

Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery in maize.

Xiaosa Xu¹, Megan Crow¹, Brian R. Rice², Forrest Li¹, Benjamin Harris¹, Lei Liu¹, Edgar Demesa-Arevalo¹, Zefu Lu³, Liya Wang¹, Nathan Fox¹, Xiaofei Wang¹, Jorg Drenkow¹, Anding Luo⁴, Si Nian Char⁵, Bing Yang^{5,6}, Anne W. Sylvester⁴, Thomas R. Gingeras¹, Robert J. Schmitz³, Doreen Ware^{1,7}, Alexander E. Lipka², Jesse Gillis¹, David Jackson^{1,8,*}

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

²Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

³Department of Genetics, University of Georgia, Athens, GA 30602, USA

⁴Department of Molecular Biology, University of Wyoming, Laramie, WY 82071, USA

⁵Division of Plant Sciences, Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA

⁶Donald Danforth Plant Science Center, St. Louis, MO 63132, USA

⁷USDA-ARS, Robert W. Holley Center, Ithaca, NY 14853, USA

⁸Lead Contact

* Correspondence: jacksond@cshl.edu

SUMMARY

Crop productivity depends on activity of meristems that produce optimized plant architectures, including that of the maize ear. A comprehensive understanding of development requires insight into the full diversity of cell types and developmental domains, and the gene networks required to specify them. Until now, these were identified primarily by morphology and insights from classical genetics, that are limited by genetic redundancy and pleiotropy. Here, we investigated the transcriptional profiles of 12,525 single cells from developing maize ears. The resulting developmental atlas provides a single-cell RNA sequencing (scRNA-seq) map of an inflorescence. We validated our results by mRNA *in situ* hybridization and by Fluorescence-Activated Cell Sorting (FACS) RNA-seq, and show how these data may facilitate genetic studies by predicting genetic redundancy, integrating transcriptional networks, and identifying candidate genes associated with crop yield traits.

KEYWORDS

Maize ear, meristem, scRNA-seq, developmental networks, GWAS, trait genes.

INTRODUCTION

Plant architecture is initiated by meristems made up of pluripotent stem cells and their descendants that are organized in distinct cell types and domains with specific functions. In developing maize ears, a series of meristems build inflorescence architecture, including spikelet pair meristems (SPMs) formed from the inflorescence meristem (IM) and spikelet meristems (SMs) made from the branching of SPMs (Irish, 1997). Mutant studies have identified key cell type or domain-specific regulators that orchestrate inflorescence architecture by specifying different developmental domains (Vollbrecht and Schmidt, 2009). For example, the homeodomain transcription factor encoded by *KNOTTED1* (*KNI*) is critical for meristem establishment and maintenance, and is expressed throughout shoot meristems (Jackson et al., 1994). The production of axillary meristems to elaborate branching architecture depends on expression of a basic helix-loop-helix transcription factor encoded by *BARREN STALK1* (*BAI*), expressed specifically in the adaxial meristem periphery where axillary meristems initiate (Gallavotti et al., 2004). Another transcription factor, *BRANCHED SILKLESS1* (*BD1*), is expressed at the boundary of meristems and glumes to control spikelet architecture by promoting meristem determinacy (Chuck et al., 2002), whereas *RAMOSA* genes, such as *RA1* and *RA3*, are expressed in an arc of cells at the base of meristems, to impose determinacy on spikelet branches (Sato-Nagasawa et al., 2006). Many of these key regulators have reshaped inflorescence architecture during evolution or domestication, and their discovery was enabled by the availability of mutants that block specific aspects of development. However, such insights are limited by genetic redundancy and pleiotropy, thus, a high-resolution expression atlas of specific cell types and domains is needed to gain further insights into the gene networks that control development.

Single-cell RNA sequencing (scRNA-seq) offers the opportunity to assay gene expression with high resolution, and to construct developmental maps of complex organs or organisms (Kulkarni et al., 2019, Potter, 2018). Recently, the 10x Genomics Chromium scRNA-seq platform has been used extensively to identify cell type or domain markers in Arabidopsis roots (Rich-Griffin et al., 2020), however the application of this technology to shoot tissues has been limited. As well as providing expression information, scRNA-seq data can be integrated with other genomic datasets, such as ChIP-seq identification of targets of transcription factors, or

surveys of chromatin status. Such datasets have been generated from developing maize inflorescences (Bolduc et al., 2012, Eveland et al., 2014, Pautler et al., 2015), however single-cell data has not yet been integrated.

Productivity of maize depends on development of the inflorescences, in particular the seed-bearing ear. Genome-Wide Association Studies have identified candidate genes associated with yield-related traits (Liu et al., 2020) that can guide breeding or trait engineering. Regulatory genes functioning in early stages of ear development show significant association with ear yield traits (Vollbrecht and Schmidt, 2009, Bommert et al., 2013, Je et al., 2016, Liu et al., 2020), yet it remains challenging to identify and validate such regulators on a genome-wide scale. To fill these gaps, we optimized a protocol using 10x Genomics scRNA-seq technology to generate a high-resolution transcriptome atlas of the developing maize ear inflorescence. We illustrate how this data can enhance maize genetics by predicting genetic redundancy, build transcriptional regulatory networks at cell-type resolution, and identify candidate loci associated with ear yield traits.

RESULTS

Construction of a Single-Cell Transcriptome Atlas of the Developing Maize Ear

To generate a single-cell atlas from developing maize ears, we used the 5-10mm stage, where major developmental and architectural decisions, including meristem initiation, maintenance and determinacy, organ specification and differentiation of vascular and ground tissues, are being made (Irish, 1997, Vollbrecht and Schmidt, 2009). We first optimized a cell wall digestion method, taking into account the different composition of grass cell walls (Ortiz Ramírez et al., 2018), that allowed us to isolate ear protoplasts within ~45 minutes (see STAR Methods). However, developing ear protoplasts were fragile, and we removed small debris and organelles from broken cells (Figure S1A) by filtration followed by FACS (see STAR Methods; Figure 1A) before loading into the 10x Genomics Chromium Controller. Then, scRNA-seq libraries were generated and sequenced on the Illumina platform (Table S1). In total, we profiled 12,525 individual cells from three independent replicates (Table S1), and detected expression

from 28,899 genes using maize V3 reference, comparable to the number detected by bulk RNA-seq of the same tissue (Eveland et al., 2014, Pautler et al., 2015).

Technical variation and sparse data in scRNA-seq make it challenging to identify reproducible clusters (groups of cells) that represent homogenous cell types across technical replicates (Crow et al., 2018). We used MetaNeighbor to ask how well the identity of cells in a given cluster of one replicate can be predicted based on their similarity to a cluster from another replicate (Crow et al., 2018), reported as the average area under the receiver operating characteristic curve (AUROC). All cluster pairs with AUROCs > 0.9 in both directions, across at least two replicates, were used to merge and identify 12 replicable cell identity clusters (hereafter referred to as meta-clusters) (Figure 1B). As some genes may be affected by protoplasting (the process to isolate protoplasts), we used mRNA-seq to compare total ear protoplasts with freshly dissected, intact developing ear tissue, and identified 713 protoplasting-responsive genes ($FDR < 0.05$, $|\log_2FC| > 2$, Table S1). After excluding these genes, $> 97\%$ of the highly variable genes used for generating clusters were unchanged, as were the meta-clusters (Figures S1B and S1C). Thus, protoplasting-responsive genes did not affect clustering, as reported previously (Ma et al., 2020).

Prediction and Validation of Meta-Cluster Identities

We visualized meta-clusters using a Uniform Manifold Approximation and Projection (UMAP) plot, where we could track the distribution of genes of interest (Figure 1C). Next, to predict identities for each meta-cluster, we compiled a list of known or predicted inflorescence development marker genes, whose expression patterns have been studied in maize or Arabidopsis (Table S1). Among them, 74 are functionally characterized by their mutant phenotype in maize. Importantly, we detected the expression of 73 of these genes, and each had enriched expression in one or more of the 12 meta-clusters (Figures 1D-1N and S2; Table S1). For example, to identify meristem cell types, we used *KN1*, which is expressed throughout the meristem as well as the developing stem and vascular tissues, but strictly excluded from the epidermis and determinate lateral organs (Jackson et al., 1994). *KN1* was expressed in multiple (10 out of 12) meta-clusters, as expected (Figure 1D). Ear architecture is governed by branching events that are controlled by genes expressed in different meristem domains. To identify meta-

clusters representing these domains, we searched for expression of characterized marker genes, including *BDI*, which is expressed at the boundary of spikelet meristems (Figures 1E and 1O) (Chuck et al., 2002), and found it to be uniquely expressed in meta-cluster 9, identifying this as a meristem boundary meta-cluster (Figure 1E). Another well-characterized marker, *BAI*, is expressed in a distinct adaxial meristem periphery domain (Figures 1F and 1O) (Gallavotti et al., 2004), and was expressed in meta-cluster 11 (Figure 1F). A third cellular domain that controls branching is marked by an arc of expression of *RAMOSA* genes at the meristem base (Figures 1G and 1O) that partially overlaps with *BAI* (Satoh-Nagasawa et al., 2006, Vollbrecht and Schmidt, 2009). Correspondingly, we found expression of *RAMOSA* genes in meta-cluster 11 (Satoh-Nagasawa et al., 2006), similar to *BAI*, as well as in meta-cluster 10 (Figure 1G). As such, we could identify three of the *KN1* expressing meta-clusters as distinct branching domains.

KN1 expression is excluded from the meristem epidermis and determinate lateral organs (Jackson et al., 1994). We found epidermis marker gene *ZmHOMEODOMAIN LEUCINE ZIPPER IV8* (*ZmHDZIV8*) (Javelle et al., 2011) was highly enriched in the two meta-clusters, 3 and 6, that did not express *KN1* (Figure 1H), while determinate lateral organ marker gene *ZmYABBY14* (*ZmYAB14*) (Strable et al., 2017) was expressed throughout meta-cluster 3 (Figure 1I), suggesting that meta-cluster 3 was a determinate lateral organ meta-cluster while meta-cluster 6 corresponded to meristem epidermis cells (Figure 1O). Consistently, meta-cluster 3 was significantly enriched for organ initiation genes from a maize shoot laser capture microdissection (LCM) study ($q < 0.001$) (Table S1) (Knauer et al., 2019). Using additional markers, we identified four distinct sub-clusters of meta-cluster 3 (Figure S2A), including determinate lateral organ epidermis domain (marker = *LIPID TRANSFER PROTEIN1* (*LTP1*)) (Takacs et al., 2012), and non-epidermis domain (marker = *ZmRIBULOSE BISPHOSPHATE CARBOXYLASE SMALL SUBUNIT 1A* (*ZmRBCS1A*)), as well as adaxial domain (marker = *DROOPING LEAF2/ZmYABBY7* (*DRL2/ZmYAB7*)) and abaxial domain (marker = Homolog of Arabidopsis *AUXIN RESPONSE FACTOR3/4* (*ZmARF3/4-LIKE1*)) (Chitwood et al., 2007).

KN1 is also expressed in developing vascular tissues (Jackson et al., 1994). To identify distinct vascular meta-clusters, we used maize homologs of Arabidopsis genes (De Rybel et al., 2016). For example, xylem marker *ZmTARGET OF MONOPTEROS 5* (*ZmTMO5*), and its paralogs *ZmTMO5-LIKE1* and 2, were expressed in meta-cluster 4 (Figures 1J, 1O and S2B). We

also found a sub-cluster of predicted maturing xylem cells in meta-cluster 4 (Figure S2B), using homologs of Arabidopsis marker genes for secondary cell walls and tracheary elements, including *ZmMYB DOMAIN PROTEIN 46* (*ZmMYB46*) (Zhong et al., 2007) and *ZmXYLEM CYSTEINE PEPTIDASE 2* (*ZmXCP2*) (Funk et al., 2002). In contrast, we found that meta-cluster 5 represented phloem cells, as shown by specific expression of a maize homolog of Arabidopsis *ALTERED PHLOEM DEVELOPMENT* (Figures 1K and 1O) (De Rybel et al., 2016). Phloem tissues include distinctive sieve element and companion cells, which were reflected in sub-clusters marked by maize homologs of Arabidopsis protophloem sieve element marker *PHLOEM EARLY DOF 1* (*PEAR1*)/*PEAR2* (Miyashima et al., 2019) and companion cell marker *PHLOEM PROTEIN 2-LIKE A1* (Figure S2C) (Guo et al., 2018). Therefore, meta-clusters 4 and 5 correspond to xylem and phloem cells, respectively, and we found significant enrichment of genes in these meta-clusters with vascular markers in a maize LCM study ($q < 0.001$) (Table S1) (Knauer et al., 2019). In addition to meta-cluster 4 and 5, cells in meta-cluster 12 also expressed vascular marker genes ($q < 0.001$) (Table S1) (Knauer et al., 2019), including bundle sheath markers, such as *ZmSHORT-ROOT* (*ZmSHR1*) (Figures 1L and 1O) (Chang et al., 2012), thus meta-cluster 12 was predicted to be bundle sheath.

The remainder of the developing ear corresponds to ground tissue, including the outer cortex and inner pith tissues (Figures 1M-1O and S2D). A maize homolog, *GRMZM2G345700*, of the most unique Arabidopsis root cortex marker, *ATIG62510*, a bifunctional lipid-transfer / 2S albumin superfamily gene (Denyer et al., 2019), had restricted expression in meta-cluster 1 (Figure 1M), as did stem cortex marker homolog *ZmNITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER FAMILY 6.4-LIKE 2* (*ZmNPF6.4-LIKE2*) (Figure S2D) (Tong et al., 2016). We thus predicted this meta-cluster to be cortex. In contrast, we predicted that meta-cluster 8 was comprised of pith cells, by specific expression of homologs of sorghum or Arabidopsis markers, such as *ZmNO APICAL MERISTEM DOMAIN CONTAINING (NAC) TRANSCRIPTION FACTOR 122* (*ZmNAC122*), *GRMZM2G430849*, a homolog of *Sobic.006G147400* (Figure 1N) (Fujimoto et al., 2018), or *GRMZM2G039074*, a homolog of *AT2G3830*, an MYB transcriptional regulator (Figure S2D) (Schürholz et al., 2018). Finally, meta-clusters 2 and 7 were highly enriched for expression of cell cycle genes, such as *ZmCYCLINB1;2* (*ZmCYCB1;2*) and *ZmHISTONE2A12* (*ZmHIS2A12*), indicating that these two meta-clusters contained dividing cells at different phases of the cell cycle (Figure S2D); similar cell cycle clusters are found in

root scRNA-seq studies (Denyer et al., 2019, Rich-Griffin et al., 2020). We calculated the percentage of cells in each meta-cluster (Figure S2E). 23% of cells were from meristem domains (meta-clusters 6, 9, 10 and 11), 21% from ground tissues (meta-clusters 1 and 8), 20% from vascular tissues (meta-clusters 4, 5 and 12) and 19% from determinate lateral organ tissues (meta-cluster 3). In summary, using maize inflorescence development markers and homologs of markers from other plants, we predicted the cell or domain identities of all 12 meta-clusters (Figure 1C), and in several cases sub-divided them into more specific cell types or developmental stages.

To validate our predicted meta-cluster identities, we first used differential expression (DE) analysis to identify marker genes with AUROCs ≥ 0.7 in at least one replicate, and identified 813 candidate markers (Table S1). The top markers of each meta-cluster were further selected based on the percentage of cells expressing the marker, and showed highly enriched expression, as expected (Figure 2A). We prioritized a set of these markers by predicted developmental roles and validated them using *in situ* hybridization (Figures 2B-2M; Table S1). For example, marker genes for meta-cluster 9, *GRMZM2G004528*, annotated as *ZmMYO-INOSITOL PHOSPHATE SYNTHASE2* (*ZmMIPS2*), predicted to act in auxin signaling and transport (Chen and Xiong, 2010), and *GRMZM2G097989*, annotated as *ZmGLUTATHIONE TRANSFERASE 41* (*ZmGST41*), involved in meristem size control (Horváth et al., 2019), showed specific expression in the meristem boundary, similar to *BDI* (Figures 2B and 2C) (Chuck et al., 2002). *ZmGST41* was also a DE marker for meta-cluster 12, and consistently showed vascular trace expression (Figure 2C, arrow). Markers of a second meristem meta-cluster, 11, included *GRMZM2G038284*, which was expressed in the adaxial meristem periphery similar to *BAI* (Figure 2D), and encodes a homolog of Arabidopsis *DROUGHT INDUCED19*, of interest because maize ear development is especially sensitive to drought stress (Nuccio et al., 2015). Two additional markers of meta-cluster 11, *GRMZM2G034152*, which encodes a *ZmPOLYAMINE OXIDASE 1* (*ZmPAOI*) (Figure 2E), and *GRMZM2G430522*, a homolog of Arabidopsis *CUP-SHAPED COTYLEDON 3* (*ZmCUC3-LIKE*) (Figure S3A) also showed restricted expression at the adaxial meristem periphery. Meta-clusters 3 and 6 were predicted to have an epidermal identity, and specific expression was observed as expected for markers such as *EF517601.1_FG016*, annotated as *MALE FLOWER SPECIFIC 18* (Figure 2F), and *GRMZM2G126397*, a *ZmPHOSPHOLIPID TRANSFER PROTEIN3* (*ZmPLTP3*) gene (Figure

2G). Moving away from the meristem, marker genes for meta-cluster 3, predicted to be determinate lateral organ, included *GRMZM2G019686*, annotated as *ZmFLOWERING PROMOTING FACTOR 1* (*ZmFPF1*) (Figure 2H), and *GRMZM2G075255*, annotated as *ZmECERIFERUM1* (*ZmCER1*) (Figure 2I), and showed expected expression patterns.

We also identified candidate vascular markers in meta-clusters 4 and 5. Predicted xylem markers included *ZmTARGET OF MONOPTEROS5-LIKE3* (*ZmTMO5-LIKE3*), *GRMZM2G176141* (Figure 2J), *ZmTRANSMEMBRANE AMINO ACID TRANSPORTER FAMILY PROTEIN* (*ZmTMAAT*), *GRMZM2G109865* (Figure S3B), and *ZmWALLS ARE THIN 1* (*ZmWAT1*), *GRMZM2G007953* (Figure S3C), and all showed specific expression identified by the distinctive cell walls of xylem vessels. Interestingly, some xylem markers were also expressed in meristem tips, mostly enriched in central zone (Figures S3D-S3F). We also confirmed the specific expression of meta-cluster 5 (phloem) marker, *GRMZM2G116079*, which encodes Zinc Finger Protein 30 (*ZmZNF30*), whose Arabidopsis homolog, *AT3G15680*, is predicted to be involved in RNA regulation (Figure 2K) (Gipson et al., 2020). Lastly, candidate markers from meta-clusters 2 and 7, predicted to be dividing cells, included several cyclin and histone encoding genes, such as *ZmCYCLINB2-4* (*ZmCYCB2-4*), *GRMZM2G061287* (Figure 2L), and *ZmHISTONE2A* (*ZmHIS2A*), *GRMZM2G305046* (Figure 2M), and had punctate expression, as expected.

FACS RNA-seq has been used to validate scRNA-seq data in Arabidopsis roots (Rich-Griffin et al., 2020). Few marker lines are available in maize, but one, *pZmYAB14-TagRFP* (Je et al., 2016), is specifically expressed in determinate lateral organs (Figure 2N). We introgressed this reporter into a *bdl;Tunicate* (*bdl;Tu*) double mutant background, which produces highly proliferative ears, to generate large amounts of ear tissue. We made protoplasts from this tissue, and used FACS to sort RFP positive cells, followed by RNA-seq to identify lateral organ domain-specific genes (Figure 2N; Table S2). We found highly enriched expression of *ZmYAB14* and other *YAB* genes (Strable et al., 2017) in the FACS sorted cells, while negative control markers, such as *KNI* (Jackson et al., 1994) were significantly depleted (Figure 2O), as expected. We identified 2,040 differentially expressed genes (FDR < 0.05) (Table S2), and as we expected the majority were differentially expressed (AUROC score 0.8) in scRNA-seq meta-cluster 3, with predicted lateral organ identity, validating our scRNA-seq data (Figures 2P and 2Q).

We also used FACS sorted *pZmYAB14-TagRFpt* expressing cells in an ATAC-seq experiment to investigate how chromatin accessibility changes during differentiation of determinate lateral organs in the ear. Genome-wide analysis of accessible chromatin regions (ACRs) found that 31% mapped within 10kb upstream of the transcription start site (TSS), and 18% localized to transcription termination sites (TTS), Untranslated Regions (UTRs), exons or introns, comparable to whole ear tissue ATAC-seq results (Figure S3G; Table S2) (Ricci et al., 2019). 60% of all maize genes had ACRs in *pZmYAB14-TagRFpt* FACS sorted cells (Figure 2R; Table S2), and this value was significantly enriched for DE genes from scRNA-seq meta-cluster 3 (71%; $p < 0.001$, chi-squared test, Figure 2R; Table S2). As expected, several *ZmYAB* genes that were significantly enriched in scRNA-seq meta-cluster 3 had accessible chromatin, (Figure S3H), and we validated the expression of two meta-cluster 3 marker genes with ACRs by *in situ* hybridization (Figures S3I and S3J), including *GRMZM2G026556*, a homolog of Arabidopsis *BLADE ON PETIOLE2*, which controls lateral organ fate (Ha et al., 2007), and *GRMZM2G004012*, homolog of Arabidopsis *PLANTACYANIN*, that plays a role in the development of reproductive organs (Dong et al., 2005). To gain further insight into these data, we subtracted the determinate lateral organ specific ATAC-seq peaks from whole developing ear ATAC-seq peaks (Ricci et al., 2019) to predict ACRs specific to indeterminate meristem tissues, as well as vascular and ground tissues (Figure S3K, Table S2). scRNA-seq marker genes corresponding to these domains (Figure S3K; Table S1) were significantly enriched with these ACR associated genes ($p < 0.001$, chi-squared test; Figure S3K; Table S2). Thus, the integration of ATAC-seq and scRNA-seq may provide insights into chromatin accessibility and its effect on gene expression in specific cell types or developmental contexts.

scRNA-seq Networks Predict Redundancy

Gene redundancy often masks the phenotype of single-gene knockouts (Lloyd and Meinke, 2012), however, distinguishing redundant from non-redundant paralogs can be challenging. In a recent study of the maize branching mutant *ramosa3* (*ra3*), we identified a *ra3* enhancer as its' paralog, *ZmTREHALOSE PHOSPHATE PHOSPHATASE 4* (*ZmTPP4*) (Claeys et al., 2019). Among 12 maize *ZmTPP* genes, two of them, *ZmTPP4* and *ZmTPP12*, are up-regulated in *ra3* mutants (Eveland et al., 2014), a common predictor of compensating redundant paralogs (Rodriguez-Leal et al., 2019). However, CRISPR knockouts of *ZmTPP12* do not affect

ear development, nor do they enhance *ra3* (Claeys et al., 2019). Neither of the two paralogs is more similar in sequence to *RA3*, so to ask why *ZmTPP4*, and not *ZmTPP12*, acts as a redundant compensator, we queried their co-expression. In an aggregate network across 89 maize bulk tissue RNA-seq datasets (Lee et al., 2020), *RA3* and its two paralogs had similar co-expression scores (Figure S4A). In contrast, *ZmTPP4* was highly co-expressed with *RA3* in our single-cell data, similar to a *RA3-RA1* positive control (Sato-Nagasawa et al., 2006), whereas *ZmTPP12* was not (Figure S4B). Thus, functional redundancy in maize ear branching could be predicted by co-expression in scRNA-seq, but not in bulk tissue RNA-seq networks. To test this idea further, we identified a small gene family of *VASCULAR PLANT ONE-ZINC-FINGER* (*ZmVOZ*) genes, whose homologs regulate flowering time in Arabidopsis (Yasui et al., 2012). Two of them, *ZmVOZ4* and *ZmVOZ5* exhibited highly similar co-expression within our scRNA-seq data (Spearman correlation 0.88, Figures S4C-S4H), and we identified a common set of high confidence genes that showed consistent co-expression with both genes (FDR < 0.05, Table S3). To ask if these paralogs acted redundantly, we made CRISPR/Cas9 knockouts of all *ZmVOZ* members, including *ZmVOZ1* and 2 that were not detected in our scRNA-seq dataset, possibly due to their low expression in ears. As predicted, single *Zmvoz* mutants had no obvious phenotype, but *Zmvoz1,2,4,5* quadruple mutants were severely delayed in the floral transition, reminiscent of *voz1,2* double mutants in Arabidopsis (Figure 3A) (Yasui et al., 2012). Therefore, these two examples highlight the utility of maize ear scRNA-seq data in predicting genetic redundancy.

Using scRNA-seq to Build Transcriptional Regulatory Networks

We next asked if scRNA-seq might aid in building transcriptional regulatory networks, given that directly modulated targets of a transcription factor (TF) should be co-expressed in the same cell types. We used our scRNA-seq data to calculate co-expression of *KN1* with its published directly modulated targets (Table S3) (Bolduc et al., 2012), and found that it was significantly higher than expected compared to a control using all maize genes ($p < 0.01$), supporting our hypothesis (Figure 3B). Thus, we next generated two additional ChIP-seq datasets, for ZmHOMEODOMAIN LEUCINE ZIPPER IV6 (*ZmHDZIV6*) (Javelle et al., 2011), which was uniquely expressed in the epidermis (Figures 3C and S4I), and *ZmMADS16* (*ZmM16*) (Bartlett et al., 2015), which was expressed in specific floral organs (Figure 3D).

Biological replicates for each TF ChIP-seq had significant overlap (Figure S4J; Table S3), and we identified 907 high confidence peaks for ZmHDZIV6, and 1,155 for ZmM16 (Table S3). ~60% of these peaks mapped to gene regions, with a preference for promoters (Figures S4K and S4L), similar to other maize ChIP-seq studies (Bolduc et al., 2012). Members of the homeodomain leucine zipper IV family bind a GCATTAAATGC consensus sequence (Nakamura et al., 2006), and we found a similar sequence in motif analysis of ZmHDZIV6 bound peaks (Figure 3E). Similarly, motif analysis of ZmM16 bound peaks found an expected MADS binding motif, CC(A/T)₆GG (Figure 3F) (Aerts et al., 2018). We were thus confident in our ZmHDZIV6 and ZmM16 bound target predictions (Table S3).

Modulated TF targets are often inferred by comparison of ChIP-seq bound targets and expression changes in mutant RNA-seq. However, since ZmHDZIV6 and ZmM16 are members of large gene families and mutant RNA-seq of them was not available. We therefore asked if we could predict modulated targets based on scRNA-seq co-expression with each TF (Table S3). We thereby identified 79 and 55 candidate modulated targets for ZmHDZIV6 and ZmM16 respectively, using a Jaccard index co-expression cutoff of ≥ 0.05 (Table S3). Among the predicted modulated targets of ZmHDZIV6, we identified five additional members of the *ZmHDZIV* family (Figures S4M and S4N; Table S3), some of which are similarly expressed in the maize SAM epidermis (Javelle et al., 2011) and might form transcriptional cascades to regulate epidermal differentiation. We validated the epidermal expression of additional candidate modulated targets, including *ZmNOD26-LIKE MEMBRANE INTRINSIC PROTEIN1* (*ZmNIP1A*, GRMZM2G041980) and *ZmPRECURSOR ELICITOR PEPTIDE1* (*ZmPROPEP1*, GRMZM5G899080) (Figures 3G-3L). Homologs of these genes in Arabidopsis function as transporters (Liu et al., 2009), or in pathogen defense (Huffaker et al., 2011), suggesting similar roles in the maize ear epidermis. Similarly, among the 55 co-expressed ChIP-seq targets identified for ZmM16, we found additional members of the MADS family, such as *ZmSEP3/ZmMADS7* and *ZmAGAMOUS-LIKE 8* (*ZmAGL8*) (Figures S4O and S4P; Table S3), indicating that these genes might act downstream of ZmM16 to form gene regulatory networks controlling inflorescence development, analogous to MADS networks in Arabidopsis (Chen et al., 2018).

scRNA-seq Identifies Genes Associated with Maize Yield Traits

Maize ear morphology is associated with yield traits (Je et al., 2016, Liu et al., 2020). To ask if the cell- or domain-specific genes identified in our scRNA-seq overlapped with candidate regulators of maize yield, we used a targeted GWAS approach, by comparing our scRNA-seq marker genes from meristem, determinate lateral organ and vascular meta-clusters against a GWAS panel of 281 maize lines phenotyped for ear morphology traits related to yield (Figures 4A-4D; Table S3) (Rice et al., 2020). Using SNPs in or within 2kb of scRNA-seq marker genes, we found the meta-cluster 3 marker gene *ZmYABBY9* (*ZmYAB9*) had two significant SNPs (at 5% FDR) for cob weight (CW) (Figure 4B). We also found two significant SNPs (at 10% FDR) associated with ear diameter (ED), with minor additive effects (Figures 4C and 4D). One was associated with *GRMZM2G361210*, a marker of meristem branching related meta-clusters 9, 10, and 11 (Figure 4C), that encodes a C2H2-type zinc finger transcription factor related to *RAMOSA1* (*RA1*), a major player in maize domestication that controls branching and grain yield traits (Sigmon and Vollbrecht, 2010). A second significant SNP for ear diameter (ED) was associated with *ZmTMO5*, *GRMZM2G043854* (Figure 4D), a xylem meta-cluster marker (Figure 1J), whose Arabidopsis homolog controls periclinal cell divisions during vascular development (De Rybel et al., 2016).

We also conducted lambda analysis (Parvathaneni et al., 2020) to ask if scRNA-seq marker SNPs were more significantly associated with yield traits compared to a random subset of maize genes (Table S3; see STAR Methods). Using SNPs in or within 2kb of genes (2kb partition), we indeed found that ear diameter (ED) was significantly associated (Figures 4A and 4E; Table S3), suggesting that scRNA-seq marker genes preferentially control this trait (Figure 4E). Given that natural variation in distal regulatory elements also controls maize domestication and yield traits (Liu et al., 2020), we also considered SNPs within 200kb of gene coding regions (200kb partition), and found again that ear diameter (ED) was significantly associated (Figure 4F), as was an additional yield related trait, seed set length (SSL) (Figures 4A and 4G).

Although lambda analysis detects significance for a target set of markers, it does not quantify the level of trait variability that is explained by those markers. Therefore, we next estimated narrow-sense heritability (h^2) for scRNA-seq markers, compared to a distribution of h^2 estimates from random subsets of markers (Table S3; see STAR Methods). We found that SNP heritability for scRNA-seq markers was consistently greater than the 95th percentile of h^2

estimates from random markers for ear diameter (ED), cob weight (CW), and seed set length (SSL) traits (Figures 4A, 4H-4I), validating our targeted GWAS and lambda analysis. Similar findings were obtained for ear length (EL) in the 2kb partition (Figure 4H), as well as for ear row number (ERN) and ear rank number (EKN) in the 200kb partition (Figure 4I).

To ask if scRNA-seq data was required for these insights, we also calculated a list of “whole ear” specific genes from bulk tissue RNA-seq data (Table S3) (Walley et al., 2016), and found a low overlap with our scRNA-seq markers (4%, Table S3), indicating that the bulk tissue RNA-seq lacked cell or developmental domain specific information, as expected. We performed GWAS analysis using the whole ear specific genes, and they were also enriched for association with ear morphology traits, however in most cases for different traits and genes (Table S3). For instance, this analysis identified *ZmYAB9*, that was also identified using scRNA-seq markers (Figure 4B). However, the ear diameter (ED) trait was associated with different genes in scRNA-seq (Figures 4C-4D) and bulk tissue RNA-seq (Table S3) datasets. Furthermore, in heritability (h^2) analysis at a 200kb partition, no traits were associated with whole ear specific genes (Table S3), whereas four of them were significantly associated with scRNA-seq marker genes (Figure 4I). In summary, scRNA-seq markers revealed associations with multiple ear traits that were not found using bulk RNA-seq data, suggesting a unique application of this approach in identifying candidates to improve crop yield traits.

DISCUSSION

Development requires programmed cell- or domain-specific expression of regulatory and effector genes that together orchestrate stereotypical patterns of morphogenesis. The maize ear has a complex morphology with multiple indeterminate meristem and determinate organ types, and optimization of morphology is important for maize yield. To identify spatial regulators of ear development, we performed scRNA-seq of ~12,500 single cells from developing ears, and predicted 12 meta-clusters that were identified using known markers. As expected, many meta-clusters were meristem-associated, including for discrete domains that control branching, and we also found distinct vasculature meta-clusters, including xylem, phloem, and bundle sheath. We identified meta-clusters from determinate lateral organs and ground tissues, and in several cases,

sub-clusters could be identified. Our method was sensitive enough to detect most maize genes, though we failed to detect the expression of *CLV3* and *WUS* orthologs, possibly due to their low expression, or to a relatively low representation of central zone and organizing center cells in our experiments using whole developing ears. Another possibility is that these cell types were not recovered with our current protocol, and further improvements or profiling of cells from more finely dissected meristem tissues may address this issue. It is also intriguing that some xylem markers were expressed in the tips of meristems. In Arabidopsis, *class III HD-ZIP* and *KANADI* genes, which are expressed in xylem and phloem, respectively, are expressed in complementary patterns in adaxial and abaxial sides of lateral organs to specify their polarity (Emery et al., 2003). This polarity is pre-patterned by their corresponding central and peripheral expression in the shoot meristem (Caggiano et al., 2017). Thus, our finding of maize ear xylem markers expressed in the meristem tip suggests that additional vascular genes specify a pre-pattern in meristems.

We validated our scRNA-seq results by mRNA *in situ* hybridization and by comparing to a FACS RNA-seq dataset, thus constructing a robust single-cell transcriptome atlas of a developing inflorescence. We also provided three applications showing how this atlas can enhance functional studies. First, we highlight how the cellular resolution of scRNA-seq data can accurately predict redundancy, a major obstacle in genetic analyses. We could predict redundancy in a family of maize TPPs that control inflorescence branching, and identified a family of redundant maize *VOZ* genes that produced a delayed floral transition phenotype through multiplex CRISPR/Cas9 mutagenesis. Flowering time is a major target of maize breeding (Liu et al., 2020), thus our findings provide candidates to fine tune flowering for crop improvement. In some cases, redundant paralogs may show only partial co-expression, as observed for SQUAMOSA PROMOTER BINDING (SBP)-box transcription factors UNBRANCHED2 (UB2) and UB3 which control initiation of lateral organ (Chuck et al., 2014, Du et al., 2020), or *AINTEGUMENTA* (*ANT*) and *ANT-LIKE6*, whose expression overlaps partially to redundantly regulate floral organ patterning and growth (Krizek, 2009). Therefore, care should be taken in use of scRNA-seq data in discerning such partially overlapping expression patterns.

In a second example, we hypothesized that the resolution of scRNA-seq could be combined with ChIP-seq to predict directly modulated targets of TFs. We created ChIP-seq datasets for two TFs that are likely to act redundantly, precluding the use of single mutants to find their modulated targets. We also integrated scRNA-seq with FACS ATAC-seq to provide evidence of spatially-regulated accessible chromatin. The limited availability of maize reporter lines for FACS may be overcome in the future by application of single-cell ATAC-seq (Rich-Griffin et al., 2020). Lastly, we hypothesized that scRNA-seq marker genes with spatially restricted expression in developing ears are enriched for regulators of ear morphology traits important for crop yields. Indeed, the scRNA-seq marker genes were significantly associated with ear morphology trait SNPs in a GWAS panel. These marker candidates could be selected in breeding programs or genetically modified to test their effects on yield.

In summary, scRNA-seq allowed valuable insights into maize ear development. The atlas can inform developmental genetics studies and breeding, and the methods we developed can be applied to studies of other complex shoot systems. As more plant scRNA-seq datasets are generated, a cross-species (eg. between maize and *Arabidopsis*) or cross-tissue (eg. between shoot and root) comparative analysis at single-cell resolution will inform how gene signatures were selected during evolution to shape the diverse morphologies that are critical to reproductive success and agricultural production.

ACKNOWLEDGMENTS

We thank Drs. Kenneth Birnbaum, Carlos Ortiz Ramírez, and Bruno Guillotin for advice on scRNA-seq and FACS RNA-seq experiments, and Drs. Bert De Rybel and Joyce Chery for discussions on vascular biology. We thank support from Dr. Jon Preall and the CSHL single-cell core facility, flow cytometry facility, microscopy facility, NGS facility, and farm management team. We thank Cassidy Danyko for assisting in FACS RNA-seq library preparation, Richelle Chen for assisting in bioinformatic analyses, and Michael Okoro for assisting in field work and sample preparation. We acknowledge funding support from NSF (IOS-1833182, IOS-1445025, IOS-1027445, IOS-1934388, IOS-1755141). R.J.S. acknowledges support from NSF (IOS-1856627). J.G. acknowledges support from R01 LM012736 and R01 MH113005. M.C. acknowledges support from K99 MH120050. B.H. acknowledges support from The Robertson Research Fund, CSHL School for Biological Sciences. We thank Drs. Munenori Kitagawa and Penelope Lindsay for their feedback on the manuscript.

AUTHOR CONTRIBUTIONS

X.X. performed experimental procedures and data analysis, except for those listed below, and wrote the draft of the manuscript. M.C., B.H., and N.F. performed scRNA-seq analysis. B.R. performed GWAS analysis. F.L. performed ChIP-seq analysis and assisted ATAC-seq analysis. L.L. performed FACS RNA-seq analysis and assisted in mRNA *in situ*. E.D.A. and S.N.C. generated *Zmvoz* CRISPR mutants. Z.L. assisted to prepare FACS ATAC-seq library and performed analysis. L.W. assisted both FACS RNA-seq and ChIP-seq analyses. X.W. assisted ChIP-seq analysis. J.D. assisted to prepare FACS RNA-seq libraries. A.L. generated ZmHDZIV6-YFP transgenic lines. D.J., J.G., D.W., A.E.L., R.S., T.R.G., A.S., and B.Y. supervised the research. D.J. co-wrote the manuscript, and all authors edited.

DECLARATION OF INTERESTS

The authors declare no competing interests.

MAIN FIGURE AND TITLES AND LEGENDS

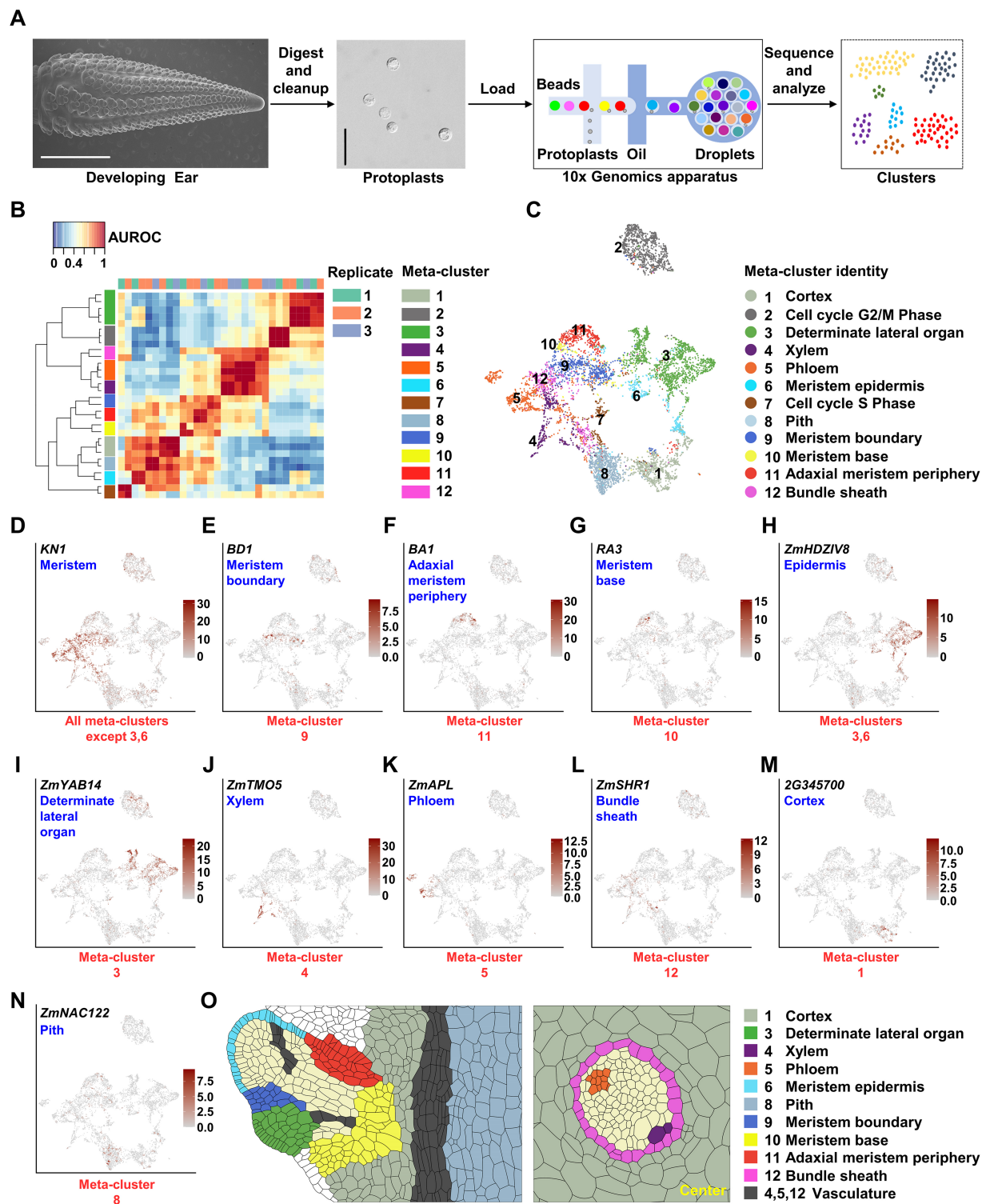


Figure 1: Isolation of Maize Ear Protoplasts to Construct a Single-Cell Transcriptomic Atlas.

(A) Experimental design, the first panel shows a scanning electron microscope image of a 5-10mm developing ear (scale bar = 2mm), second panel image of ear protoplasts, scale bar = 50μm. (B) MetaNeighbor identifies 12 reproducible meta-clusters (left color blocks) across three biological replicates (top color blocks) of single-cell RNA-seq datasets. (C) 12 meta-clusters displayed by an integrated Uniform Manifold Approximation and Projection (UMAP) plot in two dimensions, with each dot representing a cell. (D-N) UMAP plots of marker genes predicting the identities of meta-clusters, with color scale indicating normalized expression level. (D) *KNI*, meristem, all meta-clusters except 3 and 6; (E) *BDI*, meristem boundary, meta-cluster 9; (F) *BAL*, adaxial meristem periphery, meta-cluster 11; (G) *RA3*, meristem base, meta-cluster 10; (H) *ZmHDZIV8*, epidermis, meta-clusters 6 and part of 3; (I) *ZmYAB14*, determinate lateral organ, meta-cluster 3; (J) *ZmTMO5*, xylem, meta-cluster 4; (K) *ZmAPL*, phloem, meta-cluster 5; (L) *ZmSHR1*, bundle sheath, meta-cluster 12; (M) *GRMZM2G345700* (*2G345700*), cortex, meta-cluster 1; (N) *ZmNAC122*, pith, meta-cluster 8. (O) Sketches of longitudinal section of a spikelet meristem (left panel) and transverse section of vascular bundle (right panel) showing cell/domain identities in scRNA-seq meta-clusters.

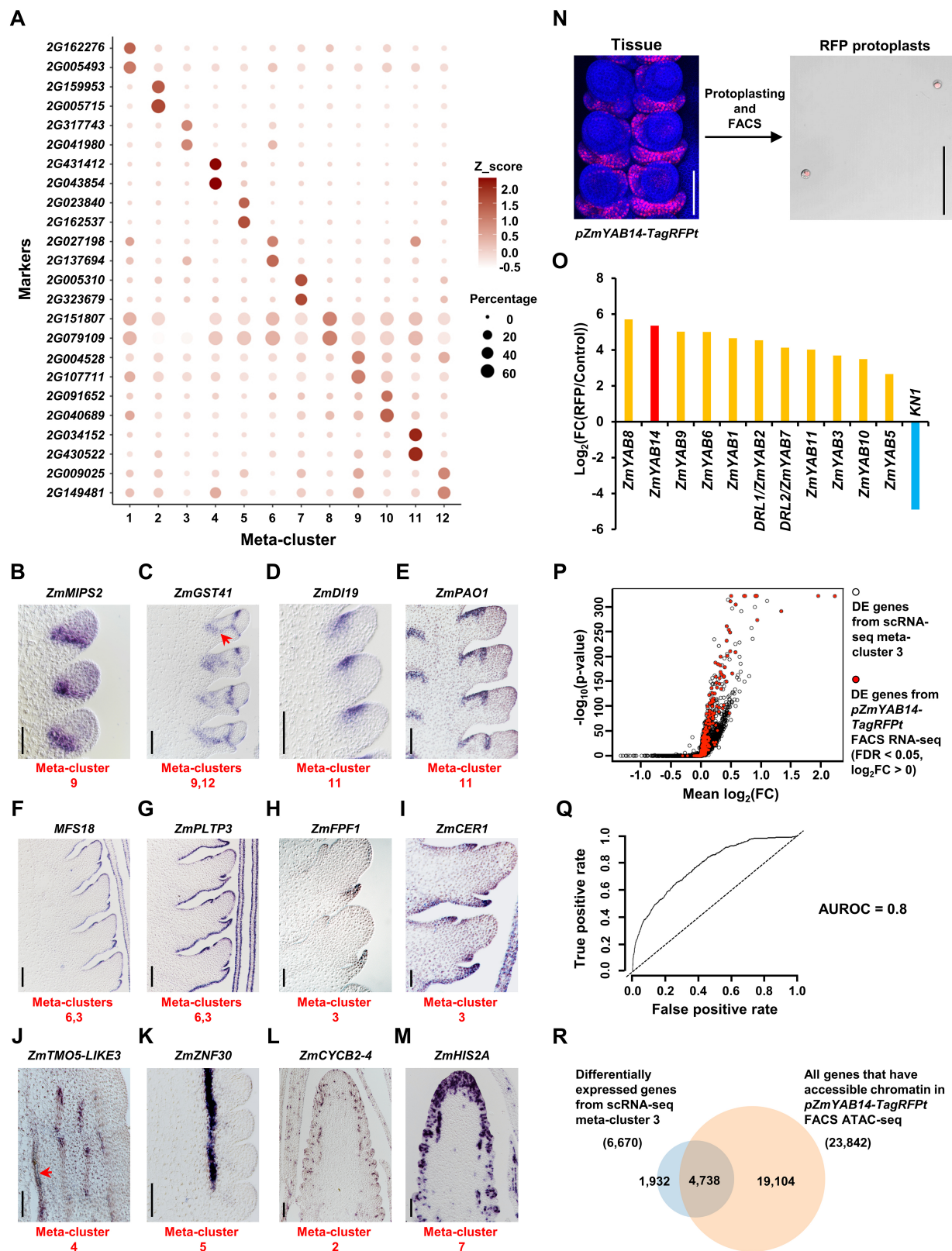


Figure 2: Validation of scRNA-seq by mRNA *in situ* and FACS RNA-seq.

(A) The top two marker genes of each meta-cluster are shown in dot plots with circle size indicating the percentage of cells expressing the marker and color representing Z_{scored} expression value. (B-M) mRNA *in situ* of meta-cluster marker genes validates the predicted identities: (B) *ZmMIPS2*, meristem boundary; (C) *ZmGST41*, meristem boundary (meta-cluster 9) and bundle sheath (meta-cluster 12, red arrow); (D-E) *ZmDII9* (D) and *ZmPAOI* (E), adaxial meristem periphery; (F-G) *MFS18* (F) and *ZmPLTP3* (G), meristem epidermis (meta-cluster 6) and determinate lateral organ (meta-cluster 3) epidermis; (H-I) *ZmFPF1* (H) and *ZmCER1* (I), determinate lateral organ; (J) *ZmTMO5-LIKE3*, xylem (red arrow indicates xylem vessels); (K) *ZmZNF30*, phloem. (L) *ZmCYCB2-4*, cell cycle G2/M phase; (M) *ZmHIS2A*, cell cycle S phase. Scale bar = 100 μ m. (N) Collection of RFP protoplasts from *pZmYAB14-TagRFPt* reporter line using FACS. Scale bar = 100 μ m. Three biological replicates were collected for FACS RNA-seq. One biological replicate was collected for FACS ATAC-seq. (O) $\text{Log}_2(\text{fold change(FC)})$ of determinate lateral organ domain enriched markers, *ZmYAB* genes, and depleted marker, *KN1*, between RFP and total control protoplasts (Control) in FACS RNA-seq. (P) Volcano plot with 1-sided test positions the hits of enriched markers from *pZmYAB14-TagRFPt* FACS RNA-seq (red dots) on the ranked list of scRNA-seq differentially expressed (DE) genes from meta-cluster 3 (black circles). X-axis indicates the mean $\text{log}_2(\text{FC})$ of DE genes between meta-cluster 3 and all other meta-clusters. Y-axis indicates corresponding $-\text{log}_{10}(\text{p-value})$. (Q) *pZmYAB14-TagRFPt* FACS RNA-seq and scRNA-seq meta-cluster 3 have concordant differential gene expression patterns with Area Under the Receiver Operating Characteristics (AUROC) score = 0.8 (indicated by curved line; dashed line indicates the null (AUROC score = 0.5)). Axes indicates the true and false positive rate, the proportion of *pZmYAB14-TagRFPt* FACS RNA-seq enriched markers that do or do not match to scRNA-seq meta-cluster 3 enriched markers, respectively. (R) meta-cluster 3 DE genes are enriched in open chromatin in *pZmYAB14-TagRFPt* FACS sorted cells, see text for details.

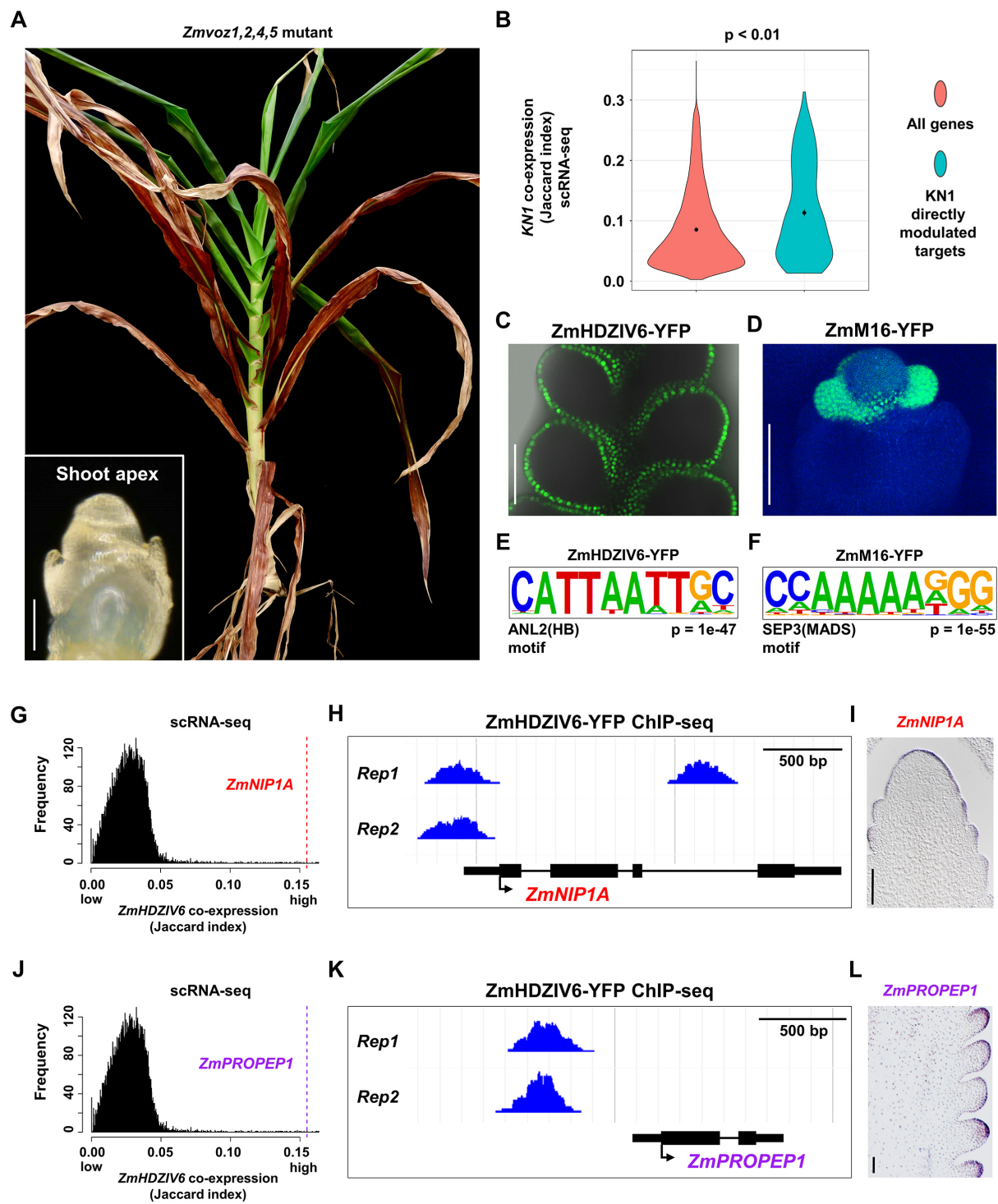


Figure 3: scRNA-seq Can Predict Genetic Redundancy and Aid in Predicting Transcriptional Regulatory Networks.

(A) Maize plant with CRISPR/cas9 knockout of four *ZmVOZ* paralogs fails to transition to flowering, as shown by two month old shoot apex (left bottom panel, scale bar = 100 μ m), and a 6 month old plant that lacks ears or tassel. (B) Directly modulated transcriptional targets of *KN1* are significantly co-expressed with *KN1* at the single-cell level; all maize genes are used as control ($p < 0.01$, one-way ANOVA with Tukey's HSD). (C-D) Expression of TF translational fusion lines, *ZmHDZIV6*-YFP (C, merge of YFP channel and bright field) and *ZmM16*-YFP (D, merge of YFP and DAPI channels), used for two biological replicates of ChIP-seq. Scale bar = 100 μ m. (E-F) Expected motifs are significantly over-represented in bound peaks of *ZmHDZIV6* (E) or *ZmM16* (F). $p = 1e-47$ (E) and $p = 1e-55$ (F). (G, J) *ZmHDZIV6* candidate modulated targets, *ZmNIP1A* (G) and *ZmPROPEP1* (J) are highly co-expressed with *ZmHDZIV6* in scRNA-seq (Jaccard index = 0.155 for both targets). (H, K) *ZmHDZIV6* bound peaks in *ZmNIP1A* (H) and *ZmPROPEP1* (K). Scale bar = 500 bp. (I, L) *ZmNIP1A* (I) and *ZmPROPEP1* (L) are specifically expressed in the epidermis, by mRNA *in situ*. Scale bar = 100 μ m.

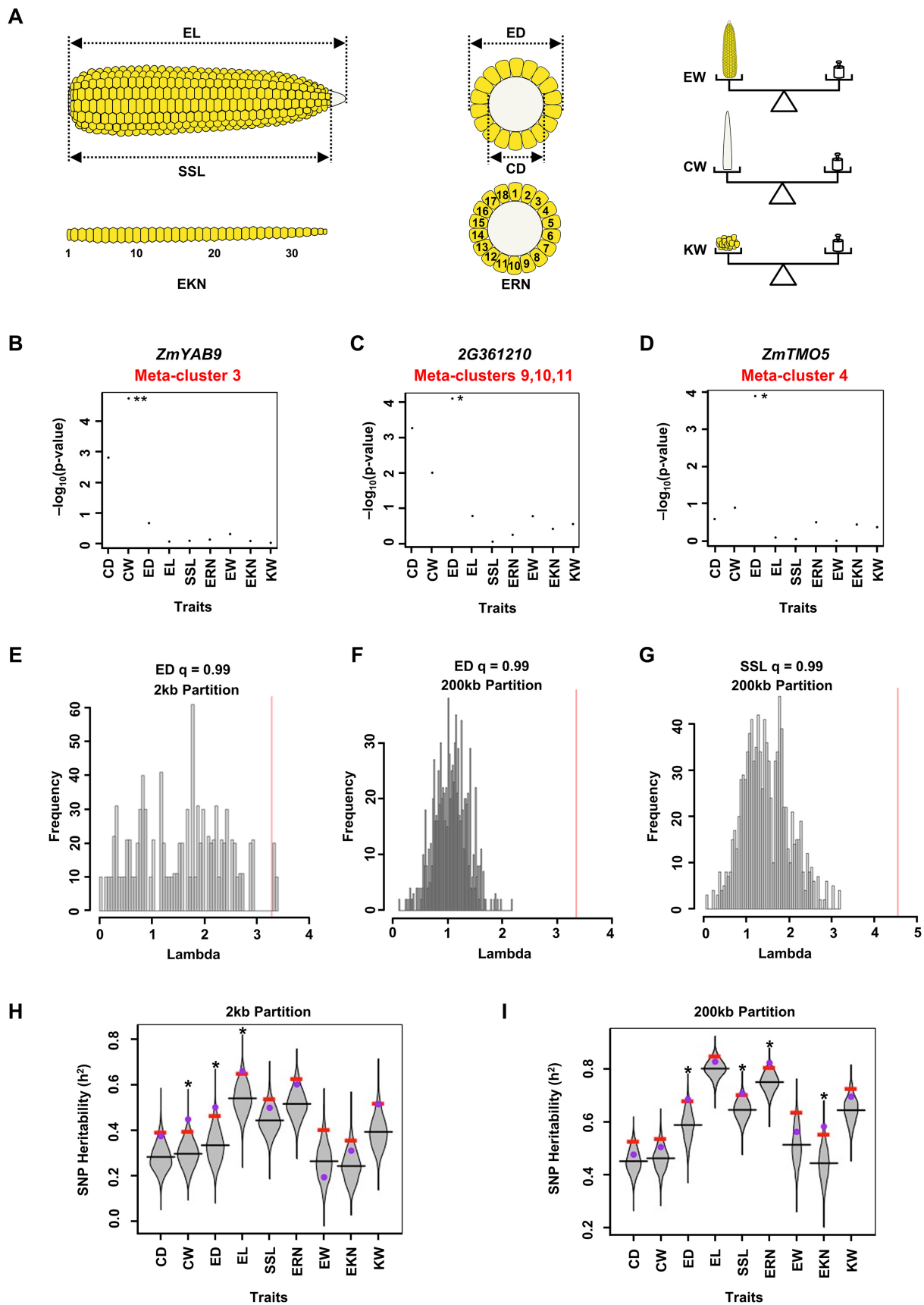


Figure 4: scRNA-seq Marker Genes Are Associated with Maize Ear Traits.

(A) Diagrams of 9 different ear traits measured for GWAS analysis: Ear Length (EL), Seed Set Length (SSL), Ear Rank Number (EKN), Ear Diameter (ED), Cob Diameter (CD), Ear Row Number (ERN), Ear Weight (EW), Cob Weight (CW), Kernel Weight (20 Seeds) (KW). (B-D) Targeted GWAS using SNPs in or within 2kb of genes reveals that scRNA-seq marker gene *ZmYAB9*, has significant SNPs for ear weight (B), and two marker genes *GRMZM2G361210* (*2G361210*) (C) and *ZmTMO5* (D) have significant SNPs for ear diameter. ** FDR threshold of 0.05, * FDR threshold of 0.1. Y-axis indicates the $-\log_{10}(\text{p-value})$ (Table S3). (E-G) Lambda values of scRNA-seq marker genes (red lines) are greater than two standard deviations from mean lambda values of 1,000 random gene sets (histogram distributions) for ear diameter (2kb partition, E, 200kb partition, F), and for seed set length trait (200kb partition, G). Lambda values are reported in Table S3. (H-I) Distributions of SNP heritability (h^2) using 2kb (H) or 200kb (I) partitions; h^2 values for scRNA-seq marker genes (purple dots) for the given traits (*) are greater than the top 5% permuted h^2 values (red bars) using 1000 random subsets of maize genes (grey violin plots). h^2 values are reported in Table S3.

STAR★METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Jackson (jacksond@cshl.edu).

Materials Availability

Requests for materials should be directed to Lead Contact, David Jackson (jacksond@cshl.edu). Requests for transgenic plant materials will require a Materials Transfer Agreement (MTA).

Data and Code Availability

The accession numbers for the raw scRNA-seq data reported in this paper are NCBI's Sequence Read Archive (SRA) BioProjects: PRJNA646989, PRJNA646996, and PRJNA647001. The accession numbers for the raw protoplasting-response bulk RNA-seq data reported in this paper is NCBI's SRA BioProject: PRJNA647196. The accession numbers for the raw pZmYAB14-TagRFPt FACS RNA-seq data reported in this paper is NCBI's SRA BioProject: PRJNA647195. The accession numbers for the raw pZmYAB14-TagRFPt FACS ATAC-seq data reported in this paper is NCBI's SRA BioProject: PRJNA647197. The accession numbers for the raw ZmHDZIV6-YFP ChIP-seq data reported in this paper is NCBI's SRA BioProject: PRJNA647198. The accession numbers for the raw ZmM16-YFP ChIP-seq data reported in this paper is NCBI's SRA BioProject: PRJNA647200. SRA BioProject IDs were also listed in Key resources table. This study used codes from published software described in QUANTIFICATION AND STATISTICAL ANALYSIS.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All analyses were performed with *Zea mays* (Maize). Maize plants were grown in the summer field (June – October) of Uplands Farm Agricultural Station at Cold Spring Harbor, New York or in the greenhouse with 16hr daytime, 26-28°C, and 8hr night, 22-24°C (Wu et al., 2020). Reference B73 inbred plants were used for single-cell experiments. The *pZmYAB14-*

TagRFPt reporter line (Je et al., 2016) and ZmM16-YFP translational fusion line (Bartlett et al., 2015) were obtained from previous studies. The ZmHDZIV6-YFP translational fusion line was constructed using *ZmHDZIV6* native promoter and coding sequence as previously described in the pTF101 Gateway-compatible vector (Je et al., 2016), primer sequences were listed in Table S4. The *pZmYAB14-TagRFPt* reporter line and ZmHDZIV6-YFP translational fusion line were introgressed into a proliferative cauliflower-like double mutant line, *bdl;Tu*, to generate a large amount of ear meristem tissue for FACS RNA-seq and ChIP-seq respectively. Primer sequences for genotyping *bdl* are listed in Table S4. *Tu* genotyping was performed as previously described (Han et al., 2012).

CRISPR/Cas9 was used to knockout *ZmVOZs* genes following *Agrobacterium*-mediated transformation of Hi-II embryos (Je et al., 2016). Guide RNAs (sgRNAs) were designed based on B73 V3 reference genome, one pair targeting *ZmVOZ1* and *ZmVOZ2*, and a second pair targeting *ZmVOZ4* and *ZmVOZ5*, Table S4. The sgRNAs were introduced by Gateway Recombination into pGW-Cas9 vector (Addgene plasmid # 50661; RRID: Addgene_50661) (Wang et al., 2014) and transferred to *Agrobacterium* (EHA101) for maize transformation. 29 plants from 8 transformation events were obtained and analyzed by PCR amplification and Sanger sequencing to identify mutations in the targeted regions. The Cas9 transgene was segregated away by crossing with B73 to recover stable mutant alleles. Primer sequences and PCR assays for genotyping were listed in Table S4.

METHOD DETAILS

Protoplast preparation and 10x Genomics library construction and sequencing

For the three biological replicates of wild type (B73) whole ear samples, protoplasts were prepared as previously described but without L-cysteine pretreatment (Ortiz Ramírez et al., 2018). 5-10mm developing ears were dissected into protoplast washing buffer and diced with a razor blade to 0.5-1mm pieces, then washed three times with protoplast washing buffer before adding enzyme solution. Tissues were protoplasted for ~ 45min at room temperature with gentle shaking. The mixture was filtered through a 30µm cell strainer (pluriStrainer, 43-50030-50) then collected by centrifugation at 500g for 3 mins at 4°C. The supernatant was gently removed without disturbing the protoplast pellet, which was washed by gentle resuspension in protoplast

washing buffer. Protoplasts were then filtered as before and purified by FACS sorting (FACS Aria II SORP with 100-micron setup, purity precision, yield mask at 32, purity mask at 32, plates voltage at 2,500, voltage centering at 20, sheath pressure at 20, and target gap at 12). Protoplasts were sorted into 1×PBS with 0.1% BSA and 0.4M mannitol, and pelleted at 400g for 2 mins at 4°C. The supernatant was carefully removed, leaving 20-40µl to gently resuspend the pellet. The protoplasts were stained with trypan blue (Thermo Fisher Scientific, 15250061) to check concentration and viability with a hemocytometer under a light microscope, and good quality protoplasts with viability $\geq 70\%$ were immediately loaded into the 10x Genomics Chromium System using V2 chemistry kits. scRNA-seq libraries were sequenced by Illumina short reads with $\sim 400\text{M}$ paired end reads per library (read1 = 28bp, read2 = 56bp) (Table S1). Raw sequencing data were deposited in NCBI's Sequence Read Archive (SRA). SRA IDs were listed in Table S1.

Anatomy and confocal microscopy

For anatomy, wild type (B73) developing ears from fresh plants were hand dissected and fixed in 4% paraformaldehyde (Electron Microscopy Sciences, 15714-s) (Jackson et al., 1994). The fixed ear tissue was dehydrated through a graded alcohol series (50%, 70%, 85%, 95%, and 100%) and a histoclear series, then embedded in paraplast (McCormick Scientific, 39503002) (Jackson et al., 1994). 5µm sections were cut using a Leica microtome, then mounted on ProbeOn Plus™ Slides (Fisher Scientific, 22-230-900) and rehydrated and stained with Calcofluor white stain (Sigma-Aldrich, 18909). Images were taken on a ZEISS LSM 710 confocal microscope using DAPI channel.

For vascular bundle anatomy, wild type (B73) developing ears were fixed overnight at 4°C in 0.1 M cacodylate buffer pH 7.4 (Electron Microscopy Sciences, 11652) with 4% paraformaldehyde (Electron Microscopy Sciences, 15714-s) and 1% glutaraldehyde (Electron Microscopy Sciences, 16537-16), washed three times with 0.1M cacodylate buffer and then dehydrated in a graded ethanol series (50%, 60%, 70%, 80%, 90%, 95%, two times 100%). The dehydrated ear samples were transferred to LR white resin / ethanol in 1:3, 1:1, and 3:1 ratios, and twice overnight in 100% LR white resin (Electron Microscopy Sciences, 905072). Tissues in LR white resin were then polymerized in a 60 °C oven for 48 h, and 0.5-1µm thick sections were

cut using 45 Diamond DiATOME Histo Knife on a Reichert Ultracut E microtome. Sections were collected and stained by toluidine blue (Sigma-Aldrich, T3260). Images were taken by Nikon DS-Ri2 microscope.

pZmYAB14-TagRFPT, *ZmM16-YFP*, and *ZmHDZIV6-YFP* lines were imaged using ZEISS LSM 710 or 780 confocal microscopes. The DAPI channel was used for capturing autofluorescence (blue color) in Figures 2N and 3D.

mRNA *in situ* hybridization

mRNA *in situ* were conducted as previously described (Jackson et al., 1994). Briefly, wild type (B73) developing ears were freshly collected, fixed in 4% paraformaldehyde (Electron Microscopy Sciences, 15714-s) (Jackson et al., 1994). The fixed ear tissue was dehydrated through a graded alcohol series (50%, 70%, 85%, 95%, and 100%) and a histoclear series, then embedded in paraplast (McCormick Scientific, 39503002) (Jackson et al., 1994). 10µm sections were cut using a Leica microtome, then mounted on ProbeOn Plus™ Slides (Fisher Scientific, 22-230-900). To prepare probes for marker genes, we added T7 promoter sequences GAGTAATACGACTCACTATAGGGAGA into reverse primers used to amplify gene-specific PCR products from cDNA templates. Then probes were synthesized by *in vitro* transcription using T7 RNA Polymerase (Sigma-Aldrich, 10881775001). Primer sequences for all genes were listed in Table S4. Probes were then applied on tissue sections and incubated at 50°C overnight. To detect the hybridization signal, we applied freshly dissolved NBT/BCIP Ready-to-Use Tablets (Roche, 11697471001) in the alkaline phosphatase reaction solution. Images were taken using a Nikon DS-Ri2 DIC microscope.

FACS and bulk RNA-seq, ATAC-seq library preparation

For three biological replicates of bulk RNA-seq to identify protoplasting-responsive genes, wild type B73 background ear tissue was used. RNA was isolated from protoplasted and equivalent non-protoplasted tissue to compare side by side. For three biological replicates of *pZmYAB14-TagRFPT* FACS RNA-seq, ear tissue of *pZmYAB14-TagRFPT/bd1;Tu* and *bd1;Tu* negative control plants were collected and digested as described earlier. Protoplasts were gently washed, filtered, and resuspended as before. *bd1;Tu* negative control protoplasts were first

loaded into FACSaria II SORP to set up the gate for identifying autofluorescence signals. *pZmYAB14-TagRFPt/bd1;Tu* protoplasts were then loaded using the same settings. RFP cells were collected based on specific signals from the mStrawberry channel, and examined under a ZEISS LSM 710 confocal microscope. RNA for RFP positive protoplast samples and control samples (total protoplasts without sorting) was extracted using Arcturus PicoPure RNA Isolation Kit (Thermo Fisher Scientific, KIT0204). RNA was examined by a RNA Bioanalyzer kit (Agilent, 5067-1513). RNA-seq libraries were built using SMART-Seq™ v4 Ultra™ Low Input RNA Kit (Takara Bio USA, Inc., 634890) and Nextera XT DNA Library Prep Kit (Illumina, FC-131-1024). Library quality and size was examined by a DNA High Sensitivity Bioanalyzer chip (Agilent, 5067-4626), and quantified using the KAPA Library Quantification Kit (Roche, KK4824) before Illumina sequencing. Raw sequencing data were deposited into NCBI's Sequence Read Archive (SRA). SRA IDs were listed in Tables S1-S2. For one biological replicate of *pZmYAB14-TagRFPt* FACS ATAC-seq, RFP protoplasts were collected as described above, and nuclei isolated for library construction and sequencing as previously (Lu et al., 2017). Raw sequencing data were deposited in NCBI's Sequence Read Archive (SRA). SRA ID was listed in Table S2.

ChIP-seq library preparation

ChIP experiments were conducted as previously described (Pautler et al., 2015) with some modifications. Briefly, two biological replicates of freshly harvested ear tissues of *ZmHDZIV6-YFP/bd1;Tu* and *ZmM16-YFP* were immediately cross-linked in buffer containing 1% formaldehyde, 10mM HEPES-NaOH PH7.4, 0.4 M sucrose, 1 mM EDTA, and 1 mM PMSF, for 20min under vacuum. Glycine was then added to a concentration of 0.1 M to for another 5 min under vacuum. Nuclei extraction was conducted using CellLytic™ PN Isolation/Extraction Kit (Sigma-Aldrich, CELLYTPN1). For immunoprecipitation, we used high-affinity GFP-Trap magnetic agarose (ChromoTek, gtma-20; RRID: AB_2631358). For building ChIP-seq libraries, we used NEXTflex ChIP-seq Kit (PerkinElmer Applied Genomics, NOVA-5143-02) with AMPure XP beads (Beckman Coulter, A63880). ChIP-seq libraries were quantified by KAPA Library Quantification Kits (Roche, KK4824) before Illumina sequencing. Raw sequencing data were deposited in NCBI's Sequence Read Archive (SRA). SRA IDs were listed in Table S3.

QUANTIFICATION AND STATISTICAL ANALYSIS

scRNA-seq analysis, clustering, and selection of marker genes

Sequencing reads of three biological replicates of wild type (B73) whole ear scRNA-seq samples were aligned to the maize v3 reference genome using STARsolo v2.7.0.f (Dobin et al., 2013). We updated the v3.31 GTF annotation file by adding four maize *CLAVATA3/EMBRYO SURROUNDING REGION-RELATED* (*ZmCLE*) genes, including *ZmCLE7* (GRMZM2G372364), *ZmCLE14* (AC191109.3_FG001), *ZmCLE25* (GRMZM2G525788), and *ZmCLEug-2* (GRMZM2G054501) (Table S4; MaizeGDB), before using it to build a STAR genome index with default parameters (Dobin et al., 2013). All downstream processing was performed in R (R Core Team, 2013), with each dataset analyzed separately. In brief, we removed droplets that lacked a protoplast using EmptyDrops with a minimum threshold of 800 UMIs (Lun et al., 2019), and removed probable doublets using DoubletFinder (McGinnis et al., 2019). Expression data was log2 normalized using scater (McCarthy et al., 2017). We identified highly variable genes using the trendVar function in scran (Lun et al., 2016), selecting genes at FDR<0.05, and we used the rsvd package (Erichson et al., 2019) to calculate approximate principal components for all cells after subsetting to highly variable genes. We then generated a nearest neighbor graph for cells with the ccd package (Marchette, 2015) using the Euclidean distance across the top 20 PCs, with k=100. To find clusters, we used the InfoMap algorithm implemented in the igraph R package (Csardi and Nepusz, 2006), resampling 100 times (Rosvall and Bergstrom, 2008). MetaNeighbor analysis was performed as previously described (Crow et al., 2018). We used the umap package to generate embeddings for visualization (Konopka, 2020). Within each UMAP, every dot represents a cell, and the color scale indicates the normalized expression level by adding a constant 75% of each cell's expression value to nearest neighbors for clear visualization. Differential expression statistics were calculated with an AUROC test on log counts using the auc multifunc function from EGAD (Ballouz et al., 2017). Differential expression genes with AUROC scores of ≥ 0.7 in at least one replicate were considered as meta-cluster marker genes (Table S1).

FACS and bulk RNA-seq, ATAC-seq analysis

Three biological replicates of FACS and bulk RNA-seq analyses were performed as previously (Wang et al., 2020) with some modifications. Bulk RNA-seq datasets for extracting ear tissue specific genes were downloaded from a previous study (Walley et al., 2016). Raw sequencing reads were firstly trimmed with Trimmomatic (Bolger et al., 2014), and then mapped with STAR (Dobin et al., 2013) using the same updated maize V3 reference as scRNA-seq analysis. edgeR (Robinson et al., 2010) was used to perform differential expression analysis. We calculated similarity between scRNA-seq and FACS RNA-seq data using the `auroc_analytic` function in EGAD (Ballouz et al., 2017), with ranked scRNA-seq meta-cluster 3 p-values as the “scores” and FACS RNA-seq DE genes ($\log_2 \text{FC} > 0$ and $\text{FDR} < 0.05$, Table S2) as the “labels”. Whole ear ATAC-seq datasets were downloaded from a previous study (Ricci et al., 2019). One biological replicate of FACS ATAC-seq analysis was performed as previously (Ricci et al., 2019), using the same reference as scRNA-seq analysis.

ChIP-seq analysis

Two biological replicates of ChIP-seq reads for both ZmHDZIV6-YFP and ZmM16-YFP datasets were trimmed using sickle (<https://github.com/najoshi/sickle>). Duplicated reads were further removed by elprep (Herzeel et al., 2015), before aligning to the same updated maize V3 reference used for scRNA-seq analysis with BWA-MEM (Li and Durbin, 2009). Alignment reads were filtered for those above a MAPping Quality (MAPQ) threshold above 40. Peak calling, peak annotation, and motif enrichment were performed with HOMER (Heinz et al., 2010), with peak calling parameters as the following: -F = 8: Fold enrichment threshold of IP tag count over input tag count, -L = 2: Fold enrichment of a putative peak’s tag count over surrounding (local) genomic region, and -LP = 0.0001: p-value cutoff for local fold enrichment to a peak to be considered. High confidence peaks between two biological replicates were determined by finding midpoints of peaks positioned within 300bp of each other (Pautler et al., 2015).

scRNA-seq co-expression analysis

89 maize bulk tissue RNA-seq datasets were used for calculating co-expression values as previously described (Lee et al., 2020). Briefly, co-expression networks were constructed using Spearman’s correlation. Three gene pairs including *RA3-ZmTPP4*, *RA3-ZmTPP12*, and

ZmTPP4-ZmTPP12 were then extracted from the datasets. The aggregated co-expression value of each gene pair was calculated and reported in Figure S4. For scRNA-seq co-expression to predict *RA3-ZmTPPs* redundancy, *ZmVOZ* co-expression, and TF ChIP-Seq directly modulated targets, the average Jaccard index was calculated for B73 whole ear datasets, using genes expressed in >1% of cells. Genes with at least 1 UMI were assigned a value of 1, and all others were assigned a value of 0. The Jaccard index was calculated for each ear dataset separately and then averaged. To calculate significance (Figure S4) we took a non-parametric approach with the null hypothesis that there is no replicable co-expression across batches, calculated by convolving the uniform distributions obtained for each batch after ranking. To compare overlaps in co-expression between two genes, we reranked and assessed similarity against the null using the same approach. FDRs were calculated by dividing cumulative nulls by cumulative empiricals.

Integration of GWAS with scRNA-seq or ear tissue bulk RNA-seq

To integrate GWAS with scRNA-seq, we firstly selected unique scRNA-seq marker genes from ear meristem, determinate lateral organ, and vasculature meta-clusters (3, 4, 5, 6, 9, 10, 11, 12) using two different cutoffs. The stringent cutoff was AUROC ≥ 0.7 across all replicates, which gave 68 scRNA-seq marker genes (Table S3). The less stringent cutoff was AUROC ≥ 0.7 in at least one replicate, which gave 241 scRNA-seq marker genes (Table S3). For targeted GWAS analysis: we used the best linear unbiased predictors (BLUPs) from nine ear phenotypes to perform a targeted GWAS using the unified mixed linear model. BLUPs estimation, GWAS model details, and genomic marker filtering procedures were described in (Rice et al., 2020). The SNPs in or within 2kb of scRNA-seq marker genes from lists of both stringent and less stringent cutoffs were used (Table S3).

For lambda analysis, the procedure has been previously described (Parvathaneni et al., 2020). Briefly, lambda was a ratio of the FDR adjusted p-values for a given set of markers when included with the full marker set compared to when considered on their own (equation below).

$$\text{Lambda} = \frac{qth_Percentile(-\log(\text{Reduced FDR Adjusted } p\text{-values}))}{qth_Percentile(-\log(\text{Genomewide FDR Adjusted } p\text{-values}))}$$

We looked at the significance of 99th percentile ($q = 99$) of FDR adjusted p-values subset of markers (SNPs in the region of genes of interest). We performed 1,000 replicates (random

subsets) and estimated the lambda distributions for each trait. Lambda values of 241 unique markers from scRNA-seq meta-clusters (3, 4, 5, 6, 9, 10, 11, 12) with less stringent cutoff (Table S3) were compared against random subsets lambda distributions to determine significance. Traits with $\lambda \geq \text{mean} \pm 2\text{SD}$ were considered to be biologically significant.

For SNP Heritability analysis, we estimated narrow-sense heritability (h^2) from the subsets of the scRNA-seq SNPs considered in the lambda analysis using the LDAK software v5.0 (Speed et al., 2012). Thus, the resulting estimate of h^2 provides an estimate of the additive genetic variance explained by genes of interest. To determine if the resulting heritability for a given trait was greater than chance, the heritability for 1000 permutations using a random subset of maize genes was estimated. For a given permutation, genes with at least one SNPs within the genic region were randomly selected. Enough genes were selected to ensure the total number in a permuted subset was ± 5 compared to the target set. A target set was declared significant for a given trait if its heritability was greater than the top 5% of permuted values.

To integrate GWAS with ear tissue bulk RNA-seq, we firstly extracted ear tissue specific genes from a previous study (Table S3) (Walley et al., 2016). Then we used these genes to perform same analyses, including targeted GWAS, lambda, and SNP Heritability as mentioned above for scRNA-seq (Table S3).

SUPPLEMENTAL EXCEL TABLE TITLES

Table S1: scRNA-seq and protoplasting-responsive genes datasets. Related to Figures 1 and S1-S3.

Table S2: scRNA-seq validation by FACS RNA & ATAC-seq. Related to Figures 2 and S3.

Table S3: scRNA-seq applications. Related to Figures 3, 4, and S4.

Table S4: Primers, sgRNAs sequences and *ZmCLE* genes' annotation. Related to all Figures.

REFERENCES

- Aerts, N., de Bruijn, S., van Mourik, H., Angenent, G.C., and van Dijk, A.D.J. (2018). Comparative analysis of binding patterns of MADS-domain proteins in *Arabidopsis thaliana*. *BMC plant biology* 18, 131-131.
- Ballouz, S., Weber, M., Pavlidis, P., and Gillis, J. (2017). EGAD: ultra-fast functional analysis of gene networks. *Bioinformatics* 33, 612-614.
- Bartlett, M.E., Williams, S.K., Taylor, Z., DeBlasio, S., Goldshmidt, A., Hall, D.H., Schmidt, R.J., Jackson, D.P., and Whipple, C.J. (2015). The Maize PI/GLO Ortholog Zmm16/sterile tassel silky ear1 Interacts with the Zygomorphy and Sex Determination Pathways in Flower Development. *Plant Cell* 27, 3081-3098.
- Bolduc, N., Yilmaz, A., Mejia-Guerra, M.K., Morohashi, K., O'Connor, D., Grotewold, E., and Hake, S. (2012). Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes & development* 26, 1685-1690.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
- Bommert, P., Nagasawa, N.S., and Jackson, D. (2013). Quantitative variation in maize kernel row number is controlled by the FASCIATED EAR2 locus. *Nature genetics* 45, 334-337.
- Caggiano, M.P., Yu, X., Bhatia, N., Larsson, A., Ram, H., Ohno, C.K., Sappl, P., Meyerowitz, E.M., Jönsson, H., and Heisler, M.G. (2017). Cell type boundaries organize plant development. *Elife* 6, e27421.
- Chang, Y.-M., Liu, W.-Y., Shih, A.C.-C., Shen, M.-N., Lu, C.-H., Lu, M.-Y.J., Yang, H.-W., Wang, T.-Y., Chen, S.C.C., Chen, S.M., *et al.* (2012). Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. *Plant physiology* 160, 165-177.
- Chen, D., Yan, W., Fu, L.-Y., and Kaufmann, K. (2018). Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana*. *Nature Communications* 9, 4534.
- Chen, H., and Xiong, L. (2010). myo-Inositol-1-phosphate synthase is required for polar auxin transport and organ development. *The Journal of biological chemistry* 285, 24238-24247.
- Chitwood, D.H., Guo, M., Nogueira, F.T.S., and Timmermans, M.C.P. (2007). Establishing leaf polarity: the role of small RNAs and positional signals in the shoot apex. *Development* 134, 813.

Chuck, G., Muszynski, M., Kellogg, E., Hake, S., and Schmidt, R.J. (2002). The control of spikelet meristem identity by the branched silkless1 gene in maize. *Science* 298, 1238-1241.

Chuck, G., Brown, P.J., Meeley, R., and Hake, S. (2014). Maize SBP-box transcription factors unbranched2 and unbranched3 affect yield traits by regulating the rate of lateral primordia initiation. *Proc Natl Acad Sci U S A* 111, 18775-18780.

Claeys, H., Vi, S.L., Xu, X., Satoh-Nagasawa, N., Eveland, A.L., Goldshmidt, A., Feil, R., Beggs, G.A., Sakai, H., Brennan, R.G., *et al.* (2019). Control of meristem determinacy by trehalose 6-phosphate phosphatases is uncoupled from enzymatic activity. *Nature Plants* 5, 352-357.

Crow, M., Paul, A., Ballouz, S., Huang, Z.J., and Gillis, J. (2018). Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nature communications* 9, 884.

Csardi, G., Nepusz, T. (2006). “The igraph software package for complex network research.” *InterJournal, Complex Systems*, 1695. <http://igraph.org>.

De Rybel, B., Mähönen, A.P., Helariutta, Y., and Weijers, D. (2016). Plant vascular development: from early specification to differentiation. *Nature Reviews Molecular Cell Biology* 17, 30-40.

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C. (2019). Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. *Developmental cell* 48, 840-852.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* 29, 15-21.

Dong, J., Kim, S.T., and Lord, E.M. (2005). Plantacyanin plays a role in reproduction in Arabidopsis. *Plant physiology* 138, 778-789.

Du, Y., Liu, L., Peng, Y., Li, M., Li, Y., Liu, D., Li, X., and Zhang, Z. (2020). UNBRANCHED3 Expression and Inflorescence Development is Mediated by UNBRANCHED2 and the Distal Enhancer, KRN4, in Maize. *PLoS Genet* 16, e1008764.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol* 13, 1768-1774.

Erichson, N.B., Voronin, S., Brunton, S.L., Kutz, J.N. (2019). Randomized Matrix Decompositions Using R. *Journal of Statistical Software* 89, 1-48. doi: 10.18637/jss.v089.i11.

Eveland, A.L., Goldshmidt, A., Pautler, M., Morohashi, K., Liseron-Monfils, C., Lewis, M.W., Kumari, S., Hiraga, S., Yang, F., Unger-Wallace, E., *et al.* (2014). Regulatory modules controlling maize inflorescence architecture. *Genome Res* 24, 431-443.

Fujimoto, M., Sazuka, T., Oda, Y., Kawahigashi, H., Wu, J., Takanashi, H., Ohnishi, T., Yoneda, J.I., Ishimori, M., Kajiya-Kanegae, H., *et al.* (2018). Transcriptional switch for programmed cell death in pith parenchyma of sorghum stems. *Proc Natl Acad Sci U S A* 115, E8783-e8792.

Funk, V., Kositsup, B., Zhao, C., and Beers, E.P. (2002). The Arabidopsis xylem peptidase XCP1 is a tracheary element vacuolar protein that may be a papain ortholog. *Plant physiology* 128, 84-94.

Gallavotti, A., Zhao, Q., Kyozuka, J., Meeley, R.B., Ritter, M.K., Doebley, J.F., Pe, M.E., and Schmidt, R.J. (2004). The role of barren stalk1 in the architecture of maize. *Nature* 432, 630.

Gipson, A.B., Giloteaux, L., Hanson, M.R., and Bentolila, S. (2020). Arabidopsis RanBP2-Type Zinc Finger Proteins Related to Chloroplast RNA Editing Factor OZ1. *Plants (Basel)* 9, 307.

Guo, P., Zheng, Y., Peng, D., Liu, L., Dai, L., Chen, C., and Wang, B. (2018). Identification and expression characterization of the Phloem Protein 2 (PP2) genes in ramie (*Boehmeria nivea* L. Gaudich). *Scientific Reports* 8, 10734.

Ha, C.M., Jun, J.H., Nam, H.G., and Fletcher, J.C. (2007). BLADE-ON-PETIOLE 1 and 2 control Arabidopsis lateral organ fate through regulation of LOB domain and adaxial-abaxial polarity genes. *Plant Cell* 19, 1809-1825.

Han, J.-J., Jackson, D., and Martienssen, R. (2012). Pod corn is caused by rearrangement at the Tunicate1 locus. *The Plant cell* 24, 2733-2744.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* 38, 576-589.

Herzeel, C., Costanza, P., Decap, D., Fostier, J., and Reumers, J. (2015). elPrep: High-Performance Preparation of Sequence Alignment/Map Files for Variant Calling. *PLoS One* 10, e0132868.

- Horváth, E., Bela, K., Holinka, B., Riyazuddin, R., Gallé, Á., Hajnal, Á., Hurton, Á., Fehér, A., and Csiszár, J. (2019). The Arabidopsis glutathione transferases, AtGSTF8 and AtGSTU19 are involved in the maintenance of root redox homeostasis affecting meristem size and salt stress sensitivity. *Plant Science* 283, 366-374.
- Huffaker, A., Dafoe, N.J., and Schmelz, E.A. (2011). ZmPep1, an ortholog of Arabidopsis elicitor peptide 1, regulates maize innate immunity and enhances disease resistance. *Plant Physiol* 155, 1325-1338.
- Irish, E.E. (1997). Class II Tassel Seed Mutations Provide Evidence for Multiple Types of Inflorescence Meristems in Maize (Poaceae). *American Journal of Botany* 84, 1502-1515.
- Jackson, D., Veit, B., and Hake, S. (1994). Expression of maize KNOTTED1 related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* 120, 405-413.
- Javelle, M., Klein-Cosson, C., Vernoud, V., Boltz, V., Maher, C., Timmermans, M., Depège-Fargeix, N., and Rogowsky, P.M. (2011). Genome-Wide Characterization of the HD-ZIP IV Transcription Factor Family in Maize: Preferential Expression in the Epidermis. *Plant Physiology* 157, 790-803.
- Je, B.I., Gruel, J., Lee, Y.K., Bommert, P., Arevalo, E.D., Eveland, A.L., Wu, Q., Goldshmidt, A., Meeley, R., Bartlett, M., *et al.* (2016). Signaling from maize organ primordia via FASCIATED EAR3 regulates stem cell proliferation and yield traits. *Nature genetics* 48, 785-791.
- Knauer, S., Javelle, M., Li, L., Li, X., Ma, X., Wimalanathan, K., Kumari, S., Johnston, R., Leiboff, S., Meeley, R., *et al.* (2019). A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits. *Genome Res* 29, 1962-1973.
- Konopka, T. (2020). umap: Uniform Manifold Approximation and Projection. R package version 0.2.6.0. <https://CRAN.R-project.org/package=umap>.
- Krizek, B. (2009). AINTEGUMENTA and AINTEGUMENTA-LIKE6 act redundantly to regulate Arabidopsis floral growth and patterning. *Plant Physiol* 150, 1916-1929.
- Kulkarni, A., Anderson, A.G., Merullo, D.P., and Konopka, G. (2019). Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Curr Opin Biotechnol* 58, 129-136.

Lee, J., Shah, M., Ballouz, S., Crow, M., and Gillis, J. (2020). CoCoCoNet: conserved and comparative co-expression across a diverse set of species. *Nucleic acids research* 48, W566-w571.

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

Liu, J., Fernie, A.R., and Yan, J. (2020). The Past, Present, and Future of Maize Improvement: Domestication, Genomics, and Functional Genomic Routes toward Crop Enhancement. *Plant Communications* 1, 100010.

Liu, Q., Wang, H., Zhang, Z., Wu, J., Feng, Y., and Zhu, Z. (2009). Divergence in function and expression of the NOD26-like intrinsic proteins in plants. *BMC Genomics* 10, 313.

Lloyd, J., and Meinke, D. (2012). A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant physiology* 158, 1115-1129.

Lu, Z., Hofmeister, B.T., Vollmers, C., DuBois, R.M., and Schmitz, R.J. (2017). Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic acids research* 45, e41.

Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* 5, 2122.

Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., Marioni, J.C., and participants in the 1st Human Cell Atlas, J. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biology* 20, 63.

Ma, X., Denyer, T., and Timmermans, M.C. (2020). PscB: A Browser to Explore Plant Single Cell RNA-Sequencing Datasets. *Plant Physiology*, pp.00250.02020.

Marchette, D.J. (2015). cccd: Class Cover Catch Digraphs. R package version 1.5. <https://CRAN.R-project.org/package=cccd>.

McCarthy, D.J., Campbell, K.R., Lun, A.T., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33, 1179-1186.

McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems* 8, 329-337.e324.

Miyashima, S., Roszak, P., Sevilem, I., Toyokura, K., Blob, B., Heo, J.-o., Mellor, N., Help-Rinta-Rahko, H., Otero, S., Smet, W., *et al.* (2019). Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature* 565, 490-494.

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., and Takahashi, T. (2006). Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in Arabidopsis. *Plant physiology* 141, 1363-1375.

Nuccio, M.L., Wu, J., Mowers, R., Zhou, H.P., Meghji, M., Primavesi, L.F., Paul, M.J., Chen, X., Gao, Y., Haque, E., *et al.* (2015). Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions. *Nat Biotechnol* 33, 862-869.

Ortiz-Ramírez, C., Arevalo, E.D., Xu, X., Jackson, D.P., and Birnbaum, K.D. (2018). An Efficient Cell Sorting Protocol for Maize Protoplasts. *Current protocols in plant biology* 3, e20072.

Parvathaneni, R.K., Bertolini, E., Shamimuzzaman, M., Vera, D.L., Lung, P.-Y., Rice, B.R., Zhang, J., Brown, P.J., Lipka, A.E., Bass, H.W., *et al.* (2020). The regulatory landscape of early maize inflorescence development. *Genome Biology* 21, 165.

Pautler, M., Eveland, A.L., LaRue, T., Yang, F., Weeks, R., Je, B.I., Meeley, R., Komatsu, M., Vollbrecht, E., and Sakai, H. (2015). FASCIATED EAR4 encodes a bZIP transcription factor that regulates shoot meristem size in maize. *The Plant Cell* 27, 104-120.

Potter, S.S. (2018). Single-cell RNA sequencing for the study of development, physiology and disease. *Nature reviews Nephrology* 14, 479-492.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Ricci, W.A., Lu, Z., Ji, L., Marand, A.P., Ethridge, C.L., Murphy, N.G., Noshay, J.M., Galli, M., Mejía-Guerra, M.K., Colomé-Tatché, M., *et al.* (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nature Plants* 5, 1237-1249.

Rice, B.R., Fernandes, S.B., and Lipka, A.E. (2020). Multi-Trait Genome-wide Association Studies Reveal Loci Associated with Maize Inflorescence and Leaf Architecture. *Plant & cell physiology* 61, 1427-1437.

Rich-Griffin, C., Stechemesser, A., Finch, J., Lucas, E., Ott, S., and Schäfer, P. (2020). Single-Cell Transcriptomics: A High-Resolution Avenue for Plant Functional Genomics. *Trends in Plant Science* 25, 186-197.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (Oxford, England) *26*, 139-140.

Rodriguez-Leal, D., Xu, C., Kwon, C.-T., Soyars, C., Demesa-Arevalo, E., Man, J., Liu, L., Lemmon, Z.H., Jones, D.S., Van Eck, J., *et al.* (2019). Evolution of buffering in a genetic circuit controlling plant stem cell proliferation. *Nature genetics* *51*, 786-792.

Rosvall, M., and Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* *105*, 1118.

Satoh-Nagasawa, N., Nagasawa, N., Malcomber, S., Sakai, H., and Jackson, D. (2006). A trehalose metabolic enzyme controls inflorescence architecture in maize. *Nature* *441*, 227-230.

Schürholz, A.-K., López-Salmerón, V., Li, Z., Forner, J., Wenzl, C., Gaillochet, C., Augustin, S., Barro, A.V., Fuchs, M., Gebert, M., *et al.* (2018). A Comprehensive Toolkit for Inducible, Cell Type-Specific Gene Expression in Arabidopsis. *Plant physiology* *178*, 40-53.

Sigmon, B., and Vollbrecht, E. (2010). Evidence of selection at the *ramosa1* locus during maize domestication. *Molecular ecology* *19*, 1296-1311.

Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. *American journal of human genetics* *91*, 1011-1021.

Strable, J., Wallace, J.G., Unger-Wallace, E., Briggs, S., Bradbury, P.J., Buckler, E.S., and Vollbrecht, E. (2017). Maize *YABBY* Genes *drooping leaf1* and *drooping leaf2* Regulate Plant Architecture. *Plant Cell* *29*, 1622-1641.

Takacs, E.M., Li, J., Du, C., Ponnala, L., Janick-Buckner, D., Yu, J., Muehlbauer, G.J., Schnable, P.S., Timmermans, M.C., Sun, Q., *et al.* (2012). Ontogeny of the maize shoot apical meristem. *Plant Cell* *24*, 3219-3234.

Tong, W., Imai, A., Tabata, R., Shigenobu, S., Yamaguchi, K., Yamada, M., Hasebe, M., Sawa, S., Motose, H., and Takahashi, T. (2016). Polyamine Resistance Is Increased by Mutations in a Nitrate Transporter Gene NRT1.3 (AtNPF6.4) in *Arabidopsis thaliana*. *Frontiers in Plant Science* *7*, 834.

Vollbrecht, E., and Schmidt, R.J. (2009). Development of the Inflorescences. In *Handbook of Maize: Its Biology*, J.L. Bennetzen, and S.C. Hake, eds. (New York, NY: Springer), pp. 13-40.

Walley, J.W., Sartor, R.C., Shen, Z., Schmitz, R.J., Wu, K.J., Urich, M.A., Nery, J.R., Smith, L.G., Schnable, J.C., Ecker, J.R., et al. (2016). Integration of omic networks in a developmental atlas of maize. *Science* 353, 814-818.

Wang, L., Lu, Z., delaBastide, M., Van Buren, P., Wang, X., Ghiban, C., Regulski, M., Drenkow, J., Xu, X., Ortiz-Ramirez, C., *et al.* (2020). Management, Analyses, and Distribution of the MaizeCODE Data on the Cloud. *Front Plant Sci* 11, 289.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80-84.

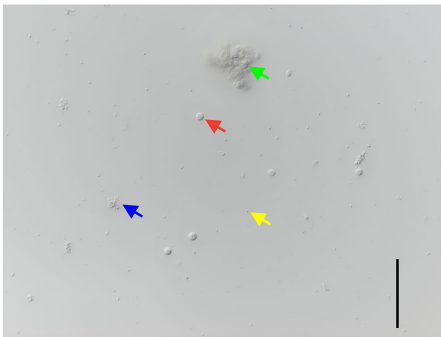
Wu, Q., Xu, F., Liu, L., Char, S.N., Ding, Y., Je, B.I., Schmelz, E., Yang, B. and Jackson, D. (2020). The maize heterotrimeric G protein β subunit controls shoot meristem development and immune responses. *Proceedings of the National Academy of Sciences* 117, 1799-1805.

Yasui, Y., Mukougawa, K., Uemoto, M., Yokofuji, A., Suzuri, R., Nishitani, A., and Kohchi, T. (2012). The phytochrome-interacting vascular plant one-zinc finger1 and VOZ2 redundantly regulate flowering in Arabidopsis. *Plant Cell* 24, 3248-3263.

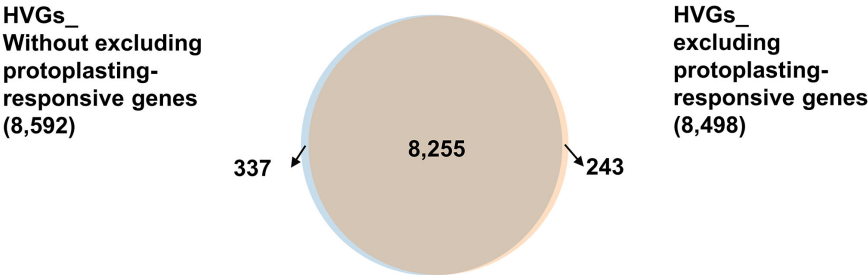
Zhong, R., Richardson, E.A., and Ye, Z.-H. (2007). The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. *The Plant cell* 19, 2776-2792.

SUPPLEMENTARY FIGURES

A



B



C

Meta-clusters without excluding protoplasting- responsive genes	Clusters excluding protoplasting- responsive genes
Meta-cluster 1	No_Proto_cluster 1
Meta-cluster 2	No_Proto_cluster 2
Meta-cluster 3	No_Proto_cluster 3
Meta-cluster 4	No_Proto_cluster 4
Meta-cluster 5	No_Proto_cluster 5
Meta-cluster 6	No_Proto_cluster 6
Meta-cluster 7	No_Proto_cluster 7
Meta-cluster 8	No_Proto_cluster 8
Meta-cluster 9	No_Proto_cluster 9
Meta-cluster 10	No_Proto_cluster 10
Meta-cluster 11	No_Proto_cluster 11
Meta-cluster 12	No_Proto_cluster 12

Figure S1: Protoplasting-responsive genes do not affect clustering. Related to Figure 1.

(A) A representative picture after enzyme digestion shows cells (red arrow), debris (blue arrow), organelles released from broken cells (yellow arrow) and some occasional undigested tissue pieces (green arrow). Scale bar = 100 μ m. (B) Overlap of Highly Variable Genes (HVGs) identified with or without excluding protoplasting-responsive genes. (C) The river plot demonstrates that the meta-clusters are not changed after excluding protoplasting-responsive genes from the analysis. The left-sided grouping represents meta-clusters that are generated without excluding protoplasting-responsive genes (as in Figures 1B and 1C), and right-sided grouping represents meta-clusters generated by excluding protoplasting-responsive genes (No_Proto cluster 1 to No_Proto cluster12). Each cell between left-sided and right-sided groupings is connected with a grey line. The grey rectangles indicate complete match between groups of cells from each side. Protoplasting-responsive genes are identified by bulk RNA-seq (STAR methods) with three biological replicates.

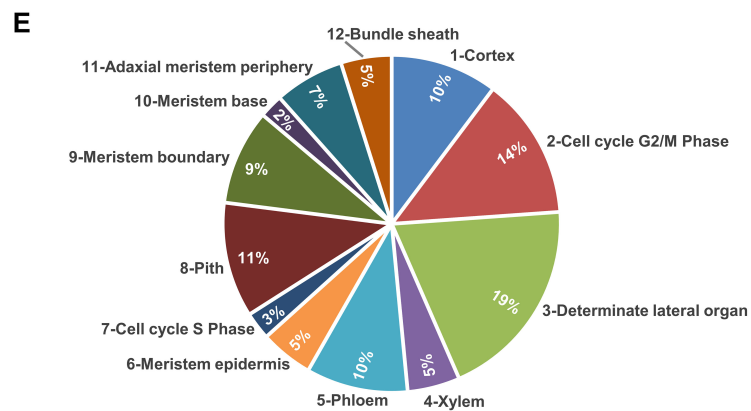
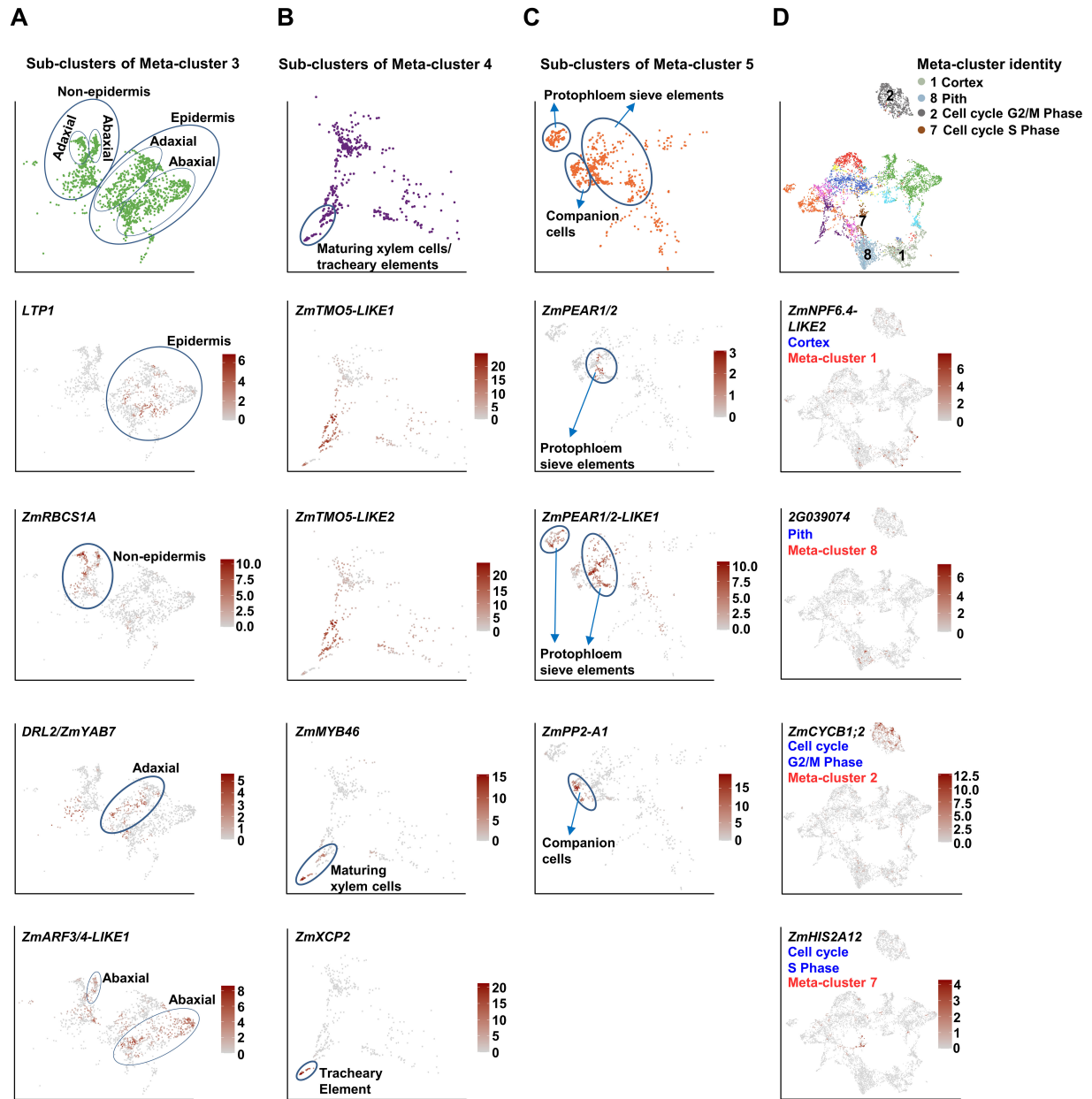


Figure S2: Sub-clusters of meta-clusters 3, 4, and 5 and markers for ground tissue and cell cycle meta-clusters. Related to Figure 1.

(A) Close-up view of meta-cluster 3 shows sub-clusters of epidermis and non-epidermis cells, and adaxial and abaxial cells, with markers shown in UMAPs: epidermis marker *LIPID TRANSFER PROTEIN1 (LTP1)*, *GRMZM2G126646*, non-epidermis marker *ZmRIBULOSE BIPHOSPHATE CARBOXYLASE SMALL SUBUNIT 1A (ZmRBCS1A)*, *GRMZM2G098520*, adaxial marker *DROOPING LEAF2/ZmYABBY7 (DRL2/ZmYAB7)*, *GRMZM2G102218*, and abaxial marker *ZmAUXIN RESPONSE FACTOR3/4-LIKE1 (ZmARF3/4-LIKE1)*, *GRMZM2G159399*. (B) Close-up view of meta-cluster 4 shows xylem markers and sub-cluster for maturing xylem cells/tracheary elements, with markers shown in UMAPs: xylem markers *ZmTARGET OF MONOPTEROS5-LIKE1 (ZmTMO5-LIKE1)*, *GRMZM2G072342* and *ZmTMO5-LIKE2*, *GRMZM2G526668*, maturing xylem marker *ZmMYB DOMAIN PROTEIN 46 (ZmMYB46)*, *GRMZM2G052606*, and tracheary element marker *ZmXYLEM CYSTEINE PEPTIDASE 2 (ZmXCP2)*, *GRMZM2G0367701*. (C) Close-up view of meta-cluster 5 demonstrates sub-clusters of protophloem sieve elements and companion cells, with markers shown in UMAPs: protophloem sieve element markers *ZmPHLOEM EARLY DOF 1/2 (ZmPEAR1/2)*, *GRMZM2G093725*, and *ZmPEAR1/2-LIKE1*, *GRMZM2G017470*, companion cell marker *ZmPHLOEM PROTEIN 2-A1 (ZmPP2-A1)*, *GRMZM2G073693*. (D) UMAP plots highlight two ground tissue meta-clusters, meta-cluster 1 (cortex) and meta-cluster 8 (pith), and two cell cycle-related meta-clusters, meta-cluster 2 (cell cycle G2/M Phase) and meta-cluster 7 (cell cycle S Phase), with markers listed below: cortex marker *ZmNITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER FAMILY 6.4-LIKE 2 (ZmNPF6.4-LIKE2)*, *GRMZM2G361652*, pith marker *GRMZM2G039074 (2G039074)*, cell cycle G2/M phase marker *ZmCYCLINB1;2 (ZmCYCB1;2)*, and cell cycle S phase marker *ZmHISTONE2A12 (ZmHIS2A12)*. Each dot represents a cell. (E) Pie chart shows the percentage of number of recovered cells across all replicates in each meta-cluster.

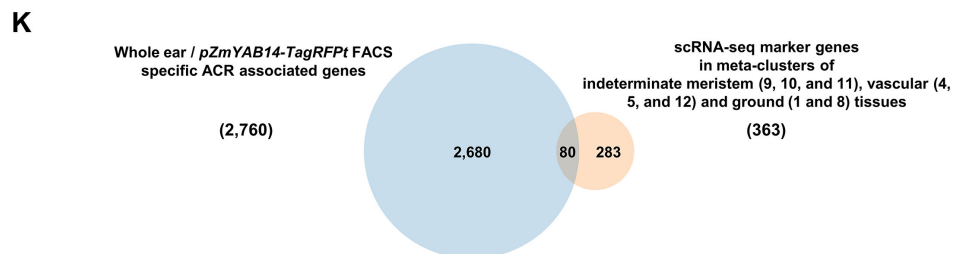
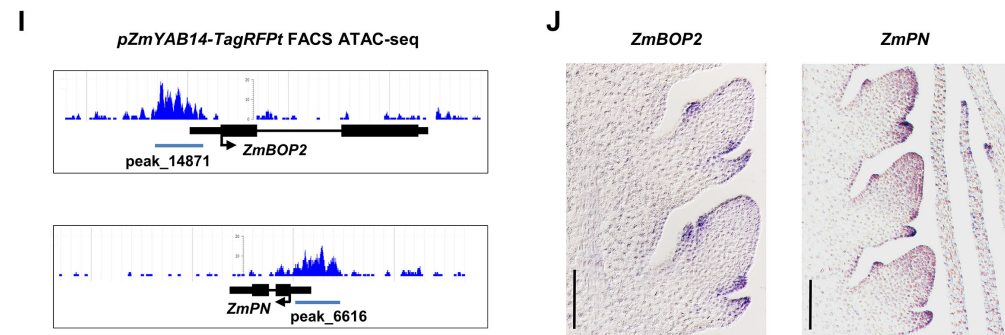
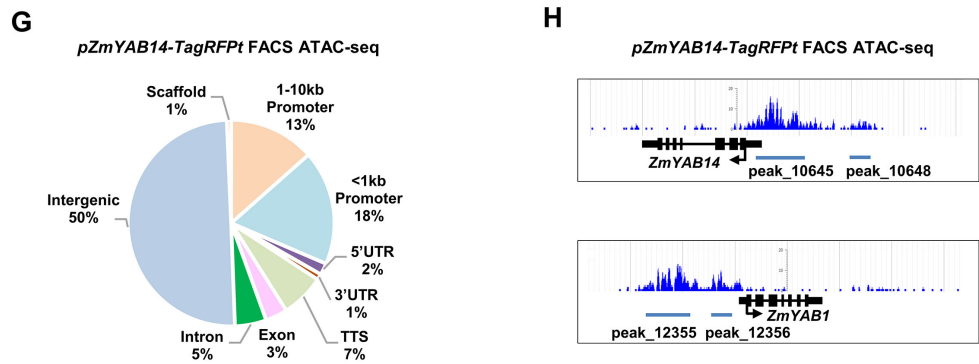
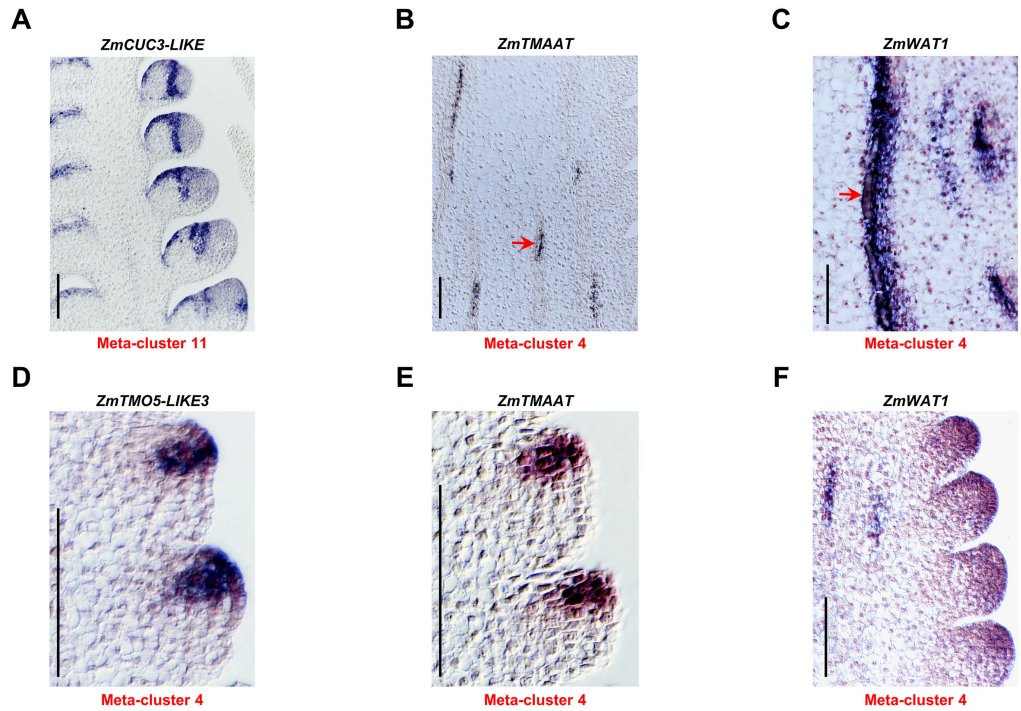


Figure S3: mRNA *in situ* for selected markers in meta-clusters 11 and 4 and integration of scRNA-seq and FACS ATAC-seq. Related to Figure 2.

(A) mRNA *in situ* of meta-cluster 11 (adaxial meristem periphery) marker *ZmCUP-SHAPED COTYLEDON 3-LIKE* (*ZmCUC3-LIKE*), GRMZM2G430522; (B-C) mRNA *in situ* of meta-cluster 4 (xylem) markers, *ZmTRANSMEMBRANE AMINO ACID TRANSPORTER FAMILY PROTEIN* (*ZmTMAAT*), GRMZM2G109865 (B) and *ZmWALLS ARE THIN 1* (*ZmWAT1*), GRMZM2G007953 (C), with red arrow indicating xylem vessels. (D-F) Examples of xylem markers identified from scRNA-seq meta-cluster 4 that also show meristem tip expression: *ZmTARGET OF MONOPTEROS5-LIKE3* (*ZmTMO5-LIKE3*), GRMZM2G176141 (D); *ZmTMAAT* (E) and *ZmWAT1* (F). Scale bar = 100µm. (G) Genome-wide distribution of *pZmYAB14-TagRFpT* FACS ATAC-seq accessible chromatin regions from one biological replicate. (H) Representative scRNA-seq meta-cluster 3 enriched genes from *ZmYAB* families have accessible chromatin in *pZmYAB14-TagRFpT* FACS sorted cells (Table S2), such as *ZmYAB14* (top panel) and *ZmYAB1* (bottom panel). (I) scRNA-seq meta-cluster 3 enriched genes, *ZmBLADE ON PETIOLE2* (*ZmBOP2*), GRMZM2G026556 (top panel), and *ZmPLANTACYANIN* (*ZmPN*), GRMZM2G004012 (bottom panel) have accessible chromatin in their promoter-TSS regions in *pZmYAB14-TagRFpT* FACS sorted cells. (J) mRNA *in situ* shows *ZmBOP2* and *ZmPN* expression in determinate lateral organs. Scale bar = 100µm. (K) scRNA-seq marker genes in meristem (9, 10, and 11), vascular (4, 5, and 12) and ground tissue (1 and 8) meta-clusters are enriched for genes with open chromatin in whole ear but not in *pZmYAB14-TagRFpT* FACS sorted lateral organ cells, see text for details.

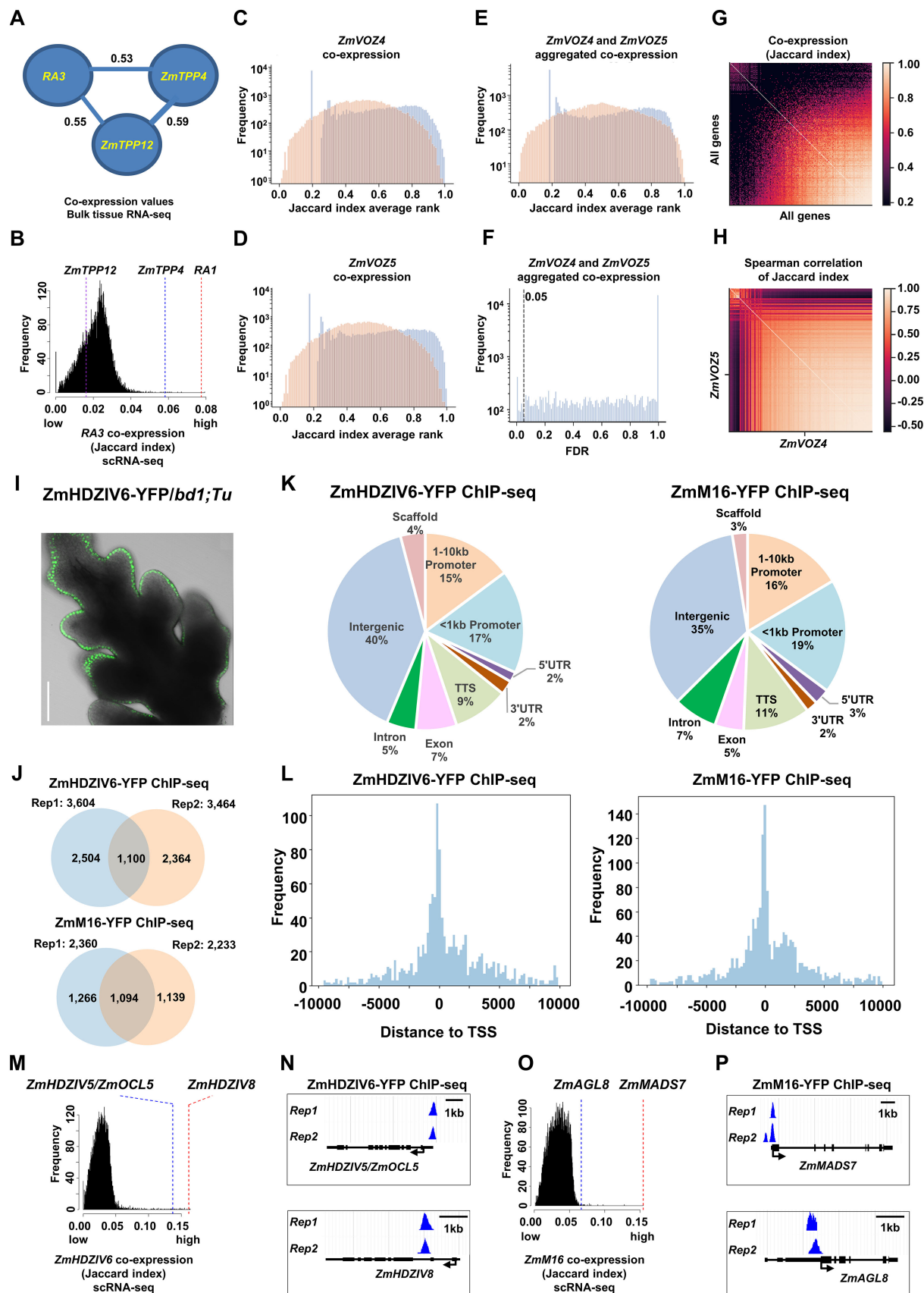


Figure S4: Application of scRNA-seq to predict genetic redundancy and build transcriptional regulatory networks. Related to Figure 3.

(A) Similar co-expression values of *RA3-ZmTPP4* (0.53), *RA3-ZmTPP12* (0.55) and *ZmTPP4-ZmTPP12* (0.59) from an aggregate network across 89 maize bulk tissue RNA-seq datasets. (B) Jaccard index co-expression values are higher between *RA3* and redundantly acting *ZmTPP4* than *ZmTPP12* in scRNA-seq datasets. X-axis represents co-expression values of *RA3* and all maize genes as reflected by the Jaccard index from low to high. Y-axis represents the frequency of each co-expression value. Colored dashed lines indicate corresponding gene co-expression values with *RA3*. (C-D) Empirical distribution (blue color) of *ZmVOZ4*, *GRMZM2G156016* (C) and *ZmVOZ5*, *GRMZM2G449165* (D) co-expression metrics. X-axis represents Jaccard index co-expression values. Y-axis represents log scaled frequency. The orange color indicates null distribution. (E-F) The overlap in co-expression distribution of *ZmVOZ4* and *ZmVOZ5* (E) identify 627 genes (Table S3) that are consistently co-expressed with both *ZmVOZ4* and *ZmVOZ5* (F) at cutoff at $FDR < 0.05$. (G) Pairwise similarity in co-expression partners using Jaccard index values ordered from lowest to the highest expression of all maize genes. (H) Spearman correlation of co-expression profiles between each pair of genes, indicating similarity in co-expression partners. *ZmVOZ4* and *ZmVOZ5* have Spearman correlation of 0.88 in their co-expression profiles. (I) Epidermis specific expression of *ZmHDZIV6*-YFP in the *bd1;Tu* line. Scale bar = 100 μ m. (J) Two biological replicates of *ZmHDZIV6*-YFP (top panel) and *ZmM16*-YFP (bottom panel) overlap significantly. (K) Genome-wide distribution of *ZmHDZIV6* (left panel) and *ZmM16* (right panel) bound peaks. (L) Bound peaks within ± 10 kb regions for both *ZmHDZIV6* (left panel) and *ZmM16* (right panel) are strongly enriched around transcription start sites (TSS). (M-N) *ZmHDZIV6* predicted direct targets, *ZmHDZIV8* and *ZmHDZIV5/OCL5*, identified by co-expression in scRNA-seq (M) and ChIP-seq (N). (O-P) Predicted *ZmM16* directly activated targets, *ZmMADS7* and *ZmAGAMOUS-LIKE 8* (*ZmAGL8*, *GRMZM2G079727*), identified by co-expression in scRNA-seq (O) and ChIP-seq (P). In the histogram of M and O, X-axis represents the co-expression values of *ZmHDZIV6* (M) or *ZmM16* (O) with all maize genes, reflected by the Jaccard index from low to high. Y-axis represents the frequency of each co-expression value. Dashed lines indicate co-expression values of predicted direct targets with *ZmHDZIV6* (M) or *ZmM16* (O).