802,' 'Chr19_3504802,' 'Chr19_35104157,' 'Chr19_35154892,' 'Chr19_35176331,' and 'Chr19_3524834,' near the known major gene *Glyma.19G101700*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The difficulty in this approach lies in considering the multifaceted environmental influences and their intersections with genetic factors. However, the successful introduction of deep learning in various domains indicates a promising approach for improving genome prediction models by integrating environmental factors (Khan et al., 2022). Our previous genome prediction models, primarily focused on genetic data, have demonstrated the efficacy of deep learning techniques. Building upon this foundation, we propose enhancements to these models by adding



environmental considerations for a more comprehensive and realistic approach. For instance, the enhanced model could be composed of two submodules: one for genes and another for environmental factors. The gene sub-model would adhere to the preprocessing structure of the SoyDNGP model, while the environment sub-model would be designed to deal with a matrix of N-dimensional environmental features, aligned with the dimensions of the gene model input (Fig. 3). The two sets of features, genetic and environmental, could be integrated using approaches including point

Fig. 3. Construction of genome prediction models combining genotypic and environmental features.

This figure depicts the architecture of genome prediction models integrating both genotypic (G) and environmental (E) factors. To predict crop phenotypes in reaction to environmental influences, the model includes three parts: a gene model, an environment model, and a gene-environment interaction model. The gene model employs a preprocessing structure similar to SoyDNGP, while the environment model uses a matrix with dimensions matching the gene model input, accommodating N-dimensional environmental features (N being the number of environmental feature dimensions). These input features are resized with a deep-learning model to align the feature map of the gene model. Subsequently, the features from both models are combined (either through point multiplication or direct addition) and incorporated into the gene-environment interaction model for training. The final output of the model is a prediction of the crop's phenotype under given environmental conditions.

multiplication or direct addition (Fig. 3). This fused feature map would be integrated into a gene-environment interaction model for training, ultimately resulting in more accurate phenotypic predictions. This approach aims to leverage the strengths of deep learning to obtain the complex interplay between genetics and environments in crop breeding.

Conclusions

AI in crop breeding represents a transformative shift accompanied by significant implications. It promises accelerated breeding cycles, refined selection precision, and adequate handling of complex genomic data. AI models, including SoyDNGP, exemplify these advantages, possibly fast-tracking the crop breeding process. However, as the methodologies we have proposed for the selection of breeding parents, prediction of multi-trait/gene aggregation effects, and integration of environmental factors, the interpretation of deep learning models call for feasibility, they also demand further validation and development by interdisciplinary groups of researchers. As we refine these technologies, AI breeders could evolve from a tool for prediction to an autonomous system able to make informed decisions throughout the breeding process. This evolution would revolutionize how breeding decisions are made and could contribute significantly to global food security and sustainable agriculture.

Author contributions

Xutong Wang, Wanjie Feng, and Pengfei Gao wrote the manuscript and developed the figures. All authors contributed to and approved of the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to express our gratitude to Dr. Tingting Guo and Dr. Hu Zhao from Huazhong Agricultural University for their valuable discussion. This research was supported by the National Key Research and Development Program of China (grant number 2022YFD1201502).

References

Akdemir, D., Beavis, W., Fritsche-Neto, R., Singh, A.K., Isidro-Sánchez, J., 2019. Multiobjective optimized genomic breeding strategies for sustainable food improvement. Heredity 122, 672–683.

- Crossa, J., de los Campos, G., Maccaferri, M., Tuberosa, R., Burgueño, J., Pérez-Rodríguez, P., 2016. Extending the marker× environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. Crop Sci. 56 (5), 2193–2209.
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., et al., 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 22 (11), 961–975.
- Cuevas, J., Crossa, J., Montesinos-López, O.A., Burgueño, J., Pérez-Rodríguez, P., de Los Campos, G., 2017. Bayesian genomic prediction with genotype× environment interaction kernel models. G3 Genes Genom. Genet. 7 (1), 41–53.
- Danilevicz, M.F., Gill, M., Anderson, R., Batley, J., Bennamoun, M., Bayer, P.E., et al., 2022. Plant genotype to phenotype prediction using machine learning. Front. Genet. 13, 822173.
- Gao, P., Zhao, H., Luo, Z., Lin, Y., Feng, W., Li, Y., et al., 2023. SoyDNGP: a web-accessible deep learning framework for genomic prediction in soybean breeding. Brief. Bioinform. 24 (6), bbad349.
- Ivanovs, M., Kadikis, R., Ozols, K., 2021. Perturbation-based methods for explaining deep neural networks: a survey. Pattern Recognit. Lett. 150, 228–234.
- Jia, Y., Jannink, J.L., 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192 (4), 1513–1522.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.
- Khan, M.H.U., Wang, S., Wang, J., Ahmar, S., Saeed, S., Khan, S.U., et al., 2022. Applications of artificial intelligence in climate-resilient smart-crop breeding. Int .J. Mol. Sci. 23 (19), 11156.
- Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., et al., 2018. A deep convolutional neural network approach for predicting phenotypes from genotypes. Planta 248, 1307–1318.
- Meuwissen, T.H., Hayes, B.J., Goddard, M., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157 (4), 1819–1829.
- Montesinos-Lopez, O.A., Montesinos-Lopez, J.C., Salazar, E., Barron, J.A., Montesinos-Lopez, A., Buenrostro-Mariscal, R., Crossa, J., 2021. Application of a Poisson deep neural network model for the prediction of count data in genome-based prediction. Plant Genome 14 (3), e20118.
- Newman, S.J., Furbank, R.T., 2021. Explainable machine learning models of major crop traits from satellite-monitored continent-wide field trial data. Nat. Plants 7, 1354–1363.
- Novakovsky, G., Dexter, N., Libbrecht, M.W., Wasserman, W.W., Mostafavi, S., 2023. Obtaining genetics insights from deep learning via explainable artificial intelligence. Nat. Rev. Genet. 24, 125–137.
- Sedivy, E.J., Wu, F., Hanzawa, Y., 2017. Soybean domestication: the origin, genetic architecture and molecular bases. New Phytol. 214, 539–553.
- Wang, K., Abid, M.A., Rasheed, A., Crossa, J., Hearne, S., Li, H., 2023. DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. Mol. Plant 16 (1), 279–293.
- Xiong, W., Reynolds, M., Xu, Y., 2022. Climate change challenges plant breeding. Curr. Opin. Plant Biol. 70, 102308.
- Xu, C., Jackson, S.A., 2019. Machine learning and complex biological data. Genome Biol. 20 (1), 4.
- Xu, Y., Zhang, X., Li, H., Zheng, H., Zhang, J., Olsen, M.S., et al., 2022. Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. Mol. Plant 15 (11), 1664–1695.
- Yan, H., Guo, H., Li, T., Zhang, H., Xu, W., Xie, J., et al., 2023. High-precision early warning system for rice cadmium accumulation risk assessment. Sci. Total Environ. 859, 160135.
- Yan, J., Wang, X., 2023. Machine learning bridges omics sciences and plant breeding. Trends Plant Sci. 28 (2), 199–210.
- Zhou, J., Troyanskaya, O.G., 2015. Predicting effects of noncoding variants with deep learning-based sequence model. Nat. Methods 12, 931–934.