Systematic mining and genetic characterization of regulatory factors for wheat spike development

Xuelei Lin[1,8], Yongxin Xu[1,2,8], Dongzhi Wang[1,8], Yiman Yang[1,3], Xiaoyu Zhang[1,2], Xiaomin Bie[4], Hongzhe Wang[1], Jiafu Jiang[3], Yiliang Ding[5], Fei Lu[1,2,7], Xueyong Zhang[6], Xiansheng Zhang[4], Xiangdong Fu[1,2], Jun Xiao[1,2,7]*

[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]Shandong Agricultural University, Tai'an, Shandong, 271018, China

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

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

[7]Centre of Excellence for Plant and Microbial Science (CEPAMS), JIC-CAS, 100101, China

[8]These authors contributed equally to this article.

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

Abstract

Spike architecture largely affects grain number embraced, which is a key factor influencing wheat grain yield. Here, we systematically explore the genetic regulation network governing wheat spike development by integration of multi-omic data with population genetics. Transcriptome and epigenome profile of shoot apex at eight developmental stages were generated. Gain-of-chromatin accessibility and changes of H3K27me3 coordinately associate with the progressive transcriptome alteration for flowering. A core transcriptional regulation network (TRN) that likely drives various meristematic cell identities transition to form spike was constructed. Integration of the

TRN and genome-wide association analysis (GWAS), 260 transcription factors (TFs) were identified, including 52 characterized factors in crops, but mostly unstudied. Multi-layer regulatory module among *TaSPL6*, *TaMADS34* and *TaMADS14* suggested by TRN was further validated experimentally. About 85 novel TFs contain high impact mutant lines in KN9204 TILLING library. Of them, 44 TFs with homozygous mutation, including NAC, bHLH, MYB, and WRKY, show altered spike architecture. In particular, TaMYB4-A positively regulates fertile spikelet number likely via regulating hormones homeostasis or signaling, while acting downstream of- and repressed by WFZP. The elite allele of *TaMYB4-A*, with higher expression and more fertile spikelet, was selected during breeding process in China. Collectively, we present invaluable resource of high-confidence regulators and novel strategy for understanding the genetic regulation of wheat spike development systematically.

Key words: Spike architecture, multi-omics, GWAS, TRN, MYB

Introduction

Wheat (*Triticum aestivum*) is a major crop worldwide and provides about 20% of the daily calories and proteins consumed by human population. Yield is a polygenic complex quantitative trait composed of multiple elements including fertile tiller per area, grain number per spike (GNPS) and grain weight (Xiao et al., 2022). GNPS is determinate by the architecture of spike/inflorescence, which involves the number and arrangement of spikelet and fertile florets within spikelet. The spike architecture is shaped by endogenous developmental programs as well as environmental conditions (Gao et al., 2019; Lee et al., 2019; Qi et al., 2019).

During seedling stage, shoot apex meristem (SAM) constitutively produce primordium of leaves, nodes and internodes. The bract leaf primordia first appeared on the growth cone to form a single edge (SR) structure, and then the spikelet meristem (SM) was formed in the upper part of the bract to enter into double ridge (DR) stage, which is also

a symbol for flowering transition. SM gradually differentiates and enlarges from the middle of inflorescence to both ends. After formation of apical SM, the maximum number of spikelets is reached. Following the completion of spikelet differentiation, the lemma primordium (floral primordia differentiation stage) is formed, then it differentiates to produce pistil primordium and pollen sacs (floral organ primordia differentiation stage), and finally enters the booting stage (Bonnett, 1936). Throughout wheat spike development, SM differentiation during and after DR stage determines the spikelet number per spike (SNS) (Dobrovolskaya et al., 2015; Li et al., 2019; Li et al., 2021).

Much progresses have been made in the molecular regulation of cereal inflorescence architecture, including rice, maize and wheat (Gao et al., 2019; Yuan et al., 2020). In general, several steps are critical for the inflorescence structure, such as the duration of inflorescence development, the initiation, arrangement and terminal of spikelets, as well as the fertility of spikelet and floret (Yuan et al., 2020). In wheat, flowering time regulator such as *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, which in turn affects the paired spikelet formation (Boden et al., 2015; Finnegan et al., 2018). *TEOSINTE BRANCHED1* (*TB1*) encoding a TCP transcription factor, controls lateral branching in maize (Doebley et al., 1997). In wheat, TaTB1 interacts with FT1 to control the production of paired spikelets via regulating *VRN1* expression (Dixon et al., 2018). *VERNALIZATION1* (*VRN1*) and the other orthologs of AP1/FUL-like factor, *FRUITFULL2* (*FUL2*), and *FUL3* are required synergistically for initiation and maintenance of the SM identity (Li et al., 2019). Overexpression of *PANICLE PHYTOMER2* (*TaPAP2*), a MADS box TF coding gene in wheat, inhibits SM formation and reduces SNS (Wang et al., 2017). While *WHEAT ORTHOLOG OF APO1* (*WAPO1*) affects SNS by regulating the timing of terminal spikelet formation (Kuzay et al., 2022; Kuzay et al., 2019). *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 spike meristem to terminal spikelets

(Debernardi et al., 2020). Overexpression of *TERMINAL FLOWER 1* (*TaTFL1*)-*2D* delayed the terminal spikelet formation and increased SNS, while *Grain Number Increase 1* (*GNI1*) played the opposite role on determing SNS (Sakuma et al., 2019; Wang et al., 2017).(Wang et al., 2017; Sakuma et al., 2019). *WFZP*, a homologous of rice *FRIZZY PANICLE* (*FZP*), determines the SNS and the floret meristem fate partially by *VRN1* and *HOMEOBOX4* (*TaHOX4*) (Li et al., 2021), as well as by inhibiting the repressor of spikelet formation gene *TaBA1* through epigenetic modification factors TaHDA1 and TaLHP1 (Du et al., 2021). Several SQUAMOSA promoter-binding protein-like (SPL) and MADS-box TFs are involved in the floral organ formation and regulate floret fertility, such as *TaSPL13-B* (Li et al., 2020), *TaSPL14* (Cao et al., 2021), *TaAGL6-A* (Kong et al., 2022). Besides of the key TFs that mediating the meristem identity switch, factors in auxin and cytokinin (CK) hemostasis also play important role in regulation of spike development in wheat, such as *TaCYP78A5* and *TaCKX2.1*, *TaCKX2.2* (Guo et al., 2022; Jablonski et al., 2020; Jablonski et al., 2021; Qi et al., 2019).

The factors identified above are mainly through homologous cloning or forward genetic mapping from mutants or bi-parental population. Recently, population genetics has flourished owning to the sequencing technology innovation (Sehgal and Dreisigacker, 2022). Large-scale GWAS analysis was performed among landmark cultivars with single nucleotide polymorphism (SNP) generated with genotyping-by-sequencing (GBS) (Pang et al., 2020), whole genome resequencing (Hao et al., 2020), or exome capture sequencing (Li et al., 2022a). Various genetic loci were associated with spike developmental related traits, such as spike length (SL), SNS, GNPS, grain setting and spike compactness (Pang et al., 2020; Sun et al., 2017). However, wheat is an allohexaploid of large genome plant with generally triple genes number, large proportion of intergenic regions, more complex sub-genome interaction, and long LD intervals (Hao et al., 2020; Li et al., 2022a; Pang et al., 2020), which greatly reduced the accuracy of association study (Huang and Han, 2014). Other types of information, such as genes expression network provides different filter for identification of key

regulators, as proved by previous studies (Li et al., 2018; VanGessel et al., 2022; Wang et al., 2017).

Here, we carried out a time serial profiling of transcriptome and epigenome at shoot apex in elite wheat cultivar Kenong 9204 (KN9204), to understand how chromatin landscape is coordinated with transcriptome dynamics during the process of spike formation. More importantly, we integrate genes co-expression and regulatory network with GWAS/QTL to identify key factors shaping wheat spike architecture, and further validate and do in-depth analysis to uncover the potential regulators' function.

Results

A transcriptome and chromatin landscape atlas for wheat spike formation

To understand the transcriptional regulation for wheat spike development, shoot tips of eight distinct stages of cultivar KN9204 were sampled, including SAM, elongation stage (EL), SR, DR, spikelet meristem initiation stage (SMI), glume primordium differentiation stage (GPD), floral meristem initiation stage (FMI), floral organ primordia differentiation stage (FOP) (Figure 1A). 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 (Figure 1A and Supplemental Figure 1A) (Buenrostro et al., 2015; Kaya-Okur et al., 2019; Zhao et al., 2022).

In total, 58,875 high-confidence genes were found to express at least in one stage (TPM > 0.5) (Supplemental Table 1). The principal component analysis (PCA) showed that the eight stages could be subdivided into two major categories, as vegetative group including SAM, EL, SR, DR and reproductive group including SMI, GPD, FMI and FOP (Figure 1B). More differential expressed genes (DEGs) between adjacent stages occurs at morphological transition points, such as DR to SMI, SMI to GPD, FMI to FOP, when spikelet meristem, glume primordial and floral organ primordial initiate

147 respectively (Figure 1C and Supplemental Table 2). Stage-specific expressed genes are
148 identified via clustering (Figure 1D). Known factors involved in flowering time
149 regulation and inflorescence development is highlighted from specific cluster. For
150 instance, flowering time gene *TaPPD1* in cluster2 is expressed before floral transition
151 from SAM to DR (Boden et al., 2015; Perez-Gianmarco et al., 2020); *WAPO1*, required
152 for the maintenance of SM activity, is highly expressed at SMI (Kuzay et al., 2022);
153 *TaTFL1* and *WFZP*, involved in terminal spikelet formation and floret fate, are
154 expressed at GPD and FMI (Li et al., 2021; Wang et al., 2017); floral organ regulator
155 *TaAGL6* is highly expressed at FOP (Kong et al., 2022) (Figure 1D). Besides, various
156 members of SPL family TF (*TaSPL9/12*, *TaSPL1*, *TaSPL14*, and *TaSPL8/16*) are
157 highly expressed at different stages (Figure 1D), consistent with recent report (Li et al.,
158 2022b). Such transcriptome pattern correlates with the morphological characteristic of
159 individual stage, indicating our sampling could capture the dynamic gene expression
160 profile during wheat spike development.
161
162 As our previous report (Zhao et al., 2022), accessible chromatin is mainly located in
163 the distal intergenic regions and promoters (Figure 1E). ATAC-seq peak intensity at
164 genic region, in particular around TSS, is positively correlated with gene expression
165 (Supplemental Figure 1B). Unsupervised PCA revealed a continuous trajectory of
166 chromatin accessibility dynamics during wheat spike development (Figure 1F). Eight-
167 stages could partition into five sub-clusters, as vegetative cluster before flowering
168 including SAM, EL, SR, flowering transition stage (DR), inflorescence initiation
169 (SMI), spikelet meristem formation (FMI, GPD) and floret meristem formation (FOP)
170 (Supplemental Figure 1C). A globally increasing of chromatin accessibility at genic
171 regions was observed from vegetative cluster to flowering transition and inflorescence
172 initiation, whereas, declined during spikelet and floret formation (Figure 1G and
173 Supplemental Figure 1D).
174
175 Based on the chromatin accessibility dynamics, SAM, DR, SMI, GPD and FOP stages
176 were chosen for histone modification profiling via Cleavage Under Targets and

Tagmentation (CUT&Tag) (Figure 1A) (Zhao et al., 2022). 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 (Figure 1E and Supplemental Figure 1B). H3K27me3 and H3K36me3 was mutually exclusive for each other, H3K4me3 and H3K27ac was positively correlated, whereas, histone variants H2A.Z was overlapped with both active H3K27ac and H3K4me3 and repressive H3K27me3, but depleted with H3K36me3 (Supplemental Figure 1E and Supplemental Table 3). The PCA and cluster analyses showed that various histone modifications had stage-specific transitions during different developmental stages (Figures 1H and 1I and Supplemental Figures 1F-1H). Of note, H3K27me3 presents continuous trajectory while others are not, suggesting higher correlation with transcriptional profile dynamics (Figure 1H). Indeed, H3K27me3 abundance and chromatin accessibility jointly affect different genes expression pattern during spike development (Figure 1J).

Taken together, we have generated transcriptomic and epigenomic profile of shoot apex at important stages during wheat spike formation, which can facilitate elucidating the transcriptional regulatory insights for shaping wheat inflorescence architecture.

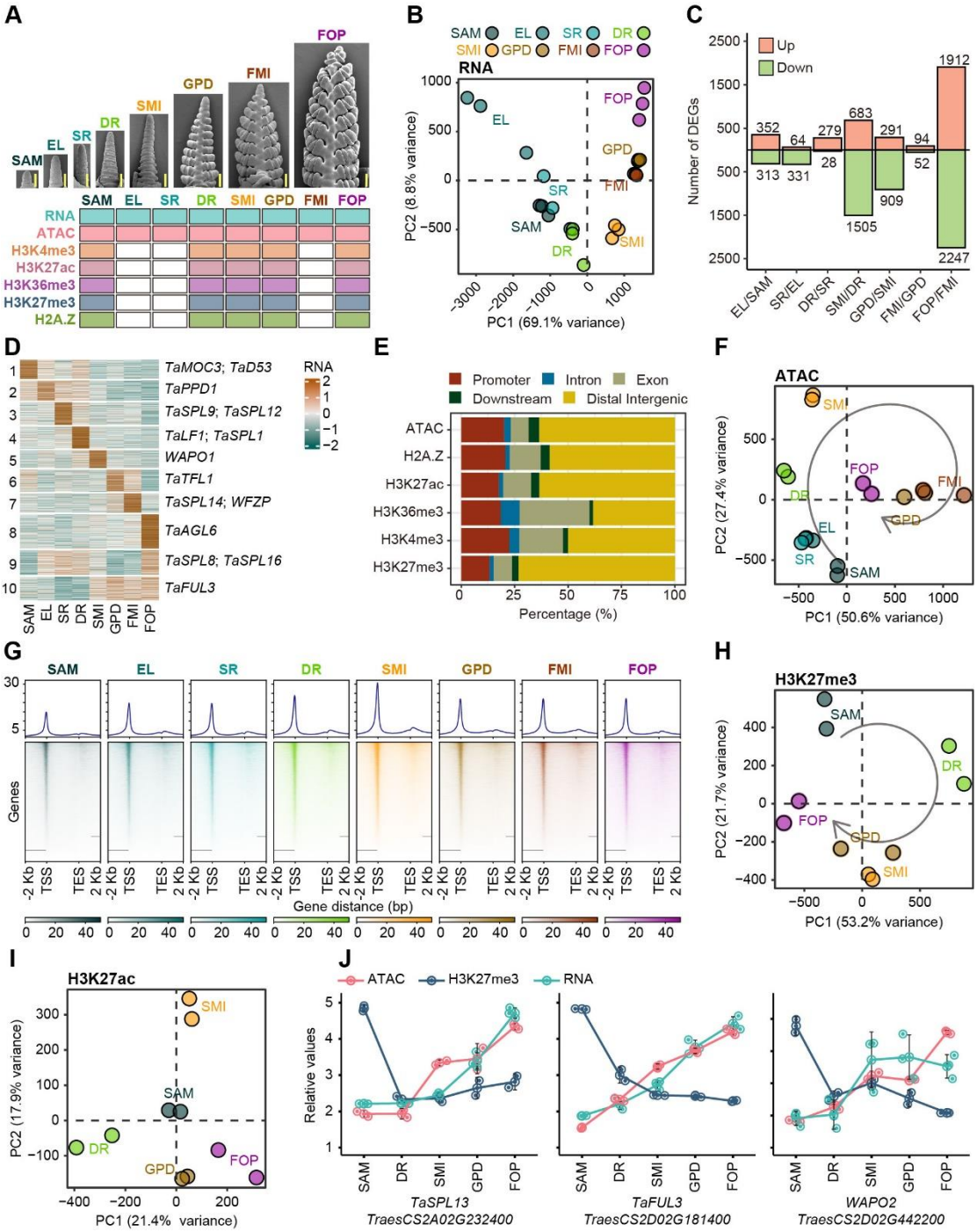


Figure 1. Charting transcriptome and chromatin landscapes during wheat spike development.

(A) Scanning electronic micrographs (SEM) of young spikes in eight developmental stages and experimental design. SAM, shoot apical aeristem; EL, elongation stage; SR, single ridge; DR, double ridge; SMI, spikelet meristem initiation; GPD, glume primordium differentiation; FMI, floral meristem initiation; FOP, floral organ primordium differentiation. Bar= 200 μm.

(B) Principal component analysis (PCA) of transcriptome showing distinct development 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 differential expressed genes (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 are 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.

A permissive chromatin environment facilitates vegetative-to-reproductive transition

As indicated, chromatin accessibility increased globally during vegetative-to-reproductive transition at shoot apex (Figure 1G). We further look at differential accessible promoter regions (DAPR) from SAM to EL, SR, DR till SMI stages (Figure 2A and Supplemental Table 4). In total, 49,153 DAPR was identified, categorized into 6 clusters. Of them, majority of DAPR showed increased accessibility at either DR or SMI stages, including clusters 2,3,4,6 (Figure 2A and Supplemental Table 5). Interestingly, genes of those four clusters show higher positive correlation between chromatin accessibility and transcriptional dynamics (R > 0.5) (Figure 2B). Indeed, genes with increased chromatin accessibility at DR or SMI stages is significantly overlapped with genes of elevated expression at DR or SMI as compared to SAM (1,920 genes within gene set I; Figure 2C, Supplemental Figure 2A and Supplemental Table 6). The degree of increased chromatin accessibility is highly correlated (R = 0.85) with fold change of elevated expression level (Figure 2D). Gene Ontology (GO) term analysis suggests that hormone biosynthesis and signaling, inflorescence meristem

identity, asymmetric cell division are enriched in gene set I (Figure 2E). This highlights the synchronous between gain-of chromatin accessibility and elevated gene expression during vegetative apical meristem to inflorescence meristem transition (Figure 2F).

Besides, open chromatin could set a 'primed' status for gene activation later on (Bonifer et al., 2017; He and Li, 2018). There are 3,435 genes with gained chromatin accessibility at DR or SMI but elevated mRNA level at stages after SMI (Gene set II; Figure 2C, Supplemental Figure 2B and Supplemental Table 6). Those genes are mainly involved in hormones metabolism, floral organ identity, meristem development (Supplemental Figure 2C), which are important for spike formation similar as gene set I. By contrast, large amount of genes with gain-of chromatin accessibility but no change of mRNA level at tested spike developmental stages (Gene set III; Figure 2C and Supplemental Table 6), such as genes participated in polarity specification, cell size regulation, regulation of translation or protein transport (Supplemental Figure 2D). Thus, other chromatin features may also involve in the gene expression regulation of different gene sets.

Indeed, 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 (Figure 2G and Supplemental Figure 2E). In details, genes in set I have relative higher chromatin accessibility at SAM stage, and sharply increased openness at DR and SMI stages, while reduced H3K27me3 level at DR as compared to SAM (Figure 2G). As example, well known flowering promoting gene *TaVRN-B1* shows increased chromatin accessibility and decreased H3K27me3 in genic region, while no change for H3K4me3 from SAM to DR and SMI stages (Figure 2H). Whereas for gene set II, though chromatin accessibility is elevated at SMI and GPD stages, but the relative openness is low. Meanwhile, H3K27me3 level doesn't change from SAM to DR, even increased at SMI stage. The combined effect may result in the 'primed' transcriptional status, which gene expression level increased later with the reduction of H3K27me3, such as *WAP3* (Figures 2G and 2H). For gene set III, lowest chromatin

272 accessibility gained and highest H3K27me3 coverage is observed, which can restrict
273 gene activation (Figure 2G), as the case for *TaEhd2* gene (Figure 2H).
274
275 In summary, from vegetative to reproductive transition, the genome-wide chromatin
276 accessibility increased and H3K27me3 reduced at selective part of genome that create
277 a permissive environment for activation of inflorescence meristem genes and initial the
278 reproductive development of shoot apex.
279

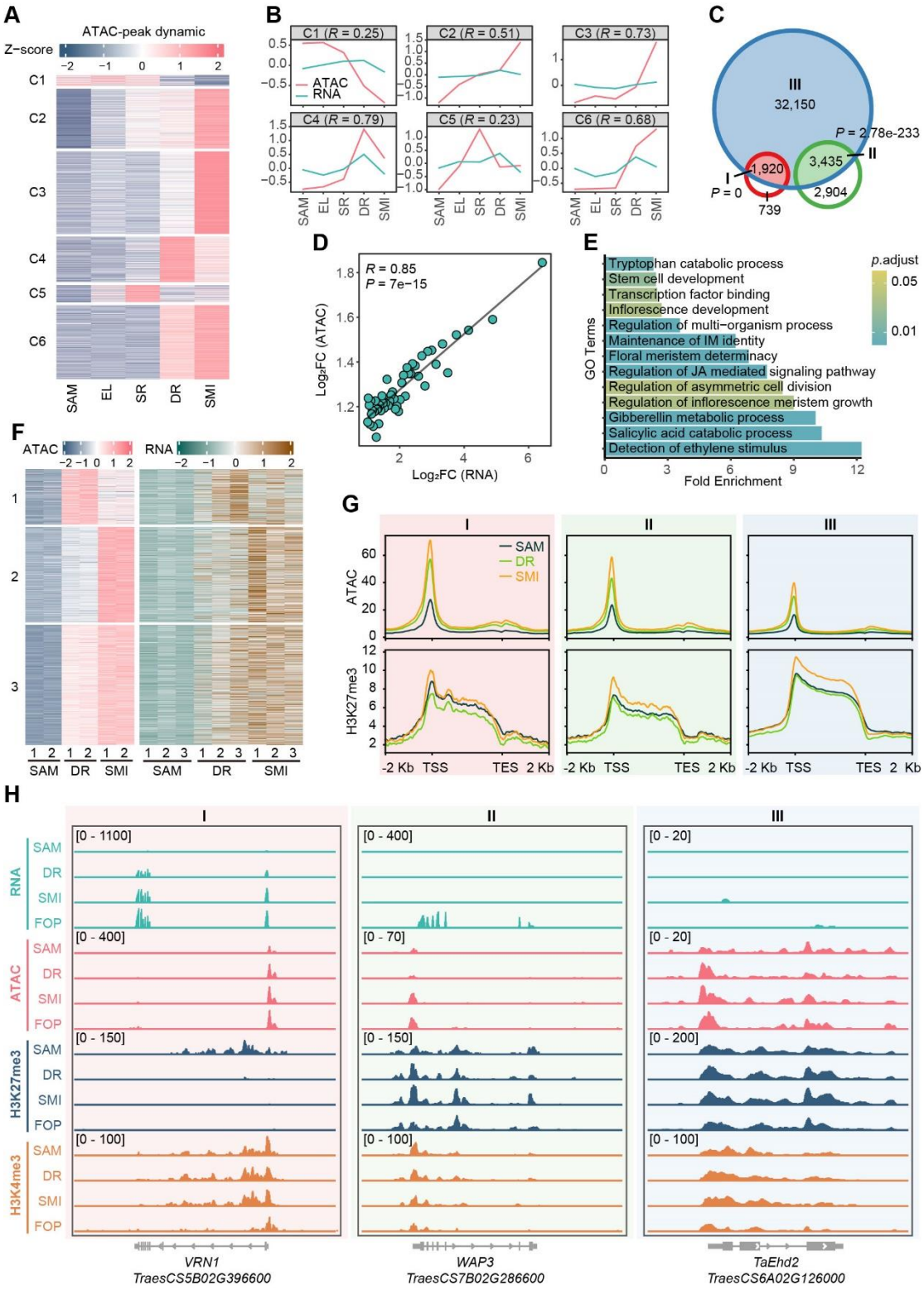


Figure 2. Chromatin accessibility and H3K27me3 dynamics facilitate vegetative-to- reproductive transition of shoot apex.

(A) K-means clustering of differential accessible promoter regions (DAPR) across the samples collected at stage from SAM to SMI.

(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 DR and SMI stage versus SAM stage (red circle) and up-regulated at GPD, FMI, FOP stage versus former stages (green circle). See also Supplemental Figures 2A and 2B. Geneset I showing overlap between red and blue circle genes; geneset II showing overlap between red and green circle; while rest genes from red circle belong to geneset III. Hypergeometric test was used to calculate P-value for the enrichment of geneset I and II.

(D) Correlation of gene expression elevation (x-axis) and chromatin accessibility gain (y-axis) from SAM to DR or SMI based on 1920 genes in geneset I in (C). Genes were ranked by fold change of elevated expression and separated into 50 bins.

(E) GO enrichment analysis of 1920 genes in gene set I.

(F) Synchronous pattern of gene set I between the gain of chromatin accessibility and elevation of gene expression from SAM to DR or SMI stage. Individual bio-replicate are shown separately.

(G) Chromatin accessibility (top) and H3K27me3 (bottom) levels of genes in gene set I, II, III at SAM, DR and SMI stages.

(H) IGV browser 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).

Gene co-expression and transcriptional regulatory network for spike architecture formation

Following the initiation of reproductive growth (DR), inflorescence undergoes several processes, including SMI, GPD, FMI, and FOP to form the spike architecture (Figure 1A and Figure 3A). This is critical for determining SNS, contributing to grain yield. We thus analyzed the genes co-expression and regulatory network to highlight the potential factors that may govern the process.

We calculated the pseudo time of stages covering SMI to FOP based on PCA distance (Figure 3A). GPD and FMI stages are close to each other, while distinct from SMI and FOP. A total of 8,200 DEG was identified in the process of spike architecture formation (any DEG among SMI, GPD, FMI, FOP), which could be clustered into 6 categories (Figure 3B and Supplemental Table 7). Floral meristem determinacy and various hormone metabolic genes are highly expressed at SMI; hormone signaling, polarity establishment genes are active at GPD and FMI; while floral organ identity, polarity

specification genes are elevated at FOP (Figures 3B and 3C). 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 SMI stage. For GPD and FMI stages, NF-Y and SBP TFs are abundantly expressed within clusters 2 and 4. Whereas, MADS box TFs are outstanding at FOP stage of cluster 5 (Figures 3B and 3C). Such expression pattern fits well with the spatiotemporal switch of meristem identities in the context of hormone regulation (Kellogg, 2022; Koppolu et al., 2022), and indicates potential importance of individual TFs family in driving the transcriptional network that governing spike formation.

Accessible chromatin regions provide docking sites for TF binding. Binding motifs of AP2/ERF, bHLH and SPL showed greatest accessibility variability during SMI to FOP (Figure 3D), suggesting the potential importance of those TFs. Based on the TF-motif recognition pairs, in combination with the co-expression pattern of TFs and downstream targets, we are able to construct a hierarchy TRN by integrating the multi-stages ATAC-seq and RNA-seq data (Supplemental Figure 3A). Totally, 36,908 pairs of TFs-target genes interactions were identified, covering a total of 5,106 genes, of which 4,916 pairs were regulated between TFs (Supplemental Figure 3B). A clear sequential regulatory relation among time-serial expressed gene clusters from SMI to FMI is observed as C6-C3-C2-C4 (Figure 3E). Whereas, FOP specific expressed genes in C1 and C5 are separated and regulated each other (Figure 3E), indicating different transcriptional regulatory networks between spikelet and floret development. TFs of ERF, B3, TCP, DOF and MIKC-MADS may play core regulatory role in this network as more targets being identified (Figure 3Fand Supplemental Figure 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., 2019), *TaFUL3* (Li et al., 2019) (Figure 3G), indicating the capability of our TRN in capturing important regulators.

To validate the prediction power of the hierarchy of transcriptional regulation among TFs, we extract a small module that contains factors being functionally studied individually in other crops but without known of regulatory relations, such as *SPL6* (Wang et al., 2018), *MADS34* (Lin et al., 2014; Meng et al., 2017; Zhu et al., 2022), *MADS15* (Wang et al., 2010; Wu et al., 2017) (Figure 3G). From their temporally expression pattern (Supplemental Figure 3D) and the presence of specific motifs in the open chromatin of each gene (Supplemental Figure 3E), we speculate a positive transcriptional regulation 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 (Figure 3H and Supplemental Figure 3F). As well, such regulation circuit is dependent on the TF-motif specific recognition (Figure 3H and Supplemental Figure 3F).

Thus, the TRN constructed could well predict the potential important regulatory factors and their transcriptional regulatory relationships during the process of spike development.

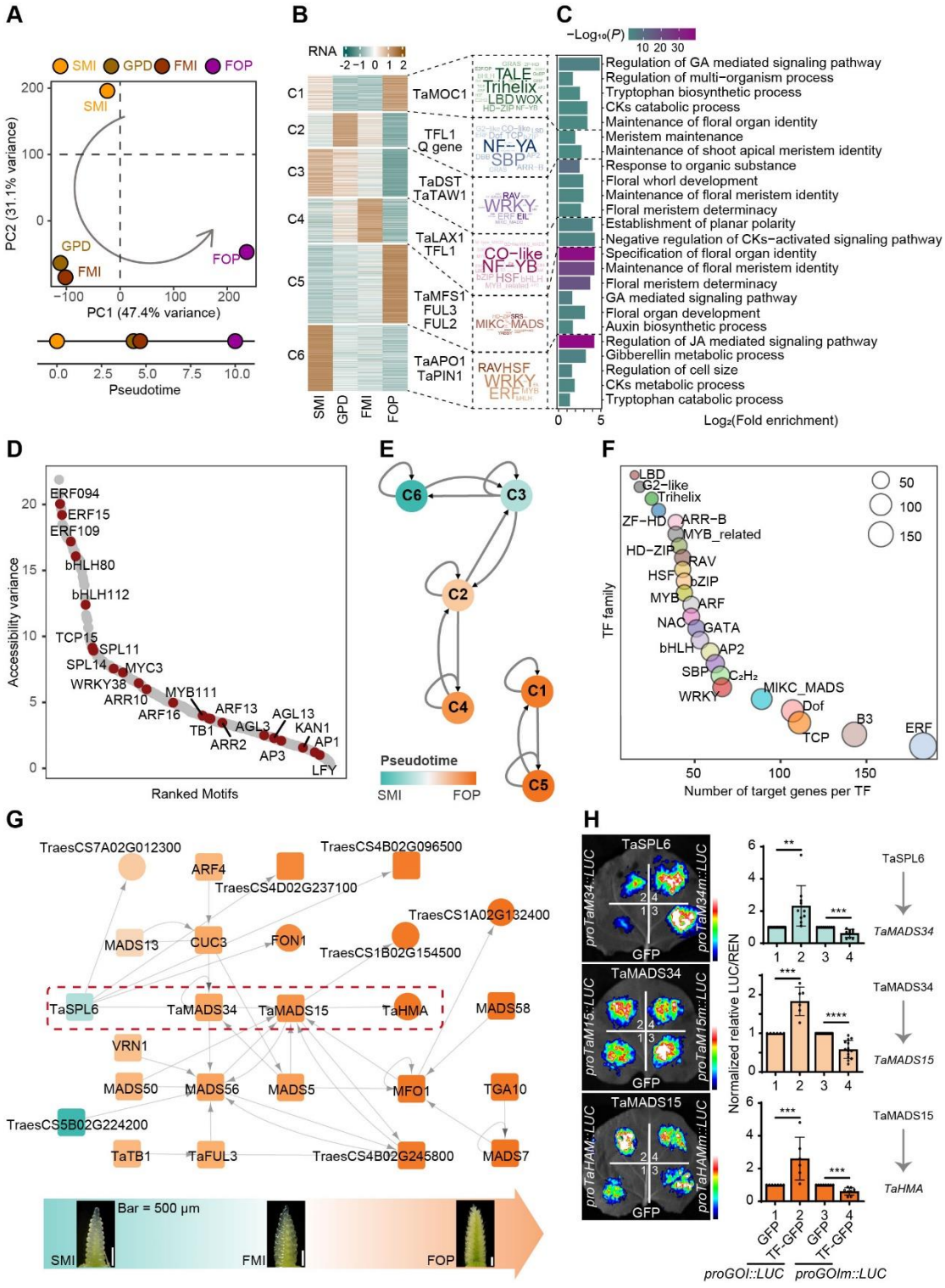


Figure 3. Transcriptional regulatory network (TRN) governing spike architecture formation.

(A) Principal-components plots of gene expression from SMI to FOP 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-mean clustering of DEGs from SMI to FOP stage and representative genes in

each cluster.

(C) TF family enrichment and GO enrichment analysis within each cluster in B.

(D) Motif accessibility variance from SMI to FOP stage.

(E) Hierarchical transcriptional regulations of sequentially expressed gene clusters from (B). *P*-value was determined by the hypergeometric test.

(F) Average target genes number of each TF family in the TRN. The size of dot represents number of target genes.

(G) TRN for represented key TFs participated in spike development, from SMI to FOP stages. Genes were roughly ranked by the expression timing from left to right as indicated by 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 tested in (H).

(H) Luciferase reporter assays validation of transcriptional regulation among representative TF-target pairs as indicated. 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 (**, $p \leq 0.01$; ***, $p \leq 0.001$).

Integration of multi-data to systematically screen regulators shaping spike development

The open region of chromatin is crucial for the establishment of transcriptional regulatory relationships (Klemm et al., 2019). Indeed, more binding motifs of TFs are significantly present in the open region of chromatin (Figure 4A), with lower DNA variation at the TFs binding footprints (Figure 4B) (Zhou et al., 2020). Furthermore, the open chromatin regions contain significantly higher GWAS signals with related to spike morphology traits from previous publication (Figure 4C and Supplemental Table 8). The QTL frequency of open chromatin regions quintupling that of the whole genome totally, with analogical results obtained for concrete spike related traits, such as spike length (SL), SNS, GNPS (Figure 4D). Take the fertile spikelet number per spike (FSPS) as an example, a total of 153 SNPs passed the significance threshold from GWAS analysis and considered as significant associated loci (Figure 4F and Supplemental Table 9). We enlarged the candidate region to 1 Mb centered on the GWAS signal peak, 2,916 genes were identified as candidate genes. Among the 2,916 genes, 1762 genes (60.43%, C1 and C4) present the highest openness of chromatin accessibility at DR or SMI, when spikelet primordia were undergoing initiation/ differentiation and considered to be critical for determination of final SNS (Figure 4E and Supplemental

Figure 4A). In addition, 438 genes (15.02%, C6) exhibit to most open at FMI, which is crucial to the floret development and seed setting (Supplemental Table 9). Analogical results were observed for SL, spikelet density (SD), and SNS (Supplemental Figure 4B). This result indicated that large proportion of GWAS candidate genes possess high chromatin openness at key developmental stage relating to the corresponding trait.

We further integrate the spike traits-associated genetic regions from population genetics study (Supplemental Table 8) with key regulators identified for floral transition, giving the importance of regulating duration of inflorescence meristem development (Boden *et al.*, 2015; Finnegan et al., 2018), as well as the regulators in the TRN network that potentially governing the formation of spike (Figure 4F). With focusing on TFs coding genes, we got 776 candidates for further screening (Figure 4F and Supplemental Table 10). Among them, more than half (404 TFs) are within the spike morphological traits-associated genetic regions (Figure 4Fand Supplemental Table 10). We further narrow-down the candidates to 260 TFs by looking for presence of SNP in the open chromatin region of those TFs (Figure 4F and Supplemental Table 10), which is likely to affect the transcriptional regulatory circuit and rewire the TRN of spike development.

Of the candidate TFs, we categorize them into three classes, termed as ‘researched’ for functional studied in wheat and ‘conserved’ as orthologue being studied in other species as well as ‘novel’ for functional unknown for wheat and other crops (Figure 4G). Among the 260 TFs, 11 factors are ‘researched’ in wheat, including *VRN1* (Li et al., 2019), *TaFUL3* (Li et al., 2019), *TaAGL6* (Kong et al., 2022), *TaSEP3* (Zhang et al., 2021). Another 41 factors are ‘conserved’, including *MOC1* (Zhang et al., 2015), *PAP2* (Kobayashi et al., 2010; Kobayashi et al., 2012), *MADS14* (Wu et al., 2017), etc. (Figure 4G and Supplemental Table 11). Of note, majority of candidates identified are ‘novel’ TFs, as many as 208, without study in wheat or other crops. These TFs are enriched in ERF, WRKY, bHLH, MYB, B3 families (Figure 4H and Supplemental Table 11). Thus, by integrating of multi-dimensional data, we have identified potential key regulators for spike development.

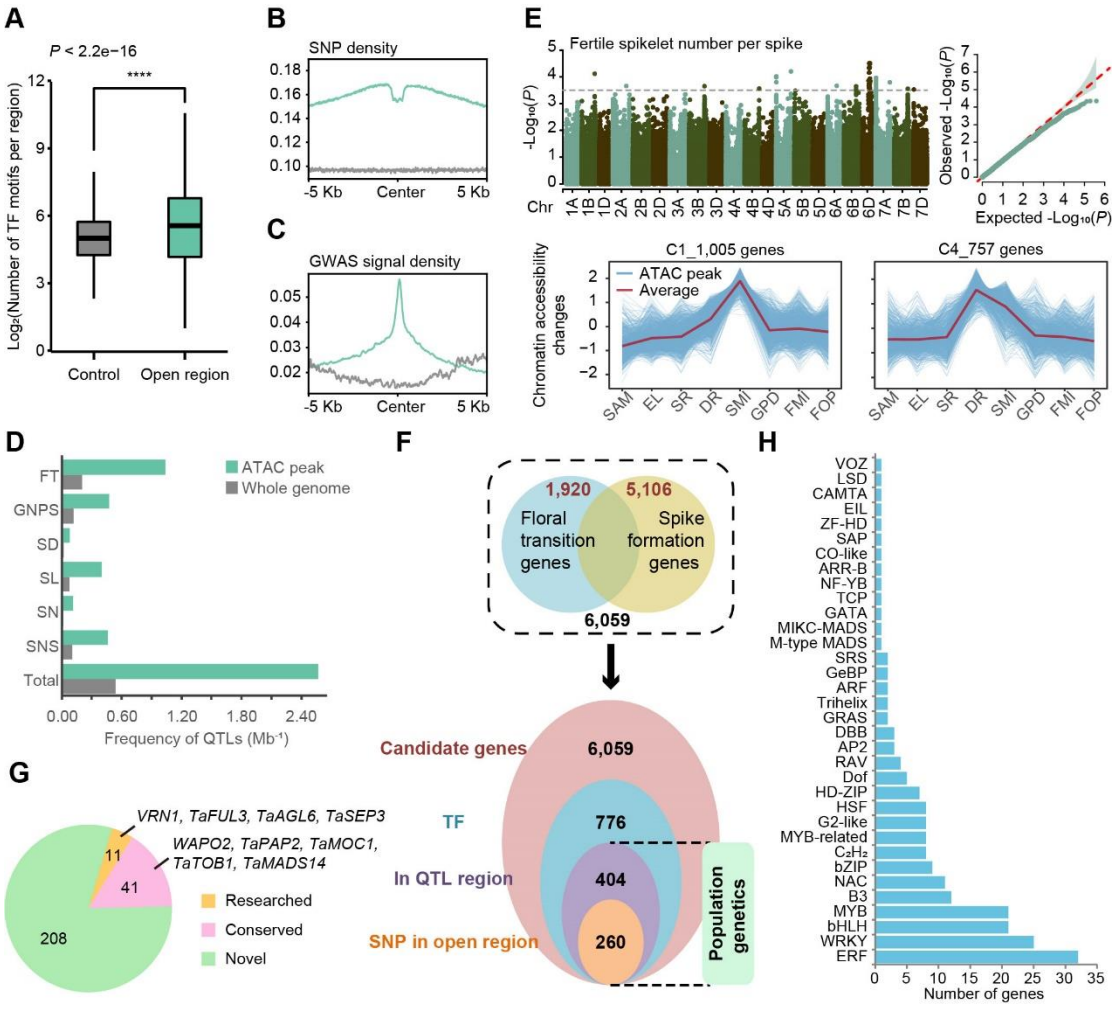


Figure 4. Screening for key TFs in spike development by integration of multi-dimensional data.

(A-C) Enrichment of TF binding motifs (A), SNP density (B) and GWAS signals of multi-traits (C) in open chromatin regions. Green and gray refer to open chromatin regions and control regions, respectively.

(D) The QTL frequency of spike morphology related traits in open chromatin regions (light green bar) and whole genome (gray bar). The QTL frequency of each trait were determined with QTL numbers divided by physical length of the open chromatin regions and whole genome, respectively.

(E) The chromatin openness patterns of fertile spikelet number per spike (FSPS) trait correlated candidate genes. Manhattan plot (up-left) and Quantile-Quantile plot (up-right) showing the significant associated signals of FSPS using a mixed linear model (MLM). The chromatin accessibility change showed GWAS signal associated genes specifically open at DR and SMI (low), with average trends highlighted in red. The y-axis represents z-scaled chromatin accessibility across developmental stages.

(F) Schematic of the strategy used for key factors identification for shaping spike

architecture.

(G) Pie chart summary of candidate factors for spike development in each category as indicated. Represented genes' name were listed.

(H) Gene number of corresponding TF families for the 208 potential novel TFs governing spike development.

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

Among the 'conserved' factors, we again found *TaSPL6*, *TaMADS34*, *TaMADS15* (Figure 4G), as showing up in the TRN governing spike development (Figures 3G and 3H). Thus, we wonder whether these factors are indeed involved in shaping spike architecture in wheat.

In situ hybridization finely characterize the spatial-temporal expression profile of these factors (Figure 5A). *TaSPL6* was weakly expressed in the SAM, fixedly expressed in the IM tip, spikelet primordia and floret meristem. *TaMADS34* and *TaMADS15* were both expressed throughout the SAM, SMI and FOD stages and highly expressed in the spikelet and floral primordia. *TaHMA*, a downstream target of TaMADS15, also expressed in the spikelet and floral primordia. 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 (Supplemental Figure 5A). As expected, *TaMADS34-RNAi* and *TaMADS15-RNAi* plants exhibit similar phenotypes, with shorter SL, decreased SNS and GNPS (Figures 5B-5D and Supplemental Figure 5B). *TaSPL6-RNAi* plants also show altered spike developmental phenotype, but with longer SL, increased SNS and GNPS (Figures 5B-5D and Supplemental Figure 5B). This is likely because that TaSPL6 is on the top layer of hierarchy TRN, it can regulate multi-targets in addition to *TaMADS34* (Figure 3G). Consistent with the LUC reporter assay *in vitro* (Figure 3H), in *TaSPL6-RNAi* transgenic wheat, we observed significant reduced expression of *TaMADS34* (Figure 5E), further indicating that TaSPL6 positively regulate *TaMADS34*. Similar expression pattern was observed for supporting

that TaMADS34 activated *TaMADS15*, and TaMADS15 promoted *TaHMA* (Figure 5E). 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 (Figure 5F). 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 (Figure 5F). Similarly, in the promoter region of *MADS34* homologous genes, the binding motif of SPL6 (CCGTACGG) exists (Supplemental Figure 5C). 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 (Figure 5F and Supplemental Figure 5C).

Thus, the ‘conserved’ factors we identified were indeed involved in regulating spike development in wheat, and the hierarchy transcriptional regulatory circuit is likely maintained within *Triticum* tribe.

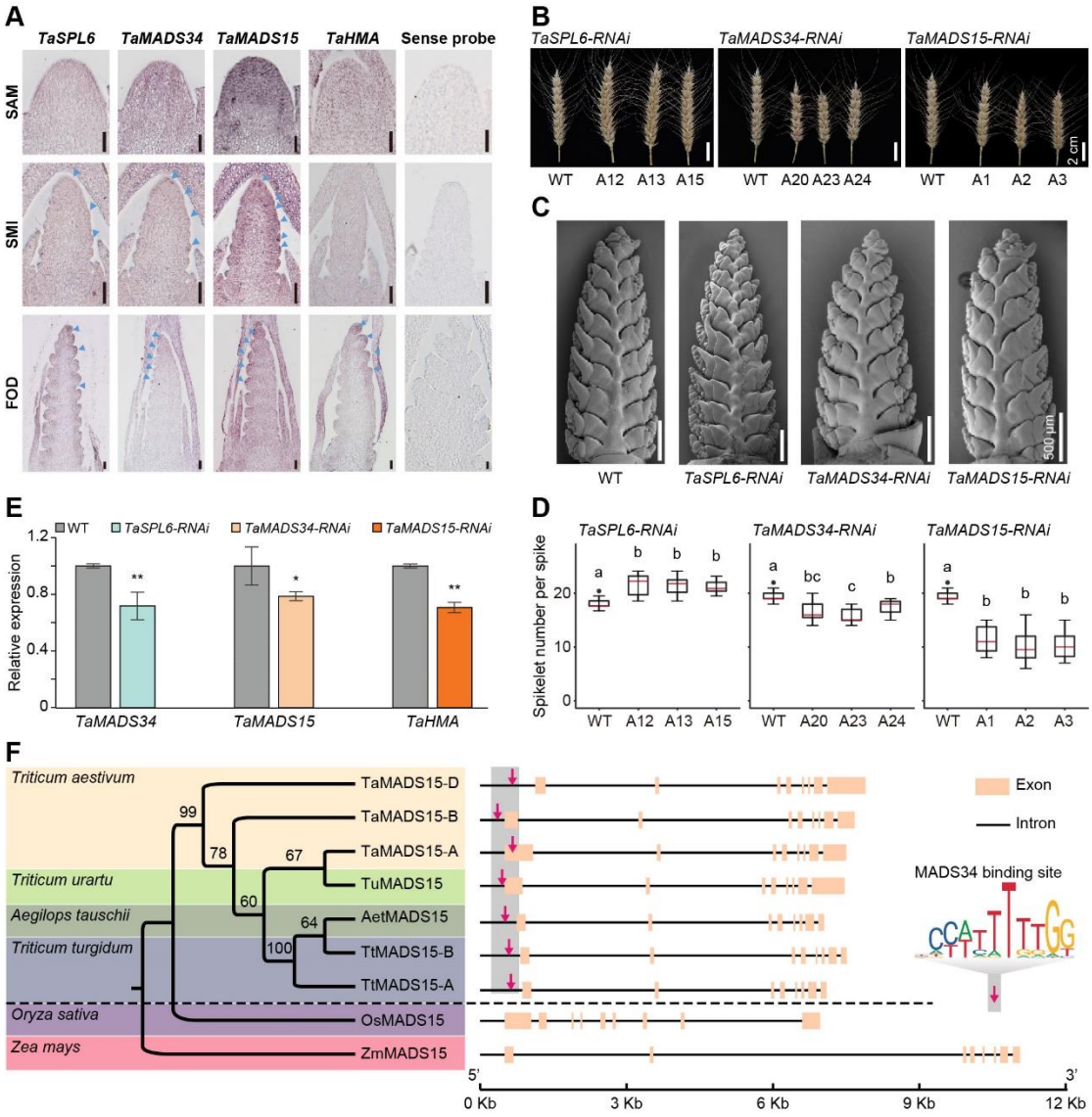


Figure 5. SPL6-MADS34-MADS15-*HMA* module regulates spike development in wheat.

(A) Spatiotemporal expression patterns of *TaSPL6*, *TaMADS34*, *TaMADS15* and *TaHMA* at different spike developmental stages, as indicated by *in situ* hybridization. Sense probe is used as negative control. 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) Scanning electron micrographs (SEM) of young spikes from WT and RNAi transgenic plants of *TaSPL6*, *TaMADS34*, *TaMADS15* at DR, GPD and FOP stages. Scale bars = 500 μm.

(D) Quantification of spikelet number per spike (SNS) between WT, *TaSPL6*-RNAi, *TaMADS34-RNAi* and *TaMADS15-RNAi* transgenic plants. Two-tailed Student's t-tests. 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 $p < 0.01$.

(E) The expression level of *TaMADS34, TaMADS14, TaHMA* in 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 is shown on the right.

Verification 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 meta TILLING mutant lines in Kronos (Krasileva et al., 2017; Uauy et al., 2009), Cadenza (Krasileva et al., 2017) and KN9204 (we generated and did exome sequenced, unpublished), of which the mutated sites were identified by whole exome sequencing.

Of the 208 novel TFs, 131 TFs were found to have at least one mutant line that containing loss-of-function mutation (Figure 6A and Supplemental Table 11). Among the 85 TFs with homozygous TILLING mutant lines, 44 TFs (51.76%) exhibited spike developmental defects within three major types, including flowering time difference (Type I, n=11), degeneration of spikelet or floret (Type II, n=24), and altered SL or SNS (Type III, n=21) (Figure 6B and Supplemental Table 11). Representative mutant lines of different types were shown (Figures 6C and 6E and Supplemental Figure 6). For instance, *TraesCS3A02G433300* (*TaCAMTA-A*) mutant showed an early flowering phenotype with about 2 weeks earlier than the control line; *TraesCS2A02G443800* (*TaWRKY37-A*) mutant showed more degenerated spikelet number per spike (DSNS) at basal part of inflorescence, resulted in decreased SNS/GNPS; *TraesCS1A02G193200* (*TabHLH009-A*) mutant showed reduced spike length and increased spikelet density (Figures 6B-6D and Supplemental Figure 6A). The similar phenotype of *TaWRKY37-A* and *TabHLH009-A* genes also exhibit in the Cadenza mutant (Supplemental Figures 6B and 6C). The loss-of-function mutant of *TraesCS5A02G500500* (*TaNAC22-A*) in

562 the tetraploid wheat Kronos (*Triticun turgidum*) exhibits increased floret number per
563 spikelet (about 10 florets in a spikelet) due to non-terminated floral primordium
564 differentiation, and most of the florets were abortive which result in decreased GNPS
565 (Figures 6E and 6F). We further confirmed the spatiotemporal expression pattern of
566 those genes by *in situ* hybridization (Figure 6G). Consistent with their morphologic
567 defects, *TaWRKY37-A*, *TabHLH009-A* and *TaNAC22-A* genes are all expressed at the
568 spikelet initiation region and the spikelet/floral primordia; while *TaWRKY37-A* is
569 highly expressed in the IM tip and the spikelet and floral meristem at the base of spike,
570 *TabHLH009-A* is highly expressed in the spikelet and floral meristem at the upper
571 region of spike, and *TaNAC22-A* has significant expression in the SAM and floral
572 primordia (Figure 6G).
573
574 Taking the advantage of available TILLING mutant lines, we have proved that novel
575 TFs identified by our integration strategy played important role in regulating spike
576 development. The strategy used for screening novel factors is efficient and the
577 identified factors were worth for in-depth study.
578

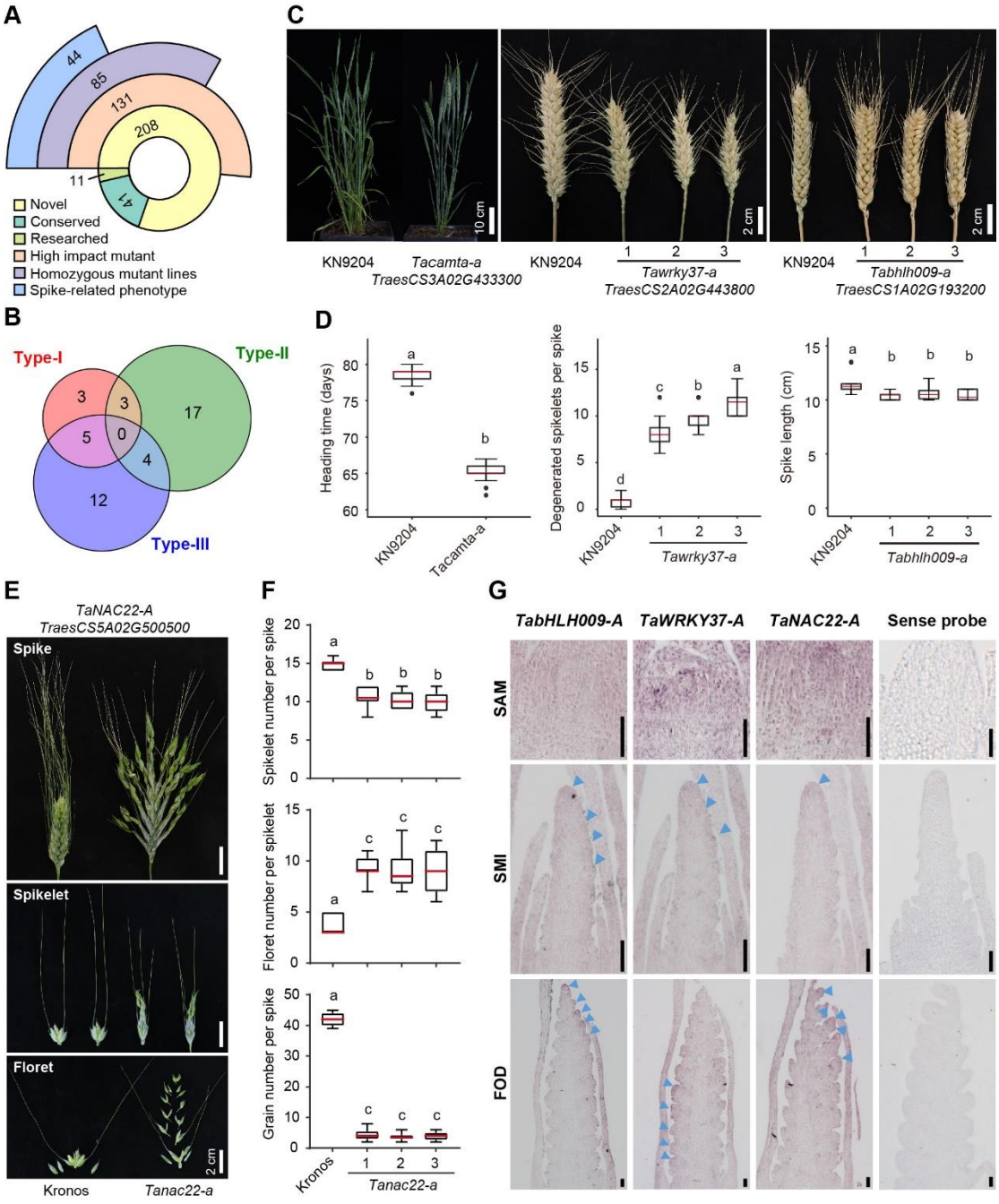


Figure 6. Novel factors identified regulate spike development.

(A) Summary of novel TFs with different categories of KN9204 TILLING mutant lines.

(B) Venn-diagram of different type of spike-related phenotype presented in KN9204 TILLING mutant lines containing homozygous mutation at novel TFs coding genes.

(C) Represented KN9204 TILLING mutant lines of type I/II/III spike developmental defects as compared to control KN9204. Type I, CAMTA (*TraesCS3A02G433300*, *Tacamta-a*), Scale bar = 10 cm; type II, WRKY (*TraesCS2A02G443800*, *Tawrky37-a*), Scale bar =2 cm; type III, bHLH (*TraesCS1A02G193200*, *Tabhlh009-a*), Scale bars = 2 cm.

(D) Quantification of different types of spike developmental defects in represented KN9204 TILLING mutant lines, as heading date of *Tacamta-a*, degenerated spikelet number per spike (DSNP) of *Tawrky37-a*, spikelet density of *Tabhlh009-a*, the Wild-type is KN9204. Two-tailed Student's *t*-tests. Different letters mean significant difference at $p < 0.01$.

(E) The spike developmental defect of *TraesCS5A02G500500* (*TaNAC22-A*) mutant lines as compared to control Kronos. Scale bars =2 cm.

(F) Quantification of floret number per spikelet (FNPS), gain number per spike (GNPS), spikelet number per spike (SNS) for represented Kronos TILLING mutant lines containing mutation of *TraesCS5A02G500500* (*Tanac22-a*). Two-tailed Student's *t*-tests. 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 $p < 0.01$.

(G) Spatiotemporal expression patterns of *TaWRKY37-A*, *TabHLH009-A* and *TaNAC22-A* at different spike developmental stages, as indicated by *in situ* hybridization. Sense probe is used as negative control. Scale bars = 100 μm.

Regulation of wheat spike architecture by novel factor TaMYB4-A

To further explore the potential 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 (Supplemental Figure 3B). It also located within a genetic region that was significantly associated with FSPS by GWAS analysis (Figure 7A and Figure 4F). Two haplotypes (C-type Vs T-type) of *TaMYB4-A* could separate a natural population of 214 wheat varieties with significant differences in FSPS (Figure 7B). Consistently, we find *TaMYB4-A* is initially expressed in the spikelet initiation region at DR stage, and highly expressed in the spikelet and floral primordia during SMI and FMI stages by *in situ* hybridization (Figure 7C), which fits its role in determining spikelet number and. Further, we obtained loss-of-function mutant lines of *TaMYB4-A* from KN9204 and Cadenza TILLING libraries (Figure 7D and Supplemental Figure 7A). Mutation of *TaMYB4-A* significantly reduced SL, SNS, GNPS (Figures 7D and 7E and Supplemental Figures 7B and 7C), which further confirmed its function in regulating spikelet development.

We further study the molecular regulation network of *TaMYB4-A* in mediating spike development. Based on the TRN generated, we extract the hierarchy transcription regulatory module containing TaMYB4-A (Supplemental Figure 7D). Among the potential upstream regulators of *TaMYB4-A*, we further confirmed that WFZP and a DOF type TF TaDOF17 (TraesCS2B02G592700) could repress or active *TaMYB4-A* in the reporter assay in tobacco leaves, respectively (Figure 7F and Supplemental Figure 7E). 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 after SMI with *TaMYB4-A* reduction after SMI stage (Supplemental Figure 7F). WFZP repression of *TaMYB4-A* is further evidenced by down regulation of *TaMYB4-A* in WFZP gain-of function transgenic wheat (Figure 7G) (Li et al., 2021). Among the numerous potential downstream targets of TaMYB4-A (Supplemental Figure 7D), several genes were confirmed to be regulated by TaMYB4-A in reporter assay in tobacco leaves (Figure 7F). This includes *TaAP2-39* (*TraesCS2D02G425700*), of which homologous 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, 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 (Figure 7F). Consistently, *TaAP2-39*, *TaI-BAK1* and *TaPILS7* showed a synchronized temporal expression pattern as *TaMYB4-A* during spike formation process (Supplemental Figure 7F).

Thus, *TaMYB4-A* positively regulates fertile spikelet likely through regulating hormones homeostasis and/or signaling, acting downstream of and repressed by WFZP (Figure 7H).

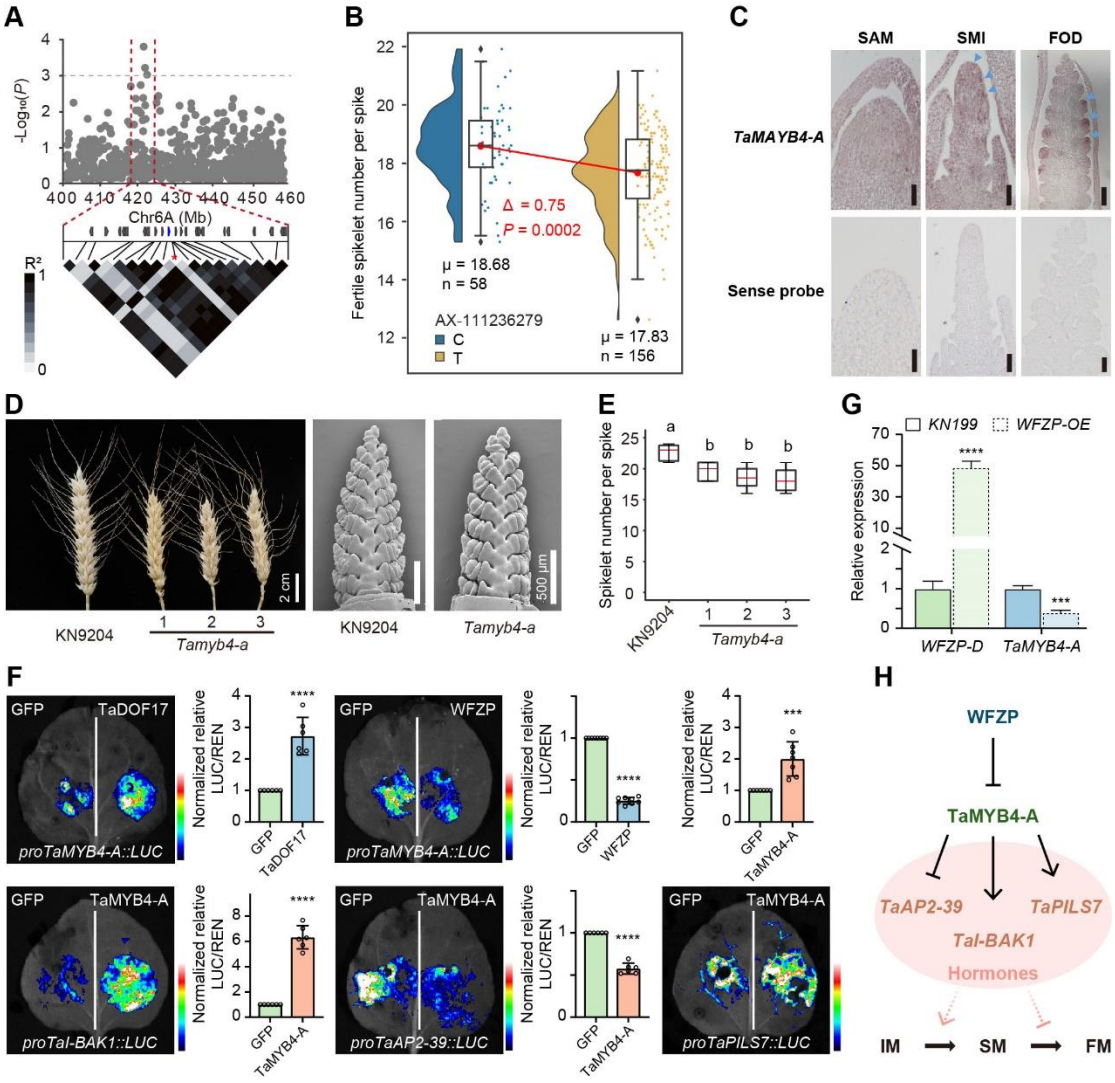


Figure 7. The novel factor TaMYB4-A regulates spike architecture.

(A) *TaMYB4-A* located in a GWAS signal associated with FSPS. A manhattan locuszoom were plot with gene models and linkage disequilibrium plot of SNPs shown below, in 5-Mb physical scale centered on the peak SNP.

(B) FSPS distribution between haplotypes (C and T) defined by the peak SNP. The bars within raincloud box plots represent 25th percentiles (haplotype T), medians, and 75th percentiles (haplotype C). Mean values of two haplotypes were linked by a red line, and one-way ANOVA was used to determine significant differences.

(C) Spatiotemporal expression pattern of *TaMYB4-A* is indicated by *in situ* hybridization. Scale bars = 100 μm.

(D) The spike phenotype and scanning electron micrographs (SEM) of *TaMYB4-A* (*TraesCS6A02G224200*) mutant from the KN9204 TILLING mutant library. Scale bars =2 cm (left), 500 μm (right).

(E) Statistics comparison of spikelet number per spike (SNS) between KN9204 and *Tamyb4-a* mutant lines. Two-tailed Student's *t*-tests. 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 $p < 0.01$.

(F) Luciferase reporter assays of *TaMYB4-A* regulatory network. Schematic diagram in the left part showing the vectors used in this array. Student's *t*-test was used for the statistical significance. **, $p < 0.001$; ****, $p < 0.0001$.
(G) The expression level of *WFZP* and *TaMYB4-A* in KN199, *WFZP*-OE transgenic plants by RT-qPCR. The error bars denote ±SD. ***, $p \leq 0.001$; ****, $p \leq 0.0001$.
(H) 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. Interestingly, the DNA variation SNP-939 (C/T) is located in the promoter region of *TaMYB4-A*, right under the WFZP core binding motif (Figure 8A). This suggested that the C/T SNP may contribute to *TaMYB4-A* expression difference via affecting WFZP binding. 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 (Figure 8B). We further selected 33 different wheat varieties of the two types (C-type and T-type) to measure the *TaMYB4-A* expression level and spike morphology. Wheat varieties of C-type showed elite agronomic traits with longer SL with more FSPS and a significant higher level of *TaMYB4-A* expression, while the T-type containing wheat varieties showed shorter SL, less FSPS and lower level of *TaMYB4-A* expression (Figure 8C). This result provides genetic evidence that the C/T SNP could cause expression level difference of *TaMYB4-A* and in turn affects SL and FSPS.

Furthermore, we wonder how this C/T SNP site within *TaMYB4-A* promoter being selected during the breeding process in China. Based on the exome capture sequencing data of the Chinese wheat mini-core collection (MCC) (Li et al., 2022a), percentages of accessions carrying the reference allele (C-type) was considerably higher in landraces (82.08%) and modern cultivar (78.18%) compared with introduced cultivar (45.45%) (Figure 8D). Of note, the frequency of C-type was significantly elevated along with the progress of breeding, indicating that the elite C-type of *TaMYB4-A* has

been widely used in the past breeding process of China, especially after year 1978 (Figure 8D). Founder parents have extensively promoted the improvement of wheat varieties in China since the 1950s, including Zhoumai16, St2422/464, Nanda2419, Nanda2419, St2422/464, Funo, Xiaoyan6, Aimengniu, Zhou8425B, Abbondanza and Lovrin10 (Yang et al., 2022). 17 out of 21 founder lines were C-type at SNP-93, expect for Aimengniu (T/T) and Lovrin10 (T/T) (Figure 8E). Importantly, the most widely grown derived varieties from Aimengniu or Lovrin10 (for instance, Jimai26, Jimai38 and Shi4185 derived from Lovrin10, and Yumai21, Yuejin5, Lumai15, Zhoumai16 and Zhoumai18 from Aimengniu) are apt to retain the C-type at SNP-939 from another parent (Figure 8E). Interestingly, the alleles showed distinct distribution characteristics in the major Chinese agro-ecological zones. The inferior T-type was relatively frequent (≥40 %) in spring wheat regions (VI, VII, VIII), followed by winter-spring mixed regions (IX and X, 22.22% and 23.08% respectively), and appears to be with low frequency (< 20%) in the winter wheat regions (I, II, III, IV, V). (Figure 8F).

Thus, the elite C-type allele of *TaMYB4-A* is likely originated from Chinese local germplasm and widely used during the breeding process but still hold the potential to be used for cultivar improvement for certain wheat production region in China.

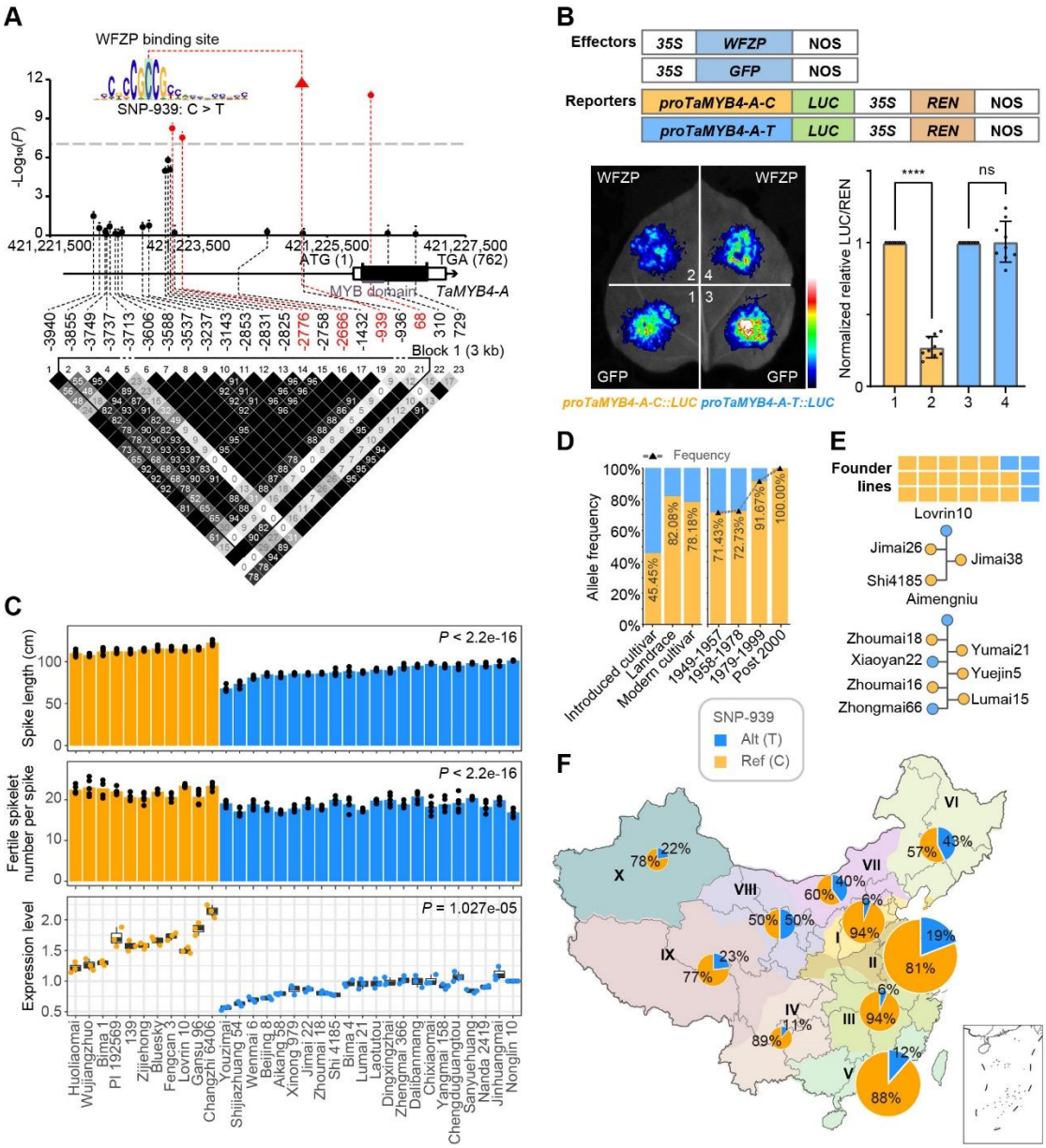


Figure 8. *TaMYB4-A* has a key SNP at WFZP binding site associate with spike architecture selection.

(A) Association analysis between genetic variations in *TaMYB4-A*. Dots connected with the blue dashed lines indicate the variants that are significantly associated with fertile spikelet number per spike (FSPS). The schematic diagram of the ~6-kb genomics region of *TaMYB4-A* is shown, followed by a LD plot with white to black represents r2 = 0-1. The SNP-939 was located in the core WFZP DNA-binding motif. Haplotype of selected 33 varieties for expression level and spike morphology determination were shown below.

(B) Luciferase reporter assays of *TaMYB4-A* promoter activity with the WFZP binding site C/T. Schematic diagram in the left part 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) Spike length and FSPS of cultivars with different haplotypes of *TaMYB4-A* based on five-point data from two years and their expression level of *TaMYB4-A*. Dots show data distribution (n=3 or 5 biologically independent samples), *p* values calculated using two-tailed *t*-test. (D)The percentages of accessions with reference and alternative allele of SNP-939 in different category (I, left) and breeding periods (I, right). The line chart shows the allele frequency of reference C-type.

(E) The genotype of wheat founder lines and their derived cultivars. The derived cultivars of Aimengniu and Lovrin10 were shown in the tree, and cultivars with reference C-type were in yellow and the T-type in blue.

(F) 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).

Discussion

Wheat is one of the first domesticated crops and domestication is linked with the modification of inflorescence architecture to ease harvesting and improve grain yield (Gauley and Boden, 2019). Increasing yields is also one of the primary goals for breeding. For cereal crops, inflorescence architecture largely determines the grain productivity via affecting the spikelet and floret development. Better understanding of the molecular mechanism that governs inflorescence architecture would facilitate the trait-designed breeding process. As compared to rice, maize and barley (Gao et al., 2019; Wang et al., 2021; Yuan et al., 2020; Zhang and Yuan, 2014), the genetic and molecular regulation of spike/inflorescence development in wheat is largely delayed. Here, we have generated a time-serial epigenomic landscapes consisting of various types of histone modifications, accessible chromatin, and transcriptomes of wheat shoot apex from vegetative development to spike architecture formation. This would be a valuable data resource for systematic study of molecular insights for wheat inflorescence/spike development and mining of key regulators for shaping spike architecture (Figure 1).

Epigenetic layer regulation of vegetative-to-reproductive transition in wheat

The meristem of shoot apex generates different primordium cells for initiation of various tissues, such as leaf primordia from SAM at vegetative stage, and spikelet

primordia from IM during flowering transition. This transition is tightly regulated by various factors, such as TaTB1, VRN1, TaFT1, Ppd1 via mediating endogenous and environmental signals (Figure 1D). Here, we found during the transition from SAM to DR/SMI (vegetative to reproductive growth), the chromatin accessibility is generally increased (Figure 1G), showing a synchronous pattern with gene activated during floral induction such as inflorescence meristem identity genes, hormone biosynthesis and signaling (Figures 2B and 2F). However, open chromatin is not sufficient to active genes expression, especially when H3K27me3 is covered at the genic region (Figures 2G and 2H). But gain-of chromatin accessibility does set a 'primed status' for later activation of genes when the H3K27me3 is removed at a late developmental stage (Figures 2G and 2H). Such chromatin status affected genes including well-known flowering time gene *VRN1*, and genes involved in spikelet meristem formation, such as *WAP3*, *TaFUL3*. Thus, chromatin layer regulation is associated with the transcriptional status of key regulators during vegetative-to-reproductive transition in wheat.

Integration of TRN with GWAS enables systematic and efficient identification of key factors in determine inflorescence architecture

The inflorescence of wheat is made up with spikelets and florets harbored within spikelet. Wheat inflorescence is determined at the time of terminal spikelet formation, which in turn affects the SNS. Whereas the arrangement of spikelet could influence the floret development inside. Thus, the initiation, distribution and termination of spikelet largely shape the inflorescence architecture and grain yield (Gao et al., 2019; Wang et al., 2021). This is likely driven by identity transition of different primordia cells in the context of hormone signaling and transcriptional regulatory network (Feng et al., 2017; Qi et al., 2019). Understanding the main regulatory network and identification of key factors that driving such network would give us potential candidates for shaping inflorescence structure. In addition, one could expect that genetic variation on such key factors within the regulatory network *per se* or variations that changing the regulatory circuit could generate influence on the outcome of inflorescence architecture.

Following this logic, by taking advantage of our time-serial profiling of transcriptome and epigenome dataset, in combination with TF-motif binding information in model plant (Castro-Mondragon et al., 2022), we build-up a transcriptional regulatory networks that likely governs the spike formation after floral transition (Figure 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, ARF, SPL families that being reported to regulate inflorescence formation in other crops (Figure 3) (Liu et al., 2022a; Ram et al., 2020; Rong et al., 2018; Wang et al., 2022; Xu et al., 2016). On top of this, we combined public available GWAS or QTL analysis with focus on the traits related to spike development to filter for those TFs located within the GWAS associated genetic regions and have SNP in the regulatory open chromatin regions that likely affects the transcriptional regulation circuit (Figure 4). This strategy identified 260 TFs, including 52 functionally analyzed in wheat or other crops. Through TILLING mutant lines screening, we confirmed mutant lines with 44 novel TFs showing spike development defects, including initiation, distribution and termination or degeneration of spikelet or even floret (Figure 6). Thus, such strategy sets a good example for batch screening factors for agronomic traits in crops.

TRN facilitates gene functional study and elite allele discovery in breeding application

TRN generated not only concentrate the attention for identification of key factors involved in spike development, but also provide guidance for gene functional study. This is evidenced by revealing the hierarchy regulation module of known individual factors such as SPL6-MADS34-MADS15-*HMA* (Figure 5). Importantly, such regulation module is relatively conserved within *Triticum* tribe, indicating a broad application for TRN generated in wheat to be used in other species. What is more, the regulatory circuit suggested by TRN could promote functional study of novel factor, for instance *TaMYB4-A* (Figure 7). In addition to the genes regulatory relation, TRN could enable the identification of critical region with high resolution that mediating transcriptional regulation. For the case of *TaMYB4-A*, we found the SNP presence in

the binding motif of upstream regulator WFZP serves as selection site during the breeding process in China (Figure 8). The elite allele (C-type) within promoter of *TaMYB4-A* is likely originated from Chinese local germplasm. The frequency of this elite allele is sharply increased during the later breeding process might because of the founder effect.

In summary, we integrated multi-omics data to reveal transcriptional regulatory network and epigenetic dynamic during wheat spike formation. With combination of GWAS analysis, we have identified dozens of novel factors that shape spike architecture and revealed that SNP under the WFZP binding site within promoter of *TaMYB4-A* is critical during the wheat breeding process in China.

Materials and 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

The winter wheat cultivar KN9204 was used to amplify gene sequences and the spring wheat cultivar Fielder was used to generate transgenic wheat plants. 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., 2022b). 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 2 consecutive years (planted in September in 2020 and 2021). Primers for genotyping are listed in Supplemental Table 12.

The phenotypic traits of mutant lines, T2 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.

RNA extraction, sequencing, quantitative PCR 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., 2022), 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). Quantitative PCR was performed using the ChamQ Universal SYBR qPCR Master Mix (Vazyme, Q711-02) by QuantStudio5 (Applied biosystems). Expression of genes of interest were normalized to Tubulin for calibration, relative expression level is calculated via the $2^{\Delta\Delta Ct}$ analysis method (Livak and Schmittgen, 2001). Primers used for qPCR are listed in Supplemental Table 12.

RNA *in situ* hybridization was carried out as described previously (Cui et al., 2010). Fresh young spikes were fixed in formalin-acetic acid-alcohol overnight at 4°C,

dehydrated through a standard ethanol series, embedded in Paraplast Plus tissue-embedding medium (Sigma-Aldrich, P3683), and sectioned at 8 μm width using a microtome (Leica Microsystems, RM2235). Digoxigenin-labeled RNA probes were synthesized using a DIG northern Starter Kit (Roche, 11277073910), according to the manufacturer's instructions. Primer sequences used for probe synthesis are listed in Supplemental Table 12.

Data Preprocessing and reads alignment

Raw reads were filtered by fastp v0.20.1 with parameter "--detect_adapter_for_pe" for adapters removing, low-quality bases trimming, and reads filtering (Chen et al., 2018). Furthermore, FastQC v0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was performed to ensure the high quality of reads.

Reads were aligned using either BWA-MEM v0.7.17 with parameter "-M" (for ATAC-seq and CUT&Tag seq) or hisat2 v2.1.0 with default parameters (for RNA-seq) to the *Triticum aestivum* (Chinese Spring) reference genome (IWGSC RefSeq v1.0, https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v1.0/) (Appels et al., 2018; Kim et al., 2019; Li and Durbin, 2009). Gene models from the IWGSC Annotation v1.1 was used as the reference and high-confidence genes were used throughout this study. The resulting SAM files were converted to BAM format, sorted, and indexed using Samtools v1.4 (Danecek et al., 2021). Sam files of RNA-seq generated from hisat2 were converted to bam files without deduplication. For ATAC-seq and CUT&Tag, SAM files were further filtered with "samtools view -bS -F 1,804 -f 2 -q 30" to filter out the low-quality mapped reads. Duplicates in the high-quality mapped reads were removed using Picard v2.23.3. Two replicates bam files were merged by samtools. To normalize and visualize the individual and merged replicate datasets, the BAM files were converted to bigwig files using bamCoverage provided by deepTools v3.3.0 with 10 bp bin size and normalized by RPKM (Reads Per Kilobase per Million mapped reads) with parameters "-bs 10 --effectiveGenomeSize 14,600,000,000 --normalizeUsing RPKM --smoothLength 50" (Ramirez et al., 2014).

921

922 **RNA-seq data analyses**

923 The number of paired reads that mapping to each gene was counted using feature
924 Counts v2.0.1 with the parameter "-p" (Liao et al., 2014). The counts files were then
925 used as inputs for DEGs (differentially expressed genes) analysis by DESeq2 v1.26.0
926 with a threshold "absolute value of Log2 Fold Change ⩾ 1 and FDR⩽ 0.05" (Love
927 et al., 2014). The raw counts were further normalized to TPM (Transcripts Per Kilobase
928 Million) for gene expression quantification. For subsequent clustering and
929 visualization, we obtained mean counts by merging three biological replicates. TPM
930 values of genes were Z-scaled and clustered by k-means method and displayed using R
931 package ComplexHeatmap (v2.4.3) (Gu et al., 2016). Functional enrichment was
932 performed using an R package clusterProfiler v3.18.1, and GO annotation files were
933 generated from IWGSC Annotation v1.1 (Yu et al., 2012).

934

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

936 ATAC-seq and CUT&Tag experiment were done follow the previous described method
937 (Zhao et al., 2022). Tn5 transposase used and tagmentation assay is done following the
938 manual (Vazyme, TD501-01). Libraries were purified with AMPure beads (Beckman,
939 A63881) and sequenced using the Illumina Novaseq platform at Annoroad Gene
940 Technology. Antibodies used for histone modifications are listed in Supplemental Table
941 13.

942

943 Data processing and reads alignment were performed as previously described (Zhao et
944 al., 2022). MACS2 v2.1.4 was used to call peaks. Parameters “-p 1e-3” was used for
945 H3K27ac, H3K4me3 and H2A.Z; parameters “--broad –broad-cutoff 0.05” were used
946 for H3K27me3 and H3K36me3 (Zhang et al., 2008). For ATAC-seq data, MACS2 was
947 used with parameters “--cutoff-analysis --nomodel --shift -100 --extsize 200”. The
948 peaks were annotated by R package ChIPseeker v1.26.2 with "annotatePeak" function
949 (Yu et al., 2015). The gene promoters are defined as 3.0 kb upstream of gene TSS.

For the identification of transcription factor footprints in ATAC-seq peaks, we used the HINT tool v0.13.2 of the Regulatory Genomics Toolbox (RGT) (Gusmao et al., 2014). Custom wheat genome was configured using IWGSC refseq v1.1 Chinese Spring genome based on the introduction of HINT software. TF motifs were downloaded from JASPAR Plantae database (https://jaspar.genereg.net/) (Castro-Mondragon et al., 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 "absolute value of Log2 Fold Change $\geqslant 1$ and FDR $\leqslant 0.05$".

Psuedotime indexing and gene regulatory network construction

We used the psuedotime 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., 2022). All of the expressed genes during spike reproductive development (From SMI to FOP) 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 GRNs construction, we only focused on DEGs with TPM values higher than 0.5 in

any stages from SMI to FOP. 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 SMI to FOP simultaneously. Secondly, we overlapped these TF-target regulations with WGCNA network constructed based on transcriptome of 8 stage from SAM to FOP, 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).

GWAS analysis

The 319, 558 SNPs (missing rates ≤ 0.1, MAF ≥ 0.05) from wheat 660K SNP array screening of 214 samples were performed to do association analysis with phenotypic data, implemented in Tassel v5.2 using the mixed linear model. The threshold for

genome-wide significance was determined by the value 1/independent number of SNPs (SNPs of weak LD with other SNPs ($r^2 < 0.5$) were regarded as independent SNPs). The genome-wide significant marker-trait associations were identified using a threshold cutoff of 3.16E-04. Manhattan plots and quantile-quantile plots were generated using R package "CMplot" (https://github.com/ YinLiLin/R-CMplot).

Gene based association analysis

The nucleotide polymorphisms in 6 kb genomic region of *TaMYB4-A*, including exons, intron regions, 4 kb promoter regions and 0.5 kb 3'-UTR regions, were identified using the whole genome exon capture sequencing data of 287 Chinese wheat mini-core collection samples (Li et al., 2022a). Tassel v5.2 was used to establish the association of polymorphisms with the FSS, and Haploview 4.2 was used to calculate the pairwise linkage disequilibrium and draw the LD plot.

Spike morphology observation by scanning electron microscopy (SEM)

Photomicrographs of young spikes were taken using a stereomicroscope (S8 APO, Leica Microsystems) 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*, *TaDOF17* were cloned into PTF101 vector to generate the effector construct *35Spro: 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 *target-pro*: LUC (see Supplemental Table 12 for primers). The plasmids were transformed into Agrobacterium GV3101. The mixture of bacterial solution

*35Spro:*TF-GFP (OD=0.5), target-pro: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*). 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 1 day' co-cultivation in dark and 2 days in light, the relative value of LUC/REN is indicated as average with standard error of multiple replicates.

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. For two groups' comparison of data, the student's t-test was used, such as Figure 3H, 5D, 5E, 7F, 7G, 8B, 8C, and Supplemental Figure 5A. For enrichment analysis, Fisher's exact test was used, such as Figure 2C, 2E, Figure3B, 3C, 4H, and Supplemental Figure 2C, 2D. For three or more independent groups comparison of data, Fisher's Least Significant Difference (LSD) was used, such as Figure 5D, 6D, 7E, and Supplemental Figure 5B, 6A, 6C, 6D, 7B, 7C.

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 (PRJCA013096) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa

Acknowledgements

We thank X.G. Liu (College of Life Sciences, Hebei Normal University) for the wheat *WFZP*-OE seeds. This research is supported by the National Natural Sciences Foundation of China (31921005), Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24010204), National Key Research and Development

Program of China (2021YFD1201500), and the Major Basic Research Program of Shandong Natural Science Foundation (ZR2019ZD15).

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 and qRT-PCR. 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; 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.

Competing interests

The authors declare no competing interests

References

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. Nature Plants **1**:1-6. 10.1038/Nplants.2014.16.

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 Bioinformatics **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. Genomics Proteomics Bioinformatics **19**:578-583. 10.1016/j.gpb.2021.08.001.

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.

Dobrovolskaya, O., Pont, C., Sibout, R., Martinek, P., Badaeva, E., Murat, F., Chosson, A., Watanabe, N., Prat, E., Gautier, N., et al. (2015). FRIZZY PANICLE drives supernumerary spikelets in bread wheat. Plant Physiol **167**:189-199. 10.1104/pp.114.250043.

Doebley, J., Stec, A., and Hubbard, L. (1997). The evolution of apical dominance in maize. Nature **386**:485-488. 10.1038/386485a0.

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 Bioinformatics **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 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 Biotechnology Journal **20**:168-182. 10.1111/pbi.13704.

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.

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.

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**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**10.3390/ijms22084142.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and 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.

Kellogg, E.A. (2022). Genetic control of branching patterns in grass inflorescences. Plant Cell **34**:2518-2533. 10.1093/plcell/koac080.

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.

Kobayashi, K., Maekawa, M., Miyao, A., Hirochika, H., and Kyozuka, J. (2010). PANICLE PHYTOMER2 (PAP2), encoding a SEPALLATA subfamily MADS-box protein, positively controls spikelet meristem identity in rice. Plant and Cell Physiology **51**:47-57. 10.1093/pcp/pcp166.

Kobayashi, K., Yasuno, N., Sato, Y., Yoda, M., Yamazaki, R., Kimizu, M., Yoshida, H., Nagamura, Y., and Kyozuka, J. (2012). Inflorescence Meristem Identity in Rice Is Specified by Overlapping Functions of Three AP1/FUL-Like MADS Box Genes and PAP2, a SEPALLATA MADS Box Gene. Plant Cell **24**:1848-1859. 10.1105/tpc.112.097105.

Kong, X.C., Wang, F., Geng, S.F., Guan, J.T., Tao, S., Jia, M.L., Sun, G.L., Wang, Z.Y., Wang, K., Ye, X.G., et al. (2022). The wheat AGL6-like MADS-box gene is a master regulator for floral organ identity and a target for spikelet meristem development manipulation. Plant Biotechnology Journal **20**:75-88. 10.1111/pbi.13696.

Koppolu, R., Chen, S., and Schnurbusch, T. (2022). Evolution of inflorescence branch modifications in cereal crops. Curr Opin Plant Biol **65**:102168. 10.1016/j.pbi.2021.102168.

Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R.H., Wang, X.D., Borrill, P., et al. (2017). Uncovering hidden variation in polyploid wheat. P Natl Acad Sci USA **114**:E913-E921. 10.1073/pnas.1619268114.

Kuzay, S., Lin, H., Li, C., Chen, S., Woods, D.P., Zhang, J., Lan, T., von Korff, M., and Dubcovsky, J. (2022). WAPO-A1 is the causal gene of the 7AL QTL for spikelet number per spike in wheat. PLoS Genet **18**:e1009747. 10.1371/journal.pgen.1009747.

Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J.A., Brown-Guedira, G., DeWitt, N., et al. (2019). Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. Theor Appl Genet **132**:2689-2705. 10.1007/s00122-019-03382-5.

Lee, Z.H., Hirakawa, T., Yamaguchi, N., and Ito, T. (2019). The Roles of Plant Hormones and Their Interactions with Regulatory Genes in Determining Meristem Activity. International Journal of Molecular Sciences **20**ARTN 4065 10.3390/ijms20164065.

Leiboff, S., and Hake, S. (2019). Reconstructing the Transcriptional Ontogeny of Maize and Sorghum Supports an Inverse Hourglass Model of Inflorescence Development. Curr Biol **29**:3410-3419 e3413. 10.1016/j.cub.2019.08.044.

Li, A., Hao, C., Wang, Z., Geng, S., Jia, M., Wang, F., Han, X., Kong, X., Yin, L., Tao, S., et al. (2022a). Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. Mol Plant **15**:504-519. 10.1016/j.molp.2022.01.004.

Li, C., Lin, H., Chen, A., Lau, M., Jernstedt, J., and Dubcovsky, J. (2019). Wheat VRN1, FUL2 and FUL3 play critical and redundant roles in spikelet development and spike determinacy. Development **146**10.1242/dev.175398.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754-1760. 10.1093/bioinformatics/btp324.

Li, L., Shi, F., Wang, G.L., Guan, Y.B., Zhang, Y.F., Chen, M.J., Chang, J.L., Yang, G.X., He, G.Y., Wang, Y.S., et al. (2022b). Conservation and Divergence of SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family between Wheat and Rice. International Journal of Molecular Sciences **23**ARTN 2099 10.3390/ijms23042099.

Li, L., Shi, F., Wang, Y., Yu, X., Zhi, J., Guan, Y., Zhao, H., Chang, J., Chen, M., Yang, G., et al. (2020). TaSPL13 regulates inflorescence architecture and development in transgenic wheat (Triticum aestivum L.). Plant Sci **296**:110516. 10.1016/j.plantsci.2020.110516.

Li, Y., Fu, X., Zhao, M., Zhang, W., Li, B., An, D., Li, J., Zhang, A., Liu, R., and Liu, X. (2018). A Genome-wide View of Transcriptome Dynamics During Early Spike Development in Bread Wheat. Sci Rep **8**:15338. 10.1038/s41598-018-33718-y.

Li, Y., Li, L., Zhao, M., Guo, L., Guo, X., Zhao, D., Batool, A., Dong, B., Xu, H., Cui, S., et al. (2021). Wheat FRIZZY PANICLE activates VERNALIZATION1-A and HOMEOBOX4-A to regulate spike development in wheat. Plant Biotechnol J 10.1111/pbi.13535.

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 Phytologist **202**:689-702. 10.1111/nph.12657.

Liu, J., Shi, X., Chang, Z., Ding, Y., and Ding, C. (2022a). Auxin Efflux Transporters OsPIN1c and OsPIN1d Function Redundantly in Regulating Rice (Oryza sativa L.) Panicle Development. Plant Cell Physiol **63**:305-316. 10.1093/pcp/pcab172.

Liu, X., Bie, X., Lin, X., Li, M., Wang, H., Zhang, X., Yang, Y., Zhang, C., Zhang, X., and Xiao, J. (2022b). Uncovering transcriptional regulatory network during regeneration for boosting wheat transformation. bioRxiv https://doi.org/10.1101/2022.10.21.513305.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) 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.

Meng, Q., Li, X., Zhu, W., Yang, L., Liang, W., Dreni, L., and Zhang, D. (2017). Regulatory network and genetic interactions established by OsMADS34 in rice inflorescence and spikelet morphogenesis. J Integr Plant Biol **59**:693-707. 10.1111/jipb.12594.

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. Journal of Experimental Botany **71**:1185-1198. 10.1093/jxb/erz483.

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 Genomics **20**:390. 10.1186/s12864-019-5726-x.

Ram, H., Sahadevan, S., Gale, N., Caggiano, M.P., Yu, X.L., Ohno, C., and Heisler, M.G. (2020). An integrated analysis of cell-type specific gene expression reveals genes regulated by REVOLUTA and KANADI1 in the Arabidopsis shoot apical meristem. Plos Genetics **16**ARTN

e1008661 10.1371/journal.pgen.1008661.

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.

Rong, X.F., Sang, Y.L., Wang, L., Meng, W.J., Zou, C.H., Dong, Y.X., Bie, X.M., Cheng, Z.J., and Zhang, X.S. (2018). Type-B ARRs Control Carpel Regeneration Through Mediating AGAMOUS Expression in Arabidopsis. Plant and Cell Physiology **59**:761-769. 10.1093/pcp/pcx187.

Sakuma, S., Golan, G., Guo, Z., Ogawa, T., Tagiri, A., Sugimoto, K., Bernhardt, N., Brassac, J., Mascher, M., Hensel, G., et al. (2019). Unleashing floret fertility in wheat through the mutation of a homeobox gene. Proc Natl Acad Sci U S A **116**:5182-5187. 10.1073/pnas.1815465116.

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.

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 Biotechnology Journal **15**:953-969. 10.1111/pbi.12690.

Uauy, C., Paraiso, F., Colasuonno, P., Tran, R.K., Tsai, H., Berardi, S., Comai, L., and Dubcovsky, J. (2009). A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biol **9**:115. 10.1186/1471-2229-9-115.

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.

Wang, C., Yang, X., and Li, G. (2021). Molecular Insights into Inflorescence Meristem Specification for Yield Potential in Cereal Crops. Int J Mol Sci **22**10.3390/ijms22073508.

Wang, K., Tang, D., Hong, L., Xu, W., Huang, J., Li, M., Gu, M., Xue, Y., and Cheng, Z. (2010). DEP and AFO regulate reproductive habit in rice. PLoS Genet **6**:e1000818. 10.1371/journal.pgen.1000818.

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. Science China-Life Sciences **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**ARTN 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., and 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, 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, D.B., and Yuan, Z. (2014). Molecular Control of Grass Inflorescence Development. Annu Rev Plant Biol **65**:553-578. 10.1146/annurev-arplant-050213-040104.

Zhang, L., Zhang, H., Qiao, L.Y., Miao, L.F., Yan, D., Liu, P., Zhao, G.Y., Jia, J.Z., and Gao, L.F. (2021). Wheat MADS-box gene TaSEP3-D1 negatively regulates heading date. Crop J **9**:1115-1123. 10.1016/j.cj.2020.12.007.

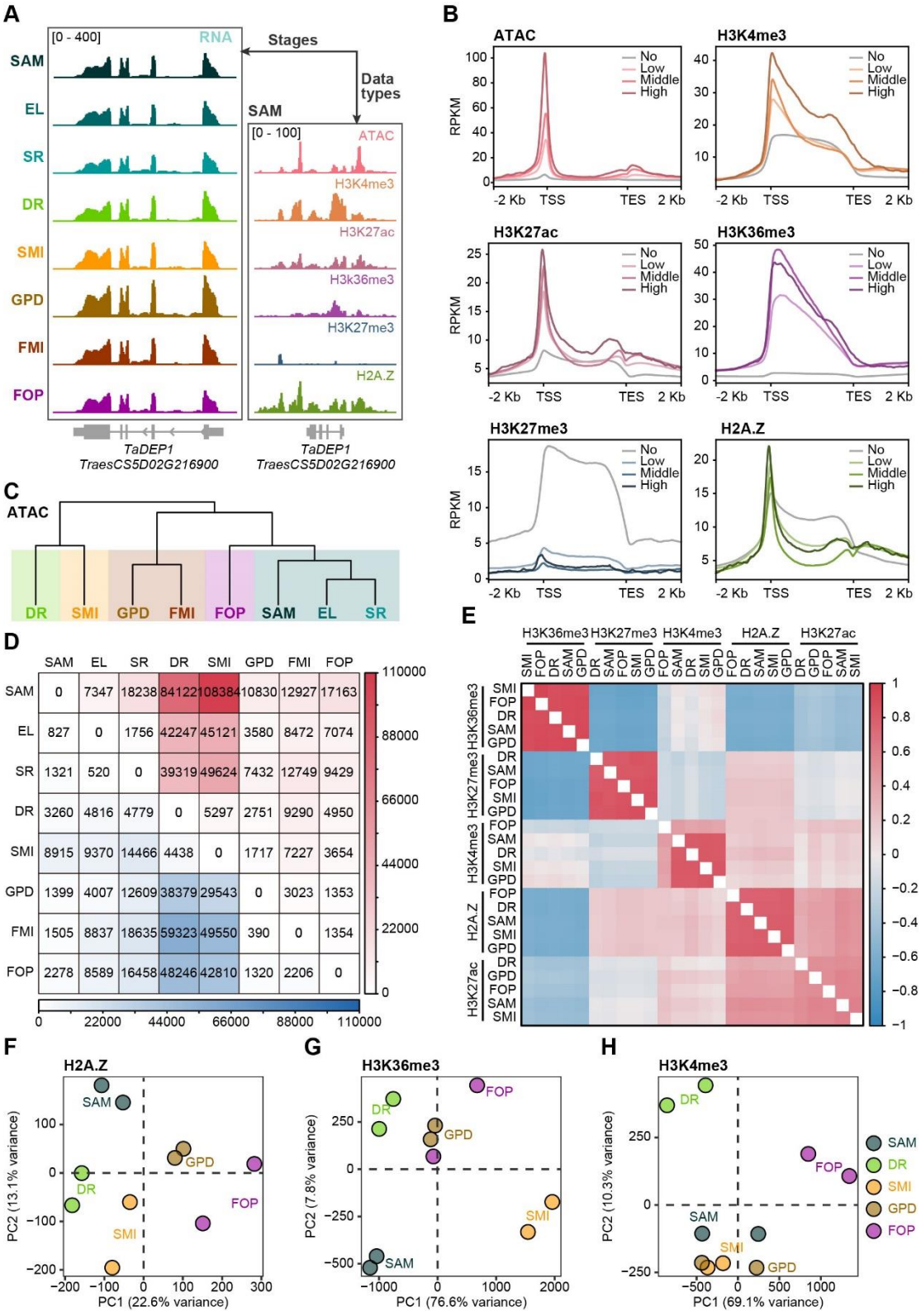
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol **9**:R137. 10.1186/gb-2008-9-9-r137.

Zhao, L., Lin, X., Yang, Y., Bie, X., Zhang, H., Chen, J., Liu, X., Wang, H., Jiang, J., Fu, X., et al. (2022). Chromatin reprogramming and transcriptional regulation orchestrate embryogenesis in hexaploid wheat. bioRxiv https://doi.org/10.1101/2022.01.21.477188.

Zhou, Y., Zhao, X.B., Li, Y.W., Xu, J., Bi, A.Y., Kang, L.P., Xu, D.X., Chen, H.F., Wang, Y., Wang, Y.G., et al. (2020). Triticum population sequencing provides insights into wheat adaptation. Nature Genetics **52**:1412-1422. 10.1038/s41588-020-00722-w.

Zhu, W.W., Yang, L., Wu, D., Meng, Q.C., Deng, X., Huang, G.Q., Zhang, J., Chen, X.F., Ferrandiz, C., Liang, W.Q., et al. (2022). Rice SEPALLATA genes OsMADS5 and OsMADS34 cooperate to limit inflorescence branching by repressing the TERMINAL FLOWER1-like gene RCN4. New Phytologist **233**:1682-1700. 10.1111/nph.17855.

Supplemental Figures



Supplemental Figure 1. Features of various histone modifications.

(A) IGV browser view at *TaDEP1* locus showing various epigenomic data types at SAM stage and transcriptome levels of different sampling stages.

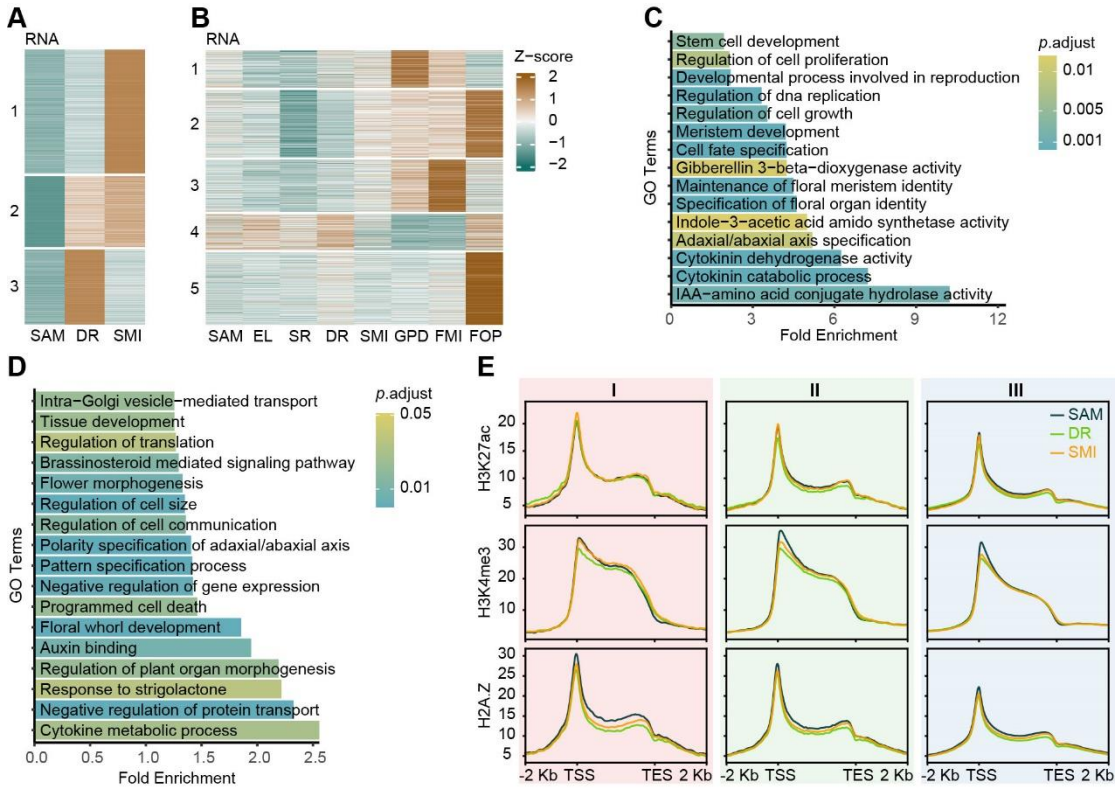
(B) Correlation between different types of histone modification profiles and gene expression levels.

(C) Cluster dendrogram of ATAC-seq data showing five distinct development clusters: vegetative cluster (SAM, EL, SR), flowering transition stage (DR), inflorescence initiation (SMI), spikelet meristem formation (GPD, FMI) and floret meristem formation (FOP).

(D) 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.

(E) 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.

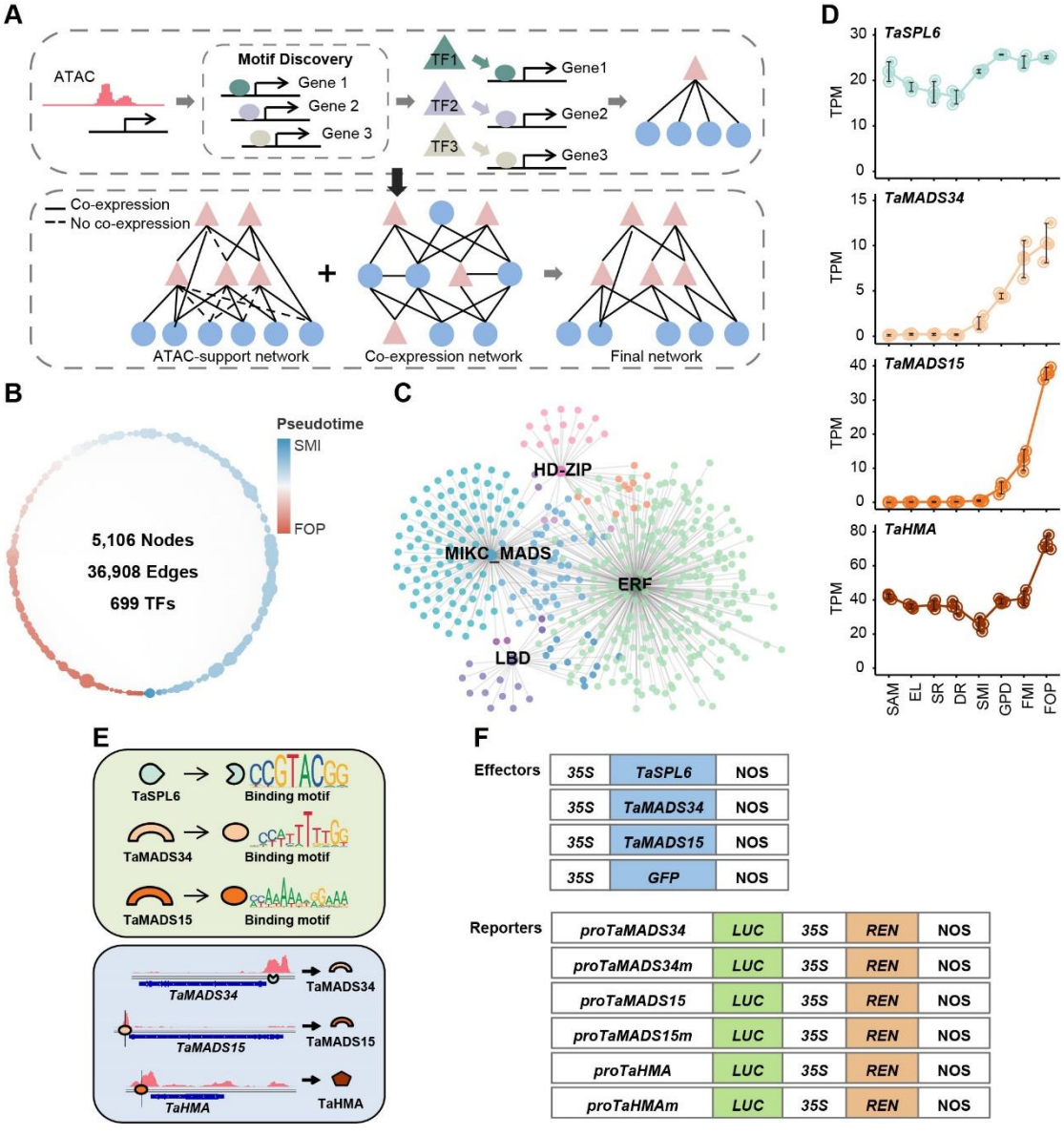
(F-H) PCA of H2A.Z (F), H3K36me3 (G) and H3K4me3 (H) samples during spike development. Each dot represents one sample; two bio-replicates are sequenced for each stage.



Supplemental Figure 2. Chromatin landscape dynamics associates with transcriptional change from vegetative to reproductive transition.

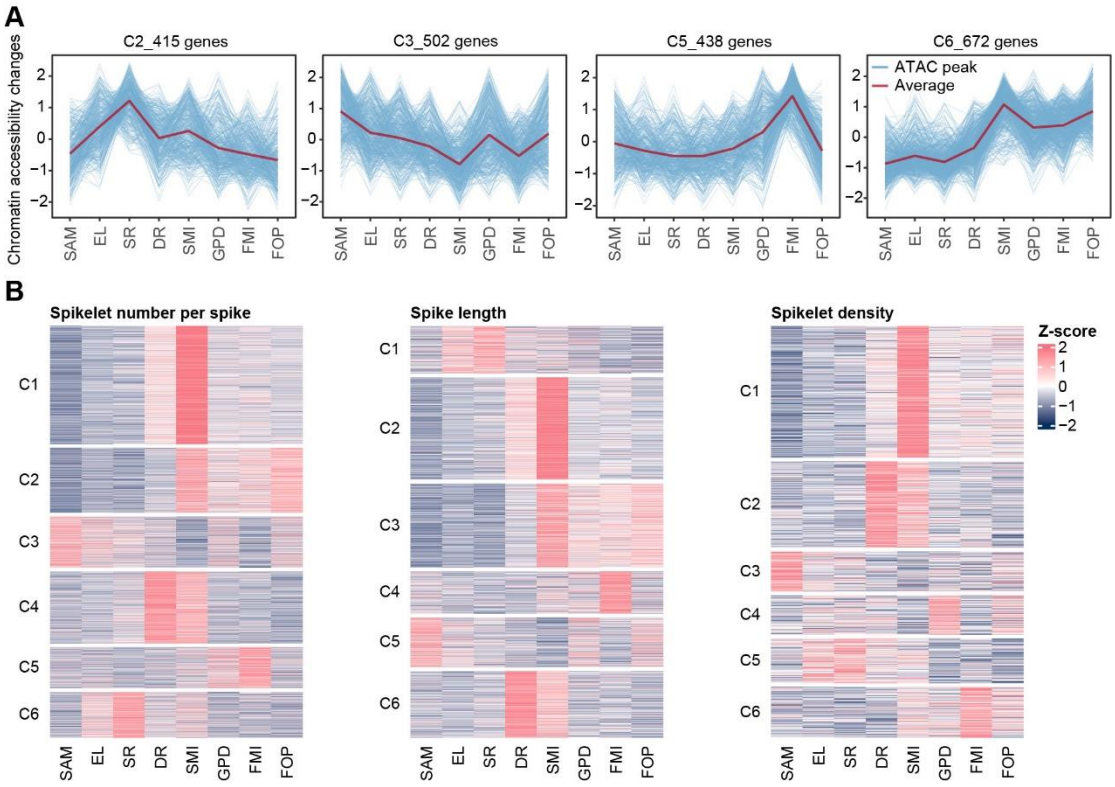
(A and B) Expression pattern of genes up-regulated at DR and SMI stage versus SAM stage (A) and genes up-regulated at later stages rather than DR or SMI stages (B). Heatmap showing k-mean clustering of gene expression.

(C and D) GO enrichment analysis of genes in gene set II (C), III (D) in Figure 2C.
(E) H3K27ac (top), H3K4me3 (middle) and H2A.Z (bottom) levels of genes in gene set I, II, III in Figure 2C at SAM, DR and SMI stages.



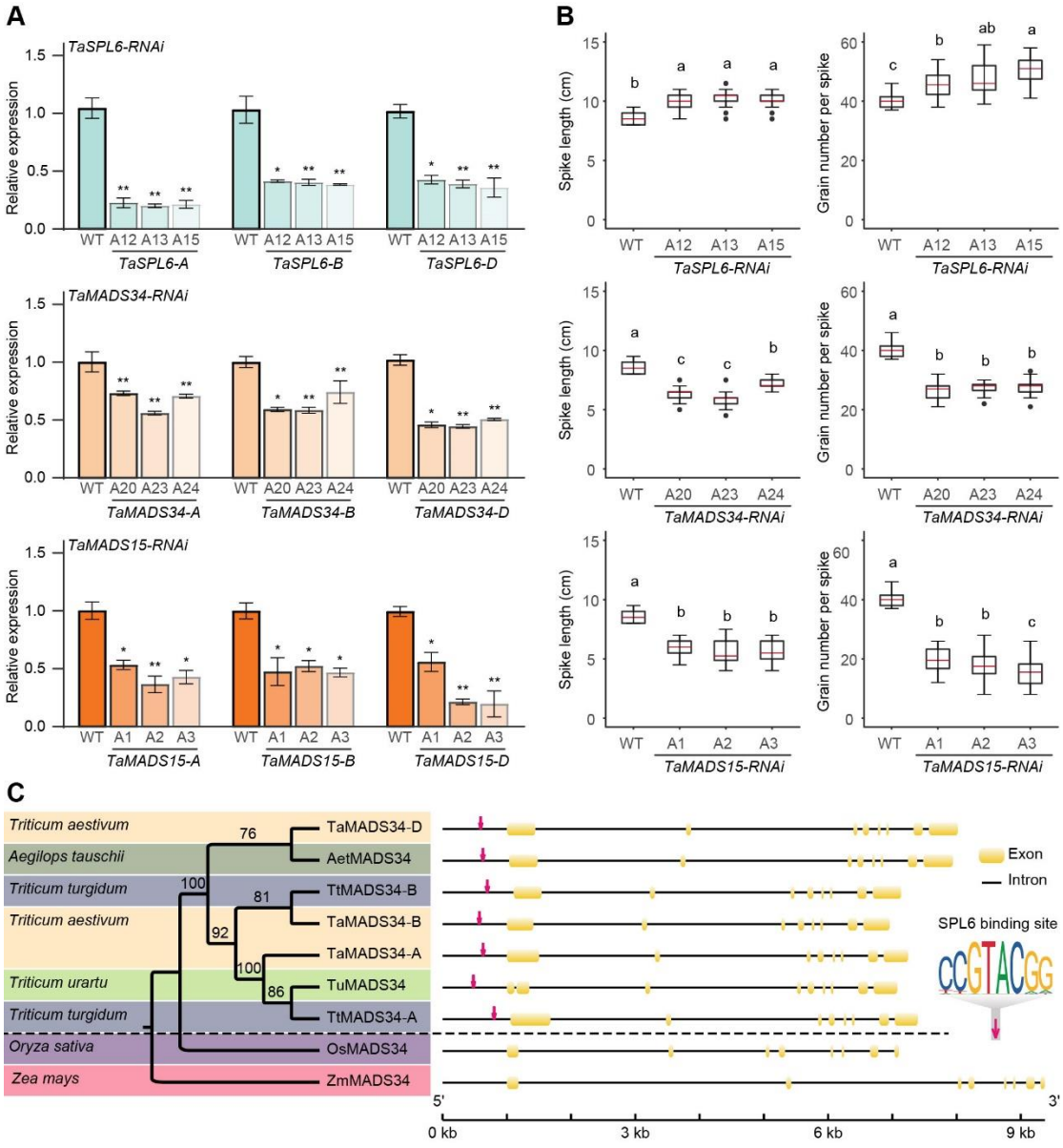
Supplemental Figure 3. Construction of transcription regulatory network (TRN).

(A) Schematic of the strategy for TRNs building (see method for detail).
(B) Total TRNs characterized for governing spike architecture formation.
(C) Map view of TF-targets for different TF families.
(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 SPL6-MADS34-MADS15-HMA regulatory module.
(F) Schematic diagram showing the vectors used in the Luciferase reporter assays of Figure 3H.



Supplemental Figure 4. The chromatin accessibility dynamics of GWAS signal associated genes.

(A and B) K-means clustering of chromatin accessibility of fertile spikelet number per spike (FSPS) (A), spikelet number per spike (SNS) (left-B), spike length(SL) (middle-B), spikelet density (SD) (right-B) related genes during spike development.

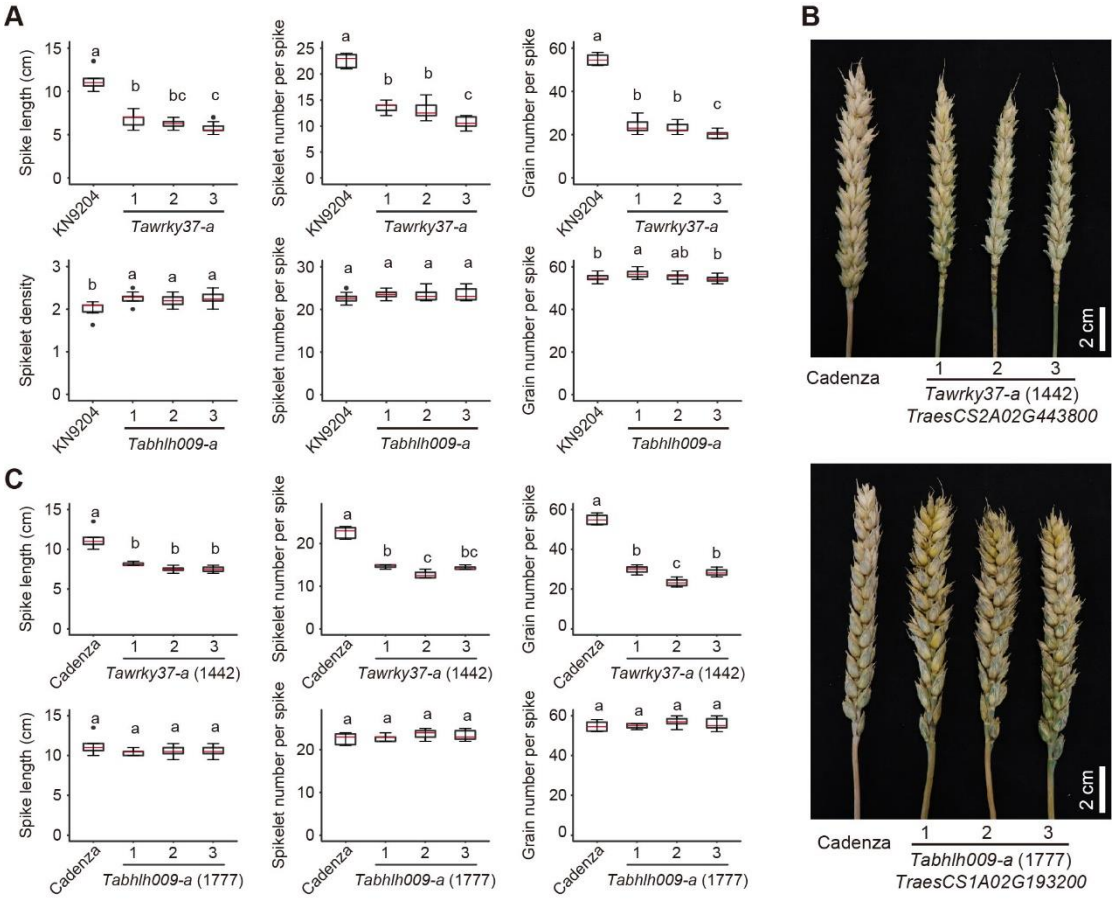


Supplemental Figure 5. SPL6-MADS34-MADS15-*HMA* module affects spike development in wheat.

(A) The knock-down efficiency of target genes in different *TaSPL6-RNAi*, *TaMADS34-RNAi*, *TaMADS15-RNAi* transgenic plants as measured by qPCR. Fielder served as wild-type (WT). 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 (*, $p \leq 0.05$; **, $p \leq 0.01$).

(B) Quantification of spike length, grain number per spike (GNPS) between WT, *TaSPL6*-RNAi, *TaMADS34-RNAi* and *TaMADS15-RNAi* transgenic plants. Two-tailed Student's t-tests. 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 $p < 0.01$.

(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.



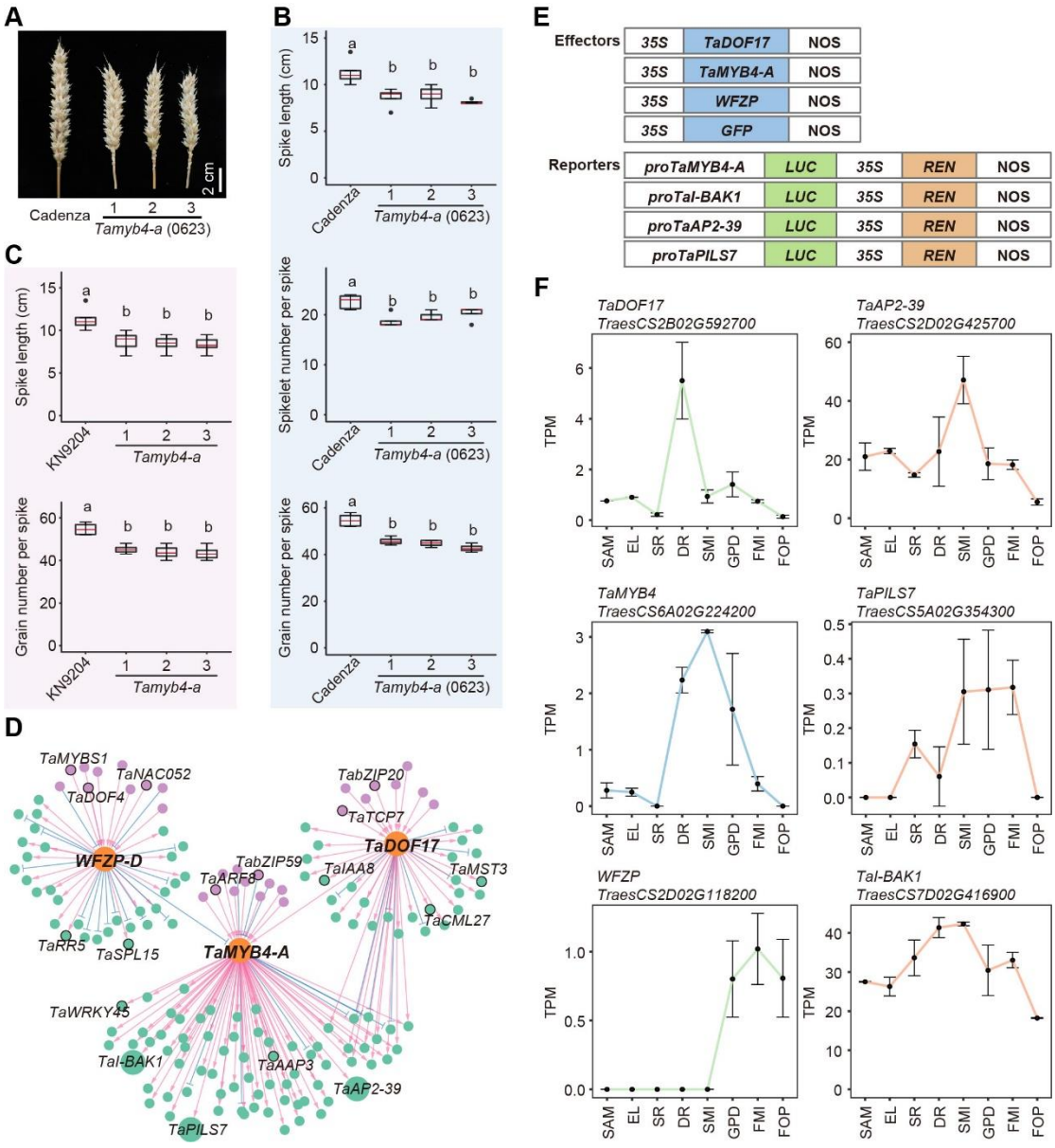
Supplemental Figure 6. The quantification data of spike phenotype associated with novel factors mutations.

(A) Quantification of spike length, spikelet number per spike (SNS), gain number per spike (GNPS) for represented KN9204 TILLING lines containing mutation of *Tabhlh009-a* or *Tawrky37-a*. Two-tailed Student's *t*-tests. 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 $p < 0.01$.

(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, spikelet number per spike (SNS), gain number per spike (GNPS) for represented Cadenza TILLING lines containing mutation of 1442 (*TaWRKY37-A*, *TraesCS2A02G443800*) and 1777 (*TabHLH009-A*, *TraesCS1A02G193200*). Two- tailed Student's *t*-tests. 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 $p < 0.01$.



Supplemental Figure 7. The regulation network of TaMYB4-A of wheat spike architecture.

(A) Comparison of the spike phenotype of Wild-type (Cadenza) and mutant lines from Cadenza mutant library. 0623 (*TaWRKY37-A*, *TraesCS2A02G443800*). Scale bars =2 cm.

(B) Statistics comparison of spike length and spikelet number per spike (SNS) between Wild-type (Cadenza) and 0623 (*TaWRKY37-A*, *TraesCS2A02G443800*) mutant lines. Two-tailed Student's *t*-tests. 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 $p < 0.01$.

(C) Statistics comparison of spike length and grain number per spike (GNPS) between KN9204 and *Tamyb4-a* mutant lines. Two-tailed Student's *t*-tests. 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 $p < 0.01$.

(D) The transcription regulatory network (TRN) of *TaMYB4-A*.

(E) Schematic diagram showing the vectors used in the Luciferase reporter assays of *TaMYB4-A* regulatory network of Figure 7F.

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