PNAS RESEARCH ARTICLE | PLANT BIOLOGY

Regulators of early maize leaf development inferred from transcriptomes of laser capture microdissection (LCM)-isolated embryonic leaf cells

Wen-Yu Liu[a], Chun-Ping Yu[a], Chao-Kang Chang[a], Hsiang-June Chen[a], Meng-Yun Li[a], Yi-Hua Chen[a], Shin-Han Shiu[b,c], Maurice S. B. Ku[d,e], Shih-Long Tu[f,1], Mei-Yeh Jade Lu[a,1], and Wen-Hsiung Li[a,g,1]

Contributed by Wen-Hsiung Li; received May 31, 2022; accepted July 28, 2022; reviewed by Alice Cheung and Nicholas Provart

The superior photosynthetic efficiency of C_4 leaves over C_3 leaves is owing to their unique Kranz anatomy, in which the vein is surrounded by one layer of bundle sheath (BS) cells and one layer of mesophyll (M) cells. Kranz anatomy development starts from three contiguous ground meristem (GM) cells, but its regulators and underlying molecular mechanism are largely unknown. To identify the regulators, we obtained the transcriptomes of 11 maize embryonic leaf cell types from five stages of pre-Kranz cells starting from median GM cells and six stages of pre-M cells starting from undifferentiated cells. Principal component and clustering analyses of transcriptomic data revealed rapid pre-Kranz cell differentiation in the first two stages but slow differentiation in the last three stages, suggesting early Kranz cell fate determination. In contrast, pre-M cells exhibit a more prolonged transcriptional differentiation process. Differential gene expression and coexpression analyses identified gene coexpression modules, one of which included 3 auxin transporter and 18 transcription factor (TF) genes, including known regulators of Kranz anatomy and/or vascular development. In situ hybridization of 11 TF genes validated their expression in early Kranz development. We determined the binding motifs of 15 TFs, predicted TF target gene relationships among the 18 TF and 3 auxin transporter genes, and validated 67 predictions by electrophoresis mobility shift assay. From these data, we constructed a gene regulatory network for Kranz development. Our study sheds light on the regulation of early maize leaf development and provides candidate leaf development regulators for future study.

Kranz anatomy | laser capture microdissection | cell-type transcriptome | gene regulatory network | TF binding site

Significance

C_4 leaves are superior to C_3 leaves in photosynthetic efficiency because they have the Kranz anatomy, in which the veins are surrounded by one layer of bundle sheath (BS) cells and one layer of mesophyll (M) cells. To identify early regulators of Kranz development, we obtained 11 transcriptomes of maize pre-BS and pre-M cells. Differential gene expression and gene coexpression analyses and determination of transcription factor (TF) binding motifs provided data for constructing a gene regulatory network (GRN) for Kranz development, which includes known regulators of Kranz anatomy and/or vein development. In situ hybridization of 11 TF genes in the GRN revealed their expression in early Kranz development. This study provides insights into the regulation of early maize leaf development.

C_4 plants are characterized by a distinct leaf structure, namely Kranz anatomy, in which the vein is surrounded by one inner layer of organelle-rich bundle sheath (BS) cells and one outer layer of mesophyll (M) cells (1–3). This spatial configuration with a short metabolite diffusion path between the two cell types and the morphological as well as biochemical specialization of BS and M cells greatly increase photosynthesis efficiency (4).

The development of Kranz anatomy can be divided into three stages: 1) initiation of procambium, 2) BS and M cell differentiation, and 3) chloroplast development and the integration of the C_4 cycle (5). The primary determinant of cell fate is the cell position within the ground meristem (GM), and cell-to-cell communication is considered essential for interpreting that position (4). Histological studies of maize leaf ontogeny revealed that the development of Kranz anatomy starts from three contiguous median ground meristem (mGM) cells (6, 7) (*SI Appendix*, Fig. S1). The three cells then give rise to five contiguous cells by two unequal periclinal divisions of the middle cell, forming four pre-BS cells and one prevein (pre-V) cell (*SI Appendix*, Fig. S1), which are denoted as 4BS+V. Then, additional cell divisions lead to five and six pre-BS cell stages (referred to as the 5BS+V and 6BS+V stages, respectively), and enlargement of BS cells occurs to form a typical Kranz anatomy (*SI Appendix*, Fig. S1). Below, we shall call the precursors of BS and V cells "pre-Kranz cells" and the precursors of M cells "pre-M" cells.

In the study of leaf Kranz anatomy, one long-standing question is how the developmental process of Kranz anatomy is regulated. With genetic approaches, some studies have examined mutants with altered vein structure or density, including maize mutants with abnormal leaf veins (8), rice mutants with increased leaf vein density (9), and sorghum mutants with decreased leaf vein density (10). Several transcription factors (TFs) that regulate Kranz anatomy development have been identified, including SCARECROW (SCR), SHORT-ROOT (SHR), and AINTEGUMENTA 1 (ANT1) (8, 11, 12). In addition, a number of systems biology–based analyses have led to the identification of genes differentially expressed between C_3 and C_4 plants or tissue types (13–15). Other studies have analyzed maize embryonic

Author contributions: W.-Y.L., M.S.B.K., S.-L.T., M.-Y.J.L., and W.-H.L. designed research; W.-Y.L., C.-P.Y., C.-K.C., H.-J.C., M.-Y.L., Y.-H.C., and M.-Y.J.L. performed research; W.-Y.L. and C.-P.Y. analyzed data; and W.-Y.L., C.-P.Y., C.-K.C., S.-H.S., M.S.B.K., S.-L.T., M.-Y.J.L., and W.-H.L. wrote the paper.

Reviewers: A.C., University of Massachusetts Amherst; and N.P., University of Toronto.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: tsl@gate.sinica.edu.tw, meiyehlu@gate.sinica.edu.tw, or whli@uchicago.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2208795119/-/DCSupplemental.

Published August 24, 2022.

leaves at different developmental stages (16) and time course transcriptomes of maize embryonic leaves (7, 17).

While these studies have contributed to a better understanding of differential regulation of gene expression at different developmental stages relevant to the development of Kranz anatomy, the tissues used are of mixed cell types with mainly M cells but substantially fewer preBS+PreV (BS+V) cells. Thus, the goal of this study was to obtain cell type–specific transcriptomes of early precursors of maize BS+V and M cells using laser capture microdissection (LCM). By examining temporal transcriptomes of developing cells at high resolution, a gene expression atlas was constructed from cell-type RNA sequencing (RNA-seq) data. A gene regulatory network (GRN) was then constructed using WGCNA (weighted gene cocorrelation network analysis) and cis-motif analysis. Based on the inferred regulatory network, we identified candidate TFs and their target cis elements that are involved in Kranz anatomy development. The implications of our findings for the regulation of early maize leaf development are discussed.

Results

LCM of Embryonic Leaves and RNA Extraction. To identify genes related to Kranz anatomy development, we used LCM to isolate two sets of cells from embryonic leaves of 3-d-old maize seedlings (Fig. 1 and *Methods*). The first set included cells that would develop into BS cells and V cells. They were obtained from five developmental stages (Fig. 1*A*): 1) seven mGM cells (Fig. 1*A*) (these cells lie between two veins of primordia 1 (P1) or P2); 2) the three–contiguous cell stage that initiates vein development, denoted by "3C"; 3) the stage of four pre-BS cells + pre-V cells (i.e., "4BS+V"); 4) the stage of five pre-BS cells + pre-V cells (i.e., "5BS+V"); and 5) the stage of six pre-BS cells + pre-V cells (i.e., "6BS+V"). The second set included pre-M cells that would develop into M cells. For each of the above five stages, the same stages of prepalisade mesophyll (PM) cells were also isolated (referred to as PM, 3PM, 4PM, and 5/6PM, respectively) (Fig. 1*B*). In addition, we also isolated the one median mesophyll (1M) cell and the two median mesophyll (2M) cells between two adjacent preveins at the stage of 5BS+V or 6BS+V.

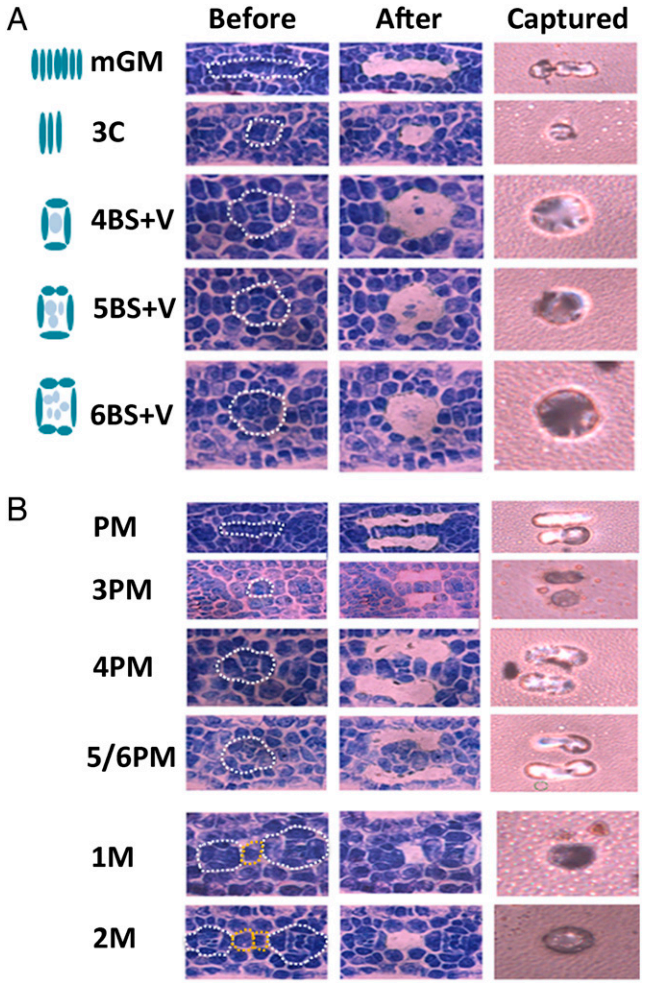

Fig. 1. Isolation of tissue-specific cells from maize embryonic leaves by LCM. (*A*) Isolation of mGM cells and early BS+V cells. The first column shows sections before LCM, the second column shows sections after LCM, and the third column shows captured cells. (*B*) Isolation of pre-M cells. White lines:pre-Kranz structure; yellow lines: M cells.

Total RNA extracted from two biological LCM replicates for the mGM, 4BS+V, 5BS+V, 6BS+V, and 5/6PM cell samples and triplicates for the 3C, 3PM, 4PM, 1M, and 2M cell samples were subjected to linear amplification. The RNA samples were used to construct complementary DNA (cDNA) libraries, which were then paired-end (PE) sequenced using an Illumina HiSeq 2500 platform (*SI Appendix*, Table S1). The RNA samples were found to be of good quality (*SI Appendix*, Fig. S2). The RNA-seq data are shown in *SI Appendix*, Table S2.

The sequencing reads were quality trimmed and mapped to the maize reference genome (B73 RefGen_v4). Among the mapped reads, 4.5 to 8.6 million reads (62.4 to 72.96% of total reads) of pre-Kranz cells and 3.4 to 12.5 million reads (54.69 to 71.52%) of pre-M cells were mapped to exonic sequences (*SI Appendix*, Table S2). The exonic reads were normalized using Cufflinks (18) and reported in terms of "fragments per kilobase of transcript per million mapped reads" (FPKMs). From the RNA-seq data, a gene was defined as expressed if it was found in at least 1 of the 11 transcriptomes; a total of 20,352 expressed genes were identified (Dataset S1).

Distinct messenger RNA (mRNA) Populations of Pre-Kranz Cells and Pre-M Cells. To study the mRNA expression differentiation between the pre-Kranz cell lineage and the pre-M cell lineage, we conducted a principal component analysis (PCA) and a hierarchical clustering analysis of the RNA-seq data (Fig. 2). The PCA shows that the mRNA population of mGM cells and that of PM cells are fairly similar (Fig. 2*A*), indicating only mild differentiation between the two cell lineages at this early phase of leaf development. Then, gene expression differentiation progresses rapidly from mGM cells to 3C cells while less rapidly from PM cells to 3PM cells. Thus, at this stage of leaf development, the pre-Kranz cell lineage develops faster than the pre-M cell lineage, and the mRNA population of 3C cells already becomes clearly distinct from those of mGM cells and 3PM cells (Fig. 2*A*). Subsequently, in the pre-Kranz lineage, gene expression differentiation becomes much slower, so that the mRNA populations of 3C, 4BS+V, 5BS+V, and 6BS+V cells are similar and clustered in one group (Fig. 2*A*). In contrast, in the pre-M cell lineage, gene expression differentiation continues rapidly between 3PM and 4PM (Fig. 2*A*); after that, however, gene expression differentiation becomes much slower. Interestingly, the mRNA populations of 1M and 2M are clustered with those of 4PM and 5/6PM, indicating that 1M and 2M cells are similar in maturity to 4PM and 5/6PM cells (Fig. 2*A*). In the hierarchical clustering analysis, the mRNA populations of pre-Kranz cells and those of pre-M cells form two separate groups after the mGM and PM stages

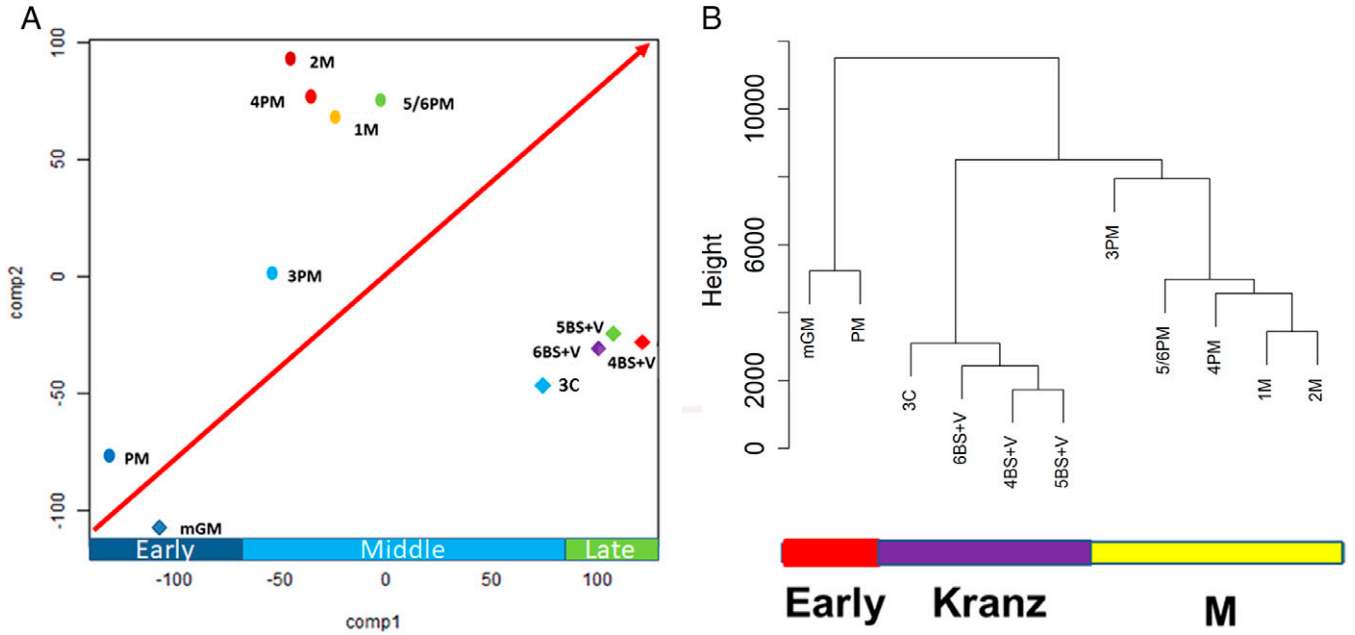

Fig. 2. Relationship between the RNA populations obtained from the pre-Kranz and pre-M cells of maize embryogenic leaves. (*A*) PCA of genes expressed in the captured pre-Kranz and pre-M cells. The mRNA populations of mGM cells and PM cells are fairly similar; those of 3C, 4BS+V, 5BS+V, and 6BS+V cells form one group, and that of 3PM cells is clearly separated from all others, whereas those of 4PM, 5/6PM, 1M, and 2M cells form one group. (*B*) A cluster dendrogram showing transcriptome relationships among samples of pre-Kranz and pre-M cells. The clustering patterns are consistent with those of PCA. For example, mGM and PM are clustered in one clade, and 3PM is well separated from 4PM and 5/6PM, which together with 1M and 2M, form one clade. The horizontal color bar indicates the developmental phases: early phase, Kranz anatomy development, and M cell development.

(Fig. 2*B*), consistent with the PCA. In summary, the differentiation of pre-Kranz cells occurs largely before the 3C stage, whereas rapid differentiation of pre-M cells continues until the 4PM stage.

Taken together, our data indicate that the BS and M cell gene expression programs diverge early on in maize leaf development and that at the 3C stage, the three neighboring cells are already committed to become BS and V cells. Thus, our gene expression data likely capture the early molecular activities relevant to establishing the characteristic Kranz anatomy development.

Expression Profiles of Cell Populations and Candidate Regulators of BS Cell Differentiation. To find genes that modulate pre-Kranz and pre-M cell differentiation in maize embryonic leaves, we identify genes that are differentially expressed between pre-Kranz cells and corresponding pre-M cells (e.g., 3C vs. 3PM) (*Methods* and Dataset S2). In Fig. 3*A*, the transcriptional profiles of differentially expressed genes (DEGs) are shown in terms of the relative expression levels normalized across the 11 cell types (*z* scores), while in Fig. 3*B*, the expression levels are normalized across genes in the same cell. In Fig. 3*A*, the 592 DEGs are clustered into eight groups (C1 to C8; the first column). In groups C1 to C6, the gene expression levels in pre-Kranz cells are low except for mGM in C6 (many genes), C3 (a number of genes), and C1 (a few genes). In C1 to C3, the expression levels are the highest in 4PM or 5/6PM. In C4, the expression levels are the highest in 2M, 4PM, or 3PM. In C5, the expression levels are the highest in 3PM. In C6, the expression levels are the highest in PM and/or 3PM. As C1 to C6 include more than two-thirds of the DEGs, there are more drastic changes in transcriptional programs in pre-M cells than pre-Kranz cells. Many of these genes are likely important for the development of M cells, although some (e.g., in C6) are probably also important for Kranz cell development when they are up-regulated. On the other hand, most C7 genes show the highest expression levels in pre-Kranz cells except mGM. These genes are also among the highest expressed genes in each of the pre-Kranz cell types as seen in Fig. 3*B*. The C8 genes, with only a few exceptions, show higher expression levels in mGM than in all other cell types. Thus, the genes in C7 and C8 are likely important for Kranz cell differentiation and development. C7 contains only 1 TF gene (*ZmPRR1*), while C8 contains 12 (*LBD33*, *ZmBBX3*, *GATA32*, *WRKY94*, *YABBY1*, *ZmANT2*, *DOT5*, *DOF25*, *ZmANT3*, *AIL6*, *ZmANT1*, *YABBY3*), which are candidate regulators of Kranz anatomy or vascular development (*Discussion*).

To find genes that are relatively highly or lowly expressed at a developmental stage, the relative expression levels of genes are sorted in the same cell type (Fig. 3*B*, vertical comparison); the genes are listed in the same order as in Fig. 3*A*. We mention some examples. First, there are genes that show dark red levels in all 11 cell types, indicating that these genes are the most highly expressed genes in all 11 cell types; they are found mainly in C4 to C6 (Fig. 3*B*). Second, there are genes (mainly in C6) that show red or dark red in mGM and PM but red or light red in all other cell types, indicating that these genes are very highly expressed at the early stage of leaf development but soon become only highly or moderately highly expressed as leaf development continues. Third, a conspicuous pattern is observed in C7 in that a cluster of genes shows light, intermediate, or dark red at 3C, 4BS+V, 5BS+V, and 6BS+V in Fig. 3*B* but shows red or light red at the same stages in Fig. 3*A*. These genes are more highly expressed than other genes in these cell types and are much more highly expressed in pre-Kranz cells than in pre-M cells at these stages of development. Fourth, in C8 there are genes that show low expression (blue or light blue color) in all cell types except mGM and/or 3C cell types (Fig. 3*A*). The implications of these observations will be discussed in *Discussion*.

Identification of Gene Coexpression Modules. In Fig. 3, the clustering of gene expression profiles does not provide regulatory relationships between genes. To understand the regulatory

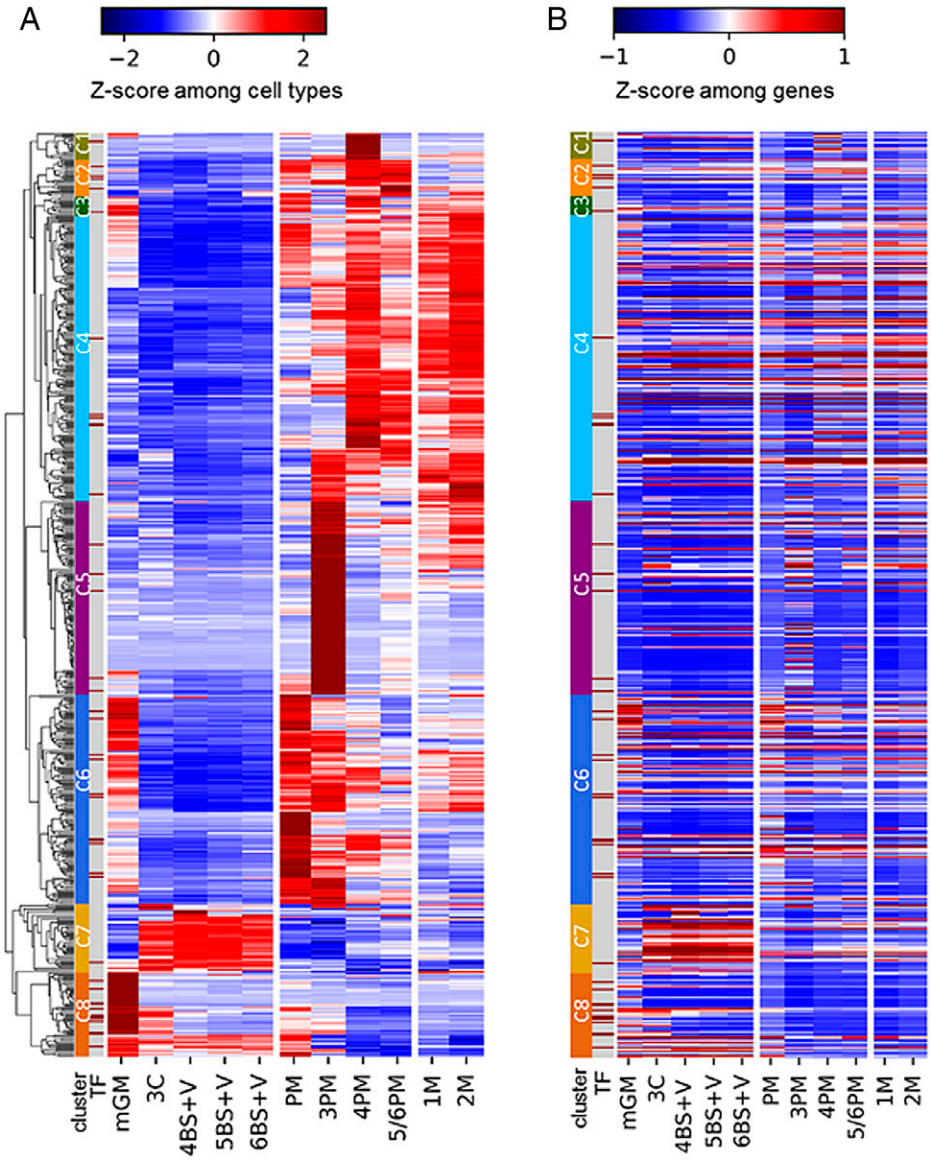

Fig. 3. Clustering of expression profiles of DEGs between pre-Kranz and pre-M cells. (*A*) Clustering of the expression profiles of DEGs in the 11 cell types. The color bar at the top shows the normalized *z* scores of gene expression levels among the 11 cell types (horizontal normalization); blue color means the lowest *z* score, while red color means the highest *z* score. The 592 DEGs are clustered into eight groups (clusters; C1 to C8) as indicated in the first column. In the second column, each red line indicates a TF gene. The *z* scores of the DEGs are shown after the second column. (*B*) Clustering of the expression profiles of DEGs across genes in the same cell type. The DEGs are listed vertically and clustered in eight groups as in *A*. However, for each cell type, the *z* scores are normalized among genes in the same cell type (vertical normalization). The color bar for the *z* scores is shown at the top. In the second column, each red line indicates a TF gene as in *A*. The identified TFs in the eight groups in *A* are listed below. C1: *ZmARF14*; C2: *ZmPDF2*, *ZmHDG8*, *EREBP059*, and *bHLH116*; C3: *GRF13*; C4: *GATA2*, *bHLH130*, *MYB126*, *bHLH96*, *bHLH82*, and *ZmCCA1*; C5: *MYB35*, *ERF81*, *DBB11*, *HSF18*, and *CO-like13*; C6: *GRF6*, *CSD2*, *MYB79*, *ERF182*, *DBB4*, *NAC49*, *HD83*, *HD17*, *HD71*, *HD91*, and *HD75*; C7: *ZmPRR1*; C8: *LBD33*, *ZmBBX3*, *GATA32*, *WRKY94*, *YABBY1*, *ZmANT2*, *DOT5*, *Dof25*, *ZmANT3*, *AIL6* (*ERF130*), *ZmANT1*, and *YABBY3*. All the DEGs and their gene identifications are given in Dataset S2.

relationships between genes, we construct gene coexpression networks by applying the WGCNA (19, 20) to the cell-type gene expression data (*Methods*). The genes are clustered into 15 coexpression modules, excluding a module (denoted by M16) consisting of transcripts that do not match the selection criteria (Fig. 4). The 15 modules contain 43 (M15) to 5,138 (M6) genes, including 0 (M15) to 222 (M6) coexpressed TF genes (Dataset S3 and SI Appendix, Fig. S3). The relationship between coexpression module and cell type is assessed by the Pearson correlation coefficient analysis and visualized using a heat map (Fig. 4*A*), and the average expression profiles of genes in the 11 cell types for each of the 16 modules are shown in Fig. 4*B*. The first 15 coexpression modules can be roughly classified into four categories according to the gene coexpression profiles across cell types. The first category consists of the M1 and M2 modules. These two modules are correlated with both mGM (r = 0.66; 0.67) and PM (r = 0.62; 0.44) cells. In both modules, especially the M2 module, the genes show a higher expression level in pre-Kranz cells than in corresponding pre-M cells and thus, may play an important role in the differentiation of pre-Kranz cells. The second category consists of a single module (M3), in which gene expression is much higher in mGM than in PM, and the expression levels are reduced in 3C cells but remain expressed thereafter. The expression patterns of genes in this module suggest that this module is important for establishing the BS+V cell fate and in subsequent pre-Kranz cell differentiation. In the third category, the member genes generally are correlated in expression levels with the genes in all BS+V cell stages (i.e., the M2 and

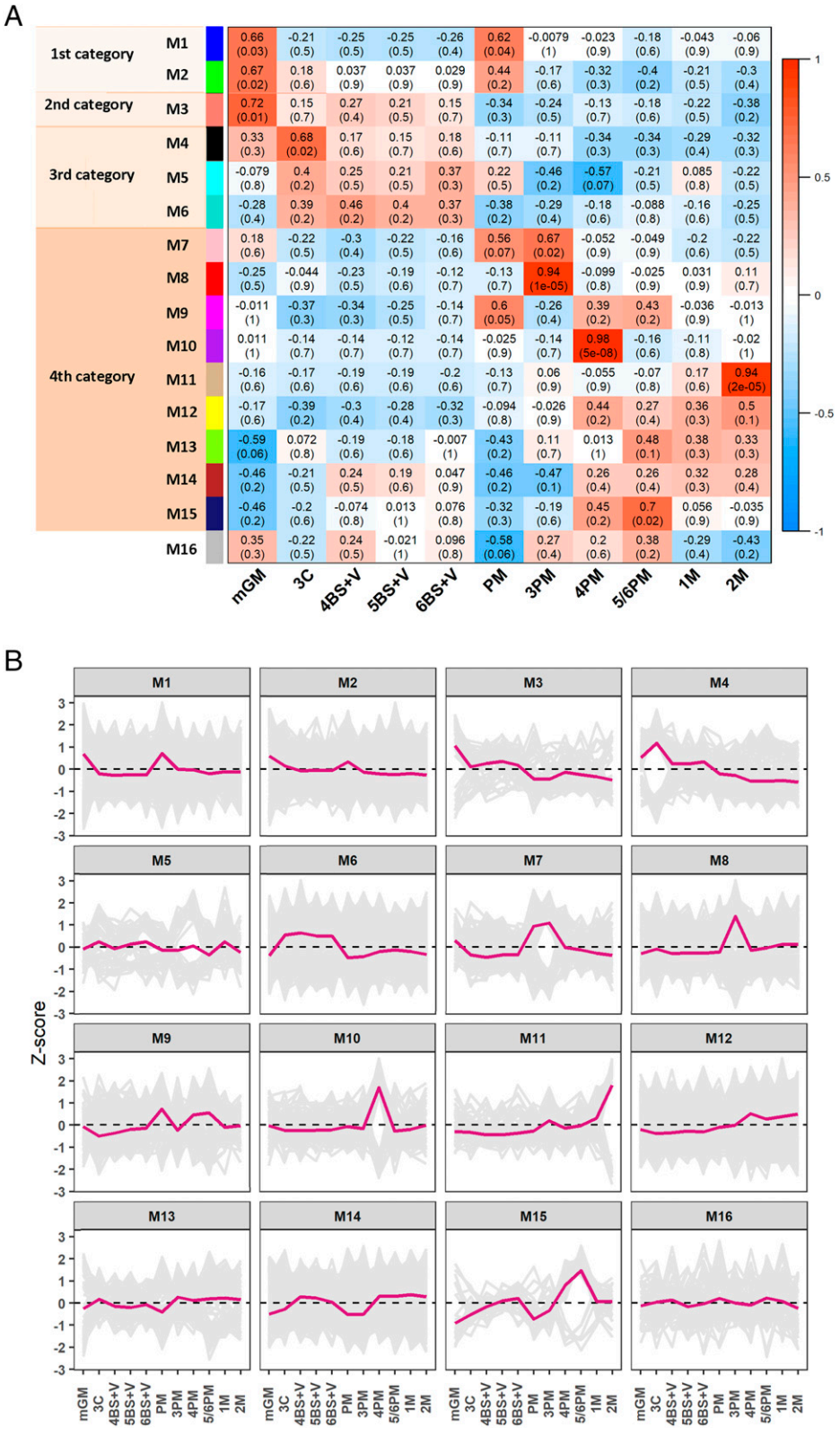

Fig. 4. Relationship and expression profiles between the gene coexpression module and cell type. (*A*) Each row corresponds to a module, while each column corresponds to a cell type. The values in the cells are "Pearson correlation coefficients (*P* values)," and the color of a cell indicates the sign and magnitude of the correlation (red, positive correlation; blue, negative correlation). The color bar for the correlation coefficient is shown on the right-hand side. (*B*) *z*-score plots showing expression profiles of all genes in the WGCNA-generated coexpression modules M1 to M16.

M4 to M6 modules). Genes in these modules are likely responsible for further BS+V cell differentiation. The fourth category modules tend to correlate with pre-M cells (i.e., M7 module with PM and 3PM, M8 with 3PM, M9 with 3PM, M10 with 4PM, M11 and M12 with 2M, and M13 to M15 with 5/6PM). These genes have higher expression levels in pre-M cells than in pre-Kranz cells and thus, are likely important for further differentiation and maturation of pre-M cells. Taken together, these results indicate that each of the pre-Kranz and pre-M cell compartments of a maize embryonic leaf is associated with one or more

coexpression modules that reflect the gene regulatory processes specific to a compartment and are indicators of the differentiation programs functioning within specific cellular compartments.

Biological Processes Enriched in Coexpression Modules of BS+V Cells. To identify the major biological processes associated with the BS+V coexpression modules, we used the Singular Enrichment Analysis in agriGO version 2.0 (21) to identify the processes that were significantly enriched (with a false discovery rate (FDR) of <0.05) in the modules that showed high correlation with BS+V developments. These included modules in category 1 (M1 and M2; mildly differentiated), category 2 (M3; BS+V cell fate), and category 3 (M4 to M6; BS+V differentiation) (Fig. 5*A*), but only M1, M2, M4, and M6 have Gene Ontology (GO) terms significantly enriched (Fig. 5*B* and Dataset S4). The BS+V-correlated M1 and M2 modules are found to be enriched in "auxin-mediated signaling pathway" and "response to auxin stimulation," with the latter GO category including the *ZmPIN1a* (Zm00001d044812) and *ZmPIN1d* (Zm00001d052442) genes, which encode two auxin efflux carriers involved in shoot and root development (22, 23). Additionally, a close inspection of genes in the M2 module also identifies *ZmLAX2* (Zm00001d028401), the paralog of auxin influx carrier (*AUX1*), which maintains the auxin distribution pattern against environmental or developmental influences (24).

To investigate the importance of auxin import in Kranz anatomy development, we study auxin distribution and the expression of *LAX2*, a major influx carrier in maize. Auxin distribution and accumulation are found in developing maize seedlings (Fig. 6 *A* and *B*), and *ZmLAX2* transcripts are present in inner cells of shoot apical meristem (SAM) and become abundant in the vascular bundles of seedlings and young leaves (Fig. 6 *C*–*F*). The combined transcription and in situ data show that the three auxin transporter genes (*ZmLAX2*, *ZmPIN1a*, and *ZmPIN1d*) are more highly expressed at the mGM stages than at the other stages (Fig. 7*A*), and they are expressed in vascular bundles. These observations suggest that auxin transport is important for Kranz anatomy development by first forming auxin maxima for rapid cell division and differentiation.

Candidate TFs Important for BS+V Cell Differentiation. We focus on the BS+V-associated M2 coexpression module to decipher a portion of the auxin-mediated GRN because auxin is important for Kranz anatomy development. We select TF genes with two criteria: coexpression with auxin transport genes in the M2 module and enriched expression in mGM cells compared with PM cells. There are 18 TFs and three auxin transporters that are specific to the M2 module (Fig. 7*A* and *SI Appendix*, Table S3). From the preferential gene expression in mGM over PM cells, we make the following observations (Fig. 7*A* and *SI Appendix*, Table S3). First, these 21 genes have an expression level 1.5 to 4 times higher in pre-Kranz cells (especially in mGM and 3C cells) than in pre-M cells. The genes that are strongly and preferentially expressed (greater than three to four times) in mGM and 3C cells over PM and 3M cells are *ANT2*, *ANT3*, *ANT3-1*, *AIL6*, *ARF20*, *C2H2*, *DOT5*, *DOF43*, *GATA21*, *LAX2*, and *PIN1a*. Second, in the descending order, the most highly expressed TF genes in mGM and 3C are *YABBY1*, *AIL6*, *ANT3*, *ANT1*, *ARF25*, *ANT3-1*, and *DOT5*. Collectively, these genes may play a key role in initiating BS cell differentiation and vein development. Third, the expression levels of several TF genes (e.g., *ANT3*, *ARF25*, and *YABBY1*) are the highest in mGM and maintain a relatively higher level of expression thereafter. These TFs may be involved in both BS cell differentiation and maintenance. Significantly, *LAX2*, *PIN1a*, and *PIN1d* are more highly expressed at the mGM and 3C stages than at the other stages, consistent with the notion that they play a key role at early cell developmental stages in forming auxin maxima and determining their cell fates.

The 18 TF genes specific to the M2 module include 5 ANT genes that are involved in cellular differentiation because their loss of function showed reduced fertility, abnormal ovules and lateral organs, and aberrant vascular development (12, 25). Two (*ARF20* and *ARF25*) of the 18 TFs are auxin-response TFs; their *Arabidopsis* orthologs are known to regulate vascular development (26, 27). The M2 module also includes three basic helix–loop–helix (bHLH) TFs (*ZmbHLH88*, *ZmbHLH105*, *ZmbHLH134*); their *Arabidopsis* orthologs regulate root hair and sperm cell development and control asymmetric cell divisions during stomatal development (28, 29). *C2H2* and *DOT5* (DEFECTIVELY ORGANIZED TRIBUTARIES 5) are two C2H2 zinc finger TFs; the *Arabidopsis* ortholog of DOT5 regulates vascular patterning (30). *DOF43* (zinc finger) TF, the ortholog of *Arabidopsis* HCA2 (HIGH CAMBIAL ACTIVITY 2), is expressed in vascular tissue (31). *GATA21* encodes a GATA TF whose *Arabidopsis* ortholog is required to position the proembryo boundary in the early embryo (32). *Mybr59* encodes the *Arabidopsis thaliana* myb/SANT domain protein. *WRKY109* encodes a member of the WRKY TF family. *YABBY1* encodes a member of the YABBY TF family whose *Arabidopsis* ortholog is involved in the abaxial cell–type specification in leaves and fruits (33).

To validate the above findings, we select 11 of the 18 candidate TF genes and 2 auxin transporter genes (*PIN1a* and *LAX2*) from the M2 module (Fig. 7*A* and *SI Appendix*, Table S3) for in situ hybridization study. All of these 13 genes exhibit Kranz-specific differential expression and are likely genes functioning within or close to the developing Kranz anatomy during early maize leaf development. The in situ hybridization data (Fig. 7*B*), as well as the transcriptomes, show that they are expressed in the BS cells and vascular bundles of seedlings, suggesting that these genes are involved in early Kranz anatomy development.

Identifying TF Binding Motifs and Inferring Regulatory Relationships between Genes. To be able to predict the target genes of the above TFs, we determine their binding motifs (transcription factor binding sites [TFBSs]) using the DNA affinity purification sequencing (DAP-seq) technique (34), which is a powerful in vitro technique for genome-wide TFBS discovery. We are able to infer the PWMs (position weight matrices) of 15 TFs from the DAP-seq data (*SI Appendix*, Table S4).

For each of the above 15 PWMs, we use FIMO (Find Individual Motif Occurrences) of The MEME Suite (35) to detect its presence in the promoter sequences of the predicted target genes in the M2 coexpression module (*Methods* and *SI Appendix*, Table S5). The presence of a PWM motif sequence in the promoter is indicated by a red line between the TF and a predicted target gene in Fig. 8, in which the TFs and their predicted target genes with TFBS support are visualized using Cytoscape (36).

We then use the electrophoresis gel mobility shift assay (EMSA) to test for the predicted binding of a TF to the promoters of its inferred target genes (*SI Appendix*, Table S5). In total, 10 TFs tagged with glutathione S-transferase (GST) fusion are successfully expressed and purified, including ZmANT1, ZmANT3, ZmAIL6, ZmDOF43, ZmbHLH88, ZmbHLH105, ZmARF25, ZmARF20, ZmGATA21, and ZmDOT5. As expected, retardation in gel mobility is clearly observed when the TF recombinant proteins are individually incubated with the biotin-labeled probes with sequence derived from the

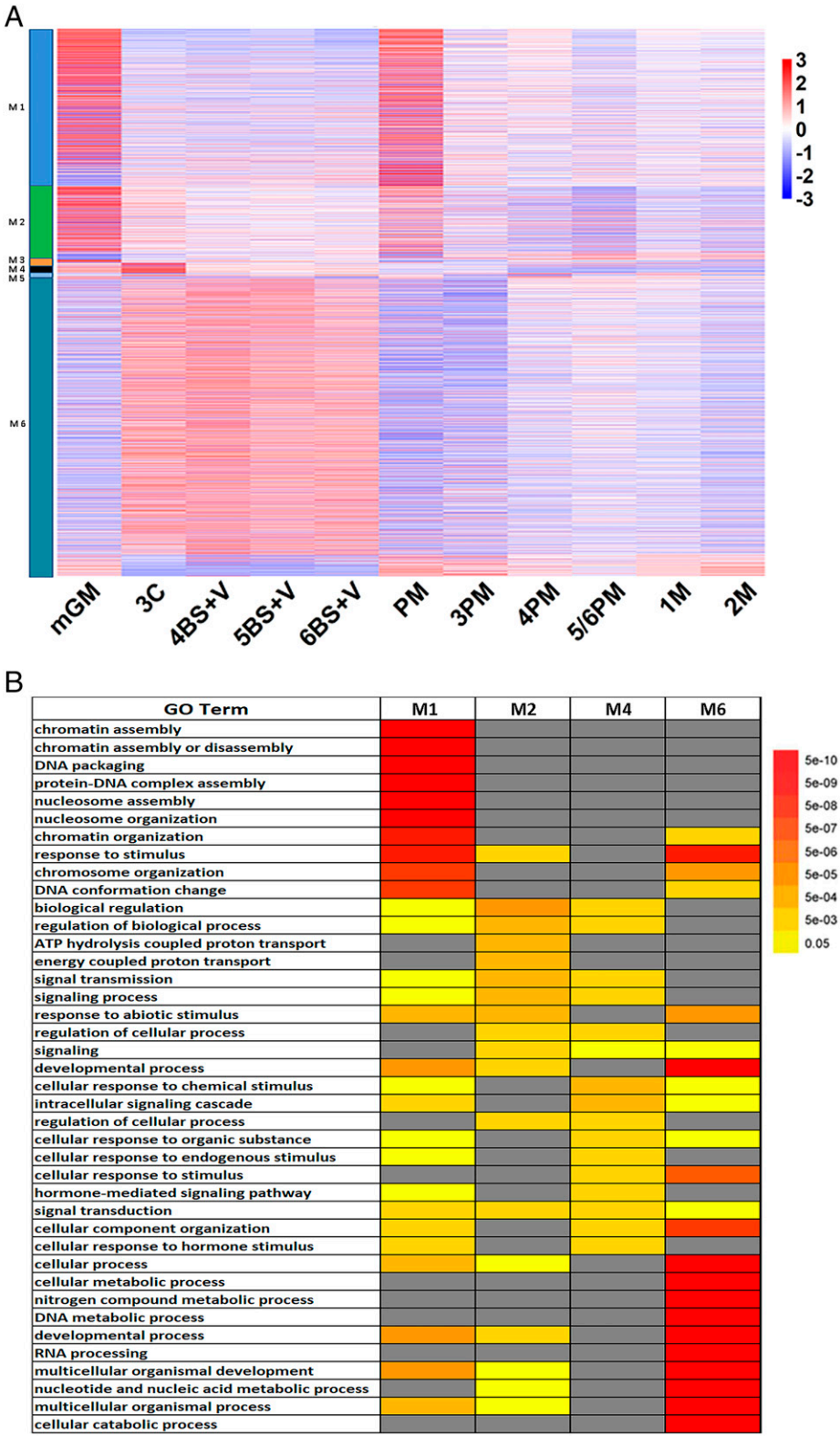

Fig. 5. Biological processes enriched in gene coexpression modules of the pre-Kranz cell types. (*A*) Heat map of the scaled FPKM values (*z* scores) of the expression levels of genes in the six gene coexpression modules of pre-Kranz genes in the 11 cell types. The color bar on the left shows the six modules (e.g., the M1 module at the top), while the color bar on the right shows the *z* scores in color. ATP :adenosine triphosphate (*B*) Enrichment of biological processes in the pre-Kranz compartment-correlated modules. The heat map shows enrichment in the top 10 biological process GO terms inferred for the M1, M2, M4, and M6 modules based on the multiple hypothesis tests (false discovery rates (FDRs) < 0.05) (see GO terms in Dataset S4). The yellow to red color indicates the level of enrichment from low to high significance, respectively, and gray indicates "not significant."

promoters in the corresponding target genes (*SI Appendix*, Fig. S4). A total of 67 transcriptional regulatory pairs are assayed, and 58 pairs are successfully validated (>85%). Only nine pairs fail (*SI Appendix*, Fig. S4 and Table S5). Some of the transcriptional regulatory pairs are supported by the literature. For example, AtANT contributes to organ polarity and regulates the growth of lateral organs in combination with YABBY genes (37). As another example, AtPIN3 is a direct target of AtARF7 (the

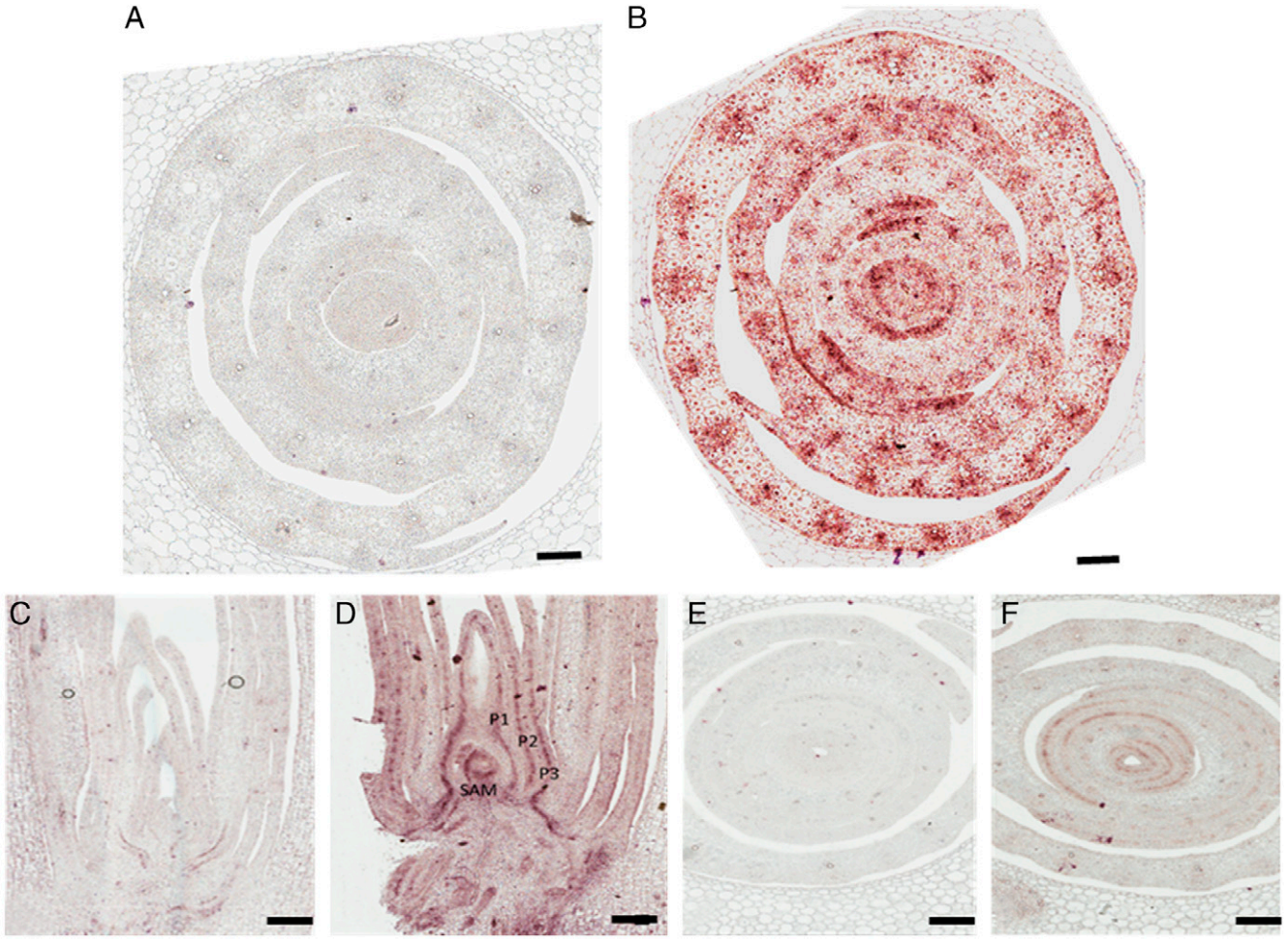

Fig. 6. Localization of indole-3-acetic acid (IAA) and *ZmLAX2* of the maize seedling shoot. (*A* and *B*) Immunohistochemical localization at the shoot tip (*A*) and vasculature (*B*), showing strong IAA signals in the shoot apical meristem (SAM) and the P1 to P3 primordia. (Scale bars: 50 μm.) (*C*–*F*) In situ hybridization of *ZmLAX2* transcript accumulation in the maize seedling shoot. Longitudinal (*C* and *D*) and transverse (*E* and *F*) sections of wild-type seedling apices reveal a punctuated, interspersed accumulation pattern of *ZmLAX2* transcripts (purple) on the vasculature of the SAM and the P1 to P3 primordia. (Scale bars: 50 μm.)

ortholog of ZmARF20) during lateral root development (38). Overall, these findings support the reliability of the predicted transcriptional networks.

Discussion

Pre-Kranz and Pre-M Cell Isolation and Transcriptome Characterization. In this study, we developed a special LCM protocol that allowed for microscopic isolation, amplification, and high-throughput sequencing of nanogram amounts of RNA from specific pre-Kranz and pre-M cells collected from fixed maize embryonic leaves (Fig. 1 and *SI Appendix*, Table S1). The high-quality RNAs isolated from the laser-captured cells (Fig. 1 and *SI Appendix*, Fig. S2 and Table S1) and the highly reproducible RNA-seq data (*SI Appendix*, Table S2) enabled us to detect the mRNAs of 44,483 genes accumulated in at least one captured cell compartment and 20,352 genes expressed in at least one compartment (Dataset S1). We used an LCM RNA-seq profiling approach to comprehensively detect mRNA populations for five pre-Kranz and six pre-M cell compartments of 3-d-old maize embryonic leaves and obtained the transcriptome for each type of cell.

By clustering analysis of gene expression profiles, we classified the DEGs into eight groups (Fig. 3). A surprising observation is that there are far more genes (six groups) that show a higher expression level in pre-M cells than in pre-Kranz cells, indicating that more genes undergo expression differentiation in pre-M cells than pre-Kranz cells. On the other hand, groups C7 and C8 show higher gene expression levels in pre-Kranz cells than in pre-M cells. While C7 includes only 1 TF, C8 includes 12 TF genes, so they might play important roles in pre-Kranz cell fate determination and subsequent differentiation. The WGCNA provided a more detailed clustering analysis and classified the genes into 15 modules (Fig. 4). Indeed, 5 (*LBD33*, *ZmBBX3*, *GATA32*, *WRKY94*, and *DOF25*) of the 12 TF genes in the C8 group in Fig. 3 belong to the M1 module (*SI Appendix*, Fig. S5), and the remaining seven genes (*YABBY1*, *ZmANT2*, *DOT5*, *ZmANT3*, *ERF130 [AIL6]*, *YABBY3*, and *ZmANT1*) belong to the M2 module (Dataset S3 and *SI Appendix*, Table S3). Thus, WGCNA provides more detailed information than the simple clustering analysis; however, the latter provides more directly interpretable groupings.

Among the 15 gene coexpression modules identified by WGCNA, the M2 module is the most interesting one as discussed in some details above. The M1 module is also of special interest because it includes genes preferentially expressed at the mGM stage (*SI Appendix*, Fig. S5), so it might play a role in the initial development of Kranz anatomy. Among the 19 TFs in M1, the PWMs of 9 TFs were found in the CIS-BP Database (39) (*SI Appendix*, Table S6). A candidate GRN of the

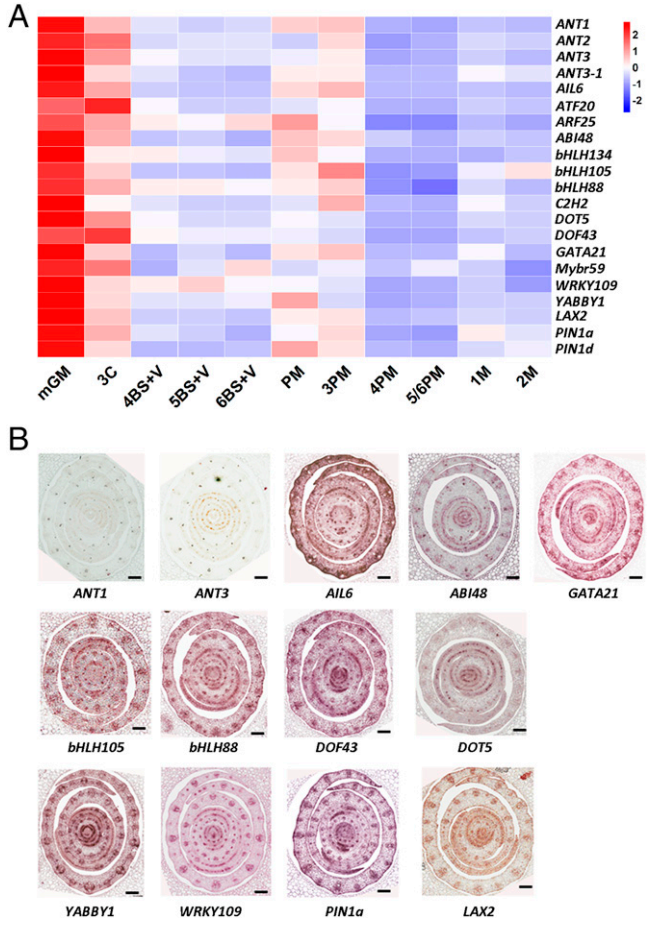

Fig. 7. Gene expression profile and localization of candidate genes of the maize seedling shoot. (*A*) The heat map of expression values (*z* scores) of the 21 candidate GRN genes in the 11 cell types. The color bar is shown on the right. (*B*) In situ hybridization of 7-d-old maize seedlings for gene transcript localization in the pre-Kranz and vascular traces of leaf primordia. The names of the 11 TFs and two auxin transporters are indicated under the figures. (Scale bars: 100 μm.)

M1 module was then inferred by combining the gene coexpression network and the scanned potential TFBSs in promoters (*SI Appendix*, Fig. S7). In this network (*SI Appendix*, Fig. S7 and Table S7), ZmSHR1 is a well-known regulator of Kranz anatomy development (11), and ZmSCL3 is a paralog of ZmSCR, which is another well-known regulator of Kranz anatomy development (8). The two TFs (*ZmDOF25* and *ZmDOF39*) at the top of the network are involved in vascular system development (40). Moreover, the GRN also includes *ZmTCP5*, *ZmTCP38*, and *ZmTCP40*, which may be involved in the regulation of cell division to control leaf morphogenesis (41). The putative functions of the remaining TFs in this GRN are described in *SI Appendix*, Table S7. Thus, this network is apparently involved in the regulation of Kranz anatomy development in maize leaves.

The M3 module is also of special interest because its genes are only preferentially expressed at the mGM stage (Fig. 4*B* and *SI Appendix*, Fig. S7), so it might play an import role in the transition from mGM to 3C. This module includes only two TFs, *ZmINDETERMINTE DOMAIN 14 (IDD14)* (Zm00001d020683) and *ZmTCP30* (Zm00001d043368), both of which are preferentially expressed at the mGM stage. Their PWMs are found in the CIS-BP Database (39) (*SI Appendix*, Table S7). ZmIDD14 is indirectly implicated in the control of vein spacing (42, 43), and the SHR/SCR/IDD network might wire auxin and brassinosteroid (BR) signaling to coordinate leaf vein density (42). The function of ZmTCP30 may be similar to that of ZmTCP5, which may be involved in the regulation of cell division to control leaf morphogenesis (41).

Transcript and Tissue Sample Clusters. For the 11 different cell-type RNA-seq libraries obtained in this study, the number of genes expressed in at least 1 library was 20,352. Construction of the regulatory network by WGCNA broadly reduced the number of major gene expression profiles. Many expression profiles are defined by genes that express much higher in one cell type than in the others. This is similar to an LCM study of endosperm cell types (44), highlighting the importance of examining molecular activities, such as gene expression, at the cellular level because pooling cells of different types together leads to loss of regulatory details and the relationship between gene expression (or function) and tissue (or cell localization). This relationship can be better expressed in terms of statistical correlations (Fig. 4).

The three stages of maize Kranz anatomy development are 1) initiation of procambium, 2) BS and M cell differentiation, and 3) chloroplast development and the integration of the C_4 cycle (5). The M1 to M6 coexpression modules contain genes preferentially expressed in mGM-, 3C-, 4BS+V-, 5BS+V-, or 6BS+V-stage cells (Fig. 4). The functional enrichment analyses revealed that the M1, M2, M4, and M6 modules are composed of genes more highly expressed at the mGM, 3C, and 4BS+V stages (Fig. 5*A*). The GO terms of mGM-associated M1 and M2 modules are related to the initiation of procambium. The M4 coexpression module may be involved in the BS and M cell differentiation at the 3C stage (Fig. 4 and Dataset S4). The M6 module may be involved in the Kranz anatomy development and the integration of the C_4 cycle at the 3BS+V to 6BS+V stages (Dataset S4). These observations indicate that the coexpression modules can be used to decipher the key regulatory programs associated with cellular differentiation in Kranz anatomy development.

Auxin-Mediated Kranz Anatomy Development. The enrichment of auxin-related genes in modules relevant to BS+V cell fate determination and differentiation is consistent with the established role of auxin as a major positional signal for vascular tissue formation (45). An antiindole-3-acetic acid antibody was employed to determine the auxin distribution and accumulation in developing maize seedling (Fig. 6 *A* and *B*). Auxin accumulation was detectable at the tip of SAM and in SAM and P1 to P3 primordia, and the *ZmLAX2* and *ZmPIN1a* transcripts are present in the vascular bundles (Fig. 7). These observations suggest that both auxin and auxin transport modulate embryonic leaf patterning and Kranz cell development.

The first stage of Kranz anatomy development is the initiation of procambium (5). As mentioned above, auxin and its transporters together provide a major positional signal for vascular tissue formation, and the canalization model of auxin flow predicts a feedback regulation of the auxin transport rate and polarity by a localized auxin source (45). This mechanism would be adequate to gradually generate more concentrated auxin channels that would determine the position of a new vasculature and explain the vasculature formation seen in leaves after wounding or in a newly initiated organ (46). To look for the early key regulators that regulate the positional signal of Kranz anatomy development, we studied the mGM and PM stages. A GO analysis of the DEGs between mGM and PM showed that they were enriched in PM for the terms "xylem and phloem pattern formation," "positive gravitropism," and "developmental process" and in mGM for the terms "response

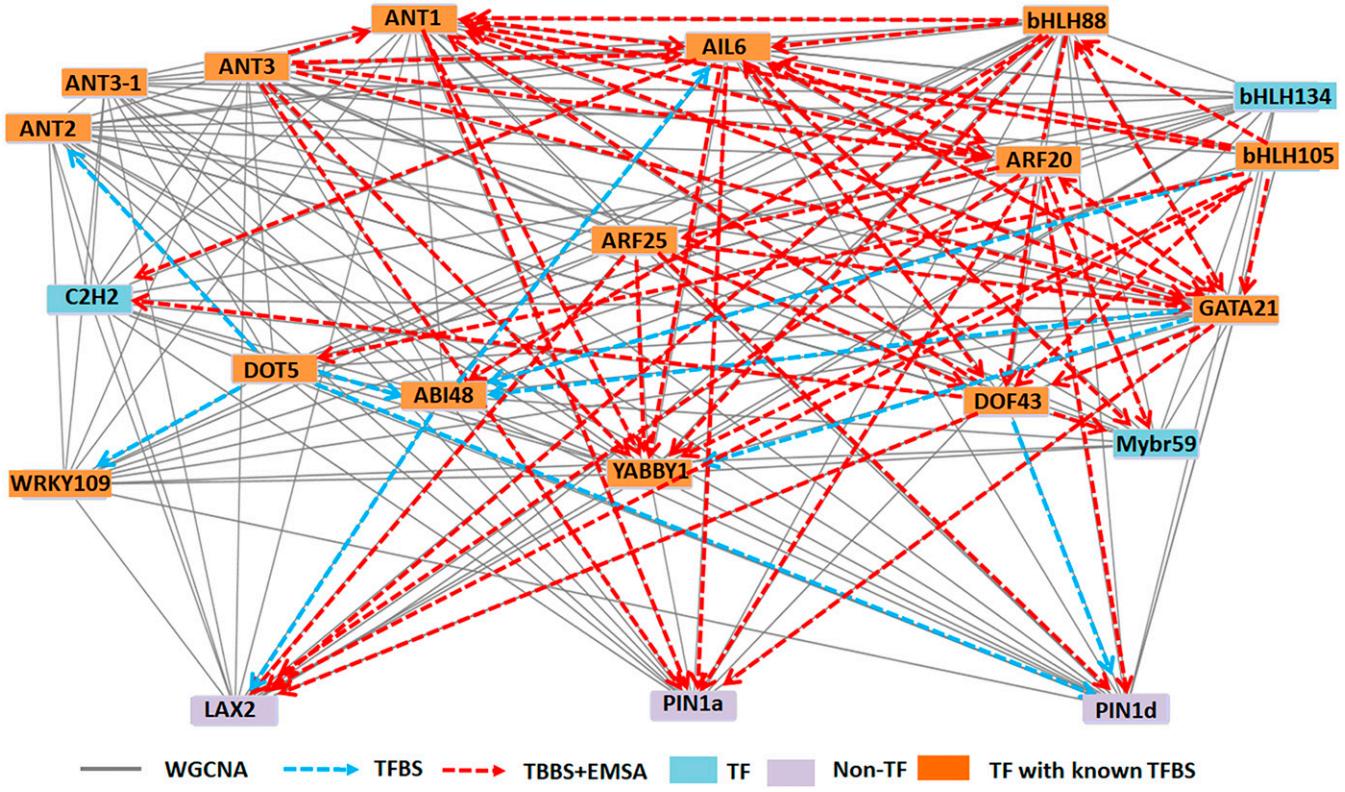

Fig. 8. Predicted directional GRN associated with the process of Kranz cell differentiation. A regulatory network of inferred connections between genes, including 18 TF genes and the genes for *LAX2*, *Pin1a*, and *Pin1d*, in the M2 coexpression module identified by WGCNA. Each TF with known TFBS is shown in an orange box, while the other TFs are in blue boxes; the non-TF genes are in purple boxes. Known TFBSs of TFs are used to check the presence of their mapped sites in the promoter sequences of their candidate downstream genes. The TFs and genes in the predicted regulatory pathway are visualized using Cytoscape [Shannon et al. (36)]. A gray line denotes a connection of coexpression inferred by WGCNA. Red and cyan dashed lines denote the presence of the TFBS in the promoter of the target gene and TF binding validated by EMSA, respectively.

to abiotic stimulus," "secondary metabolism," "photosynthesis," "cuticle development," and "wax biosynthesis," etc. (*SI Appendix*, Fig. S7). The major "supercluster" of enriched GO terms is xylem and phloem pattern formation, and it includes many enriched GO terms related to this. Indeed, we inferred an M2 regulatory module containing 18 TF genes and three auxin transporters that are likely involved in auxin-mediated Kranz anatomy development. As an example, the transcripts of *WRKY109* are seen throughout all stages of Kranz anatomy development (mGM, 3C, 4BS+V, 5BS+V, and 6BS+V stages) (*SI Appendix*, Fig. S8). It may be a central regulator for positional signaling of Kranz anatomy development, and it governs the cell fate of Kranz cells once the positional signal is determined. From these findings, we can infer that the positional signal is established at a very early stage (the 3C stage). Therefore, further studies using TF gene mutants will be helpful for determining the full GRN.

The WGCNA data in Fig. 8 imply that some of the TFs at or near the top of the network (*ANT1*, *ANT3*, *AIL6*, and *bHLH88*) modulate the initiation and/or the early stage of Kranz anatomy development. In this connection, we note that ANT1 has indeed been shown to modulate Kranz anatomy and vascular development and that ANT3 and AIL6 (ANT-like protein) are two paralogues of ANT1 (12). Moreover, *bHLH88*, the ortholog of *Arabidopsis LJRHL1-LIKE 1* (*LRL1*), interacts with two PERICYCLE FACTOR TYPE-A (PFA) proteins and PERICYCLE FACTOR TYPE-B (PFB), which is required for the competence of pericycle cells to initiate lateral root primordium formation (47). Overexpression of PFA genes confers hallmark pericycle characteristics, including specific marker gene expression and auxin-induced cell division, and multiple loss-of-function mutations in PFA genes or the repression of PFB target genes results in the loss of this specific pericycle function (47). *ZmDOF43*, the ortholog of *AtDof5.6/HCA2*, was also expressed in the pericycle of primary roots (31). Considering the reported roles of LRL1 and also, the up-regulation in mGM and PM of its maize ortholog, bHLH88, the pericycle developmental pathway may be co-opted in C_4 plants for Kranz anatomy development.

In summary, our dataset provides a high-resolution atlas of gene expression in differentiating pre-Kranz cell compartments and pre-M cell compartments of early maize embryonic leaves. This dataset provides insights into the functions of pre-Kranz cell types and identifies the coexpressed gene sets that establish the differentiated states and functions of these cell types. Furthermore, as exemplified by our analysis of the auxin-mediated regulatory module, our dataset can be used to dissect the modules regulating Kranz anatomy development. The findings from this study, therefore, constitute a significant advancement toward the identification of GRNs that regulate maize Kranz anatomy development.

Methods

Plant Growth Conditions and Sample Preparation. Seeds were surface sterilized, imbibed, and germinated on wet paper towels. Approximately 72 h after imbibition, embryonic leaves from 3-d-old seedlings were still enclosed in coleoptiles, and the embryonic leaves were fixed with ethanol/acetic acid (3:1) overnight and infiltrated with 10 to 15% sucrose prior to freezing in isopentane for generating sections suitable for LCM (48). Tissues were dehydrated through an ethanol gradient and stained in xylene before proceeding to LCM. Then, we used LCM to microdissect the cells of interest according to the manufacturer's instructions. Based on the cell diameters, the cell types were isolated using a

15-μm laser beam; the laser power was set at 40 mW, and the laser pulse duration was 20 μs. Specific cell types, including pre-Kranz cells (mGM-, 3C-, 4BS-, 5BS-, and 6BS-stage GM cells) and pre-M cells (PM, 3PM, 4PM, 5/6PM, 1M, and 2M cells), were identified by staining before being captured (Fig. 1). Then, these cells were separately harvested and captured on the cap (Fig. 1). For each cell type, we obtained more than 1,000 cells per biological replicate and collected batches in duplicates or triplicates. Total RNA was extracted from the collected cells using the MN NucleoSpin RNA XS (Macherey-Nagel). DNase I treatment was performed during RNA isolation using on-column DNase digestion. The quality and profile of the RNA samples were checked using the Agilent 2100 Bioanalyzer (Agilent Technologies). Good quality and quantities of RNA were obtained for RNA-seq (*SI Appendix*, Fig. S2 and Table S1).

Developing a Low-Input RNA-seq Library Technique and Constructing RNA-seq Libraries. Due to the scarce amount of RNA isolated from LCM-dissected cells, the RNA was subjected to preamplification before strand-specific RNA-seq library construction. aRNA amplification was conducted using the TargetAmp 1-Round antisense RNA (aRNA) Amplification Kit (Epicentre; now Lucigen). Briefly, the first-strand cDNA was synthesized by SuperScript III (Thermo Fisher Scientific) with 3′-oligo-dT primer engineered with the T7 promoter sequence, followed by second-strand synthesis to complete the double-stranded cDNA. The in vitro aRNA amplification was conducted with T7 RNA polymerase, followed by deoxyribonuclease (DNase) I treatment and purification with the MinElute kit (Qiagen). The purified aRNA samples were quantified by the Qubit RNA assay (Thermo Fisher Scientific) and profiled by the BioAnalyzer RNA Pico assay (Agilent). The resulting aRNA showed a profile from 50 bp to over 4 kb. For RNA-seq library construction, input of 40 ng of aRNA from each sample was used along with the KAPA Stranded mRNA Sample Prep kit (KAPA), starting from RNA fragmentation, cDNA synthesis, A tailing, and adaptor ligation. The barcoded libraries were examined by Qubit and the BioAnalyzer HS DNA assay (Agilent). The adaptor dimers (128 bp) in the libraries were removed by gel size selection using Blue Pippin with 1.5% agarose gel cassette (Sage Science). For equal pooling of the barcoded libraries for sequencing, the final stranded aRNA sequencing (aRNA-seq) libraries were normalized by qPCR for molar concentration using KAPA Library Quantification Kit Illumina Platforms (Kapa Biosystems).

Obtaining Transcriptomes by RNA-seq. The RNA-seq libraries were sequenced using the Illumina HiSeq2500 High Throughput v4 mode with the PE2*125nt format. The raw reads were quality checked and mapped to the maize reference genome (B73 RefGen_v4, AGPv4) (49) by Bowtie2 software (version 2.2.3) (50) and TopHat2 (version 2.0.14) (51). Expression abundances (FPKMs) for *Zea mays* B73 AGPv4 genes (49) were quantified with Cufflinks (version 2.2.1) (52). Of the reference-mapped reads (8.0 to 11.1 million, 63 to 92% of total reads), 3.2 to 11.3 million (54 to 68%) were aligned to exonic regions (*SI Appendix*, Table S2).

Identifying DEGs between Pre-Kranz and Pre-M Cells. To identify the genes that are differentially expressed between pre-Kranz cells and pre-M cells, we used NOISeq (53, 54) to compare gene expression levels in pre-Kranz and pre-M cells with $q \geqq 0.8$ as the criterion for DEGs. A total of four pairs of samples were compared, including mGM vs. PM, 3C vs. 3PM, 4BS+V vs. 4PM, and 5BS+V vs. 5/6PM. The DEGs in each comparison were collected, and their expression levels were normalized in terms of *z* scores. According to the *z* scores, genes were clustered into eight groups using the K-means method with $k = 8$.

Gene Coexpression Network Analysis. The gene coexpression network analyses were carried out using WGCNA in the R package (19, 20). Before the network construction, the proper soft-thresholding power was determined by an analysis of the network topology. We used the block-wise network construction option with a soft-thresholding power value of 13. Genes with a module membership (also known as kME (eigengene-based connectives)) of >0.3 were assigned to an eigengene module or to an M16 module if they did not meet the above criteria. The correlation between each gene pair was calculated to establish a similarity matrix. Using a range of soft-threshold values, the average connection network and the fitting evaluation network of scale-free topology models showed an approximate scale-free topology. The adjacency was converted into a topological overlap matrix (TOM), and all coding sequences were hierarchically clustered by TOM similarity. The Dynamic Tree Cut method (55), which merged highly correlated modules using a height cut less than 0.25, was used to determine the coexpression gene modules of the gene dendrogram. Finally, we confirmed the stage-specific modules according to the distinguished interrelationship between module membership and the significance gene. By using Cytoscape version 3.4.0 (36), the highly connected genes of specific modules at each developmental stage were visually identified. Differential expression was determined using NOISeq (53, 54) with $q \geqq 0.8$. GO enrichment analysis was performed using the agriGO version 2.0 (21).

In Situ Hybridization Localization of mRNAs. Embryonic leaves obtained from maize seedlings were harvested, fixed, processed, and hybridized to probes as described in Huang et al. (56). The primers used to generate the probe clones are listed in *SI Appendix*, Table S8.

DAP-seq Protocol and Inference of the TF Binding Motif. Our DAP-seq protocol is as described in ref. 12. Genomic DNA was extracted from 4-d-old millet (*Setaria italica*) leaves, fragmented, and ligated with the truncated adaptor from the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) with five cycles of PCR amplification to derive shotgun genomic libraries. For protein expression, the TF coding sequence was cloned into the pIX in vitro expression vector with the N-terminal HALO-Tag using LR Clonase (Thermo Fisher Scientific) and expressed using the TNT T7 Coupled Wheat Germ Extract System (Promega). The GST sequence was cloned into the pIX and used as an experimental protein control. Haloalkane dehalogenase (HALO)-tagged TF and GST proteins were immobilized on Magne HaloTag beads (Promega), washed, and incubated with 350 ng of the anchor-amplified shotgun genomic library. After bead washing, the bead-bound DNA fragments were eluted and amplified with the NEB indexing primers by 10 to 21 cycles of PCR. The TF-enriched libraries were sequenced on Illumina HiSeq 2500 with the PE 2- × 101-bp mode at the High Throughput Genomics Core, Academia Sinica, Taiwan.

To infer the binding motif of a TF from DAP-seq data, the PE reads were preprocessed to remove adapters, and low-quality bases were trimmed off using Trimmomatic (version 0.39) (57) with options ILLUMINACLIP:- TruSeq3-PE.fa: 2:40:12:8:true LEADING:10 SLIDINGWINDOW:4:15 MINLEN:50. The trimmed reads were aligned using Bowtie2 (version 2) (50) to the *Setaria viridis* genome (version 2.1), in which the repetitive sequences were masked. Those reads aligned to multiple loci and those duplicated reads aligned to the same locus were filtered out using SAMtools (58). The enriched binding events were identified by comparing the reads coverage from GST TFs enrichment vs. GST-only control using MACS2 (version 2.1.2) (59) with q value < 0.05. The top 500 peaks (summits.bed by MACS2) of binding regions based on the MACS2 q values were selected to extract sequences from two 100-bp flanking regions of the peaks, 200 bp in total. The top five motifs (PWMs) were then identified using MEME-ChIP (version 5.0.5) (60). The PWM having the largest number of occurrences (at least 100 sites) among the top 500 peak sequences was selected as the PWM of the TF under study.

TF Target Gene Prediction. The prediction is as described in Liu et al. (12). We used the PWM of a TF to find its potential target genes by mapping the PWM to the promoter sequences of all candidate genes using FIMO (35) with P value < 0.005. The promoter sequence of a gene was defined as the region from −1,000 to +200 bp relative to the transcription start site of the gene (61). We then added the genes with promoters covered by DAP-seq peaks (as described above), even though they did not pass the search of the PWM sequence. From this set of genes, we selected those that showed in M2 module to form a set of candidate genes. Then, following the method of Yu et al. (61), we required that a PWM sequence be found in the promoters in *Z. mays*. These genes were considered the putative target genes of the cognate TF.

EMSAs. The protocol is as described in Liu et al. (12). The primers used to generate the biotin-labeled promoter probes are listed in *SI Appendix*, Table S9.

Data, Materials, and Software Availability. Transcriptome data have been deposited in NCBI SRA (accession no. PRJNA831613) (62). All other data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank Jane Langdale and Jen Sheen for valuable suggestions. This study was supported by Academia Sinica Grant AS-TP-109-L10. S.-H.S. was partially supported by NSF Grants IOS-1546617 and DEB-1655386 and US Department of Energy Great Lakes Bioenergy Research Center Grant BER DE-SC0018409.

Author affiliations: [a]Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan; [b]Department of Plant Biology, Michigan State University, East Lansing, MI 48824; [c]Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824; [d]Department of Bioagricultural Science, National Chiayi University, Chiayi 600, Taiwan; [e]School of Biological Sciences, Washington State University, Pullman, WA 99164; [f]Institute of Plant and Microbial Biology, Academia Sinica, Taipei 115, Taiwan; and [g]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637

1. N. G. Dengler, T. Nelson, "Leaf structure and development in C4 plants" in *C4 Plant Biology*, R. F. Sage, R. K. Monson, Eds. (Academic Press, San Diego, CA, 1999), pp. 133–172.
2. P. A. Christin *et al.*, Anatomical enablers and the evolution of C4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1381–1386 (2013).
3. R. F. Sage, R. Khoshravesh, T. L. Sage, From proto-Kranz to C4 Kranz: Building the bridge to C4 photosynthesis. *J. Exp. Bot.* **65**, 3341–3356 (2014).
4. R. F. Sage, R. K. Monson, Eds., *C4 Plant Biology* (Academic Press, San Diego, CA, 1999).
5. J. P. Fouracre, S. Ando, J. A. Langdale, Cracking the Kranz enigma with systems biology. *J. Exp. Bot.* **65**, 3327–3339 (2014).
6. A. M. Bosabalidis, R. F. Evert, W. A. Russin, Ontogeny of the vascular bundles and contiguous tissues in the maize leaf blade. *Am. J. Bot.* **81**, 745–752 (1994).
7. W. Y. Liu *et al.*, Anatomical and transcriptional dynamics of maize embryonic leaves during seed germination. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3979–3984 (2013).
8. T. L. Slewinski, A. A. Anderson, C. Zhang, R. Turgeon, Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol.* **53**, 2030–2037 (2012).
9. A. B. Feldman *et al.*, Increasing leaf vein density by mutagenesis: Laying the foundations for C4 rice. *PLoS One* **9**, e94947 (2014).
10. G. Rizal *et al.*, Two forward genetic screens for vein density mutants in sorghum converge on a cytochrome P450 gene in the brassinosteroid pathway. *Plant J. Cell Mol. Biol.* **84**, 257–266 (2015).
11. T. L. Slewinski *et al.*, Short-root1 plays a role in the development of vascular tissue and kranz anatomy in maize leaves. *Mol. Plant* **7**, 1388–1392 (2014).
12. W. Y. Liu *et al.*, Maize ANT1 modulates vascular development, chloroplast development, photosynthesis, and plant growth. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21747–21756 (2020).
13. L. Wang *et al.*, Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nat. Biotechnol.* **32**, 1158–1165 (2014).
14. C. F. Huang *et al.*, Elevated auxin biosynthesis and transport underlie high vein density in C_4 leaves. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6884–E6891 (2017).
15. P. Wang, S. Kelly, J. P. Fouracre, J. A. Langdale, Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *Plant J. Cell Mol. Biol.* **75**, 656–670 (2013).
16. P. Li *et al.*, The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* **42**, 1060–1067 (2010).
17. C. P. Yu *et al.*, Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2477–E2486 (2015).
18. C. Trapnell *et al.*, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
19. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
20. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
21. T. Tian *et al.*, agriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **45**, W122–W129 (2017).
22. J. J. Zhou, J. Luo, The PIN-FORMED auxin efflux carriers in plants. *Int. J. Mol. Sci.* **19**, 2759 (2018).
23. C. Forestan, S. Varotto, PIN1 auxin efflux carriers localization studies in *Zea mays*. *Plant Signal. Behav.* **5**, 436–439 (2010).
24. K. Bainbridge *et al.*, Auxin influx carriers stabilize phyllotactic patterning. *Genes Dev.* **22**, 810–823 (2008).
25. Y. Mizukami, R. L. Fischer, Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 942–947 (2000).
26. C. Li *et al.*, Ectopic expression of a maize hybrid down-regulated gene ZmARF25 decreases organ size by affecting cellular proliferation in *Arabidopsis*. *PLoS One* **9**, e94830 (2014).
27. H. Xing *et al.*, Genome-wide identification and expression profiling of auxin response factor (ARF) gene family in maize. *BMC Genomics* **12**, 178 (2011).
28. C. A. MacAlister, K. Ohashi-Ito, D. C. Bergmann, Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature* **445**, 537–540 (2007).
29. J. Zhang *et al.*, Sperm cells are passive cargo of the pollen tube in plant fertilization. *Nat. Plants* **3**, 17079 (2017).
30. J. J. Petricka, N. K. Clay, T. M. Nelson, Vein patterning screens and the defectively organized tributaries mutants in *Arabidopsis thaliana*. *Plant J. Cell Mol. Biol.* **56**, 251–263 (2008).
31. Y. Guo, G. Qin, H. Gu, L. J. Qu, Dof5.6/HCA2, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in *Arabidopsis*. *Plant Cell* **21**, 3518–3534 (2009).
32. T. Nawy *et al.*, The GATA factor HANABA TARANU is required to position the proembryo boundary in the early *Arabidopsis* embryo. *Dev. Cell* **19**, 103–113 (2010).
33. R. Sarojam *et al.*, Differentiating *Arabidopsis* shoots from leaves by combined YABBY activities. *Plant Cell* **22**, 2113–2130 (2010).
34. A. Bartlett *et al.*, Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* **12**, 1659–1672 (2017).
35. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
36. P. Shannon *et al.*, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
37. S. Nole-Wilson, B. A. Krizek, AINTEGUMENTA contributes to organ polarity and regulates growth of lateral organs in combination with YABBY genes. *Plant Physiol.* **141**, 977–987 (2006).
38. H. Z. Wang *et al.*, Transcriptional regulation of PIN genes by FOUR LIPS and MYB88 during *Arabidopsis* root gravitropism. *Nat. Commun.* **6**, 8822 (2015).
39. M. T. Weirauch *et al.*, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
40. R. Le Hir, C. Bellini, The plant-specific dof transcription factors family: New players involved in vascular system development and functioning in *Arabidopsis*. *Front Plant Sci* **4**, 164 (2013).
41. H. Wang *et al.*, Genome-wide identification of *TCP* family transcription factors in *Medicago truncatula* reveals significant roles of *miR319*-targeted *TCPs* in nodule development. *Front Plant Sci* **9**, 774 (2018).
42. D. Kumar, E. A. Kellogg, Getting closer: Vein density in C_4 leaves. *New Phytol.* **221**, 1260–1267 (2019).
43. C. P. Coelho, P. Huang, D. Y. Lee, T. P. Brutnell, Making roots, shoots, and seeds: IDD gene family diversification in plants. *Trends Plant Sci.* **23**, 66–78 (2018).
44. J. Zhan *et al.*, RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell* **27**, 513–531 (2015).
45. T. Sachs, Cell polarity and tissue patterning in plants. *Development* **113**, 83–93 (1991).
46. J. Hajný, S. Tan, J. Friml, Auxin canalization: From speculative models toward molecular players. *Curr. Opin. Plant Biol.* **65**, 102174 (2022).
47. Y. Zhang *et al.*, Two types of bHLH transcription factor determine the competence of the pericycle for lateral root initiation. *Nat. Plants* **7**, 633–643 (2021).
48. N. M. Kerk, T. Ceserani, S. L. Tausta, I. M. Sussex, T. M. Nelson, Laser capture microdissection of cells from plant tissues. *Plant Physiol.* **132**, 27–35 (2003).
49. Y. Jiao *et al.*, Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).
50. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
51. D. Kim *et al.*, TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
52. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
53. S. Tarazona *et al.*, Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* **43**, e140 (2015).
54. S. Tarazona, F. García-Alcalde, J. Dopazo, A. Ferrer, A. Conesa, Differential expression in RNA-seq: A matter of depth. *Genome Res.* **21**, 2213–2223 (2011).
55. P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
56. C. F. Huang *et al.*, Whole-genome duplication facilitated the evolution of C4 photosynthesis in *Gynandropsis gynandra*. *Mol. Biol. Evol.* **38**, 4715–4731 (2021).
57. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
58. H. Li *et al.*; 1000 Genome Project Data Processing Subgroup, The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
60. P. Machanick, T. L. Bailey, MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
61. C. P. Yu, J. J. Lin, W. H. Li, Positional distribution of transcription factor binding sites in *Arabidopsis thaliana*. *Sci. Rep.* **6**, 25164 (2016).
62. W-Y. Liu *et al.*, Regulators of early maize leaf development inferred from transcriptomes of laser capture microdissection (LCM)-isolated embryonic leaf cells. NCBI SRA. https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA831613. Deposited 23 April 2022.