Network and epigenetic characterization of subsets of genes specifically expressed in maize bundle sheath cells

Shentong Tao, Wenli Zhang *

State Key Laboratory for Crop Genetics and Germplasm Enhancement, Collaborative Innovation Center for Modern Crop Production co-sponsored by Province and Ministry (CIC-MCP), Nanjing Agricultural University, No.1 Weigang, Nanjing, Jiangsu 210095, PR China

A R T I C L E   I N F O

A B S T R A C T

Bundle sheath (BS) cells exhibit dramatically structural differences and functional variations at physiological, biochemical and epigenetic levels as compared to mesophyll (M) cells in maize. The regulatory mechanisms controlling functional divergences between M and BS have been extensively investigated. However, BS cell-related regulatory networks are still not completely characterized. To address this, we herein conducted WGCNA-related co-expression assays using bulk M and BS cell RNA-seq data sets and identified a module containing 384 genes highly expressed in BS cells (including 20 hub TFs) instead of M cells. According to the hub TF centered regulatory network, we found that *Dof22* and *Dof30* might act as key regulators in the regulation of expression of BS-specific genes, and several *MYB* TFs exhibited a high collaboration with *Dof* TFs. By comparing the enrichment levels of histone modifications, we found that genes in the aforementioned module were more enriched with histone acetylation as compared to other BS-enriched DEGs with similar expression levels. Moreover, we found that a subset of genes functioning in photosynthesis, protein auto processing and enzymatic activities were significantly enriched with broad H3K4me3. Thus, our study provides evidence showing that regulatory network and histone modifications may play vital roles in the regulation of a subset of genes with important functions in BS cells.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Maize is a typical C_4 crop, which is important to secure food production for demanding due to increasing population worldwide, especially facing dramatic impacts of global warming on agricultural systems. C_4 crops tend to be more productive, especially in arid areas, largely due to their higher efficiency in utilization of nutrient, water and photosynthesis as compared to C_3 crops [1]. Maize and other C_4 crops have two structurally and functionally specialized cell types, termed as mesophyll (M) and bundle sheath (BS) cells, which form a "Kranz" anatomy surrounding the veins [2]. The underlying mechanisms determining functional divergences between M and BS cells have been extensively investigated in maize. Briefly, functional variations between both cells occur in multiple levels, such as physiological, biochemical and epigenetic levels. For example, C_4 acids are initially synthesized from CO_2, which is catalyzed by phosphoenolpyruvate carboxylase (PEPC) in M cells, and are then transported to BS cells to release

CO_2 again for carboxylation through stepwise enzymatic reactions [3,4]. Therefore, C_4 enzymes working in close collaboration between two cell types ensure the achievement of high photosynthesis efficiency in C_4 plants [1,5]. Quantitative proteomics assays showed that the majority of starch metabolic enzymes are more abundant in BS cells than in M cells [6]. PEPC kinase (PEPCk), which is preferentially expressed in M cells, can specifically regulate the activity of PEPC through reversible phosphorylation [7]. Transcriptomic analyses revealed that a subset of genes differentially expressed between M and BS cells may play vital roles in cell-type related functional differentiation in maize [8–11]. BS and M cell related DNase I-seq and ATAC-seq assays further indicated that differential chromatin openness causes variations in the accessibility of underlying *cis*-regulatory elements (CREs) to trans-acting factors, leading to differential expression of related genes between two cell types [11,12]. Moreover, ChIP-seq assays provided evidence showing that C_4 genes between two cell types in maize can be differentially regulated by chromatin modifications such as H3K9ac, H3K27mes and H3K4me3 [11,13]. For example, *PEPC*, *PPDK*, *PCK*, *ME* and *CA* are potentially regulated by the secondary upstream peak (R-SUP) of H3K9ac, while *NADP-ME*, *PEPC*, *Rbcs1*

* Corresponding author.
  E-mail address: wzhang25@njau.edu.cn (W. Zhang).

and *Rbcs2* are potentially regulated by H3K27me3 and H3K4me3 in a cell-specific manner.

Transcription factors (TFs) have been found to regulate several key C_4 genes in a cell-type dependent manner in maize. It has been documented that *Dof* TFs play both active and repressive roles in the regulation of several C_4 gene expression like *PPDK* and *PEPC* [14,15]. It has been reported that some of TF genes are preferentially expressed and function in BS cells [8,9,16]. Some of bHLH family can preferentially bind to *NADP-ME* and *PEPC*. For instance, bHLH80 can bind to the promoter of *PEPC*, resulting in cell-specific expression [17,18]. *ZmOrphan94* highly expressed in BS cells can act as a repressor for *PEPC* in maize protoplasts, indicative of cell-type dependent functions of *ZmOrphan94* [19]. Variations in the accessibility of underlying *cis*-regulatory elements (CREs) between the two cell types can cause cell-type related transcription of some of key C_4 genes like *PEPC*, *CA* and *NAD-ME* [20–22]. Similarly, it has been reported that TFs in the MYB-MYC module can bind to the promoter CREs of some genes, resulting in their preferential expression in bundle sheath of *Arabidopsis* [23]. Furthermore, functions of TFs in the regulation of cell-dependent gene expression have been comprehensively characterized through regulatory network analyses using single M cell RNA-seq data, leading to discoveries that several *CO-like* TFs possibly play dominant roles in leaf development and photosynthesis for M cells in maize [16,24]. However, BS cell-related regulatory networks are still not completely characterized, especially it is hard to be achieved at a single cell level due to technique limiting for isolation of enough BS cells for sequencing as compared to single M cells [16,25]. To address this question, we here conducted WGCNA-related co-expression assays using bulk M and BS cell RNA-seq data sets and identified a co-expression module containing genes highly expressed in BS cells instead of M cells. TFs in the module were further used for construction of TF centered regulatory network using GENIE3 and TF-CollaborativeNet methods. A possible epigenetic regulation of these genes specifically expressed in BS cells was also investigated by analyzing histone modification ChIP-seq data sets.

2. Results

2.1. Transcriptomic analyses of mesophyll and bundle sheath cells in maize

To explore transcriptional differences between two distinct photosynthetic cell types, M and BS cells, in maize, we isolated M and BS cells using the second and third leaves collected from 10-day-old plants. The purity of isolated M and BS cells was assessed using RT-PCR for *PEPC* (*Zm00001d046170*) and *rbcS* (*Zm00001d052595* and *Zm00001d004894*) genes (Fig. S1), which are key C_4 marker genes primarily expressed in M and BS cells, respectively. We then conducted two well-correlated biological replicates of RNA sequencing (RNA-seq) (R^2 = 0.95 for M cells and 0.99 for BS cells, Fig. S2) and obtained 73.7 and 85.3 M total reads for M and BS cells, respectively. Among 46,430 genes annotated in maize B73 reference genome (maize B73 v4), we identified 21,730 (*ca.* 47%) in M cells and 19,915 (*ca.* 43%) in BS cells expressed genes with FPKM > 0.5. After comparing the expression levels of all genes between M and BS cells, we identified 5,748 (12.38%) and 4,004 (8.62%) differentially expressed genes (DEGs) in M and BS cells, respectively, with fold changes > 4 (Fig. 1A). The majority of DEGs were present in both replicates (Fig. 1B). To confirm the representative of these identified DEGs, we chose 10 key genes responsible for C_4 photosynthesis, 5 of them highly expressed in M or BS cells, for comparisons, which have been found to be specifically expressed in M or BS cells before. We found that the data in our study were comparable with previously published

data sets (Fig. 1C). Moreover, we randomly selected 10 DEGs (5 of them highly expressed in M or BS cells) for RT-qPCR assays, and found that the expression trend of these DEGs matched well with the RNA-seq data we analyzed (Fig. 1D). Therefore, these analyses indicated that the newly generated data were reliable for downstream analyses.

Functional divergences of photosynthesis between M and BS cells have already been well characterized in maize. To assess biological implications of DEGs we identified, we conducted GO term enrichment analyses (Fig. 1E). We found that, in addition to some of GO terms shared in both cells, like responses to endogenous or biotic stimuli and secondary metabolic processes, BS-cell related DEGs were mainly enriched in GO terms associated with functions in thylakoid, plastid, photosynthesis, generation of precursor metabolites and energy, carbohydrate metabolic processes and anatomical structure development, by contrast, M-cell related DEGs were overrepresented in GO terms associated with functions in signal transduction, responses to stress and cell communication.

2.2. Identification of genes specifically expressed in BS cells through co-expression analyses

To identify genes specifically expressed in BS cells, we pooled 28 published BS and M cell related RNA-seq data (Supplementary Table S1) for Weighted Gene Co-Expression Network Analysis (WGCNA) [26]. We obtained 23 co-expression modules (Fig. S3), which were marked using different colors. Except for non-effectively grouped genes listed in the gray module, the number of genes in each module ranged from 33 (darkgreen) to 3208 (turquoise) (Supplementary Table S1). To identify genes specifically expressed in M or BS cells, we conducted correlation analyses between gene modules and M/BS cells (Fig. 2A). We found that darkgreen (R^2 = 0.83), magenta (R^2 = 0.87) and grey60 (R^2 = 0.70) modules exhibited a high correlation with BS cells. According to the module eigengene values (ME, defined as the first principal component of genes in a module), which represent the overall expression pattern of genes in the module, we found that the eigengene expression level in the magenta module was higher in the majority of BS cell related RNA-seq data, but much lower in the corresponding M cell related RNA-seq data (Fig. 2B). Therefore, we considered 384 genes in the magenta module as potential candidates for genes specifically expressed in BS cells. Moreover, we found that *RBCS1*, *RBCS2*, *NAPD-ME*, *PCK* and *MEP3* genes involved in the BS cell photosynthesis were also present in this module (Fig. 2C), suggesting that this module may be primarily related to functions of BS cells. We further conducted GO enrichment analyses, and found that 384 genes in the magenta module were mainly involved in both biological processes and cellular component, such as generation of precursor metabolites and energy, carbohydrate metabolic processes and organelle envelop (Fig. 2D). To identify potential hub TFs in the magenta module, we further screened all genes using the module membership (KME) values calculated by the WGCNA package (Fig. S4). According to the size of |KME| representing the core degree of a gene in a module, we identified 20 hub TFs, including *Dof*, *MYB*, *bZIP* and *G2-like* families (Fig. 2E).

2.3. Hub TFs related regulatory network in BS cells

To investigate how these hub TFs individually or in cooperation with other TFs or genes function in BS cells, we identified TF modules using the TF-CollaborativeNet methods [27]. TF-CollaborativeNet can divide TFs into subgroups based on the overlaps of top 100 confident target genes between TFs. We identified a total of 114 TF subclusters, and 13 of which were composed of > 10 genes (Fig. 3A). Of 20 TFs in the aforementioned magenta module, 12 TFs were present in subcluster 1 and 4 TFs were present in sub-
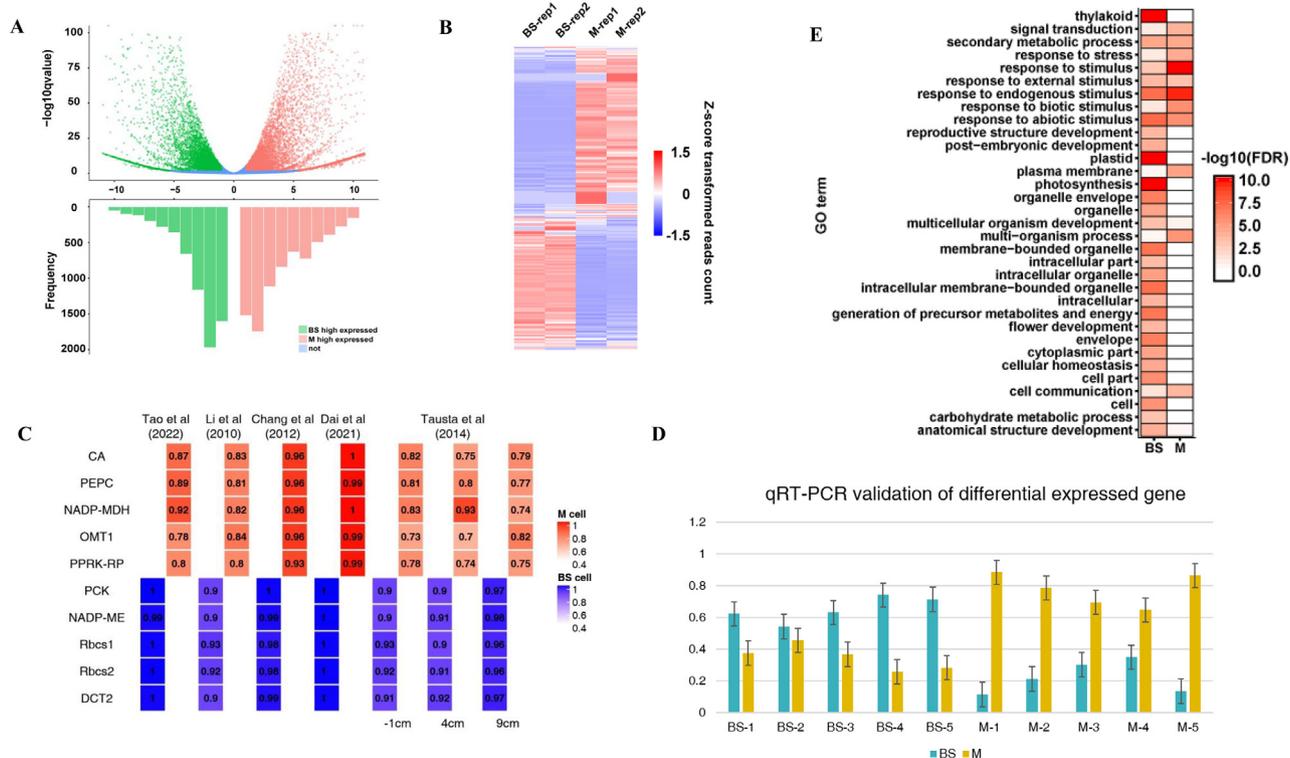
Fig. 1. Reanalyses of our RNA-seq data. A) Volcano plot showing the differentially expressed genes (DEGs) between BS and M cells. B) Heatmap showing variations in then gene expression levels between BS and M cells. C) Heatmap showing the Ri value for C4 genes between BS and M cells for our and published data. D) Validation of 10 randomly selected DEGs using qRT-PCR. E) GO enrichment analyses for BS and M-specific DEGs.
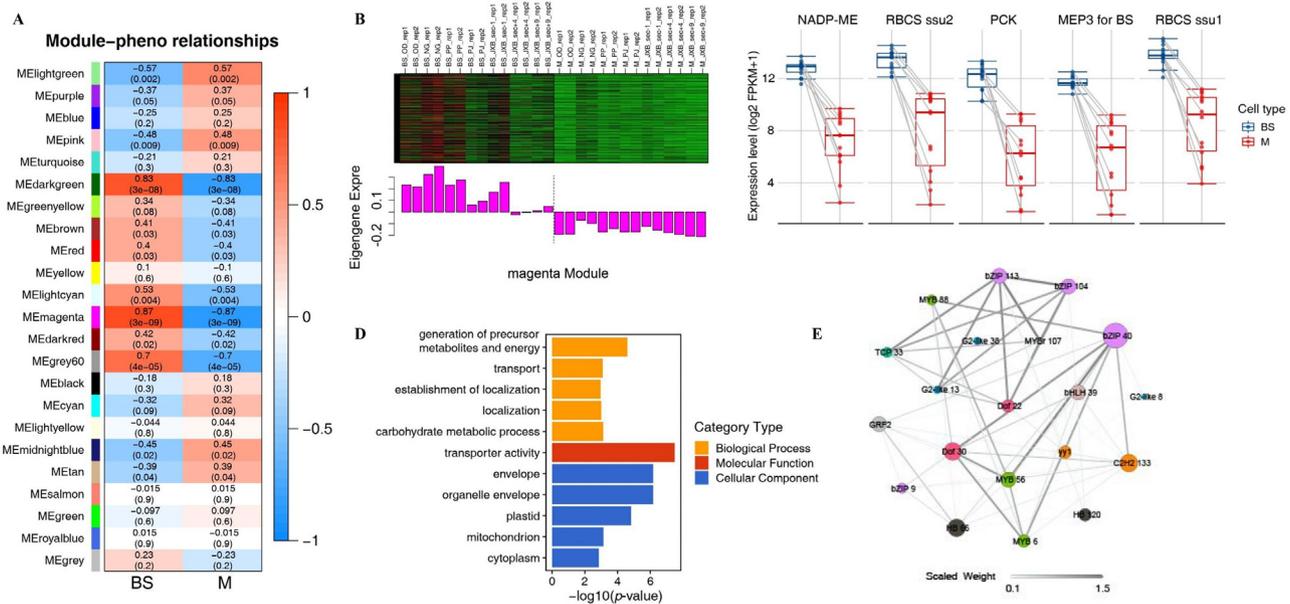
Fig. 2. Co-expression analyses using published BS and M cell RNA-seq data. A) Heatmap showing the correlation for gene module and cell types. B) Barplot combined with heatmap showing the Eigengene values for the BS-specific magenta module. C) Boxplot showing the FPKM values for key C4 genes in the magenta module between BS and M cells. D) Barplot showing the enriched GO terms for genes in the magenta module. E) Co-expression network of hub genes in the magenta module, both size and color for edges represent the weight of the link. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cluster 3 (Fig. 3B), indicating a high collaboration among these TFs. They included *Dof30 and bZIP9* and *bZIP40* TFs, which have been reported to function in BS cells [11,13,28,29]. Moreover, we found that *MYB6*, *MYB56* and *MYB88* TFs had more target genes shared with *Dof30* TF than *bZIP9* and *bZIP40* TFs, indicating that *MYB* and *Dof30* TFs may function in the same pathway in the regulation of

key C4 gene expression. A machine-learning approach was also employed to predict gene regulatory relationship between TFs and their target genes. According to the sum weight calculated for all links among all TFs, we found that *MYB6, MYB56* and *MYB88* TFs had higher sum weight and more number of links than other TFs (Fig. 3C), indicating that these genes may play important
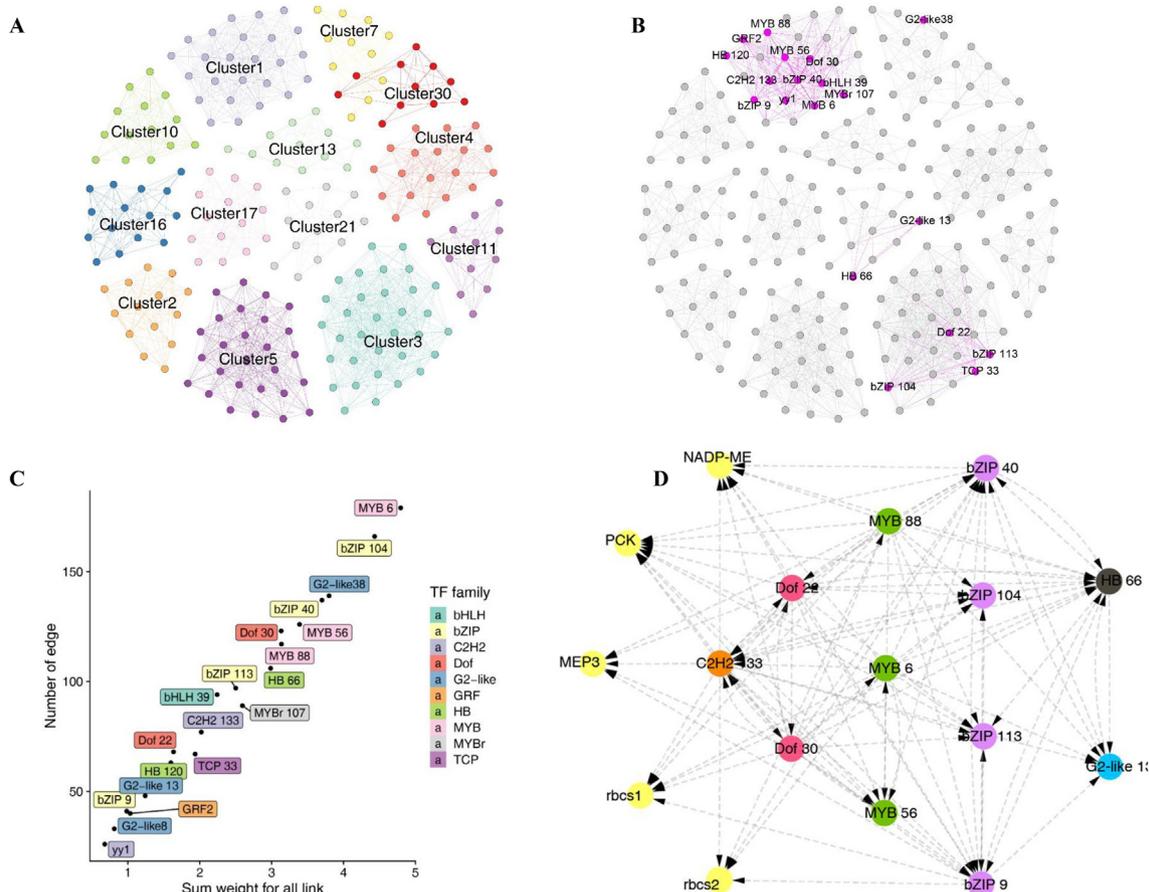
Fig. 3. Analyses of hub TFs in the magenta module. A) TF collaborative clusters for all TFs from the high variable expression matrix in WGCNA, only clusters with 10 more genes were shown. B) Highlight of the hub TFs in the magenta module in the TF collaborative clusters. C) Scatterplot showing the number of the edge and summed weight for all edges for each hub TF in the magenta module. D) Regulatory network based on TF ChIP-seq data for the hub TFs and key BS-specific C_4 genes in the magenta module. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

regulatory roles in the magenta module. To further reveal the potential regulatory relationship between the hub TFs or between the hub TFs and BS-specific C_4 genes, we used the published protoplast based 104 TF ChIP-seq data sets [24] to construct potential regulatory network between TFs and C_4 genes (Fig. 3D). We found that *Dof* TF might be regulated by *MYB*, *bZIP* and other TFs, especially *Dof30* may be specifically regulated by *C2H2-133* and *Dof22* TFs. *MYB* TFs may be self-regulated or be mainly regulated by *Dof* TF, while *bZIP* TFs may be regulated by *Dof* and *MYB* TFs. Furthermore, we found that *GLK1*, *HB 79* and *COL18* from the lightgreen and pink module were highly expressed in M cells but lowly expressed in BS cells, and they can potentially target some of BS-specific C_4 genes after analyzing the protoplast-based ChIP-seq data (Fig. S5). This result indicated that these C_4 genes can also be negatively regulated by TFs highly expressed in M cells. Thus, these analyses indicated that hub TFs might function in BS cells through complex regulatory network, and may positively or negatively regulate some of key C_4 genes in BS cells.

Collectively, these results suggest that *Dof22* may be an important regulator in the regulation of genes specifically expressed in BS cells, and *Dof30* and *MYB* TFs may cooperate with *Dof* TFs to regulate expression of important C_4 genes.

2.4. Differential enrichment of histone methylations in the genes belonging to the magenta module between M and BS cells

H3K27me3, a repressive mark, has been found to be involved in plant cell differentiation [30] and differential gene expression

between M and BS cells in maize [11]. To profile histone methylations in a subset of genes in the magenta module between M and BS cells, we analyzed the published H3K27me3 and H3K4me3 ChIP-seq data [11], and found that the actual number of genes enriched with H3K27me3 was significantly higher than the expected ones for all DEGs and a subset of genes in the magenta module. Moreover, a similar fold change between the actual and the expected number of genes was observed for all DEGs and subsets of genes in the magenta module, indicating that parts of genes in the magenta module tend to be regulated by H3K27me3 (Fig. 4A). As expected, an overall negative correlation was observed for changes in the enrichment levels of H3K27me3 and changes in the expression levels of all genes (n = 13,077), DEGs (n = 1,697) and a subset of genes (n = 194) in the magenta module enriched with H3K27me3 (Fig. 4B). Moreover, a negative correlation coefficient for genes ($R^2 = -0.27$) in the magenta module was higher than that for all genes ($R^2 = -0.19$), but lower than that for DEGs ($R^2 = -0.39$), further confirming that expression of parts of genes in the magenta module was affected by H3K27me3. We then specifically investigated the 20 hub TFs in the magenta module and found that 2 TFs (*Dof 22* and *Dof30*) in M cells and 1 TF (*HB 66*) in BS cells were highly enriched with H3K27me3 (Fig. 4C). Typical GA-repeat motifs were also found within the 2 kb upstream of the promoter regions of both *Dof* TFs (Fig. S6), which possibly act as the polycomb response elements (PREs) for the binding of BPC TFs, then recruiting PRC2 for methylating histone H3 at K27 residue [31]. We also analyzed the enrichment levels of H3K4Kme3, an active mark directly correlating with gene expression in plants
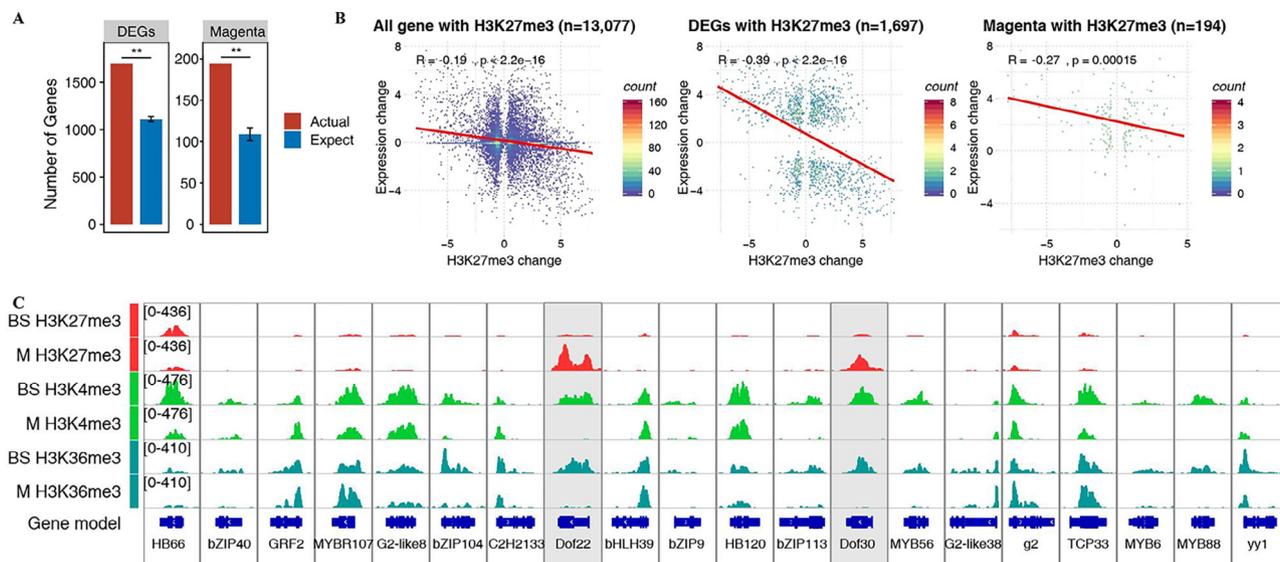
Fig. 4. Epigenetic features for genes in the magenta module. A) Barplot showing the actual and expected number of genes with H3K27me3 for DEGs and the Magenta module. The expected number were presented by mean ± SD for 50 random list with equal number. Significance test was determined using one-sample *t*-test, ** indicates *p* < 0.01. B) Scatterplot showing the correlation of the significance of expression levels and H3K27me3 enrichment changes between BS and M cells for all genes, DEGs and genes in the magenta module. C) IGV snapshot showing the variations in the epigenetic regulation of H3K27me3, H3K4me3 and H3K36me3 for hub TFs in the magenta module between two cell types. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

[32–34], and found that 8 of 20 hub TFs, such as *Dof22*, *Dof30*, *HB66*, *bZIP104*, *bZIP 113*, *MYB56*, *MYB6* and *MYB88*, were more enriched with both H3K4me3 and H3K36me3 in BS cells than in M cells (Fig. 4C). In addition, we observed co-presence of H3K27me3 and H3K4me3, which are functionally exclusive with each other in the regulation of gene expression, in *Dof22*, *Dof30* and *HB66* TF. This suggests that expression of these genes may be controlled by bivalent chromatin, which has been reported to regulate the transcription of genes involved in stress responses in plants [35] and human embryonic stem cell differentiation [36,37]. These results suggest that *Dof22* and *Dof30* may be important regulators in the regulation of BS cell specificity.

2.5. Differential enrichment of histone acetylation in genes belonging to the magenta module between M and BS cells

Histone acetylation is an active mark directly correlating with gene transcription [33,38]. To profile enrichment of histone acetylation, we conducted H3K4ac, H3K9ac and H4K12ac related ChIP-seq assays in BS cells. As expected, we observed a direct correlation of these three acetylation marks with gene expression (Fig. S7). Fold changes in the number of actual and expected genes enriched with histone acetylation tended to be higher in subsets of genes in the magenta module than DEGs in BS cells, 1.67 vs 1.31 for H3K4ac, 2.15 vs 1.86 for H3K9ac and 2.0 vs 1.55 for H4K12ac (Fig. 5A). This indicated that more percentages of genes in the magenta module were enriched with histone acetylation relative to DEGs in BS cells. We then plotted normalized read counts of each ChIP-seq across ±2 kb of the TSSs of subtypes of genes, including genes in the magenta module, BS-cell enriched DEGs, the top 75% of BS-cell enriched DEGs, M-cell enriched DEGs, all genes and random genes. After combining with gene expression levels shown in Fig. 5C, we found that each acetylation mark was generally more enriched in genes with higher expression levels than those with lower expression levels (Fig. 5B), by contrast, it was not the case for the top 75% of BS-cell enriched DEGs, genes in the magenta module, BS-cell enriched DEGs, and M-cell enriched DEGs, which exhibited a descending order as per the mean gene expression levels (Fig. 5B, C). Genes in the magenta module exhibited the highest

abundance of each acetylation mark at +1 and +2 nucleosome among all subtypes of genes examined, and subtle differences in the enrichment level of each mark occurred for BS- cell enriched DEGs, the top 75% of BS-cell enriched DEGs and M-cell enriched DEGs (Fig. 5B).

These results indicate that genes in the magenta module were more enriched with histone acetylation as compared to these BS-cell enriched DEGs with similar expression levels.

2.6. Significant enrichment of broad H3K4me3 in genes belonging to the magenta module

It has been reported that genes with broad H3K4me3 are related to photosynthesis in *Arabidopsis* [39]. It suggests that photosynthetic genes may have distinct chromatin features relative to other functional genes, which inspired us to investigate genes with broad H3K4me3 in M and BS cells. To this end, we conducted k-means clustering analyses (k value = 5) using all H3K4me3 related genes in BS and M cells (Fig. 6A). We obtained 5 subclusters with distinct profiles of H3K4me3 for both M and BS cells with similar gene numbers (Fig. 6A). Strikingly, we observed that distinct subsets of C_3 genes between both cells had much broader H3K4me3, which covered almost the whole region immediately downstream of the TSS, as compared to genes in other subclusters of both cells. After calculating H3K4me3 peaks length, we found that the majority of H3K4me3 peak length were 500–1000 bp in both cells, read counts for H3K4me3 peaks exhibited a sharp decrease when over approximate 1200 bp (Fig. 6B). We determined to set the threshold of consecutive H3K4me3 peak length as 1200 bp for broad H3K4me3, and the method for identification of genes with broad H3K4me3 can be seen in Fig. S8 (Fig. S8). We obtained 1,945 genes in BS cells and 1,706 genes in M cells marked with broad H3K4me3, and 1,344 genes shared between two cell types. To assess if genes with broad H3K4me3 in BS or M cells have any biological implications, we conducted GO term enrichment analyses, and found that these genes in both cells had primarily enriched GO terms with functions related to photosynthesis, indicating that genes marked by broad H3K4me3 may have dominant functions in C_4 photosynthesis in maize (Fig. S9).
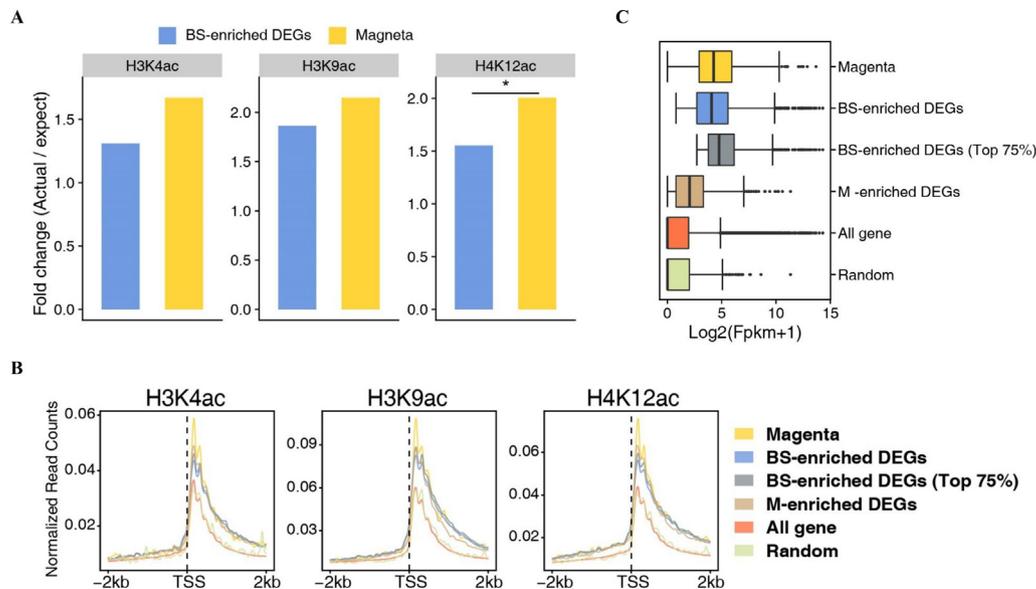
Fig. 5. Profile of histone acetylation in BS cells. A) Barplot showing the fold change between the actual and expected number of genes with histone acetylation for BS-enriched DEGs and genes in the magenta module. Significance test was determined using chisq-test, * indicates $p < 0.05$. B) Meta-plots showing the normalized read counts across the ± 2 kb of TSSs for different gene groups. C) Boxplot showing the expression levels for different gene groups. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
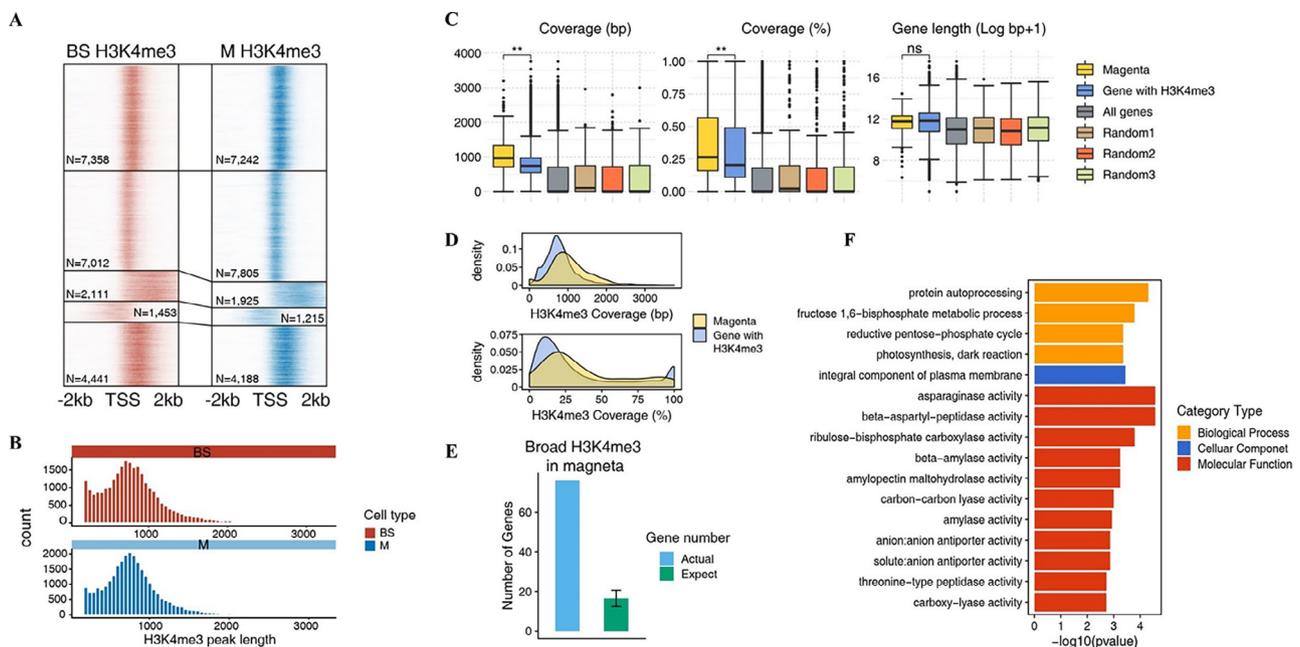
Fig. 6. Analyses of broad H3K4me3 enriched genes. A) Clustered heatmap showing the different profiles of H3K4me3 across ± 2 kb of TSS for all genes with H3K4me3. B) Histogram showing the length distributions of H3K4me3 peaks for BS and M cells. C) Boxplot showing the values of H3K4me3 coverage (bp), gene length and H3K4me3 coverage (%), significance test was determined using wilcoxon-test, ** indicates $p < 0.01$. D) Density curve of H3K4me3 coverage (both bp and %) for genes in the magenta module and genes with H3K4me3 in BS cells. E) Barplot showing the actual and expected number for genes with broad H3K4me3 in the Magenta module. The expected number was presented by mean ± SD for 50 random gene list with equal number. Significance test was determined using one-sample t-test, ** indicates $p < 0.01$. F) Barplot showing the enriched GO terms for 76 genes with broad H3K4me3 in the magenta module. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We also looked into 384 genes in the magenta module, and found that these genes exhibited a similar mean gene length, but higher H3K4me3 coverages and more percentages of H3K4me3 coverage relative to gene length than all genes or genes with H3K4me3 enrichment (Fig. 6C, 6D). This suggests that these genes in the magenta module tend to be associated with broad H3K4me3. We also used 1.2 kb as a cutoff for peak length to iden-

tify genes with broad H3K4me3 in the magenta module. We found that only 76 (*ca.* 20%) genes were marked by broad H3K4me3, by contrast, the ratio between the actual and the expected numbers was 4.6 times more in the magenta module than random controls, further indicating that genes in the magenta module are significantly marked by broad H3K4me3 (Fig. 6E). Again, we conducted GO term enrichment analyses for these 76 genes and found that

they were enriched in GO terms with functions related to photosynthesis, protein auto processing, enzymatic activities essential for sugar and protein metabolism (Fig. 6F), suggesting that these genes with broad H3K4me3 are essential for fundamental cellular and biological processes in maize.

3. Discussion

Differential gene expression on a genome-wide scale between BS and M cell has been extensively investigated [8–11,40], which can partially explain functional divergences in the photosynthetic pathway between two cell types in C_4 plants. Characterization of transcriptomic regulatory network can provide comprehensive insights into how hub TFs coordinate with key genes to result in functional divergences between two cell types. The single M-cell RNA-seq related network analyses revealed distinct roles of *COL8* and *HSF1* in the regulation of genes involved in the early and late leaf development [16,24]. However, the single BS-cell RNA-seq related network analyses are still unavailable since it is difficult to collect enough single BS cells for the single cell RNA sequencing [16,25], partially due to technique limitation and the structural nature of BS cells. Here, we for the first time conducted network analyses by analyzing bulk BS and M cell RNA-seq data. We identified a BS cell-specific co-expression module, the magenta module containing 384 genes highly expressed in BS cells instead of M cells (Fig. 2B), suggesting that these genes, at least part of them if not all, may play key roles in BS cells and underlie functional divergences between BS and M cells. Importantly, our study showed that some of hub TFs in the magenta module, such as *Dof*, *MYB*, *bZIP* and *G2-like* families, might play important roles in BS cells through regulating expression of BS-specific genes (Fig. 2E). The binding motifs and expression levels of *Dof* family have been found to be enriched in BS cells [8,9,11,16], indicative of importance of *Dof* TFs in C_4 functions in BS cells. Specifically, *Dof* TFs have been found to play both active and repressive roles in regulating photosynthesis in maize [14,15]. *Dof* TF (*Dof22*) in the magenta module and *Dof30* may regulate expression of C_4 genes due to their high expression levels in BS cells [11,13]. It has been found that functions of *Dof30* in the regulation of key C_4 genes can cooperate with *bZIP9* and *bZIP40* highly expressed in BS cells [11,28,29]. *MYB6*, *MYB56* and *MYB88* have even more inferred overlapped target genes with *Dof30* than *bZIP9* and *bZIP40*, indicating that *MYB* TFs may also cooperate with *Dof30* to regulate key C_4 genes. Moreover, by using a machine-learning approach, we found that *MYB6*, *MYB56* and *MYB88* showed both high sum weight and number of predicted links to target coding genes (Fig. 3C), further indicating that these genes may play important regulatory roles in the magenta module. Functional studies showed that functional mutation of the *golden plant 2* (*g2*) identified as a hub TF the magenta module results in both biochemical and morphological perturbations of BS cells development [41,42]. Thus, our study provides evidence to show possible regulatory mechanisms underlying functional divergences between BS and M cells in maize.

In addition, epigenetic regulation such as histone and DNA modifications has been found to regulate differential expression of C_4 genes in maize. For instance, *PEPC*, *NAPD-ME*, *PPDK*, *CA*, and *PCK* genes exhibiting cell type related expression were found to be differentially enriched with H3K4me3, an active mark directly associated with expression of overlapping genes [43,44]. Differentially methylated sites in the upstream of the *PEPC* gene resulted in differential abundance of its transcript between M and BS cells [45]. A genome-wide characterization showed that BS and M cells exhibit differential enrichment of H3K27me3, H3K4me3 and DNA methylation, indicative of the cell-specific epigenetic regulation in maize leaves [11]. Similarly, we found that only two *Dof* TFs

were differentially enriched with H3K27me3 and H3K4me3 between the two cell types in our hub TFs, leading to the possibility of these TFs controlled by bivalent chromatin, which has been found to be associated with plant development or stress responses [35,46]. Meanwhile, we found *Dof* TFs can hardly been regulated by other TFs (Fig. 3D) based on the network constructed using TF ChIP-seq [24]. *Dof22* and *Dof30* may be important activators for other TFs and the cooperation of these TFs may regulate other genes specific to BS cells. We also found that several histone acetylation markers (H3K4ac, H3K9ac, H4K12ac) were slightly more enriched in genes in the magenta module as compared to other genes, indicating that histone acetylation play vital roles in the regulation of a subset of genes in BS cells. Similarly, it has been reported that *PEPC* and *NADP-ME* are also enriched with H3K9ac in maize leaves [13,43,44,47]. In addition, we also provided evidence showing that genes with broad H3K4me3 can function in photosynthesis pathways in both BS and M cells and genes from BS-specific modules tended to be regulated by broad H3K4me3 (Fig. 6E), which may maintain the transcription consistency [48]. Genes with broad H3K4me3 has been found to be involved in several important biological processes including cell identity, expression of tumor-suppressor genes, maternal-to-zygotic cell transition in humans and mammals [48–52], and in photosynthesis and flower development in plants [39,53]. It is worth noting that this study used TF related ChIP-seq from M cells to indicate the potential binding of some of TFs in BS cells due to lack of related data in BS cells. It is necessary to be further validated by directly using TF related ChIP-seq from BS cells. Collectively, our study provides new insights regarding the regulatory network and underlying histone modifications responsible for distinct functions of BS cells.

4. Materials and methods

4.1. Preparation of BS and M cells

Isolation of M or BS cells was conducted according to our published procedures [16]. Briefly, 0.5–1 mm slices were prepared from the 10-day-old second leaf. M and BS cells were isolated using enzymatic digestion and mechanical isolation, respectively, as we described in our previous study [16]. After examining the purity using a bright-field microscope, the purified M and BS cells were stored at 80 °C until use.

4.2. Re-analyses of RNA-seq

All published RNA-seq data were download from EBI database in fastq format. RNA-seq data were mapped to the *Zea mays* AGPv4 reference genome using HISAT2 v2.1.0 [54]. The expression levels of all genes were calculated using StringTie 1.3.3b based on the ensemble B73v4.43 gene annotation [55]. Fragments per kilobase of transcript per million fragments mapped read (FPKM) values were used to measure the expression levels for all samples. DESeq2 was used to identify differential expressed genes (DEGs) between BS and M cells for our data. The threshold was set to $|\log_2^{(fold\ change)}| > 2$ with q-value <0.01 [56]. Correlation analyses for RNA-seq data were conducted using deeptools with a combined commend of mutilBamSummary and plotCorrelation [57].

4.3. Reverse transcription (RT) and quantitative RT-PCR (qRT-PCR) assays

Total RNA was extracted from BS and M cells using RNeasy plant Mini kit (Qiagen, Cat. # 74904). Approximately 200 ng total RNA without genomic DNA contamination was used for synthesis of the first strand cDNA following the manual from the SuperScript

III reverse Transcriptase kit (Cat. # 18080-044, Thermo Fisher Scientific). We randomly selected 10 differentially expressed genes (DEGs) for qPCR assays and qRT-PCR was performed according to the published procedures [58]. qPCR primers are listed in Supplementary Table S3. PCR was performed for *PEPC* and *RubisCO*, products of which were separated by running 1.5% agarose gel containing dye in 1xTBE buffer followed by visualizing and recording with gel doc. The primers used in the PCR assay are the same as the ones reported in previous studies [9].

4.4. Co-expression network analyses

FPKM matrices were merged from all samples. Genes expressed in at least 4 samples of BS or M cells were retained. Genes showing low variations in the expression levels across all samples based on mean absolute deviation (MAD) < 1 were filtered out. The matrices of 12,851 remaining genes were used to perform co-expression analyses using the WGCNA package [26]. Soft thresholding power was determined as $\beta = 9$ using pickSoftThreshold function. Scale-free network was constructed using blockwisemodule function with default settings. Trait matrices for BS and M cells were set using a binary value. Genes were considered as hub genes with | kME | > 0.75. The network was visualized using Gephi [59].

4.5. Construction of TF collaborative network

Construction of the transcriptional regulatory network was performed using the perl package TF-collaborative (https://github.com/hwei0805/TF_CollaborativeNet), which is the renamed version with the same algorithm as previously published TF-Cluster [27]. This package calculates the Shared Co-expression Connectivity Matrices for TFs and TFs, and finally clusters the TFs according to the triple-link algorithm. The analyses were performed with default settings with input files of TF list and expression matrices used in the WGCNA analyses.

4.6. Construction of gene regulatory network

Gene Regulatory Network (GRN) between TFs and target genes was constructed using GEne Network Inference with Ensemble of trees (GENIE3) v1.12.0 package, which employs an unsupervised approach [60]. GENIE3 takes expression matrix and gene list of transcript factor as input file and output predicted links between TFs and target genes. Expression matrices used in co-expression analyses were also used in GENIE3. Transcription factor list of 186 genes was determined according to the information from PlantTFDB database [61]. Regulatory relationships with weight > 0.02 were extracted. Summed weight values of all links for each TF were calculated.

4.7. ChIP-seq and data analyses

ChIP experiments and ChIP-seq library preparation for M and BS cells were conducted following the published protocols [62]. The commercial antibodies for ChIP assays are listed as below: H3K36me3 for both cell types (ab9050, abcam) and H3K9ac (07–352, Millipore), H3K4ac (07–539, Millipore), H4K12ac (07–595, Millipore) for BS cells only. Published H3K27me3 and H3K4me3 ChIP-seq data for BS and M cells were downloaded from EBI database. After sequencing on the Illumina platform, all raw ChIP-seq data sets were quality controlled (QCed) using FastQC [63] and cleaned using Cutadapt [64]. Clean ChIP-seq reads for all histone marks were aligned to the B73 reference genome using Bowtie2 with default parameters [65]. Reads with a mapQ > 20 were considered as uniquely mapped ones, and were then used for the downstream analyses. Peak calling was performed using macs2

with parameter --nomodel -q 0.01 --broad --broad-cutoff 0.1 for H3K27me3 and --nomodel -q 0.01 for other histone markers [66]. Differential enrichment analyses for BS and M cells were performed using manorm [67].

4.8. GO enrichment analyses

Online tools from agrigo v2.0 (https://systemsbiology.cau.edu.cn/agriGOv2/) were used for GO enrichment analyses. q-value of the Fisher test was < 0.01.

4.9. Statistical test

Significance test in Figs. 4A and 6E was determined using One-Sample t.test. Spearman correlation test was used in Fig. 4B. Significance test in Fig. 5A was determined using Chis-Q test. Significance test in Fig. 6C was determined using Wilcoxon test. For the expect number in Fig. 4A and Fig. 5A, we first calculated the ratio of genes with histone modifications (H3K27me3, H3K9ac, H3K4ac and H4K12ac) from all genes and considered as *p*. For any gene groups, the number of the gene group was considered as *n*, and the expect number of genes with histone modifications was calculated by multiplying *n* with *p*.

CRediT authorship contribution statement

Shentong Tao: Formal analysis, Writing – original draft. **Wenli Zhang:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author's contributions

W.L.Z. conceived and designed the initial study. S.T.T. analyzed all data sets; S.T.T. and W.L.Z. wrote the manuscript.

Availability of data and materials

ChIP-seq for BS and M cells and RNA-seq for maize were deposited in the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE206002 and GSE117982. All published data used in this study are listed in Supplementary Table S1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.07.004.

References

[1] Sage RF, Sage TL. C$_4$ plants. Encycl Biodiver 2013;75(1):361–81.
[2] Hatch MD, Agostino A. Bilevel disulfide group reduction in the activation of C$_4$ leaf nicotinamide adenine dinucleotide phosphate-malate dehydrogenase. Plant Physiol 1992;100(1):360–6.
[3] Aubry S, Kelly S, Kümpers BM, Smith-Unna RD, Hibberd JM. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C$_4$ photosynthesis. PLoS Genet 2014;10(6): e1004365.
[4] Wang L, Peterson RB, Brutnell TP. Regulatory mechanisms underlying C$_4$ photosynthesis. New Phytol 2011;190(1):9–20.
[5] Sedelnikova OV, Hughes TE, Langdale JA. Understanding the genetic basis of C$_4$ Kranz anatomy with a view to engineering C$_3$ crops. Annu Rev Genet 2018;52:249–70.
[6] Friso G, Majeran W, Huang M, Sun Q, van Wijk KJ. Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. Plant Physiol 2010;152 (3):1219–50.
[7] Agetsuma M, Furumoto T, Yanagisawa S, Izui K. The ubiquitin–proteasome pathway is involved in rapid degradation of phosphoenolpyruvate carboxylase kinase for C$_4$ photosynthesis. Plant Cell Physiol 2005;46(3):389–98.
[8] Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, et al. The developmental dynamics of the maize leaf transcriptome. Nat Genet 2010;42(12):1060–7.
[9] Chang YM, Liu WY, Shih AC, Shen MN, Lu CH, Lu MY, et al. Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiol 2012;160 (1):165–77.
[10] Tausta S, Lori Li, Pinghua Si, Yaqing, Gandotra, Neeru Liu. Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C$_4$-related processes. J Exp Bot 2014;65(13):3543–55.
[11] Dai X, Tu X, Du B, Dong P, Sun S, Wang X, et al. Chromatin and regulatory differentiation between bundle sheath and mesophyll cells in maize. Plant J 2022;109(3):675–92.
[12] Burgess SJ, Reyna-Llorens I, Stevenson SR, Singh P, Hibberd JM. Genome-wide transcription factor binding in leaves from C$_3$ and C$_4$ grasses. Plant Cell 2019;31(10):2297–314.
[13] Perduns R, Horst-Niessen I, Peterhansel C. Photosynthetic genes and genes associated with the C$_4$ trait in maize are characterized by a unique class of highly regulated histone acetylation peaks on upstream promoters. Plant Physiol 2015;168(4):1378–88.
[14] Yanagisawa S. Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. Plant J 2000;21(3):281–8.
[15] Yanagisawa S, Sheen J. Involvement of Maize Dof Zinc Finger proteins in tissue-specific and light-regulated gene expression. Plant Cell 1998;10 (1):75–89.
[16] Tao S, Liu P, Shi Y, Feng Y, Gao J, Chen L, et al. Single-cell transcriptome and network analyses unveil key transcription factors regulating mesophyll cell development in maize. Genes 2022;13(2):374.
[17] Górska AM, Gouveia P, Borba AR, Zimmermann A, Serra TS, Lourenço TF, et al. Zmb HLH 80 and Zmb HLH 90 transcription factors act antagonistically and contribute to regulate PEPC 1 cell-specific gene expression in maize. Plant J 2019;99(2):270–85.
[18] Borba AR, Serra TS, Górska A, Gouveia P, Cordeiro AM, Reyna-Llorens I, et al. Synergistic binding of bHLH transcription factors to the promoter of the maize NADP-ME gene used in C$_4$ photosynthesis is based on an ancient code found in the ancestral C$_3$ state. Mol Biol Evol 2018;35(7):1690–705.
[19] Górska AM, Gouveia P, Borba AR, Zimmermann A, Serra TS, Carvalho P, et al. ZmOrphan94 transcription factor downregulates ZmPEPC1 gene expression in maize bundle sheath cells. Front Plant Sci 2021;12:246.
[20] Reyna-Llorens I, Burgess SJ, Reeves G, Singh P, Stevenson SR, Williams BP, et al. Ancient duons may underpin spatial patterning of gene expression in C$_4$ leaves. Proc Natl Acad Sci U S A 2018;115(8):1931–6.
[21] Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, et al. cis-Regulatory elements for mesophyll-specific gene expression in the C$_4$ plant Flaveria trinervia, the promoter of the C$_4$ phosphoenolpyruvate carboxylase gene. Plant Cell 2004;16(5):1077–90.
[22] Gowik U, Schulze S, Saladié M, Rolland V, Tanz SK, Westhoff P, et al. A MEM1-like motif directs mesophyll cell-specific expression of the gene encoding the C$_4$ carbonic anhydrase in Flaveria. J Exp Bot 2017;68(2):311–20.
[23] Dickinson PJ, Knerova J, Szecowka M, Stevenson SR, Burgess SJ, Mulvey H, et al. A bipartite transcription factor module controlling expression in the bundle sheath of Arabidopsis thaliana. Nat Plants 2020;6(12):1468–79.
[24] Tu X, Mejía-Guerra M, Franco J, Tzeng D, Zhong S. Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. Nat Commun 2020;11(1):5089.
[25] Bezrutczyk M, Zollner NR, Kruse CPS, Hartwig T, Lautwein T, Kohrer K, et al. Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. Plant Cell 2021;33(3):531–47.
[26] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf 2008;9(1):1–13.
[27] Nie J, Stewart R, Zhang H, Thomson JA, Ruan F, Cui X, et al. TF-Cluster: a pipeline for identifying functionally coordinated transcription factors via network decomposition of the shared coexpression connectivity matrix (SCCM). BMC Syst Biol 2011;5(1):1–19.
[28] Bei Z, Wei C, Foley RC, Büttner M, Singh KB. Interactions between distinct types of DNA binding proteins enhance binding to ocs element promoter sequences. Plant Cell 1995;7(12):2241–52.
[29] Casati P, Walbot V. Maize lines expressing RNAi to chromatin remodeling factors are similarly hypersensitive to UV-B radiation but exhibit distinct transcriptome responses. Epigenetics 2008;3(4):216–29.
[30] Lafos M, Kroll P, Hohenstatt ML, Thorpe FL, Clarenz O, Schubert D. Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. PLoS Genet 2011;7(4):e1002040.
[31] Xiao J, Jin R, Yu X, Shen M, Wagner JD, Pai A, et al. Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. Nat Genet 2017;49(10):1546–52.
[32] Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di-and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol 2009;10(6):1–14.
[33] Wang X, Elling AA, Li X, Li N, Peng Z, He G, et al. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. Plant Cell 2009;21 (4):1053–69.
[34] He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, et al. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell 2010;22(1):17–33.
[35] Liu N, Fromm M, Avramova Z. H3K27me3 and H3K4me3 chromatin environment at super-induced dehydration stress memory genes of Arabidopsis thaliana. Mol Plant 2014;7(3):502–13.
[36] Blanco E, González-Ramírez M, Alcaine-Colet A, Aranda S, Di Croce L. The bivalent genome: characterization, structure, and regulation. Trends Genet 2020;36(2):118–31.
[37] Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006;125(2):315–26.
[38] Dion MF, Altschuler SJ, Wu LF, Rando OJ. Genomic characterization reveals a simple histone H4 acetylation code. Proc Natl Acad Sci U S A 2005;102 (15):5501–6.
[39] Brusslan JA, Bonora G, Rus-Canterbury AM, Tariq F, Jaroszewicz A, Pellegrini M. A genome-wide chronological study of gene expression and two histone modifications, H3K4me3 and H3K9ac, during developmental leaf senescence. Plant Physiol 2015;168(4):1246–61.
[40] Sawers RJ, Peng L, Anufrikova K, Hwang JG, Brutnell TP. A multi-treatment experimental system to examine photosynthetic differentiation in the maize leaf. BMC Genomics 2007;8(1):12.
[41] Hall LN, Rossini L, Cribb L, Langdale JA. GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. Plant Cell 1998;10 (6):925–36.
[42] Rossini L, Cribb L, Martin DJ, Langdale JA. The maize golden2 gene defines a novel class of transcriptional regulators in plants. Plant Cell 2001;13 (5):1231–44.
[43] Heimann L, Horst I, Perduns R, Dreesen B, Offermann S, Peterhansel C. A common histone modification code on C4 genes in maize and its conservation in Sorghum and Setaria italica. Plant Physiol 2013;162(1):456–69.
[44] Sascha Offermann BD, Horst I, Danker T, Jaskiewicz M, Peterhansel C. Developmental and environmental signals induce distinct histone acetylation profiles on distal and proximal promoter elements of the C$_4$-Pepc gene in Maize. Genetics 2008;179(4):1891–901.
[45] Langdale JA, Taylor WC, Nelson T. Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site >3 kb upstream of the gene. Mol Gen Genet 1991;225(1):49–55.
[46] Zheng B, Chen X. Dynamics of histone H3 lysine 27 trimethylation in plant development. Curr Opin Plant Biol 2011;14(2):123–9.
[47] Dong X-M, Li Y, Chao Q, Shen J, Gong X-J, Zhao B-G, et al. Analysis of gene expression and histone modification between C$_4$ and non-C$_4$ homologous genes of PPDK and PCK in maize. Photosynth Res 2016;129(1):71–83.
[48] Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, et al. H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell 2014;158(3):673–88.
[49] Dahl JA, Jung I, Aanes H, Greggains GD, Manaf A, Lerdrup M, et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature 2016;537(7621):548–52.
[50] Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature 2016;537(7621):553–7.
[51] Liu X, Wang C, Liu W, Li J, Li C, Kou X, et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature 2016;537(7621):558–62.
[52] Dhar SS, Zhao D, Lin T, Gu B, Pal K, Wu SJ, et al. MLL4 is required to maintain broad H3K4me3 peaks and super-enhancers at tumor suppressor genes. Mol cell 2018;70(5):825–41.

[53] Zhang A, Wei Y, Shi Y, Deng X, Gao J, Feng Y, et al. Profiling of H3K4me3 and H3K27me3 and their roles in gene subfunctionalization in allotetraploid cotton. Front Plant Sci 2021;12:2901.

[54] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12(4):357–60.

[55] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015;33(3):290–5.

[56] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15(12):550.

[57] Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42(Web Server issue):W187–91.

[58] Zheng D, Wang L, Chen L, Pan X, Lin K, Fang Y, et al. Salt-responsive genes are differentially regulated at the chromatin levels between seedlings and roots in rice. Plant Cell Physiol 2019;60(8):1790–803.

[59] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. Proceedings of the international AAAI conference on web and social media 2009;3:361–2.

[60] Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS ONE 2010;5 (9):e12776.

[61] Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 2016;45(D1):D1040–5.

[62] Zhao H, Zhang W, Chen L, Wang L, Marand AP, Wu Y, et al. Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. Plant Physiol 2018;176(4):2789–803.

[63] Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. 2010.

[64] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 2011;17(1):10–2.

[65] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2011;9(4):357–9.

[66] Yong Z, Tao L, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9(9):1–9.

[67] Zhen S, Zhang Y, Yuan GC, Orkin SH, Waxman DJ. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome Biol 2012;13(3):R16.