

# Comprehensive mapping and modelling of the rice regulome landscape unveils the regulatory architecture underlying complex traits

Zhu T.<sup>1</sup>, Xia C.J.<sup>2</sup>, Yu R.R.<sup>1</sup>, Zhou X.K.<sup>1</sup>, Xu X.B.<sup>2</sup>, Wang L.<sup>1</sup>, Zong Z.X.<sup>2</sup>, Yang J.J.<sup>2</sup>, Liu Y.M.<sup>2</sup>, Ming L.C.<sup>2</sup>, You Y.X.<sup>1</sup>, Chen D.J.<sup>1\*</sup>, Xie W.B.<sup>2\*\*</sup>

<sup>1</sup> State Key Laboratory of Pharmaceutical Biotechnology, Department of Gastroenterology, Nanjing Drum Tower Hospital, National Resource Center for Mutant Mice, School of Life Sciences, Nanjing University, Nanjing, China

<sup>2</sup> National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, China

\* [dijunchen@nju.edu.cn](mailto:dijunchen@nju.edu.cn) (Dijun Chen); \*\* [weibo.xie@mail.hzau.edu.cn](mailto:weibo.xie@mail.hzau.edu.cn) (Weibo Xie)

**Key words:** rice; ATAC-seq; RNA-seq; GWAS; deep learning

*Motivation and Aim:* Unraveling the regulatory mechanisms that govern complex traits is pivotal for advancing crop improvement [1]. However, our understanding of the regulatory mechanisms governing complex traits in rice remains incomplete. Given that diverse traits manifest across distinct developmental stages and tissues, systematic annotation of noncoding regulatory variants in rice is currently hindered by the lack of a comprehensive epigenome map across various tissues and growth stages.

*Methods and Algorithms:* UMI-ATAC-seq [2] was used to capture chromatin open regions in rice. cisDyNet [3] was used for data preprocessing and downstream advanced analysis of ATAC-seq. RNA-seq data were quantified using RSEM [4]. Basenji [5] deep learning model was used to train with prediction of tissue-specific chromatin accessibility and used to assess the effect of non-coding region variants.

*Results:* we systematically mapped chromatin accessibility profiles in various tissues across the life cycle of three representative rice cultivars using the UMI-ATAC-seq method [2], a modified ATAC-seq (assay for transposase accessible chromatin-sequencing) protocol developed in our lab. Through analysis of 145 ATAC-seq datasets, we obtained a total of 117,176 unique open chromatin regions (OCRs), accounting for ~15 % of the rice genome. By integration of RNA-seq data from matched tissues, we predicted potential target genes for OCRs based on the correlation of gene expression and adjacent chromatin accessibility across tissues. Through TF footprinting analysis, we inferred tissue- or stage-specific regulatory networks and identified cultivar-polymorphic/trait-associated OCRs by comparing the regulatory landscapes between indica and japonica rice subspecies. Notably, our analysis unveiled a preference for GWAS-associated variants within tissue-specific OCRs, enabling the identification of causal associations between 209 complex agronomic traits and noncoding regulatory variants using this OCR landscape. Utilizing optimized deep learning models, we decoded the regulatory grammar through modeling of tissue-specific chromatin accessibility and across-variety predictions from sequences. The modeling approach sheds light on the key genetic alterations contributing to cis-regulatory divergence. Overall, these data not only serve as a cornerstone resource for the plant research

community but also provide valuable regulatory variants for precision molecular breeding.

*Conclusion:* We used UMI-ATAC-seq to map chromatin accessibility in three rice varieties. Combined with RNA-seq, we effectively established associations between regulatory regions and target genes. Using these resources, we integrated published GWAS data and cloned the gene *Osbzip06*, which affects seed germination rate. We optimized deep learning models that can effectively predict tissue-specific chromatin accessibility and help to identify important genetic loci associated with agronomic traits. Overall, our study establishes a foundational resource for rice functional genomics and precision molecular breeding, providing valuable insights into regulatory mechanisms governing complex traits.

*Funding:* This study was supported by the National Natural Sciences Foundation of China (No. 32070656, 31922065).

### References

1. Fu L.-Y. et al. ChIP-Hub provides an integrative platform for exploring plant regulome. *Nat Commun.* 2022;13(1):3413
2. Zhu T. et al. ATAC-seq with unique molecular identifiers improves quantification and footprinting. *Commun Biol.* 2020;3(1):675
3. Zhu T. et al. cisDynet: An integrated platform for modeling gene-regulatory dynamics and networks. *iMeta.* 2023;2(4):e152
4. Li B., Dewey C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12(1):323
5. Kelley D.R. et al. Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Res.* 2018;28(5):739-750