A multiplexed transcriptomic analysis of a plant embryonic hourglass

Hao Wu[1*], Ruqiang Zhang[1], Michael J. Scanlon[1*]

[1]School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

[*] Corresponding authors

Abstract

The embryo defines the primordial developmental stage of multicellular organisms, from fertilization of the zygote to the formation of tissue layers and organs. Animal zoologists [1, 2, 3] described a period of maximum morphological similarity between embryos of different species within related taxa, later termed the phylotypic stage[4] of mid-embryogenesis. Modern transcriptomic studies identified a phylotypic pattern of gene expression in Drosophila, zebrafish, and the flowering plant Arabidopsis, featuring enhanced accumulation of evolutionarily older and conserved transcripts during mid-embryogenesis[5, 6, 7]. Such plots of transcriptomic age and divergence conform to the "hourglass models" of morphological similarities described by 19th century embryologists. Chronological descriptions of the phylotypic stage are problematic for plants, where embryogenesis extends throughout the life cycle and the mid-point of seed maturation is morphologically and chronologically variable among taxa[7]. Here we use multiplexed spatial-transcriptomic analyses to investigate the homology of the grass embryo, which evolved ~ 150 million years since the divergence of eudicots and grasses[8]. Our data support a model where the single maize embryonic leaf (cotyledon) is a novel, bi-partite organ derived within the monocotyledonous grasses, after the

evolution of the ancestral angiosperm cotyledon from a modified embryonic leaf [9, 10]. We describe a maize embryonic-organ-initiation genetic-network that uses ancient and conserved genes, comprising a phylotypic pattern of "mid embryogenesis" during the ancestrally conserved initiation of histogenesis and organogenic growth in plant and animal embryos.

Introduction

Following fertilization and formation of the diploid zygote, embryogenesis in flowering plants comprises three widely conserved stages[11]: (1) polarized and asymmetric cell divisions to create the embryo proper; (2) patterning of discrete histological tissue-layers; and (3) organogenesis of lateral organs alongside the formation of pluripotent meristems. Persistent stem-cell reservoirs comprising the shoot apical meristem (SAM) and root apical meristem (RAM) enable plants to extend organogenesis long-beyond embryonic stages. Seed maturation and quiescence, the onset of which is widely-variable and correlated with the mode of seed dispersion, is a commonly-utilized but not universally-conserved strategy of plant embryogenesis[11, 12].

Embryos from the eudicot *Arabidopsis thaliana* and the monocot grass *Zea mays* ssp. *mays* both use asymmetric cell divisions to establish the embryo proper that will generate the body plan and the suspensor (Fig. 1a), a temporary supporting structure that is later consumed by programmed cell death[13]. Histological layering of the globular embryo proper forms an outer protoderm surrounding layers of internal tissue(s) in grass and eudicot embryos[14, 15, 16, 17]. Thereafter, Arabidopsis embryos develop paired, embryonic leaves called cotyledons and organogenic meristems in the shoot and root poles, whereafter organogenesis is suspended until seed germination[10, 11]. In contrast, the morphology of the grass embryo is dissimilar to that of

Arabidopsis. The first lateral organ formed in the maize embryo is the scutellum (Fig. 1b), which at maturity forms a shield-shaped, glandular, non-photosynthetic structure that absorbs nutrients from the germinating seed[13, 17]. Subsequently, the coleoptile emerges as the second maize embryonic structure (Fig. 1c), and later forms a cylindrical sheath that protects the underlying, fragile shoot during seedling germination. Unlike Arabidopsis, grasses extend embryonic organogenesis to form several foliar leaves before seed quiescence. Upon seed germination, lateral organ production resumes at the SAM, and continues throughout the life cycle[14, 15]. The homology of the single cotyledon in monocotyledonous grasses is an unresolved topic of botanical debate[10, 18, 19, 20].

Result

Single-cell profiling of the transcriptomic landscapes in Stage 1 maize embryo

Historically, homology is inferred on the basis of developmental morphology and/or ontogenetic expression of conserved genetic networks[21, 22]. Stage 1 maize embryos are approximately 1 mm long and contain a prominent, ovately-flattened scutellum, an emerged coleoptile that is beginning to surround the SAM, and the newly-initiated primordium of the first foliar leaf[15] (Fig. 1d). 10X Genomics Chromium™ was used to profile the single-cell transcriptomes of 11,467 protoplasts digested from over 400 Stage 1 embryos in two replicates (Fig. 1e). Twenty-three cell clusters were identified; median values of 10,594 UMI per cell and 4,030 genes per cell were obtained.

A first challenge encountered in analyses of sc-RNAseq data is the identification of the tissue/organ-specific origin of each cell-cluster. Since prior RNA-seq analyses in maize have

focused on seedlings or inflorescences[23], cell-clusters derived from the SAM, embryonic stem, and vasculature[24] are readily identified, as is the residual starchy endosperm (Fig. 1e; Clusters 6, 15 and 20 respectively). Relatively few studies have analyzed the transcriptomic patterns within the grass scutellum or coleoptile, such that the origins of several cell-clusters were not immediately identifiable. To annotate the UMAP cell-clusters, we utilized a multiplexed transcriptomic approach combining these sc-RNAseq data with spatial-transcriptomics and laser-microdissection Ranse (LM-RNAseq) to obtain spatial information for all the cell-clusters, as described below (Fig. 2; Extended Data Fig. 1-2, Extended Data Table 1).

Eight individual Stage 1 embryo sections were processed using the 10X Genomics Visium™ spatial transcriptomics pipeline (Fig. 2a-d). Nine spatially resolved cell-clusters were identified and categorized into four supergroups: (1) shoot-root axis; (2) scutellum epidermis; (3) internal scutellum; and (4) endosperm/embryo boundary (Fig. 2e). Hypergeometric testing analyzed the overlap of cluster marker genes between the single-cell and spatial-transcriptomic data, suggesting that 14 of the 23 single-cell clusters (0-5, 7-9, 11, 13, 14, 18 and 19) are derived from the scutellum (Fig. 2h). Comparisons of transcript localizations of known marker genes within the UMAP and the spatial-transcriptomic plots suggest that the resolution of Visium™ technology may comprise 5-10 embryonic cells. Thus, it was not yet possible to spatially-distinguish transcripts among the coleoptile, the SAM, and leaf1. To increase the resolution of our embryo transcriptomics, LM-Ranse was utilized in micro-dissected transcriptomic analyses of the SAM, adaxial scutellum, upper/lower coleoptile, and leaf primordium in Stage 1 embryos (Fig. 2f-g). Principal component analysis (PCA) shows that the transcriptomic profiles of these five embryonic organ/structures are well-separated (Extended Data Fig. 2a). In particular, 9,913

genes are identified as differential-expressed in at least one embryonic structure, and there are 89, 1058, 246, 84 and 53 genes preferentially expressed in these same five embryonic regions (Extended Data Fig. 2b-f). Thus, this multiplexed transcriptomic approach enabled identification of the tissue/organ origin of each cell-cluster in the maize Stage 1 embryonic UMAP (Fig. 1e; Fig. 2i).

Temporal and spatial analysis of gene expression via multiplexed RNA-targeting

Homologous organs utilize homologous genes regulating developmental pattern[21, 25]. With an emphasis on candidate regulatory genes differentially expressed in our transcriptomic analyses of Stage 1 embryos, we targeted 100 maize transcripts for temporal and spatial analyses of multiplexed transcript *in-situ* localization using the 10X Xenium™ protocol on longitudinal sections of (i) Transition-staged embryos just forming a SAM and scutellum (Fig. 1b), (ii) Coleoptilar-staged embryos (Fig. 1c), and (iii) Stage 1 embryos initiating a leaf (Fig. 1d). We first examined the correlations of our sc-RNAseq data, spatial-transcriptomic plots, and Xenium™ *in situ* RNA-targeting data via comparisons of 16 genes preferentially expressed in one or two single-cell clusters (Extended Data Fig. 1). In all cases examined, the *in situ* hybridization patterns are consistent with the spatial expression data. Satisfied with the consistency of these multiplexed *in situ* data, we next analyzed the dynamic co-localization patterns of some key regulators of maize shoot development during serial stages in grass embryo ontogeny.

Consistent with prior studies[26, 27], the spatial-temporal expression of the indeterminacy marker *KNOTTED1* (*KN1*; Fig. 3a-d) is first-detected during SAM initiation at the Transition stage (Fig.

1b). At later stages, *KN1* accumulates in the SAM and subtending embryonic stem, but is downregulated in lateral organs[26, 27]. *NARROW SHEATH1* (*NS1*) and *WUSCHEL-RELATED HOMEOBOX 3A* (*WOX3A*) encode homeodomain transcription factors required for mediolateral expansion of maize lateral organs[25], and are co-expressed at edges of the developing coleoptile and initiating leaf primordia as described previously[17, 21, 28] (Fig. 3b-c, f-g). Undetected in prior studies of embryo ontogeny, Fig. 3a reveals *NS1* transcript accumulation in the initiating scutellum at the Transition stage. *CUP-SHAPED COTYLEDON2* (*CUC2*) marks developmental boundaries throughout maize ontogeny[29]; *CUC2* transcripts are detected in-between the expression domains of *KN1* and *NS1* during scutellum initiation (Fig. 3a), and in the boundaries between the SAM and later-staged embryonic lateral organs (Fig. 3b-c,e). *DROOPING LEAF1* (*DRL1*) regulates leaf architecture and is expressed in incipient and developing foliar leaf primordia[30] but is not detected in the coleoptile or scutellum at any embryo stage examined (Fig. 3c,h). These combinatorial *in situ* RNA-targeting analyses enable identification of previously undescribed combinatorial patterns of dynamic, marker gene expression during maize embryo development.

Other candidate genes analyzed by Xenium™ displayed patterns of transcript accumulation during embryo ontogeny that offer insights into the developmental homologies of maize embryonic lateral organs. For example, *BLADE-ON-PETIOLEa* (*BOPa*) accumulates throughout the developing coleoptile and in the proximal region of the initiating leaf at Stage 1 (Fig. 3i-k), consistent with the proposed function of *BOP* homologs as leaf proximal-identity genes[31]. Notably, *BOPa* is predominantly undetected throughout scutellum ontogeny; similar accumulation patterns are noted for the *BOPa* paralog *TASSELS REPLACE UPPER EARS1*

(*TRU1*) (Extended Data Fig. 3). Transcripts of the *DRL1*-paralog *YABBY14* (*YAB14*) are densely accumulated in the apical tip of the scutellum at the transition stage and in the elongating scutellum and newly emerged coleoptile at the coleoptilar stage. During Stage 1, a new *YAB14* expression maximum is organized at the primordial leaf tip (Fig. 3l-n). Moreover, the pleiotropic gene *INDETERMINATE GAMETOPHYTE1* (*IG1*) is implicated to function during leaf polarity and patterning of maize leaves[32, 33]. *IG1* accumulates on the top (i.e. adaxial) surfaces of initiating leaf primordia during Stage 1, but also in adaxial layers of the emerging coleoptile and in the upper-adaxial region of the transition-staged scutellum (Fig. 3o-t) No significant *IG1* expression is noted during later stages of embryonic scutellum development (Fig. 3p-q, s-t). Another leaf polarity regulator, *AUXIN RESPONSE FACTOR 3* (*ARF3*)[34], is expressed on the abaxial region of the transition-staged embryo during scutellum initiation, and in the abaxial domains of the newly-emerged coleoptile and leaf during later embryo stages (Fig. 3u-z). Interestingly, the expression of *ARF3* in the scutellum moves from abaxial domains in transition stage to adaxial region at Stage 1 (Fig. 3z).

The initiation of maize embryonic lateral organs utilizes a shared genetic network

Homologous organs are inferred to utilize evolutionarily-conserved developmental genetic networks[22]. By Stage 1, the three grass embryonic lateral organs have already begun to diverge morphologically, and will each diverge further before seed quiescence[15]. Interrogations of lateral organ homology require investigation throughout embryo ontogeny[22]. We therefore combined the LM-RNAseq data from the Stage 1 embryonic organs generated in the current study with our previous LM-RNAseq data from serial stages of maize embryonic development[24], to generate a more complete sampling of grass embryogenesis. Specifically, these prior analyses generated

transcriptomic data precisely from the newly initiated embryonic organ and the adjacent SAM, which were serially sampled at the transition stage (scutellum +SAM), the coleoptilar stage (coleoptile +SAM) and Stage 1 (leaf 1 +SAM). Analyses of pairwise differential gene expression among the five embryonic organs reveals a patterned expression of 53 leaf-preferential genes in Stage 1 embryos (Extended Data Fig. 2e), wherein their overall expression is highest in leaf 1 (L) and less in the lower coleoptile (LC), with steadily decreased expression in the upper coleoptile (UC) and scutellum (S) (i.e. L > LC > UC > S).

Notably, the expression of those leaf-preferential genes during the transition stage and the coleoptilar stage is similar to that discovered in leaf 1 during Stage 1 (Extended Data Table 2)[24]. These data suggest that genes involved in foliar leaf initiation at Stage 1 are also associated with the initiation of the scutellum and coleoptile at successively earlier stages in embryo ontogeny. To further interrogate this putative maize embryonic-organ-initiation network, a weighted gene co-expression network analysis (WGCNA)[35, 36] was performed on our Stage 1 embryonic LM-RNAseq data to identify co-expression modules correlated with the expression pattern L > LC > UC > S. A total of 5,797 gene transcripts with high expression variability were selected; all except one non-clustered transcript (MEgrey module) were clustered into twelve co-expression modules (ME1 through ME12). Two co-expression modules (ME2 and ME3) comprising 963 genes emerged as correlated (Extended Data Fig. 4a; Fig. 4a). Intersection of these 963 genes with transcripts previously found via LM-RNAseq of the transition-staged apex[24] (excluding putative SAM-specific genes) identified 130 genes comprising the maize embryonic-organ-initiation network (Fig. 4a, Extended Data Fig. 4b). Expression of these 130 network genes is significantly and positively correlated with the initiation of the transition-staged scutellum, the

coleoptilar-staged coleoptile, and the Stage 1 leaf and lower-coleoptile (Extended Data Fig. 4c-e). During Stage 1, the expression levels of this embryonic-organ-initiation network significantly decrease in the scutellum and upper coleoptile, but not in the lower coleoptile or leaf 1 (Extended Data Fig. 4b, Fig. 4b). Gene ontology (GO) enrichment analyses showed that these network genes are replete for the GO term Organ Development, whereas the transcriptomic networks of the Stage 1 scutellum and coleoptile are enriched for the GO terms Carbohydrate and Lipid Metabolism, and Organ Development/Lipid Metabolism, respectively (Fig. 4c). The data suggests that during initiation, all maize embryonic lateral organs utilize a similar genetic network, the expression of which decreases in later embryonic stages as the scutellum, coleoptile, and leaf differentiate to perform distinct functions in later plant development (Fig. 4d).

A central question in the controversial homology of grass embryonic organs is the identity of the coleoptile as either a foliar leaf[18], the entire grass cotyledon[19], or the proximal component of a bipartite, single grass cotyledon[20]. A second WGNCA analysis was performed on the LM-RNAseq data obtained from the Stage 1 embryonic shoot apex. In addition, we integrated published LM-RNAseq data[37] of the Plastochron 0 (P0) and P1 leaf from 14-day after planting maize seedlings as a reference for later leaf ontogeny. Batch effects were removed (Extended Data Fig. 5a,b), and the merged dataset was filtered to select genes with variability in transcript accumulation levels across embryonic organs. In the analysis, 5,128 out of 5,137 selected genes were clustered into 15 co-expression modules (LM1, LM2, …, LM15), whereas the remaining 9 transcripts resolved to an unclustered module (Extended Data Fig. 5c; Fig 4d).

Pearson correlation tests between module eigengenes and organ-specific transcriptomes identified associations between specific co-expression modules and the corresponding embryonic organs (Fig. 4d, Extended Data Fig. 5c). Transcriptomic relationships among the co-expression modules are reflected in a tree graph (Fig. 4d), in which the y-axis (height) scales the length of organ-specific branches, indicating transcriptomic distances between these modules (Fig. 5a). Two major branches comprise (1) the Stage 1 leaf, SAM, P0 and P1, and (2) the coleoptile and the scutellum.

To investigate the transcriptomic relationships among all Stage 1 maize embryonic lateral organs at single-cell resolution, we then performed hdWGCNA[38] on the scRNAseq data. A total of 5,026 genes transcripts were clustered into 11 co-expression networks (M1, M2, …, M11). A hypergeometric test analyzed the overlap between the putative single-cell cluster marker-gene transcripts and those highly correlated with corresponding co-expression modules. The data illustrate that the SAM and the leaf cluster marker-genes significantly overlap with the closely related modules M9 and M10, respectively. In contrast, marker genes from the scutellum and the coleoptile cluster overlap with multiple co-expression modules (M5, M7 and M8; Fig. 5b). In summary, these hdWGCNA analyses of embryo scRNA-seq data suggest close transcriptomic relationships between the initiating/expanding foliar leaves and the SAM, and between the scutellum and the coleoptile.

The embryonic-organ-initiation genetic network stage is conserved in Arabidopsis embryos

To investigate the transcriptomic similarities of embryonic lateral organs of maize and those of the eudicot model Arabidopsis, we used comparative genomic toolkits [GRAMENE

(https://www.gramene.org/), TAIR (https://www.arabidopsis.org/), BLAST tools (https://blast.ncbi.nlm.nih.gov/Blast.cgi)] to select Arabidopsis orthologs of the 130 transcripts identified to comprise the maize, embryonic-organ-initiation network, and interrogated the temporal, gene expression atlas of Arabidopsis embryo development[39]. Nomenclature describing embryo stages differs in eudicots and grasses[11]. In Arabidopsis, histological layering and SAM initiation occur during the globular stage, cotyledon initiation and emergence defines the heart stages, whereafter the cotyledons elongate during the torpedo and bent embryo stages.

Out of 130 genes in the maize embryonic-organ-initiation network, forty-nine have corresponding Arabidopsis orthologs expressed during embryo development, commencing during cotyledon initiation at the early heart stage and persisting until the bent cotyledon stage (Supplementary Dataset; Extended Data Fig. 6). Notably, the stage-specific accumulation patterns of these homologous maize transcripts in the emerging scutellum (Transition stage), coleoptile (Coleoptilar stage), and foliar leaf (Stage 1) are significantly correlated with the reported expression of the corresponding Arabidopsis orthologs during specific stages in cotyledon initiation and development in this eudicot embryo[39] (Fig. 4f, Extended Data Fig. 6). In addition, transcripts accumulating in the expanding Stage 1 scutellum correlate with expression of Arabidopsis orthologs during later stages of cotyledon growth (i.e. late-torpedo stage[39]). Likewise, maize transcripts accumulating in the elongating Stage 1 coleoptile are significantly correlated with expression of orthologous Arabidopsis genes during earlier stages of cotyledon expansion (i.e. early heart stage[39]) (Fig. 4f). These data suggest that despite the morphological dissimilarities of these grass and eudicot “mid-staged” embryos, the development of the maize and Arabidopsis embryonic lateral organs share a homologous regulatory network.

The embryonic organ initiation network comprises a phylotypic stage

We show that the maize embryonic-organ-initiation network is conserved in *Arabidopsis thaliana*, where it correlates with the transcriptomic signatures of the eudicot heart and torpedo stage embryos. Prior studies revealed that during mid-embryogenesis, old and conserved genes are more highly expressed than young and divergent genes, marking the phylotypic stage of Arabidopsis embryogenesis[7]. To investigate if the maize embryonic-organ-initiation network is associated with a phylotypic stage, we analyzed the age (TAI) and divergence (TDI) of the genes in the maize embryonic-organ-initiation network.

In detail, to evaluate gene age, we assigned a phylostratum to each protein-coding gene in the maize genome (PS) by performing BLAST sequence alignment analyses on each protein sequence in the embryonic-organ-initiation network to the protein sequences of 350 other species, while searching for the most distant phylogenetic node containing at least one species with a homologous protein domain.

We obtained 13 phylostrata, PS1-PS13 (Fig. 5a; Supplementary Dataset), from the oldest genes with domains homologous to prokaryotic genes to the youngest genes with no homologs identified in any other species. We utilized the transcriptome age index (TAI) to measure the average gene age, adjusted by level of gene expression, at each developmental stage sampled. As shown in Fig 5b and Extended Data Fig. 7a-c, the TAI undergoes a dramatic decrease beginning at the proembryo stage and reaches a minimum value in the Stage 1 scutellum. This hourglass pattern, characterized by low TAI values, continues through the Coleoptilar stage and Stage 1

followed by an increase at Stage 2 of embryogenesis and continued increased TAI values in the newly initiated leaves of germinated maize seedlings (Fig. 5b).

To evaluate the gene sequence divergence, calculated the K_a/K_s ration between maize and its sister species *Zea diploperennis.* K_a and K_s refer to the number of non-synonymous and synonymous substitutions, respectively, between each orthologous pair. The transcriptome divergence index (TDI) measures the average divergence of genes weighted, by transcript accumulation level, at different developmental stages. Consistent with the TAI results, the TDI also showed a reduction from the proembryo stage to the initiation of the first lateral organs at the Transition and Coleoptilar stages, and until Stage 1, followed by a dramatic increase at Stage 2 and in germinated seedling leaf primordia (Fig. 5c; Extended Data Fig. 7d-e). Thus, measures of both transcriptional gene age and gene sequence divergence show an “hourglass” pattern during maize embryogenesis, indicating that the processes of tissue layer organization and lateral organ initiation are associated with evolutionarily older and more conserved genes.

Discussion

A multiplexed transcriptomic approach was applied to investigate the developmental homologies of grass embryonic organs. Spatial transcriptomics complements the ambiguities encountered in cell-cluster identification that can complicate scRNA-seq analyses; the precision of LM-RNAseq can differentiate the tissue/organ-specific transcriptomes of embryonic micro-structures that are unresolved using spatial transcriptomics alone (Fig. 2). Moreover, high-throughput, multiplexed *in situ* RNA-targeting technologies enable localizations of multiple transcripts simultaneously and combinatorically, throughout embryo ontogeny. RNA-targeting analyses confirm the expression of several key regulators comprising a transcriptomic network that functions at the

initiation of the grass scutellum, coleoptile, and foliar leaf (Fig. 3). Thus, the combinatorial usage of these four transcriptomic approaches complements one another to build a high throughput, highly resolved, and spatially informed transcriptomic profile of grass embryo development.

Although the evolution of the plant cotyledon from conifer and/or dicot ancestral leaves is still unclear[10], genetic and morphological evidence support models wherein the cotyledon evolved via modification of the first embryonic leaf[11, 40]. Comparisons of transcriptomic profiles from all Stage 1 embryonic lateral organs identified a conserved embryonic-organ-initiation network, suggesting that the scutellum, coleoptile, and leaf 1 are homologous organs (Fig. 4a; Extended Data Fig. 4a-b). Accumulation of transcripts comprising this embryonic-organ-initiation network is diminished during later developmental stages (Fig. 4b; Extended Data Fig. 4a-b), whereupon each embryonic lateral organ is specifically modified to perform discrete functions during seed germination[13, 17]. Organ-specific changes in gene expression during later embryonic stages are observed in coleoptile and leaf primordia, wherein expression of the embryonic-organ-initiation network decreases as these organs exercise differential gene-expression to form functionally divergent organs.

Classical plant morphologists have proposed disparate models for the homology of maize embryonic lateral organs[reviewed in 10]. One model suggests that the scutellum is homologous to the cotyledon, and the coleoptile is a modified foliar leaf[18, 19]. Others proposed that the coleoptile comprises the maize cotyledon, and the scutellum is an evolutionarily unique haustorium that arose within the grasses[19]. A synthetic, bipartite hypothesis suggests that the scutellum and the coleoptile together, comprise the single grass cotyledon, a derived character that arose within the

grasses. In this view the scutellum comprises the apical region, and the coleoptile forms the sheathing base of the singular cotyledon[20].

Quality characteristics of Arabidopsis cotyledons include their unique status as the first-initiated embryonic organs and their ultimate function in supporting the germinating seedling. The scutellum is the first lateral organ initiated in grass embryos, whereas both the scutellum and coleoptile satisfy the second criterion of seedling support at maturity, albeit in very different ways[15, 41]. Transcriptomic analyses of Arabidopsis cotyledon development reveal significantly correlated expression of orthologs of the maize embryonic-organ-initiation network, beginning during embryo histogenesis in the globular-staged embryo and persisting until later-staged cotyledon growth and expansion (Fig. 4f). Expression of the embryonic-organ-initiation network during maize coleoptile and scutellum development at Stage 1 is significantly correlated with that of mid-stages of Arabidopsis cotyledon development, suggesting that the Arabidopsis cotyledon is homologous to the maize scutellum *and* coleoptile.

Our WGCNA analyses support a model where the maize cotyledon is a bi-partite embryonic organ comprising the scutellum and the coleoptile (Fig. 4). The evolution of a single grass cotyledon as a composite embryonic organ raises interesting questions. Grass phyllotaxy is distichous; successive organs initiate ~180° apart and in two hierarchical ranks about the stem[41]. If the scutellum and coleoptile were separate embryonic organs initiated successively from the embryonic axis, they are predicted to emerge from opposite sides of shoot. However, the scutellum and coleoptile are both initiated on the same side of the grass embryo, although the

scutellum precedes the coleoptile[15] (Fig. 1a-d). Our composite data suggest that the scutellum and coleoptile comprise different domains of the single grass cotyledon[42] (Fig. 6).

Foliar leaves of grasses contain a distal, photosynthetic blade and a proximal sheath that surrounds and overlaps the stem[41]. A morphological model for grass cotyledon homology[19, 20] proposes that the scutellum and coleoptile are homologous to the distal blade and the sheathing base of the foliar leaf, respectively (Fig. 6). Consistent with this model, cells contributing to the distal leaf blade emerge from the shoot apex prior to the emergence of leaf sheath progenitor cells[25, 43, 44, 45]. This pattern mirrors the ontogenetic timing of scutellum and coleoptile initiation. Furthermore, expression of the paralogous sheath marker genes *BOPa* and *TRU1* is restricted to the maize leaf sheath[29, 46] and to the Stage 1 coleoptile (Fig. 3i-k, Extended Data Fig. 4), in support of models for a bipartite grass cotyledon. Higher-order *Osbop* mutants delete the rice leaf sheath[31], supporting hypotheses for BOP-mediated development of sheathing leaf bases. Taken together, these data suggest that the coleoptile and the leaf sheath comprise the sheathing bases of homologous lateral organs in grasses.

Our multiplexed transcriptomic analysis of maize embryogenesis identifies a network of genes expressed during initiation of embryonic organs that comprises a phylotypic period. As previously described in the eudicot plant Arabidopsis and in animal embryos of fruit flies (*Drosophila melanogaster*) and zebrafish (*Danio rerio*)[5, 6, 7], the transcriptomic phylotypic period is defined by the preferential expression of conserved and evolutionarily ancient genes (Fig 5b-c). We note that this transcriptomic phylotypic period in maize embryogenesis commences when tissue layers are established during the late proembryo stage and forms an extended hourglass

shape corresponding to successive initiations of the scutellum (T stage), coleoptile (C stage), and leaf 1 (Stage 1). Striking decreases in both transcript age and conservation occur during later stages of maize embryogenesis (Stage 2) and upon initiation of seedling leaves post-germination (Fig. 5b-c). Considering the number of genes represented, most of the 130 transcripts in this phylotypic maize embryonic-organ -initiation-network evolved in either phylostratum PS1 (unicellular eukaryotes and prokaryotes), PS2 (all eukaryotes), or PS4 (Bryophyta, i.e. the first land plants). Less than 10 percent of this maize embryonic network comprises newer transcripts, from PS5-PS13, that arose later in plant evolution (Extended Data Fig. 7b; Extended Data Fig. 8a; Supplementary Dataset). Notably, the majority of genes in this network evolved within PS4, which marks the first appearance of the plant embryo[8]. In contrast, just 12 embryonic transcripts (9.23%) are identified from algae (PS3), comprising plants that do not contain embryos. Relatively, gene expression numbers during the non-phylotypic Proembryo stage are more dominated by later evolved genes from PS9-PS11(i.e. the Poaceae or grasses, the panicoid grasses, and the genus *Zea*; Fig. 5a), whereas transcripts from Stage 2 embryos and seedling leaves are enriched for newly evolved genes from PS12 and PS13 and the genus *Zea* (Extended Data Fig. 8a). Thus, as in the eudicot Arabidopsis and in the animal models Drosophila and Zebrafish, the transcriptomic hourglass pattern of the phylotypic period of maize is decidedly embryonic, and is not identified during later, seemingly equivalent organogenic events that continue throughout the life cycle of plants.

The maize phylotypic period displays higher TAI values (i.e. expression of newer genes) at the Transition and Coleoptilar stages than during Stage 1 (Fig 5b; Extended Data Fig. 8b-i). These data likely reflect the relatively recent innovation of the bipartite grass cotyledon (the scutellum

and coleoptile) that arose long after the evolution of the plant cotyledon from ancestral embryonic leaves[10,20]. In contrast, the initiating scutellum at the Transition stage uses highly conserved genes, corresponding to the phylotypic trough (Fig. 5c). These data suggest that once the scutellum first appeared, the expression of highly conserved genes during scutellum initiation ensured that this morphological novel, embryonic organ was retained in subsequent generations. Relatedly, the slight decrease in TDI value during the Coleoptilar stage is attributed to changes in gene expression level in the initiating coleoptile, which may ensure that the bipartite components of the single maize cotyledon are morphologically distinct.

Historically, the morphological and transcriptomic phylotypic stages are described as occurring during mid-embryogenesis in plants and animals[4, 5, 6, 7], although this chronological staging is especially confounding in plants where embryo-morphogenic outcomes can vary dramatically among taxa[7]. The phylotypic peak for transcriptomic age (TAI) in maize occurs during Stage 1, a mid-embryogenic stage when the scutellum and coleoptile are young primordia and the first foliar leaf is just initiating (Fig. 1d). Similarly in Arabidopsis, the transcriptomic phylotypic hourglass began its decent during histological patterning at the globular stage, and peaked during the torpedo stage when the two cotyledon primordia are expanding[7]. Likewise in Drosophila embryogenesis[5, 6],gene age (TAI) peaked during germband elongation[5], at the onset of embryo segmentation and initiation of the organogenic imaginal discs[47]. Notably, the trend toward usage of older gene transcripts in Drosophila also began earlier in embryogenesis, during gastrulation and the consummate formation of tissue layers [5]. In this way, the phylotypic period of *Drosophila melanogaster* also resembles an extended hourglass. A similar pattern is observed in zebrafish (*Danio rerio*), wherein the phylotypic hourglass begins its descent toward older gene

transcripts during gastrulation and then peaks in the subsequent late segmentation/early pharyngula stages, when organ primordia initiate and grow[5]. A later reduction in relative transcriptomic gene age is noted during the zebrafish larval stage, which is accompanied by additional morphogenesis. Taken in summary, the phylotypic period in plants and animals commences during tissue layer organization and extends to include organogenesis and initial growth of embryonic organs.

Plants and animals evolved before the advent of multicellularity. Notably and perhaps surprisingly, the earliest steps in animal and plant embryogenesis, after fertilization and during the establishment of apical basal polarity, are not conserved within taxa and utilize new and divergent genes to form an embryo proper (Fig. 5) [5, 6, 7]. We suggest that embryo histogenesis and organogenesis correlates with asymmetric cell divisions that generate morphology during the phylotypic period. In contrast, after apical-basal polarization, the initial cell divisions that generate the globular embryo proper are primarily symmetrical and non-phylotypic. Despite their evolutionary distances, we demonstrate that in maize, as in Arabidopsis, fruit flies, and fish, the processes of embryonic histogenesis and initial organogenesis utilize old and conserved genes, comprising anciently evolved, developmental genetic networks that help form the fundamental body plan. At the end of this phylotypic period, plant and animal organs each express newer and divergent genes in species-specific ways, to generate the vast morphological diversity found in nature.

References

1. Haeckel E. *Generelle Morphologie der Organismen. Allgemeine Grundzüge der Organischen Formen-Wissenschaft, Mechanisch begründet durch die von Charles Darwin reformierte Descendenz-Theorie, vols I and II*. Verlag Georg Reimer, Berlin (1866).

2. Meckel JF. *Entwurf einer Darstellung der zwischen dem Embryozustande der höheren Tiere und dem permanenten der niederen stattfindenden Parallele: Beyträge zur vergleichenden Anatomie*. Carl Heinrich Reclam, Leipzig (1811).

3. Serrés ERA. Théorie des formations organiques, ou rescherches d'anatomie transcendante sur les lois de l'organogénie, appliquée à l'anatomie pathologique. *Annales de la Societe des Sciences Naturelles et d'Archeologie de Toulon et du Var* **21**, 5-49 (1827).

4. Sander K. *In Development and Evolution: the sixth Symposium of the British Society for Developmental Biology.* . Cambridge University Press. (1983).

5. Domazet-Loso T, Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815-818 (2010).

6. Kalinka AT, *et al.* Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811-814 (2010).

7. Quint M, Drost HG, Gabel A, Ullrich KK, Bonn M, Grosse I. A transcriptomic hourglass in plant embryogenesis. *Nature* **490**, 98-101 (2012).

8. Soltis PS, Soltis DE, Chase MW, Endress PK, Crane PR. The diversification of flowering plants. In: *Assembling the tree of life,* (eds Cracraft J, Michael J D). Oxford University Press (2004).

9. Dean E. Upcoming changes in flowering plant family names: those pesky taxonomists are at it again! *Fremontia* **30**, 3-12 (2002).

10. Chandler JW. Cotyledon organogenesis. *J Exp Bot* **59**, 2917-2931 (2008).

11. Kaplan DR, Cooke TJ. Fundamental concepts in the embryogenesis of dicotyledons: a morphological interpretation of embryo mutants. *Plant Cell* **9**, 1903–1919 (1997).

12. Kaplan DR. The monocotyledons: their evolution and comparative biology vii. the problem of leaf morphology and evolution in the monocotyledons. *The Quarterly Review of Biology* **48**, 437-457 (1973).

13. Chen M, *et al.* Comparative analysis of embryo proper and suspensor transcriptomes in plant embryos with different morphologies. *Proc Natl Acad Sci U S A* **118**, (2021).

14. Randolph LF. Developmental morphology of the caryopsis in maize. *J Agric Res* **53**, 881-916 (1936).

15. Abbe EC, Stein OL. The growth of the shoot apex in maize: embryogeny. *American Journal of Botany* **41**, 285-293 (1954).

16. Bommert P, Werr W. Gene expression patterns in the maize caryopsis: clues to decisions in embryo and endosperm development. *Gene* **271**, 131-142 (2001).

17. Nardmann J, Werr W. Patterning of the maize embryo and the perspective of evolutionary developmental biology. In: *Handbook of Maize: Its Biology* (eds Jeff L. Bennetzen, Hake SC). Springer (2009).

18. Weatherwax P. Position of Scutellum and Homology of Coleoptile in Maize. *Botanical Gazette* **69**, 179-182 (1920).

19. Boyd L. Evolution in the monocotyledonous seedling: a new interpretation of the morphology of the grass embryo. *Transactions of the Botanical Society of Edinburgh* **30**, 286-303 (1931).

20. Kaplan DR. Early plant development: From seed to seedling to established plant. In: *Principles of Plant Morphology*). Copy Central (1996).

21. Satterlee JW, *et al.* A Wox3-patterning module organizes planar growth in grass leaves and ligules. *Nat Plants* **9**, 720-732 (2023).

22. Kaplan DR. The concept of homology and its central role in the elucidation of plant systematic relationships. In: *Cladistics: Perspectives on the Reconstruction of Evolutionary History* (eds Duncan T, Stuessy TF). First Edition edn. Columbia University Press (1984).

23. Xu X, *et al.* Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev Cell* **56**, 557-568 e556 (2021).

24. Takacs EM, *et al.* Ontogeny of the maize shoot apical meristem. *Plant Cell* **24**, 3219-3234 (2012).

25. Poethig RS, Szymkowiak EJ. Clonal analysis of leaf development in maize. *Maydica* **40**, 67–76 (1995).

26. Smith LG, Jackson D, Hake S. Expression of knotted1 marks shoot meristem formation during maize embryogenesis. *Developmental Genetics* **16**, 344-348 (1995).

27. Vollbrecht E, Reiser L, Hake S. Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, knotted1. *Development* **127**, 3161-3172 (2000).

28. Nardmann J, Ji J, Werr W, Scanlon MJ. The maize duplicate genes narrow sheath1 and narrow sheath2 encode a conserved homeobox gene function in a lateral domain of shoot apical meristems. *Development* **131**, 2827-2839 (2004).

29. Johnston R, Wang M, Sun Q, Sylvester AW, Hake S, Scanlon MJ. Transcriptomic analyses indicate that maize ligule development recapitulates gene expression patterns that occur during lateral organ initiation. *Plant Cell* **26**, 4718-4732 (2014).

30. Strable J, *et al.* Maize YABBY genes drooping leaf1 and drooping leaf2 regulate plant architecture. *Plant Cell* **29**, 1622-1641 (2017).

31. Toriba T, *et al.* BLADE-ON-PETIOLE genes temporally and developmentally regulate the sheath to blade ratio of rice leaves. *Nat Commun* **10**, 619 (2019).

32. Evans MM. The indeterminate gametophyte1 gene of maize encodes a LOB domain protein required for embryo Sac and leaf development. *Plant Cell* **19**, 46-62 (2007).

33. Leiboff S, Strable J, Johnston R, Federici S, Sylvester AW, Scanlon MJ. Network analyses identify a transcriptomic proximodistal prepattern in the maize leaf primordium. *New Phytol* **230**, 218-227 (2021).

34. Fukushima K, Hasebe M. Adaxial-abaxial polarity: the developmental basis of leaf shape diversity. *Genesis* **52**, 1-18 (2014).

35. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

36. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* **1**, 54 (2007).

37. Knauer S, *et al.* A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits. *Genome Res* **29**, 1962-1973 (2019).

38. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep Methods* **3**, 100498 (2023).

39. Hofmann F, Schon MA, Nodine MD. The embryonic transcriptome of Arabidopsis thaliana. *Plant Reprod* **32**, 77-91 (2019).

40. Spurr AR. Histogenesisand organizationof the embryo in Pinusstrobus L. . *Am J Bot* **36**, 629-641 (1949).

41. Kiesselbach TA. The structure and reproduction of corn. *Uni of Nebraska Agric Stat Res Bull* **161**, 1-96 (1949).

42. Zimmermann R, Werr W. Transcription of the putative maize orthologue of the Arabidopsis DORNROSCHEN gene marks early asymmetry in the proembryo and during leaf initiation in the shoot apical meristem. *Gene Expr Patterns* **7**, 158-164 (2007).

43. Poethig RS. Cellular parameters of leaf morphogenesis in maize and tobacco. In: *Contemporary problems in plant anatomy* (eds White R, Dickison W). Academic Press (1984).

44. Scanlon MJ. Developmental complexities of simple leaves. *Curr Opin Plant Biol* **3**, 31-36 (2000).

45. Scanlon MJ. The polar auxin transport inhibitor N-1-naphthylphthalamic acid disrupts leaf initiation, KNOX protein regulation, and formation of leaf margins in maize. *Plant Physiol* **133**, 597-605 (2003).

46. Richardson AE, *et al.* Evolution of the grass leaf by primordium extension and petiole-lamina remodeling. *Science* **374**, 1377-1381 (2021).

47. Gilbert SF. *Developmental Biology*, 6th edn. Sunderland (MA): Sinauer Associates (2000).

48. Ciceri P, Locatelli F, Genga A, Viotti A, Schmidt RJ. The activity of the maize Opaque2 transcriptional activator is regulated diurnally. *Plant Physiol* **121**, 1321-1328 (1999).

Figure Legends

Fig. 1 Multiplexed spatial transcriptomics of the maize Stage 1 embryo.

Differential interference contrast (DIC) microscopy of maize embryos at the Proembryo stage (**a**), Transition stage (**b**) Coleoptilar stage, (**c**) and Stage 1 (**d**). Immunohistological analyses of

the Class I KNOX proteins mark the shoot apical meristem (SAM). ep, embryo proper; en, endosperm; su, suspensor; sc, scutellum; c, coleoptile; L, leaf; *, SAM. Scale bar: 50 μm (**a** and **b**) and 100 μm (**c** and **d**). **e**. Uniform Manifold Approximation and Projection (UMAP) of Stage1 embryo single-cell transcriptomic profiling with cell/tissue types assigned for each cell cluster.

Fig. 2 Spatial transcriptomics of maize Stage 1 embryo

Longitudinal cryosections of Stage 1 embryos embedded in OCT (optimal cutting temperature compound; **a-b**) or in paraffin (**c-d)** for 10x Visium^{TM} spatial transcriptomics analysis. **a, c** Toluidine Blue-stained sections for 10x Visium^{TM} spatial transcriptomics assay are overlayed with the10x Visium^{TM} spatial map (**b, d**). The various color-coded circles correspond to the capture spots where spatial cDNA libraries were generated. Scale bar = 100 μm. **e.** Longitudinal sections of paraffin-embedded Stage 1 embryo for use in laser microdissection RNA-seq analysis (LM-RNAseq). **f.** Enlarged view of a Stage 1 embryo showing regions microdissected for LM-RNAseq). Figure labels: scu, scutellum; uc, upper coleoptile; lc, lower coleoptile. **e, f** Scale bar=100 μm. **g**. UMAP of the Visium^{TM} spatial transcriptomic analysis of Stage 1 embryos cell clusters. The spatial, embryonic origins of each cluster are indicated by the color-coded embryo cartoon; numbers correspond to the cluster designations on the UMAP. **h**. Heatmap illustrating the correlation between the single-cell (sc) clusters (y-axis) and Visium^{TM} spatial clusters (x=axis). The significance level is marked by $-\log_{10}p$. **i**. Expression of candidate genes identified in microdissected embryonic organs. Cell clusters exhibiting elevated candidate gene expression of the specific markers are indicated; expression values are $\log_2$ transformed, i.e. $\log_2(TPM+1)$.

Fig. 3 RNA-targeting analyses of Transition-staged, Coleoptilar-staged, and Stage 1 embryos

a-c. Signal point map of *KN1*, *CUC2*, *NS1*, *WOX3A* and *DRL1* at the Transition stage (**a**), Coleoptilar stage (**b**) and Stage 1 (**c**). **d-h.** Expression density map of *KN1* (**d**), *CUC2* (**e**), *NS1* (**f**), *WOX3A* (**g**) and *DRL1* (**h**) at Stage 1. **i-k, l-n, o-q** and **u-w**. Signal point map of *BOPa* (**i-k**), *YAB14* (**l-n**), *IG1* (**o-q**), and *ARF3* (**u-w**) at the Transition stage, Coleoptilar stage and Stage 1. **r-t** and **x-z**. Expression density map of *IG1* (**r-t**) and *ARF3* (**x-z**) at the corresponding stages. Figure labels: T, Transition stage; C, Coleoptilar stage; L: Stage 1. The point colors are arbitrarily determined. Yellow asterisks mark the adaxial expression of *IG1* and *ARF3*, and red asterisks mark the abaxial expression of *ARF3*. The green asterisk in panel **k** marks the expression maximum of *BOPa* at the emerging leaf sheath. The yellow asterisks in panel **r-t** and **z** mark the expression maxima of *IG1* at the adaxial side of the scutellum (**r**), the coleoptile (**s**), the coleoptile and the leaf (**t**), as well as the expression maxima of *ARF3* at the adaxial side of the scutellum (**z**), respectively. The red asterisks in panel **x-z** mark the abaxial expression of *ARF3* at the scutellum (**x, y**), the coleoptile (**y, z**) and the leaf (**z**). Scale bar for all panels = 50 μm.

Fig. 4 Dynamics of the maize embryo organ initiation network

a. Venn diagram identifies 130 genes in the embryo-organ-initiation-network. **b**. Normalized expression of these 130 genes at the Proembryo stage (PE), Transition stage (T), Coleoptilar stage (C), Stage 1 scutellum (Scu), Stage 1 upper coleoptile (UC), Stage 1 lower coleoptile (LC), Stage 1 leaf (L), as well as P0 and P1 of the 14 day-after-germination seedling. The letters above the boxplot represent significance levels (non-overlapping letters indicate significant differences

at $p<0.05$) by Tukey-Kramer HSD test. **c**. Heatmap of the GO terms of the putative initiation network, as well as the single-cell (sc) cluster of the Stage 1 scutellum, coleoptile and leaf. The heatmap color is determined by $-\log_{10}$FDR (false discovery rate). **d.** WGCNA tree graph of the transcriptomic relationships among the co-expression modules specific to corresponding maize organs. The height (y-axis) scales the length of organ-specific branches, indicating the transcriptomic distance between the modules. **e**. Heatmap illustrating the correlation between the single-cell (sc) clusters (y-axis) and hdWGCNA modules (x=axis). The significance level is marked by $-\log_{10}p$. **f**. Heatmap illustrating the dynamic correlation of the initiation network expression between Maize (y-axis) and Arabidopsis (x-axis) across the embryogenesis. The heatmap color is determined by correlation coefficient (r) via Pearson correlation test with the significance level $p<0.05$ (*) or $p<0.01$ (**). Scu, Scutellum; Col, coleoptile; hypoc, hypocotyl.

Fig. 5 Phylotypic stage analysis of the maize embryogenesis

a. Phylostratigraphic map of maize. All maize protein-coding transcripts were assigned to 13 phylostrata (PS1-PS13), Numbers in parenthesis refer to the number of genes per phylostratum, and the taxonomic groupings are marked above the corresponding PS. **b-c**. The transcriptome age index (TAI; **b**) and transcriptome divergence index profiles (TDI; **c**) of maize embryogenesis and seedling leaves. Figure labels: PE, proembryo stage; T, transition stage; C, coleoptilar stage; L, Stage 1 leaf; Scu, Stage 1 scutellum; Col, Stage 1 coleoptile; P0 and P1, Plastochron 0 and 1 leaf of the 14-day after planted seedlings. The gray lines represent the standard deviation estimated by permutation analysis. The overall patterns of TAI and TDI profiles are significant, as measured by flat line test ($P_{TAI}=3.56\times10^{-6}$, $P_{TDI}=4.80\times10^{-4}$), reductive hourglass test ($P_{TAI}=1.49\times10^{-4}$, $P_{TDI}=1.53\times10^{-3}$), and reductive early conservation test ($P_{TAI}=1$, $P_{TDI}=0.998$).

Fig. 6 Model for the homology of the grass cotyledon

The model describes that the single maize cotyledon as composed of two domains, the apical scutellum (homologous to the foliar leaf blade; dashed arrow line) and the basal coleoptile (homologous to the foliar leaf sheath; solid arrow line)

Methods

Plant Materials and growth conditions

B73 inbred lines, the *ns* 1:1 lines (segregate 1/2 wild type NS1+/*ns1 ns2/ns2* plants and 1/2 *ns1/ns1 ns2/ns2* double mutant[1], and *log7*/+ (obtained from the Maize Genetics COOP Stock Center: UFMu-02863 mu1030680::Mu) were grown in the Gutermann greenhouse facility, Cornell University (Ithaca, NY). Transition staged, coleoptilar staged, Leaf1 staged and Leaf2 staged embryos were collected at 8-9 DAP, 10-11 DAP, 12-13 DAP, 14-15 DAP, respectively. *ns* 1:1 seeds were planted in soil consisting of a 1:1 mixture of Turface and LM111 in Percival A100 growth chamber at Weill Hall, Cornell University (Ithaca, NY) with the set condition: day 29.4 °C/16 hours, night 23.9 °C/8 hours, humidity 50%. The seedlings were collected on the 7th day after planting for coleoptile comparisons between WT and *ns1;2* mutants.

Single-cell transcriptomic analysis

Two batches of B73 Leaf1 staged embryos were collected in summer 2021 and winter 2022. For each batch, 200 embryos were dissected from developing seeds for protoplast isolation as previously described[2]. Protoplast (1000-2000 cells per µL) were loaded into the 10X Genomics

Chromium platform to generate a Gel Bead-In-Emulsion (GEM). The GEM-captured RNA was reverse transcribed and amplified to make 3′ cDNA libraries, followed by the high throughput RNA sequencing on the Illumina NextSeq 2000 sequencer at the Cornell University Biotechnology Resource Center.

The resulting raw base call (BCL) files were demultiplexed to generate FASTQ files via CellRanger mkfastq v6.0 (10x Genomics, Pleasanton, CA). FASTQ reads were trimmed, aligned (to B73 reference genome V3[3]) and assigned to cell barcodes to generate expression count matrices via CellRanger count with default settings. The expression count matrices were then analyzed using Seurat v4[4, 5, 6]. In detail, the matrices were filtered with the setting (number of genes > 1500, UMI > 6000) to remove low quality cells. The filtered matrices were normalized and log transformed via SCTransform function. The normalized matrices were merged and batch effects were removed via R package Harmony v1.2.0[7]. The cell-cycle associated genes were scored and regressed via CellCycleScoring and ScaleData function. Matrix dimension was performed using RunPCA. The cell clusters and the Uniform Manifold Approximation and Projection (UMAP) plot were generated with the setting (dim = 1:40, resolution = 1, n. neighbors = 40, min. dist = 0.01, spread = 3). The differential expressed genes and cell cluster markers were identified via FindAllMarkers function with the log2 fold change threshold = 0.2.

Spatial transcriptomic analysis

For use in spatial transcriptomics, B73 Leaf1 staged embryos were processed for either paraffin-embedding and microtome sectioning, or cryoembedding and cryosectioning. For paraffin -embedding, two embryos were fixed in Farmer's Fixative (Ethanol:Acetic acid = 3:1), followed

by dehydration in an ethanol concentration gradient (75%, 85%, 95%, 100%) and clearing with Ethanol:Xylene gradient of 3:1, 1:1, 1:3, and 100% Xylene at room temperature. Cleared embryos were embedded in Paraplast Plus® (Leica Biosystems, Deer Park, IL). The embedded blocks were microtome-sectioned at 10 um at room temperature; sections were subsequently mounted onto 10x Visium™ Spatial Gene Expression Slides (10x Genomics, Pleasanton, CA). Slides were deparaffinized with two changes of 100% Xylene and 95% Ethanol for 10 min and 2 min, respectively. For cryoembedding, two kernels were dissected longitudinally to remove left and right margin of the kernel; the central region containing the embryo was embedded in Tissue-Tek® Optimal Cutting Temperature (O.C.T., Sakura Finetek, Torrance, CA), and snap-frozen in the dry ice. Frozen tissue blocks were cryosectioned at10 µm, and the sections were mounted on 10x Visium™ Spatial Gene Expression Slide (10x Genomics, Pleasanton, CA).

The 10x Visium™ Spatial Gene Expression Slides were processed for spatial library construction following the manufacturer instructions. Briefly, the sections were baked at 37 ^{o}C for 1 min followed by re-fixation in 100% methanol at -20 ^{o}C for 30 min. Subsequently, the sections were stained with Toluidine Blue for histological imaging. Tissue permeabilization was performed at 37 ^{o}C for 6 min to release RNA from the fixed tissue, followed by first strand cDNA synthesis. Subsequently, second strand cDNA was synthesized and amplified for library construction and sequencing. High-throughput RNA sequencing was conducted on an Illumina NextSeq 2000 sequencer at the Cornell University Biotechnology Resource Center.

Raw reads were demultiplexed, mapped to the Toluidine Blue stained images, and aligned to the B73 reference genome V3[3] to generate spatial expression count matrices using SpaceRanger v2.0

(10x Genomics, Pleasanton, CA). The matrices were analyzed using Seurat v4[4, 5, 6]. In detail, the matrices were normalized and log transformed via SCTransform function. The normalized matrices were merged followed by dimensionality reduction. Cell clusters and the UMAP were generated with the following settings (dim = 1:40; resolution = 1.5; neighbors = 10; minimum distance =0.01). The spatial data were visualized with DimPlot and SpatialPlot; differential expressed genes and cell cluster markers were identified via FindAllMarkers function with the log2 fold change threshold = 0.1.

Laser-microdissection RNAseq analysis

Sixty maize Leaf1 staged kernels in two replicates were dissected longitudinally to remove the left and right margins of the kernels. This embryo-containing central region was fixed in Farmer's Fixative (Ethanol:Acetic acid=3:1). After fixation, the samples were dehydrated, cleared, embedded and sectioned as described previously[8, 9] . Tissue sections were mounted on PEN membrane slides (Leica Microsystems, Deer Park, IL) pre-treated at 180 ^{o}C for 4 hours followed by UV irradiation at 254 nm for 30 minutes. The sections were deparaffinized with two changes of 100% Xylene and 100% Ethanol for 10 min and 2 min, respectively. Laser microdissection was performed using a PALM Laser Microbeam System (Zeiss Group, Oberkochen, Germany). Adaxial scutellum, SAM, leaf, as well as upper and lower coleoptiles were microdissected (total area for each captured organ is over 0.3 mm^2) and RNA extraction was performed using the Arcturus^{TM} PicoPure^{TM} RNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA). The extracted RNA was amplified using Arcturus^{TM} RiboAmp^{TM} HS PLUS RNA Amplification Kit (Thermo Fisher Scientific, Waltham, MA). RNA quality control, library construction and sequencing were performed at the Cornell University Biotechnology Resource

Center. The raw reads quality control was conducted via FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw reads were mapped to the. B73 reference genome V3[3] by HISAT2 (v2.1.0)[10, 11]. Read counts and the normalized number of transcripts (transcript per million reads, TPM) were calculated as previously described[12]. DEseq2 (v1.26.0)[13] was used to identify differentially expressed genes (DEGs) with the cutoff criteria of log2 fold change >1 or <1 with the false discovery rate (FDR) < 0.05.

Multiplex *in situ* hybridization

Transition staged, coleoptilar staged and Leaf1 staged kernels (two replicate for each stage) were dissected longitudinally to remove the left and right kernel margins. The central region containing the embryo was fixed in 4% paraformaldehyde (PFA); fixed samples were dehydrated, cleared, embedded and sectioned as described previously[14, 15]. Sections were mounted on 10x Genomics Xenium™ slides, deparaffinized and decrosslinked following the manufacturer's instructions. The Xenium custom gene expression panel containing probes of 99 maize genes were applied on each section to perform hybridization, ligation and rolling circle amplification, followed by cycles of fluorescent probe hybridization, imaging and decoding. The data was visualized by Xenium Explorer v1.3. The Xenium™ multiplex *in situ* hybridization assays were conducted at Memorial Sloan Kettering Cancer Center (New York, NY).

Histology and immunohistolocalizations

Developing kernels from the proembryo stage, transition stage, coleoptilar stage, and the Stage 1 were fixed in FAA, dehydrated, cleared, embedded in paraplast, sectioned and stained as described previously[15]. Immunohistolocalizations were performed as described[1].

WGCNA and hdWGCNA

For WGCNA to pursue to initiation network, log2-transformed TPM data from LM-RNAseq (SAM data was excluded) were filtered following these criteria: row mean > 0.2 and coefficient of variation > 0.4. The general pipeline was performed as per R package instructions[16] with minor modifications as follows. The network type was set to “signed”. The soft power, the minimum module size ,and the merging threshold were set to 13, 20 and 0.3, respectively. For the analysis to pursue transcriptomic relationship between organs, the LM-RNAseq data in this study was merged with the LM-RNAseq data of the 14-day seedling P0 and P1 data[17]. The data was normalized and batched effect removed as described, and filtered following the same criteria as the WGCNA for the initiation network. The network type was set to “signed”. The soft power, the minimum module size ,and the merging threshold were set to 16, 20 and 0.4, respectively.

For hdWGCNA, the single-cell RNAseq data was used to perform the analysis. The data was SCTransformed and normalized. The general pipeline was conducted as the R package instructions[18] with the following minor modifications. The network type was set to “signed”, and the soft power and minimum module size were set to 10 and 30, respectively. Hypergeometric test was performed to analyze the correlation between the hdWGCNA modules and the single-cell clusters following the previously described procedures[19].

Gene Ontology (GO) enrichment analysis

GO enrichment analysis of the putative initiation network as well as the single-cell scutellum, coleoptile and leaf clusters was performed by AgriGOv2

(http://systemsbiology.cau.edu.cn/agriGOv2/) with the locus ID v3.30 (Gramene Release 50) datasets as GO annotation background. The significant enrichment cutoff FDR was set to 0.05.

Statistical analysis

Hypergeometric tests were performed to analyze the correlation between single-cell clusters and spatial transcriptomic clusters, as well as between the single-cell clusters and hdWGCNA modules. We followed previously algorithms[19] to statistically evaluate if the two groups of data were significantly correlated by calculating the number of overlapping gene relative to the total number of protein-coding genes in the maize B73 reference genome V3[3]. The p values were transformed by –log10p, and visualized as heatmap built by the R package gplots (https://github.com/cran/gplots).

LM-RNAseq read count data were normalized to fit the generalized linear model (GLM), and Empirical Bater shrinkage estimation was used for calculating dispersions, expression fold changes and adjust p values. In addition, Pearson correlation test was used to calculate the correlation between the coexpression modules and the corresponding clusters (microdissected organs or putative initiation network) in WGCNA, and between the maize and Arabidopsis embryogenesis genes.

To statistically compare the expression of the initiation network between the transition, coleoptilar and the stage1, the TPM data was log2 transformed [log2(TPM+1)] and compared using Tukey-Kramer HSD test by JMP Pro 17 (SAS Institute, Cary, NC) with default settings.

For the correlation test, the transition, coleoptilar and the stage1 TPM data was double log2 transformed, i.e. log2[log2(TPM+1)+1], to fix the data skewness caused by the outliers. The Arabidopsis TPM data was log2 transformed. The Pearson correlation was performed in R v4.3.2 with the significance cutoff value $p<0.05$.

Calculating TAI and TDI for phylotypic period

The procedures of constructing a phylostratigraphic map, calculating TAI and TDI, and statistical tests, have been presented previously[20, 21, 22, 23]. For each of the 13 phylostrata shown in Fig. 5a, the amino acid sequences of all 350 species with completely sequenced genomes were extracted from NCBI, Ensembl genomes, Treegene (https://treegenesdb.org), Fernbase (https://fernbase.org), or Maizegdb (https://maizegdb.org). A database of these sequences was generated, and each of the 63,031 amino acid sequences of *Zea mays* ssp. *mays* (v3) with a minimum length of 30 amino acids was blasted against this database using BLASTp (BLAST version 2.15.0) with an E-value cut-off of 10^{-5}. The corresponding gene of *Zea mays* ssp. *mays* was assigned to the phylogenetically most distant (oldest) phylostratum containing at least one species with at least one blast hit, otherwise the corresponding gene of *Zea mays* ssp. *mays* was assigned to phylostratum 13 (PS13) where no blast hit was identified.

Orthologous gene pairs of *Zea mays* ssp. *mays* and *Zea mays* ssp. *mexicana* were determined with the method of best hits using BLASTp. Amino acid sequence alignments of each pair generated with MAFFT[24] (L-INS-i option) were used for codon alignments generated with PAL2NAL[25] to compute sequence divergence levels (K_a/K_s) with KaKs Calculator (version 3.0)[26]. Gene pairs with $K_a<0.5$, $K_s<5$ and K_a/K_s ratios < 2 were retained.

Expression levels of maize development stages were collected from 4 data sources 1) LM-RNAseq data of proembryo stage[27], 2) organ initiation network (this study), 3) RNAseq data of Stage 2 embryo[28] and 4) LM-RNAseq data of 14-day after germination seedling P0 and P1 leaves[17]. The datasets were processed with batch effect removal and normalization as described above. The TAI and the TDI are weighted means of evolutionary age and sequence divergence, respectively, and defined previously[20, 22]. The transcriptome age index TAI_s and the transcriptome divergence index TDI_s of developmental stage s (s = proembryo, transition, coleoptilar, stage 1 leaf, stage 1 scutellum, stage 1 coleoptile, seedling leaves P0 and seedling leaves P1) are the weighted mean of the evolutionary age (phylostratum) weighted by the expression level of each gene at developmental stage s, and the weighted mean of the Ka/Ks ratio of each gene weighted by the expression level of each gene at developmental stage *s*, respectively. Low/high PS values correspond to evolutionarily old/young genes, so low/high TAI values correspond to evolutionarily old/young transcriptomes. Low/high K_a/K_s ratios correspond to conserved/divergent genes, so low/ high TDI values correspond to conserved/divergent transcriptomes. For gene models with multiple predicted peptides, TAI was calculated with the most conserved phylostratum assigned to that locus and TDI was calculated with the conserved codon divergence strata assigned to that locus.

To determine the statistical significance of the TAI and TDI profiles, the flat line test[22], reductive hourglass test and reductive early conservation test were performed via myTAI package (v1.0.1)[23]. The flat line test is a permutation test based on the variance of the TAI or TDI values of a given TAI or TDI profile. To reductive hourglass test, the developmental stages were partitioned into three modules, early (proembryo stage), mid (transition stage, coleoptilar stage, Stage 1 leaf, Stage 1 scutellum, and Stage 1 coleoptile), and late (stage 2 and seedling leaves),

and permutation test was performed for the minimum differences of the mean TAI or TDI values between early and mid-modules, and between late and mid modules. To reductive early conservation test, permutation test was performed for the minimum differences of the mean TAI or TDI values between mid and early modules, and between late and early modules. The detailed procedures were described in Drost, et al., 2015[23].

Data accessions

The single-cell RNAseq, Visium™ spatial transcriptomics, and LM-RNAseq data were deposited in the NCBI Sequence Read Archive (SRA) with the BioProject ID: PRJNA1045558. The Xenium™ multiplex *in situ* hybridization data was deposited in Science Data Bank with the link: https://www.scidb.cn/s/F7ZbEz, which could be visualized by Xenium Explorer v1.3. The code for TAI and TDI calculations were uploaded to https://gitlab.com/rz444/maizephylper.git.

Methods references

1. Scanlon MJ, Freeling M. The narrow sheath leaf domain deletion : a genetic tool used to reveal developmental homologies among modified maize organs. *Plant Journal* **13**, 547-561 (1998).

2. Satterlee JW, Scanlon MJ. Protoplast isolation from undifferentiated maize seedling shoot tissue. *Methods Mol Biol* **2464**, 123-130 (2022).

3. Schnable PS, *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112-1115 (2009).

4. Hao Y, *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*, (2023).

5. Stuart T, *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902 e1821 (2019).

6. Hao Y, *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

7. Korsunsky I, *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019).

8. Scanlon MJ, Ohtsu K, Timmermans MC, Schnable PS. Laser microdissection-mediated isolation and in vitro transcriptional amplification of plant RNA. *Curr Protoc Mol Biol* **Chapter 25**, Unit 25A 23 (2009).

9. Zhang S, Thakare D, Yadegari R. Laser-capture microdissection of maize kernel compartments for rna-seq-based expression analysis. *Methods Mol Biol* **1676**, 153-163 (2018).

10. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915 (2019).

11. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357-360 (2015).

12. Wu H, *et al.* NAKED ENDOSPERM1, NAKED ENDOSPERM2, and OPAQUE2 interact to regulate gene networks in maize endosperm development. *Plant Cell*, (2023).

13. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

14. Laureyns R, *et al.* An in situ sequencing approach maps PLASTOCHRON1 at the boundary between indeterminate and determinate cells. *Plant Physiol* **188**, 782-794 (2022).

15. Jackson D. In situ hybridization in plants. In: *Molecular Plant Pathology: A Practical Approach*, (eds D. J. Bowles, S. J. Gurr, McPherson M). Oxford University Press (1991).

16. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

17. Knauer S, *et al.* A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits. *Genome Res* **29**, 1962-1973 (2019).

18. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep Methods* **3**, 100498 (2023).

19. Zhan J, *et al.* RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *The Plant Cell* **27**, 513-531 (2015).

20. Domazet-Loso T, Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815-818 (2010).

21. Domazet-Loso T, Brajkovic J, Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet* **23**, 533-539 (2007).

22. Quint M, Drost HG, Gabel A, Ullrich KK, Bonn M, Grosse I. A transcriptomic hourglass in plant embryogenesis. *Nature* **490**, 98-101 (2012).

23. Drost HG, Gabel A, Grosse I, Quint M. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol Biol Evol* **32**, 1221-1231 (2015).

24. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33**, 511-518 (2005).

25. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609-612 (2006).

26. Zhang Z. KaKs_Calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinformatics* **20**, 536-540 (2022).

27. Takacs EM, *et al.* Ontogeny of the maize shoot apical meristem. *Plant Cell* **24**, 3219-3234 (2012).

28. Chen J, *et al.* Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol* **166**, 252-264 (2014).

Acknowledgements

We thank L. Evans and R. Ragas for helpful discussions of the data and comments on the manuscript. Special thanks to Cornell Genomics Facility for performing single-cell and Visium™ sequencing, to the Cornell Transcriptional Regulation and Expression Facility for RNA sequencing, and the Cornell Guterman Bioclimatic Labs for our plants. Thanks to the Integrated Genomics Operation (IGO) of Memorial Sloan Kettering Cancer Center for performing Xenium assays. H. Wu was supported by a grant from the National Science Foundation (IOS-2016021). R. Zhang is supported by another grant from the National Science Foundation (IOS-2210259).

Author contributions

Author Contributions M.J.S, H.W. and R.Q. conceived the study. M.J.S. supervised the project. H.W and R.Q. generated and analyzed the data. M.J.S, H.W. and R.Q. interpreted the results, and M.J.S, H.W. and R.Q wrote the manuscript.

Ethics declaration

Competing interests

The authors declare no competing interests.

Supplementary information

Supplementary Dataset 1. Single-cell RNAseq cluster marker genes

Supplementary Dataset 2. Spatial transcriptomic cluster markers

Supplementary Dataset 3. Normalized gene expression (TPM) of LM-RNAseq

Supplementary Dataset 4. Normalized gene expression (log2(TPM+1)) of tissue specific classic genes

Supplementary Dataset 5. Xenium™ multiplex in situ hybridization probes and result

Supplementary Dataset 6. WGCNA: modules with gene expression pattern: Scu<UpCol<LoCol<Leaf

Supplementary Dataset 7. Putative initiation network and the log2-transformed TPM in the lateral organs of the transition, coleoptilar, Stage 1 (L1) and 14-day after germination seedlings

Supplementary Dataset 8. Gene ontology term category and corresponding FDR

Supplementary Dataset 9. WGCNA: Cell-type specific modules

Supplementary Dataset 10. hdWGCNA module table

Supplementary Dataset 11. Maize initiation network vs Arabidopsis orthologs

Supplementary Dataset 12. Gene expression for TAI and TDI calculation

Supplementary Dataset 13. Phylostrata of maize genes

Supplementary Dataset 14. *Ka/Ks* ratio of maize genes expressed at the embryogenesis and 14-day seedling

Supplementary Dataset 15. Genomes used for phylogeny reconstruction and phylostratigraphic map.

Supplementary Dataset 16. The gene numbers of each phylostratum in embryo stages.

Corresponding author

Michael J. Scanlon: mjs298@cornell.edu

Hao Wu: haowu2021@gmail.com

Extended data figure legends

Extended Data Fig. 1 Verification of single-cell cluster marker genes.

Left panel: the single-cell cluster marker expression on a UMAP. Purple dots mark cells with corresponding gene expression data. Middle panel: the verification of the single-cell cluster marker genes by Visium™. Red to green spots marks the location of the gene expression with high to low level, respectively. Right panel: the verification of the single-cell cluster marker genes by multiplex *in situ* hybridization. Note: that *GRMZM2G166141* (**k**), *GRMZM2G094353* (**m**) and *GRMZM2G099454* (**p**) lack the multiplex *in situ* hybridization data. Scale bar = 100 μm.

Extended Data Fig. 2 Differential expression between laser microdissected RNAseq samples.

a. Principal component analysis of laser-microdissected RNAseq samples. Each dot represents a laser-microdissected embryonic shoot sample. **b-f**. Pairwise comparison of the gene expression between laser-microdissected RNAseq samples. The number of overlapping, differential-expressed genes (DEGs) is showed in the Venn diagram. The circled numbers represent the genes preferentially expressed in the SAM (**b**), upper coleoptile (**c**), scutellum (**d**), leaf (**e**) and lower coleoptile (**f**).

Extended Data Fig. 3 Multiplex *in situ* hybridization of *TRU1*.

The spatial expression of *TRU1* is showed in signal point map (**a-c**) and density map (**d-f**) from the transition stage (**a and d**), the coleoptilar stage (**b and e**) and the Stage 1 (**c and f**). The red arrow marks the expression of *TRU1* at the emerging leaf sheath. Scale bar = 100 μm.

Extended Data Fig. 4 Expression of the embryo organ initiation network.

a. WGCNA to identify modules with overall highest expression in the Stage 1 leaf (L) and less in the lower coleoptile (LC), with steadily decreased expression in the upper coleoptile (UC) and scutellum (S) (i.e. L > LC > UC > S). The heatmap shows the correlation coefficient (r) and the corresponding p value (in the parentheses) of each module (ME1-12). **b**. The expression of the 130 genes in the initiation network at the transition stage (T), coleoptilar stage (C) and the Stage 1 scutellum (Scu), upper coleoptile (UC), lower coleoptile (LC) and leaf (L). **c-e**. The correlation of the expression of the 130 genes between initiation stage of the scutellum, coleoptile and leaf.

Extended Data Fig. 5 WGCNA Module-organ correlation.

a-b. Correlation between datasets of the 14-day after pollination seedling and the embryo before and after batch effect removal. The correlation of the normalized overall gene expression between the 14 day after germination (DAG) seedling and embryo datasets before (**a**) and after (**b**) batch effect removal. **c**. Heatmap illustrating the correlation between co-expression modules (y-axis) and corresponding organs (x-axis. The colors are determined by the correlation coefficient (r) with the p value (numbers in the parentheses)

Extended Data Fig. 6 Correlation between the maize initiation network genes and the Arabidopsis orthologs.

The correlation of expression between the maize initiation network genes at transition (**a**) and coleoptilar (**b**) stage, and the Arabidopsis orthologs at early heart stage.

Extended Data Fig. 7. TAI and TDI for embryo stages under different conditions

a-c. Distribution of all maize genes and genes in each embryo stage across phylostrata and TAI profiles for different E-value cut-offs (ranging from 1E-2 to 1E-20). This analysis investigates whether and how E-values affect the hourglass pattern. **a**. Variation of the E-value cut-off for the BLASTp analysis demonstrates minimal import on the distribution of maize genes among the phylostrata, indicating an overall pattern independent of the E-values. The y-axis is presented on a logarithmic scale. **b**. Similarly, the variation of the E-value cut-off for the BLASTp analysis does not much severely alter the distribution of maize genes in each embryo stage among the phylostrata. **c**. the hourglass pattern of the TAI profile remains consistent across various E-value cut-offs. **d-e**. Transcriptome divergence indices based on two different genome comparisons: *Zea mays* ssp. mays vs. *Zea mays* ssp. mexicana (**d**); *Zea mays* ssp. mays vs. *Zea diploperennis* (**e**). The overall patterns of TDI profiles are significant, as measured by flat line test ($P_{\text{TDI-Mexicana}}=1.07\times10^{-3}$, $P_{\text{TDI-diploperennis}}=4.80\times10^{-4}$), reductive hourglass test ($P_{\text{TDI Mexicana}}=0.027$, $P_{\text{TDI-diploperennis}}=1.53\times10^{-3}$), and reductive early conservation test ($P_{\text{TDI Mexicana}}=0.68$, $P_{\text{TDI-diploperennis}}=0.998$).

Extended Data Fig. 8. The gene expression in each embryo stage per phylostratum.

a. The stacked bars in percentage of gene numbers of each phylostratum in embryo stages. **b-i.** The relative gene expression (normalized to the maximal level in each embryo stage) of maize embryo genes, including in proembryo stage (**b, c**), in Initiation-Network stages (transition stage, coleoptilar stage, and stage 1 leaf) (**d, e**), in stage 1 (stage 1 leaf, stage 1 scutellum, stage 1 coleoptile) (**f, g**), and after stage 2 and in seedling leaves (**h, i**). The data to after stage 2 is from Chen et al., 2014, and were normalized in data source in (**h**).

Extended Data Table 1 Identity of single-cell clusters.

Single cell clusters	Identity	Evidence
0	Basal scutellum	Fig. 1h; Extended Data Fig. 1a
1	Abaxial or adaxial scutellum epidermis	Fig. 1h; Extended Data Fig. 1b
2	Abaxial scutellum epidermis	Fig. 1h; Extended Data Fig. 1c
3	Apical scutellum	Fig. 1h; Extended Data Fig. 1d
4	Adaxial scutellum epidermis	Fig. 1h; Extended Data Fig. 1e
5	Abaxial scutellum epidermis	Fig. 1h; Extended Data Fig. 1f
6	SAM and hypocotyl	Fig. 1h; Extended Data Fig. 1g; known marker genes[24, 29]
7	Abaxial scutellum epidermis	Fig. 1h; Extended Data Fig. 1f
8	Apical scutellum	Fig. 1h; Extended Data Fig. 1d
9	Possible vasculature	Fig. 1h; Extended Data Fig. 1h
10	Coleoptile epidermis	Fig. 1h,n; Extended Data Fig. 1i
11	Basal scutellum	Fig. 1h; Extended Data Fig. 1a
12	Possible embryo-endosperm boundary region	Fig. 1h; Extended Data Fig. 1j
13	Apical scutellum	Fig. 1h; Extended Data Fig. 1k
14	Abaxial scutellum epidermis	Fig. 1h; Extended Data Fig. 1l
15	Vasculature	Extended Data Fig. 1m; known marker genes[24]
16	Boundary of the lateral organs	Fig. 1h; Extended Data Fig. 1n

17	Coleoptile internal tissue	Fig. 1h,i,j,k,n; Extended Data Fig. 1o
18	Possible Tip of scutellum	Fig. 1h; Extended Data Fig. 1p
19	Apical scutellum or suspensor (residue)	Fig. 1h; Extended Data Fig. 1q
20	Endosperm (residue)	Fig. 1h; known marker genes[48]
21	Leaf	Fig. 1h,n; Extended Data Fig. 1r
22	Suspensor (residue)	Fig. 1h; Extended Data Fig. 1s

Evidence based on previous published marker genes, Visium™ and multiplex *in situ* hybridization (Extended Data Fig. 1)

Extended Data Table 2 Gene expression of selected leaf genes in the transition, coleoptilar and Leaf1 stage in the LM-RNAseq data.

Genes	**Transition**	**Coleoptilar**	**Leaf1**
YAB14	60.85	59.62	82.24
FPF1L	83.16	178.63	103
IG1	33.85	21.49	26.94
NS1	8.37	19.55	10.28

Figures and Extended Data Figures

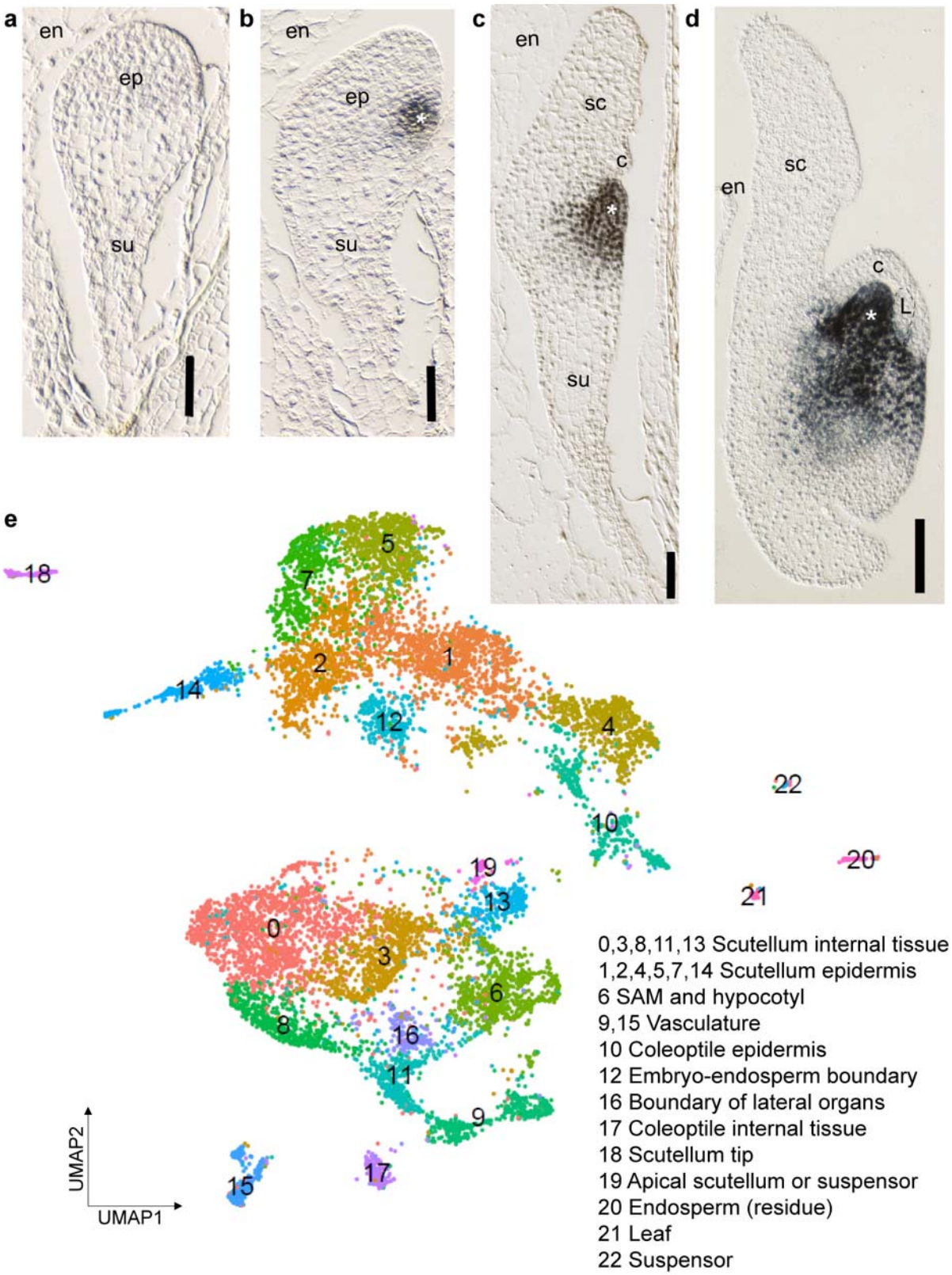
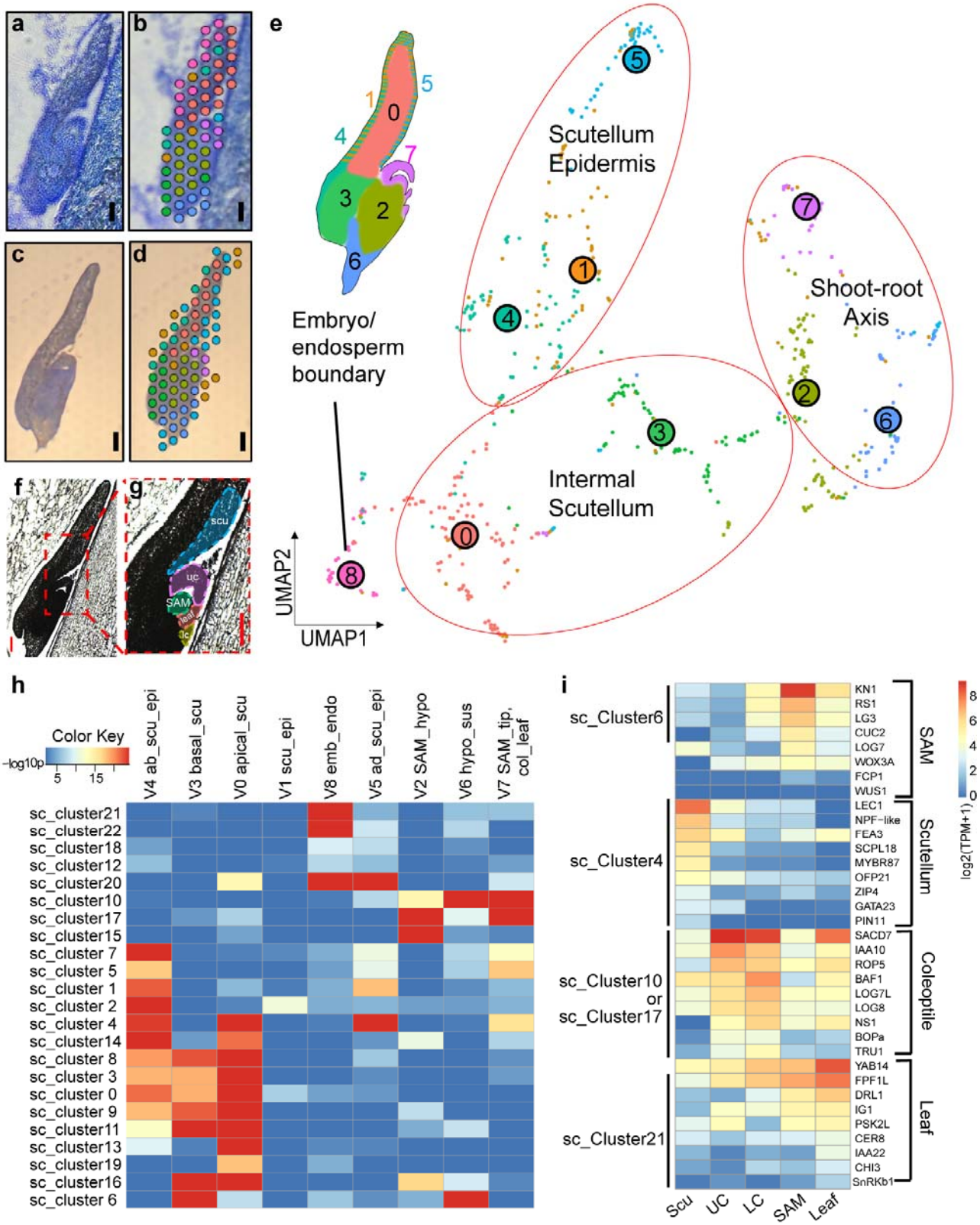


Fig. 1 Multiplexed spatial transcriptomics of the maize Stage 1 embryo.

Differential interference contrast (DIC) microscopy of maize embryos at the Proembryo stage (**a**), Transition stage (**b**) Coleoptilar stage, (**c**) and Stage 1 (**d**). Immunohistological analyses of the Class I KNOX proteins mark the shoot apical meristem (SAM). ep, embryo proper; en, endosperm; su, suspensor; sc, scutellum; c, coleoptile; L, leaf; *, SAM. Scale bar: 50 μm (**a** and **b**) and 100 μm (**c** and **d**). **e**. Uniform Manifold Approximation and Projection (UMAP) of Stage1 embryo single-cell transcriptomic profiling with cell/tissue types assigned for each cell cluster.


a
b
c
d
e
f
g
scu
uc
SAM
lcul
lc
Embryo/
endosperm
boundary
Scutellum
Epidermis
Shoot-root
Axis
Intermal
Scutellum
UMAP2
UMAP1
h
Color Key
-log10p
5
15
V4 ab_scu_epi
V3 basal_scu
V0 apical_scu
V1 scu_epi
V8 emb_endo
V5 ad_scu_epi
V2 SAM_hypo
V6 hypo_sus
V7 SAM_tip, col_leaf
sc_cluster21
sc_cluster22
sc_cluster18
sc_cluster12
sc_cluster20
sc_cluster10
sc_cluster17
sc_cluster15
sc_cluster 7
sc_cluster 5
sc_cluster 1
sc_cluster 2
sc_cluster 4
sc_cluster14
sc_cluster 8
sc_cluster 3
sc_cluster 0
sc_cluster 9
sc_cluster11
sc_cluster13
sc_cluster19
sc_cluster16
sc_cluster 6
i
sc_Cluster6
sc_Cluster4
sc_Cluster10
or
sc_Cluster17
sc_Cluster21
KN1
RS1
LG3
CUC2
LOG7
WOX3A
FCP1
WUS1
LEC1
NPF-like
FEA3
SCPL18
MYBR87
OFP21
ZIP4
GATA23
PIN11
SACD7
IAA10
ROP5
BAF1
LOG7L
LOG8
NS1
BOPa
TRU1
YAB14
FPF1L
DRL1
IG1
PSK2L
CER8
IAA22
CHI3
SnRKb1
SAM
Scutellum
Coleoptile
Leaf
log2(TPM+1)
Scu
UC
LC
SAM
Leaf

1043

1044

Fig. 2 Spatial transcriptomics of maize Stage 1 embryo

Longitudinal cryosections of Stage 1 embryos embedded in OCT (optimal cutting temperature compound; **a-b**) or in paraffin (**c-d)** for 10x Visium^{TM} spatial transcriptomics analysis. **a, c** Toluidine Blue-stained sections for 10x Visium^{TM} spatial transcriptomics assay are overlayed with the10x Visium^{TM} spatial map (**b, d**). The various color-coded circles correspond to the capture spots where spatial cDNA libraries were generated. Scale bar = 100 μm. **e.** Longitudinal sections of paraffin-embedded Stage 1 embryo for use in laser microdissection RNA-seq analysis (LM-RNAseq). **f.** Enlarged view of a Stage 1 embryo showing regions microdissected for LM-RNAseq). Figure labels: scu, scutellum; uc, upper coleoptile; lc, lower coleoptile. **e, f** Scale bar=100 μm. **g**. UMAP of the Visium^{TM} spatial transcriptomic analysis of Stage 1 embryos cell clusters. The spatial, embryonic origins of each cluster are indicated by the color-coded embryo cartoon; numbers correspond to the cluster designations on the UMAP. **h**. Heatmap illustrating the correlation between the single-cell (sc) clusters (y-axis) and Visium^{TM} spatial clusters (x=axis). The significance level is marked by $-\log_{10}p$. **i**. Expression of candidate genes identified in microdissected embryonic organs. Cell clusters exhibiting elevated candidate gene expression of the specific markers are indicated; expression values are $\log_2$ transformed, i.e. $\log_2(TPM+1)$.

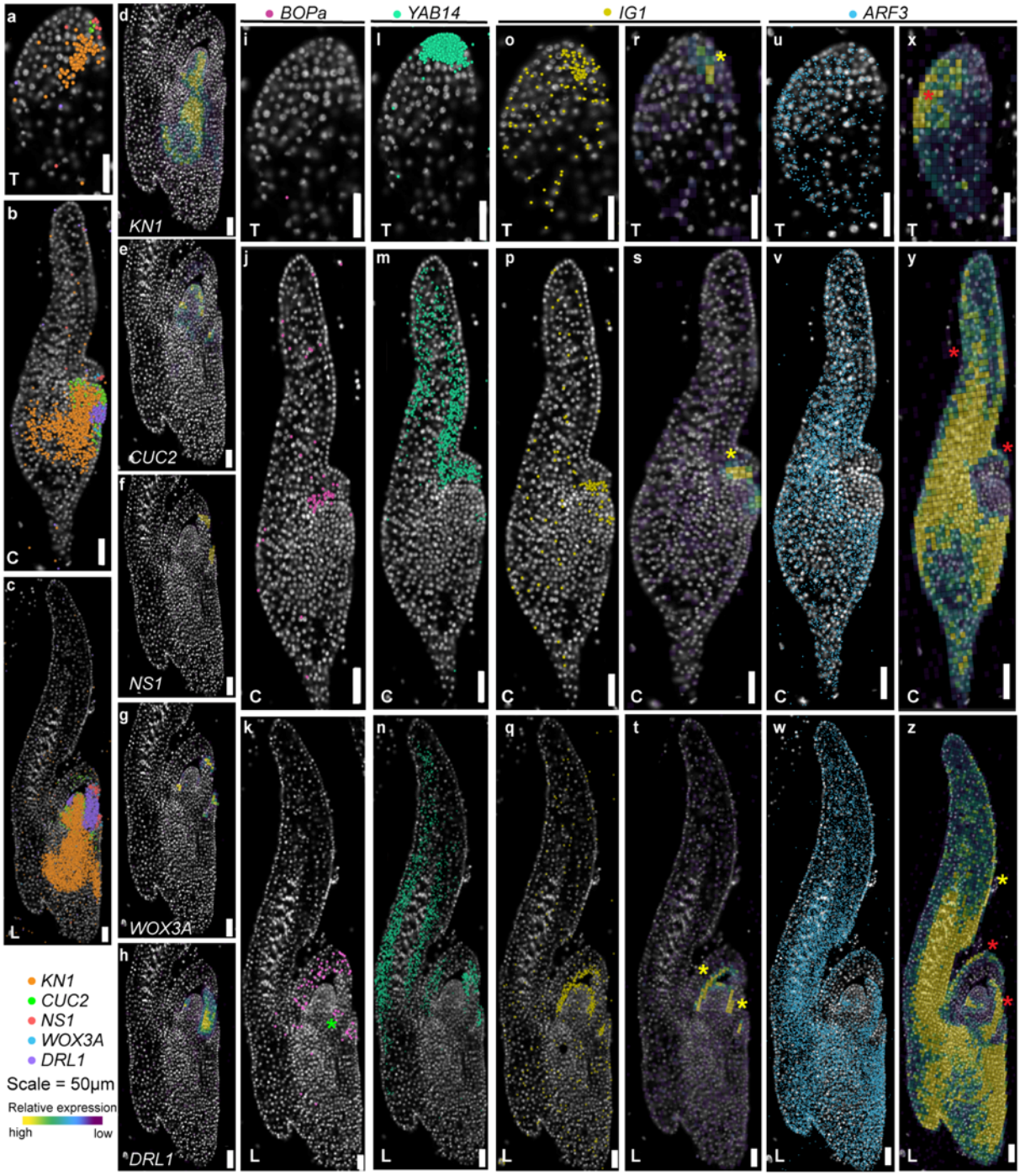

BOPa
YAB14
IG1
ARF3
KN1
CUC2
NS1
WOX3A
DRL1
KN1
CUC2
NS1
WOX3A
DRL1
Scale = 50µm
Relative expression
high
low

Fig. 3 RNA-targeting analyses of Transition-staged, Coleoptilar-staged, and Stage 1 embryos

a-c. Signal point map of *KN1*, *CUC2*, *NS1*, *WOX3A* and *DRL1* at the Transition stage (**a**), Coleoptilar stage (**b**) and Stage 1 (**c**). **d-h.** Expression density map of *KN1* (**d**), *CUC2* (**e**), *NS1* (**f**), *WOX3A* (**g**) and *DRL1* (**h**) at Stage 1. **i-k, l-n, o-q** and **u-w**. Signal point map of *BOPa* (**i-k**), *YAB14* (**l-n**), *IG1* (**o-q**), and *ARF3* (**u-w**) at the Transition stage, Coleoptilar stage and Stage 1. **r-t** and **x-z**. Expression density map of *IG1* (**r-t**) and *ARF3* (**x-z**) at the corresponding stages. Figure labels: T, Transition stage; C, Coleoptilar stage; L: Stage 1. The point colors are arbitrarily determined. Yellow asterisks mark the adaxial expression of *IG1* and *ARF3*, and red asterisks mark the abaxial expression of *ARF3*. The green asterisk in panel **k** marks the expression maximum of *BOPa* at the emerging leaf sheath. The yellow asterisks in panel **r-t** and **z** mark the expression maxima of *IG1* at the adaxial side of the scutellum (**r**), the coleoptile (**s**), the coleoptile and the leaf (**t**), as well as the expression maxima of *ARF3* at the adaxial side of the scutellum (**z**), respectively. The red asterisks in panel **x-z** mark the abaxial expression of *ARF3* at the scutellum (**x, y**), the coleoptile (**y, z**) and the leaf (**z**). Scale bar for all panels = 50 μm.

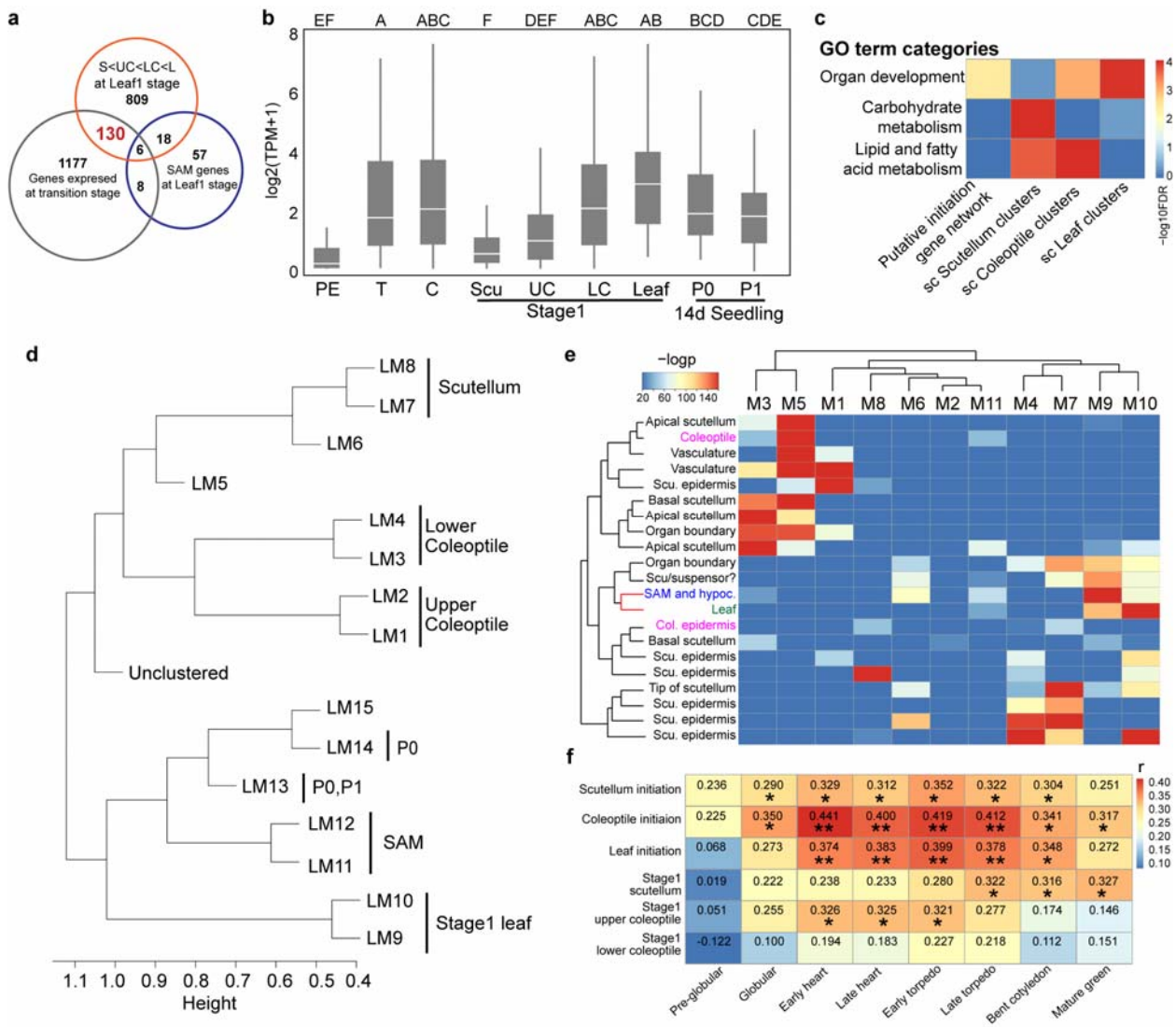
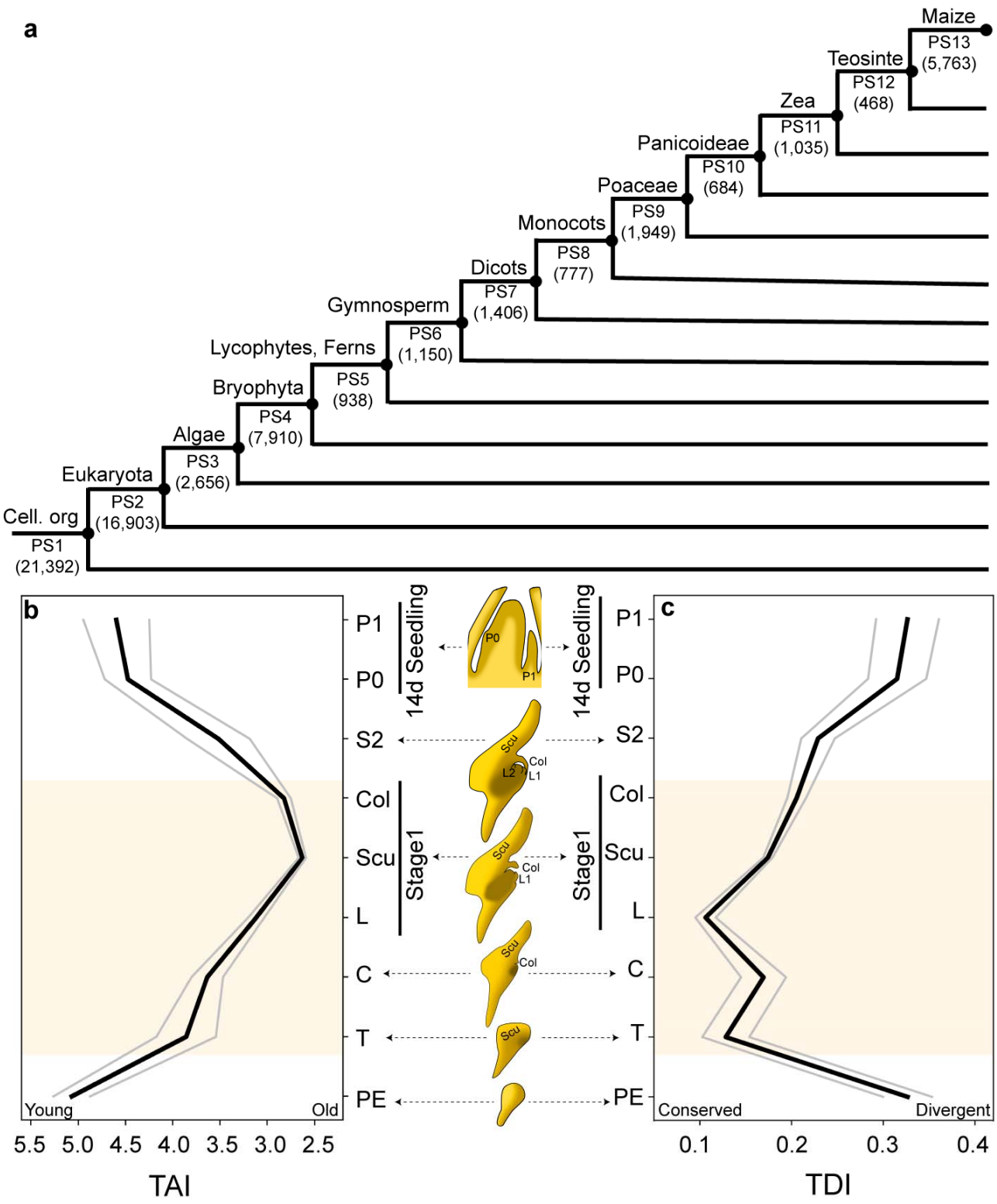


Fig. 4 Dynamics of the maize embryo organ initiation network

a. Venn diagram identifies 130 genes in the embryo-organ-initiation-network. **b**. Normalized expression of these 130 genes at the Proembryo stage (PE), Transition stage (T), Coleoptilar stage (C), Stage 1 scutellum (Scu), Stage 1 upper coleoptile (UC), Stage 1 lower coleoptile (LC), Stage 1 leaf (L), as well as P0 and P1 of the 14 day-after-germination seedling. The letters above the boxplot represent significance levels (non-overlapping letters indicate significant differences at $p<0.05$) by Tukey-Kramer HSD test. **c**. Heatmap of the GO terms of the putative initiation network, as well as the single-cell (sc) cluster of the Stage 1 scutellum, coleoptile and leaf. The

heatmap color is determined by $-\log_{10}$FDR (false discovery rate). **d.** WGCNA tree graph of the transcriptomic relationships among the co-expression modules specific to corresponding maize organs. The height (y-axis) scales the length of organ-specific branches, indicating the transcriptomic distance between the modules. **e**. Heatmap illustrating the correlation between the single-cell (sc) clusters (y-axis) and hdWGCNA modules (x=axis). The significance level is marked by $-\log_{10}$p. **f**. Heatmap illustrating the dynamic correlation of the initiation network expression between Maize (y-axis) and Arabidopsis (x-axis) across the embryogenesis. The heatmap color is determined by correlation coefficient (r) via Pearson correlation test with the significance level $p<0.05$ (*) or $p<0.01$ (**). Scu, Scutellum; Col, coleoptile; hypoc, hypocotyl.



1101

1102

1103

Fig. 5 Phylotypic stage analysis of the maize embryogenesis

a. Phylostratigraphic map of maize. All maize protein-coding transcripts were assigned to 13 phylostrata (PS1-PS13), Numbers in parenthesis refer to the number of genes per phylostratum, and the taxonomic groupings are marked above the corresponding PS. **b-c**. The transcriptome age index (TAI; **b**) and transcriptome divergence index profiles (TDI; **c**) of maize embryogenesis and seedling leaves. Figure labels: PE, proembryo stage; T, transition stage; C, coleoptilar stage; L, Stage 1 leaf; Scu, Stage 1 scutellum; Col, Stage 1 coleoptile; P0 and P1, Plastochron 0 and 1 leaf of the 14-day after planted seedlings. The gray lines represent the standard deviation estimated by permutation analysis. The overall patterns of TAI and TDI profiles are significant, as measured by flat line test ($P_{TAI}=3.56\times10^{-6}$, $P_{TDI}=4.80\times10^{-4}$), reductive hourglass test ($P_{TAI}=1.49\times10^{-4}$, $P_{TDI}=1.53\times10^{-3}$), and reductive early conservation test ($P_{TAI}=1$, $P_{TDI}=0.998$).

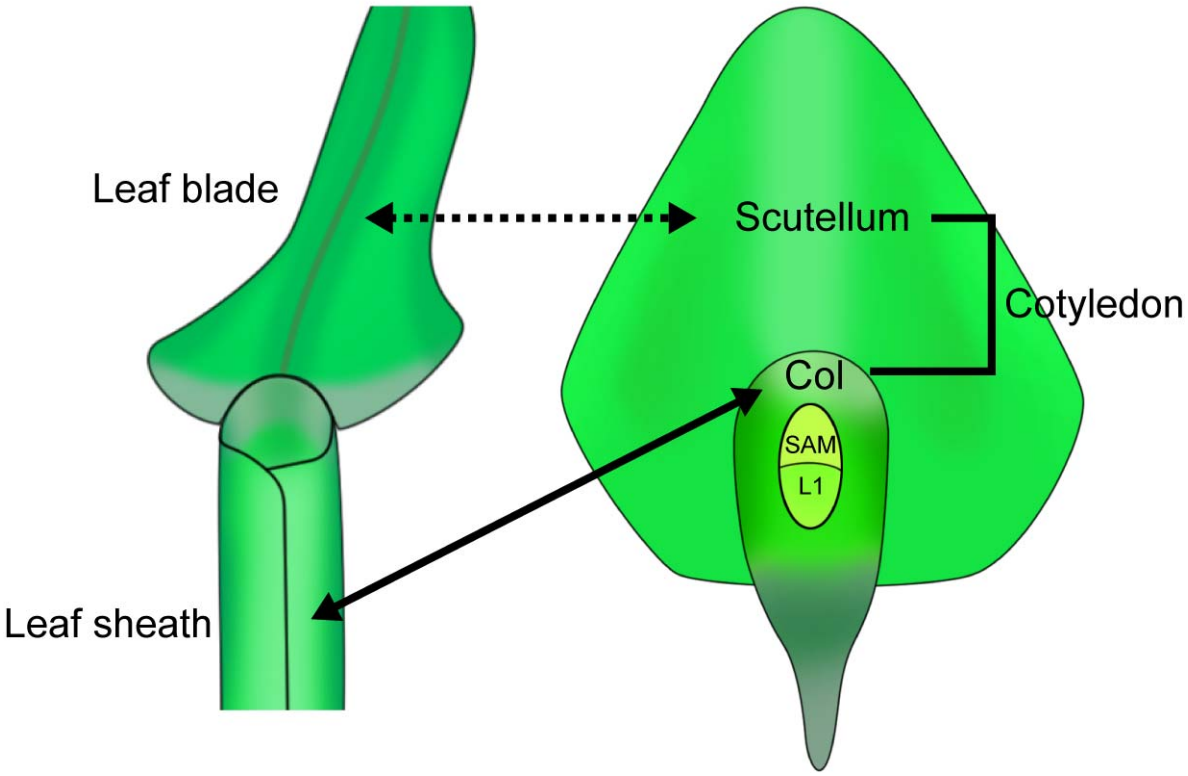
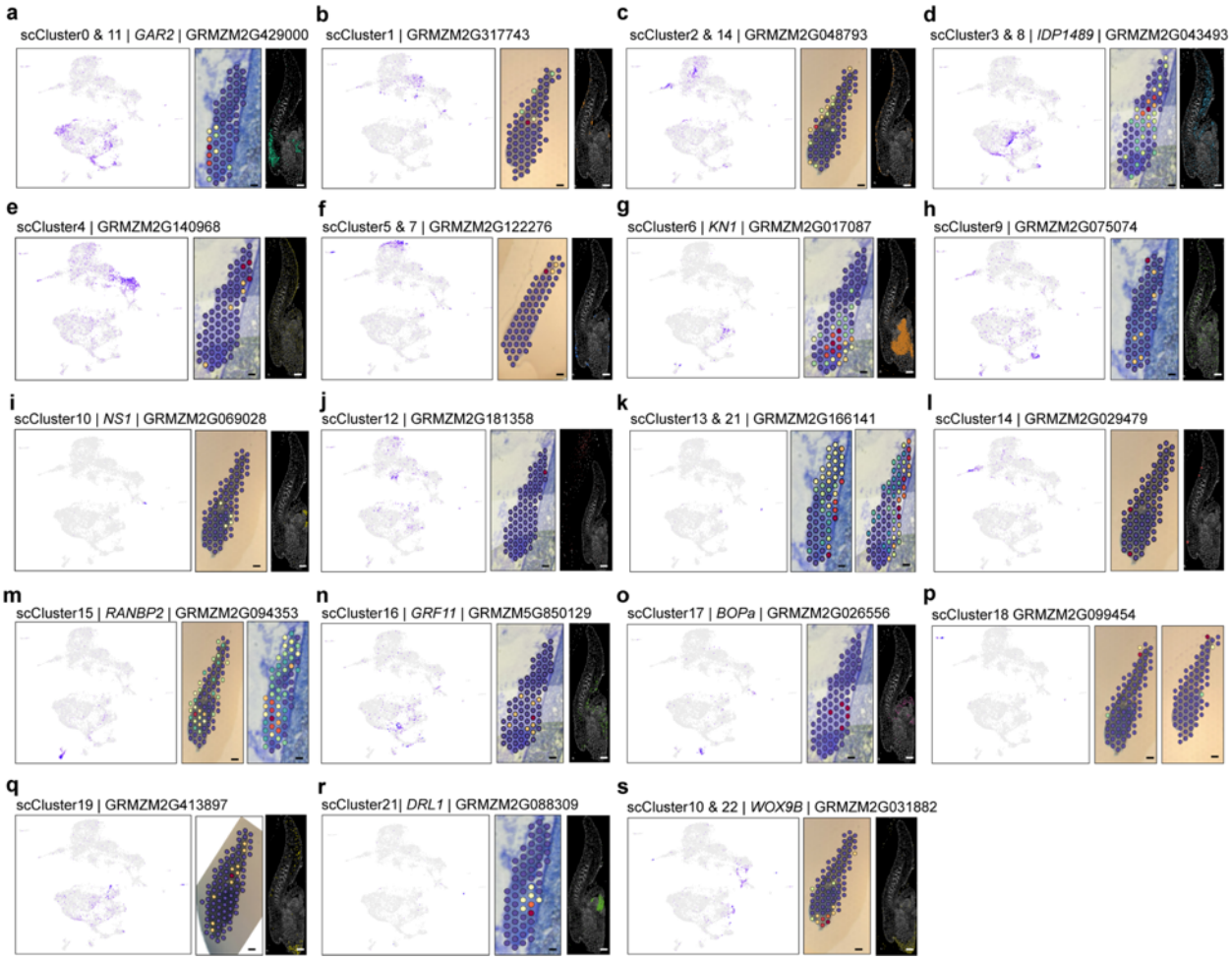


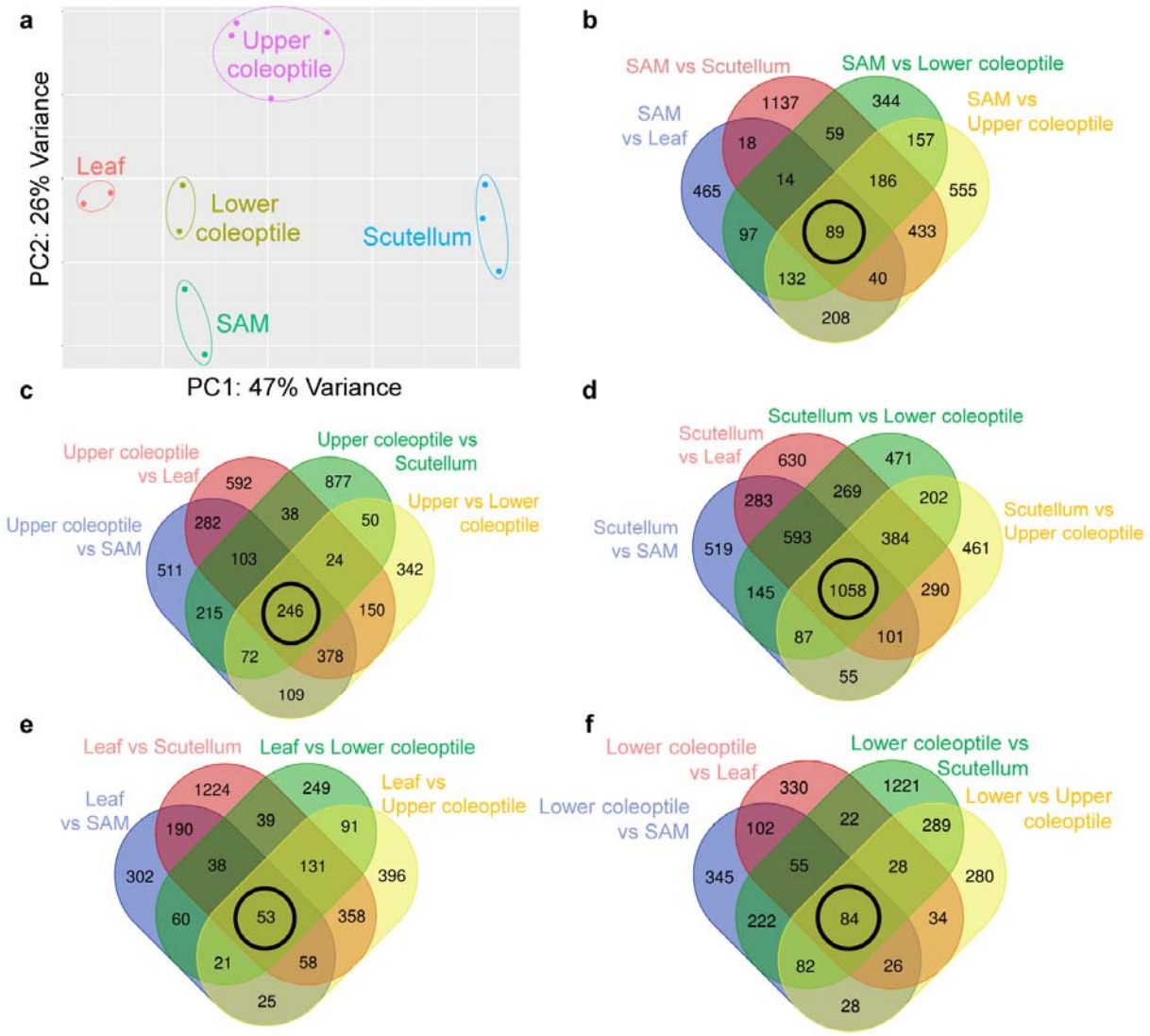
Fig. 6 Model for the homology of the grass cotyledon

The model describes that the single maize cotyledon as composed of two domains, the apical scutellum (homologous to the foliar leaf blade; dashed arrow line) and the basal coleoptile (homologous to the foliar leaf sheath; solid arrow line)



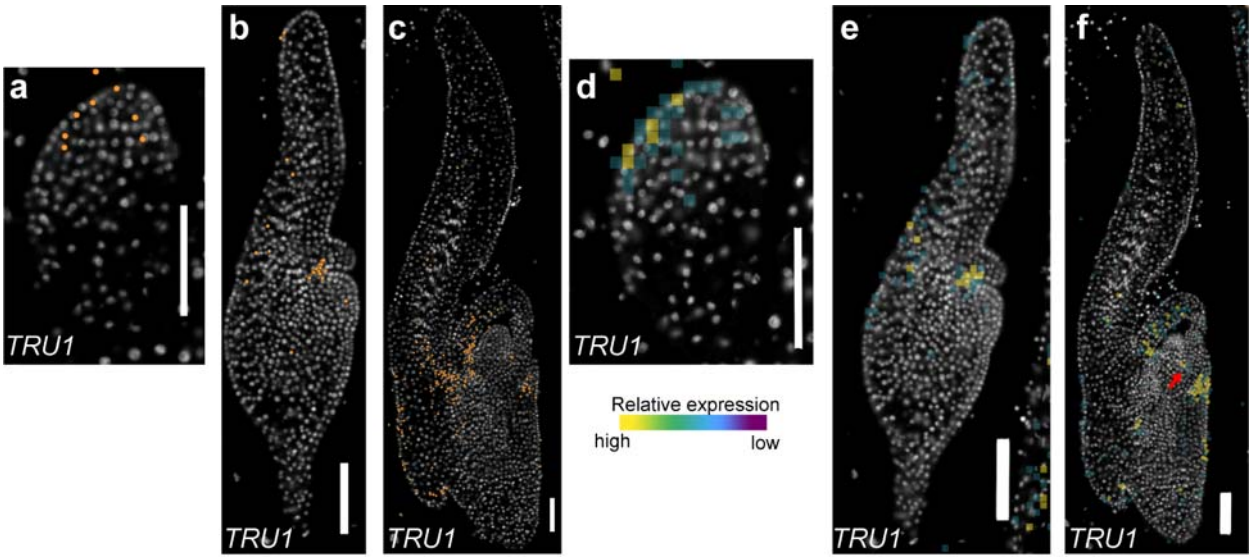
Extended Data Fig. 1 Verification of single-cell cluster marker genes.

Left panel: the single-cell cluster marker expression on a UMAP. Purple dots mark cells with corresponding gene expression data. Middle panel: the verification of the single-cell cluster marker genes by Visium™. Red to green spots marks the location of the gene expression with high to low level, respectively. Right panel: the verification of the single-cell cluster marker genes by multiplex *in situ* hybridization. Note: that *GRMZM2G166141* (**k**), *GRMZM2G094353* (**m**) and *GRMZM2G099454* (**p**) lack the multiplex *in situ* hybridization data. Scale bar = 100 μm.



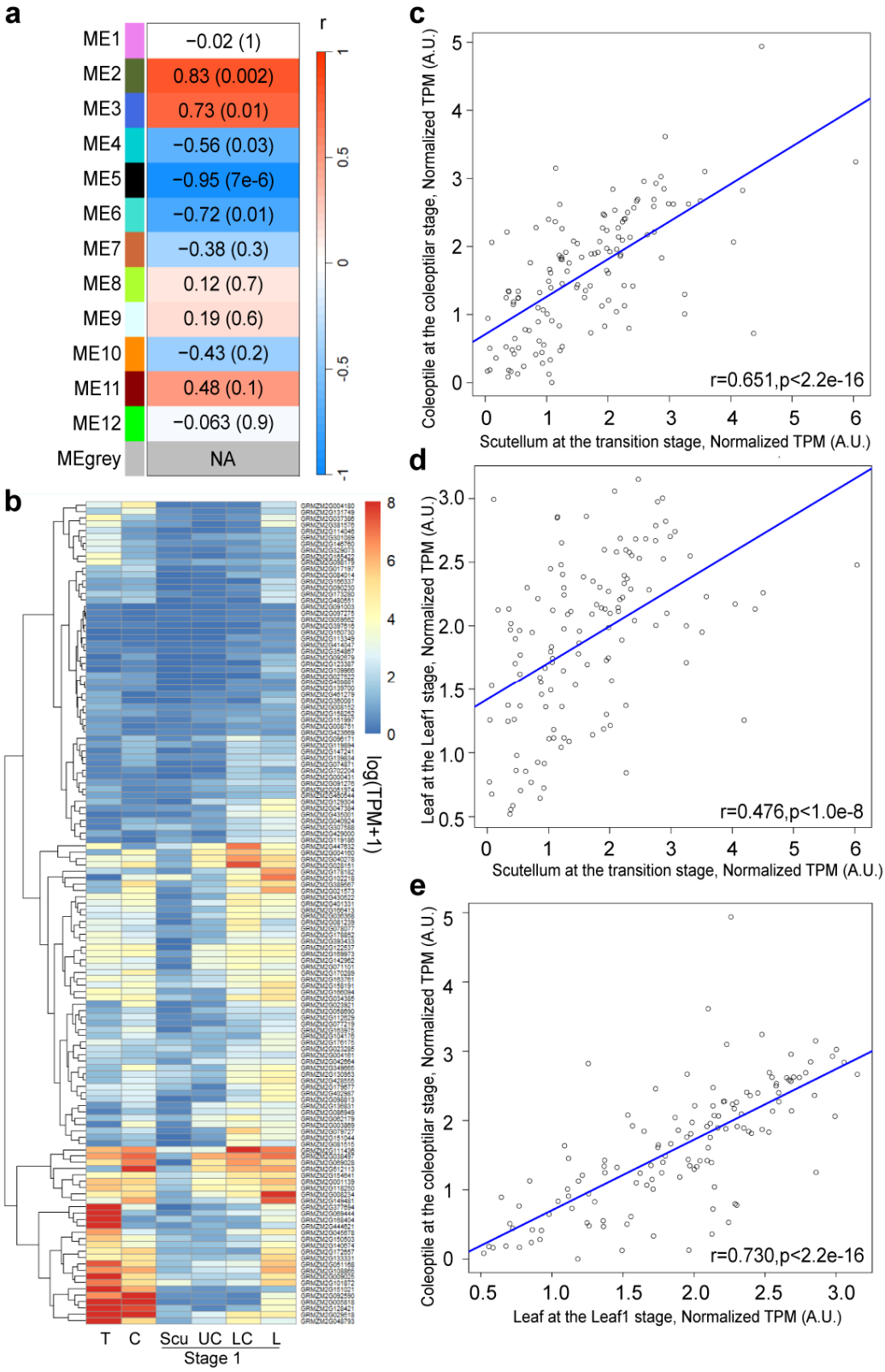
Extended Data Fig. 2 Differential expression between laser microdissected RNAseq samples.

a. Principal component analysis of laser-microdissected RNAseq samples. Each dot represents a laser-microdissected embryonic shoot sample. **b-f**. Pairwise comparison of the gene expression between laser-microdissected RNAseq samples. The number of overlapping, differential-expressed genes (DEGs) is showed in the Venn diagram. The circled numbers represent the genes preferentially expressed in the SAM (**b**), upper coleoptile (**c**), scutellum (**d**), leaf (**e**) and lower coleoptile (**f**).



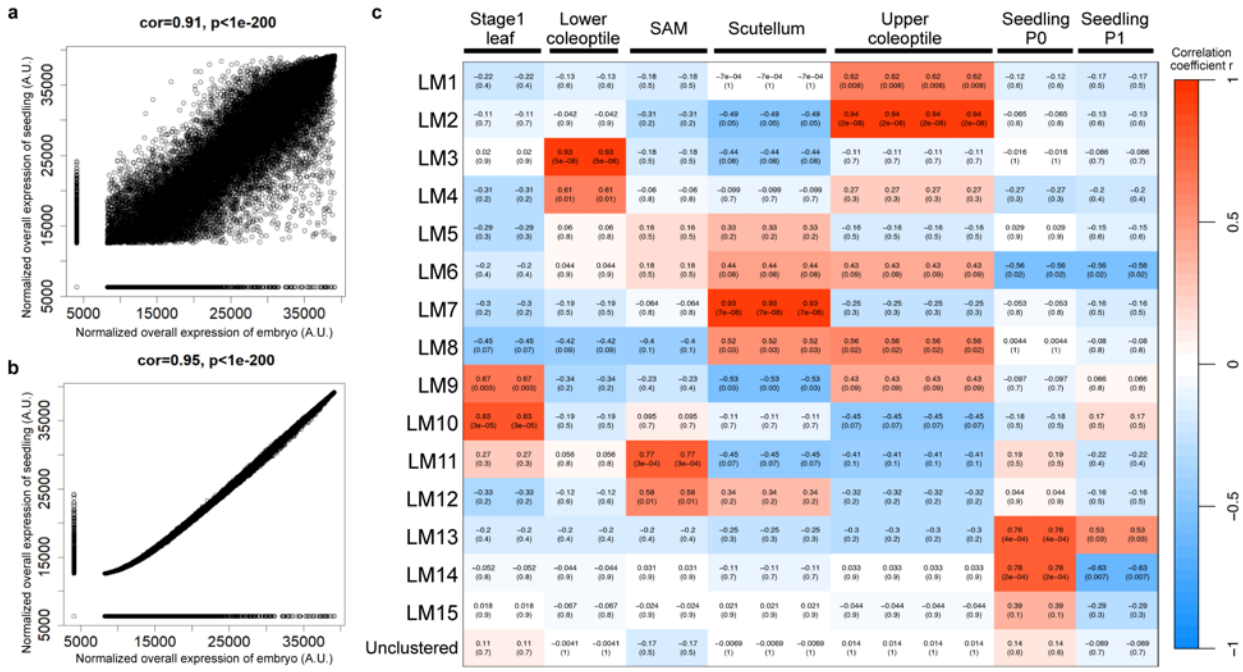
Extended Data Fig. 3 Multiplex *in situ* hybridization of *TRU1*.

The spatial expression of *TRU1* is showed in signal point map (**a-c**) and density map (**d-f**) from the transition stage (**a and d**), the coleoptilar stage (**b and e**) and the Stage 1 (**c and f**). The red arrow marks the expression of *TRU1* at the emerging leaf sheath. Scale bar = 100 μm.


a
ME1 −0.02 (1)
ME2 0.83 (0.002)
ME3 0.73 (0.01)
ME4 −0.56 (0.03)
ME5 −0.95 (7e-6)
ME6 −0.72 (0.01)
ME7 −0.38 (0.3)
ME8 0.12 (0.7)
ME9 0.19 (0.6)
ME10 −0.43 (0.2)
ME11 0.48 (0.1)
ME12 −0.063 (0.9)
MEgrey NA
r
b
log(TPM+1)
T C Scu UC LC L
Stage 1
c
Coleoptile at the coleoptilar stage, Normalized TPM (A.U.)
Scutellum at the transition stage, Normalized TPM (A.U.)
r=0.651,p<2.2e-16
d
Leaf at the Leaf1 stage, Normalized TPM (A.U.)
Scutellum at the transition stage, Normalized TPM (A.U.)
r=0.476,p<1.0e-8
e
Coleoptile at the coleoptilar stage, Normalized TPM (A.U.)
Leaf at the Leaf1 stage, Normalized TPM (A.U.)
r=0.730,p<2.2e-16

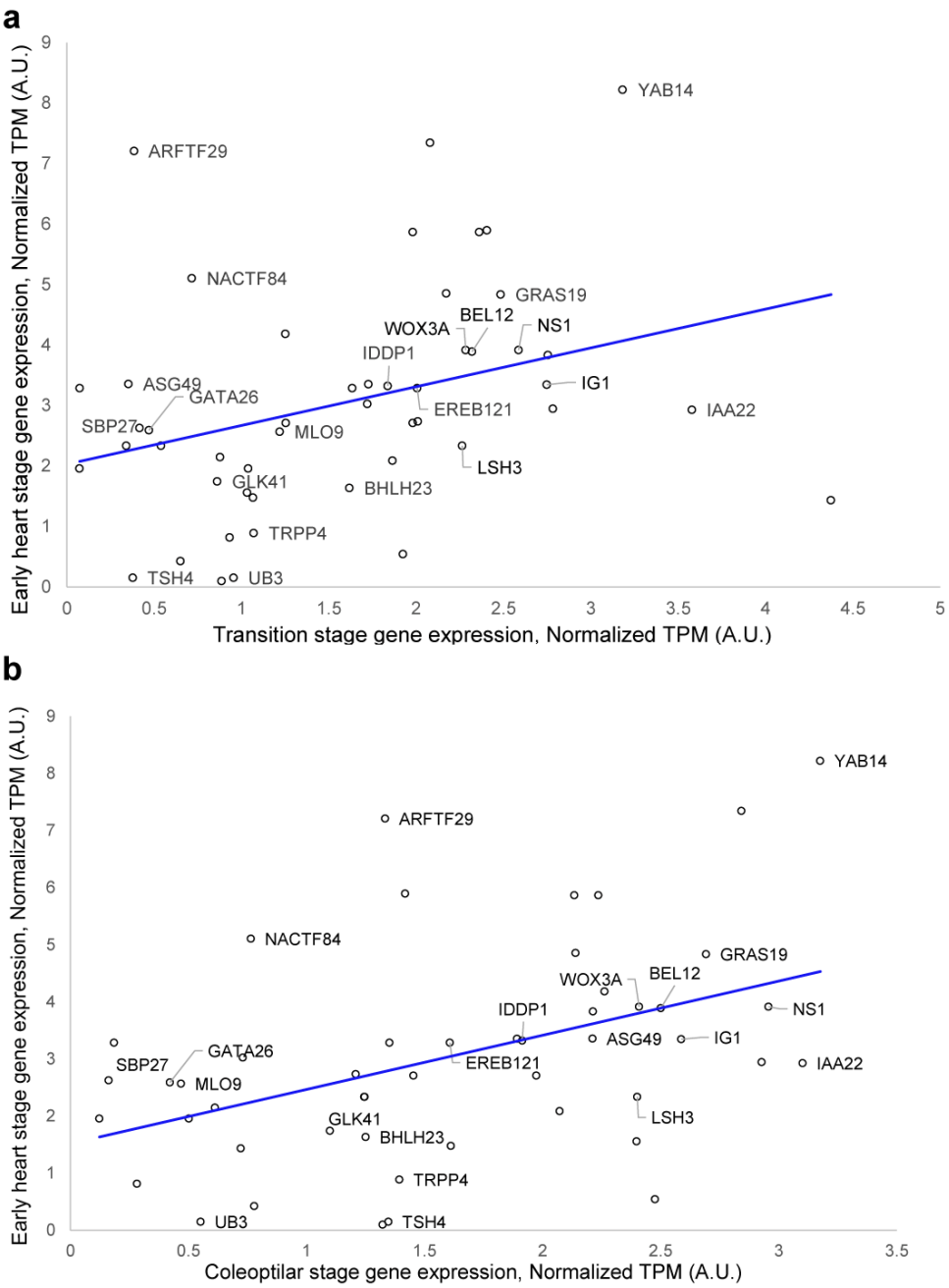
Extended Data Fig. 4 Expression of the embryo organ initiation network.

a. WGCNA to identify modules with overall highest expression in the Stage 1 leaf (L) and less in the lower coleoptile (LC), with steadily decreased expression in the upper coleoptile (UC) and scutellum (S) (i.e. L > LC > UC > S). The heatmap shows the correlation coefficient (r) and the corresponding p value (in the parentheses) of each module (ME1-12). **b**. The expression of the 130 genes in the initiation network at the transition stage (T), coleoptilar stage (C) and the Stage 1 scutellum (Scu), upper coleoptile (UC), lower coleoptile (LC) and leaf (L). **c-e**. The correlation of the expression of the 130 genes between initiation stage of the scutellum, coleoptile and leaf.



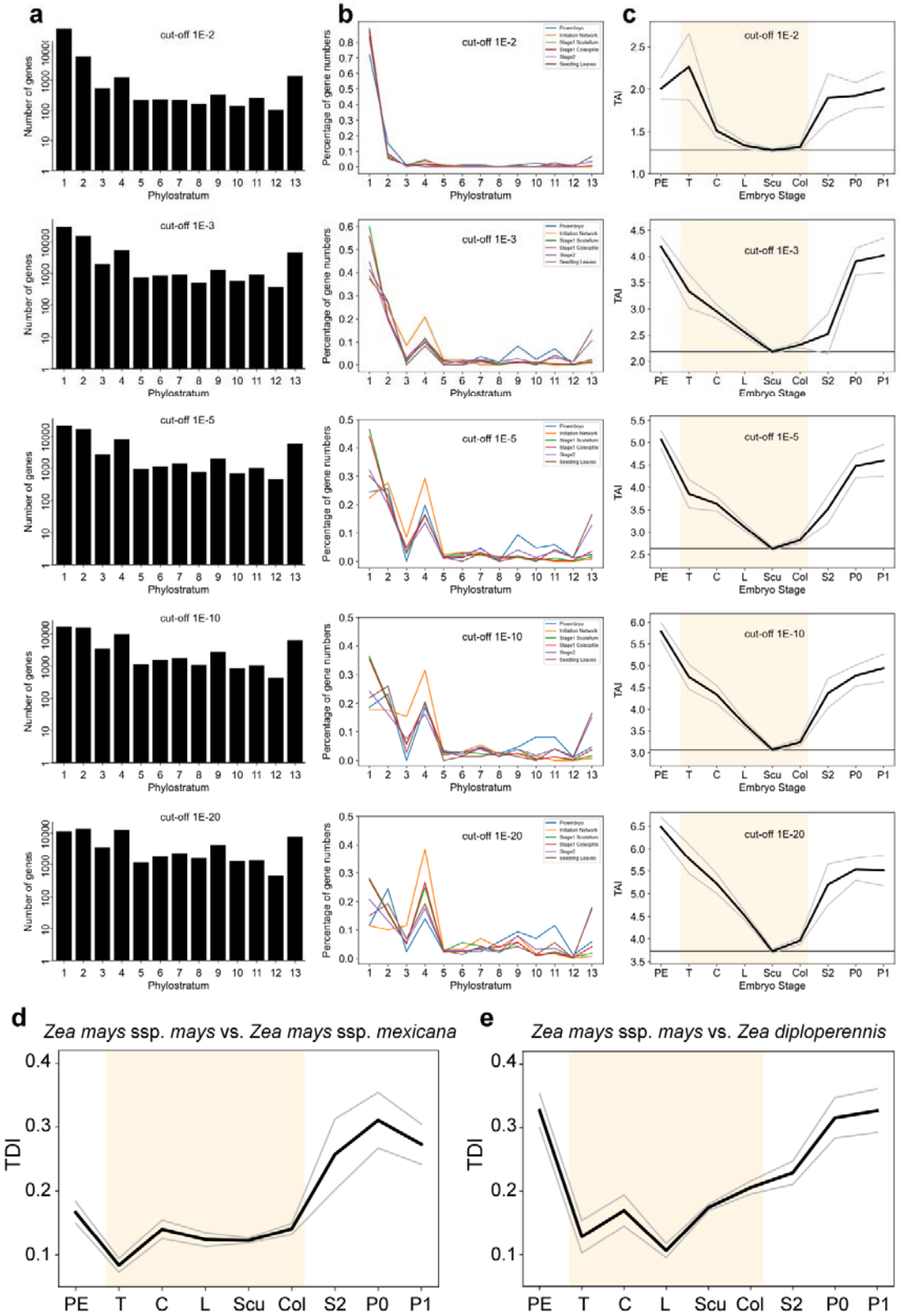
Extended Data Fig. 5 WGCNA Module-organ correlation.

a-b. Correlation between datasets of the 14-day after pollination seedling and the embryo before and after batch effect removal. The correlation of the normalized overall gene expression between the 14 day after germination (DAG) seedling and embryo datasets before (**a**) and after (**b**) batch effect removal. **c**. Heatmap illustrating the correlation between co-expression modules (y-axis) and corresponding organs (x-axis. The colors are determined by the correlation coefficient (r) with the p value (numbers in the parentheses)



Extended Data Fig. 6 Correlation between the maize initiation network genes and the Arabidopsis orthologs.

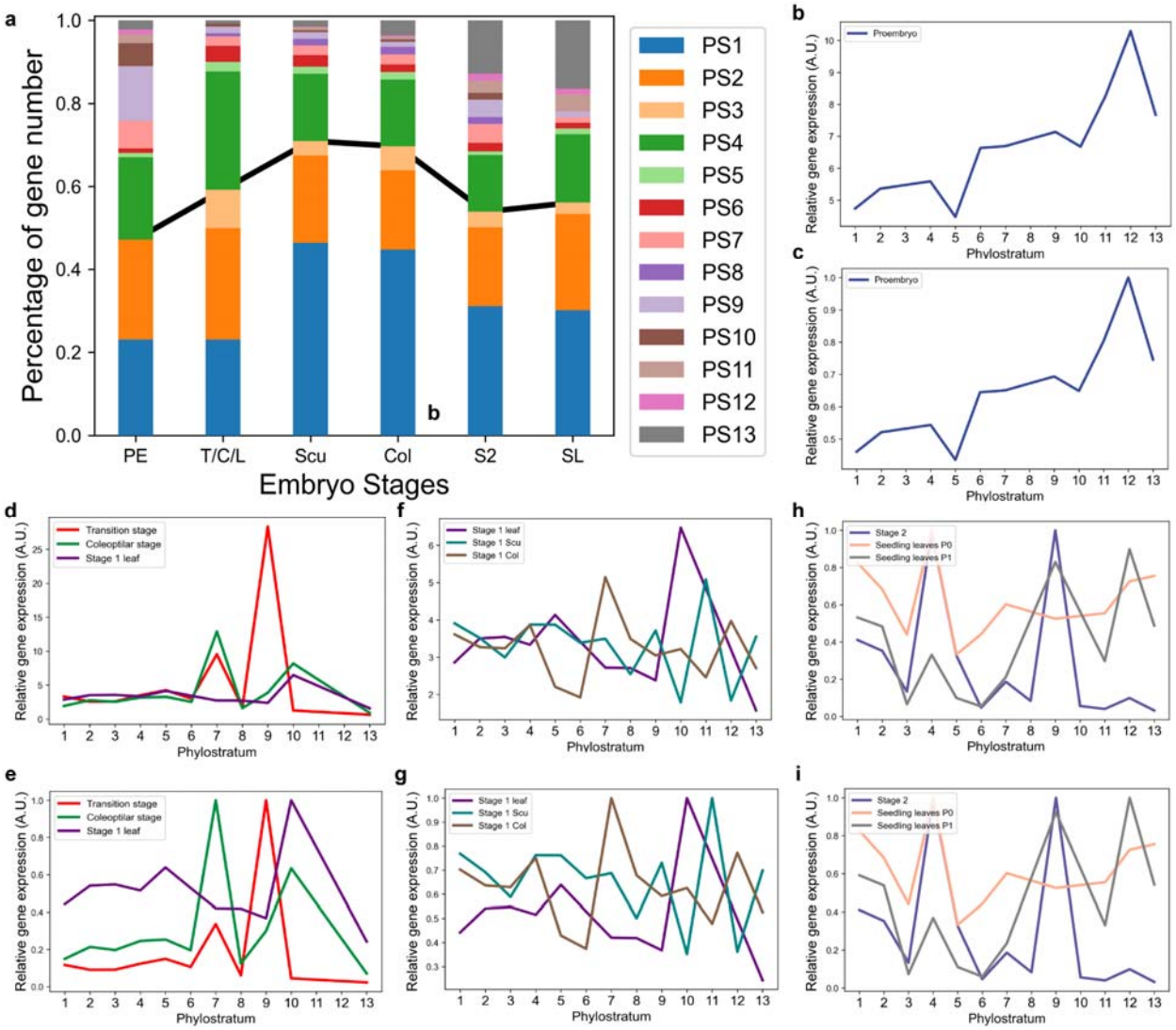
The correlation of expression between the maize initiation network genes at transition (**a**) and coleoptilar (**b**) stage, and the Arabidopsis orthologs at early heart stage.


a
b
c
cut-off 1E-2
cut-off 1E-3
cut-off 1E-5
cut-off 1E-10
cut-off 1E-20
Number of genes
Percentage of gene numbers
TAI
Phylostratum
Embryo Stage
PE T C L Scu Col S2 P0 P1
d
Zea mays ssp. mays vs. Zea mays ssp. mexicana
e
Zea mays ssp. mays vs. Zea diploperennis
TDI

1215

Extended Data Fig. 7. TAI and TDI for embryo stages under different conditions

a-c. Distribution of all maize genes and genes in each embryo stage across phylostrata and TAI profiles for different E-value cut-offs (ranging from 1E-2 to 1E-20). This analysis investigates whether and how E-values affect the hourglass pattern. **a**. Variation of the E-value cut-off for the BLASTp analysis demonstrates minimal import on the distribution of maize genes among the phylostrata, indicating an overall pattern independent of the E-values. The y-axis is presented on a logarithmic scale. **b**. Similarly, the variation of the E-value cut-off for the BLASTp analysis does not much severely alter the distribution of maize genes in each embryo stage among the phylostrata. **c**. the hourglass pattern of the TAI profile remains consistent across various E-value cut-offs. **d-e**. Transcriptome divergence indices based on two different genome comparisons: *Zea mays* ssp. Mays vs. *Zea mays* ssp. mexicana (**d**); *Zea mays* ssp. mays vs. *Zea diploperennis* (**e**). The overall patterns of TDI profiles are significant, as measured by flat line test ($P_{\text{TDI-Mexicana}}=1.07\times10^{-3}$, $P_{\text{TDI-diploperennis}}=4.80\times10^{-4}$), reductive hourglass test ($P_{\text{TDI Mexicana}}=0.027$, $P_{\text{TDI-diploperennis}}=1.53\times10^{-3}$), and reductive early conservation test ($P_{\text{TDI Mexicana}}=0.68$, $P_{\text{TDI-diploperennis}}=0.998$).



Extended Data Fig. 8. The gene expression in each embryo stage per phylostratum.

a. The stacked bars in percentage of gene numbers of each phylostratum in embryo stages. **b-i.** The relative gene expression (normalized to the maximal level in each embryo stage) of maize embryo genes, including in proembryo stage (**b, c**), in Initiation-Network stages (transition stage, coleoptilar stage, and stage 1 leaf) (**d, e**), in stage 1 (stage 1 leaf, stage 1 scutellum, stage 1 coleoptile) (**f, g**), and after stage 2 and in seedling leaves (**h, i**). The data to after stage 2 is from Chen et al., 2014, and were normalized in data source in (**h**).