Systematic mining and functional analysis of factors regulating wheat spike development for breeding selection

Xuelei Lin[1#], Yongxin Xu[1,2#], Dongzhi Wang[1#], Yiman Yang[1,3], Xiaoyu Zhang[1,2], Xiaomin Bie[4], Hongzhe Wang[1], Jiafu Jiang[3], Zhongxu Chen[5], Lixuan Gui[5], Yiliang Ding[6], Yuqing Che[7], Aili Li[7], Xueyong Zhang[7], Fei Lu[1,2,8], Xiangyu Zhao[4], Xiansheng Zhang[4], Xiangdong Fu[1,2], Jun Xiao[1,2,8]*

[1]Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Nanjing Agricultural University, Nanjing, Jiangsu, 210095, China

[4]Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University, Tai'an, Shandong, 271018, China

[5]Department of Life Science, Tcuni Inc., Chengdu, Sichuan, 610000, China

[6]Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

[7]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, 100081, China

[8]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, CAS, Beijing, 100101, China

[#]These authors contributed equally to this article.

*Correspondence: Jun Xiao (jxiao@genetics.ac.cn)

Abstract

Spike architecture largely affects grain number embraced, which is an important trait to optimize during wheat breeding process. In this study, we incorporated multi-omic data and investigated the genetic regulation of wheat spike development. We generated transcriptome and epigenome profiles of shoot apex at eight developmental stages and

observed that chromatin accessibility and H3K27me3 abundance coordinated with transcriptome alteration for flowering transition. A core transcriptional regulatory network (TRN) that drove the formation of the wheat spike was constructed. Integration of the TRN and genome-wide association analysis (GWAS) led to the identification of 227 transcription factors (TFs) with both known (42) and unknown (185) functions. Multi-layer regulatory module among TaSPL6, TaMADS34 and TaMADS15 identified by TRN was validated experimentally. Of the 85 novel TFs with homozygous mutant lines in TILLING library, 38 showed altered spike architecture or flowering time, including bHLH, MYB, and WRKY. Furthermore, we demonstrated that TaMYB4-A regulates fertile spikelet number likely via regulating hormones homeostasis and signaling, and is downstream of, and repressed by WFZP. The elite allele of *TaMYB4-A*, with a C-SNP at the WFZP binding site, was selected for balancing spikelet number and fertility during breeding process in China. Collectively, our findings provide high-confidence novel regulators and an efficient strategy for studying the genetic basis of wheat spike development for practical breeding applications.

Key words: wheat spike architecture, TRN, GWAS, MYB, breeding selection

Introduction

Wheat (*Triticum aestivum*) is a major crop worldwide, providing approximately 20% of daily caloric and protein consumption for the human population. The primary objective of crop domestication and breeding has been to increase productivity, with inflorescence architecture playing a major role in determining grain yields. Various crops display different inflorescence morphologies; for instance, rice forms a panicle inflorescence that produces primary and secondary branches with one fertile floret spikelet. In contrast, wheat forms an inflorescence with multiple sessile spikelets along an unbranched and determinate compound spike (Gao et al., 2019; Gauley and Boden, 2019; Koppolu and Schnurbusch, 2022). The development of spike in wheat is a lengthy process that involves several stages. Initially, the shoot apex meristem (SAM)

constitutively produces the primordium of leaves, nodes and internodes before flowering. Following this, the vegetative SAM changes into the transition apex in response to environmental cues. From here, the apex gradually elongates, producing double-ridged structures. The lower ridges develop into leaf primordia, while the upper swelling ridges give rise to the spikelet primordia. Enlargement of these primordia results in glume, lemma, and floret primordia appearing successively. With the onset of apical spikelet differentiation, the maximum number of spikelets is formed, and stamen primordia begin to emerge, initiating the process of floral organ differentiation. Subsequently, the pistil and floret awns appear and elongate. Mature spikelets typically produce 4-6 fertile florets (Bonnett, 1936; Waddington et al., 1983).

The grain number per spike (GNPS) is an important factor contributing to grain yield (Xiao et al., 2022). GNPS is determined by the inflorescence architecture, which includes the arrangement and number of spikelets and fertile florets per spikelet. In recent studies, modifying spike architecture has been shown to significantly increase grain yield in wheat, emphasizing the importance of this trait (Wang et al., 2022; Zhang et al., 2022). Significant advances have been made in understanding the molecular regulation of cereal inflorescence architecture (Gao et al., 2019; Yuan et al., 2020; Luo et al., 2023). In wheat, Photoperiod-1 (Ppd-1), FLOWERING LOCUS T1 (FT1), FT2 coordinately regulate the transition of SAM to inflorescence meristem (IM) and the duration of spike development (Boden et al., 2015; Finnegan et al., 2018). VERNALIZATION1 (VRN1) and other paralogs such as FRUITFULL2 (FUL2) and FUL3 are required synergistically for initiation and maintenance of the spikelet meristem (SM) identity (Li et al., 2019a). Wheat TEOSINTE BRANCHED1 (TaTB1) interacts with FT1 to control the production of paired spikelets via regulating *VRN1* expression (Dixon et al., 2018). WHEAT ORTHOLOG OF APO1 (WAPO1) affects spike number per spikelet (SNS) by regulating the timing of terminal spikelet formation (Kuzay et al., 2019; Kuzay et al., 2022). *Q* (*AP2L5*), which encodes an APETALA 2 (AP2) TF, can significantly reduce the SNS when loss-of-function, possibly due to the premature transition of the inflorescence meristem to terminal spikelet (Debernardi et al., 2020). *WFZP*, a homolog of rice *FRIZZY PANICLE* (*FZP*), determines the SNS and

SM fate partially by *VRN1* and *HOMEOBOX4* (*TaHOX4*) (Li et al., 2021), and by inhibiting the repressor of spikelet formation gene *BARREN STALK1* (*TaBA1*) through epigenetic modification (Du et al., 2021). Furthermore, Squamosa Promoter-Binding Protein-Like (SPL) and MADS-box transcription factors (TFs) such as TaSPL13-B, TaSPL14, and TaAGL6-A are involved in SNS determination (Su et al., 2019; Li et al., 2020; Cao et al., 2021; Kong et al., 2022). Moreover, the balance between auxin and cytokinin (CK) hemostasis plays a vital role in regulating spike development in wheat, including factors such as TaCYP78A5 and TaCKX2.1, TaCKX2.2 (Qi et al., 2019; Jablonski et al., 2020; Jablonski et al., 2021; Guo et al., 2022). However, while significant progress has been made in understanding the regulation of inflorescence architecture, functional genomics research in wheat is not as advanced due to limited methods for gene identifications (Gupta et al., 2019, Xiao et al., 2022).

Population genetics has experienced a significant boost, thanks to the recent advancements in sequencing technology (Sehgal and Dreisigacker, 2022). Notably, large-scale GWAS analysis has been employed to study landmark cultivars using single nucleotide polymorphism (SNP) data obtained from various sequencing techniques such as genotyping-by-sequencing (GBS), whole genome resequencing, or exome capture sequencing (Hao et al., 2020; Pang et al., 2020; Li et al., 2022). This has led to the association of several genetic loci with spike developmental traits such as spike length (SL), SNS, GNPS, grain setting and spike compactness (Sun et al., 2017; Pang et al., 2020). However, GWAS analysis is often time-consuming and prone to inaccuracies when dealing with large linked intervals. As such, gene co-expression networks provide different filter for identification of key regulators, as demonstrated in previous studies based on stage-specific transcriptome and population-wide RNA-seq (Wang et al., 2017; Li et al., 2018; VanGessel et al., 2022). Aside from gene expression *per se*, chromatin features have also been explored in wheat for different tissues, revealing their association with the regulation of tissue-specific or stress-induced gene expression and subgenome-biased expression patterns of homeologs (Li et al., 2019b; Concia et al., 2020; Yuan et al., 2022; Zhao et al., 2023). Integrating gene expression, chromatin features, and transcription factor binding sites has enabled the prediction of

gene regulatory networks, as exemplified by recent studies in wheat (Chen et al., 2023; Tang et al., 2023). Combining the information obtained from meta agronomic traits GWAS analysis with gene regulatory networks could potentially enhance the identification of candidate genes linked to associated regions (Chen et al., 2023). Nonetheless, previous studies have not exploited the spike-specific gene regulatory network due to the lack of chromatin landscape data from tissues during wheat spike development, thus difficult to uncover novel regulators for spike architecture.

In this study, we conducted a time-serial profiling of the transcriptome and epigenome at the shoot apex in the elite wheat cultivar Kenong 9204 (KN9204) to gain a better understanding of how the chromatin landscape is coordinated with transcriptome dynamics during the process of spike formation. More importantly, we integrated co-expression and regulatory networks of spike developmental specific genes with GWAS to identify the key factors that shape wheat spike architecture. We also validated and performed in-depth analysis in order to uncover the potential regulators' function and reveal the allelic selection that occurs during the breeding process.

Results

A transcriptome and chromatin landscape atlas for wheat spike development

To understand the transcriptional regulation for wheat spike development, shoot apex at eight distinct developmental stages of winter wheat cultivar KN9204 were sampled, including SAM, transition apex (W1.5), early double ridge (W2), double ridge (W2.5), glume primordium present (W3), lemma primordium present (W3.25), floret primordium present (W3.5), and stamen primordium present (W4) (Fig. 1a, description as Waddington et al., 1983). RNA-seq and the Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) were used to profile global gene expression, chromatin accessibility with two or three bio-replicates (Fig. 1a, Extended Data Fig. 1a) (Buenrostro et al., 2015; Kaya-Okur et al., 2019; Zhao et al., 2023).

In total, 58,875 high-confidence genes were found to express at least in one stage (TPM > 0.5) (Table S1). The principal component analysis (PCA) and hierarchical clustering of RNA-seq showed that the eight stages could be subdivided into two major groups, as vegetative to flowering transition group including SAM, W1.5, W2, W2.5 and reproductive growth phase group including W3, W3.25, W3.5 and W4 stages (Fig. 1b, Extended Data Fig. 1b). More differentially expressed genes (DEGs) between adjacent stages occur at morphological transition points, such as W2.5 to W3, W3 to W3.25, W3.5 to W4, when glume primordium, lemma primordium and stamen primordium initiate respectively (Fig. 1c, Table S2). Stage-specific expressed genes are identified via clustering (Fig. 1d). Known factors involved in flowering time regulation and inflorescence development are highlighted from specific clusters. For instance, flowering time gene *TaPpd-1*, with a Photoperiod-insensitivity allele *TaPpd-D1a* in KN9204, is expressed before flowering transition from SAM to W2.5 stage (Boden et al., 2015; Perez-Gianmarco et al., 2020); *WAPO1*, required for the maintenance of SM activity, is highly expressed at W3 stage (Kuzay et al., 2022); *TaTFL1*, positively correlated with SNS, is expressed at W3.25 and W3.5 stages (Wang et al., 2017; Sun et al., 2023); floral organ regulator *TaAGL6* is highly expressed at W4 stage (Kong et al., 2022) (Fig. 1d). Such transcriptome pattern correlates with the morphological characteristic of individual stage, indicating our sampling could capture the dynamic gene expression profile during wheat spike development.

As previous report (Zhao et al., 2023), accessible chromatin is mainly located in the distal intergenic regions and promoters (Fig. 1e). ATAC-seq peak intensity at genic region, in particular around transcriptional start site (TSS), is positively correlated with gene expression (Extended Data Fig. 1c). Unsupervised PCA analysis revealed a roughly continuous trajectory of chromatin accessibility dynamics during wheat spike development (Fig. 1f). Eight-stages could partition into five sub-clusters, as vegetative cluster including SAM, W1.5, W2, flowering transition stage (W2.5), inflorescence initiation (W3), spikelet differentiation (W3.25 and W3.5), floral organ formation (W4) (Extended Data Fig. 1d). A globally increase of chromatin accessibility at genic regions

was observed from vegetative cluster to flowering transition and inflorescence initiation, whereas declined afterwards (Fig. 1g, Extended Data Fig. 1e).

Based on the chromatin accessibility dynamics, SAM, W2.5, W3, W3.25 and W4 stages were chosen for histone modification profiling via Cleavage Under Targets and Tagmentation (CUT&Tag) (Fig. 1a) (Zhao et al., 2023). In general, H3K27ac, H3K4me3 and H3K36me3 intensity around TSS are positively correlated with gene expression, while H3K27me3 is enriched at gene body of no/low expressed genes (Fig. 1e, Extended Data Fig. 1c). H3K27me3 and H3K36me3 were mutually exclusive for each other, H3K4me3 and H3K27ac were positively correlated, whereas, histone variants H2A.Z was overlapped with both active H3K27ac and H3K4me3 and repressive H3K27me3, but depleted with H3K36me3 (Extended Data Fig. 1f, Table S3). The PCA and cluster analyses showed that various histone modifications had stage-specific transitions during different developmental stages (Fig. 1h, i, Extended Data Fig. 1g, h, i). Of note, H3K27me3 presents a continuous trajectory while others are not, suggesting a higher correlation with transcriptional profile dynamics (Fig. 1h). Indeed, H3K27me3 abundance and chromatin accessibility jointly affect different gene expression patterns during spike development (Fig. 1j). Taken together, we have generated transcriptomic and epigenomic profile of shoot apex at important stages during wheat spike formation.

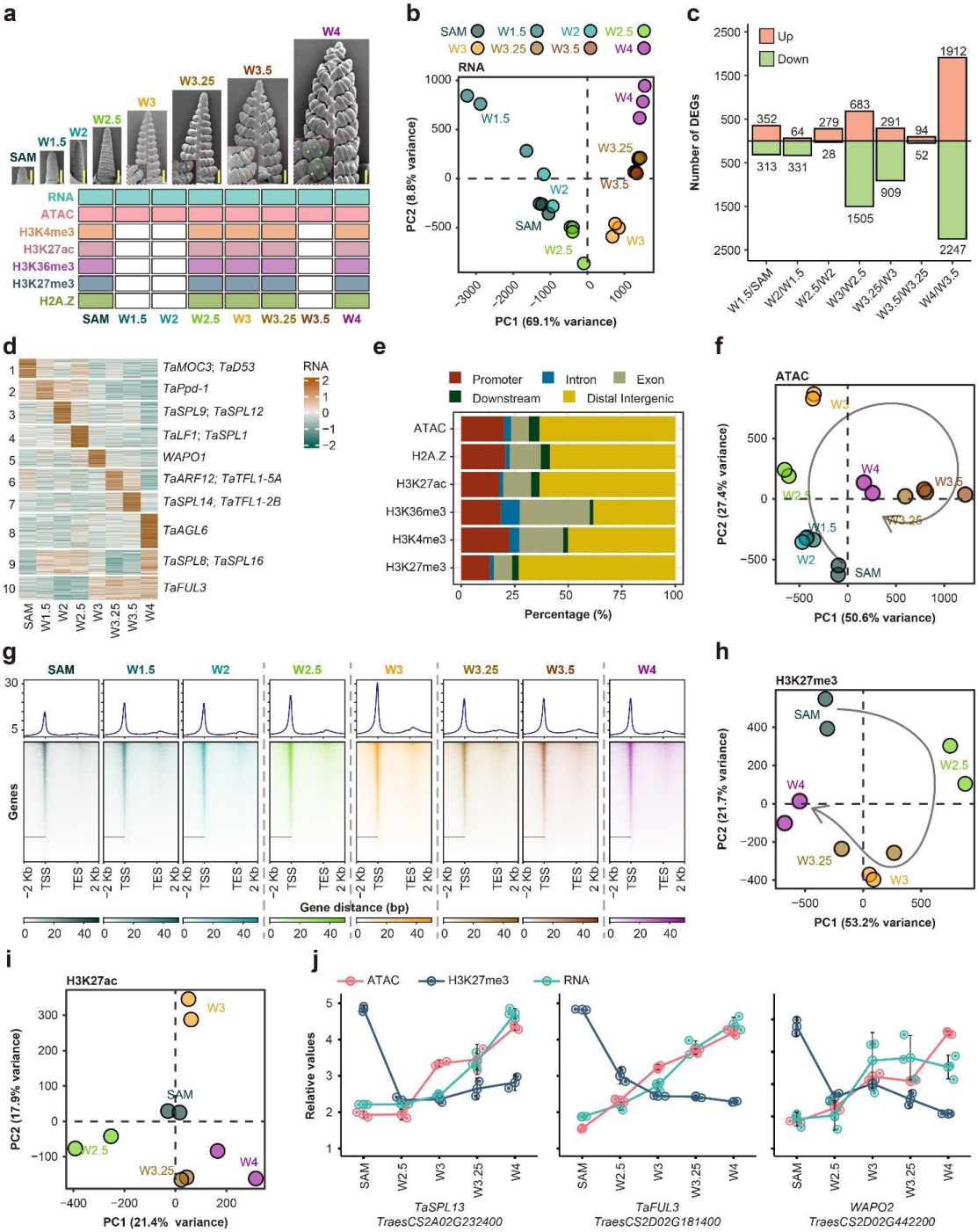


Fig. 1: Charting transcriptome and chromatin landscapes during wheat spike development.

a, Scanning electronic micrographs (SEM) of young spikes in eight developmental stages and experimental design. SAM, shoot apical meristem stage; W1.5, transition apex stage; W2, early double ridge stage; W2.5, double ridge stage; W3, glume

primordium present stage; W3.25, lemma primordium present stage; W3.5, floret primordium present stage; W4, stamen primordium present stage. Scale bars = 200 μm.

b, PCA of transcriptome showing distinct developmental stages. Each dot represents one sample. Color represents the different stages during spike development. Three bio-replicates were sequenced for each stage.

c, Number of DEGs between adjacent developmental stages. DEGs were defined by the threshold |log2 (Fold Change) | ≥ 1 and FDR ≤ 0.05 by DESeq2.

d, Heatmap of expressed genes sorted by k-mean clustering across the samples collected at different developmental stages. The representative genes of each cluster were listed on the right.

e, Peak distribution of ATAC-seq and various histone modifications relative to genes.

f, PCA of ATAC-seq samples during spike development. Each dot represents one sample. Color represents different stages. Two bio-replicates were sequenced for each stage.

g, Chromatin accessibility dynamics in proximal regions (promoter and genic regions) during wheat spike development.

h–i, PCA of H3K27me3 (**h**) and H3K27ac (**i**) samples during spike development. Each dot represents one sample. Two bio-replicates were sequenced for each stage.

j, Gene expression, chromatin accessibility and H3K27me3 changes at representative genes during spike development. The Y-axis indicates relative values of Z-scaled gene expression, chromatin accessibility and H3K27me3 levels.

Permissive chromatin status facilitates vegetative-to-reproductive transition

As indicated, chromatin accessibility increased globally during vegetative-to-reproductive transition at shoot apex (Fig. 1g). We further look at differentially accessible promoter regions (DAPR) from SAM to W1.5, W2, W2.5 till W3 stages (Fig. 2a, Table S4). In total, 49,153 DAPR was identified, categorized into 6 clusters. Of them, majority of DAPR showed increased accessibility at either W2.5 or W3 stage, including clusters 2, 3, 4, 6 (Fig. 2a, Table S5). Interestingly, genes of those four

clusters show higher positive correlations between chromatin accessibility and transcriptional dynamics ($R > 0.5$) (Fig. 2b). Indeed, genes with increased chromatin accessibility at W2.5 or W3 stage significantly overlapped with genes of elevated expression at W2.5 or W3 stage as compared to SAM (1,920 genes within gene set I; Fig. 2c, Extended Data Fig. 2a, Table S6). This highlights the synchronous between gain-of chromatin accessibility and elevated gene expression (Fig. 2d). The degree of increased chromatin accessibility is highly correlated ($R = 0.85$) with fold change of elevated expression level (Fig. 2e). Gene Ontology (GO) enrichment analysis suggests that hormone biosynthesis and signaling, inflorescence meristem identity, asymmetric cell division are enriched in gene set I (Fig. 2f).

Besides, open chromatin could set a 'primed' status for gene activation later on (Bonifer and Cockerill, 2017; He and Li, 2018). There are 3,435 genes with gained chromatin accessibility at W2.5 or W3 stage but elevated mRNA level at stages after W3 (Gene set II; Fig. 2c, Extended Data Fig. 2b, Table S6). Those genes are mainly involved in hormone metabolism, floral organ identity, meristem development (Extended Data Fig. 2c), which are important for spike formation similar as gene set I. By contrast, large number of genes with gain-of chromatin accessibility but no change of mRNA level at tested spike developmental stages (Gene set III; Fig. 2c, Table S6), such as genes participated in polarity specification, cell size regulation, regulation of translation or protein transport (Extended Data Fig. 2d), indicating that open chromatin is not sufficient to active gene.

In addition to chromatin accessibility, H3K27me3 was closely related to transcriptional pattern of gene set I, II, III, while other histone modifications such as H3K27ac, H3K4me3, H2A.Z were not specifically correlated (Fig. 2g, Extended Data Fig. 2e). In details, genes in set I have relatively higher chromatin accessibility at SAM stage, and sharply increased openness at W2.5 and W3 stages, while reduced H3K27me3 level at W2.5 as compared to SAM stage (Fig. 2g). As an example, well-known flowering promoting gene *VRN1* shows increased chromatin accessibility and decreased H3K27me3 in genic region, while no change for H3K4me3 from SAM to W2.5 and W3 stages (Fig. 2h). Whereas for gene set II, though chromatin accessibility is elevated

at W3 and W3.25 stages, but the relative openness is low. Meanwhile, H3K27me3 level doesn't change from SAM to W2.5, even increased at W3 stage. The combined effect may result in the 'primed' transcriptional status, which gene expression level increased later with the reduction of H3K27me3, such as wheat *APETALA3* (*WAP3*) (Fig. 2g, h). For gene set III, lowest chromatin accessibility gained and highest H3K27me3 coverage is observed, which can restrict gene activation (Fig. 2g), as the case for *TaEhd2* gene (Fig. 2h).

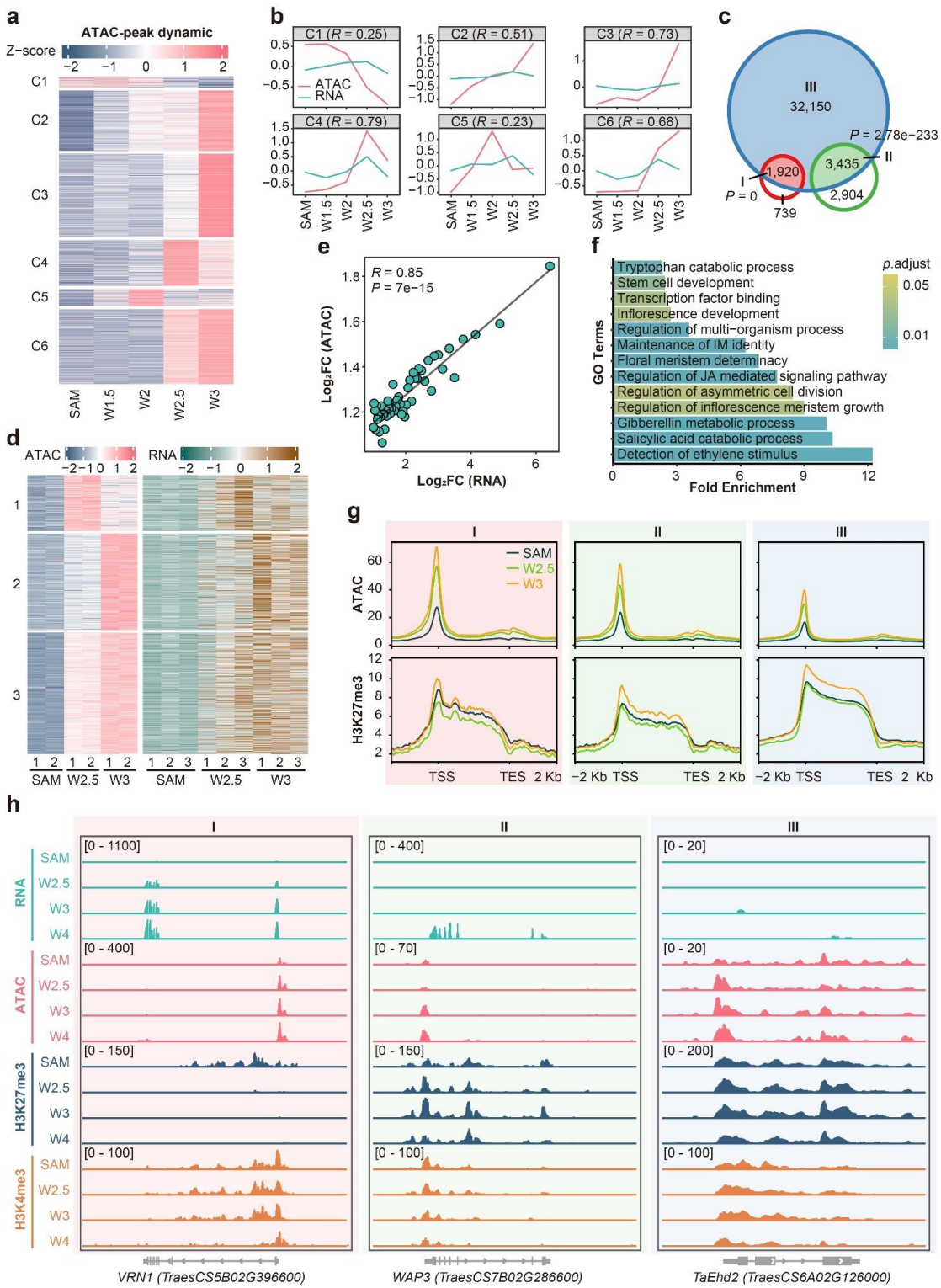


Fig. 2: Chromatin accessibility and H3K27me3 dynamics facilitate vegetative-to-reproductive transition.

a, K-means clustering of DAPR across the samples collected at stages from SAM to W3.

b, The chromatin accessibility and gene expression tendency for each cluster in (**a**). *R* value stands for the pearson correlation coefficients.

c, Overlap between genes with gained-chromatin accessibility of cluster 2, 3, 4, 6 in (**a**) (blue circle), elevated expression at W2.5 and W3 stage versus SAM stage (red circle) and up-regulated at W3.25, W3.5, W4 stage versus former stages (green circle). Gene set I showing overlap between red and blue circle genes; gene set II showing overlap between red and green circle; while rest genes from red circle belong to gene set III. Hypergeometric test was used to calculate *P* value for the enrichment of gene set I and II.

d, Synchronous pattern of gene set I between the gain of chromatin accessibility and elevation of gene expression from SAM to W2.5 and W3 stage. Individual bio-replicate are shown separately.

e, Pearson correlation of gene expression elevation (X-axis) and chromatin accessibility gain (Y-axis) from SAM to W2.5 or W3 stage based on 1,920 genes in gene set I in (**c**). Genes were ranked by fold change of elevated expression and separated into 50 bins. *P* values are determined by the two-sided Pearson correlation coefficient analysis

f, GO enrichment analysis of 1,920 genes in gene set I.

g, Chromatin accessibility (top) and H3K27me3 (bottom) levels of genes in gene set I, II, III at SAM, W2.5 and W3 stages.

h, Genomic tracks showing expression, chromatin accessibility and histone modifications change at representative genes of *VRN1* (gene set I, left), *WAP3* (gene set II, middle) and *TaEhd2* (gene set III, right).

Transcriptional regulatory network for spike architecture formation

After flowering transition, inflorescence undergoes several developmental stages, including W3, W3.25, W3.5, and W4 to form the spike architecture (Fig. 1a, 3a), which is critical for determining SNS, contributing to grain yield. We calculated the pseudo time of stages covering W3 to W4 based on PCA distance (Fig. 3a). W3.25 and W3.5 stages are close to each other, while distinct from W3 and W4 stages. A total of 8,200

DEGs were identified and clustered into 6 categories (Fig. 3b, Table S7). Floral meristem determinacy and various hormone metabolic genes are highly expressed at W3 stage; hormone signaling, polarity establishment genes are active at W3.25 and W3.5 stages; while floral organ identity, polarity specification genes are elevated at W4 stage (Fig. 3b, c). Among those stage-specific DEGs, we looked for enriched TFs. Within clusters 3 and 6, ERF and WRKY TFs are enriched and highly expressed at W3 stage. For W3.25 and W3.5 stages, NF-Y and SBP TFs are abundantly expressed within clusters 2 and 4. Whereas, MADS-box TFs are outstanding at W4 stage of cluster 5 (Fig. 3b, c).

Accessible chromatin regions provide docking sites for TFs. Based on the TF-motif recognition pairs, in combination with the co-expression pattern of TFs and downstream targets, we construct a hierarchy TRN by integrating the multi-stages ATAC-seq and RNA-seq data (Extended Data Fig. 3a). Totally, 36,908 pairs of 5,106 TF-target gene interactions were identified, of which 4,916 pairs were regulated between TFs (Extended Data Fig. 3b). To better visualize and explore the TRN, a website was constructed with access to the publics (http://39.98.48.156:8800/#/, see methods for details). A clear sequential regulatory relation among time-serial expressed gene clusters from W3 to W3.5 stages is observed as C6-C3-C2-C4 (Fig. 3d). Whereas, W4 stage specific expressed genes in C1 and C5 are separated and regulated each other (Fig. 3d), indicating distinct TRN between spikelet and floret development. TFs of ERF, B3, TCP, DOF and MIKC-MADS families may play core regulatory roles as more targets being identified (Extended Data Fig. 3c). Some functionally characterized factors that involved in wheat spike development can be found in the network, such as *TaTB1* (Dixon et al., 2018), *VRN1* (Li et al., 2019a), *TaFUL3* (Li et al., 2019a) (Fig. 3e).

Further, we extract a module that contains factors being functionally studied individually in other crops but without known of regulatory relation to evaluate the prediction power of TRN, including *SPL6* (Wang et al., 2018), *MADS34* (Lin et al., 2014), *MADS15* (Wu et al., 2017) (Fig. 3e). The temporal expression pattern and the presence of specific motifs in the open chromatin regions (Extended Data Fig. 3d, e)

suggest a positive regulatory hierarchy following SPL6-MADS34-MADS15-*HMA* module. Indeed, we observed a positive transcriptional regulatory circuit among them by using luciferase reporter assay in tobacco leaves (Fig. 3f, Extended Data Fig. 3f). As well, such regulation is dependent on the TFs-cis motifs specific recognition (Fig. 3f, Extended Data Fig. 3f). Thus, the TRN constructed could well predict the potential important factors and the regulatory relationships.

Integration with population genetics to screen regulators shaping wheat spike

The open region of chromatin is crucial for the establishment of transcriptional regulatory circuit (Klemm et al., 2019). Indeed, there are significantly more TFs recognition motifs presence in the open chromatin region (Extended Data Fig. 3g). The TF binding footprints within open chromatin regions showed lower DNA variation frequency than flanking regions and random genomic background (Fig. 3g), indicating the functional importance of such sites. Alteration of DNA sequence in the spike tissue-specific open chromatin regions might influence the inflorescence architecture. Consistent with this hypothesis, the open chromatin regions contain significantly higher GWAS signals related to spike morphological traits (Hao et al., 2020; Pang et al., 2020; Abou-Elwafa and Shehzad, 2021) (Fig. 3h). Thus, we further integrate the spike traits-associated DNA variations from population genetics study (Table S8) with key regulators identified for floral transition, giving the importance of flowering transition duration effects on inflorescence development (Boden et al., 2015; Finnegan et al., 2018), as well as regulators in the TRN that potentially governing the formation of spike (Fig. 3i).

Started with 1,920 genes from gene set I for flowering transition (Fig. 2c) and 5,106 TFs-targets within TRN (Extended Data Fig. 3b), we focused on TFs coding genes and got 776 candidate TFs (Fig. 3i, Table S9). Among them, more than 65% (518 TFs) have DNA variations in the promoter open chromatin region or with non-synonymous mutation in the coding region (Fig. 3i, Table S9), which is likely to affect the transcriptional regulatory circuit and rewire the TRN of spike development. Finally,

227 TFs harbor DNA variations that are significantly associated with spike related traits, within 2 Mb regions centered of GWAS signals (Fig. 3i, Table S9). Of these TFs, 8 factors are functional ‘researched’ in wheat (Fig. 3j, Table S10), including *VRN1-5A/5D* (Li et al., 2019a), *TaFUL3-2B* (Li et al., 2019a), *WAP3-7A* (Hama et al., 2004). Another 34 factors are ‘conserved’ as orthologue being studied for inflorescence development in other species, including *WAPO2* (*APO2*, Ikeda-Kawakatsu et al., 2012), *TaMADS34* (*OsMADS34/PAP2*, Kobayashi et al., 2010), *TaMOC1* (*MOC1*, Zhang et al., 2015) (Fig. 3j, Table S10). Of note, majority of candidate TFs (n = 185) identified are ‘novel’ for functionally unknown in wheat and other crops (Fig. 3j). These TFs are enriched in ERF, MYB, WRKY, bHLH, B3 families (Extended Data Fig. 3h, Table S10).

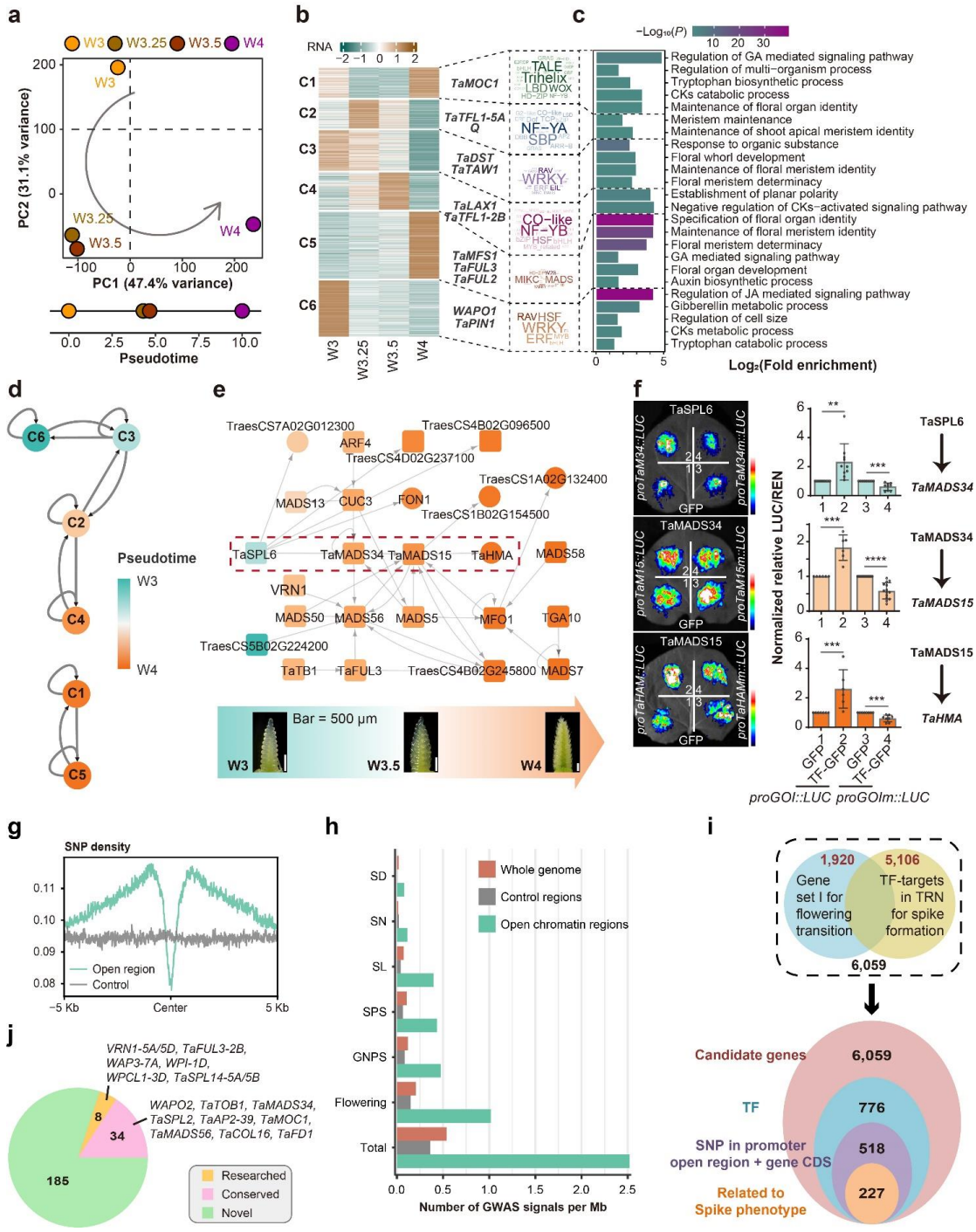


Fig. 3: Transcriptional regulatory network (TRN) governing spike architecture formation.

a, Principal-components plots of gene expression from W3 to W4 stages. The developmental time units (DTU) values calculated based on scaled straight distance between two adjacent stages was shown in the lower panel.

b, K-means clustering of DEGs from W3 to W4 stages and representative genes in each cluster.

c, TF family enrichment and GO enrichment analysis within each cluster in (**b**).

d, Hierarchical transcriptional regulations of sequentially expressed gene clusters from (**b**). *P* value was determined by the hypergeometric test.

e, TRN for represented key TFs participated in spike development, from W3 to W4 stages. Genes were roughly ranked by the expression timing from left to right as indicated different colors gradients. TFs and non-TF coding target genes were in solid rectangle or circle, respectively. The dashed red rectangle frame indicates a four-layer regulation module.

f, Luciferase reporter assays validation of transcriptional regulation among representative TF-target pairs as indicated in the red rectangle frame of (**e**). Mutations of the TF binding sites were introduced in the promoter region of target genes separately. Student's *t* test was used for the statistical significance (**, $P \leq 0.01$; ***, $P \leq 0.001$).

g, SNP density in 5 Kb regions flanking open chromatin regions (green) and control regions (gray). Open chromatin regions show low sequence diversity in contrast to the control.

h, The number of GWAS signals per Mb region of spike morphology related traits in open chromatin regions (light green bar), control regions (gray bar) and whole genome (orange bar). The values of each trait were determined with GWAS signal numbers divided by physical length of the open chromatin regions, control regions and whole genome, respectively.

i, Schematic of the strategy used for key factors identification for shaping spike architecture.

j, Pie chart summary of candidate genes for spike development in each category as indicated. Represented genes were listed.

Validation of the genetic regulatory hierarchy of 'conserved' factors for wheat spike development

Among the ‘conserved’ factors, we again found TaMADS34 (Fig. 3j), as showing up in the TRN governing spike development (Fig. 3e). Thus, we wonder whether TaMADS34 and the identified TaSPL6-TaMADS34-TaMADS15-*TaHMA* module (Fig. 3f) are indeed involved in shaping spike architecture in wheat.

In situ hybridization finely characterizes the spatial-temporal expression profile of these factors. *TaSPL6* was weakly expressed in the SAM, fixedly expressed in the IM tip (red triangle shown at W3 stage) and spikelet meristem (blue triangles shown at W3 and W3.5 stage) (Fig. 4a). *TaMADS34* and *TaMADS15* were highly expressed in the spikelet meristem. *TaHMA*, a downstream target of TaMADS15, was also expressed in the spikelet meristem (Fig. 4a, blue triangles). Similar spatiotemporal expression pattern supports their transcriptional regulatory hierarchy and potential role in spikelet or floret development. Furthermore, we generate RNAi transgenic wheat of *TaSPL6*, *TaMADS34* and *TaMADS15*, respectively (Extended Data Fig. 4a). As expected, *TaMADS34-RNAi* and *TaMADS15-RNAi* plants exhibit similar phenotypes, with shorter SL, decreased SNS and GNPS (Fig. 4b, c, d, Extended Data Fig. 4b). *TaSPL6-RNAi* plants also show altered spike developmental phenotype, but with longer SL, increased SNS and GNPS (Fig. 4b, c, d, Extended Data Fig. 4b). This is likely because that TaSPL6 is on the top layer of hierarchy TRN, it can regulate multi-targets in addition to *TaMADS34* (Fig. 3e). Consistent with the LUC reporter assay *in vitro* (Fig. 3f), *TaMADS34* was significantly down-regulated in *TaSPL6-RNAi* transgenic wheat (Fig. 4e). Similar expression pattern was observed for supporting that TaMADS34 activated *TaMADS15*, and TaMADS15 promoted *TaHMA* by quantitative polymerase chain reaction with reverse transcription (RT–qPCR）(Fig. 4e). Therefore, the spike developmental defects of TF-RNAi lines, similar spatiotemporal expression pattern and relative genes expression level in transgenic wheat prove that TaSPL6-TaMADS34-TaMADS15-*TaHMA* functions as a regulatory module during wheat spike development. We further analyzed the conservation of this hierarchy regulatory module among TaSPL6-TaMADS34-TaMADS15-*TaHMA* in *Triticum* and *Gramineae* from an evolutionary perspective (Fig. 4f). In diploid wheat AA, DD, tetraploid wheat AABB,

and hexaploid wheat AABBDD, there is a MADS34 binding motif (CCATTTTTGG) in the similar promoter region of the *MADS15* homologous genes (Fig. 4f). Similarly, in the promoter region of *MADS34* homologous genes, the binding motif of SPL6 (CCGTACGG) exists (Extended Data Fig. 4c). This indicates that this regulatory module may be conserved within *Triticum*. However, in rice (*Oryza sativa)* or maize (*Zea mays*), the promoter region of *MADS15* lacks the MADS34 binding motif, the promoter region of *MADS34* lacks the SPL6 binding motif, indicating that this regulatory module may be differentiated in grasses (Fig.4f, Extended Data Fig. 4c). Thus, the ‘conserved’ factors identified were indeed involved in regulating spike development in wheat, and the hierarchy transcriptional regulatory circuit is likely maintained within *Triticum*.

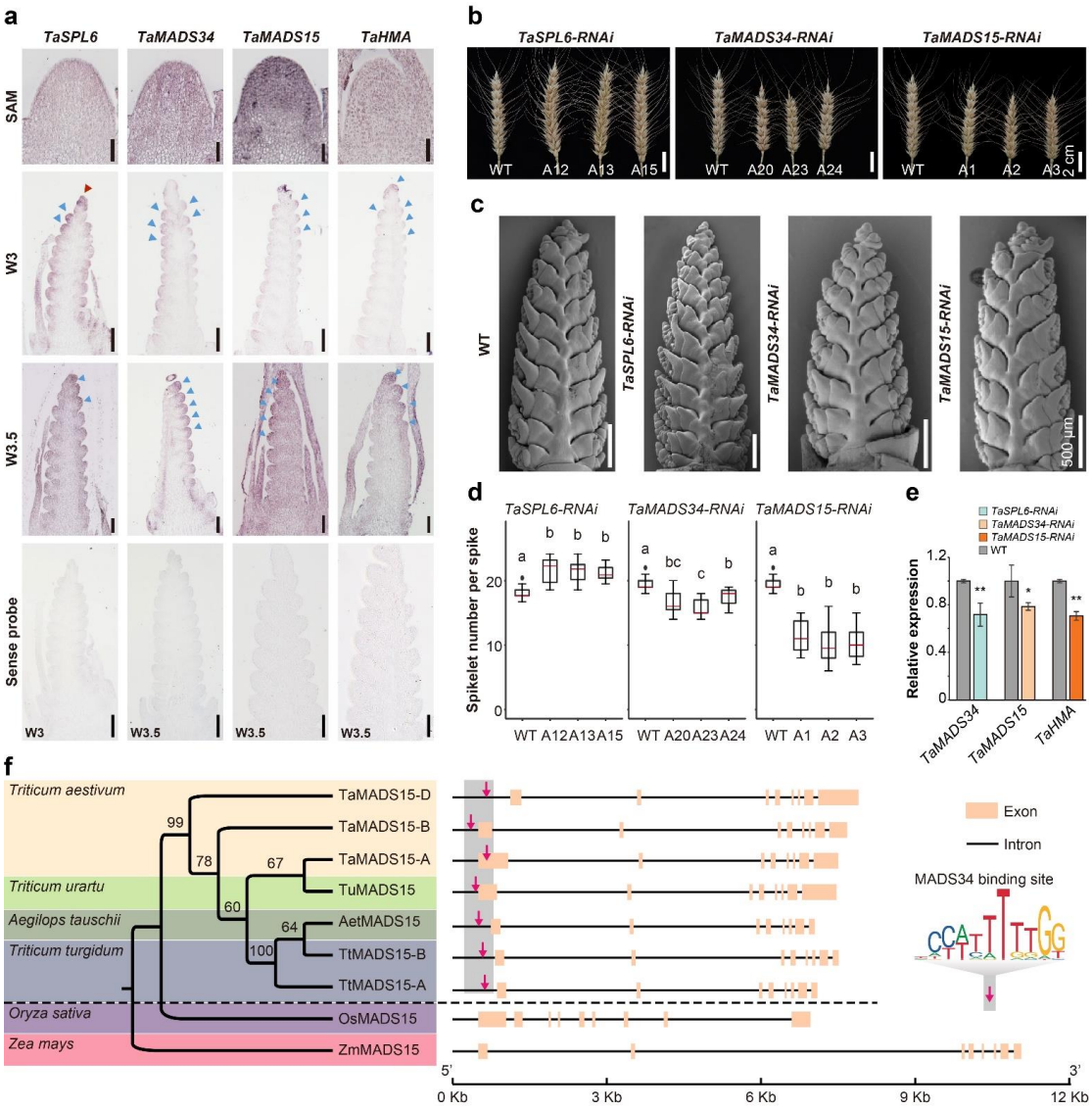


Fig. 4: SPL6-MADS34-MADS15-*HMA* module regulates spike development in wheat.

a, Spatiotemporal expression patterns of *TaSPL6*, *TaMADS34*, *TaMADS15* and *TaHMA* at different spike developmental stages, as indicated by *in situ* hybridization. Sense probe of each gene was detected in the stage with high expression level as negative control. The blue triangles indicate gene expression position. Scale bars = 100 μm.

b, Spike developmental defects of T2 RNAi transgenic plants of *TaSPL6*, *TaMADS34*, *TaMADS15* as compared to wild-type (WT) Fielder. Scale bars = 2 cm.

c, SEM of young spikes from WT and RNAi transgenic plants of *TaSPL6*, *TaMADS34*, *TaMADS15* at W4 stage. Scale bars = 500 μm.

d, Quantification of SNS between WT, *TaSPL6*-RNAi, *TaMADS34-RNAi* and *TaMADS15-RNAi* transgenic plants. Tukey's honestly significant difference (HSD) was used for the statistical significance. In box plots, the box limits indicate the 25th and 75th percentiles, the whiskers indicate the full range of the data, and the centre line indicates the median. Different letters mean significant difference at $P \leq 0.05$.

e, The expression level of *TaMADS34*, *TaMADS14*, *TaHMA* in young spike at W3.5 stage of WT or *TaSPL6-RNAi*, *TaMADS34-RNAi*, *TaMADS15-RNAi* transgenic plants. Expression level of genes in WT is set as 1.0, the relative expression of each gene in RNAi plants is shown as average ± SD of three replicates. Student's *t* test was used for the statistical significance (*, $P \leq 0.05$; **, $P \leq 0.01$).

f, The conservation of TaMADS34 binding motif (CCATTTTTGG) at chromatin accessible region of *MADS15* orthologs in different *Triticum*, but not *Oryza sativa* or *Zea mays*. Phylogenetic tree of corresponding species is indicated on the left. Schematic diagram of gene structure and presence or absence of MADS34 binding CArG motif sites are shown on the right.

Identification of novel regulators for spike development

To evaluate our strategy of identifying novel regulators for wheat spike development, we investigated the spike developmental defects of TILLING mutant lines in hexaploid wheat Cadenza (Krasileva et al., 2017) and KN9204 (Wang et al., 2023), of which the mutated sites were identified by whole exome sequencing.

Of the 185 novel TFs, 122 TFs were found to have at least one mutant line that containing loss-of-function mutation (Fig. 5a, Table S10), with 85 TFs had homozygous mutant lines. In total, 38 TFs (44.71%) of 133 mutant lines exhibited spike developmental defects within three major types, including flowering time difference (Type-I, n = 11), degeneration of spikelet or floret (Type-II, n = 20), and altered SL or SNS (Type-III, n = 18) (Fig. 5b, Table S10). Representative mutant lines of different

types were shown (Fig. 5c, d, Extended Data Fig. 5). For instance, *TraesCS1B02G282300* (*TaERF99-B*) mutant flowered two weeks earlier than the control line; *TraesCS2A02G443800* (*TaWRKY37-A*) mutant showed more sterile spikelets per spike (SSPS) at basal part of inflorescence, resulted in decreased SNS/GNPS; *TraesCS1A02G193200* (*TabHLH009-A*) mutant showed reduced SL and increased SD (Fig. 5c, d, e, Extended Data Fig. 5a). The similar phenotype of *TaWRKY37-A* and *TabHLH009-A* genes also exhibit in the Cadenza mutant (Extended Data Fig. 5b, c). We further confirmed the spatiotemporal expression pattern of those genes by *in situ* hybridization (Fig. 5f). Consistent with their morphologic defects, *TaWRKY37-A* and *TabHLH009-A* are both expressed at the spikelet initiation region (blue triangles at W2.5 stage) and the spikelet meristem (blue triangles at W3.5 stage); while *TaWRKY37-A* is highly expressed in the spikelet meristem at the base of spike, *TabHLH009-A* is highly expressed in the spikelet meristem at the upper region of spike (Fig. 5f).

Taking the advantage of available TILLING mutant lines, we have shown that numerous novel TFs identified by the integration strategy played roles in regulating spike development, which were worth for in-depth study.

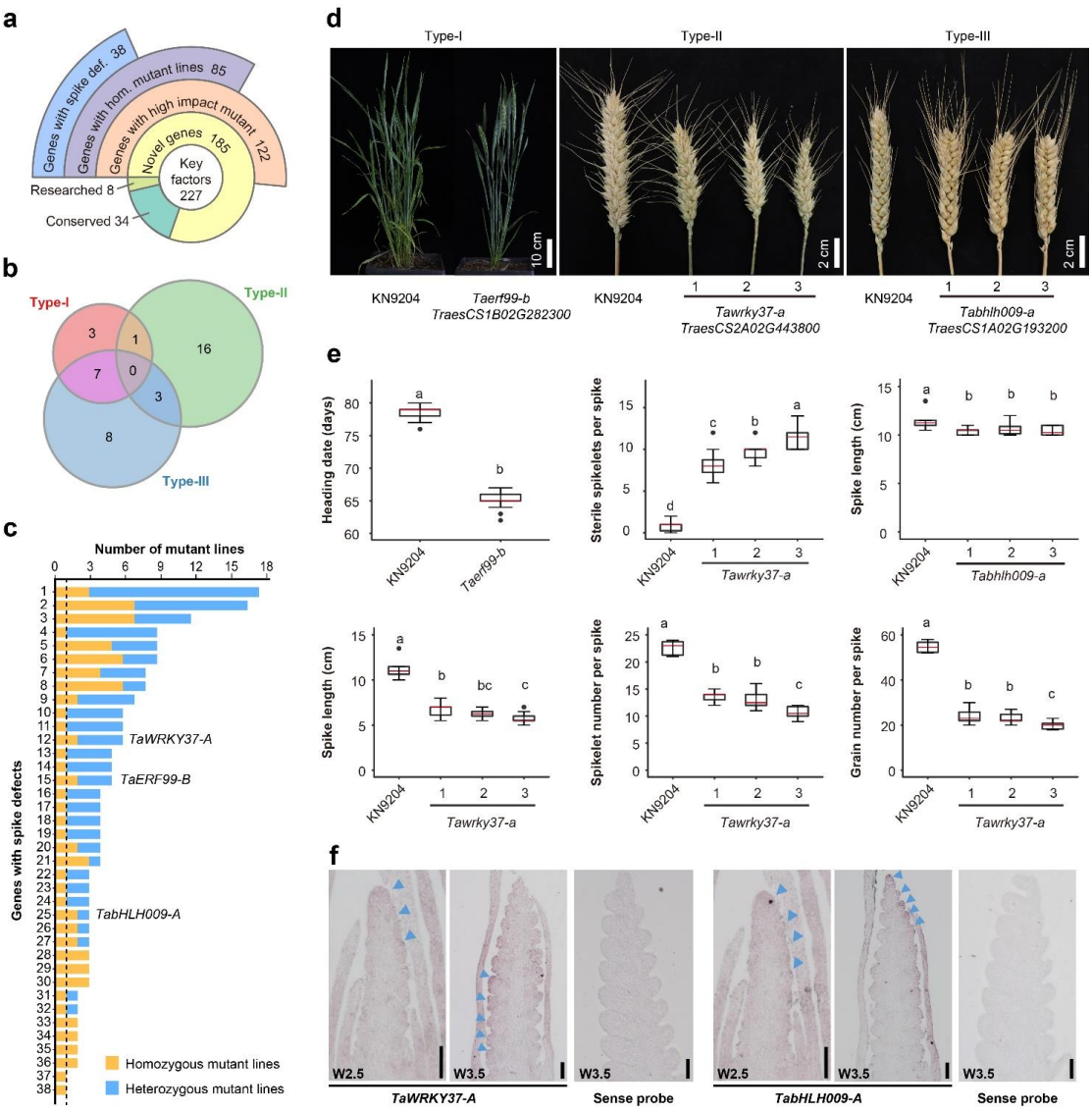


Fig. 5: Novel factors identified regulate wheat spike development.

a, Summary of novel TFs with different categories of KN9204 TILLING mutant lines. def., defects; hom., homozygous.

b, Venn-diagram of different type of spike-related phenotype presented in KN9204 TILLING mutant lines containing homozygous mutation at novel TFs coding genes. Type-I, flowering time difference; Type-II, degeneration of spikelet or floret; Type-III, SL or SNS difference.

c, The overview of 38 novel TFs with spike-related phenotypes. The number of homozygous (orange) and heterozygous (blue) mutant lines were present in stacked bar

chart, with a horizontal dashed line indicating TFs with multiple mutant lines. Three novel TFs used for phenotypic analysis were labeled.

d, Represented KN9204 TILLING mutant lines of type-I/II/III spike developmental defects as compared to control KN9204. Type-I, *Taerf99-b*, *TraesCS1B02G282300*, Scale bar = 10 cm; Type-II, *Tawrky37-a*, *TraesCS2A02G443800*, Scale bar = 2 cm; Type-III, *Tabhlh009-a*, *TraesCS1A02G193200*, Scale bar = 2 cm. To better visualize the arrangement of the spikelets, the side-view of spike was photographed.

e, Quantification of different types of spike developmental defects in represented KN9204 TILLING mutant lines, as heading date of *Taerf99-b*; SSPS, SL, SNS and GNPS of *Tawrky37-a*; SL of *Tabhlh009-a*, the Wild-type is KN9204. Tukey's HSD was used to determine the differences for three or more independent groups comparison of data. Student's *t* test was used for the heading time of *Taerf99-b*. In box plots, the box limits indicate the 25th and 75th percentiles, the whiskers indicate the full range of the data, and the centre line indicates the median. Different letters mean significant difference at $P \leq 0.05$.

f, Spatiotemporal expression patterns of *TaWRKY37-A* and *TabHLH009-A* at W2.5 and W3.5 stages, as indicated by *in situ* hybridization. The blue triangles indicate gene expression position. Sense probe of each gene was detected as negative control. Scale bars = 100 μm.

Regulation of wheat spike architecture by novel factor TaMYB4-A

To further explore the application of integrated TRN and GWAS analysis in dissecting molecular function of individual gene in wheat spike development, we took a novel regulator *TaMYB4-A* for in-depth study.

TraesCS6A02G224200 (*TaMYB4-A*) was one of the TFs from the TRN (Extended Data Fig. 3b). It was also located within a genetic region that significantly associated with SSPS by GWAS analysis using a previously described population (Li et al., 2022) (Extended Data Fig. 6a, Table S11). Two haplotypes (G-type Vs A-type) of peak SNP chr6A:421426337 could separate the natural population with significant differences in

SSPS, explaining 4.82% SSPS variation in the population (Extended Data Fig. 6a, Table S11). Consistently, we find *TaMYB4-A* is initially expressed in the spikelet initiation region (blue triangles) at W2.5 stage, highly expressed in the spikelet and floral meristem (blue triangles) during W3.5, W4 and W4.25 (pistil primordium present) stages, and also detected high signals in the stamen (green stars), pistil primordium (red triangle) and carpel primordium (yellow star) during the floret organ development from W4 to W4.5 (carpel primordium present) stages by *in situ* hybridization (Fig. 6a), which fits its role in determining spikelet number and floret development. Further, we obtained loss-of-function mutant lines of *TaMYB4-A* with a premature stop codon at 108 aa and 89 aa from KN9204 and Cadenza TILLING libraries, respectively (Fig. 6b, Extended Data Fig. 6b, c). The multiple TILLING mutant lines were backcrossed with wild-type to reduce the mutation load and phenotype-genotype co-segregation analysis was performed by Sanger sequencing (Fig. 6b, Extended Data Fig. 6b, c). The homozygous *Tamyb4-a* mutant plants, basically distributed in the smallest quartile of spike length in BC_2F_2 of Cadenza, exhibit significantly reduced SL and SNS but increased spikelet fertility ratio compared with heterozygous and wild-type plants (Fig. 6b, c, Extended Data Fig. 6d, e). The final GNPS is significantly less in *Tamyb4-a* as compared to WT, mainly due to the drastic reduction of SNS (~10%) (Fig. 6c). Similar trends were obtained in the co-segregation test of BC_1F_2 population in KN9204 background for SL, SNS, and spikelet fertility ratio, though GNPS reduction is not that dramatic (Extended Data Fig. 6b, c, d, f, g).

Based on the TRN generated, we extract the hierarchy transcription regulatory module containing TaMYB4-A (Extended Data Fig. 7a). Among the potential upstream regulators of *TaMYB4-A*, WFZP and a DOF type TF TaDOF17 (TraesCS2B02G592700) could repress or active *TaMYB4-A* in the reporter assay in tobacco leaves, respectively (Fig. 6d, Extended Data Fig. 7b). This also fits with the temporal expression pattern of *WFZP*, *TaDOF17* and *TaMYB4-A* during spike formation process. *TaDOF17* and *TaMYB4-A* showed similar pattern but *TaDOF17* changed ahead of *TaMYB4-A*, whereas *WFZP* elevated with *TaMYB4-A* reduction after W3 stage (Extended Data Fig. 7c). *TaMYB4-A* was significantly down-regulated in the

WFZP overexpression transgenic line while up-regulated in *wfzp-a/d* CRISPR lines, further validating that *TaMYB4-A* was repressed by WFZP (Fig. 6e). Among the numerous potential downstream targets of TaMYB4-A (Extended Data Fig. 7a), several genes were confirmed to be regulated by TaMYB4-A in reporter assay in tobacco leaves (Fig. 6d, Extended Data Fig. 7b). This includes *TaAP2-39* (*TraesCS2D02G425700*), of which orthologues being proved to regulate the development of tiller and panicle/ inflorescence by controlling the balance of ABA/GA in rice (Yaish et al., 2010), and *TaI-BAK1* (*TraesCS7D02G416900*), with homologous encoding a Brassinosteroid insensitive 1-associated kinase 1 (BAK1), reported to increase panicle length and grain number per panicle when overexpressed in rice (Khew et al., 2015), as well as *TaPILS7* (*TraesCS5A02G354300*), encoding an auxin efflux carrier component (Fig. 6f). Consistently, *TaAP2-39*, *TaI-BAK1* and *TaPILS7* showed a synchronized temporal expression pattern as *TaMYB4-A* during spike formation process (Extended Data Fig. 7c). Thus, *TaMYB4-A* regulates spikelet number and fertility likely through regulating hormones homeostasis and/or signaling, acting downstream of and repressed by WFZP (Fig. 6f).

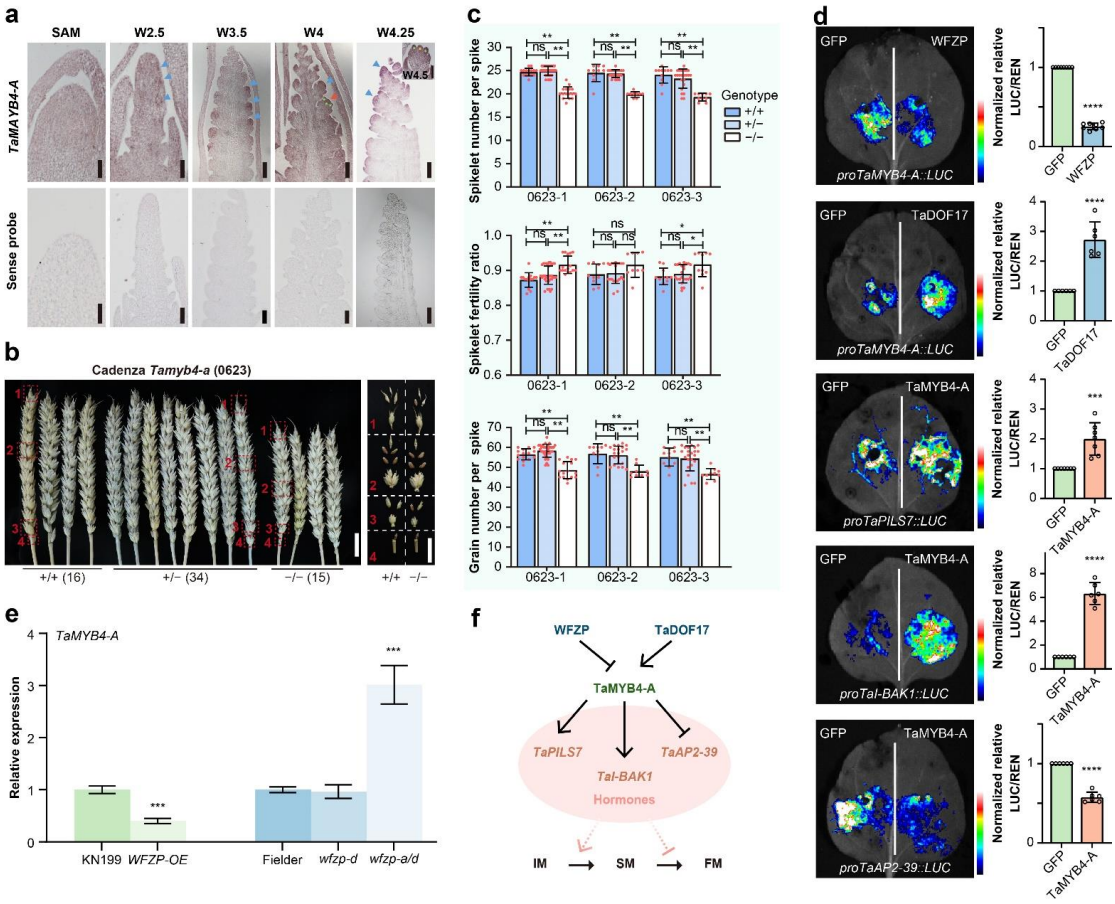


Fig. 6: The novel factor TaMYB4-A regulates spike architecture.

a, Spatiotemporal expression pattern of *TaMYB4-A* is indicated by *in situ* hybridization. The blue/red triangles and green/ yellow stars indicate gene expression position. Sense probe was detected as negative control. Scale bars = 100 μm.

b, The spike phenotype of *Tamyb4-a* (0623) mutant from the Cadenza mutant library. The *Tamyb4-a* was backcrossed with wild-type Cadenza, and the BC_2F_2 were used to perform the co-segregation test. The SL of BC_2F_2 consistent a single gene segregation ratio of 1:2:1 and homozygous *Tamyb4-a* mutant plants were with shorter spikes. Spikelets corresponding to the numbers and positions (red dotted box) were shown on the right for fertility. Scale bars = 2 cm.

c, Statistics comparison of SNS, spikelet fertility ratio and GNPS between Cadenza and *Tamyb4-a* mutant lines. The error bars denote ± SD, SNS and spikelet fertility ratio of each plant were presented in a dot. Student's *t*-test was used for the statistical significance. *, $P \leq 0.05$; **, $P \leq 0.01$; ns, no significant difference.

d, Luciferase reporter assays of TaMYB4-A regulatory network. Student's *t*-test was used for the statistical significance. ***, $P \leq 0.001$; ****, $P \leq 0.0001$.

e, The expression level of *TaMYB4-A* at W3.5 stage in KN199, *WFZP*-OE transgenic plant and Fielder, *wfzp-d*, *wfzp-a/d* CRISPR lines by RT-qPCR. The error bars denote ±SD. Student's *t*-test was used for the statistical significance. ***, $P \leq 0.001$.

f, The working model representing the possible function and genetic network of TaMYB4-A in wheat.

SNP at WFZP binding site within *TaMYB4-A* promoter mediates spike architecture selection during breeding process in China

Next, we wonder how the different haplotypes of *TaMYB4-A* could affect its function and regulation of spike architecture. Among several DNA variations at *TaMYB4-A* locus, SNP-939 (C/T) variation is mostly associated to SSPS, with reference allele (C-type) showing similar SNS, less SSPS, higher spikelet fertility ratio and more GNPS of elite agronomic trait (Fig. 7a, Extended Data Fig. 7d). Interestingly, the SNP-939 (C/T) is located in the promoter region of *TaMYB4-A*, right under the WFZP core binding motif (Fig. 7a). This suggested that the C/T SNP may contribute to *TaMYB4-A* expression difference via affecting WFZP regulation. Indeed, we confirmed the WFZP repressed *TaMYB4-A* via the recognition of the conserved motif, and such repression is abolished with C-to-T mutation in reporter assay in tobacco leaves (Fig. 7b). We further selected 37 different wheat accessions of the two types (24 accessions with C-type and 13 with T-type) to quantify the *TaMYB4-A* expression level in young spike at W3.5 stage and measure the spike morphology. Wheat cultivars of C-type and T-type presented a similar SNS, but T-type was with a lower ratio of fertile spikelets thus less GNPS and a significantly higher level of *TaMYB4-A* expression, while the C-type containing wheat cultivars had higher spikelet fertility ratio, more GNPS and lower level of *TaMYB4-A* expression (Fig. 7c, Extended Data Fig. 7e). Thus, the C/T variation could affect the expression level of *TaMYB4-A* likely through influencing WFZP repression and in turn

associated with varied spikelet fertility, similar as mutation of *Tamyb4-a*, however without significant effects on spikelet numbers.

Furthermore, we wonder how this C/T SNP site within *TaMYB4-A* promoter is selected during the breeding process in China. Percentages of accessions carrying the C-type were considerably higher in landraces (82.08%) and modern cultivars (78.18%) compared with introduced cultivars (45.45%) (Fig. 7d, left). Of note, the frequency of C-type was significantly elevated during the progress of breeding in China. Synchronized with elevation of C-type frequency, the spikelet fertility ratio and GNPS were steadily increased, while SNS represented a decreasing tendency (Fig. 7d, right). Interestingly, the alleles showed distinct distribution characteristics in the major agro-ecological zones of China. The frequency of C-type in humid and semi-humid area regions (basically I-VI zones) was higher than arid and semi-arid regions (basically VII-X zones), and winter wheat regions (I-V) higher than winter-spring mixed regions (IX and X) and then spring wheat regions (VI-VIII) (Fig. 7e). Further analysis indicates that the frequency of C-type was negatively correlated with SL and SNS, while positively correlated with spikelet fertility ratio and GNPS of different agro-ecological zones (Fig. 7f, Extended Data Fig. 7g). The regions I-V have less average SNS or shorter SL while regions VI-X have more SNS or longer SL. Correspondingly, for regions I-V, higher C-type of *TaMYB4-A* could increase the spikelet fertility ratio, thus gain more GNPS. While for regions VI-X, more SNS could reduce the demand for high fertile ratio (Fig. 7g, Extended Data Fig. 7f).

Thus, the elite C-type allele of *TaMYB4-A* is likely originated from Chinese local germplasm and widely selected during the breeding process probably due to the requirement of balancing between spikelet number and fertility to adapt for different agro-ecological zones (Fig. 7h).

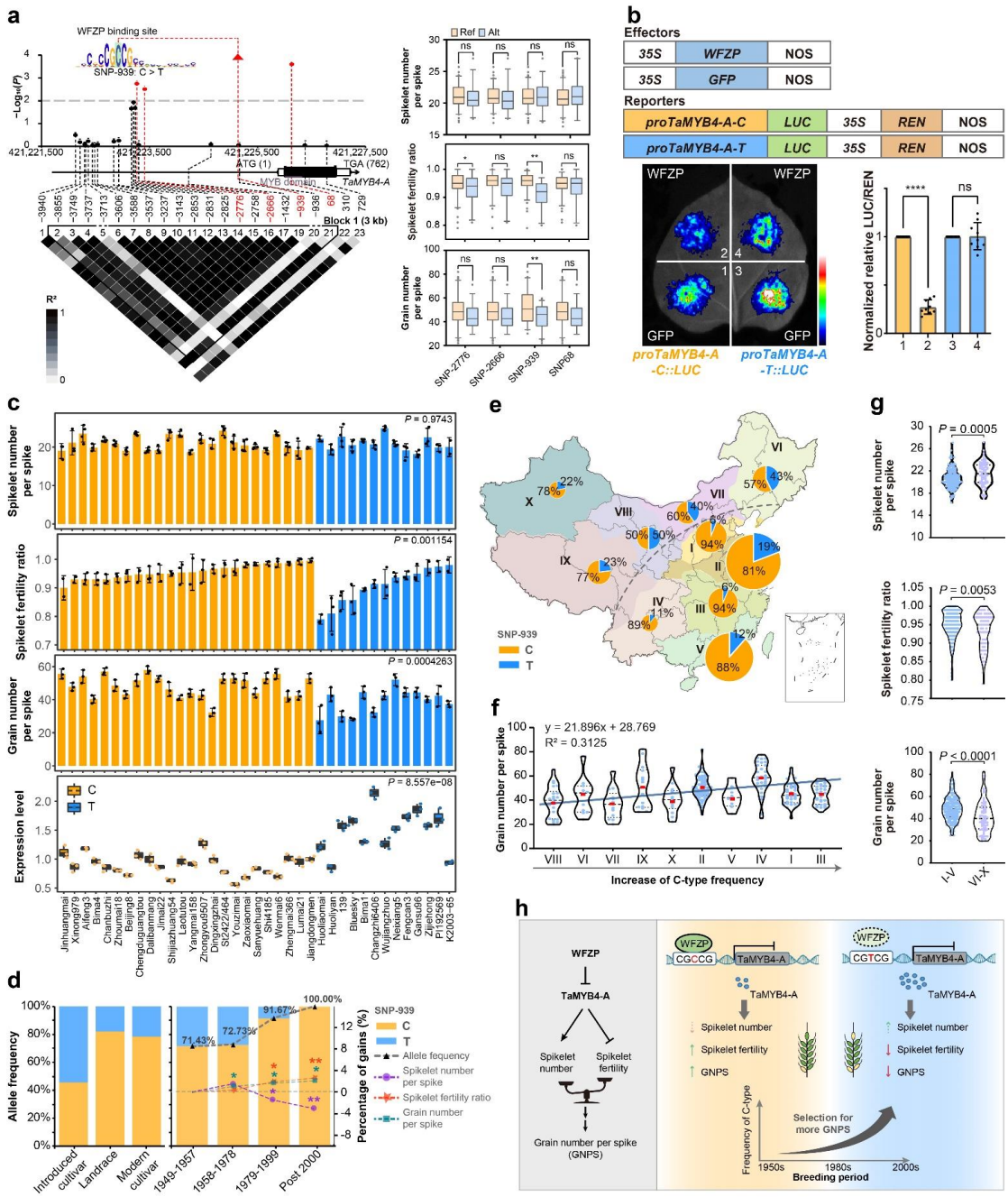


Fig. 7: Key SNP of *TaMYB4-A* at WFZP binding site is associate with breeding selection

a, Association analysis between genetic variations in *TaMYB4-A* and SSPS. The schematic diagram of the ~6 Kb genomics region of *TaMYB4-A* is shown, followed by a LD plot with white to black represents r^2 = 0-1. Dots connected with the red dashed lines indicate the variants that beyond the significance threshold of -$\log_{10}(P)$ =2, with the most significant associated SNP-939 marked by a triangle. The SNP-939 was

located in the core WFZP DNA-binding motif. The box plot illustrated the comparison of SNS, spikelet fertility ratio and GNPS between haplotypes defined by four significantly associated SNPs. Wilcoxon rank-sum test was used to determine the statistical significance; *, $P < 0.05$; **, $P < 0.01$.

b, Luciferase reporter assays of *TaMYB4-A* promoter activity with the WFZP binding site C/T. Schematic diagram showing the vectors used in this array. Student's *t*-test was used for the statistical significance; ****, $P < 0.0001$; ns, no significant difference.

c, The SNS, spikelet fertility ratio, GNPS and *TaMYB4-A* expression level of 37 cultivars with different haplotypes of *TaMYB4-A*. The expression level of *TaMYB4-A* was determined by RT-qPCR using young spike at W3.5 stage, with three biological replicates for each cultivar. Cultivars with C-type and T-type were marked in yellow and blue, respectively. The SNS, spikelet fertility ratio and GNPS were shown in bar with data from each environment in a dot, and significant difference of cultivars between two haplotypes was determined by Student's *t*-test.

d, The percentages of accessions with reference and alternative allele of SNP-939 in different category (left) and breeding periods (right). The line in different colors indicates the allele frequency of reference C-type (black, triangle), SNS (purple, dot), spikelet fertility ratio (red, star), and GNPS (blue, square), respectively. Wilcoxon rank-sum test was used to determine the statistical significance with accessions 1949-1957. Only those that reached the significance threshold are labeled (with the color corresponding to the lines). *, $P < 0.05$; **, $P < 0.01$.

e, The percentages of accessions with reference and alternative allele of SNP-939 in different ecological zones of China. The size of pie charts in the geographical map showing the number of cultivars, with percentage of the two SNP haplotypes in different color (C-type, yellow; T-type, blue). A gray dotted line was used to distinguish the zones I-V (with frequency of C-type more than 80%) and VI-X (with frequency of C-type less than 80%).

f, The distribution of GNPS in ten ecological zones and the correlation of GNPS with the frequency of C-type at SNP-939 of *TaMYB4-A*. From left to right, the C-type frequency in ecological zones increases in turn. The GNPS is positively correlative with the C-type

frequency with the Pearson correlation analysis (n = 9, the zone X was excluded due to small wheat accessions).

g, The zones I-V has lower SNS but higher spikelet fertility ratio and GNPS than zones VI-X. Wilcoxon rank-sum test was used to determine the statistical significance between two groups.

h, A proposed working model of *TaMYB4-A*. The *TaMYB4-A* was regulated by WFZP and regulating the GNPS by balancing spikelet number and spikelet fertility (left part). Natural variations in the promoter contribute to *TaMYB4-A* expression difference, via affecting WFZP binding, and mediates spike architecture selection during breeding process in China. Blunt (⊥) and sharp arrows (→) in the figure separately represent inhibition and promotion effects; the intensities of inhibition effects are indicated by the thickness of the arrows. The fertile and sterile spikelets were indicated in green and yellow, respectively.

Discussion

Inflorescence architecture largely determines the grain productivity via affecting the spikelet and floret development. In this study, we have profiled a time-serial epigenomic landscape consisting of various types of histone modifications, chromatin accessibility, and transcriptomes of wheat shoot apex from vegetative development (SAM) to spike architecture formation (Fig. 1). This allows us to analyze the transcriptional regulation that governs the flowering transition and spikelet initiation, patterning and termination as well as the formation of floret, which are factors affecting wheat inflorescence architecture (Boden et al., 2015; Luo et al., 2023).

Chromatin control of vegetative-to-reproductive transition in wheat

The meristem of shoot apex produces different primordium for the initiation of various tissues, such as leaf primordia from SAM at vegetative stage, and spikelet primordia from IM during flowering transition. Numerous factors control this process, including TaTB1, VRN1, TaFT1, and Ppd-1, mediated by internal and external signals (Luo et al., 2023). Gene expression and cell identity are intimately related to chromatin status (Perino and Veenstra, 2016). During the transition from SAM to W2.5/W3 stage

(vegetative-to-reproductive growth), chromatin accessibility increases (Fig. 1g). However, most of the opened chromatin does not lead to increased gene expression (Fig. 2c), particularly when the genic region is covered by H3K27me3 (Fig. 2g). But chromatin accessibility gains exhibit simultaneous patterns with genes activated during floral induction, including inflorescence meristem identity genes, hormone biosynthesis and signaling (Fig. 2d, f). Moreover, open chromatin could create a primed state for genes to activate at a later developmental stage when H3K27me3 is removed (Fig. 2g, h). The flowering time gene *VRN1* and *TaEhd2*, as well as genes associated with spikelet meristem formation, such as *WAP3*, are among the genes affected by chromatin status. Therefore, a permissive chromatin environment is linked to the activation of key regulators for initiating reproductive development in wheat.

Systematic and efficient identification of novel regulators for spike development

The initiation, distribution and termination of spikelet largely depend on the identity transition of different primordia cells in the context of hormone signaling and transcriptional regulatory network (Feng et al., 2017; Qi et al., 2019). Unraveling the regulatory networks would give us potential candidates for shaping inflorescence structure. In addition, one could expect that genetic variation of such key factors or variations that changing the regulatory circuit could influence the outcome of inflorescence architecture. Following this logic, by taking advantage of our time-serial profiling of transcriptome and chromatin accessibility dataset, in combination with TF-motif binding information in model plant (Castro-Mondragon et al., 2022), we build-up a transcriptional regulatory network that likely governs the spike formation after floral transition (Fig. 3). Numerous TFs are identified to take part in the TRN, including functional studied factors such as *VRN1*, *TaTB1*, *TaFUL3* in wheat and TFs from MADS-box, SBP, YABBY, CO-like, AP2-ERF, NF-YB families that being reported to regulate inflorescence formation in other crops (Fig. 3) (Gao et al., 2019; Yuan et al., 2020; Luo et al., 2023). We have also combined publicly available GWAS analysis with trait focus related to spike development to identify TFs within associated genetic regions. This strategy has led to the identification of 227 TFs, including 42 that have

been functionally analyzed in wheat or other crops. Through TILLING mutant lines screening, we have confirmed the functionality of 38 novel TFs showing spike development defects (Fig. 5). This approach offers an efficient way of screening regulators of agronomic traits in crops. To expand our knowledge, we urge further exploration of TRN for inflorescence development in different ploidy wheat, such as emmer, durum wheat, and wheat relatives like *Brachypodium distachyon*, for discovering novel regulators with elite alleles that are carried in these underutilized germplasm resources. Cases such as TaCOL-B5 and DUO-B1 demonstrate the potential for such discoveries (Wang et al., 2022; Zhang et al., 2022).

Novel regulator for wheat spike development with elite allele enriched during breeding process in China

TRN-generated not only aids in the identification of novel factors involved in inflorescence development but also provides guidance for gene functional study. This is exemplified by the revelation of a hierarchical regulation module involving known individual factors such as TaSPL6-TaMADS34-TaMADS15-*TaHMA* (Fig. 4). Moreover, this regulation module appears to be relatively conserved within the *Triticum* species, suggesting that the TRN generated in wheat has broad application across other species. The suggested regulatory circuit by TRN could further facilitate the functional study of novel factors, including the identification of upstream regulators and downstream targets for TaMYB4-A (Fig. 6). TRN can additionally enable the identification of critical regions involved in transcriptional regulation with high resolution. For example, in the case of *TaMYB4-A*, we discovered that the presence of SNP in the binding motif of the upstream regulator WFZP serves as a selection site during the breeding process in China (Fig. 7). Improving the GNPS was and would be one of the main breeding targets of Chinese breeders to improve grain yield potential, although distinct breeding strategies were used in different agro-ecological zones. Under this premise, the C-type of *TaMYB4-A*, with slightly reduced SNS but increased spikelet fertility and more final GNPS, was selected and the frequency of this elite allele drastically increased. Similar strategy could be used to pinpoint elite alleles of key

regulators for breeding selection and precision genome editing could be employed to stack multiple elite loci for stimulating grain yield.

A valuable database for studying genetic regulation of wheat spike formation

Two recent studies, Wheat-RegNet (Tang et al., 2023) and wGRN (Chen et al., 2023), have been reported, which integrate gene expression, chromatin status and GWAS analysis to predict gene regulation in wheat. However, the limitations of these studies include the lack of tissue specificity and in-depth validation of their predictions in planta. In contrast, we have developed a gene regulatory network based on RNA-seq and ATAC-seq data from the same tissues of different stages of wheat spike development. We have also validated our predictions using mutants or transgenic plants, as well as in-depth functional assays, boosting the biological relevance of our results. Our TRN has uncovered novel biological findings about potential transcription factors (TFs) that shape wheat spike architecture and balance the requirements for breeding selection process.

To facilitate the use of this data source, we have created the Wheat Spike Multi-Omic Database (WSMOD) (http://39.98.48.156:8800/#/) for visualizing gene-centered data such as stage-specific expression, epigenetic landscape, co-expression network, and TRN (Extended Data Fig. 8). This database also includes links to numerous spike developmental trait-related GWAS analyses and the KN9204 TILLING mutant library with >98% mutation coverage (Wang et al., 2023). This would create further opportunities to explore the molecular insights of inflorescence development beyond TFs. Finally, to improve the accuracy and resolution of our TRN and facilitate gene functional studies, we are generating single-cell omic datasets to study the heterogeneous cell populations at individual spike developmental stages.

Methods

Plant materials, growth condition and sampling

The winter wheat cultivar KN9204 was used in this study. The germinated seeds were treated at 4 °C for 30 days. The seedlings were transplanted into soil and grown in the greenhouse at 22 °C/20 °C day/night, under long day conditions (16 h light/8 h dark). The stage-specific shoot apex of wheat was dissected under the stereomicroscope based on the anatomic and morphological features, and immediately frozen in liquid nitrogen and stored at -80 °C. About 10 to 50 spikes were pooled for each of biological replicate for RNA-seq (three replicates), ATAC-seq and CUT&Tag (two replicates) analysis at eight or five development stages.

Transgenic wheat plant generation and spike related morphological trait observation

To obtain RNAi transgenic wheat plants, the specific fragment of *TaSPL6*, *TaMADS34* and *TaMADS15* was separately amplified and inserted into pc336 (*Ubi:GWRNAi:NOS*) vector using gateway cloning method. All constructed vectors were transformed into callus to generate the transgenic plants as described previously (Liu et al., 2023). The pc336 vector was kindly provided by Dr. Daolin Fu at College of Agronomy, Shandong Agricultural University, Tai'an, Shandong, China.

The transgenic lines and mutant lines were grown at the Experimental Station of Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Changping, Beijing for two consecutive years (planted in September 2020 and 2021). Primers for genotyping are listed in Table S12.

The phenotypic traits of mutant lines, T_2 generation transgenic and control plants (lines transformed with an empty vector) which were planted under natural conditions were recorded for 20 to 40 randomly selected transgenic plants 30 days after flowering.

The heading date was defined as the days from sowing to 50% of the spikes fully emerged from the boot. Spike-related traits were measured 15 days after flowering with 15~25 prominent tiller spikes. Spike length was measured from the base of the spike to the tip (excluding the awns), and the fertile and sterile spikelet number per spike was counted separately to obtain the SNS, FSPS, and SSPS. The spikelet fertility ratio is defined as the ratio of fertile and total spikelet numbers. After maturity, these spikes

were harvested and used to measure the GNPS.

Identification of loss-of-function mutations

To investigate the function of novel genes, loss-of-function mutants of *TaERF99-B*, *TaWRKY37-A*, *TabHLH009-A* and *TaMYB4-A* were obtained from sequenced EMS mutant populations of the hexaploid wheat cv. KN9204 (Wang et al., 2023) and Cadenza (Krasileva et al., 2017). All these mutants were premature stop-gained mutations, of which multiple lines (≥ 3 lines) were selected of *TaWRKY37-A*, *TabHLH009-A* and *TaMYB4-A* for KN9204 mutants. The mutants were genotyped by amplifying and Sanger sequencing the genomic regions containing mutation sites in mutant and Wild-type plants (Subgenome-specific amplification primers were described in Table S12). All these mutants were backcrossed with Wild-type for one or two times to eliminate noise from the genetic background prior to phenotyping. For *TaMYB4-A*, the hybrids heterozygous were self-pollinated to generate a segregating population (BC_1F_2 for KN9204 mutants, and BC_2F_2 for Cadenza mutants). From this, all progeny plants were genotyped and conducted to the spike-related phenotypic investigation. The heterozygous plants exhibit similar SL to wild-type, but markedly longer than that of homozygous mutant, with a roughly long: short = 3:1 single gene genetic segregation ratio.

RNA extraction, sequencing, RT-qPCR and *in situ* hybridization

Total RNA was extracted using HiPure Plant RNA Mini Kit according to the manufacturer's instructions (Magen, R4111-02). RNA-seq libraries' construction and sequencing platform were the same as previous description (Zhao et al., 2023), by Annoroad Gene Technology.

First-strand cDNA was synthesized from 2 μg of DNase I-treated total RNA using the TransScript First Strand cDNA Synthesis SuperMix Kit (TransGen, AT301-02). RT-qPCR was performed using the ChamQ Universal SYBR qPCR Master Mix (Vazyme, Q711-02) by QuantStudio5 (Applied biosystems). The expression of interested genes were normalized to Tubulin for calibration, and the relative expression level is calculated via the $2^{-\Delta\Delta CT}$ analysis method (Livak and Schmittgen, 2001). Primers used for RT-qPCR are listed in Table S12.

878 RNA *in situ* hybridization was carried out as described previously (Cui et al., 2010).
879 Fresh young spikes were fixed in formalin-acetic acid-alcohol overnight at 4 °C,
880 dehydrated through a standard ethanol series, embedded in Paraplast Plus tissue-
881 embedding medium (Sigma-Aldrich, P3683), and sectioned at 8 μm width using a
882 microtome (Leica Microsystems, RM2235). Digoxigenin-labeled RNA probes were
883 synthesized using a DIG northern Starter Kit (Roche, 11277073910), according to the
884 manufacturer's instructions. Primers used for probe synthesis are listed in Table S12.

885 **Data Preprocessing and reads alignment**

886 Raw reads were filtered by fastp v0.20.1 with parameter "--detect_adapter_for_pe" for
887 adapters removing, low-quality bases trimming, and reads filtering (Chen et al., 2018).
888 Furthermore, FastQC v0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects
889 /fastqc/) was performed to ensure the high quality of reads.
890 Reads were aligned using either BWA-MEM v0.7.17 with parameter "-M" (for ATAC-
891 seq and CUT&Tag seq) or hisat2 v2.1.0 with default parameters (for RNA-seq) to the
892 reference genome of *Triticum aestivum* cv. Chinese Spring (IWGSC RefSeq v1.0,
893 https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v1.0/)
894 (Appels et al., 2018; Kim et al., 2019; Li and Durbin, 2009). Gene models from the
895 IWGSC Annotation v1.1 were used as the reference and high-confidence genes were
896 used throughout this study. The resulting SAM files were converted to BAM format,
897 sorted, and indexed using Samtools v1.4 (Danecek et al., 2021). Sam files of RNA-seq
898 generated from hisat2 were converted to bam files without deduplication. For ATAC-
899 seq and CUT&Tag, SAM files were further filtered with "samtools view -bS -F 1,804 -
900 f 2 -q 30" to filter out the low-quality mapped reads. Duplicates in the high-quality
901 mapped reads were removed using Picard v2.23.3. Two replicates bam files were
902 merged by samtools. To normalize and visualize the individual and merged replicate
903 datasets, the BAM files were converted to bigwig files using bamCoverage provided
904 by deepTools v3.3.0 with 10 bp bin size and normalized by RPKM (Reads Per Kilobase
905 per Million mapped reads) with parameters "-bs 10 --effectiveGenomeSize
906 14,600,000,000 --normalizeUsing RPKM --smoothLength 50" (Ramirez et al., 2014).

907 **RNA-seq data analyses**

908 The number of paired reads that mapped to each gene was counted using feature Counts
909 v2.0.1 with the parameter "-p" (Liao et al., 2014). The counts files were then used as
910 inputs for DEGs (differentially expressed genes) analysis by DESeq2 v1.26.0 with a
911 threshold "|Log2 Fold Change| ≥ 1 and FDR ≤ 0.05" (Love et al., 2014). The raw counts
912 were further normalized to TPM (Transcripts Per Kilobase Million) for gene expression
913 quantification. For subsequent clustering and visualization, we obtained mean counts
914 by merging three biological replicates. TPM values of genes were Z-scaled and
915 clustered by k-means method and displayed using R package ComplexHeatmap (v2.4.3)
916 (Gu et al., 2016). Functional enrichment was performed using an R package
917 clusterProfiler v3.18.1, and GO annotation files were generated from IWGSC
918 Annotation v1.1 (Yu et al., 2012).

919 **Cut&Tag and ATAC-seq experiment and data analyses**

920 ATAC-seq and CUT&Tag experiments followed the previously described method
921 (Zhao et al., 2023). Tn5 transposase used and tagmentation assay is done following the
922 manual (Vazyme, TD501-01). Libraries were purified with AMPure beads (Beckman,
923 A63881) and sequenced using the Illumina Novaseq platform at Annoroad Gene
924 Technology. Antibodies used for histone modifications are listed in Table S13.

925 Data processing and reads alignment were performed as previously described (Zhao et
926 al., 2023). MACS2 v2.1.4 was used to call peaks. Parameters “-p 1e-3” was used for
927 H3K27ac, H3K4me3 and H2A.Z; parameters “--broad –broad-cutoff 0.05” were used
928 for H3K27me3 and H3K36me3 (Zhang et al., 2008). For ATAC-seq data, MACS2 was
929 used with parameters “--cutoff-analysis --nomodel --shift -100 --extsize 200”. The
930 peaks were annotated by R package ChIPseeker v1.26.2 with "annotatePeak" function
931 (Yu et al., 2015). The gene promoters are defined as 3.0 Kb upstream of gene TSS.

932 For the identification of transcription factor footprints in ATAC-seq peaks, we used the
933 HINT tool v0.13.2 of the Regulatory Genomics Toolbox (RGT) (Gusmao et al., 2014).
934 The custom wheat genome was configured using IWGSC refseq v1.1 Chinese Spring
935 genome based on the introduction of HINT software. TF motifs were downloaded from
936 the JASPAR Plantae database (https://jaspar.genereg.net/) (Castro-Mondragon et al.,
937 2022).

Differential chromatin modification enriched regions detection

For Cut&Tag and ATAC-seq, reads count and CPM normalized value of peaks were calculated by R package DiffBind v2.16.2 with the setting "DBA_SCORE_TMM_READS_EFFECTIVE_CPM". DiffBind was also used to identify differentially accessible regions and histone modification enriched regions with parameters "method = DBA_DESEQ2" and a threshold " |Log2 Fold Change| ≥ 1 and FDR ≤ 0.05".

Pseudotime indexing and gene regulatory network construction

We used the pseudotime indexing method to analyze gene expression as described in previous studies with some modifications (Hao et al., 2021; Leiboff and Hake, 2019; Zhao et al., 2023). All of the expressed genes during spike reproductive development (Stages from W3 to W4) were used to separate samples on a PCA plot. Then, each developmental stage was assigned a location and the Euclidean distance between adjacent stages was calculated and scaled from 0.0 to 10.0. For each gene, we calculated the fitted curve and interpolated the curve into 500 points based on gene expression using the “loess” function in R. We further performed PCA for each gene based on the standardized expression data and used atan2 function in R to order genes based on the time of expression.

For TRNs construction, we only focused on DEGs with TPM values higher than 0.5 in any stage from W3 to W4. For one gene, its potential upstream regulatory TFs was predicted based on the motif present at gene regulatory region. Here we firstly used HINT tool v0.13.2 to identify footprints within ATAC-seq peaks and motifs within the footprints. Then matched the motifs to TFs based on JASPAR Plantae database. TFs in wheat were mapped to TFs of JASPAR Plantae database (Castro-Mondragon et al., 2022) used blastp (v 2.10.1) (Camacho et al., 2009) with criteria “evalue < 1e-10 and identity > 40%”. In this way, we obtained the regulatory relationship between TFs and target genes. We further filtered the obtained TF-target regulation according to the following criteria: Firstly, the TPM values of TFs and target genes must be higher than 0.5 at any stage from W3 to W4 simultaneously. Secondly, we overlapped these TF-target regulations with WGCNA network constructed based on transcriptome of 8

stages from SAM to W4, and only retained TF-target regulations also supported by WGCNA network. In this way we got the final TF-target gene regulatory network. We used *k*-means function in R to cluster genes into 6 categories and performed hypergeometric test to calculate *P*-value of regulation among gene categories.

Phylogenetic tree construction

Wheat and maize orthologs were identified by reciprocal BLAST of rice MADS34 and MADS15 protein sequences. Sequence alignment was performed using MUSCLE v3.8.1551 (Multiple Protein Sequence Alignment, http://www.drive5.com/muscle) with default settings (Edgar, 2004). We only retained amino acid positions that were present in at least 50% of sequences using trimAl v1.4.rev15 (http://trimal.cgenomics.org) with parameters "-gt 0.5" (Capella-Gutierrez et al., 2009). We used RAxML v8.2.12 to create maximum likelihood phylogenetic trees using model PROTGAMMAGTR and 100 rapid bootstraps (Stamatakis, 2014).

Genome-wide and Gene-based association analysis

A previously characterized population (Li et al., 2022), consisting of the Chinese wheat mini-core collection (MCC, 262 accessions) and 25 currently widely grown cultivars, was used to investigate the genetic basis underlying SSPS variation in wheat. Field experiments were carried out in Beijing and Shijiazhuang (Hebei province, China) during three successive cropping seasons (2019-2022), using a completely randomized design with two repetitions. Sixteen seeds of each accession were planted in a 150-cm-long row with 23 cm between rows. At maturity, spikes of the main stem from 10 plants in the center of the row were used for investigates the spike characters. Spikelet without any grains was counted as sterile ones, and the average sterile spikelets number of 10 spikes were defined as the sterile spikelets per spike (SSPS). The best linear unbiased prediction (BLUP) value for the SSPS trait was performed to GWAS analysis, with 984,034 high quality SNPs (missing rates $\leq$ 0.2, MAF $\geq$ 0.05) captured by exome sequencing (Li et al., 2022), using the EMMAX software package (Kang et al., 2010). The *P* value of 1E-04 was determined to be the suggestive threshold for declaring a significant association according to the adjusted Bonferroni method. All nucleotide polymorphisms (unfiltered SNP and short InDels) in 6 Kb genomic region of *TaMYB4*-

A, including exons, intron regions, 4 Kb promoter regions and 0.5 Kb 3'-UTR regions, were performed to the gene-based association with SSPS using Tassel v5.2. Haploview 4.2 was used to calculate the pairwise linkage disequilibrium coefficient (r^2) between SNPs and draw the LD plot. Manhattan plots and quantile-quantile plots were generated using R package "CMplot" (https://github.com/ YinLiLin/R-CMplot).

Spike morphology observation by scanning electron microscopy (SEM)

Photomicrographs of young spikes were taken using a stereomicroscope (Leica Microsystems, S8 APO) equipped with a digital camera (Canon, A640). For SEM, young spikes from each stage were fixed in 2.5% glutaraldehyde at 4°C. After dehydration in a series of ethanol solutions and substitution with 3-methylbutyl acetate, the samples were subjected to critical point drying, coated with platinum, and observed using a variable pressure scanning electron microscope (Hitachi S-3000N).

Luciferase (LUC) reporter assay

For LUC analyses, full-length coding sequences of *TaSPL6*, *TaMADS34*, and *TaMADS15*, *TaMYB4-A*, *WFZP* and *TaDOF17* were cloned into PTF101 vector to generate the effector construct *35S: TF*-GFP, and about 3 Kb promoter fragment of *TaMADS34*, *TaMADS15*, *TaHMA*, *TaMYB4-A*, *TaAP2-39, TaI-BAK1* and *TaPILS7* were amplified and fused in-frame with the CP461-LUC vector to generate the reporter construct *proGOI*: LUC (see Table S12 for primers). The plasmids were transformed into *Agrobacterium* GV3101. The mixture of bacterial solution *35S:*TF-GFP or *35S:* GFP (OD = 0.5), *proGOI:LUC* or *proGOIm:LUC* (OD = 0.5) and P19 (OD = 0.3) in activation buffer (10 mM MES, 150 μM AS, 10 mM $MgCl_2$) was injected to tobacco (*Nicotiana benthamiana*) leaf. pSUPER-GFP, target-pro:LUC and P19 as control. Firefly luciferase (LUC) and Renilla luciferase (REN) activities were measured using dual luciferase assay reagent (Promega, VPE1910) after one day' co-cultivation in dark and two days in light, the relative value of LUC/REN is indicated as average with standard error of multiple replicates.

The establishment of Wheat Spike Multi-Omics Database (WSMOD)

The Wheat Spike Multi-Omics Database (http://39.98.48.156:8800/#/) integrates a core transcription regulatory network during wheat spike development, encompassing multi-

omics data such as gene expression, chromatin accessibility, multiple histone modifications, SNP information, and spike-related GWAS signals (Extended Data Fig. 8). Additionally, a co-expression network has been constructed using the expression data of 8 stages during spike development. Users can access the expression, regulatory network, and other relevant information by inputting the gene ID via six functional modules: Search by Gene ID, WGCNA, TRN, Epigenomic Status, GWAS, and KN9204 EMS Mutants. For search by gene ID function, WSMOD provides gene expression patterns, orthologs in rice and Arabidopsis thaliana, open chromatin regions and footprint, and other information. WGCNA function provides gene-gene co-expression relationships and network visualization. TRN function provides information of genes upstream TFs, TF binding sites and motifs, as well as downstream target genes. Epigenomic status function offers genome visualization of transcription, chromatin accessibility, various histone modifications, variations and QTL sites. GWAS function provides spike-related QTL regions and gene information in regions. KN9204 EMS Mutants function provides a quick link to KN9204 TILLING Database in which users can conveniently query and order the EMS mutants of the interested gene. All search results are presented in images and interactive tables that can be downloaded for further analysis. These multi-omics data sources and transcription regulatory networks would deepen our understanding of genetic regulation in wheat spike development and provide information for gene function analysis.

Statistics and data visualization

R (https://cran.r-project.org/;version 4.0.2) was used to compute statistics and generate plots if not specified. The Integrative Genomics Viewer (IGV) was used for the visual exploration of genomic data (Robinson et al., 2011). The correlation between two groups of data was conducted with the Pearson analysis (Fig. 2b, e, 7f, Extended Data Fig. 7g) and *P* values are determined by the two-sided Pearson correlation coefficient analysis (Fig. 2e). For enrichment analysis, Hypergeometric test (Fig. 2c, 3d) and Fisher's exact test were used (Fig. 2f, 3c, Extended Data Fig. 2c, d). For two groups' comparison of data that fit a normal distribution, the Student's two-tailed t-test was used (Fig. 3f, 4e, 5e, 6c, d, e, 7b, c, Extended Data Fig. 4a, 6e, f, 7e). For two groups'

comparison of data that does not fit a normal distribution, the Wilcoxon rank-sum test was used (Fig. 7a, d, g, Extended Data Fig. 3g, 6a, 7d, f). Tukey's honestly significant difference (HSD) was used to determine the differences for three or more independent groups comparison of data (Fig. 4d, 5e, Extended Data Fig. 4b, 5a, c).

Data availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Chen et al., 2021) in National Genomics Data Center (CNCB-NGDC Members and Partners, 2022), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (CRA008877) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Code availability

Code used for all processing and analysis is available at Github (https://github.com/yx-xu/A-systematic-framework-to-discover-gene-regulatory-networks-of-wheat-spike-development-for-yield-improvement).

Funding

This research is supported by the National Natural Sciences Foundation of China (31921005) to J.X., Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24010204) to J.X., National Key Research and Development Program of China (2021YFD1201500) to J.X., and the Major Basic Research Program of Shandong Natural Science Foundation (ZR2019ZD15) to X.-Y. Zhao.

Author contributions

J.X. designed and supervised the research, J.X., X.-L. L., Y.-X.X., D.-Z.W. wrote the manuscript. X.-L. L. did the sample collection and *in situ* hybridization; X.-L. L. and X.-Y.Z. did plasmid construction, RT-qPCR and co-segregation experiment. X.-M.B. did wheat transformation; X.-L. L. and Y.-M.Y. performed CUT&Tag, ATAC-seq and RNA-seq experiments; X.-Y. Z. and H.-Z. W. did the reporter assay; Y.-L.D., X.-Y.Z., F.L., X.-S.Z. and X.-D.F. provide some raw data or plant materials; A.-L. L. and Y.-Q. C. performed the GWAS; X.-Y. X., D.-Z. W. performed bio-informatics analysis; J.-F. J., X.-Y.Z., F.L., X.-S. Z. and X.-D.F. polished the manuscript; X.-L. L., Y.-X. X., D.-

Z.W., Y.-M.Y. and J.X. prepared all the figures. All authors discussed the results and commented on the manuscript.

Acknowledgements

We thank Dr. X.G. Liu (College of Life Sciences, Hebei Normal University) and Drs. Z.F. Ni and J. Liu (College of Agronomy and Biotechnology, China Agricultural University) for the *WFZP* transgenic wheat materials and mutants.

Competing interests

The authors declare no competing interests

References

Abou-Elwafa, S. F., Shehzad, T. (2021). Genetic diversity, GWAS and prediction for drought and terminal heat stress tolerance in bread wheat (*Triticum aestivum* L.). Genet Resour Crop Evol **68**:711-728. 10.1007/s10722-020-01018-y.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C.J., Choulet, F., Distelfeld, A., Poland, J., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science **361**:661. eaar7191 10.1126/science.aar7191.

Boden, S.A., Cavanagh, C., Cullis, B.R., Ramm, K., Greenwood, J., Finnegan, E.J., Trevaskis, B., and Swain, S.M. (2015). *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. Nat Plants **1**:1-6. 10.1038/Nplants.2014.16.

Bonifer, C., Cockerill, P.N. (2017). Chromatin priming of genes in development: Concepts, mechanisms and consequences. Exp Hematol. **49**:1-8. doi: 10.1016/j.exphem.2017.01.003

Bonnett, O.T. (1936). The development of the wheat spike. J Agric Res **53**:0445-0451.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol **109**:21.29.21-21.29.29. 10.1002/0471142727.mb2129s109.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinform **10**:421. 10.1186/1471-2105-10-421.

Cao, J., Liu, K., Song, W., Zhang, J., Yao, Y., Xin, M., Hu, Z., Peng, H., Ni, Z., Sun, Q., et al. (2021). Pleiotropic function of the *SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE* gene *TaSPL14* in wheat plant architecture. Planta **253**:44. 10.1007/s00425-020-03531-x.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972-1973. 10.1093/bioinformatics/btp348.

Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R.B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Perez, N., et al. (2022). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids

Res **50**:D165-D173. 10.1093/nar/gkab1113.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**:i884-i890. 10.1093/bioinformatics/bty560.

Chen, T., Chen, X., Zhang, S., Zhu, J., Tang, B., Wang, A., Dong, L., Zhang, Z., Yu, C., Sun, Y., et al. (2021). The genome sequence archive family: toward explosive data growth and diverse data types. Genom Proteom Bioinf **19**:578-583. 10.1016/j.gpb.2021.08.001.

Chen, Y., Guo, Y., Guan, P., Wang, Y., Wang, X., Wang, Z., Qin, Z., Ma, S., Xin, M., Hu, Z., Yao, Y., Ni, Z., Sun, Q., Guo, W., Peng, H. (2023). A wheat integrative regulatory network from large-scale complementary functional datasets enables trait-associated gene discovery for crop improvement. Mol Plant **16**:393-414. doi: 10.1016/j.molp.2022.12.019.

Concia, L., Veluchamy, A., Ramirez-Prado, J.S., et al. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biol **21**:104. 10.1186/s13059-020-01998-1.

Cui, R., Han, J., Zhao, S., Su, K., Wu, F., Du, X., Xu, Q., Chong, K., Theissen, G., and Meng, Z. (2010). Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). Plant J **61**:767-781. 10.1111/j.1365-313X.2009.04101.x.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience **10**10.1093/gigascience/giab008.

Debernardi, J.M., Greenwood, J.R., Jean Finnegan, E., Jernstedt, J., and Dubcovsky, J. (2020). *APETALA 2*-like genes *AP2L2* and *Q* specify lemma identity and axillary floral meristem development in wheat. Plant J **101**:171-187. 10.1111/tpj.14528.

Dixon, L.E., Greenwood, J.R., Bencivenga, S., Zhang, P., Cockram, J., Mellers, G., Ramm, K., Cavanagh, C., Swain, S.M., and Boden, S.A. (2018). *TEOSINTE BRANCHED1* regulates inflorescence architecture and development in bread wheat (*Triticum aestivum*). Plant Cell **30**:563-581. 10.1105/tpc.17.00961.

Du, D., Zhang, D., Yuan, J., Feng, M., Li, Z., Wang, Z., Zhang, Z., Li, X., Ke, W., Li, R., et al. (2021). *FRIZZY PANICLE* defines a regulatory hub for simultaneously controlling spikelet formation and awn elongation in bread wheat. New Phytol **231**:814-833. 10.1111/nph.17388.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform **5**:113. 10.1186/1471-2105-5-113.

Feng, N., Song, G.Y., Guan, J.T., Chen, K., Jia, M.L., Huang, D.H., Wu, J.J., Zhang, L.C., Kong, X.Y., Geng, S.F., et al. (2017). Transcriptome Profiling of Wheat Inflorescence Development from Spikelet Initiation to Floral Patterning Identified Stage-Specific Regulatory Genes. Plant Physiol **174**:1779-1794. 10.1104/pp.17.00310.

Finnegan, E.J., Ford, B., Wallace, X., Pettolino, F., Griffin, P.T., Schmitz, R.J., Zhang, P., Barrero, J.M., Hayden, M.J., Boden, S.A., et al. (2018). Zebularine treatment is associated with deletion of *FT-B1* leading to an increase in spikelet number in bread wheat. Plant Cell Environ **41**:1346-1360. 10.1111/pce.13164.

Gao, X.Q., Wang, N., Wang, X.L., and Zhang, X.S. (2019). Architecture of Wheat Inflorescence: Insights from Rice. Trends Plant Sci **24**: 802-809. 10.1016/j.tplants.2019.06.002.

Gauley, A., and Boden, S.A. (2019). Genetic pathways controlling inflorescence architecture and development in wheat and barley. J Integr Plant Biol **61**:296-309. 10.1111/jipb.12732.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in

multidimensional genomic data. Bioinformatics **32**:2847-2849. 10.1093/bioinformatics/btw313.

Guo, L.J., Ma, M., Wu, L.N., Zhou, M.D., Li, M.Y., Wu, B.W., Li, L., Liu, X.L., Jing, R.L., Chen, W., et al. (2022). Modified expression of *TaCYP78A5* enhances grain weight with yield potential by accumulating auxin in wheat (*Triticum aestivum* L.). Plant Biotechnol J **20**:168-182. 10.1111/pbi.13704.

Gupta, P.K., Kulwal, P.L., Jaiswal, V. (2019). Association mapping in plants in the post-GWAS genomics era. Adv. Genet. **104**: 75e154. 10.1016/bs.adgen.2018.12.001.

Gusmao, E.G., Dieterich, C., Zenke, M., and Costa, I.G. (2014). Detection of active transcription factor binding sites with the combination of DNase hypersensitivity and histone modifications. Bioinformatics **30**:3143-3151. 10.1093/bioinformatics/btu519.

Hama, E., Takumi, S., Ogihara, Y., and Murai, K. (2004). Pistillody is caused by alterations to the class-B MADS-box gene expression pattern in alloplasmic wheats. Planta **218**:712-720. 10.1007/s00425-003-1157-6.

Hao, C.Y., Jiao, C.Z., Hou, J., Li, T., Liu, H.X., Wang, Y.Q., Zheng, J., Liu, H., Bi, Z.H., Xu, F.F., et al. (2020). Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. Mol Plant **13**:1733-1751. 10.1016/j.molp.2020.09.001.

Hao, Z., Zhang, Z., Xiang, D., Venglat, P., Chen, J., Gao, P., Datla, R., and Weijers, D. (2021). Conserved, divergent and heterochronic gene expression during *Brachypodium* and *Arabidopsis* embryo development. Plant Reprod **34**:207-224. 10.1007/s00497-021-00413-4.

He, Y, Li, Z. (2018) Epigenetic environmental memories in plants: establishment, maintenance, and reprogramming. Trends Genet **34**:856-866. doi: 10.1016/j.tig.2018.07.006.

Ikeda-Kawakatsu, K., Maekawa, M., Izawa, T., Itoh, J.I., and Nagato, Y. (2012). *ABERRANT PANICLE ORGANIZATION 2/RFL,* the rice ortholog of *Arabidopsis LEAFY,* suppresses the transition from inflorescence meristem to floral meristem through interaction with *APO1*. Plant Journal **69**:168-180. 10.1111/j.1365-313X.2011.04781.x.

Jablonski, B., Ogonowska, H., Szala, K., Bajguz, A., Orczyk, W., and Nadolska-Orczyk, A. (2020). Silencing of *TaCKX1* Mediates Expression of other *TaCKX* genes to increase yield parameters in wheat. Int J Mol Sci **21**: 4809. 10.3390/ijms21134809.

Jablonski, B., Szala, K., Przyborowski, M., Bajguz, A., Chmur, M., Gasparis, S., Orczyk, W., and Nadolska-Orczyk, A. (2021). *TaCKX2.2* genes coordinate expression of other *TaCKX* family members, regulate phytohormone content and yield-related traits of wheat. Int J Mol Sci **22**: 4142. 10.3390/ijms22084142.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun **10**:1930. 10.1038/s41467-019-09982-5.

Khew, C.Y., Teo, C.J., Chan, W.S., Wong, H.L., Namasivayam, P., and Ho, C.L. (2015). Brassinosteroid insensitive 1-associated kinase 1 (OsI-BAK1) is associated with grain filling and leaf development in rice. J Plant Physiol **182**:23-32. 10.1016/j.jplph.2015.05.003.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol **37**:907-915. 10.1038/s41587-019-0201-4.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat Rev Genet **20**:207-220. 10.1038/s41576-018-0089-8.

1212 **Kobayashi, K., Maekawa, M., Miyao, A., Hirochika, H., and Kyozuka, J.** (2010). *PANICLE*
1213 *PHYTOMER2* (*PAP2*), encoding a SEPALLATA subfamily MADS-box protein, positively controls
1214 spikelet meristem identity in rice. Plant and Cell Physiol **51**:47-57. 10.1093/pcp/pcp166.
1215 **Kong, X.C., Wang, F., Geng, S.F., Guan, J.T., Tao, S., Jia, M.L., Sun, G.L., Wang, Z.Y., Wang, K.,**
1216 **Ye, X.G., et al.** (2022). The wheat *AGL6*-like MADS-box gene is a master regulator for floral organ
1217 identity and a target for spikelet meristem development manipulation. Plant Biotechnol J **20**:75-88.
1218 10.1111/pbi.13696.
1219 **Koppolu, R., Chen, S., and Schnurbusch, T.** (2022). Evolution of inflorescence branch modifications
1220 in cereal crops. Curr Opin Plant Biol **65**:102168. 10.1016/j.pbi.2021.102168.
1221 **Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J.,**
1222 **Ramirez-Gonzalez, R.H., Wang, X.D., Borrill, P., et al.** (2017). Uncovering hidden variation in
1223 polyploid wheat. Proc Natl Acad Sci USA **114**:E913-E921. 10.1073/pnas.1619268114.
1224 **Kuzay, S., Lin, H., Li, C., Chen, S., Woods, D.P., Zhang, J., Lan, T., von Korff, M., and Dubcovsky,**
1225 **J.** (2022). *WAPO-A1* is the causal gene of the 7AL QTL for spikelet number per spike in wheat.
1226 PLoS Genet **18**:e1009747. 10.1371/journal.pgen.1009747.
1227 **Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J.A., Brown-Guedira,**
1228 **G., DeWitt, N., et al.** (2019). Identification of a candidate gene for a QTL for spikelet number per
1229 spike on wheat chromosome arm 7AL by high-resolution genetic mapping. Theor Appl Genet
1230 **132**:2689-2705. 10.1007/s00122-019-03382-5.
1231 **Leiboff, S., and Hake, S.** (2019). Reconstructing the Transcriptional Ontogeny of Maize and Sorghum
1232 Supports an Inverse Hourglass Model of Inflorescence Development. Curr Biol **29**:3410-3419
1233 e3413. 10.1016/j.cub.2019.08.044.
1234 **Li, A., Hao, C., Wang, Z., Geng, S., Jia, M., Wang, F., Han, X., Kong, X., Yin, L., Tao, S., et al.**
1235 (2022). Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant
1236 architecture and grain yield. Mol Plant **15**:504-519. 10.1016/j.molp.2022.01.004.
1237 **Li, C., Lin, H., Chen, A., Lau, M., Jernstedt, J., and Dubcovsky, J.** (2019a). Wheat *VRN1*, *FUL2* and
1238 *FUL3* play critical and redundant roles in spikelet development and spike determinacy.
1239 Development **146** 10.1242/dev.175398.
1240 **Li, H., and Durbin, R.** (2009). Fast and accurate short read alignment with Burrows-Wheeler transform.
1241 Bioinformatics **25**:1754-1760. 10.1093/bioinformatics/btp324.
1242 **Li, L., Shi, F., Wang, Y., Yu, X., Zhi, J., Guan, Y., Zhao, H., Chang, J., Chen, M., Yang, G., et al.**
1243 (2020). TaSPL13 regulates inflorescence architecture and development in transgenic wheat
1244 (*Triticum aestivum* L.). Plant Sci **296**:110516. 10.1016/j.plantsci.2020.110516.
1245 **Li, M.X., Yeung, J.M., Cherny, S.S., Sham, P.C.** (2012). Evaluating the effective numbers of
1246 independent tests and significant *p*-value thresholds in commercial genotyping arrays and public
1247 imputation reference datasets. Hum Genet **131**:747-756. doi: 10.1007/s00439-011-1118-2.
1248 **Li, Y., Fu, X., Zhao, M., Zhang, W., Li, B., An, D., Li, J., Zhang, A., Liu, R., and Liu, X.** (2018). A
1249 genome-wide view of transcriptome dynamics during early spike development in bread wheat. Sci
1250 Rep **8**:15338. 10.1038/s41598-018-33718-y.
1251 **Li, Y., Li, L., Zhao, M., Guo, L., Guo, X., Zhao, D., Batool, A., Dong, B., Xu, H., Cui, S., et al.** (2021).
1252 Wheat *FRIZZY PANICLE* activates *VERNALIZATION1-A* and *HOMEOBOX4-A* to regulate spike
1253 development in wheat. Plant Biotechnol J **19**: 1141-1154. 10.1111/pbi.13535.
1254 **Li, Z., Wang, M., Lin, K., Xie, Y., Guo, J., Ye, L., Zhuang, Y., Teng, W., Ran, X., Tong, Y., Xue, Y.,**
1255 **Zhang, W., Zhang, Y.** (2019b) The bread wheat epigenomic map reveals distinct chromatin

architectural and evolutionary features of functional genetic elements. Genome Biol **20**:139. 10.1186/s13059-019-1746-8.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**:923-930. 10.1093/bioinformatics/btt656.

Lin, X.L., Wu, F., Du, X.Q., Shi, X.W., Liu, Y., Liu, S.J., Hu, Y.X., Theissen, G., and Meng, Z. (2014). The pleiotropic *SEPALLATA*- like gene *OsMADS34* reveals that the 'empty glumes' of rice (*Oryza sativa*) spikelets are in fact rudimentary lemmas. New Phytol **202**:689-702. 10.1111/nph.12657.

Liu, X., Bie, X., Lin, X., Li, M., Wang, H., Zhang, X., Yang, Y., Zhang, C., Zhang, X., and Xiao, J. (2023). Uncovering transcriptional regulatory network during regeneration for boosting wheat transformation. Nat Plants. 10.1038/s41477-023-01406-z.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods **25**:402-408. 10.1006/meth.2001.1262.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**:550. 10.1186/s13059-014-0550-8.

Luo, X.M., Yang, Y.M., Lin, X.L., Xiao, J. (2023). Deciphering spike architecture formation towards yield improvement in wheat. J Genet Genomics S1673-8527(23)00054-1. 10.1016/j.jgg.2023.02.015.

Pang, Y., Liu, C., Wang, D., St Amand, P., Bernardo, A., Li, W., He, F., Li, L., Wang, L., Yuan, X., et al. (2020). High-Resolution Genome-wide Association Study Identifies Genomic Regions and Candidate Genes for Important Agronomic Traits in Wheat. Mol Plant **13**:1311-1327. 10.1016/j.molp.2020.07.008.

Perez-Gianmarco, T.I., Severini, A.D., and Gonzalez, F.G. (2020). *Photoperiod-sensitivity genes* (*Ppd-1*): quantifying their effect on the photoperiod response model in wheat. J Exp Bot **71**:1185-1198. 10.1093/jxb/erz483.

Perino, M., Veenstra, G.J. (2016). Chromatin Control of Developmental Dynamics and Plasticity. Dev Cell **38**: 610-620. 10.1016/j.devcel.2016.08.004.

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. 10.1186/s12864-019-5726-x.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res **42**:W187-191. 10.1093/nar/gku365.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol **29**:24-26. 10.1038/nbt.1754.

Sehgal, D., and Dreisigacker, S. (2022). GWAS case studies in wheat. Methods Mol Biol **2481**:341-351. 10.1007/978-1-0716-2237-7_19.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312-1313. 10.1093/bioinformatics/btu033.

Su, Y., Liu, J., Liang, W., Dou, Y., Fu, R., Li, W., Feng, C., Gao, C., Zhang, D., Kang, Z., et al. (2019). Wheat *AGAMOUS LIKE 6* transcription factors function in stamen development by regulating the expression of *TaAPETALA3*. Development **146**: dev177527. 10.1242/dev.177527.

Sun, C.W., Zhang, F.Y., Yan, X.F., Zhang, X.F., Dong, Z.D., Cui, D.Q., and Chen, F. (2017). Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat

from the Yellow and Huai Valley of China. Plant Biotechnol J **15**:953-969. 10.1111/pbi.12690.

Sun, J., Bie, X.M., Chu, X.L., Wang, N., Zhang, X.S., Gao, X.Q. (2023). Genome-edited *TaTFL1-5* mutation decreases tiller and spikelet numbers in common wheat. Front Plant Sci **14**:1142779. 10.3389/fpls.2023.1142779.

Tang, T., Tian, S., Wang, H., Lv, X., Xie, Y., Liu, J., Wang, M., Zhao, F., Zhang, W., Li, H., Zhang, Y. (2023). Wheat-RegNet: An encyclopedia of common wheat hierarchical regulatory networks. Mol Plant. **16**:318-321. doi: 10.1016/j.molp.2022.12.018.

VanGessel, C., Hamilton, J., Tabbita, F., Dubcovsky, J., and Pearce, S. (2022). Transcriptional signatures of wheat inflorescence development. Sci Rep **12**:17224. 10.1038/s41598-022-21571-z.

Waddington, S. R., Cartwright, P. M., and Wall, P. C. (1983). A quantitative scale of spike initial and pistil development in barley and wheat. Ann Bot **51**: 119-130.

Wang, D., Li, Y., Wang, H., Xu, Y., Yang, Y., Zhou, Y., Chen, Z., Zhou, Y., Gui, L., Guo, Y., Zhou, C., Tang, W., Zheng, S., Wang, L., Guo, X., Zhang, Y., Cui, F., Lin, X., Jiao, Y., He, Y., Li, J., He, F., Liu, X., Xiao, J. (2023). Boosting wheat functional genomics via an indexed EMS mutant library of KN9204. Plant Commun 10.1016/j.xplc.2023.100593.

Wang, Q.L., Sun, A.Z., Chen, S.T., Chen, L.S., and Guo, F.Q. (2018). SPL6 represses signalling outputs of ER stress in control of panicle cell death in rice. Nat Plants **4**:280-288. 10.1038/s41477-018-0131-z.

Wang, Y., Yu, H., Tian, C., Sajjad, M., Gao, C., Tong, Y., Wang, X., and Jiao, Y. (2017). Transcriptome association identifies regulators of wheat spike architecture. Plant Physiol **175**:746-757. 10.1104/pp.17.00694.

Wang, Y., Du, F., Wang, J., Wang, K., Tian, C., Qi, X., Lu, F., Liu, X., Ye, X., and Jiao, Y. (2022). Improving bread wheat yield through modulating an unselected AP2/ERF gene. Nat Plants 10.1038/s41477-022-01197-9.

Wu, F., Shi, X., Lin, X., Liu, Y., Chong, K., Theissen, G., and Meng, Z. (2017). The ABCs of flower development: mutational analysis of *AP1/FUL*-like genes in rice provides evidence for a homeotic (A)-function in grasses. Plant J **89**:310-324. 10.1111/tpj.13386.

Xiao, J., Liu, B., Yao, Y.Y., Guo, Z.F., Jia, H.Y., Kong, L.R., Zhang, A.M., Ma, W.J., Ni, Z.F., Xu, S.B., et al. (2022). Wheat genomic study for genetic improvement of traits in China. Sci China Life Sci **65**:1718-1775. 10.1007/s11427-022-2178-7.

Xu, M., Hu, T., Zhao, J., Park, M.Y., Earley, K.W., Wu, G., Yang, L., and Poethig, R.S. (2016). Developmental Functions of miR156-Regulated *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) Genes in *Arabidopsis thaliana*. PLoS Genet **12**:e1006263. 10.1371/journal.pgen.1006263.

Yaish, M.W., El-kereamy, A., Zhu, T., Beatty, P.H., Good, A.G., Bi, Y.M., and Rothstein, S.J. (2010). The APETALA-2-like transcription factor OsAP2-39 controls key interactions between abscisic acid and gibberellin in rice. Plos Genetics **6**: e1001098. 10.1371/journal.pgen.1001098.

Yang, Z., Wang, Z., Wang, W., Xie, X., Chai, L., Wang, X., Feng, X., Li, J., Peng, H., Su, Z., et al. (2022). ggComp enables dissection of germplasm resources and construction of a multiscale germplasm network in wheat. Plant Physiol **188**:1950-1965. 10.1093/plphys/kiac029.

Yu, G., Wang, L.G., He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics **31**:2382-2383. 10.1093/bioinformatics/btv145.

Yu, G.C., Wang, L.G., Han, Y.Y., and He, Q.Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. Omics **16**:284-287. 10.1089/omi.2011.0118.

Yuan, J., Sun, H., Wang, Y., Li, L., Chen, S., Jiao, W., Jia, G., Wang, L., Mao, J., Ni, Z., Wang, X., Song, Q. (2022). Open chromatin interaction maps reveal functional regulatory elements and chromatin architecture variations during wheat evolution. Genome Biol **23**:34. doi: 10.1186/s13059-022-02611-3.

Yuan, Z., Persson, S., and Zhang, D. (2020). Molecular and genetic pathways for optimizing spikelet development and grain yield. aBIOTECH **1**:276-292. 10.1007/s42994-020-00026-x.

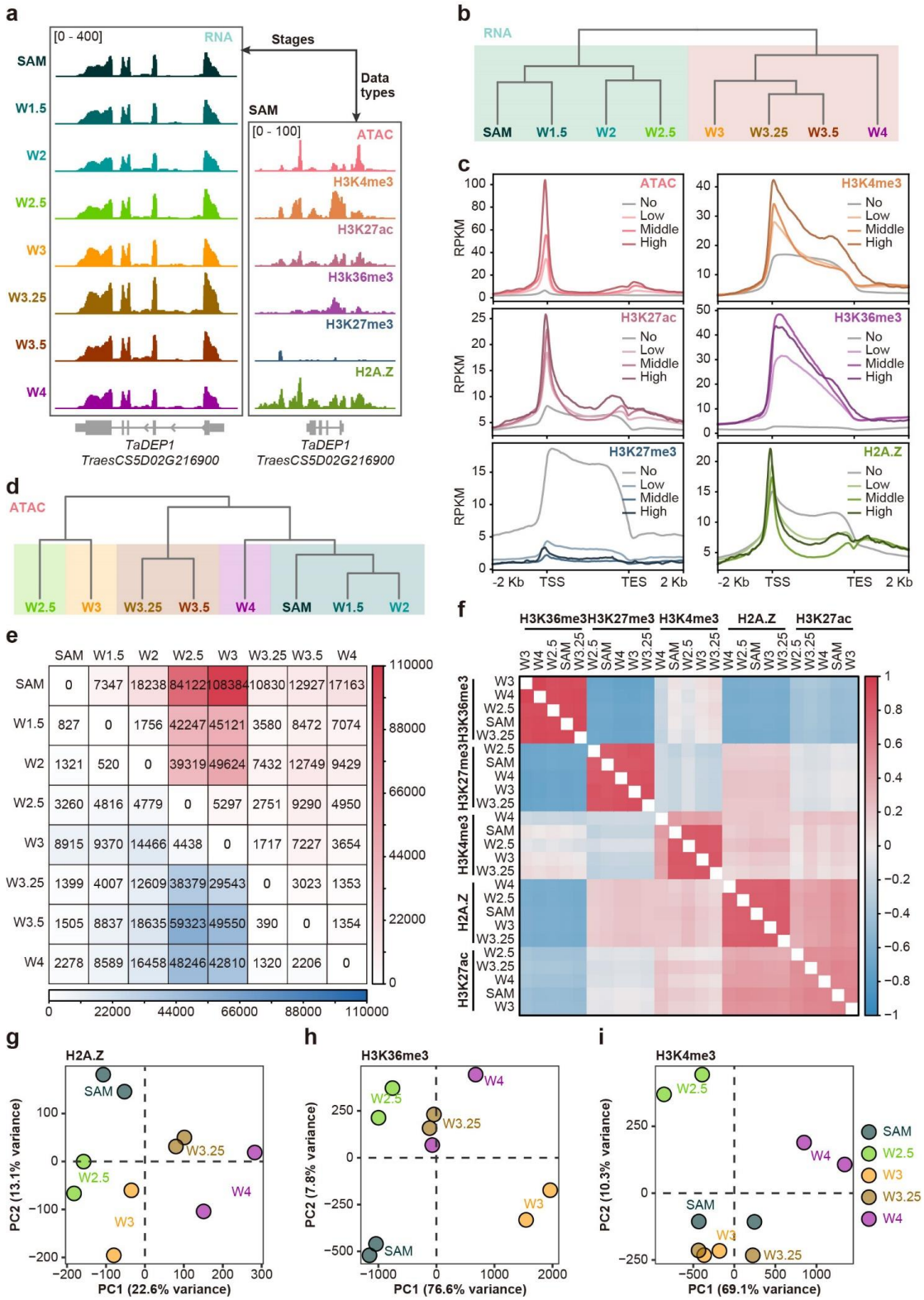
Zhang, B., Liu, X., Xu, W., Chang, J., Li, A., Mao, X., Zhang, X., and Jing, R. (2015). Novel function of a putative *MOC1* ortholog associated with spikelet number per spike in common wheat. Sci Rep **5**:12211. 10.1038/srep12211.

Zhang, X.Y., Jia, H.Y., Li, T., Wu, J.Z., Nagarajan, R., Lei, L., Powers, C., Kan, C.C., Hua, W., Liu, Z.Y., et al. (2022). *TaCol-B5* modifies spike architecture and enhances grain yield in wheat. Science **376**:180-183. 10.1126/science.abm0717.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol **9**:R137. 10.1186/gb-2008-9-9-r137.

Zhao, L., Yang, Y., Chen, J., Lin, X., Zhang, H., Wang, H., Wang, H., Bie, X., Jiang, J., Feng, X., Fu, X., Zhang, X., Du, Z., Xiao, J. (2023). Dynamic chromatin regulatory programs during embryogenesis of hexaploid wheat. Genome Biol **24**:7. 10.1186/s13059-022-02844-2.

Extended Data



Extended Data Fig. 1: Features of various histone modifications.

a, Example browser view at *TaDEP1* locus showing various epigenomic data types at SAM stage and transcriptome levels of different sampling stages.

b, Cluster dendrogram of RNA-seq data showing two distinct development clusters: vegetative cluster (SAM, W1.5, W2 and W2.5 stages), reproductive clusters (W3, W3.25, W3.5 and W4 stages).

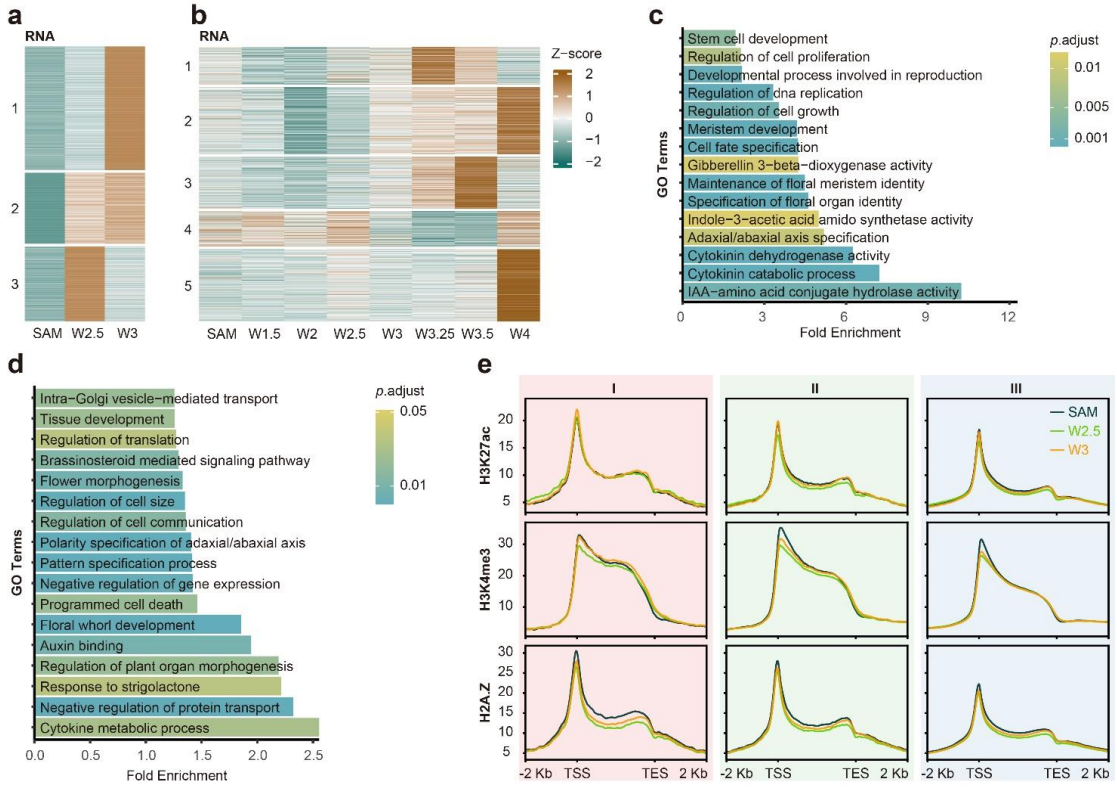
c, Correlation between different types of histone modification profiles and gene expression levels.

d, Cluster dendrogram of ATAC-seq data showing five distinct development clusters: vegetative cluster (SAM, W1.5, W2 stages), flowering transition stage (W2.5 stage), inflorescence initiation (W3 stage), spikelet formation (lemma primordium, W3.25 and floret primordium visible, W3.5), floret organ formation (stamen primordium formation, W4 stage).

e, The matrix of differentially accessible regions (DARs) numbers among developmental stages. The number of decreased and increased chromatin accessibility compared with former stages were represented in the lower-triangle (number in light blue) and upper-triangle panel (number in light red), respectively. A region with $|\log_2(\text{Fold Change})| \geq 1$ and FDR ≤ 0.05 by DiffBind between any two stages was considered as DAR.

f, Pair-wise correlation map among different histone modification profiles. Jaccard index was calculated based on the peaks overlap, and then Pearson correlation scores were generated.

g-i, PCA of H2A.Z (**g**), H3K36me3 (**h**) and H3K4me3 (**i**) samples during spike development. Each dot represents one sample; two bio-replicates are sequenced for each stage.

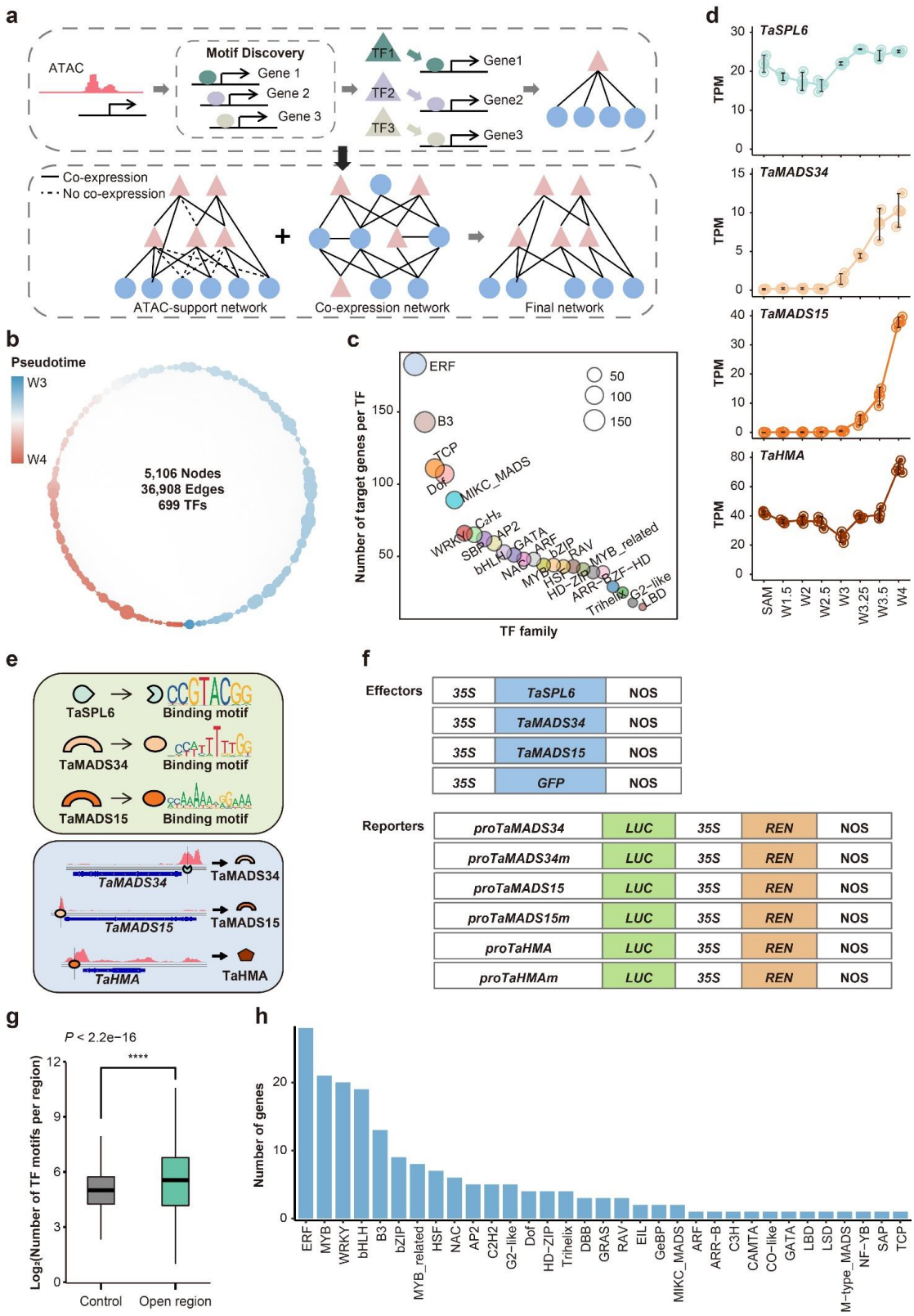


Extended Data Fig. 2: Chromatin landscape dynamics associate with transcriptional change from vegetative-to-reproductive transition.

a-b, Expression pattern of genes up-regulated at W2.5 and W3 stages versus SAM stage (**a**) and genes up-regulated at later stages rather than W2.5 or W3 stage (**b**). Heatmap showing *k*-mean clustering of gene expression.

c-d, GO enrichment analysis of genes in gene set II (**c**), III (**d**) in Fig. 2(**c**).

e, H3K27ac (top), H3K4me3 (middle) and H2A.Z (bottom) levels of genes in gene set I, II, III in Fig. 2 (c) at SAM, W2.5 and W3 stages.



Extended Data Fig. 3: Construction of transcription regulatory network (TRN).

a, Schematic of the strategy for TRNs building (see method for detail).

b, Total TRNs characterized for governing spike architecture formation.

c, Average target genes number of each TF family in the TRN. The size of dot represents number of target genes.

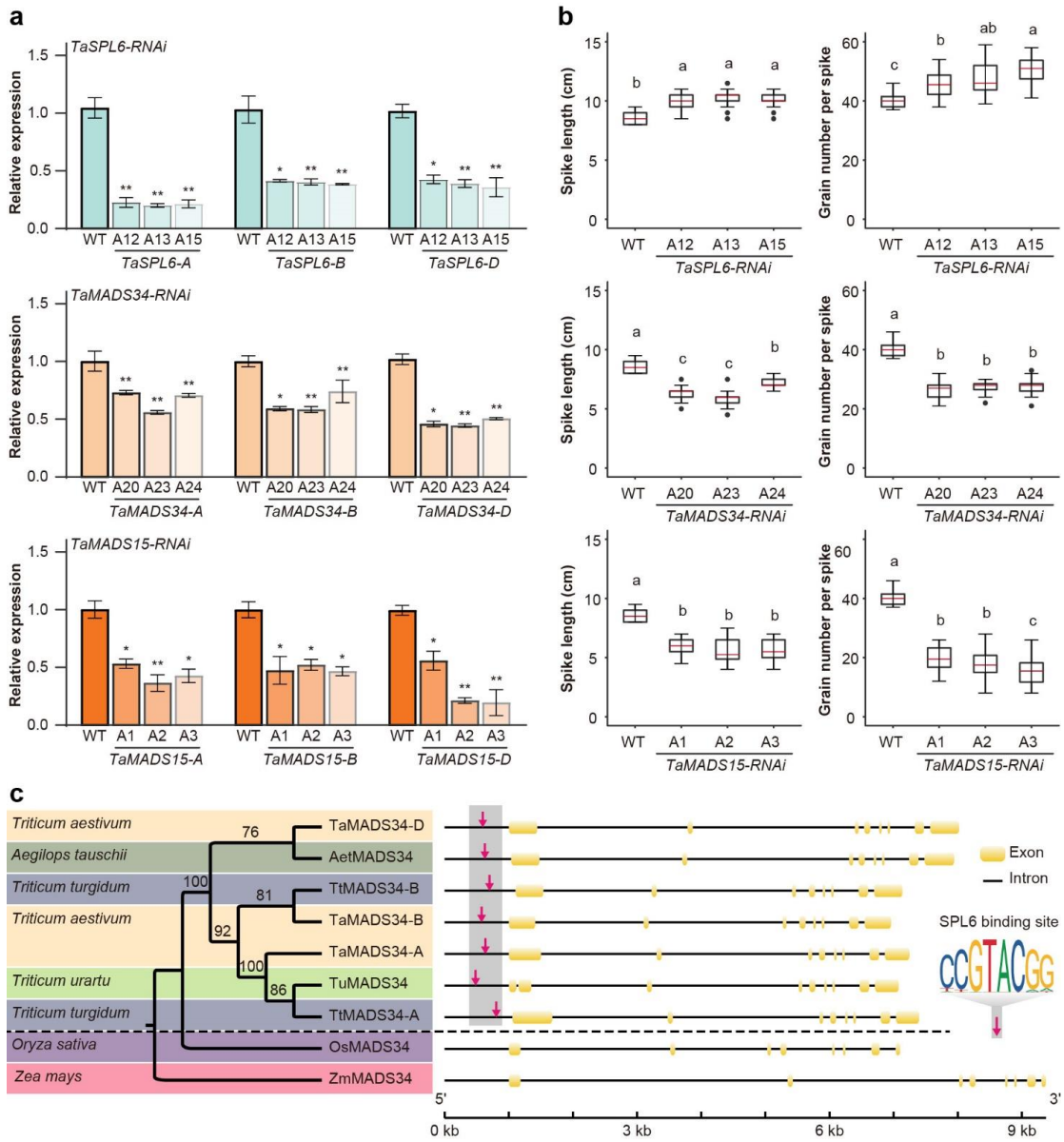
d, Dynamic expression profile of *TaSPL6*, *TaMADS34*, *TaMADS15* and *TaHMA* at different stages during spike development.

e, Presence of different TF binding motifs in the target gene's open chromatin regions from the TaSPL6-TaMADS34-TaMADS15-*TaHMA* regulatory module.

f, Schematic diagram showing the vectors used in the Luciferase reporter assays of Fig. 3 (**f**).

g, Enrichment of TF binding motifs in open chromatin regions. Green and gray refer to open chromatin regions and control regions, respectively. The bars within box plots represent 25th percentiles, medians, and 75th percentiles and Wilcoxon rank sum test was used to determine significant differences.

h, Gene number of corresponding TF families for the 185 potential novel TFs governing spike development.

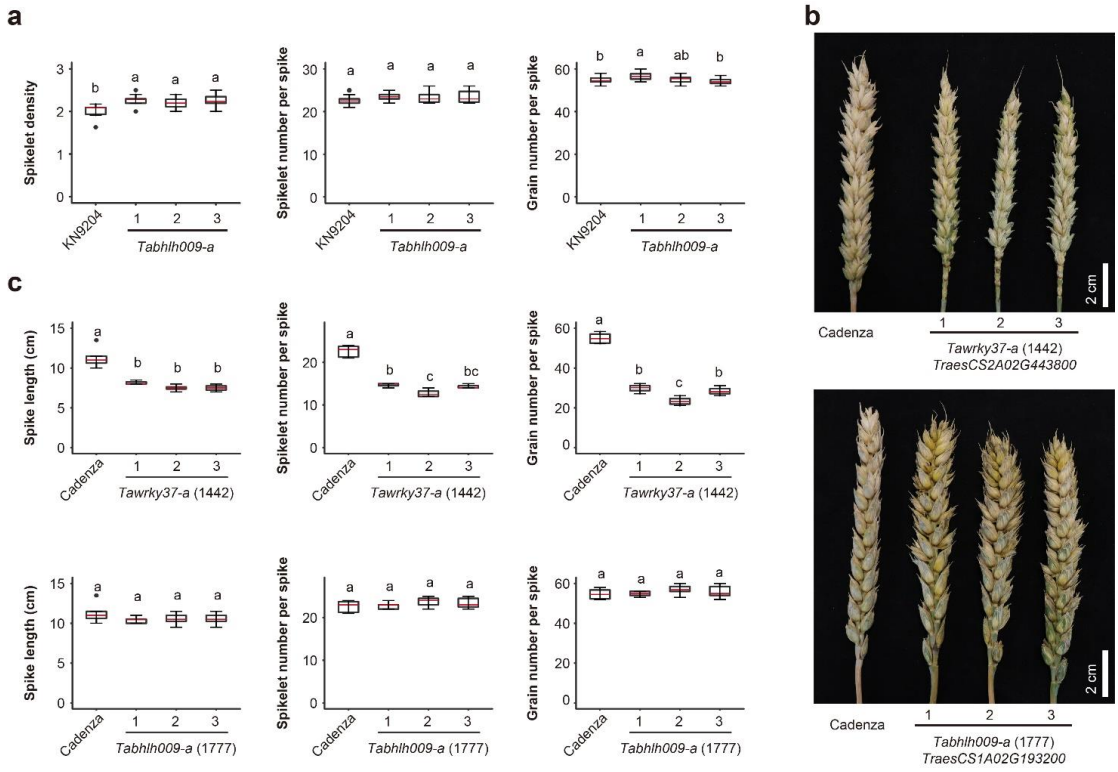


Extended Data Fig. 4: SPL6-MADS34-MADS15-*HMA* module affects spike development in wheat.

a, The knock-down efficiency of target genes in different *TaSPL6-RNAi*, *TaMADS34-RNAi*, *TaMADS15-RNAi* transgenic plants as measured by RT-qPCR. Fielder served as wild-type (WT). Expression level of genes in WT young spike at W3.5 stage is set as 1.0, the relative expression of each gene in RNAi plants young spike at W3.5 stage is shown as average ± SD of three replicates. Student's *t* test was used for the statistical significance (*, $P \leq 0.05$; **, $P \leq 0.01$).

b, Quantification of SL, GNPS between WT, *TaSPL6*-RNAi, *TaMADS34-RNAi* and *TaMADS15-RNAi* transgenic plants. Tukey's HSD was used for the statistical significance. In box plots, the box limits indicate the 25th and 75th percentiles, the whiskers indicate the full range of the data, and the centre line indicates the median. Different letters mean significant difference at $P \leq 0.05$.

c, TaSPL6 binding motif (CCGTACGG) at chromatin accessible region of *MADS34* orthologs in different *Triticum* was consedved, but not in *Oryza sativa* or *Zea mays*. Phylogenetic tree of corresponding species is indicated on the left. Schematic diagram of gene structure and presence or absence of SPL6 binding motif sites is shown on the right.



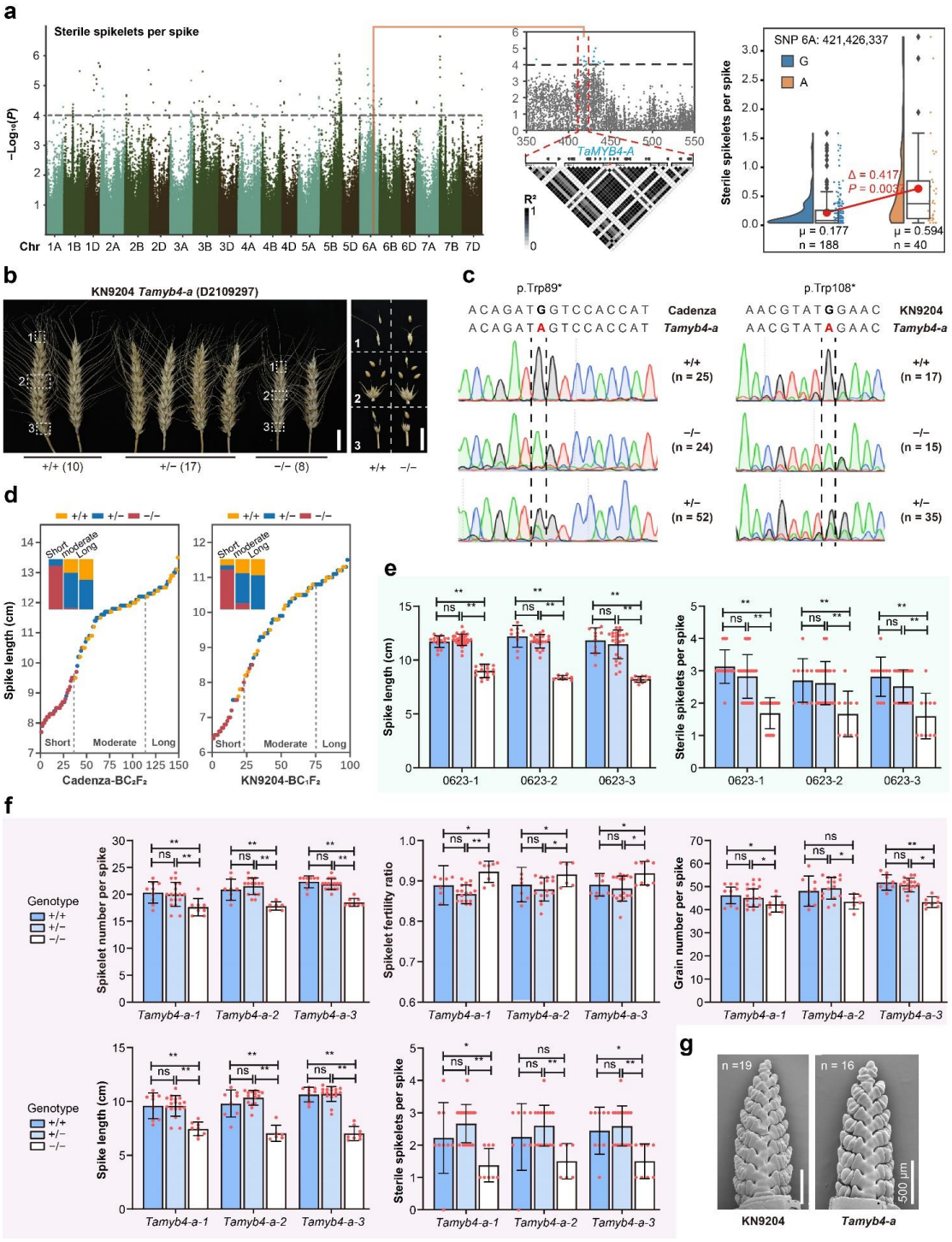
Extended Data Fig. 5: The quantification data of spike phenotype associated with novel factors mutations.

a, Quantification of spike density, SNS, GNPS for represented KN9204 TILLING lines containing mutation of *Tabhlh009-a*. Tukey's HSD was used for the statistical significance. In box plots, the box limits indicate the 25th and 75th percentiles, the whiskers indicate the full range of the data, and the centre line indicates the median.

Different letters mean significant difference at $P \leq 0.05$.

b, The spike developmental defect of *TraesCS2A02G443800* (*TaWRKY37-A*, line 1442) and *TraesCS1A02G193200* (*TabHLH009-A*, line 1777) mutant lines as compared to control Cadenza. Scale bars = 2 cm.

c, Quantification of spike length, SNS, GNPS for represented Cadenza TILLING lines containing mutation of *Tawrky37-a* (1442) and *Tabhlh009-a* (1777). Tukey's HSD was used for the statistical significance. Different letters mean significant difference at $P \leq 0.05$.



Extended Data Fig. 6: The identification and characterization of *TaMYB4-A* mutants.

a, Manhattan plot showing the significant GWAS signals for SSPS on chromosome 6A. Dashed horizontal line indicates the genome-wide significance threshold (-$\log_{10} P$ = 4.0). Local Manhattan plot of SNPs in chr6A: 350-550 Mb was shown, with a LD

heatmap showing the LD of 6 Mb physical interval flanking the peak SNP 6A:421426337. The distribution and comparison of SSPS between haplotypes defined by the peak SNP was shown in a raincloud plot. The bars within raincloud box plots represent 25th percentiles, medians, and 75th percentiles. Mean values of two haplotypes were linked by a red line, and Wilcoxon rank-sum test was used to determine the significance of SSPS between two haplotypes.

b, The spike phenotype of *TaMYB4-A* mutant from the KN9204 TILLING mutant library. The *Tamyb4-a* was backcrossed with wild-type KN9204, and the BC_1F_2 were used to perform the phenotyping. Spikelets corresponding to the numbers and positions (white dotted box) were shown on the right for fertility. Scale bars = 2 cm.

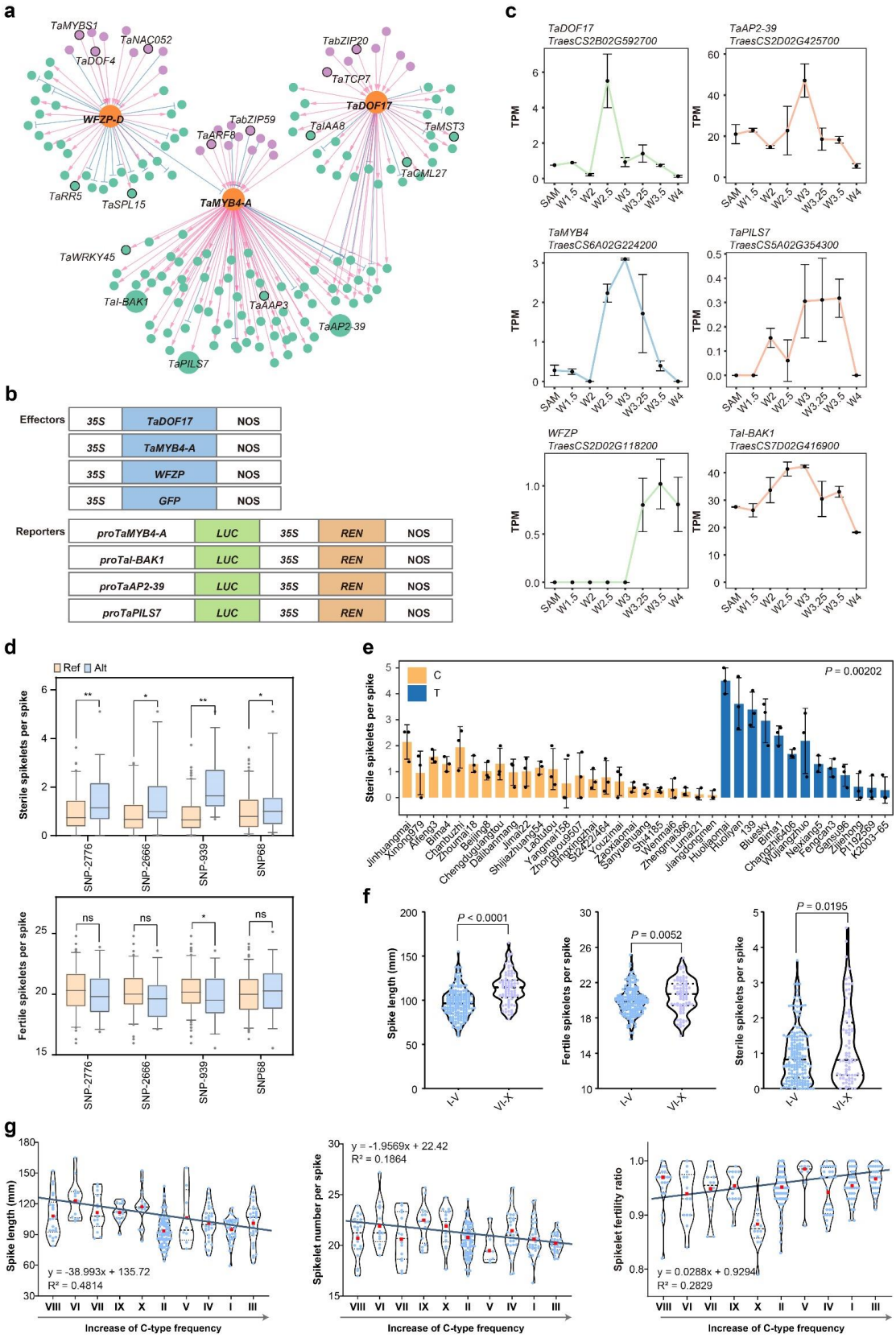
c, Mutant genotypes of *Tamyb4-A* (Cadenza mutant and KN9204 mutant) were identified by Sanger sequencing, and the ratio was consistent with Wild type: heterozygote: mutant = 1:2:1.

d, The homozygous *Tamyb4-a* mutant plants were basically distributed in the smallest quartile of spike length in both BC_2F_2 of Cadenza and BC_1F_2 of KN9204.

e, Statistics comparison of SL and SSPS between Cadenza and *Tamyb4-a* mutant lines. The error bars denote ± SD, and data of each plant were presented in a dot. Student's *t*-test was used for the statistical significance. **, $P \leq 0.01$; ns, no significant difference.

f, Statistics comparison of SNS, spikelet fertility ratio, GNPS, SL, and SSPS, between KN9204 and *Tamyb4-a* mutant lines. The error bars denote ± SD, and data of each plant were presented in a dot. Student's *t*-test was used for the statistical significance. *, $P \leq 0.05$; **, $P \leq 0.01$; ns, no significant difference.

g, SEM of KN9204 and Tamyb4-a spike at W4 stage. Scale bars = 500 μm.



Extended Data Fig. 7: The expression regulation and phenotype analysis based on SNPs of *TaMYB4-A*.

a, The TRN of TaMYB4-A.

b, Schematic diagram showing the vectors used in the luciferase reporter assays of TaMYB4-A regulatory network of Fig. 6 (**d**).

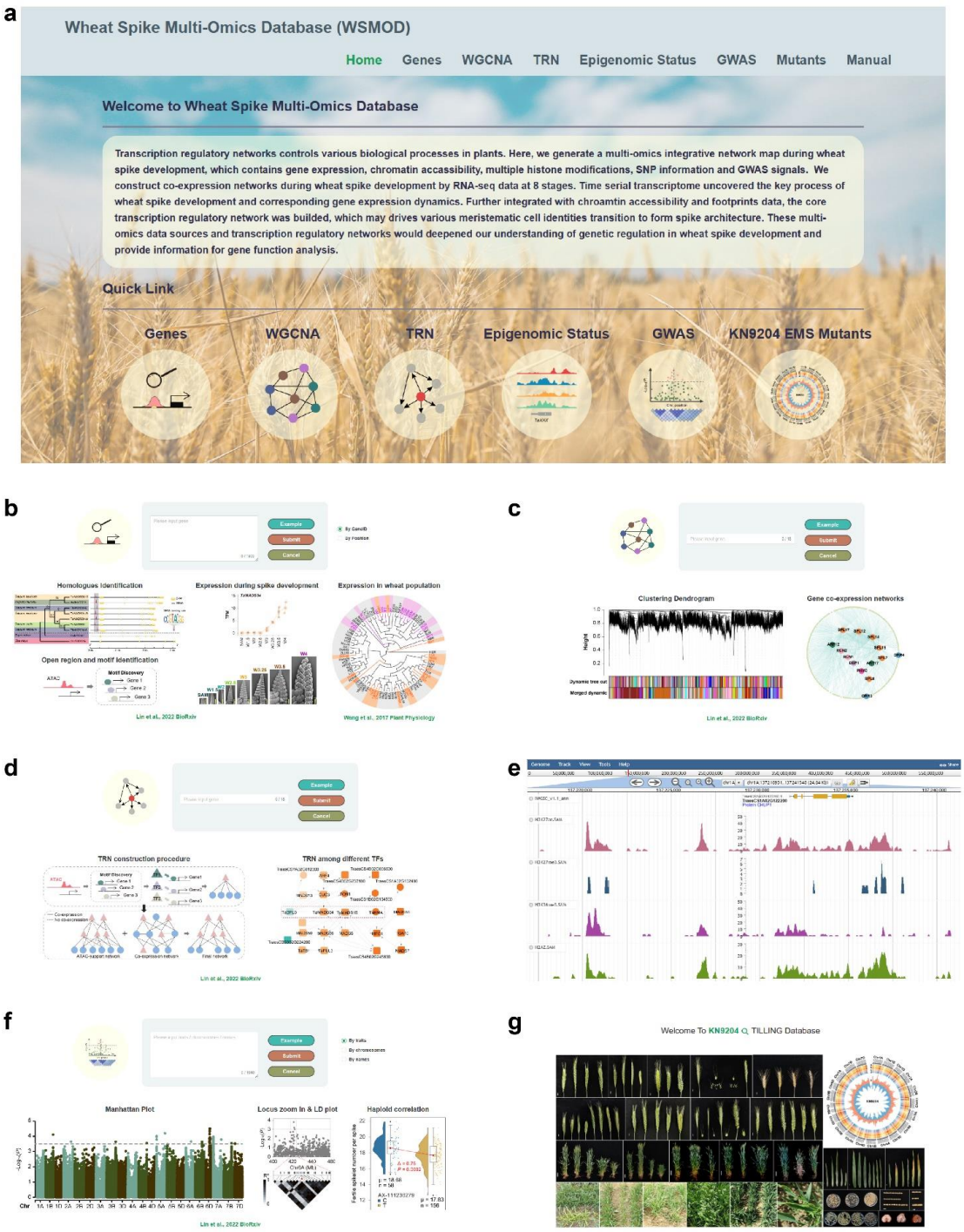
c, The expression level of *TaDOF17*, *TaMYB4-A*, *WFZP*, *TaAP2-39*, *TaI-BAK1*, *TaPILS7*.

d, The comparison of SSPS and FSPS between haplotypes defined by four significantly associated SNPs. Wilcoxon rank-sum test was used to determine the statistical significance; *, $P < 0.05$; **, $P < 0.01$.

e, The SSPS of 37 cultivars with different haplotypes of *TaMYB4-A.* Cultivars with C-type and T-type were marked in yellow and blue, respectively. The SSPS were shown in bar with data from each environment in a dot, and significant difference of cultivars between two haplotypes was determined by Student's *t*-test.

f, The zones I-V has smaller SL, FSPS and SSPS than zones VI-X. Wilcoxon rank-sum test was used to determine the statistical significance between two groups.

g, The distribution of SL, SNS and spikelet fertility ratio in ten ecological zones and their correlation with the frequency of C-type at SNP-939 of TaMYB4-A. From left to right, the C-type frequency in ecological zones increases in turn. Pearson correlation analysis (n = 9, the zone X was excluded due to small wheat accessions) were used to determine the correlation value (R^2).



Extended Data Fig. 8: The overview of Wheat Spike Multi-Omics Database (WSMOD).

a, The homepage of WSMOD.

b-g, The six primary functional modules that comprise WSMOD. (**b**) Search by gene ID function provides gene expression patterns, WGCNA modules, orthologs in rice and

Arabidopsis thaliana, open chromatin regions and footprints, QTL information. (**c**) Co-expression network function provides gene-gene co-expression relationships. (**d**) The wheat spike TRN function provides information of genes upstream TFs, TF binding sites and motifs, as well as downstream target genes. (**e**) Epigenomic status function offers visualization of transcription, chromatin accessibility, various Histone modifications, variations and QTL site. (**f**) GWAS function provides spike-related QTL regions and gene information in regions. (**g**) KN9204 EMS Mutants function provides a quick link to KN9204 TILLING Database in which users can conveniently query and order the EMS mutants of the interested gene.