

Wheat-RegNet: An encyclopedia of common wheat hierarchical regulatory networks

Common wheat (*Triticum aestivum*, $2n = 6 \times = 42$, AABBDD) is the staple crop worldwide. Elucidating the gene regulatory network provides essential information for mechanism studies and targeted manipulation of gene activity for breeding. However, it is a challenging task given the extremely large (16 Gb) and complicated allohexaploid genome of common wheat. Integrating multi-omics data is a compelling approach to construct the hierarchical regulatory network. Here, we collected 189 transcription factor (TF) binding profiles, 90 epigenomic datasets, 2356 transcriptomes, and genome-wide association study (GWAS) for 144 agronomic traits in common wheat, which were further integrated using machine learning approach to infer direct target genes and the hierarchical regulatory network. Wheat-RegNet, a web-based platform, is further developed providing four major functions: (i) to identify regulatory elements regulating input gene(s), and to infer the tissue and environmental response specificities; (ii) to identify the TFs responsible for regulating input gene(s) or locus/loci, as well as the associated GWAS traits; (iii) to construct the hierarchical regulatory network regulating input gene(s); and (iv) to browse hundreds of TF binding, epigenomic, and transcriptomic profiles of input region or gene. Well-organized results and multiple tools for interactive visualization are available through a user-friendly web interface. Wheat-RegNet is a highly useful resource for exploring gene regulatory information and for targeted manipulation, facilitating both hypothesis-driven research and breeding research in common wheat. Wheat-RegNet is freely available at <http://bioinfo.sibs.ac.cn/Wheat-RegNet>.

Construction of multi-omics-based regulatory networks

The publication of multi-omics data provided rich information that can be used to detect regulatory relationships (International Wheat Genome Sequencing Consortium et al., 2018; Ramirez-Gonzalez et al., 2018; Li et al., 2019; Concia et al., 2020; Wang et al., 2021). Several databases for retrieving wheat omics data are available. WheatOmics (Ma et al., 2021) is a typical web-based primary database that collects public datasets and has an interface to visualize the data. Recently developed wheat co-expression network could retrieve expression-related genes of interest (Li et al., 2022). However, different types of data represent information from different perspectives, which need to be integrated for a comprehensive understanding of the hierarchical regulatory networks.

We propose a machine learning strategy to integrate multi-omics data from common wheat in order to determine the regulatory relationship (Figure 1). Specifically, we collected genome-wide binding profiles of 189 TFs generated by DNA affinity purification sequencing (DAP-seq) (Zhang et al., 2022), 90 epigenetic profiles of three types of histone modifications (H3K27me3, H3K9ac, H3K4me3) generated from chromatin immunoprecipitation followed by sequencing (ChIP-seq) datasets that we published previously (Li et al., 2019; Wang et al., 2021), 2356 transcriptome

profiles from public databases, and GWAS intervals for 144 agricultural traits (Blake et al., 2016) (all data sources listed in Supplemental Table 1). All the original data were carefully organized and integrated to construct multi-evidence-based regulatory networks (Figure 1A). The TF binding profiles were used to construct the main body of the hierarchical regulatory network. The *cis*-regulatory elements (CREs) were determined by the combination of TF bindings and epigenomic architecture. The co-expression relationships of genes were deduced from transcriptomic datasets. Considering the ambiguity of assigning target genes to CREs due to the large intergenic regions in the common wheat genome, we designed a machine learning approach to calculate the regulatory potential score (RPS). RPS represents the regulatory potential of distal enhancers to target genes, which is calculated by integrating functional interactions reflected by coordinated epigenetic changes (Wang et al., 2021) and the physical interactions characterized by high-throughput chromosome conformation capture (Hi-C) (Concia et al., 2020). Furthermore, given that a large fraction of trait-associated variations identified in GWAS localized to intergenic regulatory regions (Blake et al., 2016), a comprehensive collection of CREs within trait-associated regions could help rapidly narrow down the search space for experimental validation. The constructed regulatory networks encompass 105 319 genes, 189 TFs and 224 829 CREs for which the tissue and treatment specificity were determined based on the spatial-temporal activity of the epigenetic modifications (Supplemental Figure 1).

Web-based interface for functional tools

Wheat-RegNet provides four modules based on CREs, TF binding, JBrowse visualization, and the hierarchical regulatory network, respectively (Figure 1B). For the best display effect, the Chrome and Firefox web browsers are recommended.

The CRE-based module: For any input genomic locus (loci) or gene(s), Wheat-RegNet returns CREs that regulate the input gene(s). The tissue specificity of these CREs can be visualized by generating a heatmap that reflects the tissue-specific epigenetic activities.

TF binding-based module: For any input gene(s), the TFs present or enriched for regulating the input gene(s) are listed. The regulatory relationship can be visualized via the interactive network. Detailed information for each gene can be visualized by clicking the node in the hierarchical regulatory network.

Hierarchical regulatory module: For any input gene(s), the TF regulation as well as the co-expression regulatory information can be retrieved via the interactive regulatory network.

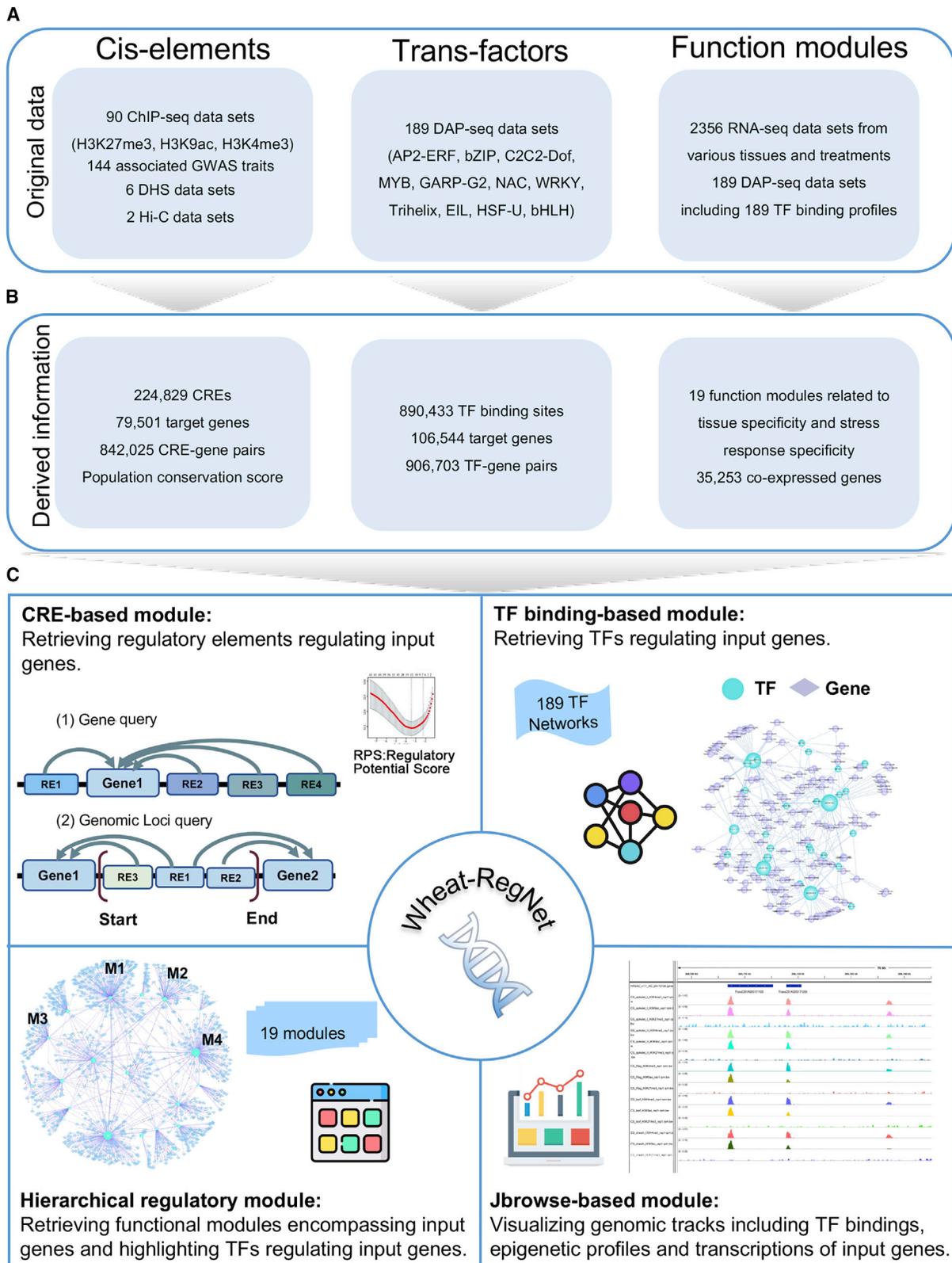


Figure 1. Data collection, integration, and function module curation in Wheat-RegNet.

(A) Summary of the datasets stored in Wheat-RegNet.

(B) The stored data were curated and organized into gene sets and loci that had common properties, including loci or genes targeted by the same transcription factors (TFs), genes that were coexpressed, and cis-regulatory element –gene pairs.

(C) Regulatory potential score-based interface for retrieving *cis*-regulatory elements, TFs, TF binding motifs, and hierarchical regulatory network for an input gene or gene list.

Correspondence

JBrowse-based module: The genomic tracks, which include 90 epigenetic and 30 transcriptomic data that surround the CREs, can be visualized in JBrowse. TF binding motifs that are present or enriched in the input region(s)/gene(s) can be retrieved.

Wheat-RegNet integrates orthologous genes from 34 whole-genome sequenced reference plants (21 from Triticeae) (Supplemental Table 1), and for any gene from these other species, the regulatory information of the homologous gene in common wheat can be retrieved.

Case study 1: Detecting CREs for input gene(s)

Wheat-RegNet can be used to detect CREs that regulate input gene(s) or locus/loci. Here, we describe a search for CREs of Q gene (TraesCS5A02G473800), which confers the free-threshing character and is essential for wheat domestication. Seventeen CREs of the Q gene were retrieved, and detailed information for some of these CREs is listed in Supplemental Figure 2A and 2B, including their coordinates, epigenetic features depicting activating or repressing functions, tissue- and treatment-specificities, and the RPS reflecting the regulatory potential of CREs on input genes. All the retrieved CREs belonged to the active functional type. The epigenetic features of these CREs and the expression of their target genes can be visualized using JBrowse (Supplemental Figure 2C). Most of the motifs in CREs of Q gene tend to be bound by AP2/ERF or C2H2 TF families (Supplemental Figure 2D). Since subgenome-common and -unique CREs harbor important information for elucidating the regulatory evolution in polyploid wheat, the CREs of the orthologs are also presented (Supplemental Figure 3). In addition, for input gene or CREs localized within reported GWAS regions, the relevant information, including the coordinates, potentially regulated agronomic traits, as well as the significance p values, are also listed (Supplemental Figure 3). CREs inside the GWAS interval represent the most plausible functional sites and can be the preferred candidates for experimental validation.

Case study 2: Searching for regulatory TFs for input gene(s)

To illustrate the application of the TF module, we used the set of abscisic acid (ABA)-induced genes published previously as input (Qi et al., 2019). The top enriched TFs that target the input gene set are from various families, including AP2-ERF, bZIP, and HSF, as determined by DAP-seq (Supplemental Figure 4A). These TFs are known to be involved in ABA signaling-related pathways (Song et al., 2016). The TFs in the output table can be interactively selected to construct the hierarchical regulatory network encompassing the selected TFs and input genes (Supplemental Figure 4B). Users can click on each node of the network to retrieve the functions or domains from the gene ontology, InterPro, Pfam, and KEGG databases (Supplemental Figure 4C).

Case study 3: Illustrating the hierarchical regulatory network for input gene(s)

To illustrate the use of the hierarchical regulatory network, we retrieved the target genes of the DOF4 TF using DOF4 as input under the TF module (Supplemental Figure 5). The retrieved targets were used as input to hierarchical regulatory module,

and the upstream regulators and coexpressed genes were obtained. The DOF4 targets include genes that control endogenous ABA levels (e.g., *NCED2/3*), genes that modulate reactive oxygen species status, genes involved in the leaf senescence process (e.g., *CAT2*), and genes that promote flower organ abscission (e.g., *PGAZAT*). These findings are consistent with the reported roles of DOF4 in Arabidopsis (Xu et al., 2020).

In summary, the Wheat-RegNet database is a helpful tool for investigating the complicated regulatory networks in common wheat, which has clear applications in common wheat genetics, to narrow the search windows for *cis*- and *trans*-regulatory factors within the large genome, thereby facilitating hypothesis-driven mechanistic studies and targeted manipulation for wheat breeding.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This study was supported by the National Science Fund for Excellent Young Scholars (32022012), National Natural Science Foundation of China (31921005), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB27010302), and Seed Industry Revitalization Project of Jiangsu Province (JBGS [2021] 006). We thank Dr. Jizeng Jia from Chinese Academy of Agricultural Sciences for insightful comments. We thank Huang Tao for his help in maintaining the high-performance computing server.

AUTHOR CONTRIBUTIONS

Y.J.Z., W.L.Z., and H.L. conceived and designed the project. Z.J.L. performed the experiments. T.F.T. and S.L.T. collected and analyzed the data. T.F.T. and F.Z. designed and maintained Wheat-RegNet. Y.J.Z. wrote the manuscript with input from all authors. T.F.T., X.T.L., H.L., and Y.J.Z. revised the manuscript.

ACKNOWLEDGMENTS

No conflict of interest is declared.

Received: November 8, 2022

Revised: December 22, 2022

Accepted: December 23, 2022

Published: December 26, 2022

Tengfei Tang^{1,2,6,7}, Shilong Tian^{2,3,7},
Haoyu Wang^{1,2,7}, Xiaotong Lv^{2,3,7}, Yilin Xie^{2,3},
Jinyi Liu^{2,3}, Meiyue Wang^{2,3}, Fei Zhao^{2,3},
Wenli Zhang^{4,*}, Hao Li^{1,*} and Yijing Zhang^{5,*}

¹State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, College of Agriculture, Henan University, Kaifeng, Henan 457004, China

²National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China

³University of the Chinese Academy of Sciences, Beijing 100049, China

⁴State Key Laboratory for Crop Genetics and Germplasm Enhancement, CIC-MCP, Nanjing Agricultural University, No.1 Weigang, Nanjing, Jiangsu 210095, China

⁵State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Biochemistry, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, China

⁶China Center of Industrial Culture Collection (CICC), China National Research Institute of Food and Fermentation Industries, Beijing 100015, China

⁷These authors contributed equally to this article.

*Correspondence: Wenli Zhang (wzhang25@njau.edu.cn), Hao Li (lihao@henu.edu.cn), Yijing Zhang (zhangyijing@fudan.edu.cn)
<https://doi.org/10.1016/j.molp.2022.12.018>

REFERENCES

- Blake, V.C., Birkett, C., Matthews, D.E., Hane, D.L., Bradbury, P., and Jannink, J.L. (2016). The triticeae toolbox: combining phenotype and genotype data to advance small-grains breeding. *Plant Genome* **9**:1–10. <https://doi.org/10.3835/plantgenome2014.12.0099>.
- Concia, L., Veluchamy, A., Ramirez-Prado, J.S., Martin-Ramirez, A., Huang, Y., Perez, M., Domenichini, S., Rodriguez Granados, N.Y., Kim, S., Blein, T., et al. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biology* **21**:104. <https://doi.org/10.1186/s13059-020-01998-1>.
- International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C.J., Choulet, F., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**:eaar7191. <https://doi.org/10.1126/science.aar7191>.
- Li, Z., Hu, Y., Ma, X., et al. (2022). WheatCENet: a database for comparative co-expression networks analysis of allohexaploid wheat and its progenitors. *Dev. Reprod. Biol.* <https://doi.org/10.1016/j.gpb.2022.04.007>.
- Li, Z., Wang, M., Lin, K., Xie, Y., Guo, J., Ye, L., Zhuang, Y., Teng, W., Ran, X., Tong, Y., et al. (2019). The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol.* **20**:139. <https://doi.org/10.1186/s13059-019-1746-8>.
- Ma, S., Wang, M., Wu, J., Guo, W., Chen, Y., Li, G., Wang, Y., Shi, W., Xia, G., Fu, D., et al. (2021). WheatOmics: a platform combining multiple omics data to accelerate functional genomics studies in wheat. *Mol. Plant* **14**:1965–1968. <https://doi.org/10.1016/j.molp.2021.10.006>.
- Qi, P.-F., Jiang, Y.-F., Guo, Z.-R., Chen, Q., Ouellet, T., Zong, L.-J., Wei, Z.-Z., Wang, Y., Zhang, Y.-Z., Xu, B.-J., et al. (2019). Transcriptional reference map of hormone responses in wheat spikes. *BMC Genom.* **20**:390–413.
- Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* **361**:eaar6089. <https://doi.org/10.1126/science.aar6089>.
- Song, L., Huang, S.S.C., Wise, A., Castanon, R., Nery, J.R., Chen, H., Watanabe, M., Thomas, J., Bar-Joseph, Z., and Ecker, J.R. (2016). A transcription factor hierarchy defines an environmental stress response network. *Science* **354**:aag1550. <https://doi.org/10.1126/science.aag1550>.
- Wang, M., Li, Z., Zhang, Y., Zhang, Y., Xie, Y., Ye, L., Zhuang, Y., Lin, K., Zhao, F., Guo, J., et al. (2021). An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. *Plant Cell* **33**:865–881. <https://doi.org/10.1093/plcell/koab028>.
- Xu, P., Chen, H., and Cai, W. (2020). Transcription factor CDF4 promotes leaf senescence and floral organ abscission by regulating abscisic acid and reactive oxygen species pathways in *Arabidopsis*. *EMBO Rep.* **21**:e48967.
- Zhang, Y., Li, Z., Liu, J., Zhang, Y., Ye, L., Peng, Y., Wang, H., Diao, H., Ma, Y., Wang, M., et al. (2022). Transposable elements orchestrate subgenome-convergent and -divergent transcription in common wheat. *Nat. Commun.* **13**:6940. <https://doi.org/10.1038/s41467-022-34290-w>.