

# Transcriptome Profiling of Developing Leaf and Shoot Apices to Reveal the Molecular Mechanism and Coexpression Genes Responsible for the Wheat Heading Date

**Yuxin Yang**

Institute of Crop Sciences, CAAS

**Xueying Zhang**

Institute of Crop Sciences, CAAS

**Lichao Zhang**

Institute of Crop Sciences, CAAS

**Guoxiang Liu**

Institute of Crop sciences, CAAS

**Chuan Xia**

Institute of Crop Sciences, CAAS

**Xu Liu**

Institute of Crop Sciences, CAAS

**Xiuying Kong** (✉ [kongxiuying@caas.cn](mailto:kongxiuying@caas.cn))

Institute of Crop Sciences, CAAS

---

## Research article

**Keywords:** Wheat, Heading date, Gene expression diversity, Transcriptional factors, Weighted gene co-expression network analysis

**Posted Date:** August 19th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-51787/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Wheat is one of the most widely planted crops worldwide. The heading date is important for wheat environmental adaptability; it not only controls flowering time but also determines the yield component in terms of the grain number per spike.

**Results:** In this research, homozygous genotypes with early and late heading dates derived from backcrossed progeny were selected to conduct RNA-seq analysis at the double ridge stage and androgynous primordium differentiation stage of the leaf and apical meristem, respectively. In total, 18,352 differentially expressed genes (DEGs) were identified. Five common GO terms and 214 common metabolic pathways were obtained by MapMan software. In addition, 1,225 DEGs were annotated to 52 transcription factor gene families, and we discovered that most of the LFY, SBP, and MADS-box genes were highly expressed during the apical development of wheat with the early heading date. Additionally, weighted gene coexpression network analysis (WGCNA) found that the DEGs clustered into 17 coexpression modules, and 16, 336, 446, and 124 DEGs had biological connections with *Vrn1-5A*, *Vrn3-7B*, *Ppd-1D*, and *WSOC1*, respectively. The 16 genes connected with *Vrn1-5A* were divided into three types of expression patterns. We further identified one important flowering time gene, *TraesCS2D02G181400*, which is a MADS-MIKC transcriptional factor and has a biological connection with *Vrn1-5A*. Gene function showed that this gene was differentially expressed at the A2.0 and L3.5 stages between the genotypes with early and late heading.

**Conclusions:** The RNA-seq analysis results in our present research are a valuable bioinformatics resource for further study of the molecular mechanism of the wheat heading date. Overall, our results shed light on the genetic architecture of wheat flowering time.

## Background

Wheat (*Triticum aestivum* L.) is one of the most important and widely distributed food crops in human society. The wide global planting area of bread wheat is attributed to its high adaptability to the natural environment. Heading date (flowering time) is an important adaptive trait for crop genetic breeding.

Heading date regulation networks mainly include vernalization, the photoperiod, and hormones (such as gibberellin) [1, 2]. There are three important regulatory genes in the vernalization pathway: *Vernalization1* (*Vrn1-5A*, *Vrn1-5B*, *Vrn1-5D*) [3-6], *Vernalization2* (*Vrn2*) [1, 4], and *Vernalization3* (*Vrn3*) [3-6]. *Vrn1* can promote flowering and has three homologous genes in wheat: *Vrn1-5A*, *Vrn1-5B*, and *Vrn1-5D*. *Vrn-2* is the main flower suppressor gene, and it is downregulated by vernalization and short-day treatment [1, 4, 7]. *Vrn-3* is a mobile signal protein that moves from the leaf to the apical meristem and induces flowering time [1, 8]. The wheat response to photoperiod is mainly controlled by the *photoperiod* (*Ppd*) genetic locus, and genes participating in the photoperiod pathway mainly include *Ppd-A1*, *Ppd-B1*, and *Ppd-D1* [9, 10]. Previous studies reported that *Ppd-1* plays a key role in the inflorescence morphological structure and young spike development process [11], and *Ppd-1* accelerates flowering by upregulating *Vrn3* expression

under long-day conditions [12, 13]. Wheat *SOC1* (*WSOC1*) is a key regulatory gene in the gibberellin regulation pathway and affects the heading date in polyploid wheat [14]; upregulation of *WSOC1* can accelerate spike development [15].

Flower development in higher plants mainly occurs in three stages: the floral induction phase, the floral primordia phase, and the floral organ development phase. The double ridge stage and androgynous primordium differentiation stage are key stages for flower induction and floral meristem development in wheat. According to a previous study, at the double ridge stage (W1.5-W2.0), the emergence of the first spikelet primordium forms the apical meristem and determines the number of spikelets on each spike of wheat; at the differentiation stage of the androgynous primordium, the first floret primordium at the base of the spikelet begins to differentiate, and the stem begins to extend (W3.5) [16]. Thus, to explore the molecular mechanism of the heading date, we performed RNA-seq analysis of two wheat accessions with a significant difference in heading date at W2.0 and W3.5.

In recent years, with the completion of hexaploidy bread wheat reference genomes (IWGSC, 2018), use of the bioinformatic method to discover flowering time genes and to examine molecular characterization has significantly accelerated the study of wheat flowering time. More than nine hundred flowering time genes have been identified in wheat by BLAST searches and OrthoMCL clustering methods [17]. Researchers have discovered eighty-four SNPs that are highly related to spike numbers based on GWAS [18]. Additionally, cleistogamous wheat is a new flowering type that was discovered by transcriptome analysis [19].

However, wheat is a hexaploid species and has a complex genome, which poses a challenge for discovering flowering time genes and revealing their genetic characteristics. Here, homozygous genotypes with early and late heading dates from the BC<sub>3</sub>F<sub>4</sub> population resulting from backcrossed progenies of YZ4110 × *m605* were selected for RNA-seq. We aimed to investigate transcriptional regulation during wheat heading date development and to identify the key flowering time genes using weighted gene coexpression network construction. We anticipate that through transcriptome analysis, we can further explore the genetic characteristics of wheat flowering time and promote the wheat improvement process.

## Results

### Generation of RNA-seq data

In this study, we constructed 24 RNA-seq libraries of the 8 treatments, WHd-A2.0, MHd-A2.0, WHd-L2.0, MHd-L2.0, WHd-A3.5, MHd-A3.5, WHd-L3.5 and MHd-L3.5, and each treatment included 3 biological repeats. After high-throughput sequencing, each sample contained 40 – 59.4 million reads; the data size of each sample was between 6 – 8.9 Gb, the Q30 value of each sequencing sample exceeded 87%, the GC content distribution was 50 – 56%, and sequencing alignment analysis results showed that 89.35 – 97.79% of the reads could be mapped to IWGSC RefSeq v1.1 (Table S1). Principal component analysis

(PCA) of the eight raw sequencing datasets indicated that the samples could be clustered into four groups according to their genotype, which showed good repeatability between samples that could be used for subsequent analysis (Fig. S2).

### Identification of differentially expressed genes

To identify genes that differed significantly between the early and late heading genotypes during the development of young panicles, we analyzed the differentially expressed genes (DEGs) with strict quality control. The DEGs of L2.0, L3.5, A2.0, and A3.5 were compared between the genotypes with an early heading time (WHd) and a late heading time (MHd). Finally, we identified 18,325 unique differentially expressed genes (Fig. 1a, Additional Table S2). Several reported flowering time genes were differentially expressed, including *Vrn1-5A* [3], *Vrn3-7B* [5], *Ppd-1D* [9, 10], and *WSOC1* [14]. In addition, we discovered that when the growth cone developed to the W2.0 stage, there were 12,941 DEGs in the apical meristem between WHd-A2.0 and MHd-A2.0, of which 9,990 DEGs were upregulated and 2,951 DEGs were downregulated (Fig. S3a). When the growth cone developed to the W3.5 stage, in the apical meristem, there were 1,557 DEGs between WHd-A3.5 and MHd-A3.5, of which 1,406 were upregulated and 151 were downregulated (Fig. S3b). In the leaf tissue, there were 6,390 DEGs between WHd-L2.0 and MHd-L2.0, of which 2,735 genes were upregulated and 3,655 genes were downregulated (Fig. S3c). In leaf tissue at the W3.5 stage, there were 1,040 DEGs, of which 764 were upregulated and 276 were downregulated in WHd-L3.5 vs. MHd-L3.5 (Fig. S3d). We screened the common differentially expressed genes at different developmental stages in the same tissue and found that 1,084 genes were common DEGs between WHd-A2.0 and MHd-A2.0 and between WHd-A3.5 and MHd-A3.5 (Fig. 1b). In addition, 215 genes were common DEGs between WHd-L2.0 and MHd-L2.0 and between WHd-L3.5 and MHd-L3.5 (Fig. 1c). We speculated that these genes play an important role in spikelet development at the W2.0 and W3.5 stages. Additionally, we explored the distribution of differentially expressed genes along the three wheat subgenomes, and the DEGs were mainly distributed at the end of the chromosome arm (Fig. S4).

### GO enrichment and MapMan metabolic pathway analysis of differentially expressed genes

To explore the regulatory pathways of DEGs, GO enrichment and MapMan metabolic pathway analysis were conducted between the early heading date and the late heading date accessions. We finally identified 657 significant GO terms (Table S3) involved in biological process (BP), molecular function (MF), and cellular component (CC) (Fig. 2; Fig. S5; Fig. S6; Fig. S7). There were 224, 94, 210, and 129 significant GO terms in A2.0, A3.5, L2.0, and L3.5, respectively (Table S3). Among them, we discovered several flowering time-related GO terms, including GO:0005975 (carbohydrate metabolism) [20], GO:0005991 (trehalose metabolic process) [21], and GO:0019684 (photosynthesis, light reaction) [22]. The metabolic pathways of four groups of DEGs were visualized by MapMan software (Fig. 3; Fig. S8; Fig. S9; Fig. S10), and a total of 328 significant metabolic pathways were finally identified. There were 224, 94, 210, and 129 significant MapMan metabolic pathway terms in A2.0, A3.5, L2.0, and L3.5, respectively (Table S4). Additionally, we found 5 common GO terms, i.e., GO:0003700 (transcription factor activity, sequence-specific DNA binding), GO:0001071 (nucleic acid binding transcription factor activity),

GO:0003824 (catalytic activity), GO:0043565 (sequence-specific DNA binding), and GO:0016491 (oxidoreductase activity), and 214 common metabolic pathways, including bin:1.1.1 (light reaction), bin:11.8 (lipid metabolism), and bin:13.2.3.5 (amino acid metabolism) (Fig. 4). We suggest that these common regulatory pathways can provide ideas for further exploring the variation in the molecular mechanisms of wild and mutant wheat.

### Transcriptome factor classification and identification

To identify the transcription factors (TFs) involved in the heading time development process, the DEGs were subjected to transcription factor analysis using iTAK software. In our study, a total of 1,225 transcription factors were classified into 56 transcription factor families (Table S5). A large number of differentially expressed genes were bHLH (129, 10.51%), WRKY (109, 8.88%), NAC (91, 7.41%), AP2/ERF-ERF (88, 7.17%), and MYB (64, 5.21%) transcription factors (Table S5). Among them, we also identified several transcription factor families involved in the plant flowering process, including three LFY transcription factors [23], eighteen SBP transcription factors, thirty-six MADS-MIKC transcription factors and ten MADS-M-type transcription factors (Table S5). Gene expression level analysis showed that transcription factors were expressed at the critical period that determines flowering time. For example, we found that the three LFY transcription factor genes (*TraesCS2A02G443100*, *TraesCS2B02G464200*, *TraesCS2D02G442200*) were highly expressed in the apical meristem, especially in WHd-A3.5 and WHd-A2.0 (Fig. 5a). Most of the SBP transcription factor genes were highly expressed in WHd-A2.0 and WHd-A3.5 (Fig. 5b). Additionally, in the MADS-box gene family, we discovered that the *TraesCS5A02G391700* (*Vrn1-5A*) gene always showed a high expression level in both wild and mutant wheat, *TraesCS3B02G612600* was highly expressed in leaf tissue, and two genes (*TraesCS3D02G284200*, *TraesCS3A02G284400*) were highly expressed in the apical meristem (Fig. 5c). GO enrichment analysis of LFY transcription factor genes revealed that the genes were mainly significantly involved in the GO:0006355 pathway (regulation of transcription, DNA-template,  $p = 6.60 \times 10^{-23}$ ), and the SBP transcription factor genes significantly participated in the GO:0003677 pathway (DNA binding,  $p = 9.70 \times 10^{-20}$ ). Additionally, the MADS-box transcription factors were mainly involved in the GO:0019219 pathway (regulation of nucleobase-containing compound metabolic process,  $p = 4.10 \times 10^{-36}$ ).

### Construction of the flowering gene regulatory network

In our study, a total of 18,352 DEGs were used to construct the coexpression network. In selecting the soft power value, we chose  $\beta = 14$  to build the coexpression network (Fig. S11a; Fig. S11b). We calculated the average gene connectivity under different soft-thresholding powers and found that when  $\beta = 14$ , the coexpression network had scale-free characteristics (Fig. S11c; Fig. S11d). Finally, a total of 17 coexpression modules were obtained by the “dynamic tree cut” method. Each branch in the cluster tree represents a gene set, and different modules were distinguished by different colors (Fig. S11e). The correlation value between coexpression modules and samples was also calculated (Fig. S12). The reported flowering time genes *Vrn1-5A* (*TraesCS5A02G391700*), *Vrn3-7B* (*TraesCS7B02G013100*), *Ppd-1D* (*TraesCS2D02G079600*), and *WSOC1* (*TraesCS4D02G341700*) were selected to build a gene coexpression

network, and genes connected to the reported flowering time genes were extracted from the coexpression modules. We finally detected 16, 336, 446 and 124 DEGs with biological connections to *Vrn1-5A* (Fig. 6a), *Vrn3-7B* (Fig. S13a), *Ppd-1D* (Fig. S14a), and *WSOC1* (Fig. S15a), respectively. The completed gene list related to the reported flowering time genes is summarized in Table S6.

### Analysis of gene expression patterns coexpressed with flowering genes

To further study the gene expression pattern of the differentially expressed genes coexpressed with the wheat heading date gene, we analyzed the 16 genes that had biological connections with *Vrn1-5A*. We found the genes that were coexpressed with *Vrn1-5A* could be divided into three patterns according to their gene expression level. First, *TraesCS2D02G181400* is a MADS-MIKC transcription factor that was highly expressed in WHd-A3.5, MHd-A3.5, MHd-L3.5, and WHd-A2.0 but expressed at low levels in WHd-L2.0, MHd-L2.0, MHd-A2.0, and WHd-L3.5 (Fig. 6b). *TraesCS2D02G181400* was upregulated between WHd-A2.0 and MHd-A2.0 and downregulated between WHd-L3.5 and MHd-L3.5 (Fig. 6c). Additionally, the other five genes (*TraesCS1D02G339300*, *TraesCS2A02G562100*, *TraesCS3A02G007200*, *TraesCS3B02G361000*, *TraesCS4D02G337000*) were expressed at low levels at all stages in both the early and late heading date accessions (Fig. 6b). In addition, the remaining ten differentially expressed genes (*TraesCS3A02G231100*, *TraesCS3B02G152300*, *TraesCS3B02G213900*, *TraesCS3B02G260400*, *TraesCS5D02G386500*, *TraesCS6A02G146500*, *TraesCS7A02G439300*, *TraesCS7A02G489000*, *TraesCS7D02G429000*, *TraesCS7D02G475300*) showed higher expression levels in WHd-L2.0, WHd-L3.5, and MHd-L2.0 but presented lower expression in WHd-A2.0, WHd-A3.5, MHd-A2.0, MHd-A3.5, and MHd-L3.5 (Fig. 6b). Additionally, the genes that were highly related to *Vrn3-7B* were mainly highly expressed in the leaf tissues (Fig. S13b). Genes connected with *Ppd-1D* tended to be expressed in the leaf (Fig. S14b). Some genes that were coexpressed with *WSOC1* were preferentially expressed in leaf tissues (Fig. S15b). Furthermore, by gene function annotation, we found some vital candidate flowering time genes in the coexpression network, such as *TraesCS1A02G220300*, *TraesCS2D02G181400*, *TraesCS3A02G143100*, and *TraesCS5A02G166100* (Table 1).

## Discussion

Bread wheat is a hexaploid crop with a large genome (17 Gb) and three subgenomes (AABBDD). Flowering time is a quantitative trait that is controlled by many genetic loci, which means that some flowering time genes still need to be discovered. In our research, transcriptome sequencing revealed that four reported flowering genes, *Vrn1-5A*, *Vrn3-7B*, *Ppd-1D*, and *WSOC1*, were differentially expressed. This result is consistent with previous studies that vernalization, photoperiod, and endogenous hormones have strong effects on wheat heading date [3, 9, 14].

WGCNA can effectively identify highly reliable flowering time genes through coexpression network construction. We systematically constructed a comprehensive coexpression network of *Vrn1-5A*, *Vrn3-7B*, *Ppd-1D*, and *WSOC1* using WGCNA. Finally, we identified 16 genes connected to *Vrn1-5A* by WGCNA analysis (Fig. 6a). Gene function annotation revealed some important candidate flowering time genes.

Among them, we found that *TraesCS2D02G181400* is a MADS-MIKC transcription factor, and its homologous gene in rice is *OsMADS18*, which is an AP1/FUL-like MADS-box gene that determines the formation of inflorescence meristem characteristics by interacting with PAP2 in the floral meristem [24]. We also discovered one basic helix-loop-helix (bHLH) transcription factor (*TraesCS5D02G386500*). Its homologous gene in *Arabidopsis* is *AT2G43010*, and it mainly negatively regulated the red light response mediated by phyB; thus, we speculated that red light could affect the expression of *Vrn1-5A*. In addition, *TraesCS6A02G146500* is a glucose-6-phosphate 1-dehydrogenase (G6DPH) that can provide energy and other metabolites for plant growth and development, and we suggest that the expression of vernalization genes requires other genes to provide energy. The above analysis shows that coexpression analysis is an effective method to detect candidate flowering time genes, and we suggest that the identified flowering time genes connected with *Vrn1-5A* can be used for further research to reveal their biological functions.

In the other coexpression network of the reported flowering time genes, including *Vrn3-7B*, *Ppd-1D*, and *WSOC1*, we also identified some important flowering time genes and explored the gene expression pattern. In the *Vrn3-7B* gene network (Fig. S13a), we identified two CONSTANS-like proteins (*TraesCS5A02G166100*, *TraesCS5D02G170700*). The homolog of both *TraesCS5A02G166100* and *TraesCS5D02G170700* in *Arabidopsis* is *AT5G15840*, which is a zinc finger transcription factor-like protein that acts upstream of *FT* and *SOC1* and is mainly involved in flowering regulation under long-day conditions [25]. Through *Ppd-1D* coexpression network construction (Fig. S14a), we identified one CONSTANS-like protein gene, *TraesCS1A02G220300*, and gene function annotation showed that its homologous gene in *Arabidopsis* is *AT5G57660*, which is involved in the flowering development process [26]. The *WSOC1* coexpression network (Fig. S15a) revealed three flowering locus T genes (*TraesCS3A02G143100*, *TraesCS3B02G162000*, *TraesCS3D02G144500*), which is consistent with previous reports that gibberellin can activate the expression of Flowering Locus T [14]. Based on the above analysis, we suggest that the discussed genes (*TraesCS2D02G181400*, *TraesCS5D02G386500*, *TraesCS6A02G146500*, *TraesCS5A02G166100*, *TraesCS5D02G170700*, *TraesCS1A02G220300*) connected with the reported flowering time genes (*Vrn1-5A*, *Vrn3-7B*, *Ppd-1D*, *WSOC1*) may play a key role in the wheat heading date regulation pathway.

Transcription factors play important roles in the regulation of plant heading/flowering stage development. Gene expression analysis identified that most of the LFY, SBP, and MADS-box genes were highly expressed in the apical meristem of wild-type wheat (WHd-A2.0, WHd-A3.5), which means that these types of transcription factors positively regulate the floral primordia phase. For example, *TraesCS2A02G443100* is an LFY transcription factor; its homologous gene in rice is *Os04g0598300*, which mainly controls the heading date and determines the development of branches and tiller in the ear, and overexpression of *Os04g0598300* can promote the heading date in rice [27]. In addition, one of the SBP transcription factors, *TraesCS7A02G246500*, was highly expressed at the WHd-A2.0 and WHd-A3.5 stages, and its orthologous gene in *Arabidopsis* is *AT SPL9* (*AT2G42200*). Previous research showed that *AT SPL9* can regulate flowering time by promoting the transcription of MADS-box genes, including *FUL* (FRUITFULL), *SOC1* (SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1), and *AGL42* (AGAMOUS-LIKE) [28, 29]. *TraesCS3A02G284400* is a MIKC-type MADS-box gene, and its orthologous gene in rice is

*OsMADS32*. *OsMADS32* is uniquely expressed in the early stage of the inflorescence meristem, i.e., spikelet primordium, and it mainly maintains the characteristics of flower organs and regulates the development of rice flowers [30, 31]. Furthermore, we discovered that some MADS-type transcription factors were highly expressed in the leaf of both wild and mutant wheat, for example, *Vrn1-5A* (*TraesCS5A02G391700*), which indicates that this type of transcription factor can regulate flowering through different tissues, such as the leaf and apical meristem.

## Conclusions

Heading date is an important agronomic trait. Knowledge of its regulatory mechanism in wheat is limited. In this study, homozygous early and late heading genotypes from BC<sub>3</sub>F<sub>4</sub> backcrossed progenies of YZ4110 × *m605* were selected to conduct RNA-seq analysis. Based on the transcriptome analysis, we discovered 18,352 differentially expressed genes (DEGs), which belongs to 52 transcriptional factor gene families. Additionally, we identified 5 common GO terms and 214 common metabolic pathways. Through the weighted gene coexpression network analysis (WGCNA) of the DEGs, we found that 16, 336, 446, and 124 DEGs had biological functions associated with *Vrn1-5A*, *Vrn3-7B*, *Ppd-1D*, and *WSOC1*, respectively. In addition, we discovered one important flowering time gene, *TraesCS2D02G181400*. The above results will help to reveal the molecular mechanism of heading date regulation and facilitate the application of important flowering genes in wheat breeding.

## Methods

### Plant materials

*m605* is a late heading mutant that was identified from the ethyl methanesulfonate mutant library of YZ4110. Previous studies conducted in our laboratory identified *m605* as a late heading mutant compared with its wild-type parent YZ4110 [32]. To reveal the molecular mechanism of the genes that regulate the heading date, backcrossed BC<sub>3</sub>F<sub>4</sub> progeny of YZ4110 × *m605* was subjected to RNA-seq analysis, and the purpose of using the backcrossed segregating population was to reduce the false positive problems caused by the genetic background. Seeds of each wheat accession were planted in flowerpots and grown under 16 h light/22 °C and 8 h dark/18 °C in an artificial culture room with a long-day environment. The plants were transferred to a 4 °C environment for vernalization treatment for 4 weeks where they developed to the one-leaf stage and were then cultured in a greenhouse until they were sampled. A total of twenty-four wheat samples were selected, including the apex of the wild/mutant type at the W2.0/W3.5 stage (WHd-A2.0, WHd-A3.5, MHd-A2.0, MHd-A3.5) and leaves of the wild/mutant type at the W2.0/W3.5 stage (WHd-L2.0, WHd-L3.5, MHd-L2.0, MHd-L3.0).

### RNA extraction, library construction, and sequencing

The sampling period included the double ridge stage (W2.0) and androgynous primordium differentiation stage (W3.5) (Fig. S1), and the sampling position included the latest unfolding leaf and growth cone of

the main stem. We used clean, liquid nitrogen-frozen forceps to quickly clamp the growth cone and place it in an RNase-free centrifuge tube (the centrifuge tube was first filled with liquid nitrogen). Newly unfolded leaves of the main stem corresponding to the growth cone were mixed at the same time; the sampling position was 2 cm away from the tip of the leaf, and a cut was made 4 cm away from the tip of the leaf to obtain a tissue sample with a length of 2 cm. All samples were rapidly frozen in liquid nitrogen and stored in a -80°C environment. The TRIzol method was used to extract the RNA from each sample, the quality of RNA was detected by 1% agarose gel electrophoresis, and the concentration of RNA was detected by a Nanodrop 2000 instrument. High-quality RNA samples were sent to the ONMATH company (Chengdu, China) to construct the sequencing library following Illumina's standard pipeline. The constructed sequence libraries were sequenced with the Illumina HiSeq™ 4000 platform for 150 bp paired sequencing, and raw data were saved as fastq files. To remove the adaptors and lower quality reads at the head and end, we used Trimmomatic (version 0.36) [33] to conduct quality control of sequencing data with the following parameters (java -jar trimmomatic -0.35.jar PE -threads 20 forward.fastq reverse.fastq forward\_paired.fastq forward\_unpaired.fastq reverse\_paired.fastq reverse\_unpaired.fastq ILLUMINACLIP:1.adapter.list:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:1:10 MINLEN:50). FaQCs software (version 2.08) [34] was used to calculate the quality score, Q30 distribution, and GC distribution with the default parameters (FaQCs -1 R1.clean.fastq -2 R2.clean.fastq -d. -qc\_only). PCA of RNA-seq data was carried out by OmicShare tools (<http://www.omicshare.com/tools>).

### **Gene expression and analysis of differentially expressed genes**

RNA-seq analysis of the sequencing libraries was performed according to the method reported by Perter et al. [35]. (1) Hisat2 software (version 2.5.3a) was used to construct an index of the Chinese Spring reference genome (IWGSC RefSeq v1.1) with default parameters. (2) Raw reads were mapped to IWGSC RefSeq v1.1 by hisat2 with the parameters hisat2 -x reference.genome.index -p 8 -X 400 -no-unal -dta -1 input.R1.clean.fastq.gz -2 input.R2.clean.fastq.gz -S input.sam, and the mapping results of reads were stored in a bam file. (3) Stringtie software (version 1.3.3b) [36] was used to carry out transcript assembly of alignments with the following parameters: stringtie -e -p 16 -B -G reference.genome.gff3 input.bam -o input.gtf. (4) The R package Ballgown [35] was used to calculate the gene expression level, and we used FPKM (fragments per kilobase of transcript per million mapped reads) to measure the gene expression value at the whole-genome level. In addition, we used HTseq-count software (version 0.11.1) [37] to count the read number of each bam file with the following command line: htseq-count -format bam -order pos -mode union -stranded no -type exon input.bam reference.genome.gtf > reference.counts.txt. The count files of reads were used to calculate the differentially expressed genes by R package egdeR [38]. The evaluation criteria of differentially expressed genes were defined as an absolute log<sub>2</sub> (fold change) greater than 1 and a false discovery rate (FDR) less than 0.05 [39,40]. The distribution of DEGs along three wheat subgenomes was determined by chromPlot software [41].

### **Gene function annotation of DEGs**

Differentially expressed genes were submitted to the agrigo analysis toolkit [42,43] for GO enrichment analysis. Annotation information provided by the IWGSC reference genome (IWGSC RefSeq v1.1) was used as a reference database. We used the singular enrichment analysis method through Fisher's exact and Bonferroni multiple tests to screen significant GO terms. In addition, the differential expression spectrum was mapped to the metabolic regulation pathway in detail using MapMan (version 3.5.1) [44]. Considering that MapMan software lacks the mapping file from the Chinese Spring reference genome, we first extracted the coding sequence of differentially expressed genes and then uploaded this sequence to Mercator (<http://www.plabipd.de/portal/mercator-sequence-annotation>) for gene annotation and to obtain the corresponding mapping file. Then, the mapping file containing the gene expression value was imported into MapMan to analyze metabolic regulation [45,46]. To explore the transcriptome factor changes during wheat heading date development, we extracted the protein sequences of DEGs from IWGSC RefSeq v1.1 and used the iTAK [47] pipeline (<http://bioinfo.bti.cornell.edu/tool/itak>) for transcriptome factor prediction and functional classification.

### **Weighted gene coexpression network analysis and flowering time regulatory network construction**

The R package WGCNA [48, 49] was used to construct the flowering time coexpression network and identify candidate flowering time genes. First, the “goodSamplesGenes” function was used to check the quality of the gene expression data, and the “pickSoftThreshold” function was used to determine the soft power. Second, according to the soft power value, we calculated the “adjacency matrix” and “topological overlap matrix”. Third, the initial gene expression matrix was obtained by hierarchical clustering of the dissimilarity matrix with the “hclust” function with the default parameters (minModuleSize: 30, networkType: “signed”, TOMType: “unsigned”, mergeCutHeight: 0.25). Based on the connectivity between coexpressed genes, Cytoscape software (version 3.7.2) [50] was used to visualize the flowering-related gene regulation network.

## **Abbreviations**

DEG, differentially expressed genes; RNA-seq, RNA sequencing; GO terms, Gene ontology terms; WGCNA, Weighted gene co-expression network analysis; EMS, ethyl methanesulfonate; PCA, principal component analysis; FPKM, Fragment per kilobase of transcripts effective length per million fragments mapped to all transcripts; AGEV: Gene expression values; SD: Standard deviation; CV: coefficient of variation.

## **Declarations**

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

## Availability of data and materials

The datasets used and analyzed during the current study available from the corresponding authors on reasonable request.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work was supported by the National Key Research and Development Program of China (2016YFD0101802), the National Natural Science Foundation of China (31371619), the National Transgenic Research Project (2016ZX08009001001-004) and Talent Program and Innovation Program of CAAS.

## Authors' contributions

YY, XZ, XL and XK designed the research. YY, XZ, GL, LZ and CX performed the experiments. YY and XK wrote the paper. All the authors have read the manuscript and approve it.

## Declarations of Competing Interest

The authors declare no conflicts of interest.

## Acknowledgement

The authors thank all who have contributed to this study.

## References

1. Distelfeld A, Li C, Dubcovsky J. Regulation of flowering in temperate cereals. *Current Opinion in Plant Biology*. 2009;12(2):178–
2. Shi C, Zhao L, Zhang X, Lv G, Pan Y, Chen F. Gene regulatory network and abundant genetic variation play critical roles in heading stage of polyploidy wheat. *BMC Plant Biology*. 2019;19(1):6.
3. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. *Proc Natl Acad Sci U S A*. 2003;100(10):6263–6268.
4. Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science*. 2004;303(5664):1640–1644.
5. Yan L, Fu D, Li C, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J. The wheat and barley vernalization gene *VRN3* is an orthologue of FT. *Proc Natl Acad Sci U S A*. 2006;103(51):19581–19586.

6. Kippes N, Zhu J, Chen A, Vanzetti L, Lukaszewski A, Nishida H, Kato K, Dvorak J, Dubcovsky J. Fine mapping and epistatic interactions of the vernalization gene *VRN-D4* in hexaploid wheat. *Molecular Genetics and Genomics*. 2014;289(1):47–62.
7. Dubcovsky J, Loukoianov A, Fu D, Valarik M, Sanchez A, Yan L. Effect of photoperiod on the regulation of wheat vernalization genes *VRN1* and *VRN2*. *Plant Mol. Biol.* 2006;60(4):469–
8. Chen A, Dubcovsky J. Wheat TILLING mutants show that the vernalization gene *VRN1* down-regulates the flowering repressor *VRN2* in leaves but is not essential for flowering. *PLoS Genetics*. 2012;8(12):e1003134.
9. Turner A, Beales J, Faure S. The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. *Science*. 2005;310(5750):1031–1034.
10. Beales J, Turner A, Griffiths S, Snape JW, Laurie DA. A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum*). *Theor. Appl. Genet.* 2007;115(5):721–733.
11. Boden SA, Cavanagh C, Cullis BR, Ramm K, Greenwood J, Jean FE, Trevaskis B, Swain SM. *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nature Plants*. 2015;1(2):14016.
12. Shaw LM, Turner AS, Laurie DA, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. The impact of photoperiod insensitive *Ppd-1a* mutations on the photoperiod pathway across the three genomes of hexaploid wheat (*Triticum aestivum*). *The Plant Journal*. 2012;71(1):71–84.
13. Nishida H, Yoshida T, Kawakami K. Structural variation in the 5' upstream region of photoperiod insensitive alleles *Ppd-A1a* and *Ppd-B1a* identified in hexaploid wheat (*Triticum aestivum*) and their effect on heading time. *Mol Breeding*. 2013;31(1):27–37.
14. Shitsukawa N, Ikari C, Mitsuya T. Wheat SOC1 functions independently of WAP1/VRN1, an integrator of vernalization and photoperiod flowering promotion pathways. *Physiologia Plantarum*. 2007;130(4):627–636.
15. Shitsukawa N, Takagishi A, Ikari C, Takumi S, Murai K. *WFL*, a wheat *FLORICAULA/LEAFY* ortholog, is associated with spikelet formation as lateral branch of the inflorescence meristem. *Genes Genet Syst.* 2006;81(1):13–20.
16. Waddington SR, Cartwright PM, Wall PC. A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany*. 1983;51(1):119–130.
17. Peng FY, Hu Z, Yang R. Genome-Wide Comparative analysis of flowering-related genes in *Arabidopsis*, wheat, and barley. *International Journal of Plant Genomics*. 2015;2015:1–17.
18. Liu J, Xu Z, Fan X, Zhou Q, Cao J, Wang F, Ji G, Yang L, Feng B, Wang T. A genome-wide association study of wheat spike related traits in China. *Frontiers in Plant Science*. 2018;9:1584.
19. Tang C, Li M, Cao M, Lu R, Zhang H, Liu C, Huang S, Zhang P, Hua H, Zhao W. Transcriptome analysis suggests mechanisms for a novel flowering type: Cleistogamous wheat. *The Crop Journal*. 2020;8(2):313–326.

20. Wahl V, Ponnu J, Schlereth A. Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. *Science*. 2013;339(6120):704–707.
21. Ponnu J, Wahl V, Schmid M. Trehalose-6-phosphate: connecting plant metabolism and development. *Frontiers in Plant Science*. 2011;2:70.
22. Ananyev G, Gates C, Kaplan A, Dismukes GC. Photosystem II-cyclic electron flow powers exceptional photoprotection and record growth in the microalga *Chlorella ohadii*, *Biochimica et Biophysica Acta (BBA)-Bioenergetics*. 2017;1858(11):873–883.
23. Goslin K, Zheng B, A SM, Rae L, Ryan PT, Kwaśniewska K, Thomson B, Ó'Maoiléidigh DS, Madueño F, Wellmer F, Graciet E. Transcription factor interplay between LEAFY and APETALA1/CAULIFLOWER during floral initiation. *Plant Physiology*. 2017;174(2):1097–1109.
24. Kobayashi K, Yasuno N, Sato Y, Yamazaki R, Kimizu M, Yoshida H, Nagamura Y, Kyozuka J. 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. *The Plant Cell*. 2012;24(5):1848–1859.
25. Yu Y, Qiao L, Chen J, Rong Y, Zhao Y, Cui X, Xu J, Hou X, Dong C. *Arabidopsis* REM16 acts as a B3 domain transcription factor to promote flowering time via directly binding to the promoters of *SOC1* and *FT*. *The Plant Journal*. 2020.
26. Hassidim M, Harir Y, Yakir E, Kron I, Green RM. Over-expression of *CONSTANS-LIKE 5* can induce flowering in short-day grown *Arabidopsis*. *Planta*. 2009;230(3):481–491.
27. Nagashree KL, Ahmed MF. Electrocatalytic oxidation of methanol on Pt modified polyaniline in alkaline medium. *Synthetic Metals*. 2008;158(15):610-616.
28. Wu G, Poethig RS. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target *SPL3*. *Development*. 2006;133(18):3539–
29. Yamaguchi A, Wu M, Yang L, Wu G, Poethig RS, Wagner D. The microRNA-regulated SBP-Box transcription factor *SPL3* is a direct upstream activator of *LEAFY*, *FRUITFULL*, and *APETALA1*. *Developmental Cell*. 2009;17(2):268–
30. Wang H, Zhang L, Cai Q, Hu Y, Jin Z, Zhao X, Fan W, Huang Q, Luo Z, Chen M, et al. *OsMADS32* interacts with PI-like proteins and regulates rice flower development. *Journal of Integrative Plant Biology*. 2015;57(5):504–
31. Sang X, Li Y, Luo Z, Ren D, Fang L, Wang N, Zhao F, Ling Y, Yang Z, Liu Y. *CHIMERIC FLORAL ORGANS1*, encoding a monocot-specific MADS box protein, regulates floral organ identity in rice. *Plant Physiology*. 2012;160(2):788–
32. Zhang X, Liu G, Zhang L, Xia C, Zhao T, Jia J, Liu X, Kong X. Fine mapping of a novel heading date gene, *TaHdm605*, in hexaploid wheat. *Frontiers in Plant Science*. 2018;9:1059.
33. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*. 2014;30(15):2114–
34. Lo CC, Chain PSG. Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*. 2014;15(1):366.

35. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*. 2016;11(9):1650.
36. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*. 2015;33(3):290.
37. Anders S, Pyl PT, Huber W. HTSeq - a Python framework to work with high-throughput sequencing data, *Bioinformatics*. 2015;31(2):166–169.
38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.
39. Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Research*. 1997;7(10):986–
40. Mariani TJ, Budhraj V, Mecham BH, Gu CC, Watson MA, Sadovsky Y. A variable fold change threshold determines significance for expression microarrays. *The FASEB Journal*. 2003;17(2):321–323.
41. Oróstica KY, Verdugo RA. chromPlot: visualization of genomic data in chromosomal context, *Bioinformatics*. 2016;32(15):2366–
42. Du Z, Zhou X, Ling Y, Zhang Z, Su Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research*. 2010;38(suppl\_2):W64–
43. Tian T, Liu Y, Yan H. agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*. 2017;45(W1):W122–W129.
44. Thimm O, Bläsing O, Gibon Y. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal*. 2004;37(6):914–939.
45. Klie S, Nikoloski Z. The choice between MapMan and Gene Ontology for automated gene function prediction in plant science, *Frontiers in Genetics*. 2012;3:115.
46. Chandran AKN, Lee GS, Yoo YH, Yoon UH, Ahn BO, Yun DW, Kim JH, Choi HK, An G, Kim TH, et al. Functional classification of rice flanking sequence tagged genes using MapMan terms and global understanding on metabolic and regulatory pathways affected by dxr mutant having defects in light response. *Rice*. 2016;9(1):1–12.
47. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases, *Molecular Plant*. 2016;9(12):1667–1670.
48. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*. 2005;4(1).
49. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559.
50. Shannon P, Markiel A, Ozier O. Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Research*. 2003;13(11):2498–

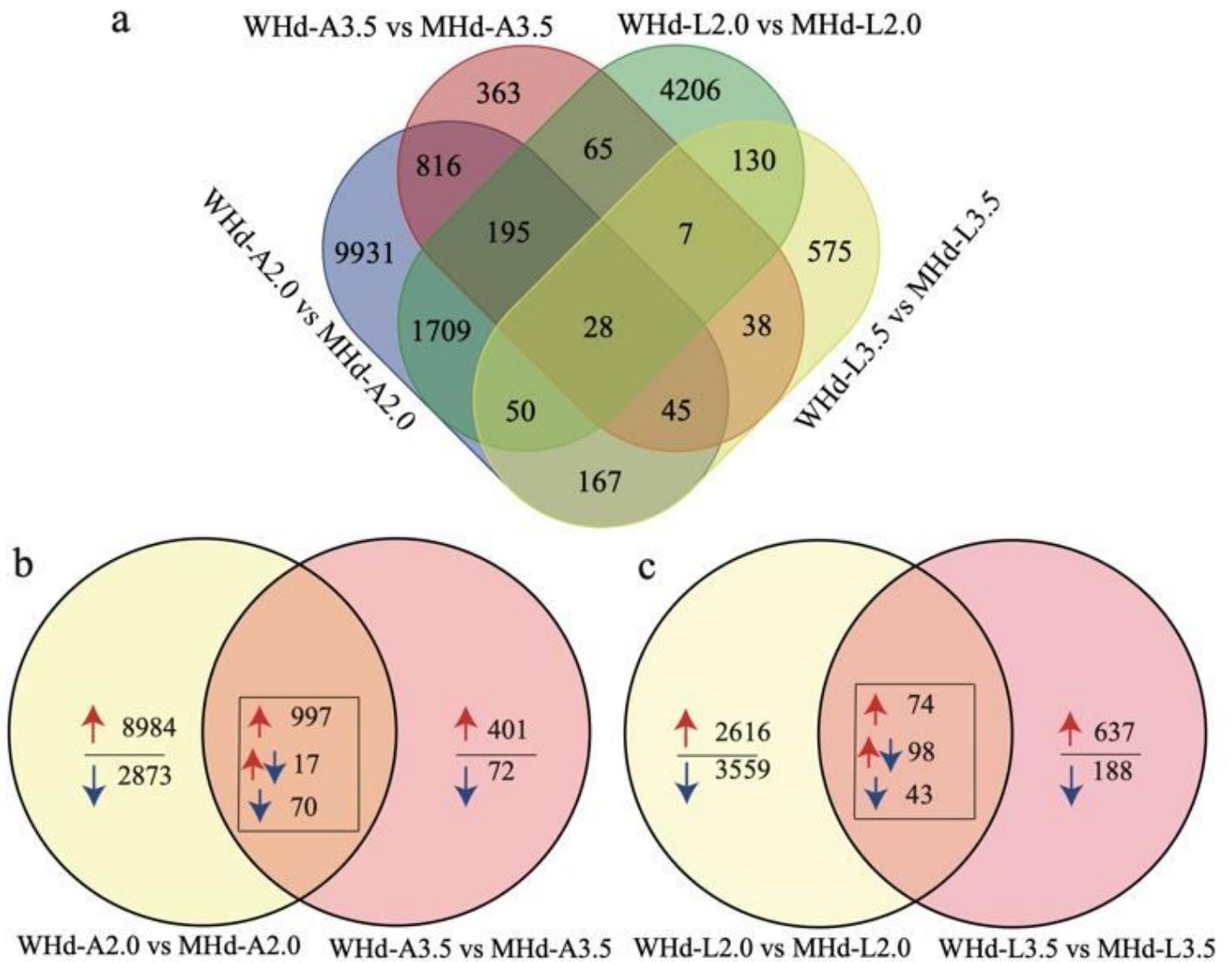
# Tables

**Table 1** Functional annotation of genes coexpressed with flowering genes.

Flowering gene	Candidate flowering time genes	Position in IWGSC1.1/Mb	Function annotation	Log2(FC)	p-value
<i>Vrn1-5A</i>	<i>TraesCS2D02G181400</i>	2D: 126.124 ~ 126.149	MADS-box transcription factor	4.72	3.90 × 10 <sup>-18</sup>
<i>Vrn3-7B</i>	<i>TraesCS5A02G166100</i>	5A: 355.071 ~ 355.072	CONSTANS-like protein	3.66	8.66 × 10 <sup>-5</sup>
	<i>TraesCS2D02G147100</i>	2D: 88.087 ~ 88.089	Gibberellin receptor GID1A	3.88	3.05 × 10 <sup>-5</sup>
	<i>TraesCS5D02G170700</i>	5D: 267.762 ~ 267.764	CONSTANS-like protein	4.16	3.56 × 10 <sup>-5</sup>
<i>Ppd-1D</i>	<i>TraesCS1A02G220300</i>	1A: 389.129 ~ 389.131	CONSTANS-like protein	-2.39	6.10 × 10 <sup>-4</sup>
	<i>TraesCS1B02G326800</i>	1B: 553.024 ~ 553.025	Early light-induced protein	-3.15	1.27 × 10 <sup>-5</sup>
	<i>TraesCS1B02G326900</i>	1B: 553.135 ~ 553.136	Early light-induced protein	-7.17	1.02 × 10 <sup>-12</sup>
	<i>TraesCS1A02G314800</i>	1A: 506.473 ~ 506.476	Early light-induced protein	-4.09	3.21 × 10 <sup>-3</sup>
<i>WSOC1</i>	<i>TraesCS3A02G143100</i>	3A:124.172 ~ 124.176	Flowering locus T	2.74	1.69 × 10 <sup>-7</sup>
	<i>TraesCS2B01G200800</i>	2B: 180.040 ~ 180.060	MADS-box transcription factor	2.43	8.77 × 10 <sup>-7</sup>
	<i>TraesCS3D01G144500</i>	3D: 132.976 ~ 132.977	Flowering locus T	2.28	2.95 × 10 <sup>-4</sup>
	<i>TraesCS3B02G162000</i>	3B: 158.715 ~ 158.719	Flowering locus T	1.13	8.54 × 10 <sup>-4</sup>

<i>TraesCS5A02G515500</i>	5A: 747.807 ~ 757.809	MADS-box transcription factor	1.39	8.45 × 10 <sup>-4</sup>
<i>TraesCS3D01G144500</i>	3D: 132.976 ~ 132.977	Flowering locus T	2.28	2.95 × 10 <sup>-4</sup>

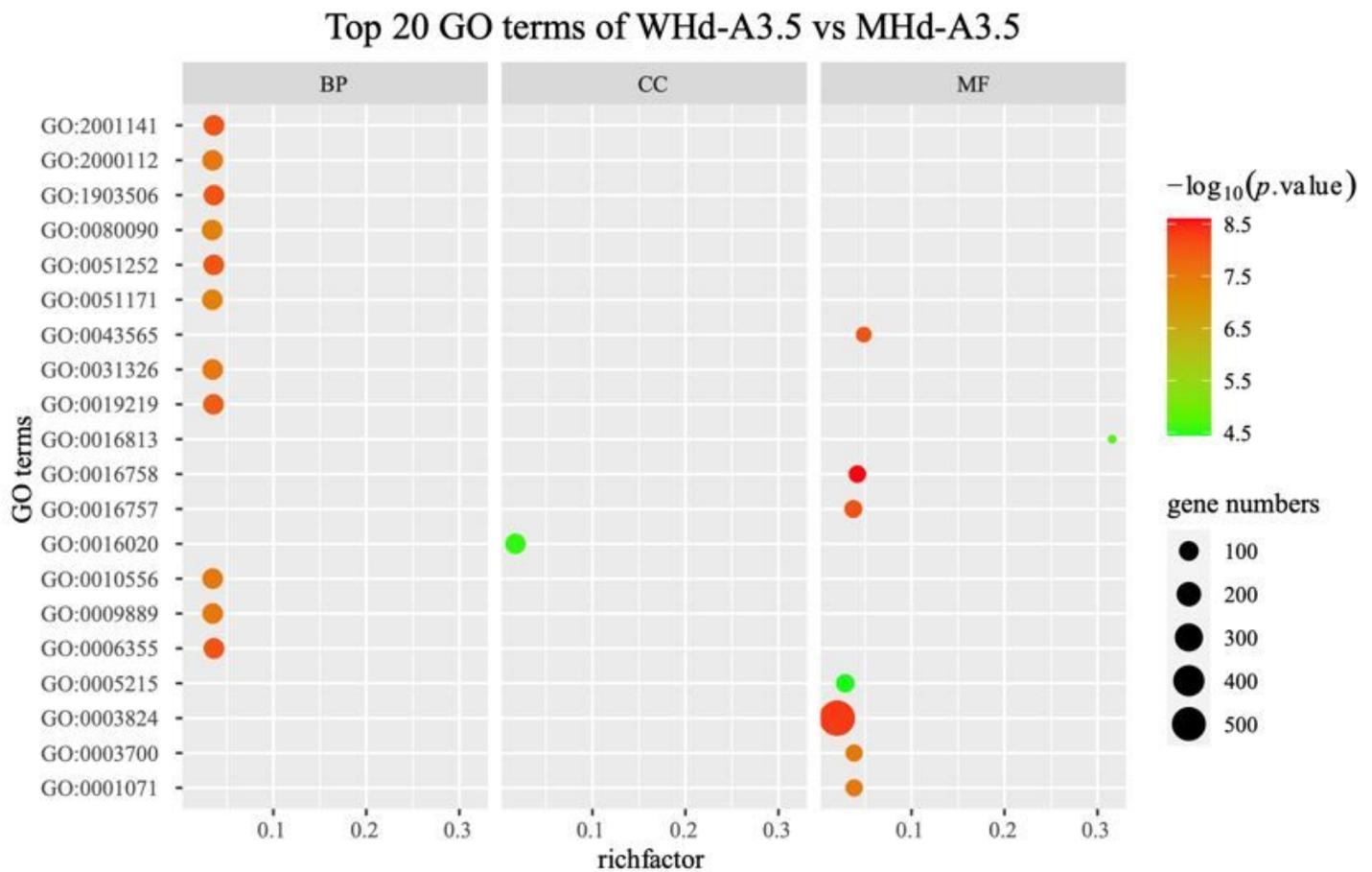
## Figures



**Figure 1**

Venn diagram of overall differentially expressed genes. (a) The DEGs between WHd and Mhd in different tissues and development periods. (b) Common DEGs in the apex of WHd vs. Mhd between W2.0 and

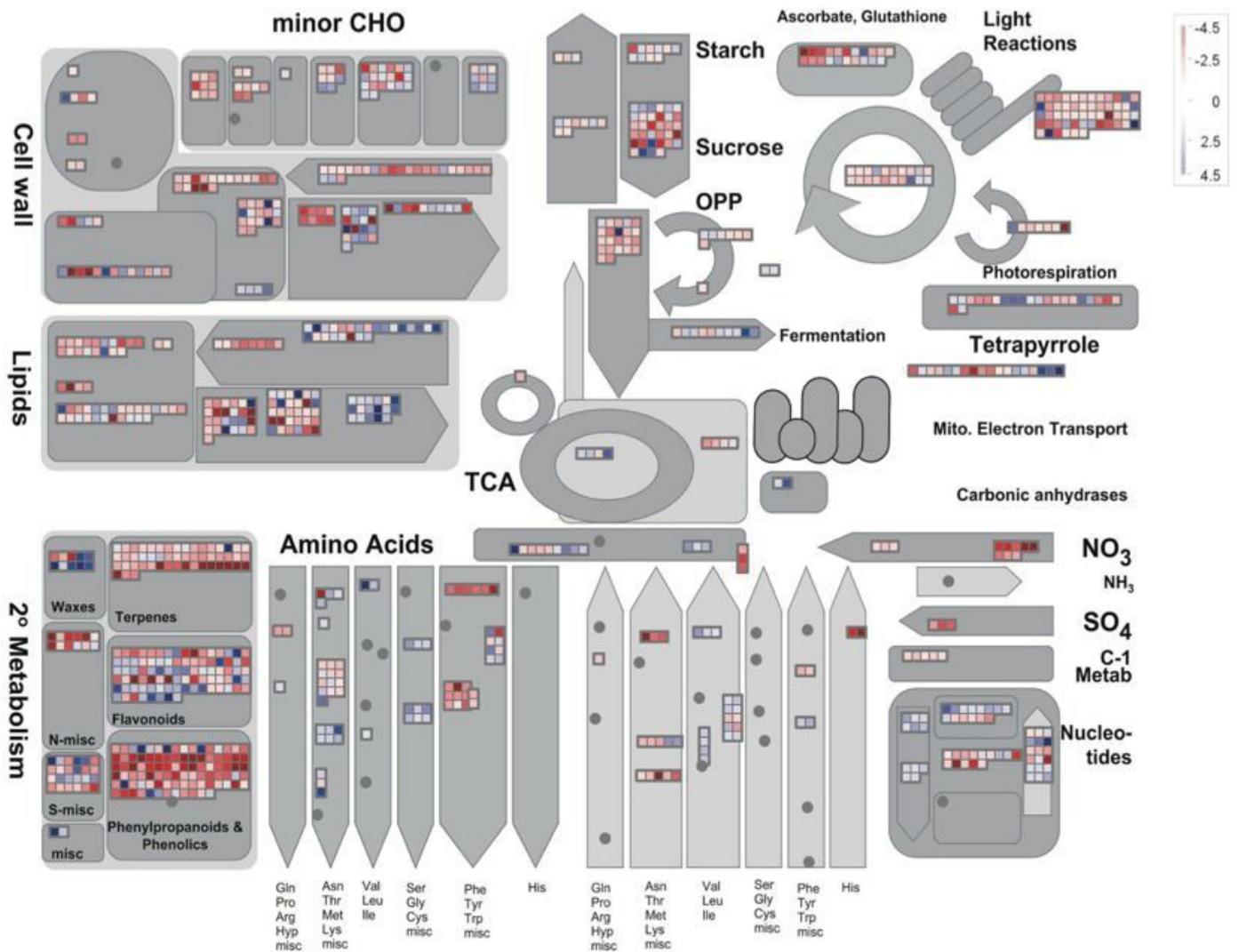
W3.5. (c) Common DEGs in leaves of WHd vs. MHd between W2.0 and W3.5. Note: the red arrow indicates upregulated genes, and the blue arrow indicates downregulated genes.



**Figure 2**

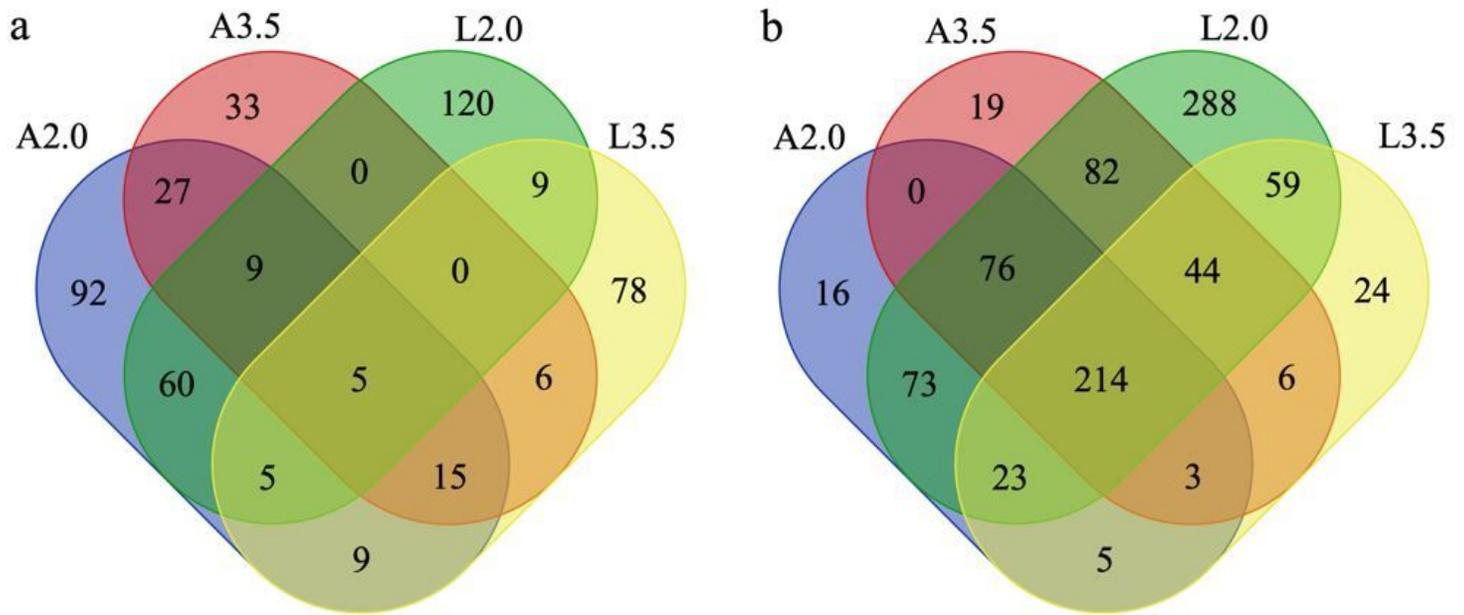
GO enrichment analysis of DEGs between WHd-A3.5 and MHd-A3.5. Each row corresponds to a significant GO term, and columns represent the rich factor (gene numbers of differentially expressed genes enriched in the pathway/all gene numbers in the background gene set). The bubble size represents the gene numbers, and the color gradient represents the  $-\log_{10}(p\text{-value})$ .

## MapMan analysis of WHd-L2.0 vs. MHd-L2.0



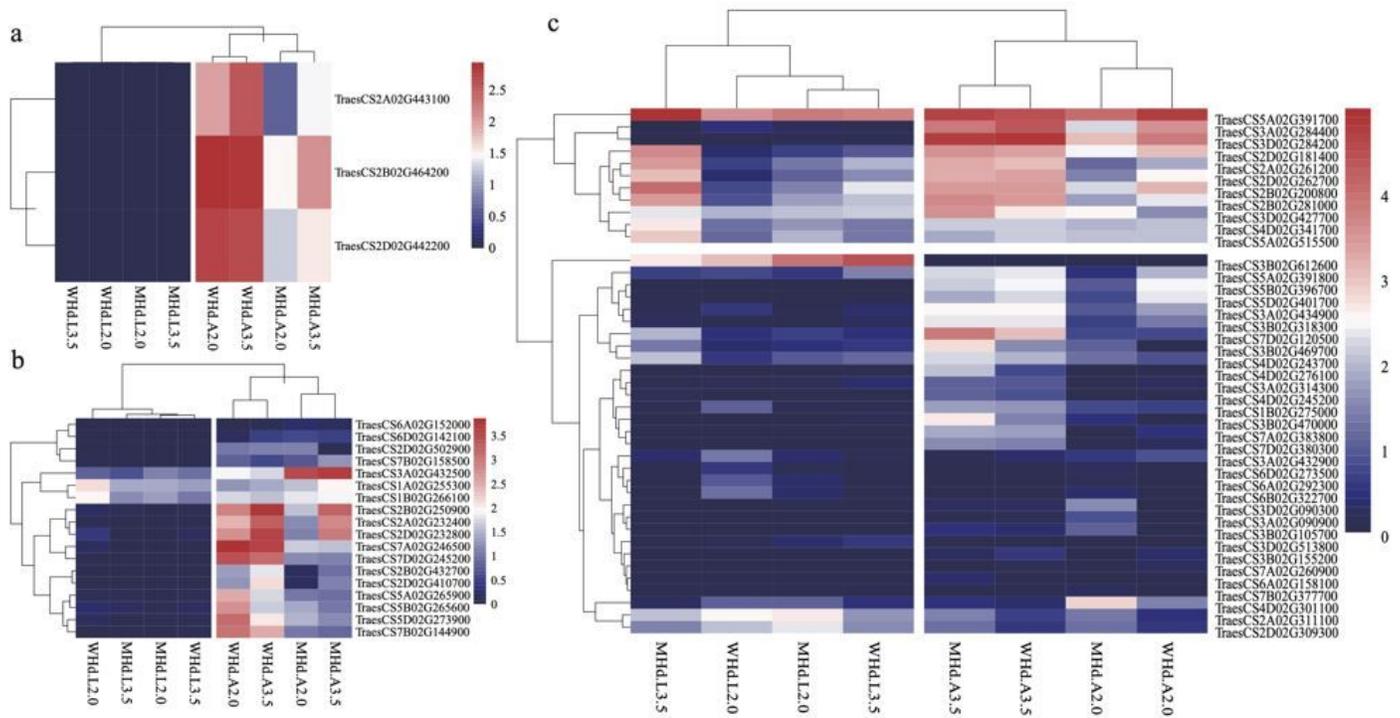
**Figure 3**

An overview of the metabolic pathway of differentially expressed genes between WHd-L2.0 and MHd-L2.0. Each inset presents a differentially expressed gene. The red lattice represents the upregulated genes, and the blue lattice represents the downregulated genes. The color scale presents the fold change value of DEGs.



**Figure 4**

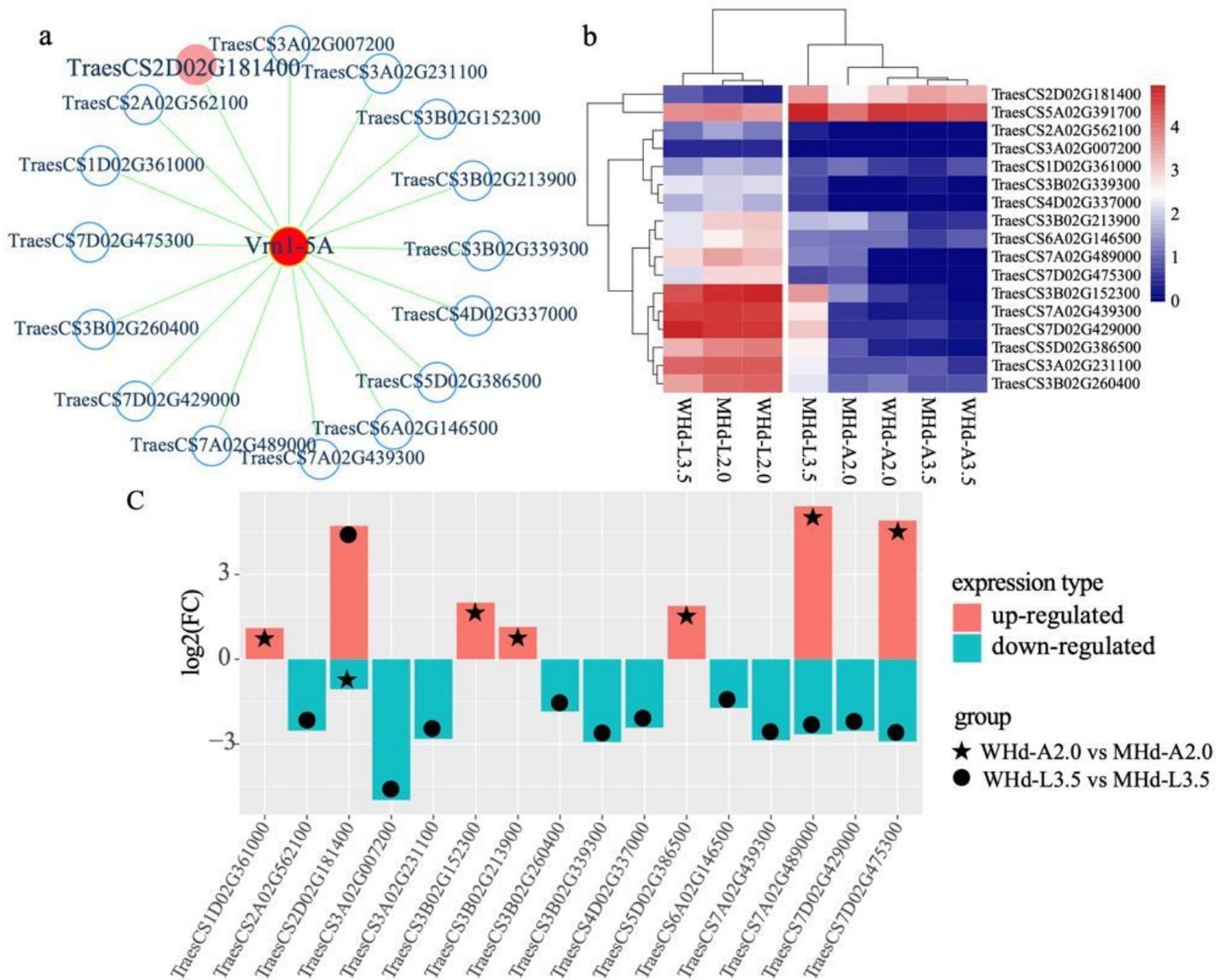
Venn plot showing the overlap of the number of significant regulatory pathways. A2.0: WHd-A2.0 vs. MHD-A2.0, A3.5: WHd-A3.5 vs. MHD-A3.5, L2.0: WHd-L2.0 vs. MHD-L2.0, L3.5: WHd-A3.5 vs. MHD-A3.5. (a) Overlap of the significant GO numbers. (b) Overlap of the significant MapMan metabolic pathways.



**Figure 5**

The gene expression level of three transcription factor gene families. (a) The LFY transcription factor gene family. (b) The SBP transcription factor gene family. (c) The MADS-box transcription factor gene

family.



**Figure 6**

Construction of the flowering time regulatory network and expression level of the differential genes. The red nodes in the network indicate high-confidence genes involved in the heading date. (a) The Vrn1-5A flowering time regulatory network. (b) The expression heatmap of differentially expressed genes that were coexpressed with Vrn1-5A.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile.docx](#)