A transcriptional atlas of early Arabidopsis seed development suggests mechanisms for inter-tissue coordination

Caroline A. Martin[1,2], Kylee R. Cogdill[1], Alesandra L. Pusey[1], Mary Gehring[1,2,3,4]

1 Whitehead Institute for Biomedical Research, Cambridge MA 02142
2 Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139
3 Department of Bioengineering, Massachusetts Institute of Technology, Cambridge MA 02139
4 Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Cambridge MA 02142

Author for correspondence: MG (mgehring@wi.mit.edu)

Abstract

Successful seed development is essential for flowering plant reproduction and requires the coordination of three genetically distinct tissues: the embryo and endosperm, which are the products of fertilization, and the maternal seed coat. Our understanding of the transcriptional programs underlying tissue-specific functions and inter-tissue coordination in seeds remains incomplete. To address this, we performed single nucleus RNA-sequencing on *Arabidopsis thaliana* seeds at 3, 5, and 7 days after pollination. We characterize all major seed cell or nuclei types, further refine transcriptional states in the endosperm, and map signatures of selection on cell type-specific genes. Among other findings, our analyses reveal the compartmentalization of genes involved in brassinosteroid-responsive transcription factor activation, abundant endosperm expression of genes that encode short, secreted peptides (SSPs), and an enrichment of rapidly evolving genes in the chalazal endosperm, illuminating the cell type and species specificity of seed genes.

14 **Main**

The three genetically distinct tissues of the seed are developmentally coordinated to
16 ensure propagation of the next generation[1]. This process begins when the egg cell and central cell of the female gametophyte are fertilized by sperm to generate the embryo and endosperm,
18 respectively. After fertilization the growth, development, and differentiation of the embryo and the more rapidly-developing endosperm are accompanied by growth and development of the
20 ovule integuments, which become the seed coat[2,3,4,5]. The close synchronization of developmental transitions in the seed suggests widespread signaling between tissues, even
22 though inter-tissue signals must cross cell walls and membranes as the embryo, endosperm, and seed coat are isolated symplastic fields[6,7]. Mechanisms underlying coordination within and
24 between tissues are beginning to be elucidated[8]. For example, seed coat development requires endosperm-derived auxin, and embryo morphogenesis relies on auxin from the early seed
26 coat[9,10,11]. The complete mechanisms for these interactions are unclear, and they are only two components of an extensive molecular dialogue.

28 The embryo-nourishing endosperm is a dynamic tissue that has been implicated in several axes of inter-tissue signaling[12,13,14]. It begins as a coenocyte, then undergoes
30 cellularization before embryo invasion and consumption during the final stages of seed development[15]. Arabidopsis endosperm was previously transcriptionally characterized
32 throughout seed development by laser-capture microdissection followed by microarray analysis, and was initially transcriptionally characterized at the single-nucleus level at 4 days after
34 pollination (DAP)[16,17]. Endosperm nuclei that are at key tissue interfaces show the highest transcriptional distinction, namely the embryo-proximal micropylar endosperm (MCE), and the
36 chalazal endosperm (CZE), which sits at the maternal-offspring interface and acts as a gateway for maternal resources into the embryo sac[16,17,18]. How endosperm compartments locally
38 coordinate processes at tissue interfaces is incompletely understood. However, some short, secreted peptides (SSPs) from the micropylar endosperm are well-characterized inter-tissue
40 signals. For example, the embryo-derived TWS1 SSP is processed in the endosperm by the ALE1 subtilase, and the mature peptide directs cuticle development in the embryo[19,20]. Given
42 the symplastic isolation of seed compartments, SSP signaling could be a frequent conduit for inter-tissue signals.

44 Throughout seed development, maternal resources are deposited into, and distributed from, the chalazal seed coat (CZSC), a specialized region of the seed coat at the terminus of
46 the vasculature[21,5]. Transporter- and channel-dense cells import nutrients and hormones and unload them into integument symplastic domains[22]. Recent evidence suggests that maternally

48 synthesized auxin and abscisic acid are imported from the funiculus and control seed size and dormancy, respectively[23,24]. The CZSC is a morphologically complex region and the degree of
50 functional specialization among CZSC cell types is unknown.

We present a timepoint-resolved snRNA-seq atlas of early Arabidopsis seed
52 development. Among other findings, our analyses revealed transcriptionally defined CZSC subtypes with complementary functions, and concentration of brassinosteroid synthesis in a
54 micropylar seed coat subtype. In the endosperm, we report genes that underlie transcriptional polarity within the CZE cyst, many of them SSPs, which is consistent with a general enrichment
56 of SSP expression in the CZE and MCE. Finally, we show that genes functioning in protein degradation and nucleic acid-related processes in the CZE appear to be under positive
58 selection. Our atlas is available for exploration at https://seedatlas.wi.mit.edu/.

60 **Results**

A transcriptional atlas of early Arabidopsis seed development

62 To capture the most dynamic period of endosperm development, we isolated and sequenced RNA from individual nuclei from 3, 5, and 7 DAP Col-0 seeds using the 10x
64 Genomics platform (Fig. 1A, Methods). At 3 DAP, the embryo is at the globular stage and the endosperm is coenocytic. At 5 DAP, the endosperm begins to cellularize at the micropylar end,
66 and at 7 DAP cellularization is complete and the embryo expands rapidly (Fig. 1A)[15]. For each time point we collected two biological replicates. After raw snRNA-seq data filtering and
68 correction (Methods) we identified an optimal clustering resolution based on both the number of clusters and cluster neighborhood purity (Supplementary Fig. 1). We assigned clusters to cell
70 types by analyzing the expression of known published marker genes, differential expression (DE) and GO term enrichment analysis, and by hybridization chain reaction RNA-fluorescent *in*
72 *situ* hybridization (HCR RNA-FISH) of cluster marker genes (Supplementary Tables 1-4, Methods). At the highest level of resolution, which we refer to as level 3 annotation (L3), we
74 identified 28, 29, and 21 clusters at 3, 5, and 7 DAP, respectively (Fig. 1B-D, Extended Data 1, Supplementary Table 3). The L3 clusters were then assigned level 2 (L2) and level 1 (L1)
76 annotations, which were harmonized across timepoints (Fig. 1B, C). Once each timepoint was annotated separately, the datasets were integrated into a final atlas dataset (Fig. 1B-D). In total,
78 our atlas contains 55,340 profiles (24,617 at 3 DAP, 16,431 at 5 DAP, 14,292 at 7 DAP) and is approximately 11% embryo, 24% endosperm, 63% seed coat, and 2% unfertilized ovule and
80 funiculus (Fig. 1E, Extended Data 2).

To assess atlas completeness, we determined whether we had captured rare embryonic cell types (Fig. 1F). The majority of embryo nuclei were isolated from seeds at 7 DAP (5,141), with 345 nuclei from 3 DAP and 414 nuclei from 5 DAP (Fig 1G, Supplementary Table 3). Although embryo nuclei were identified as single protoderm clusters at 3 and 5 DAP, at 7 DAP we identified clusters corresponding to the inner cotyledon, vascular primordium, ground tissue initials, cortical initials, and hypophysis (Fig. G)[25–30]. Correspondence between clusters and cell types was validated by HCR RNA-FISH (Fig. 1H, Supplementary Table 1-2). The shoot apical meristem (SAM) and suspensor represent rare embryonic cell types, which respectively specifically express *CUC1* and *WOX8* (Fig 1F)[31,32]. We detected 14 *CUC1*+ (5-7 DAP) nuclei and 31 *WOX8*+ (3-5 DAP) nuclei (Fig. 1I). This indicates that ~55k nuclei enable the detection of rare cell populations in the seed, although no nuclei in our dataset express *WUS*, a known marker for a subpopulation of the SAM expressed early in development[26,33]. Overall, the clusters we defined include representatives of all previously anatomically and morphologically defined seed cell types (Fig. 1, Extended Data 1, Supplementary Table 3).

A high-resolution census of the developing seed coat

The seed coat is the predominant seed tissue type at early stages (Fig. 1B,E), comprised of the five-layered testa, which in past studies has been referred to as the “general seed coat” (GSC), and the chalazal seed coat (CZSC), or the cell types near maternal vascular terminals (Fig. 2A)[17,34]. Each layer of the GSC is one cell thick, which has made comprehensive transcriptional profiling of these cell types difficult, although many layer-specific genes have been identified[35–40]. Using published markers, we identified all five layers of the seed coat in our L2 annotations across all timepoints, except for 7 DAP, for which there is a single ii1’/ii2 (inner integument 1’/inner integument 2) cluster (Extended Data 3, Supplemental Table 1, Supplemental Table 3).

Whereas the upper seed coat contains the five principal layers, the lower seed coat is a densely heterogeneous region where maternal vasculature, the CZSC, and layers of the general seed coat meet (Fig. 2A). It serves roles in hormonal signaling and nutrient uptake in the developing seed[23,41–43]. To identify lower seed coat subtypes, we used the published marker *ARR22,* which is enriched in the CZSC but is also detected in adjacent lower seed coat tissues[44]. We identified 19 *ARR22*+ clusters across all timepoints (Fig. 2B), including subtypes of the GSC, CZSC, and the chalazal proliferating tissue (CPT), a small population derived from the nucellus and that lies beneath the chalazal endosperm cyst[45]. After fertilization, the CPT is partially degraded, characterized by “persistent” and “transient” CPT subtypes, and then fully

degraded between 5-7 DAP[46,47]. We used the markers *SWEET4* and *PLL24*[48] to identify the
116 persistent and transient populations, respectively, and further identified a highly specific marker for the putative persistent CPT, *AGO3*, which we validated by HCR RNA-FISH (Fig. 2B-C,
118 Supplementary Table 4). A GO term analysis using the top persistent CPT DE genes (log_2FC > 1, adj. pval <0.05) indicated that it is a hotspot for nucleoside catabolism in the seed, consistent
120 with ongoing programmed cell death (PCD) in this region. Seven nucleoside catabolism genes showed high specificity for the persistent CPT seed-wide (Fig. 2D, Extended Data 3). DE
122 analysis between the 3 DAP persistent and transient CPT clusters implicated genes associated with cell death and metal response (Fig. 2E). Notably, *DELTAVPE*, a contributor to cell death in
124 the inner integument[39], is DE in the transient CPT, while *MSL10*, an ion channel that positively regulates programmed cell death in a mechanically sensitive manner[49], is DE in the persistent
126 population, suggesting differing cell death triggers in the transient and persistent CPT. This represents the first transcriptomic characterization of the CPT in Arabidopsis.
128

Cell types in the placentochalazal region of the chalazal seed coat have complementary
130 **functions**

The CZSC is appreciated for being the primary site of active nutrient transfer into the
132 seed from maternal vascular terminals, facilitated by SWEET, UMAMIT, and PHO1 transporters [41,42,50,51]. The cells closest to the maternal vasculature comprise the placentochalazal region and
134 specifically express the gene encoding the UMAMIT14 amino acid transporter [50](Fig. 2A, B). Based on *UMAMIT14* and high *ARR22* expression, we identified six putative CZSC clusters
136 across all timepoints, which were grouped into three subtypes based on the expression of *SWEET10*, *TET5*, and *AAP6* (Fig. 2B, F, Supplemental Table 4). Of these, we found two
138 putative placentochalazal subtypes at 3 and 5 DAP CZSC, which were labeled specifically by *SWEET10* and *TET5* and shared *UMAMIT14* expression (Fig. 2B, F, G). We identified DE
140 genes in these subtypes compared to all other clusters at 3 DAP (average log_2FC > 1, adj. p-value >0.05) (Fig. 2G, Extended Data 5). *TET5*+ cluster DE genes are associated with the GO
142 term “(1->3)-beta-D-glucan (callose) metabolic process”. Callose deposits are known to regulate plasmodesmata activity and serve an insulating role between cell types of the developing
144 ovule[52]. A callose-rich “phloem end” has recently been described in the CZSC[53]. Gene set enrichment analysis (GSEA) using all genes associated the GO term “callose biosynthesis”
146 revealed high, significant enrichment in *TET5*+ clusters across all timepoints (Extended Data 4). Several putative callose biosynthesis genes are specifically expressed in this population, the
148 strongest being *CALS8* (Fig. 2B, Extended Data 4).

In contrast, top exclusive GO terms for the *SWEET10+* cluster DE genes include
150 "phosphate ion transport", and "phosphatase activity", supported by the differential expression of *PHO1;H1*; and *TPPE, TPPB,* and *TPPG*, respectively (Fig. 2G, Extended Data 4). *PHO1;H1* is a
152 phosphate exporter with demonstrated activity in the *Arabidopsis* CZSC[42]. Additionally, trehalose phosphatase expression suggests that these cells participate in trehalose-6-
154 phosphate signaling, which directs sugar utilization[54–57]. The invertase *cwINV4* sugar transporter is also specifically expressed in this cluster (Extended Data 4). These results indicate that the
156 *SWEET10+* cluster is likely the primary site of nutrient transfer in the placentochalazal CZSC, and the *TET5+* and *SWEET10+* populations have complementary functions influencing the
158 permeability of maternal vascular terminals and surrounding cells.

160 **Brassinosteroid biosynthesis, homeostasis, and response genes show concentrated expression in the micropylar region of the seed**

162 Brassinosteroids (BR) are hormones that act widely in plant physiology and are known to direct organ formation and cell expansion in reproductive tissues[58–64]. In the seed they promote
164 endosperm proliferation by reducing the physical resistance of the seed coat through cell wall weakening[64]. Although maternal seed coat-derived BR are hypothesized to interact directly with
166 the endosperm, a recent study indicates that BR controls endosperm proliferation through cell autonomous effects in the seed coat[64]. To identify sites of BR biosynthesis in the developing
168 seed, we performed a GSEA for the GO term gene sets "BR biosynthesis" and "BR homeostasis." Across all timepoints, the ii2 (inner integument 2) seed coat layer showed the
170 greatest enrichment for BR biosynthesis (Fig. 3B). Additionally, partitioning by timepoint and L3 annotation revealed that a putative micropylar oi (outer integument) subtype (Fig. 3A-C) exhibits
172 the most significant enrichment for genes associated with BR biosynthesis and homeostasis atlas-wide. We annotated an oi micropylar cluster based on the specific expression of
174 *SWEET12* (Fig. 2B), which has previously been located to the micropylar end of the seed coat, and because the cluster shows high transcriptional similarity to oi clusters, with a slight bias
176 toward oi1 (Extended Data 5)[51]. An inspection of the genes underlying the micropylar oi enrichment scores revealed that the enzymes that catalyze the last two steps of BR
178 biosynthesis, *BR6OX1* and *BR6OX2*, are upregulated in this subtype (Fig. 3D). We propose that the micropylar oi is a key site for brassinosteroid production at early to intermediate stages of
180 seed development.

BZR1/BES1 transcription factors are activated by BR via the BRI receptor-like kinase,
182 although BR-independent activation of *BES1* by the SSP TPD1 through EMS1-SERK1/2

signaling has also been observed[65]. Once activated, they promote the expression of genes
184 involved in cell elongation, light-regulated development, and cell wall remodeling[65–70]. The
activity of some BZR1/BES1 family members has been characterized in the endosperm:
186 constitutively active *BES1* or *BZR1* causes reduced proliferation of endosperm nuclei, while
pentuple mutants of BZR1/BES1 family members show no endosperm phenotype, suggesting
188 redundant mechanisms for BZR1/BES1-family contributions to endosperm development[64]. To
map BZR1/BES1 activity, we performed GSEA for the six BZR1/BES1 genes detected in the
190 atlas, revealing high and specific expression in the 5 DAP MCE, driven by *BES1*, *BEH3*, *BZR1,*
and *BEH4* (Fig. 3E). The transmembrane BR receptor *BRI1* shows broad expression throughout
192 the seed coat, but is depleted in the 5 DAP MCE. To determine whether BZR1/BES1 may be
activated in a BR-independent manner in the MCE, we mapped the expression of *EMS1*, *TPD1*,
194 and *SERK1/2*. Although *EMS1* and *SERK1/2* show overlapping spatiotemporal expression
patterns with MCE BZR1/BES1 transcription factors, *TPD1* is nearly undetectable atlas-wide.
196 To test whether the expression of *BES1* targets show increased expression after *BES1*
upregulation in the MCE, we performed GSEA of all *BES1* targets that are DE in the MCE,
198 which indicated significant upregulation from 5 to 7 DAP (Fig. 3F-G) when ordered by reversed
7 DAP-anchored MCE pseudotime[71]. The co-occurrence of BR biosynthesis and response
200 genes in proximal micropylar cell types, the absence of *TPD1*, and the timing of BES1 target
upregulation supports the hypothesis that BRs are transported from outer integument 1 to the
202 micropylar endosperm. However, it is plausible that *EMS1* and *SERK1/2* may be activated by an
alternative TPD1-like SSP, in which case, BZR1/BES1 activation might occur cell autonomously
204 in the MCE (Supplementary Fig. 4).

206 **The micropylar-to-Embryo Surrounding Region shift is characterized by UMAMIT and nitrate transporter expression**

208 The induction of BZR1/BES1 transcription factors precedes a dramatic transcriptional
shift in the endosperm between 5-7 DAP. In a correlation analysis of L2 endosperm clusters
210 throughout development, we observed that the 7 DAP endosperm is the most transcriptionally
divergent (Extended Data 6). Within 7 DAP endosperm clusters, the *GLIP6*+ MCE (referred to
212 as the “embryo surrounding region” (ESR) at this stage), is the most distinct, driven by the
upregulation of *UMAMIT*s and nitrate transporters (Extended Data 6). The ESR is known to
214 shape embryo viability by controlling the formation of two barriers that support successful
germination: the embryonic cuticle and sheath, which respectively prevent embryo dehydration
216 and endosperm adherence[19,72]. These processes are coordinated with programmed cell death

through the transcription factor *ZOU/RGE1*[73–76]. At 7 DAP, we observed two clusters within the
218 *ZOU+* population: one expressing *KRS,* a signaling peptide that directs embryo sheath formation, and another specifically expressing *NAC074* and enriched for *NAC087,* which directs
220 programmed cell death in the endosperm[72,75](Extended Data 6). *ZOU* induces *KRS* expression[73] and likely controls PCD in partially parallel pathways with NAC transcription factors[75], and they
222 appear to be distributed on a continuum of MCE nuclei states at 7 DAP. Other genes follow this pattern, such as the nitrate transporters, which are enriched in the *KRS+* cluster (Extended Data
224 6).

226 **Transcriptional polarity in the chalazal endosperm**

The chalazal endosperm (CZE) is composed of an unusual population of nuclei that
228 does not cellularize[76]. It is important for maternal resource allocation in the developing seed due to its position at the maternal-filial interface[77]. A subset of CZE nuclei form the multi-nucleate
230 cyst, a coenocytic sac isolated by the central vacuole, and another subset form the nodules, which lie above the cyst[78](Fig. 4A). Developmental time course and live imaging studies of
232 Arabidopsis endosperm indicate that the cyst grows in part through fusion with proximal nuclei from the nodule[3,78,79]. The cyst and nodule were transcriptionally characterized in our previous
234 snRNA-seq study of 4 DAP endosperm, which also described a “nodule-like” chalazal population that was not morphologically defined[16]. This atlas contains 3,140 CZE nuclei,
236 enabling higher-powered differential expression analyses. We first identified CZE nuclei populations using marker gene sets for the chalazal cyst, nodule, and nodule-like populations
238 defined from 4 DAP, which include AT2G44240 (cyst), *CYCD4;2* (nodule), and *MEA* (nodule-like) (Fig. 4D)[16,80]. These markers were not detected at 7 DAP, so *NPF4.5* was used to define
240 the 7 DAP CZE (Fig. 4D, Supplementary Table 1)[17].

Differential expression analysis of the 3 DAP L2 clusters revealed that the SSP gene
242 *RAPID ALKALINIZATION FACTOR-LIKE 3* (*RALFL3*) was a highly specific marker for a subpopulation of CZE nuclei (Fig. 4B, Supplementary Fig. 2-3). *RALFL3* HCR RNA-FISH
244 indicated transcript accumulation either throughout or at the base of the chalazal cyst (Fig. 2C, Fig. 4C, Extended Data 7). We compared the *RALFL3* HCR signal with that of AT3G49307, an
246 SSP gene more broadly expressed in the CZE, and found AT3G49307 signal in the apical region of the cyst and nodules throughout seed development (Fig. 4C, Extended Data. 7). We
248 propose that, rather than being a structure of uniform gene expression, there is a transcriptional apical-basal axis of the chalazal cyst, with AT3G49307 and *RALFL3* labeling the apical and
250 basal regions, respectively. At 3 DAP, the basal state predominates in the cyst, but at 5 DAP the

apical/basal distinction is pronounced and two cyst states are detectable as subclusters (Fig. 4D-F). A correlation analysis of aggregated gene expression for cell types within and adjacent to the CZE showed high transcriptional correlation between the basal cyst and proximal maternal tissues, the ii1 and CPT (Fig. 4A, E, Extended Data 8). Based on the enrichment of the basal cyst nuclei at early timepoints, *RALFL3* HCR labeling, and its transcriptional similarity to both the cyst and proximal maternal tissues, we hypothesize that the basal cyst nuclei represent the "founder" nuclei that migrate to the chalazal region at early timepoints, to which subsequent nuclei fuse to generate the chalazal cyst (Fig. 5F)[79].

At 5 DAP, the CZE nuclei cyst, nodule, and nodule-like populations show the greatest transcriptional distinction from one another (Fig. 4E). To order these closely-related nuclei types on a single trajectory and identify genes that may underlie the transition from a nodule-like to cyst state, we performed a pseudotime analysis of 5 DAP PEN and CZE nuclei, anchored in the basal cyst (Fig. 4G). Pseudotime analysis positioned the nodule-like population at the greatest distance from the cyst, suggesting that this transcriptional state may represent the transition from the peripheral to the cyst-bound state (Fig. 4H). We identified 107 transcription factors that vary on this trajectory, including those that peak at transitions between nuclei states, such as *HDG8* and *GRF2* (Extended Data 8, Supplementary Table 5).

Discrete families of short, secreted peptides are enriched in endosperm subtypes

SSPs are abundant in plant genomes and play roles in reproduction, development, and innate immunity[81–86]. They are characterized by an N-terminal secretory signal sequence and are less than 250 amino acids long (Fig. 5A)[87–89]. Arabidopsis seeds devote more of their transcriptomes to genes that encode SSPs than any other tissue, but little is known about the functions of seed-specific SSPs (Extended Data 9). This is in part due to their absence from existing stage-resolved transcriptional atlases of seed development. For example, the Belmonte 2013 atlas is based on ATH1 microarray data, which only had probes for ~50% of SSP genes in the Araport11 reference (Extended Data 9)[17].

DE analysis of all Arabidopsis SSPs revealed that defensin-like (DEFL) peptides, low molecular weight cysteine-rich (LCR) peptides, and lipid transfer proteins (LTP), are the three predominant SSP families in the atlas (Fig. 5B, Supplementary Table 9). Although most have not been functionally characterized, members of both families have been implicated in defense and signaling; some play roles in pollen-pistil interactions[84]. We found that the 3 DAP MCE, 3 DAP CZE, and 5 DAP CZE are hubs of SSP expression, both by expression level and number of unique SSPs expressed (Fig. 5C, D). The SSPs enriched in the MCE and CZE belong to

families with characterized roles in cell-cell signaling (*TPD*, *CLE*, *LCR*), as well as inhibitory
286 roles (*PMEI*)[90–93]. Furthermore, many exhibit upregulation after fertilization (Fig. 5C, Methods).
SSP enrichment and diversity in the CZE and MCE are compelling from a signaling point of view
288 because these regions are important interfaces: the CZE is a gateway for maternal resources
into the seed and the MCE is the most embryo-proximal seed tissue (Fig. 1A).

290 Based on their involvement in male/female and host/pathogen interactions, some SSP
families are expected to be rapidly evolving, which coincides with observations that the CZE and
292 MCE are enriched for expression of genes under diversifying selection[94–96]. We hypothesized
that SSPs contribute to this pattern. We calculated the dN/dS, a measure of the direction of
294 natural selection, for every gene in the Arabidopsis genome using orthologs from *Arabidopsis*
lyrata in pairwise comparisons and the Nei-Gojobori 1986 method (Methods, Supplementary
296 Table 7)[97]. Overall, dN/dS is low genome-wide (mean dN/dS = 0.22). However, a comparison of
mean dN/dS of all L2 markers showed that genes specifically expressed in the CZE, MCE, and
298 PEN are under weaker purifying selection (Fig. 6A). This pattern follows the temporal
enrichment of SSP expression (Fig. 5C). We compared the dN/dS of genes DE in the 3-5 DAP
300 MCE and CZE with all other genes and SSPs in Arabidopsis and observed that MCE and CZE-
specific SSPs have disproportionately higher dN/dS than MCE and CZE-specific genes when
302 compared to all other genes (Fig. 6B). To rigorously test whether seed SSPs are under positive
selection, we used codon-substitution site models of positive, negative, and neutral selection
304 implemented in the codeml program in the PAML package to calculate the likelihood of positive
selection for all protein coding sites in the Arabidopsis genome (Methods). This analysis
306 produces a likelihood ratio test statistic (LRT) for each gene which indicates the goodness-of-fit
of its phylogeny to a positive (M2a) or nearly neutral (M1a) model of selection. We generated
308 LRTs for 22,307 genes and found that 63 seed SSPs have statistically high M2a/M1a LRT, only
10 of which are DE in the CZE (Supplementary Table 7). Thus, a unique set of sequence-
310 divergent SSPs are concentrated in the CZE and MCE, but the majority of them are
experiencing weak or diffuse selection.

312

Protein degradation and nucleic acid-associated genes show evidence of positive
314 **selection in the chalazal endosperm**

Of the ~1300 genes that have significant M2a/M1a LRTs ("M2a/M1a-sig") in the
316 Arabidopsis genome, 316 and 247 of them are DE ($\log_2$FC > 1, adj. pval <0.05) in the 3-5 DAP
PEN and CZE, respectively, and 155 are shared (Supplementary Table 7). M2a/M1a-sig genes
318 show the highest expression levels in the 3-5 DAP CZE compared to all other seed cell types

(Fig. 6C). Within the CZE, apical subtypes show greater M2a/M1a-sig enrichment than basal, and the 5 DAP apical cyst shows the highest enrichment overall (Fig. 6C). A GO term analysis of the M2a/M1a-sig genes DE in the CZE identified 33 GO terms, most of which could be grouped into two broad categories: protein degradation and nucleic acid-associated (Fig. 6D, Supplementary Table 7). All genes underlying the protein degradation GO terms encoded either E3 ligases from the ARIADNE family, ubiquitin ligases/hydrolases, or F-box proteins (Fig. 6E), and have similar M2a/M1a LRTs (Fig. 6F). Sites with strong evidence for positive selection (Bayes Empirical Bayes (BEB) > 0.95) in the *ARIADNE* family were concentrated in either an N-terminal intrinsically disordered region (IDR) or In Between Ring (IBR) finger domain (Fig. 6G). The list of M2a/M1a-sig nucleic acid-associated genes is more heterogeneous than that for protein degradation, although many of them are involved in nucleic acid binding and modification. A subset of nucleic acid-associated genes showed strong evidence for positive selection within or proximal to DNA interacting domains, mostly in intrinsically disordered subdomains (Fig. 6H). Taken together, a defining feature of the CZE is its expression of genes with evolutionary novelty, both through weaker purifying selection (dN/dS) and likelihood of positive selection (M2a/M1a LRTs). Furthermore, a subset of genes involved in CZE-specific functions of protein degradation and DNA or RNA binding appear to be rapidly evolving (Extended Data 10).

Discussion

We present a comprehensive atlas of early seed development that illuminates aspects of cell-cell signaling, functional compartmentalization, and gene diversification in transcriptionally distinct cell or nuclei types. This adds to a growing compendium of single cell or nuclei transcriptional atlases for various stages of development in Arabidopsis and other plant species[98–103]. We revealed additional insights into nutrient transport within the seed as well as sites of callose and brassinosteroid biosynthesis. Our high coverage of endosperm nuclei allowed for the identification of a rare nuclei population in the CZE, which clarifies the organization of the CZE cyst. Furthermore, this atlas will enable the identification of promoters with high spatial and temporal specificity and will serve as a community resource for seed research.

Our study revealed additional complexity within the chalazal endosperm. We propose that *RALFL3*+ CZE nuclei may be the “founder” CZE population to which subsequent nuclei fuse to create the cyst, generating an apical-basal gradient of gene expression at early timepoints. Notably, cytological features following this polarity have been observed in the CZE

cyst: mitochondria and thylakoid stacks are enriched in the basal and apical regions,
354 respectively[104]. The transcriptional similarity of the basal cyst with the CPT and ii1 suggests congruence or coordination with maternal tissue transcriptomes. Interestingly, “tentacle-like
356 processes” embedded in maternal tissues have been described at the base of the Arabidopsis cyst[104] and in other species the chalazal endosperm takes on haustorial properties, which could
358 facilitate such coordination[16]. How the *RALFL3*+ nuclei are established at the chalazal pole, and the extent to which its DE SSPs contribute to this process, remains to be studied.

360 Prior to this work, we had only partial understanding of the extent of SSP expression and diversity in seed cell types. Although the MCE is a well-appreciated source of signaling SSPs,
362 most of the MCE/CZE-specific SSPs were not assayed in existing transcriptional seed atlases, which might have led to an underestimation of the contribution of SSPs to seed development.
364 Some of the MCE/CZE SSPs that are also expressed in the ovule could function in fertilization, but the majority are more highly expressed after fertilization, suggesting alternative functions.
366 Many of the seed SSPs are annotated as defensins. This class of cysteine-rich peptides is typically thought of as acting as anti-microbial peptides in seeds, particularly against fungi[105].
368 However, few seed-expressed defensins have been specifically evaluated for anti-microbial activity. It is possible that defensins, and other seed SSPs, act as ligands that are perceived by -
370 receptor-like kinases or receptor-like proteins to activate a variety of signaling pathways. Although MCE/CZE-specific SSPs remain enigmatic, their expression specificity prompts follow-
372 up investigation to elucidate their targets in the seed. The atlas will further allow the evaluation of the expression of potential receptors in various seed compartments. The enrichment for SSP
374 expression at the embryo-endosperm and maternal-offspring interfaces is consistent with cell non-autonomous functions, although this remains to be demonstrated for individual peptides.

376 Although SSPs in endosperm subtypes exhibit relatively high dN/dS compared to other genes, they exhibit a low likelihood of site-specific selection based on maximum likelihood
378 analysis, which suggests that selection pressure on SSPs is relaxed or diffused within subsets of SSP families. However, many other differentially expressed genes in the seed have
380 significantly high M2a/M1a LRTs, and the PEN and CZE appear to be hotspots for their expression. A subset of nucleic acid- and protein degradation-associated genes differentially
382 expressed in the 5 DAP apical cyst contain sites under positive selection. We focused on *ARIADNE* genes, a class of E3 ubiquitin ligase, and nucleic acid-associated gene groups for
384 this analysis. We observed accumulation of positively selected sites in intrinsically disordered domains in both groups, likely due in part to the lack of structural constraint on these
386 domains[106,107]. The structural flexibility of IDRs allows them to engage in diverse interactions,

and as a result, they are often implicated in cell signaling and gene expression regulation.
388 However, it is unclear to what extent the rapid evolution observed in these IDRs is due to their putative functions or lack of structural constraint[108]. Positively selected sites in *ARIADNE* genes
390 that are not in IDRs occurred in the TRIAD domain, which is comprised of RING finger domains that mediate protein-protein interactions. Within the nucleic acid associated gene list, we found
392 positively selected sites in regions that interact with RNA and DNA, leading to our hypothesis that rapid evolution at these sites may reflect substrate and binding partner diversity. Functional
394 studies are needed to discern whether these genes underlie the roles and diverse morphologies of the CZE[109].

396 Taken together, we provide a transcriptional atlas with a high-resolution annotation that will serve as the basis for future studies of early seed development. To this end, we have
398 created an online resource for exploring the data at https://seedatlas.wi.mit.edu/.

400 **Methods**

402 **Generation of single nucleus transcriptomes**

404 *Plant material*

406 All plants (*Arabidopsis thaliana* (L.) Heynh., Col-0) used in HCR RNA-FISH and snRNA-seq experiments were grown at 22 ºC in a glasshouse on soil under long-day conditions (16 h light:
408 8 h dark). For timed crosses, we emasculated flower buds and 2 days later pollinated stigma that had mature papillae.

410

Tissue preparation and nuclei extraction for snRNA-seq

412

Two biological replicates, (different plants hand-pollinated and processed for snRNA-seq on
414 different days) each containing seeds isolated from 10-15 siliques (500-800 seeds), were collected for each timepoint, producing six replicates total. Seeds were dissected into 150 μL
416 cold extraction buffer on ice (1x Partec CyStain UV Precise P nuclei extraction buffer (Sysmex #05-5002), 4% BSA, 1mM DTT, 1:100 protease inhibitor cocktail for plants (Sigma #P9599),
418 and 1:30 Protector RNase inhibitor (Fisher Scientific #NC1877809)). Seeds were mechanically dissociated in 1.5 mL tubes by grinding with an Axygen pestle (Corning #PES-15-B-SI) for 10
420 turns. The nuclei suspension was filtered through a 30 μm cell strainer (Fisher Scientific #

NC9682496), prewet with Partec CyStain UV Precise P staining buffer (Sysmex #05-5002), into
422 a 5 mL tube for fluorescence activated nuclei sorting (FANS). The strainer was rinsed with
Partec CyStain UV Precise P staining buffer into the 5 mL tube to bring the final suspension
424 volume to 1 mL. 1 μL of 1mg/mL DAPI (ThermoFisher Scientific #62248) was added to increase
nuclear signal. Nuclei were purified and concentrated by FANS on a BD FACS ARIA Cell Sorter
426 using a 70 μm nozzle chip. We gated on 2C, 3C, 4C, 6C, 8C, and 16C peaks (Supplementary
Fig. 5). Nuclei were sorted into 30-50 μL collection buffer (PBS-4% BSA) in a 1.5 mL tube and
428 concentration was assessed using the ARIA nuclei count and by manual counting on a
Neubauer Improved hemocytometer. Each biological replicate was processed individually using
430 one Chromium Next GEM Single Cell 3' Kit (v3.1) reaction with a target nuclei recovery of 10k,
except for one 3 DAP biological replicate, which was split across two reactions. Libraries were
432 sequenced on a NovaSeqS2 with 50x50 paired end reads. In total, 7 libraries were generated in
this study.
434

Computational analysis of single-nucleus transcriptomes

436

Raw data pre-processing, integration and clustering

438

Raw sequencing data were processed using Cell Ranger v7.1.0 (10X Genomics). The
440 *Arabidopsis thaliana* TAIR10 genome sequence (Athaliana_447_TAIR10.fa) and the Araport11
annotation (Araport11_GFF3_genes_transposons.filtered.201606.gtf)[110] were used as inputs to
442 "cellranger mkref" v6.0.02. "cellranger count" was used with default parameters and STAR
2.7.1a.
444

All scripts for the following analysis steps are deposited in the Gehring Lab GitHub:
446 https://github.com/Gehring-Lab/seed_atlas_2025.

448 The "filtered_feature_bc_matrix" for each library was individually preprocessed with the
"per_library_QCs.R" script, which implements Seurat v5.0.0 for object generation and SoupX
450 v1.6.2 for background RNA correction[111,112]. Genes detected in less than 10 nuclei were
removed, and profiles with less than 250 genes filtered out. Following the recommendation from
452 Heumos et al. 2023, we further identified and removed "outlier" nuclei as those with a
gene/nucleus or a transcript/nucleus that differed by 5 median absolute deviations from the rest
454 of the sample[113]. We then performed an initial clustering analysis and removed clusters that had

significantly lower genes/nucleus than the mean. After low-quality profile removal, we ran
456 scDblFinder v1.12.0 on each library in random mode to generate a per-nucleus doublet score,
and removed putative doublets based on the recommended threshold generated by
458 scDblFinder, outlined in the "score_doublets.R" script[114].

460 The preprocessed libraries were then merged and further quality controlled, outlined in the
"merge_libraries_and_batch_effects_harmony.R" script, which is executed in two rounds. In the
462 first round, libraries are merged by timepoint and an initial clustering analysis is performed to
generate cluster-level QCs. Additionally, the JackStraw and ScoreJackStraw functions in Seurat
464 are used to detect statistically significant principal components (PCs) for downstream
analysis[112]. Then, the clusters are visually assessed for batch effects. In the second round, we
466 perform a final dimensionality reduction involving log normalization followed by a principal
component analysis using the number of PCs determined by JackStraw analysis, k-nearest
468 neighbor graph construction, and UMAP projection. Timepoints that exhibited clustering by
biological replicate were integrated with Harmony 1.2.0, which was required for 5 DAP
470 (Supplementary Fig. 6)[115]. At this step, cell cycle genes from Menges et al. 2005 were used to
categorize nuclei into G1/0, S, and M-phases using the CellCycleScoring function in Seurat[116].
472
To identify an appropriate initial clustering resolution for each timepoint, we performed a
474 parameter sweep of the Seurat FindClusters function, varying the resolution parameter from 1 to
2 in increments of 0.1 (see the "clustering.R" script). For each resolution, an average cluster
476 purity and silhouette score was calculated using the neighborPurity and approxSilhouette
bluster 1.8.0 functions, respectively[117]. For each timepoint, we selected a resolution that
478 preceded the greatest drop in neighborhood purity on a resolution vs. average neighborhood
purity plot. This clustering, referred to as the "*de novo"* clustering, served as the basis for the
480 level 3 annotation (L3), the highest annotation resolution (Fig. 1E, Supplementary Fig. 1,
Supplementary Table 3).
482
Cluster annotation
484
Using both published markers (Supplementary Table 1) and markers identified in this study with
486 HCR-RNA FISH validation (Supplementary Table 2), we classified all clusters, giving each L3 a
full name that includes a number, descriptor, and most informative marker gene(s)
488 (Supplementary Table 3, see "manual_annotation.R" script). *RALFL3*+ nuclei, localized to the

CZE cyst by HCR RNA-FISH, were manually annotated based on *RALFL3* expression in the 3-5 DAP datasets, and an endosperm-only clustering analysis revealed nodule and nodule-like clusters at 3 and 5 DAP. We appended the nodule and nodule-like annotations to the full 3-5 DAP datasets (Supplemental Fig. 2-3). Clusters with negligible differential expression differences in the oi1 were merged into one cluster. We annotated timepoints separately and integrated them into a final atlas dataset using robust principal component analysis (RPCA), see "atlas_merging_rpca.R" script[112].

GO term and gene list GSEA and statistical testing

Based on GO term analysis for all DE genes for each atlas cluster across all annotation levels using clusterProfiler 4.7.1.002 (see "clusterprofiler_intermediate.R" script), we identified gene sets for follow-up enrichment analysis[118]. We retrieved gene lists for GO terms from uniprot (https://www.uniprot.org/) using GO term IDs filtered by the *Arabidopsis thaliana* taxon ID (taxonomy_id:3702). Gene lists were used as input to the Seurat AddModuleScore function with default settings (number of bins, "nbin" = 24; number of control genes per bin, "nctrl" = 100), which implements the gene set enrichment approach described in Tirosh 2016[119][112]. The resulting "module score" is the difference in average expression between the gene set of interest and a randomly-generated gene set with matched expression level variability. All gene lists are deposited with the scripts used for generating scores, which include "signalling_transport_gene_enrichment.R", "peptide_enrichment.R", and "protein_catabolism_enrichment.R". The same approach was used for curated gene lists of SSPs and BZR/BES1 transcription factors.

To statistically test module scores by cluster, we compared the mean module scores of nuclei split by biological replicate against the mean of all other nuclei split by biological replicate in a Student's T-test. After performing this test for all clusters for a given module score, we adjusted p-values using Benjamini-Hochberg correction. The code for statistically testing module scores is at the end of the "protein_catabolism_enrichment.R" script. All p-values indicated in the main figures are from statistical analyses using the integrated atlas dataset (6 replicates) in which each timed L2 cluster was statistically tested for enrichment.

Spearman correlation analysis of pseudobulked clusters

To quantify similarity between cell types, we pseudobulked clusters using the Seurat function
524 AggregateExpression using default arguments (normalization.method = “LogNormalize”, scale.factor = 10000) and only the top 4000 variable genes from the integrated atlas dataset[112].
526 We used Spearman correlation to generate pairwise ρ for all clusters.

528 *Pseudotime analysis*

530 To order the 5 DAP PEN/CZE and the 3-7 DAP MCE datasets by pseudotime, we integrated all samples by timepoint using Harmony, if necessary, and performed another dimensionality
532 reduction analysis on the subset, regressing out cell cycle genes during data scaling and centering[115]. To identify only one trajectory, we assigned a single partition to all nuclei in each
534 dataset, and cells were ordered and assigned a pseudotime value using the “learn_graph” and “order_cells” functions in monocle3 1.3.7[120]. See “level_2_pseudotime_merged_timepoint.R”
536 and “harmony_chalazal_endosperm_trajectory.R” scripts for details[120].

538 *Identifying differentially expressed genes*

540 Differential expression analysis was performed on each timepoint and the integrated atlas object using the Seurat FindAllMarkers function with default arguments, generating cluster vs. all other
542 nuclei results for each gene and p-values from a Wilcoxon Rank Sum test, see “differential_expression.R” script for details. To determine whether a gene is upregulated after
544 fertilization (Fig. 5C), we used the published bulk RNA-sequencing data from fertilized and unfertilized ovules in Zhang et al. 2023[121]. We calculated the $\log_2$ fold change (Log_2FC) of TPM
546 values for all genes between the fertilized and unfertilized conditions. Genes with a Log_2FC greater than 1 in the fertilized condition were classified as upregulated.
548

<u>Marker validation using HCR RNA-FISH</u>

550 To localize clusters and marker genes *in situ*, we used HCR-RNA FISH based on a modified version of the whole mount HCR protocol outlined in Huang 2023 and the general
552 HCR-RNA FISH protocol developed by Molecular Instruments: https://www.molecularinstruments.com/hcr-gold-rnafish[122]. All HCR probes and fluorescent
554 hairpins were synthesized by Molecular Instruments (Supplemental Table 2). Hand-pollinated or unstaged siliques were collected and fixed with 4% paraformaldehyde (PFA) in 1X PBS through
556 vacuum infiltration. After overnight fixation at 4C, samples were washed in 1x PBS twice before

embedding in 3% agarose for vibratome sectioning. Whole siliques were sectioned
558 longitudinally to 60-150 μm thickness and stored in PBS on ice. Sections were subject to a
second 4% PFA fixation for 30 minutes at room temperature before two washes in PBS, then
560 transferred to 100% methanol and stored at -20 ºC. To permeabilize tissue before probe
hybridization, samples were subject to alternating ethanol and methanol incubations, following
562 Huang 2023, with a clearing step using 50% Histoclear (National Diagnostics #HS-200)/50%
ethanol halfway through the permeabilization washes[122]. Samples were rehydrated with an
564 ethanol/PBS-Tween-20 series (25%, 50%, 75%, 100%). Then, samples were incubated in
preheated Probe Hybridization Buffer (Molecular Instruments) for 30 minutes before exchange
566 into preheated Probe Hybridization Buffer (Molecular Instruments) with probes. For targets with
< 5 probe sets, we used 64 μL of 1 μM probe in 400 μL Probe Hybridization Buffer, and for all
568 other probes we used 24 μL probe 1 μM in 400 μL. Probes were hybridized to samples in 1.5mL
tubes overnight in a 37 ºC water bath. Probe washes closely followed the Molecular Instruments
570 HCR protocol: we performed four 15 minute buffer exchanges with Probe Wash Buffer
(Molecular Instruments) preheated to 37 ºC in a water bath, followed by two quick washes with
572 5X SSC-Tween-20. Samples were incubated in Amplification Buffer (Molecular Instruments) for
30 minutes at room temperature before exchange to Amplification Buffer containing snap-cooled
574 hairpins (8 μL of 3 μM stock per hairpin per 400 μL reaction volume). We used hairpins
containing B2 and B3 adapter sequences with Alexa-488 or Alexa-647 dyes (Molecular
576 Instruments). After an overnight incubation in the dark, excess probes were washed off with four
5x SSC-Tween-20 exchanges. Samples were stored for up to two weeks at 4 ºC before
578 imaging.

To prepare samples for confocal microscopy, we counterstained nuclei with 1μg/mL
580 DAPI and mounted silique sections on thin 2% agar pads on glass slides, suspended in water.
We imaged Z-stacks of samples on a Zeiss LSM 710 confocal and generated maximum
582 intensity projection images using FIJI.

584 **SSP detection and annotation**

To detect all SSPs in the Arabidopsis genome, we filtered all Araport11 protein
586 sequences for those less than 250 residues and used this as input to SignalP6 to identify those
with N-terminal secretory signal sequences, see “signalp6_command.sh” for details. We
588 transferred the SSP annotations described in Ghorbani et al 2015 to the SSPs detected by
SignalP6[87,123]. For all SSPs detected by SignalP6 expressed in the atlas that did not have an

590 annotation from Ghorbani et al 2015, we used the Araport11 annotation categorize these sequences into SSP families (Supplementary Table 9).

592 To characterize the compartmentalization and expression dynamics of SSPs in seed cell types, we first produced an updated list of all Arabidopsis SSPs using SignalP6 to detect <=250
594 amino acid proteins with N-terminal secretory signal sequences in the Araport11 reference proteome[123]. We then assigned all SSPs detected in the atlas to an SSP family, if possible,
596 combining the signaling SSP annotation published in Ghorbani et al. 2015 with Araport11 gene annotations for SSPs in families implicated in inhibitory, storage, and defense roles[87,110].

598

Genome-wide evolutionary analysis

600

dN/dS calculations between A. thaliana *and* A. lyrata

602

To calculate dN/dS for all genes in the *Arabidopsis thaliana* genome, we identified the
604 reciprocal BLAST best hit for all coding sequences (CDSs) in the Araport11 annotation against all coding sequences in the *Arabidopsis lyrata* reference (Phytozome v2.1)[124,125]. dN/dSs
606 between 1:1 orthologs were estimated using the Nei and Gojobori 1986 method implemented by the orthologr 0.4.2 R package[126,97]. See the "orthologr.R" script for all details. *A. thaliana* genes
608 without orthologs or with dS < 0.01 were removed from further analysis.

610 *Maximum likelihood estimation of positive selection at sites in all* A. thaliana *genes*

612 To implement site models of codon evolution using codeml/PAML 4.9[127], we followed the procedure outlined in the Bioinformatics Workbook[128], based on the analyses in Petersen et al.
614 2007[129,130]. We identified orthologs between the translated coding sequences (CDS) of *Arabidopsis thaliana* (Araport11, Phytozome genome ID: 447*)*, *Arabidopsis lyrata (*v2.1,
616 Phytozome genome ID: 384*)*, *Capsella grandiflora* (v1.1, Phytozome genome ID: 266), and *Brassica rapa* (CAAS_Brap_v3.01, RefSeq GCF_000309985.2) using OrthoFinder 2.5.4[131].
618 The resulting ortholog gene trees were then pruned to retain the best ortholog from each species to the *Arabidopsis thaliana* ortholog (minimum distance from *Arabidopsis thaliana*). This
620 produced one gene tree for each *Arabidopsis thaliana* gene that contained one ortholog from at least two other species. The protein sequences from gene trees were aligned with clustalo[132].
622 The protein alignments and CDSs from each gene tree were used as input to the pal2nal.pl script[133] to produce codon alignments. Codon alignments and pruned gene trees were used as

input to codeml, with arguments for performing maximum likelihood estimation of site models of codon evolution that remove all alignment sites with gaps (runmode = 0, seqtype = 1, CodonFreq = 2, model = 0, NSsites = 0 1 2 7 8, fix_kappa = 0, kappa = 2, fix_omega = 0, omega = 0.4, cleandata = 1). We calculated both M2a/M1a and M8/M7 LRTs ($2*(\ln L_{alt} - \ln L_{null})$), but proceeded with M2a/M1a for higher stringency. See “orthofinder_to_codeml.sh” script for more details. LRTs were significance tested using the chi square test function pchisq in R. All p-values were adjusted within M2a/M1a and M8/M7 comparisons using Benjamini-Hochberg correction. The Bayes Empirical Bayes results from M2a/M1a analyses for individual sites were extracted from codeml outputs and evaluated only for genes with significant M2a/M1a LRTs.

Data Availability

All sequencing data is available in NCBI GEO GSE295007. Scripts for analyses are deposited in GitHub at https://github.com/Gehring-Lab/seed_atlas_2025. A browser for the data is available at https://seedatlas.wi.mit.edu/

638 **References**

(1) Nowack, M. K.; Ungru, A.; Bjerkan, K. N.; Grini, P. E.; Schnittger, A. Reproductive Cross-
640 Talk: Seed Development in Flowering Plants. *Biochem. Soc. Trans.* **2010**, *38* (2), 604–612. https://doi.org/10.1042/BST0380604.
642 (2) Berger, F. Endosperm: The Crossroad of Seed Development. *Curr. Opin. Plant Biol.* **2003**, *6* (1), 42–50. https://doi.org/10.1016/S1369526602000043.
644 (3) Boisnard-Lorig, C.; Colon-Carmona, A.; Bauch, M.; Hodge, S.; Doerner, P.; Bancharel, E.; Dumas, C.; Haseloff, J.; Berger, F. Dynamic Analyses of the Expression of the
646 HISTONE::YFP Fusion Protein in Arabidopsis Show That Syncytial Endosperm Is Divided in Mitotic Domains. *Plant Cell* **2001**, *13* (3), 495–510.
648 (4) Becker, M. G.; Hsu, S.-W.; Harada, J. J.; Belmonte, M. F. Genomic Dissection of the Seed. *Front. Plant Sci.* **2014**, *5*. https://doi.org/10.3389/fpls.2014.00464.
650 (5) Khan, D.; Millar, J. L.; Girard, I. J.; Belmonte, M. F. Transcriptional Circuitry Underlying Seed Coat Development in Arabidopsis. *Plant Sci. Int. J. Exp. Plant Biol.* **2014**, *219–220*,
652 51–60. https://doi.org/10.1016/j.plantsci.2014.01.004.
(6) Ingram, G. C. Family Life at Close Quarters: Communication and Constraint in Angiosperm
654 Seed Development. *Protoplasma* **2010**, *247* (3–4), 195–214. https://doi.org/10.1007/s00709-010-0184-y.
656 (7) Stadler, R.; Lauterbach, C.; Sauer, N. Cell-to-Cell Movement of Green Fluorescent Protein Reveals Post-Phloem Transport in the Outer Integument and Identifies Symplastic Domains
658 in Arabidopsis Seeds and Embryos. *Plant Physiol.* **2005**, *139* (2), 701–712. https://doi.org/10.1104/pp.105.065607.
660 (8) Doll, N. M.; Ingram, G. C. Embryo-Endosperm Interactions. *Annu. Rev. Plant Biol.* **2022**, *73*, 293–321. https://doi.org/10.1146/annurev-arplant-102820-091838.
662 (9) Robert, H. S.; Park, C.; Gutièrrez, C. L.; Wójcikowska, B.; Pěnčík, A.; Novák, O.; Chen, J.; Grunewald, W.; Dresselhaus, T.; Friml, J.; Laux, T. Maternal Auxin Supply Contributes to
664 Early Embryo Patterning in Arabidopsis. *Nat. Plants* **2018**, *4* (8), 548–553. https://doi.org/10.1038/s41477-018-0204-z.
666 (10) Figueiredo, D. D.; Köhler, C. Auxin: A Molecular Trigger of Seed Development. *Genes Dev.* **2018**, *32* (7–8), 479–490. https://doi.org/10.1101/gad.312546.118.
668 (11) Figueiredo, D. D.; Batista, R. A.; Roszak, P. J.; Hennig, L.; Köhler, C. Auxin Production in the Endosperm Drives Seed Coat Development in Arabidopsis. *eLife* **2016**, *5*, e20542.
670 https://doi.org/10.7554/eLife.20542.
(12) Lafon-Placette, C.; Köhler, C. Embryo and Endosperm, Partners in Seed Development.
672 *Curr. Opin. Plant Biol.* **2014**, *17*, 64–69. https://doi.org/10.1016/j.pbi.2013.11.008.
(13) Xu, W.; Sato, H.; Bente, H.; Santos-González, J.; Köhler, C. Endosperm Cellularization
674 Failure Induces a Dehydration-Stress Response Leading to Embryo Arrest. *Plant Cell* **2023**, *35* (2), 874–888. https://doi.org/10.1093/plcell/koac337.
676 (14) Song, J.; Xie, X.; Chen, C.; Shu, J.; Thapa, R. K.; Nguyen, V.; Bian, S.; Kohalmi, S. E.; Marsolais, F.; Zou, J.; Cui, Y. LEAFY COTYLEDON1 Expression in the Endosperm Enables
678 Embryo Maturation in Arabidopsis. *Nat. Commun.* **2021**, *12* (1), 3963. https://doi.org/10.1038/s41467-021-24234-1.
680 (15) Brown, R. C.; Lemmon, B. E.; Nguyen, H.; Olsen, O.-A. Development of Endosperm in Arabidopsis Thaliana. *Sex. Plant Reprod.* **1999**, *12* (1), 32–42.
682 https://doi.org/10.1007/s004970050169.
(16) Picard, C. L.; Povilus, R. A.; Williams, B. P.; Gehring, M. Transcriptional and Imprinting
684 Complexity in Arabidopsis Seeds at Single-Nucleus Resolution. *Nat. Plants* **2021**, *7* (6), 730–738. https://doi.org/10.1038/s41477-021-00922-0.

(17) Belmonte, M. F.; Kirkbride, R. C.; Stone, S. L.; Pelletier, J. M.; Bui, A. Q.; Yeung, E. C.; Hashimoto, M.; Fei, J.; Harada, C. M.; Munoz, M. D.; Le, B. H.; Drews, G. N.; Brady, S. M.; Goldberg, R. B.; Harada, J. J. Comprehensive Developmental Profiles of Gene Activity in Regions and Subregions of the Arabidopsis Seed. *Proc. Natl. Acad. Sci.* **2013**, *110* (5), E435–E444. https://doi.org/10.1073/pnas.1222061110.

(18) van Ekelenburg, Y. S.; Hornslien, K. S.; Van Hautegem, T.; Fendrych, M.; Van Isterdael, G.; Bjerkan, K. N.; Miller, J. R.; Nowack, M. K.; Grini, P. E. Spatial and Temporal Regulation of Parent-of-Origin Allelic Expression in the Endosperm. *Plant Physiol.* **2023**, *191* (2), 986–1001. https://doi.org/10.1093/plphys/kiac520.

(19) Doll, N. M.; Royek, S.; Fujita, S.; Okuda, S.; Chamot, S.; Stintzi, A.; Widiez, T.; Hothorn, M.; Schaller, A.; Geldner, N.; Ingram, G. A Two-Way Molecular Dialogue between Embryo and Endosperm Is Required for Seed Development. *Science* **2020**, *367* (6476), 431–435. https://doi.org/10.1126/science.aaz4131.

(20) Royek, S.; Bayer, M.; Pfannstiel, J.; Pleiss, J.; Ingram, G.; Stintzi, A.; Schaller, A. Processing of a Plant Peptide Hormone Precursor Facilitated by Posttranslational Tyrosine Sulfation. *Proc. Natl. Acad. Sci.* **2022**, *119* (16), e2201195119. https://doi.org/10.1073/pnas.2201195119.

(21) Millar, J. L.; Khan, D.; Becker, M. G.; Chan, A.; Dufresne, A.; Sumner, M.; Belmonte, M. F. Chalazal Seed Coat Development in *Brassica Napus*. *Plant Sci.* **2015**, *241*, 45–54. https://doi.org/10.1016/j.plantsci.2015.09.019.

(22) Pegler, J. L.; Grof, C. P.; Patrick, J. W. Sugar Loading of Crop Seeds – a Partnership of Phloem, Plasmodesmal and Membrane Transport. *New Phytol.* **2023**, *239* (5), 1584–1602. https://doi.org/10.1111/nph.19058.

(23) Liu, H.; Luo, Q.; Tan, C.; Song, J.; Zhang, T.; Men, S. Biosynthesis- and Transport-Mediated Dynamic Auxin Distribution during Seed Development Controls Seed Size in Arabidopsis. *Plant J. Cell Mol. Biol.* **2023**, *113* (6), 1259–1277. https://doi.org/10.1111/tpj.16109.

(24) Chen, X.; Bezodis, W.; González-Suárez, P.; Knitlhoffer, V.; Goldson, A.; Lister, A.; Macaulay, I.; Penfield, S. Trans-Generational Adaptation to Maternal Climate through Hormone Transport in Plants. bioRxiv October 5, 2024, p 2024.10.04.616646. https://doi.org/10.1101/2024.10.04.616646.

(25) Kao, P.; Schon, M. A.; Mosiolek, M.; Enugutti, B.; Nodine, M. D. Gene Expression Variation in Arabidopsis Embryos at Single-Nucleus Resolution. *Dev. Camb. Engl.* **2021**, *148* (13), dev199589. https://doi.org/10.1242/dev.199589.

(26) Capron, A.; Chatfield, S.; Provart, N.; Berleth, T. Embryogenesis: Pattern Formation from a Single Cell. *Arab. Book* **2009**, *7*, e0126. https://doi.org/10.1199/tab.0126.

(27) Wysocka-Diller, J. W.; Helariutta, Y.; Fukaki, H.; Malamy, J. E.; Benfey, P. N. Molecular Analysis of SCARECROW Function Reveals a Radial Patterning Mechanism Common to Root and Shoot. *Development* **2000**, *127* (3), 595–603. https://doi.org/10.1242/dev.127.3.595.

(28) Welch, D.; Hassan, H.; Blilou, I.; Immink, R.; Heidstra, R.; Scheres, B. Arabidopsis JACKDAW and MAGPIE Zinc Finger Proteins Delimit Asymmetric Cell Division and Stabilize Tissue Boundaries by Restricting SHORT-ROOT Action. *Genes Dev.* **2007**, *21* (17), 2196–2204. https://doi.org/10.1101/gad.440307.

(29) Willemsen, V.; Bauch, M.; Bennett, T.; Campilho, A.; Wolkenfelt, H.; Xu, J.; Haseloff, J.; Scheres, B. The NAC Domain Transcription Factors FEZ and SOMBRERO Control the Orientation of Cell Division Plane in *Arabidopsis* Root Stem Cells. *Dev. Cell* **2008**, *15* (6), 913–922. https://doi.org/10.1016/j.devcel.2008.09.019.

(30) Zhang, T.; Ge, Y.; Cai, G.; Pan, X.; Xu, L. WOX-ARF Modules Initiate Different Types of Roots. *Cell Rep.* **2023**, *42* (8), 112966. https://doi.org/10.1016/j.celrep.2023.112966.

(31) Hibara, K.; Takada, S.; Tasaka, M. CUC1 Gene Activates the Expression of SAM-Related Genes to Induce Adventitious Shoot Formation. *Plant J.* **2003**, *36* (5), 687–696. https://doi.org/10.1046/j.1365-313X.2003.01911.x.

(32) Ueda, M.; Zhang, Z.; Laux, T. Transcriptional Activation of *Arabidopsis* Axis Patterning Genes *WOX8/9* Links Zygote Polarity to Embryo Development. *Dev. Cell* **2011**, *20* (2), 264–270. https://doi.org/10.1016/j.devcel.2011.01.009.

(33) Su, Y. H.; Zhao, X. Y.; Liu, Y. B.; Zhang, C. L.; O'Neill, S. D.; Zhang, X. S. Auxin-Induced WUS Expression Is Essential for Embryonic Stem Cell Renewal during Somatic Embryogenesis in Arabidopsis. *Plant J. Cell Mol. Biol.* **2009**, *59* (3), 448–460. https://doi.org/10.1111/j.1365-313X.2009.03880.x.

(34) Beeckman, T.; De Rycke, R.; Viane, R.; Inzé, D. Histological Study of Seed Coat Development in Arabidopsis Thaliana. *J. Plant Res.* **2000**, *113* (2), 139–148. https://doi.org/10.1007/PL00013924.

(35) Francoz, E.; Ranocha, P.; Pernot, C.; Le Ru, A.; Pacquit, V.; Dunand, C.; Burlat, V. Complementarity of Medium-Throughput in Situ RNA Hybridization and Tissue-Specific Transcriptomics: Case Study of Arabidopsis Seed Development Kinetics. *Sci. Rep.* **2016**, *6*, 24644. https://doi.org/10.1038/srep24644.

(36) Mizzotti, C.; Ezquer, I.; Paolo, D.; Rueda-Romero, P.; Guerra, R. F.; Battaglia, R.; Rogachev, I.; Aharoni, A.; Kater, M. M.; Caporali, E.; Colombo, L. SEEDSTICK Is a Master Regulator of Development and Metabolism in the Arabidopsis Seed Coat. *PLoS Genet.* **2014**, *10* (12), e1004856. https://doi.org/10.1371/journal.pgen.1004856.

(37) Windsor, J. B.; Symonds, V. V.; Mendenhall, J.; Lloyd, A. M. Arabidopsis Seed Coat Development: Morphological Differentiation of the Outer Integument. *Plant J. Cell Mol. Biol.* **2000**, *22* (6), 483–493. https://doi.org/10.1046/j.1365-313x.2000.00756.x.

(38) Creff, A.; Brocard, L.; Ingram, G. A Mechanically Sensitive Cell Layer Regulates the Physical Properties of the Arabidopsis Seed Coat. *Nat. Commun.* **2015**, *6*, 6382. https://doi.org/10.1038/ncomms7382.

(39) Nakaune, S.; Yamada, K.; Kondo, M.; Kato, T.; Tabata, S.; Nishimura, M.; Hara-Nishimura, I. A Vacuolar Processing Enzyme, δVPE, Is Involved in Seed Coat Formation at the Early Stage of Seed Development. *Plant Cell* **2005**, *17* (3), 876–887. https://doi.org/10.1105/tpc.104.026872.

(40) Muller, B. Characterization of UmamiTs in Arabidopsis: Amino Acid Transporters Involved in Amino Acid Cycling, Phloem Unloading and the Supply of Symplasmically Isolated Sink Tissues. Doctoral defense, Universität Regensburg, 2016. https://epub.uni-regensburg.de/34834/1/Doktorarbeit%20Benedikt%20M%C3%BCller.pdf.

(41) Sanden, N. C. H.; Kanstrup, C.; Crocoll, C.; Schulz, A.; Nour-Eldin, H. H.; Halkier, B. A.; Xu, D. An UMAMIT-GTR Transporter Cascade Controls Glucosinolate Seed Loading in Arabidopsis. *Nat. Plants* **2024**, *10* (1), 172–179. https://doi.org/10.1038/s41477-023-01598-4.

(42) Vogiatzaki, E.; Baroux, C.; Jung, J.-Y.; Poirier, Y. PHO1 Exports Phosphate from the Chalazal Seed Coat to the Embryo in Developing *Arabidopsis* Seeds. *Curr. Biol.* **2017**, *27* (19), 2893-2900.e3. https://doi.org/10.1016/j.cub.2017.08.026.

(43) Li, M.; Li, Q.; Li, S.; Niu, X.; Xu, H.; Li, P.; Bian, X.; Chen, Z.; Liu, Q.; Zhang, H.; Liu, Y.; Wu, S. SHORT-ROOT Specifically Functions in the Chalazal Region to Modulate Assimilate Partitioning into Seeds. *Plant J. Cell Mol. Biol.* **2024**, *120* (5), 2031–2044. https://doi.org/10.1111/tpj.17096.

(44) Horák, J.; Grefen, C.; Berendzen, K. W.; Hahn, A.; Stierhof, Y.-D.; Stadelhofer, B.; Stahl, M.; Koncz, C.; Harter, K. The Arabidopsis Thaliana Response Regulator ARR22 Is a Putative AHP Phospho-Histidine Phosphatase Expressed in the Chalaza of Developing Seeds. *BMC Plant Biol.* **2008**, *8*, 77. https://doi.org/10.1186/1471-2229-8-77.

(45) Lu, J.; Magnani, E. Seed Tissue and Nutrient Partitioning, a Case for the Nucellus. *Plant Reprod.* **2018**, *31* (3), 309–317. https://doi.org/10.1007/s00497-018-0338-1.
(46) Xu, W.; Fiume, E.; Coen, O.; Pechoux, C.; Lepiniec, L.; Magnani, E. Endosperm and Nucellus Develop Antagonistically in Arabidopsis Seeds. *Plant Cell* **2016**, *28* (6), 1343–1360. https://doi.org/10.1105/tpc.16.00041.
(47) Lu, J.; Le Hir, R.; Gómez-Páez, D.-M.; Coen, O.; Péchoux, C.; Jasinski, S.; Magnani, E. The Nucellus: Between Cell Elimination and Sugar Transport. *Plant Physiol.* **2021**, *185* (2), 478–490. https://doi.org/10.1093/plphys/kiaa045.
(48) Xu, W.; Gomez-Paez, D.-M.; Choinard, S.; Iannaccone, M.; Maricchiolo, E.; Peaucelle, A.; Voxeur, A.; Haas, K. T.; Pompa, A.; Magnani, E. A Change in the Cell Wall Status Initiates the Elimination of the Nucellus in Arabidopsis. bioRxiv April 29, 2024, p 2024.04.23.590775. https://doi.org/10.1101/2024.04.23.590775.
(49) Veley, K. M.; Maksaev, G.; Frick, E. M.; January, E.; Kloepper, S. C.; Haswell, E. S. Arabidopsis MSL10 Has a Regulated Cell Death Signaling Activity That Is Separable from Its Mechanosensitive Ion Channel Activity. *Plant Cell* **2014**, *26* (7), 3115–3131. https://doi.org/10.1105/tpc.114.128082.
(50) Müller, B.; Fastner, A.; Karmann, J.; Mansch, V.; Hoffmann, T.; Schwab, W.; Suter-Grotemeyer, M.; Rentsch, D.; Truernit, E.; Ladwig, F.; Bleckmann, A.; Dresselhaus, T.; Hammes, U. Z. Amino Acid Export in Developing Arabidopsis Seeds Depends on UmamiT Facilitators. *Curr. Biol. CB* **2015**, *25* (23), 3126–3131. https://doi.org/10.1016/j.cub.2015.10.038.
(51) Chen, L.-Q.; Lin, I. W.; Qu, X.-Q.; Sosso, D.; McFarlane, H. E.; Londoño, A.; Samuels, A. L.; Frommer, W. B. A Cascade of Sequentially Expressed Sucrose Transporters in the Seed Coat and Endosperm Provides Nutrition for the Arabidopsis Embryo. *Plant Cell* **2015**, *27* (3), 607–619. https://doi.org/10.1105/tpc.114.134585.
(52) Pinto, S. C.; Leong, W. H.; Tan, H.; McKee, L.; Prevost, A.; Ma, C.; Shirley, N. J.; Petrella, R.; Yang, X.; Koltunow, A. M.; Bulone, V.; Kanaoka, M. M.; Higashyiama, T.; Coimbra, S.; Tucker, M. R. Germline β-1,3-Glucan Deposits Are Required for Female Gametogenesis in Arabidopsis Thaliana. *Nat. Commun.* **2024**, *15* (1), 5875. https://doi.org/10.1038/s41467-024-50143-0.
(53) Liu, X.; Nakajima, K. P.; Adhikari, P. B.; Wu, X.; Zhu, S.; Okada, K.; Kagenishi, T.; Kurotani, K.; Ishida, T.; Nakamura, M.; Sato, Y.; Kawakatsu, Y.; Xie, L.; Huang, C.; He, J.; Yokawa, K.; Sawa, S.; Higashiyama, T.; Bradford, K. J.; Notaguchi, M.; Kasahara, R. D. Fertilization-Dependent Phloem End Gate Regulates Seed Size. *Curr. Biol.* **2025**, *0* (0). https://doi.org/10.1016/j.cub.2025.03.033.
(54) Ponnu, J.; Wahl, V.; Schmid, M. Trehalose-6-Phosphate: Connecting Plant Metabolism and Development. *Front. Plant Sci.* **2011**, *2*, 70. https://doi.org/10.3389/fpls.2011.00070.
(55) Lunn, J. E.; Feil, R.; Hendriks, J. H. M.; Gibon, Y.; Morcuende, R.; Osuna, D.; Scheible, W.-R.; Carillo, P.; Hajirezaei, M.-R.; Stitt, M. Sugar-Induced Increases in Trehalose 6-Phosphate Are Correlated with Redox Activation of ADPglucose Pyrophosphorylase and Higher Rates of Starch Synthesis in Arabidopsis Thaliana. *Biochem. J.* **2006**, *397* (1), 139–148. https://doi.org/10.1042/BJ20060083.
(56) Eastmond, P. J. SUGAR-DEPENDENT1 Encodes a Patatin Domain Triacylglycerol Lipase That Initiates Storage Oil Breakdown in Germinating Arabidopsis Seeds. *Plant Cell* **2006**, *18* (3), 665–675. https://doi.org/10.1105/tpc.105.040543.
(57) Gómez, L. D.; Gilday, A.; Feil, R.; Lunn, J. E.; Graham, I. A. AtTPS1-Mediated Trehalose 6-Phosphate Synthesis Is Essential for Embryogenic and Vegetative Growth and Responsiveness to ABA in Germinating Seeds and Stomatal Guard Cells. *Plant J. Cell Mol. Biol.* **2010**, *64* (1), 1–13. https://doi.org/10.1111/j.1365-313X.2010.04312.x.

(58) Jiang, W.-B.; Huang, H.-Y.; Hu, Y.-W.; Zhu, S.-W.; Wang, Z.-Y.; Lin, W.-H. Brassinosteroid Regulates Seed Size and Shape in Arabidopsis. *Plant Physiol.* **2013**, *162* (4), 1965–1977. https://doi.org/10.1104/pp.113.217703.
(59) Cai, H.; Liu, L.; Huang, Y.; Zhu, W.; Qi, J.; Xi, X.; Aslam, M.; Dresselhaus, T.; Qin, Y. Brassinosteroid Signaling Regulates Female Germline Specification in Arabidopsis. *Curr. Biol. CB* **2022**, *32* (5), 1102-1114.e5. https://doi.org/10.1016/j.cub.2022.01.022.
(60) Liao, K.; Peng, Y.-J.; Yuan, L.-B.; Dai, Y.-S.; Chen, Q.-F.; Yu, L.-J.; Bai, M.-Y.; Zhang, W.-Q.; Xie, L.-J.; Xiao, S. Brassinosteroids Antagonize Jasmonate-Activated Plant Defense Responses through BRI1-EMS-SUPPRESSOR1 (BES1). *Plant Physiol.* **2020**, *182* (2), 1066–1082. https://doi.org/10.1104/pp.19.01220.
(61) Lima, R. B.; Figueiredo, D. D. Sex on Steroids: How Brassinosteroids Shape Reproductive Development in Flowering Plants. *Plant Cell Physiol.* **2024**, *65* (10), 1581–1600. https://doi.org/10.1093/pcp/pcae050.
(62) Jia, D.; Chen, L.-G.; Yin, G.; Yang, X.; Gao, Z.; Guo, Y.; Sun, Y.; Tang, W. Brassinosteroids Regulate Outer Ovule Integument Growth in Part via the Control of INNER NO OUTER by BRASSINOZOLE-RESISTANT Family Transcription Factors. *J. Integr. Plant Biol.* **2020**, *62* (8), 1093–1111. https://doi.org/10.1111/jipb.12915.
(63) Vogler, F.; Schmalzl, C.; Englhart, M.; Bircheneder, M.; Sprunck, S. Brassinosteroids Promote Arabidopsis Pollen Germination and Growth. *Plant Reprod.* **2014**, *27* (3), 153–167. https://doi.org/10.1007/s00497-014-0247-x.
(64) Lima, R. B.; Pankaj, R.; Ehlert, S. T.; Finger, P.; Fröhlich, A.; Bayle, V.; Landrein, B.; Sampathkumar, A.; Figueiredo, D. D. Seed Coat-Derived Brassinosteroid Signaling Regulates Endosperm Development. *Nat. Commun.* **2024**, *15* (1), 9352. https://doi.org/10.1038/s41467-024-53671-x.
(65) Chen, W.; Lv, M.; Wang, Y.; Wang, P.-A.; Cui, Y.; Li, M.; Wang, R.; Gou, X.; Li, J. BES1 Is Activated by EMS1-TPD1-SERK1/2-Mediated Signaling to Control Tapetum Development in Arabidopsis Thaliana. *Nat. Commun.* **2019**, *10* (1), 4164. https://doi.org/10.1038/s41467-019-12118-4.
(66) Yin, Y.; Wang, Z. Y.; Mora-Garcia, S.; Li, J.; Yoshida, S.; Asami, T.; Chory, J. BES1 Accumulates in the Nucleus in Response to Brassinosteroids to Regulate Gene Expression and Promote Stem Elongation. *Cell* **2002**, *109* (2), 181–191. https://doi.org/10.1016/s0092-8674(02)00721-3.
(67) Zhu, T.; Wei, C.; Yu, Y.; Zhang, Z.; Zhu, J.; Liang, Z.; Song, X.; Fu, W.; Cui, Y.; Wang, Z.-Y.; Li, C. The BAS Chromatin Remodeler Determines Brassinosteroid-Induced Transcriptional Activation and Plant Growth in Arabidopsis. *Dev. Cell* **2024**, *59* (7), 924-939.e6. https://doi.org/10.1016/j.devcel.2024.01.021.
(68) Wang, W.; Lu, X.; Li, L.; Lian, H.; Mao, Z.; Xu, P.; Guo, T.; Xu, F.; Du, S.; Cao, X.; Wang, S.; Shen, H.; Yang, H.-Q. Photoexcited CRYPTOCHROME1 Interacts with Dephosphorylated BES1 to Regulate Brassinosteroid Signaling and Photomorphogenesis in Arabidopsis. *Plant Cell* **2018**, *30* (9), 1989–2005. https://doi.org/10.1105/tpc.17.00994.
(69) Oh, E.; Zhu, J.-Y.; Ryu, H.; Hwang, I.; Wang, Z.-Y. TOPLESS Mediates Brassinosteroid-Induced Transcriptional Repression through Interaction with BZR1. *Nat. Commun.* **2014**, *5*, 4140. https://doi.org/10.1038/ncomms5140.
(70) Wang, T.; Li, M.; Yang, J.; Li, M.; Zhang, Z.; Gao, H.; Wang, C.; Tian, H. Brassinosteroid Transcription Factor BES1 Modulates Nitrate Deficiency by Promoting NRT2.1 and NRT2.2 Transcription in Arabidopsis. *Plant J. Cell Mol. Biol.* **2023**, *114* (6), 1443–1457. https://doi.org/10.1111/tpj.16203.
(71) O'Malley, R. C.; Huang, S.-S. C.; Song, L.; Lewsey, M. G.; Bartlett, A.; Nery, J. R.; Galli, M.; Gallavotti, A.; Ecker, J. R. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **2016**, *165* (5), 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038.

(72) Doll, N. M.; Bovio, S.; Gaiti, A.; Marsollier, A.-C.; Chamot, S.; Moussu, S.; Widiez, T.; Ingram, G. The Endosperm-Derived Embryo Sheath Is an Anti-Adhesive Structure That Facilitates Cotyledon Emergence during Germination in Arabidopsis. *Curr. Biol. CB* **2020**, *30* (5), 909-915.e4. https://doi.org/10.1016/j.cub.2019.12.057.

(73) Moussu, S.; Doll, N. M.; Chamot, S.; Brocard, L.; Creff, A.; Fourquin, C.; Widiez, T.; Nimchuk, Z. L.; Ingram, G. ZHOUPI and KERBEROS Mediate Embryo/Endosperm Separation by Promoting the Formation of an Extracuticular Sheath at the Embryo Surface. *Plant Cell* **2017**, *29* (7), 1642–1656. https://doi.org/10.1105/tpc.17.00016.

(74) Yang, S.; Johnston, N.; Talideh, E.; Mitchell, S.; Jeffree, C.; Goodrich, J.; Ingram, G. The Endosperm-Specific ZHOUPI Gene of Arabidopsis Thaliana Regulates Endosperm Breakdown and Embryonic Epidermal Development. *Dev. Camb. Engl.* **2008**, *135* (21), 3501–3509. https://doi.org/10.1242/dev.026708.

(75) Doll, N. M.; Van Hautegem, T.; Schilling, N.; De Rycke, R.; De Winter, F.; Fendrych, M.; Nowack, M. K. Endosperm Cell Death Promoted by NAC Transcription Factors Facilitates Embryo Invasion in Arabidopsis. *Curr. Biol. CB* **2023**, *33* (17), 3785-3795.e6. https://doi.org/10.1016/j.cub.2023.08.003.

(76) Olsen, O.-A. Nuclear Endosperm Development in Cereals and Arabidopsis Thaliana. *Plant Cell* **2004**, *16 Suppl* (Suppl), S214-227. https://doi.org/10.1105/tpc.017111.

(77) Povilus, R. A.; Gehring, M. Maternal-Filial Transfer Structures in Endosperm: A Nexus of Nutritional Dynamics and Seed Development. *Curr. Opin. Plant Biol.* **2022**, *65*, 102121. https://doi.org/10.1016/j.pbi.2021.102121.

(78) Baroux, C.; Fransz, P.; Grossniklaus, U. Nuclear Fusions Contribute to Polyploidization of the Gigantic Nuclei in the Chalazal Endosperm of Arabidopsis. *Planta* **2004**, *220* (1), 38–46. https://doi.org/10.1007/s00425-004-1326-2.

(79) Ali, M. F.; Shin, J. M.; Fatema, U.; Kurihara, D.; Berger, F.; Yuan, L.; Kawashima, T. Cellular Dynamics of Coenocytic Endosperm Development in Arabidopsis Thaliana. *Nat. Plants* **2023**, *9* (2), 330–342. https://doi.org/10.1038/s41477-022-01331-7.

(80) Nakamura, M.; Katsumata, H.; Abe, M.; Yabe, N.; Komeda, Y.; Yamamoto, K. T.; Takahashi, T. Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in Arabidopsis. *Plant Physiol.* **2006**, *141* (4), 1363–1375. https://doi.org/10.1104/pp.106.077388.

(81) Murphy, E.; Smith, S.; De Smet, I. Small Signaling Peptides in Arabidopsis Development: How Cells Communicate Over a Short Distance. *Plant Cell* **2012**, *24* (8), 3198–3217. https://doi.org/10.1105/tpc.112.099010.

(82) Matsubayashi, Y. Posttranslationally Modified Small-Peptide Signals in Plants. *Annu. Rev. Plant Biol.* **2014**, *65* (Volume 65, 2014), 385–413. https://doi.org/10.1146/annurev-arplant-050312-120122.

(83) Higashiyama, T. Peptide Signaling in Pollen–Pistil Interactions. *Plant Cell Physiol.* **2010**, *51* (2), 177–189. https://doi.org/10.1093/pcp/pcq008.

(84) Marshall, E.; Costa, L. M.; Gutierrez-Marcos, J. Cysteine-Rich Peptides (CRPs) Mediate Diverse Aspects of Cell–Cell Communication in Plant Reproduction and Development. *J. Exp. Bot.* **2011**, *62* (5), 1677–1686. https://doi.org/10.1093/jxb/err002.

(85) Olsson, V.; Joos, L.; Zhu, S.; Gevaert, K.; Butenko, M. A.; Smet, I. D. Look Closely, the Beautiful May Be Small: Precursor-Derived Peptides in Plants. *Annu. Rev. Plant Biol.* **2019**, *70* (Volume 70, 2019), 153–186. https://doi.org/10.1146/annurev-arplant-042817-040413.

(86) Takahashi, F.; Hanada, K.; Kondo, T.; Shinozaki, K. Hormone-like Peptides and Small Coding Genes in Plant Stress Signaling and Development. *Curr. Opin. Plant Biol.* **2019**, *51*, 88–95. https://doi.org/10.1016/j.pbi.2019.05.011.

(87) Ghorbani, S.; Lin, Y.-C.; Parizot, B.; Fernandez, A.; Njo, M. F.; Van de Peer, Y.; Beeckman, T.; Hilson, P. Expanding the Repertoire of Secretory Peptides Controlling Root

Development with Comparative Genome Analysis and Functional Assays. *J. Exp. Bot.* **2015**, *66* (17), 5257–5269. https://doi.org/10.1093/jxb/erv346.
(88) Hellmann, E. MtSSPdb: A New Database for the Small Secreted Peptide Research Community. *Plant Physiol.* **2020**, *183* (1), 31–32. https://doi.org/10.1104/pp.20.00376.
(89) Hu, X.-L.; Lu, H.; Hassan, M. M.; Zhang, J.; Yuan, G.; Abraham, P. E.; Shrestha, H. K.; Villalobos Solis, M. I.; Chen, J.-G.; Tschaplinski, T. J.; Doktycz, M. J.; Tuskan, G. A.; Cheng, Z.-M. (Max); Yang, X. Advances and Perspectives in Discovery and Functional Analysis of Small Secreted Proteins in Plants. *Hortic. Res.* **2021**, *8* (1), 1–14. https://doi.org/10.1038/s41438-021-00570-7.
(90) Takeuchi, H.; Higashiyama, T. A Species-Specific Cluster of Defensin-Like Genes Encodes Diffusible Pollen Tube Attractants in Arabidopsis. *PLOS Biol.* **2012**, *10* (12), e1001449. https://doi.org/10.1371/journal.pbio.1001449.
(91) Coculo, D.; Lionetti, V. The Plant Invertase/Pectin Methylesterase Inhibitor Superfamily. *Front. Plant Sci.* **2022**, *13*. https://doi.org/10.3389/fpls.2022.863892.
(92) Huang, J.; Zhang, T.; Linstroth, L.; Tillman, Z.; Otegui, M. S.; Owen, H. A.; Zhao, D. Control of Anther Cell Differentiation by the Small Protein Ligand TPD1 and Its Receptor EMS1 in Arabidopsis. *PLOS Genet.* **2016**, *12* (8), e1006147. https://doi.org/10.1371/journal.pgen.1006147.
(93) Song, X.-F.; Hou, X.-L.; Liu, C.-M. CLE Peptides: Critical Regulators for Stem Cell Maintenance in Plants. *Planta* **2021**, *255* (1), 5. https://doi.org/10.1007/s00425-021-03791-1.
(94) Geist, K. S.; Strassmann, J. E.; Queller, D. C. Family Quarrels in Seeds and Rapid Adaptive Evolution in Arabidopsis. *Proc. Natl. Acad. Sci.* **2019**, *116* (19), 9463–9468. https://doi.org/10.1073/pnas.1817733116.
(95) Bircheneder, S.; Dresselhaus, T. Why Cellular Communication during Plant Reproduction Is Particularly Mediated by CRP Signalling. *J. Exp. Bot.* **2016**, *67* (16), 4849–4861. https://doi.org/10.1093/jxb/erw271.
(96) Takeuchi, H.; Higashiyama, T. A Species-Specific Cluster of Defensin-Like Genes Encodes Diffusible Pollen Tube Attractants in Arabidopsis. *PLOS Biol.* **2012**, *10* (12), e1001449. https://doi.org/10.1371/journal.pbio.1001449.
(97) Nei, M.; Gojobori, T. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Mol. Biol. Evol.* **1986**, *3* (5), 418–426. https://doi.org/10.1093/oxfordjournals.molbev.a040410.
(98) Shahan, R.; Hsu, C.-W.; Nolan, T. M.; Cole, B. J.; Taylor, I. W.; Greenstreet, L.; Zhang, S.; Afanassiev, A.; Vlot, A. H. C.; Schiebinger, G.; Benfey, P. N.; Ohler, U. A Single-Cell *Arabidopsis* Root Atlas Reveals Developmental Trajectories in Wild-Type and Cell Identity Mutants. *Dev. Cell* **2022**, *57* (4), 543-560.e9. https://doi.org/10.1016/j.devcel.2022.01.008.
(99) Ke, Y.; Pujol, V.; Staut, J.; Pollaris, L.; Seurinck, R.; Eekhout, T.; Grones, C.; Saura-Sanchez, M.; Van Bel, M.; Vuylsteke, M.; Ariani, A.; Liseron-Monfils, C.; Vandepoele, K.; Saeys, Y.; De Rybel, B. A Single-Cell and Spatial Wheat Root Atlas with Cross-Species Annotations Delineates Conserved Tissue-Specific Marker Genes and Regulators. *Cell Rep.* **2025**, *44* (2), 115240. https://doi.org/10.1016/j.celrep.2025.115240.
(100) Guillotin, B.; Rahni, R.; Passalacqua, M.; Mohammed, M. A.; Xu, X.; Raju, S. K.; Ramírez, C. O.; Jackson, D.; Groen, S. C.; Gillis, J.; Birnbaum, K. D. A Pan-Grass Transcriptome Reveals Patterns of Cellular Divergence in Crops. *Nature* **2023**, *617* (7962), 785–791. https://doi.org/10.1038/s41586-023-06053-0.
(101) Lee, T. A.; Nobori, T.; Illouz-Eliaz, N.; Xu, J.; Jow, B.; Nery, J. R.; Ecker, J. R. A Single-Nucleus Atlas of Seed-to-Seed Development in Arabidopsis. bioRxiv September 2, 2023, p 2023.03.23.533992. https://doi.org/10.1101/2023.03.23.533992.
(102) Guo, X.; Wang, Y.; Zhao, C.; Tan, C.; Yan, W.; Xiang, S.; Zhang, D.; Zhang, H.; Zhang, M.; Yang, L.; Yan, M.; Xie, P.; Wang, Y.; Li, L.; Fang, D.; Guang, X.; Shao, W.; Wang, F.;

Wang, H.; Sahu, S. K.; Liu, M.; Wei, T.; Peng, Y.; Qiu, Y.; Peng, T.; Zhang, Y.; Ni, X.; Xu, Z.; Lu, H.; Li, Z.; Yang, H.; Wang, E.; Lisby, M.; Liu, H.; Guo, H.; Xu, X. An Arabidopsis Single-Nucleus Atlas Decodes Leaf Senescence and Nutrient Allocation. *Cell* **2025**, *0* (0). https://doi.org/10.1016/j.cell.2025.03.024.

(103) Zhang, X.; Luo, Z.; Marand, A. P.; Yan, H.; Jang, H.; Bang, S.; Mendieta, J. P.; Minow, M. A. A.; Schmitz, R. J. A Spatially Resolved Multi-Omic Single-Cell Atlas of Soybean Development. *Cell* **2025**, *188* (2), 550-567.e19. https://doi.org/10.1016/j.cell.2024.10.050.

(104) Nguyen, H.; Brown, R. C.; Lemmon, B. E. The Specialized Chalazal Endosperm inArabidopsis Thaliana andLepidium Virginicum (Brassicaceae). *Protoplasma* **2000**, *212* (1), 99–110. https://doi.org/10.1007/BF01279351.

(105) Vriens, K.; Cammue, B. P. A.; Thevissen, K. Antifungal Plant Defensins: Mechanisms of Action and Production. *Molecules* **2014**, *19* (8), 12280–12303. https://doi.org/10.3390/molecules190812280.

(106) Wenzel, D. M.; Klevit, R. E. Following Ariadne's Thread: A New Perspective on RBR Ubiquitin Ligases. *BMC Biol.* **2012**, *10* (1), 24. https://doi.org/10.1186/1741-7007-10-24.

(107) Mladek, C.; Guger, K.; Hauser, M.-T. Identification and Characterization of the ARIADNEGene Family in Arabidopsis. A Group of Putative E3 Ligases. *Plant Physiol.* **2003**, *131* (1), 27–40. https://doi.org/10.1104/pp.012781.

(108) Singleton, M. D.; Eisen, M. B. Evolutionary Analyses of Intrinsically Disordered Regions Reveal Widespread Signals of Conservation. *PLoS Comput. Biol.* **2024**, *20* (4), e1012028. https://doi.org/10.1371/journal.pcbi.1012028.

(109) BROWN, R. C.; LEMMON, B. E.; NGUYEN, H. Comparative Anatomy of the Chalazal Endosperm Cyst in Seeds of the Brassicaceae. *Bot. J. Linn. Soc.* **2004**, *144* (4), 375–394. https://doi.org/10.1111/j.1095-8339.2003.00263.x.

(110) Cheng, C.-Y.; Krishnakumar, V.; Chan, A. P.; Thibaud-Nissen, F.; Schobel, S.; Town, C. D. Araport11: A Complete Reannotation of the Arabidopsis Thaliana Reference Genome. *Plant J.* **2017**, *89* (4), 789–804. https://doi.org/10.1111/tpj.13415.

(111) Young, M. D.; Behjati, S. SoupX Removes Ambient RNA Contamination from Droplet-Based Single-Cell RNA Sequencing Data. *GigaScience* **2020**, *9* (12), giaa151. https://doi.org/10.1093/gigascience/giaa151.

(112) Hao, Y.; Stuart, T.; Kowalski, M. H.; Choudhary, S.; Hoffman, P.; Hartman, A.; Srivastava, A.; Molla, G.; Madad, S.; Fernandez-Granda, C.; Satija, R. Dictionary Learning for Integrative, Multimodal and Scalable Single-Cell Analysis. *Nat. Biotechnol.* **2024**, *42* (2), 293–304. https://doi.org/10.1038/s41587-023-01767-y.

(113) Heumos, L.; Schaar, A. C.; Lance, C.; Litinetskaya, A.; Drost, F.; Zappia, L.; Lücken, M. D.; Strobl, D. C.; Henao, J.; Curion, F.; Schiller, H. B.; Theis, F. J. Best Practices for Single-Cell Analysis across Modalities. *Nat. Rev. Genet.* **2023**, 1–23. https://doi.org/10.1038/s41576-023-00586-w.

(114) Germain, P.-L.; Lun, A.; Garcia Meixide, C.; Macnair, W.; Robinson, M. D. Doublet Identification in Single-Cell Sequencing Data Using scDblFinder. *F1000Research* **2022**, *10*, 979. https://doi.org/10.12688/f1000research.73600.2.

(115) Korsunsky, I.; Millard, N.; Fan, J.; Slowikowski, K.; Zhang, F.; Wei, K.; Baglaenko, Y.; Brenner, M.; Loh, P.; Raychaudhuri, S. Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony. *Nat. Methods* **2019**, *16* (12), 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

(116) Menges, M.; de Jager, S. M.; Gruissem, W.; Murray, J. A. H. Global Analysis of the Core Cell Cycle Regulators of Arabidopsis Identifies Novel Genes, Reveals Multiple and Highly Specific Profiles of Expression and Provides a Coherent Model for Plant Cell Cycle Control. *Plant J. Cell Mol. Biol.* **2005**, *41* (4), 546–566. https://doi.org/10.1111/j.1365-313X.2004.02319.x.

(117) *bluster*. Bioconductor. http://bioconductor.org/packages/bluster/ (accessed 2025-04-28).

(118) Xu, S.; Hu, E.; Cai, Y.; Xie, Z.; Luo, X.; Zhan, L.; Tang, W.; Wang, Q.; Liu, B.; Wang, R.; Xie, W. Using clusterProfiler to Characterise Multi-Omics Data. *Nat. Protoc.* **2024**, *19* (11), 3292–3320.

(119) Tirosh, I.; Izar, B.; Prakadan, S. M.; Wadsworth, M. H.; Treacy, D.; Trombetta, J. J.; Rotem, A.; Rodman, C.; Lian, C.; Murphy, G.; Fallahi-Sichani, M.; Dutton-Regester, K.; Lin, J.-R.; Cohen, O.; Shah, P.; Lu, D.; Genshaft, A. S.; Hughes, T. K.; Ziegler, C. G. K.; Kazer, S. W.; Gaillard, A.; Kolb, K. E.; Villani, A.-C.; Johannessen, C. M.; Andreev, A. Y.; Van Allen, E. M.; Bertagnolli, M.; Sorger, P. K.; Sullivan, R. J.; Flaherty, K. T.; Frederick, D. T.; Jané-Valbuena, J.; Yoon, C. H.; Rozenblatt-Rosen, O.; Shalek, A. K.; Regev, A.; Garraway, L. A. Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq. *Science* **2016**, *352* (6282), 189–196. https://doi.org/10.1126/science.aad0501.

(120) Cao, J.; Spielmann, M.; Qiu, X.; Huang, X.; Ibrahim, D. M.; Hill, A. J.; Zhang, F.; Mundlos, S.; Christiansen, L.; Steemers, F. J.; Trapnell, C.; Shendure, J. The Single-Cell Transcriptional Landscape of Mammalian Organogenesis. *Nature* **2019**, *566* (7745), 496–502. https://doi.org/10.1038/s41586-019-0969-x.

(121) Zhang, Y.; Maruyama, D.; Toda, E.; Kinoshita, A.; Okamoto, T.; Mitsuda, N.; Takasaki, H.; Ohme-Takagi, M. Transcriptome Analyses Uncover Reliance of Endosperm Gene Expression on Arabidopsis Embryonic Development. *FEBS Lett.* **2023**, *597* (3), 407–417. https://doi.org/10.1002/1873-3468.14570.

(122) Huang, T.; Guillotin, B.; Rahni, R.; Birnbaum, K. D.; Wagner, D. A Rapid and Sensitive, Multiplex, Whole Mount RNA Fluorescence in Situ Hybridization and Immunohistochemistry Protocol. *Plant Methods* **2023**, *19*, 131. https://doi.org/10.1186/s13007-023-01108-9.

(123) Teufel, F.; Almagro Armenteros, J. J.; Johansen, A. R.; Gíslason, M. H.; Pihl, S. I.; Tsirigos, K. D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. *Nat. Biotechnol.* **2022**, *40* (7), 1023–1025. https://doi.org/10.1038/s41587-021-01156-3.

(124) Hu, T. T.; Pattyn, P.; Bakker, E. G.; Cao, J.; Cheng, J.-F.; Clark, R. M.; Fahlgren, N.; Fawcett, J. A.; Grimwood, J.; Gundlach, H.; Haberer, G.; Hollister, J. D.; Ossowski, S.; Ottilar, R. P.; Salamov, A. A.; Schneeberger, K.; Spannagl, M.; Wang, X.; Yang, L.; Nasrallah, M. E.; Bergelson, J.; Carrington, J. C.; Gaut, B. S.; Schmutz, J.; Mayer, K. F. X.; Van de Peer, Y.; Grigoriev, I. V.; Nordborg, M.; Weigel, D.; Guo, Y.-L. The Arabidopsis Lyrata Genome Sequence and the Basis of Rapid Genome Size Change. *Nat. Genet.* **2011**, *43* (5), 476–481. https://doi.org/10.1038/ng.807.

(125) Rawat, V.; Abdelsamad, A.; Pietzenuk, B.; Seymour, D. K.; Koenig, D.; Weigel, D.; Pecinka, A.; Schneeberger, K. Improving the Annotation of Arabidopsis Lyrata Using RNA-Seq Data. *PLOS ONE* **2015**, *10* (9), e0137391. https://doi.org/10.1371/journal.pone.0137391.

(126) Drost, H.-G.; Gabel, A.; Grosse, I.; Quint, M. Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. *Mol. Biol. Evol.* **2015**, *32* (5), 1221–1231. https://doi.org/10.1093/molbev/msv012.

(127) Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24* (8), 1586–1591. https://doi.org/10.1093/molbev/msm088.

(128) *Positive, Neutral, Negative Selection with Codeml using Multiple Genome Annotations*. Bioinformatics Workbook. https://bioinformaticsworkbook.org/dataAnalysis/ComparativeGenomics/Finding_Positive_Selection_With_Codeml.html (accessed 2025-04-28).

(129) Petersen, L.; Bollback, J. P.; Dimmic, M.; Hubisz, M.; Nielsen, R. Genes under Positive Selection in Escherichia Coli. *Genome Res.* **2007**, *17* (9), 1336–1343. https://doi.org/10.1101/gr.6254707.

(130) Petersen, L.; Bollback, J. P.; Dimmic, M.; Hubisz, M.; Nielsen, R. *Genes under positive selection in Escherichia coli*. https://doi.org/10.1101/gr.6254707.

(131) Emms, D. M.; Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* **2019**, *20* (1), 238. https://doi.org/10.1186/s13059-019-1832-y.
(132) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. https://doi.org/10.1038/msb.2011.75.
(133) Shakya, M.; Ahmed, S. A.; Davenport, K. W.; Flynn, M. C.; Lo, C.-C.; Chain, P. S. G. Standardized Phylogenetic and Molecular Evolutionary Analysis Applied to Species across the Microbial Tree of Life. *Sci. Rep.* **2020**, *10* (1), 1723. https://doi.org/10.1038/s41598-020-58356-1.

Acknowledgements
We thank Souraya Khouider and Elizabeth Hemenway for assistance with timed seed isolation for snRNA-seq. We thank Troy Whitfield, Sumeet Gupta, and Alex Dionisio for guidance on GSEA statistical analyses, snRNA-seq analysis, and seed atlas app development, respectively. Additionally, we thank Jennifer Love, Stephen Mraz III, and Amanda Chilaka at the WIBR Genome Technology Core for all snRNA-seq library preparation and sequencing. We also thank Patrick Autissier and Aditya Rathee for performing FANS procedures at the WIBR flow core facility. This research was supported by The Manton Foundation, The Dr. Vincent J. Ryan Orphan Plant Project, and a National Science Foundation Graduate Research Fellowship to CAM. MG is an Investigator of the Howard Hughes Medical Institute.

Competing Interests
The authors declare no competing interests.

Extended Data
Extended Data 1. Level 3 annotation key
Extended Data 2. General cluster characteristics and similarity
Extended Data 3. Informative markers for seed coat layers and nucleoside metabolism enrichment in the chalazal proliferating tissue.
Extended Data 4. Cell types with complementary functions in the 3 DAP CZSC.
Extended Data 5. Level 3 seed coat cluster similarity at 3 and 5 DAP.
Extended Data 6. The 7 DAP MCE is the most distinctive endosperm subtype.
Extended Data 7. Differentially expressed genes along the apical/basal axis of the CZE.
Extended Data 8. Genes that vary along the CZE nodule-to-cyst trajectory.
Extended Data 9. SSPs are highly expressed in seed tissues and were absent from existing transcriptional atlases of seed development.
Extended Data 10. GO families containing rapidly evolving genes are enriched in the CZE and PEN.

Supplementary Information
Supplementary Table 1. All markers used to annotate clusters in this study
Supplementary Table 2. HCR-FISH probe information
Supplementary Table 3. Level 3 cluster quality controls and metadata
Supplementary Table 4. Differential expression analyses for levels 1-3 within all timepoints
Supplementary Table 5. Transcription factors that are vary along 5 DAP PEN-CZE pseudotime
Supplementary Table 6. Student's T-test results for all GSEA in this study
Supplementary Table 7. Genome-wide dN/dS and M2a/M1a LRT results
Supplementary Table 8. Plotting key for Mergner atlas tissue samples
Supplementary Table 9. SSPs detected and annotated in this study

1140 Supplementary Figure 1. Identifying an initial clustering resolution for level 3 annotations.
Supplementary Figure 2. Isolating chalazal endosperm subtypes at 3 DAP through a
1142 subclustering analysis.
Supplementary Figure 3. Isolating chalazal endosperm subtypes at 5 DAP through a
1144 subclustering analysis.
Supplementary Figure 4. TPD1-like peptides are expressed in the 3 DAP endosperm.
1146 Supplementary Figure 5. FANS gating scheme for purifying seed nuclei before snRNA-seq.
Supplementary Figure 6. 5 DAP biological replicates before and after Harmony integration.

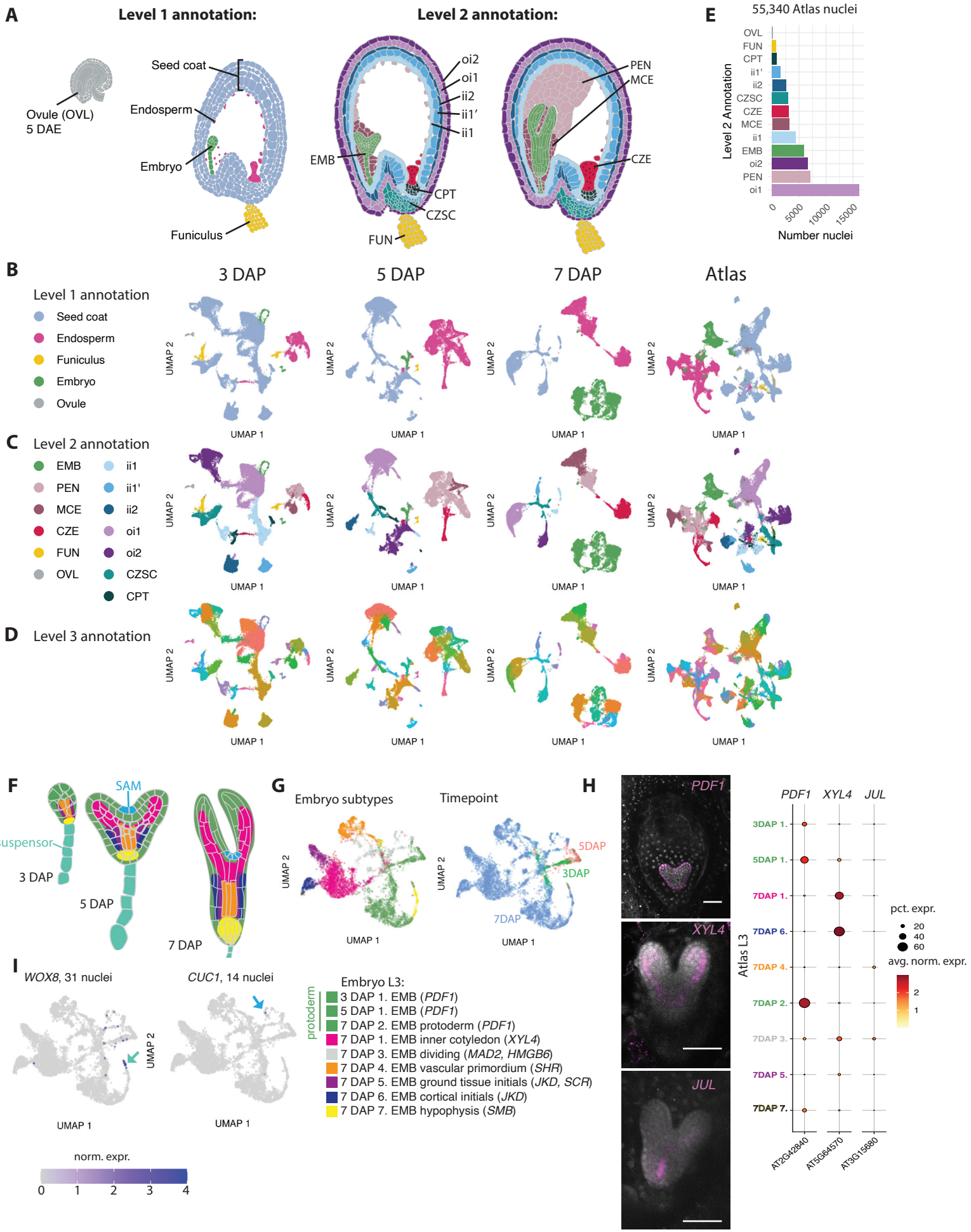

A
Level 1 annotation:
Level 2 annotation:
Ovule (OVL)
5 DAE
Seed coat
Endosperm
Embryo
Funiculus
oi2
oi1
ii2
ii1′
ii1
EMB
CPT
CZSC
FUN
PEN
MCE
CZE
E
55,340 Atlas nuclei
OVL
FUN
CPT
ii1'
ii2
CZSC
CZE
MCE
ii1
EMB
oi2
PEN
oi1
Level 2 Annotation
0
5000
10000
15000
Number nuclei
B
3 DAP
5 DAP
7 DAP
Atlas
Level 1 annotation
Seed coat
Endosperm
Funiculus
Embryo
Ovule
UMAP 2
UMAP 1
C
Level 2 annotation
EMB
PEN
MCE
CZE
FUN
OVL
ii1
ii1'
ii2
oi1
oi2
CZSC
CPT
D
Level 3 annotation
F
SAM
suspensor
3 DAP
5 DAP
7 DAP
G
Embryo subtypes
Timepoint
5DAP
3DAP
7DAP
H
PDF1
XYL4
JUL
3DAP 1.
5DAP 1.
7DAP 1.
7DAP 6.
7DAP 4.
7DAP 2.
7DAP 3.
7DAP 5.
7DAP 7.
Atlas L3
AT2G42840
AT5G64570
AT3G15680
pct. expr.
20
40
60
avg. norm. expr.
2
1
I
WOX8, 31 nuclei
CUC1, 14 nuclei
norm. expr.
0
1
2
3
4
Embryo L3:
protoderm
3 DAP 1. EMB (PDF1)
5 DAP 1. EMB (PDF1)
7 DAP 2. EMB protoderm (PDF1)
7 DAP 1. EMB inner cotyledon (XYL4)
7 DAP 3. EMB dividing (MAD2, HMGB6)
7 DAP 4. EMB vascular primordium (SHR)
7 DAP 5. EMB ground tissue initials (JKD, SCR)
7 DAP 6. EMB cortical initials (JKD)
7 DAP 7. EMB hypophysis (SMB)

Figure 1: A transcriptional atlas of Arabidopsis seed development. A, Seed developmental stages profiled in this study and Level 1 (L1) and Level 2 (L2) annotations. Outer integument 2 (oi2), outer integument 1 (oi1), inner integument 2 (ii2), inner integument 1′ (ii1′), inner integument 1 (ii1), embryo (EMB), chalazal proliferating tissue (CPT), chalazal seed coat (CZSC), funiculus (FUN), peripheral endosperm (PEN), micropylar endosperm (MCE), chalazal endosperm (CZE), ovule (OVL). **B-D**, Level 1-3 annotations across all datasets. See Supplementary Table 3 for complete descriptions of all L3 annotations. **E**, Cell type proportions for L2 annotations in the full atlas dataset. **F**, Morphological stages of the embryo. **G**, Integrated 3-7 DAP embryo datasets. 3 and 5 DAP embryo are a single cluster labeled with protoderm markers. **H**, Left, HCR validation of the protoderm (*PDF1+*), inner cotyledon (*XYL4+*), and vascular primordium (*JUL+*) in the embryo. Scale bar = 50µm. Right: Expression of genes corresponding to HCR probes in the integrated embryo dataset, split by L3 annotation. **I**, Detection of rare nuclei types in the integrated embryo dataset, the teal arrow points to the putative suspensor population (*WOX8+*) and the blue arrow points to the putative SAM population (*CUC1+*).

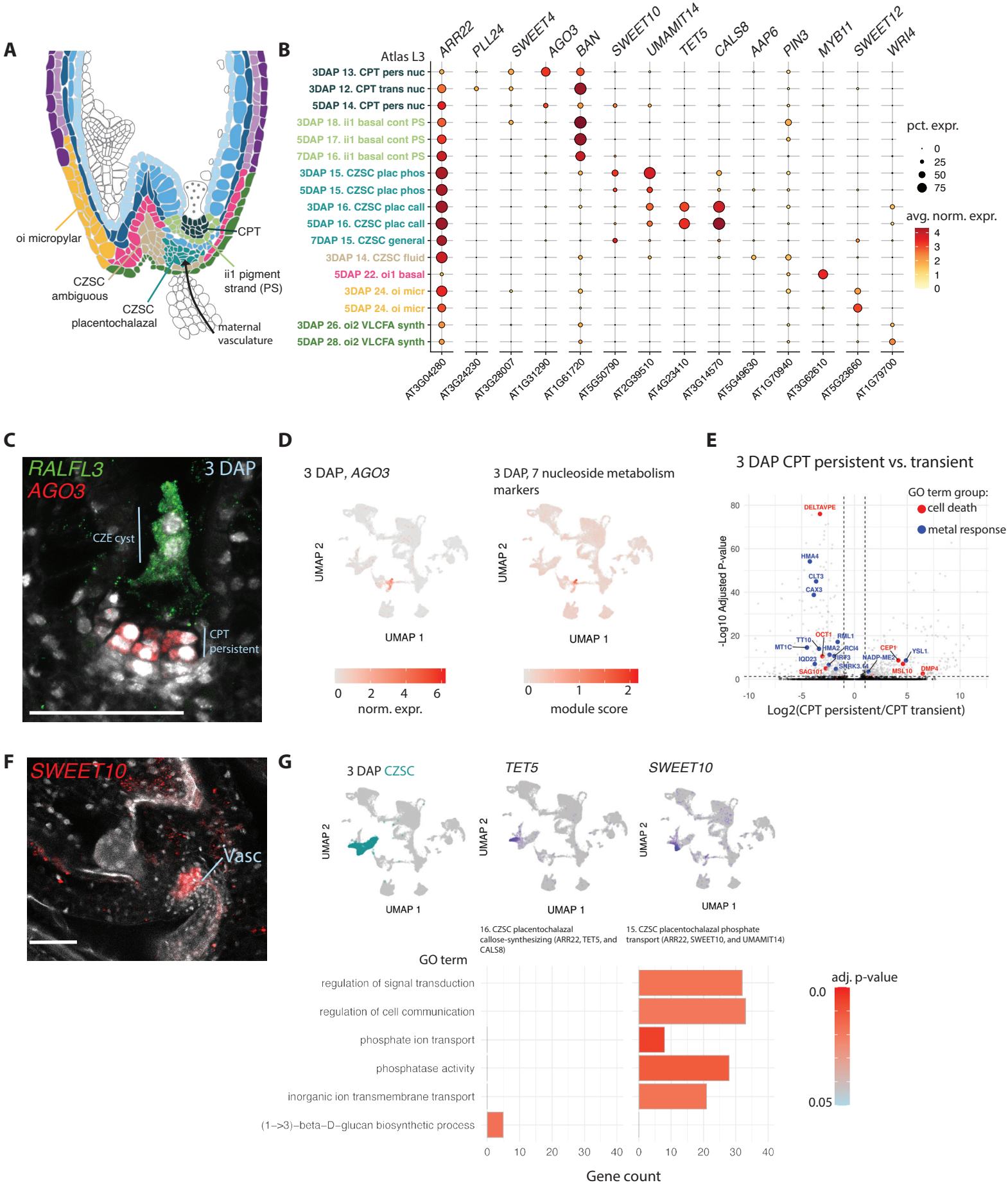

A
oi micropylar
CPT
CZSC ambiguous
ii1 pigment strand (PS)
CZSC placentochalazal
maternal vasculature
B
Atlas L3
ARR22
PLL24
SWEET4
AGO3
BAN
SWEET10
UMAMIT14
TET5
CALS8
AAP6
PIN3
MYB11
SWEET12
WRI4
3DAP 13. CPT pers nuc
3DAP 12. CPT trans nuc
5DAP 14. CPT pers nuc
3DAP 18. ii1 basal cont PS
5DAP 17. ii1 basal cont PS
7DAP 16. ii1 basal cont PS
3DAP 15. CZSC plac phos
5DAP 15. CZSC plac phos
3DAP 16. CZSC plac call
5DAP 16. CZSC plac call
7DAP 15. CZSC general
3DAP 14. CZSC fluid
5DAP 22. oi1 basal
3DAP 24. oi micr
5DAP 24. oi micr
3DAP 26. oi2 VLCFA synth
5DAP 28. oi2 VLCFA synth
AT3G04280
AT3G24230
AT3G28007
AT1G31290
AT1G61720
AT5G50790
AT2G39510
AT4G23410
AT3G14570
AT5G49630
AT1G70940
AT3G62610
AT5G23660
AT1G79700
pct. expr.
0
25
50
75
avg. norm. expr.
4
3
2
1
0
C
RALFL3
AGO3
3 DAP
CZE cyst
CPT persistent
D
3 DAP, AGO3
3 DAP, 7 nucleoside metabolism markers
UMAP 2
UMAP 1
0 2 4 6
norm. expr.
0 1 2
module score
E
3 DAP CPT persistent vs. transient
GO term group:
cell death
metal response
-Log10 Adjusted P-value
Log2(CPT persistent/CPT transient)
DELTAVPE
HMA4
CLT3
CAX3
OCT1
RML1
TT10
MT1C
HMA2
RCI4
CEP1
YSL1
IQD23
IRF3
NADP-ME2
SNRK3.14
SAG101
MSL10
DMP4
F
SWEET10
Vasc
G
3 DAP CZSC
TET5
SWEET10
16. CZSC placentochalazal callose-synthesizing (ARR22, TET5, and CALS8)
15. CZSC placentochalazal phosphate transport (ARR22, SWEET10, and UMAMIT14)
GO term
regulation of signal transduction
regulation of cell communication
phosphate ion transport
phosphatase activity
inorganic ion transmembrane transport
(1–>3)–beta–D–glucan biosynthetic process
Gene count
adj. p-value
0.0
0.05

Figure 2: Specialized cell types of the lower seed coat. A, A model for the distribution of lower seed coat nuclei types identified in this study at 5 DAP. Xylem and phloem terminate in the CZSC. **B**, Expression patterns of published markers (Supplementary Table 1) and HCR-validated markers (*SWEET10*, *AGO3*) across all L3 annotations for the lower seed coat. See Supplementary Table 3 to match abbreviated L3 names to their full descriptions. **C**, *AGO3* labels a subpopulation of the CPT at 3 DAP by HCR, while *RALFL3* labels the CZE. Scale bar = 50 µm. **D**, *AGO3* is highly specific to CPT (left), and seven nucleoside metabolism markers follow the same expression pattern (right). **E**, Differential expression between the 3 DAP CPT persistent and transient nuclei. **F**, *SWEET10* is localized to the placentochalazal region by HCR, and appears to be surrounding maternal vascular (Vasc) terminals. Scale bar = 50 µm. **G**, Top: *TET5* and *SWEET10* are the best markers for the L3 subtypes of the 3 DAP placentochalazal CZSC. Bottom: a subset of GO terms exclusively associated with each placentochalazal cluster**.**

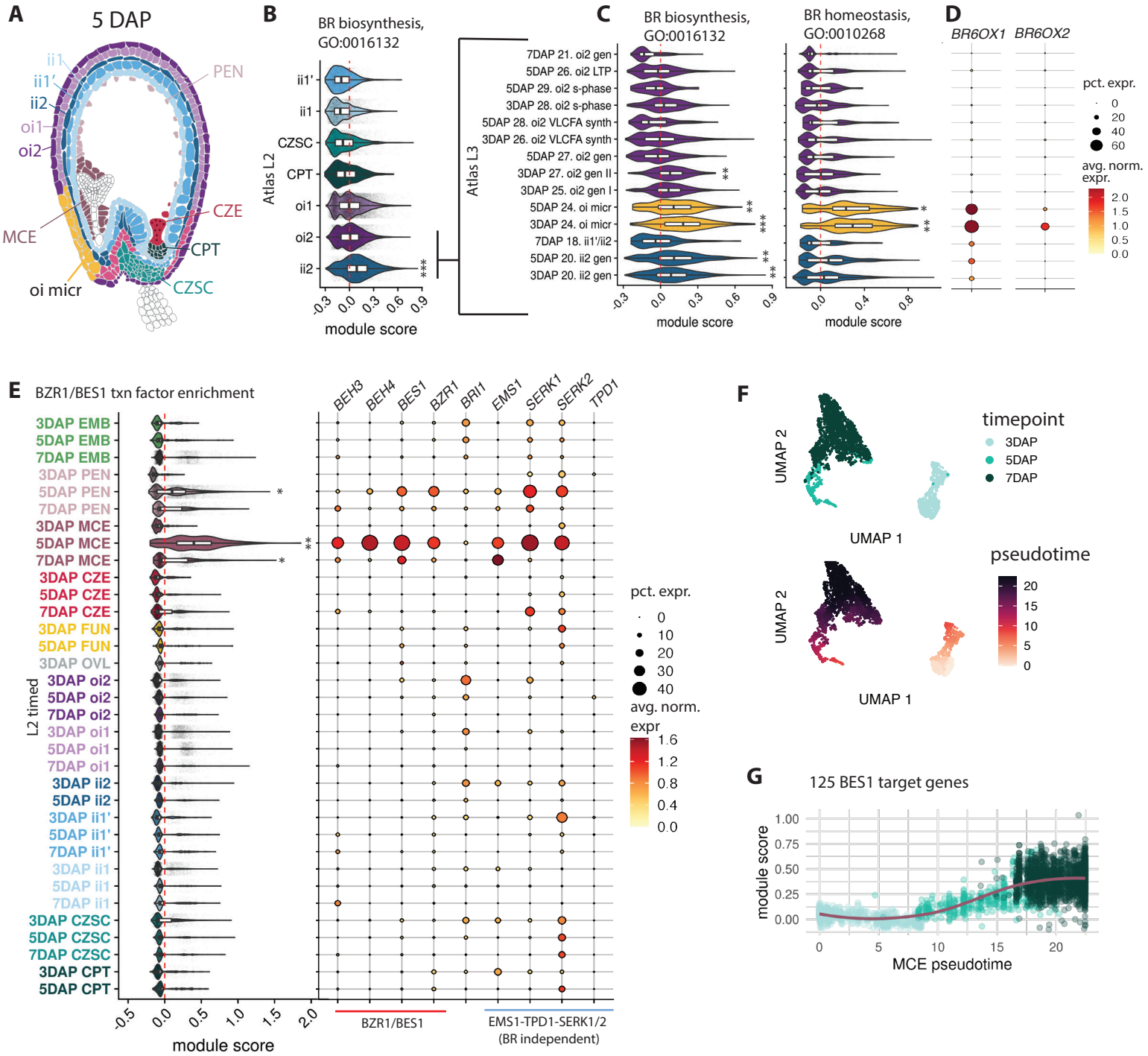


Figure 3: Genes that underlie BZR/BES1 transcription factor activity are expressed in the micropylar region of the seed. A, A subset of 5 DAP L2 annotations with the L3 oi1 micropylar region highlighted. **B**, Gene set enrichment for all genes associated with brassinosteroid (BR) biosynthesis (GO: 0016132) in the L2 seed coat across all timepoints. Only significant positive scores are labelled (p-value *<= 0.05,** <= 0.01, ***<= 0.001, and ****<= 0.0001 for all figures). **C**, Gene set enrichment for BR biosynthesis and homeostasis (GO:0010268) in a subset of L3 seed coat clusters. See Supplementary Table 3 to match abbreviated L3 names to their full descriptions. **D**, Gene expression pattern for *BR6OX1* and *BR6OX2*. **E**, Gene set enrichment for all five BZR/BES1 transcription factors detected in the atlas (*BES1*, *BEH1*, *BEH2*, *BEH3*, *BEH4*, *BZR1*). *BES1*, *BEH3*, *BZR1*, and *BEH4* show the strongest expression and are enriched in 5 DAP MCE. Components of the BR-independent BES1 activation pathway (*EMS1*, *SERK1/2*) show overlapping expression with *BES1*, although *TPD1* shows low, nonspecific expression throughout seed development. **F**, UMAP of the MCE-only dataset colored by time (top) and reversed 7 DAP-anchored MCE pseudotime (bottom). **G**, Gene set enrichment for 125 *BES1* targets (identified in O'Malley 2016 and DE in the MCE) increases through reversed 7 DAP-anchored MCE pseudotime

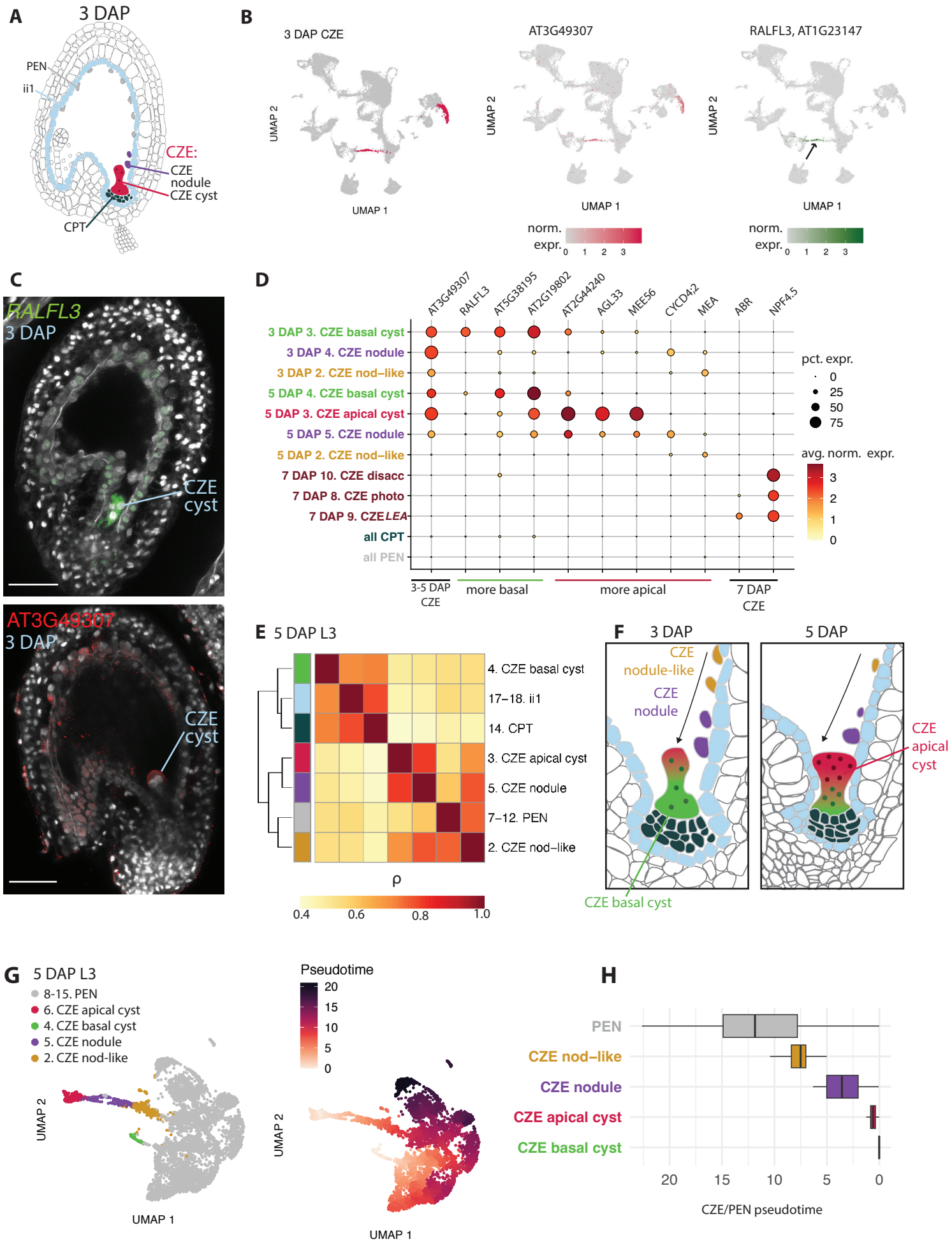

A
3 DAP
PEN
ii1
CZE:
CZE nodule
CZE cyst
CPT
B
3 DAP CZE
AT3G49307
RALFL3, AT1G23147
UMAP 2
UMAP 1
norm. expr.
0 1 2 3
C
RALFL3
3 DAP
CZE cyst
AT3G49307
3 DAP
CZE cyst
D
AT3G49307
RALFL3
AT5G38195
AT2G19802
AT2G44240
AGL33
MEE56
CYCD4;2
MEA
ABR
NPF4.5
3 DAP 3. CZE basal cyst
3 DAP 4. CZE nodule
3 DAP 2. CZE nod–like
5 DAP 4. CZE basal cyst
5 DAP 3. CZE apical cyst
5 DAP 5. CZE nodule
5 DAP 2. CZE nod–like
7 DAP 10. CZE disacc
7 DAP 8. CZE photo
7 DAP 9. CZE LEA
all CPT
all PEN
pct. expr.
0
25
50
75
avg. norm. expr.
3
2
1
0
3-5 DAP CZE
more basal
more apical
7 DAP CZE
E 5 DAP L3
4. CZE basal cyst
17–18. ii1
14. CPT
3. CZE apical cyst
5. CZE nodule
7–12. PEN
2. CZE nod–like
ρ
0.4
0.6
0.8
1.0
F
3 DAP
5 DAP
CZE nodule-like
CZE nodule
CZE apical cyst
CZE basal cyst
G 5 DAP L3
8-15. PEN
6. CZE apical cyst
4. CZE basal cyst
5. CZE nodule
2. CZE nod-like
UMAP 2
UMAP 1
Pseudotime
20
15
10
5
0
H
PEN
CZE nod–like
CZE nodule
CZE apical cyst
CZE basal cyst
20
15
10
5
0
CZE/PEN pseudotime

Figure 4: Transcriptional polarity in the chalazal endosperm. A, L2 annotations of chalazal-proximal tissues, including the CZE, CPT, PEN, and ii1. The chalazal nodule and cyst are morphologically characterized compartments of the CZE. **B**, Left: the CZE highlighted in the 3 DAP dataset. Middle, right: AT3G49307, a CZE marker, and *RALFL3* (arrow), a marker for a subtype of the CZE at 3 DAP. **C**, HCR validation of *RALFL3* (top) and AT3G49307 (bottom) expression in 3 DAP seeds. Scale bar = 50 µm. **D**, Expression patterns of CZE subtype markers from Picard et al. 2021 (*MEA*, *CYCD4;2*, AT2G44240) and this study in all L3 clusters for CZE, PEN, and CZE-proximal tissues at all timepoints. See Supplementary Table 3 to match abbreviated L3 names to their full descriptions. **E**, Clustered heatmap of the Spearman correlation coefficients for aggregated expression of the 5 DAP L3 clusters for CZE, PEN, and CZE-proximal maternal subtypes. **F**, A model for CZE cyst growth through fusion with proximal nuclei based on live imaging studies and our transcriptional characterization. **G**, Left: L3 annotations in the 5 DAP CZE and PEN dataset used for pseudotime analysis. Right: pseudotime analysis indicates that the basal and apical CZE cyst are proximal in pseudotime. **H**, Ordered PEN and CZE nuclei types based on mean pseudotime. The apical/basal axis of the CZE is bounded by the basal cyst and the CZE nodule-like populations.

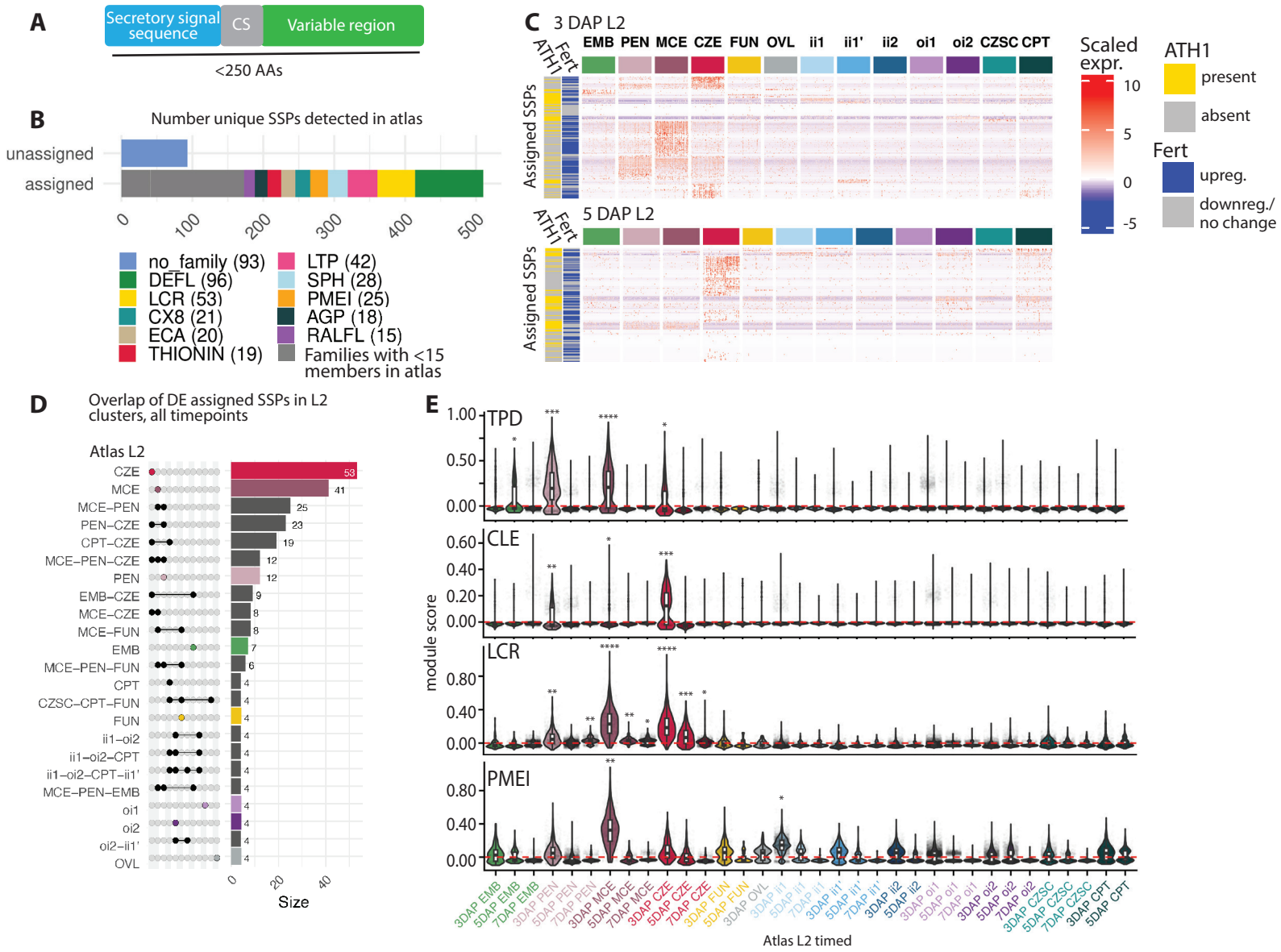


Figure 5: Short, secreted peptides are enriched in subtypes of the endosperm. A, Conventional structure of short, secreted peptides (SSPs), and the criteria used for SSP detection. Cleavage site (CS). **B**, Number of SSPs detected in the atlas assigned to characterized families, and the number of each peptide family represented among the assigned SSPs. The grey area in the "assigned" bar consists of all peptides in families with less than 15 SSPs detected in the atlas. **C** Clustered heatmaps of scaled gene expression for variable SSPs detected in the atlas. The row annotations "ATH1" and "Fert" indicates presence or absence from the ATH1 microarray, and whether the SSP was upregulated after fertilization (p-value < 0.05, log2FC >1) in the bulk RNA sequencing data from Zhang et al 2023. **D**, Upset plot indicating the proportion of unique SSPs DE (p-value < 0.05, log2FC >1) at any time in development for the L2 clusters. **E**, The results of gene set enrichment analysis for four SSP families with homology to members involved in signaling (TPD, CLE, and LCRs) as well as inhibitory roles (PMEI). P-values from pairwise Student's T-test with Bonferroni correction (p-value **<= 0.01, ***<= 0.001, and ****<= 0.0001 for all figures).

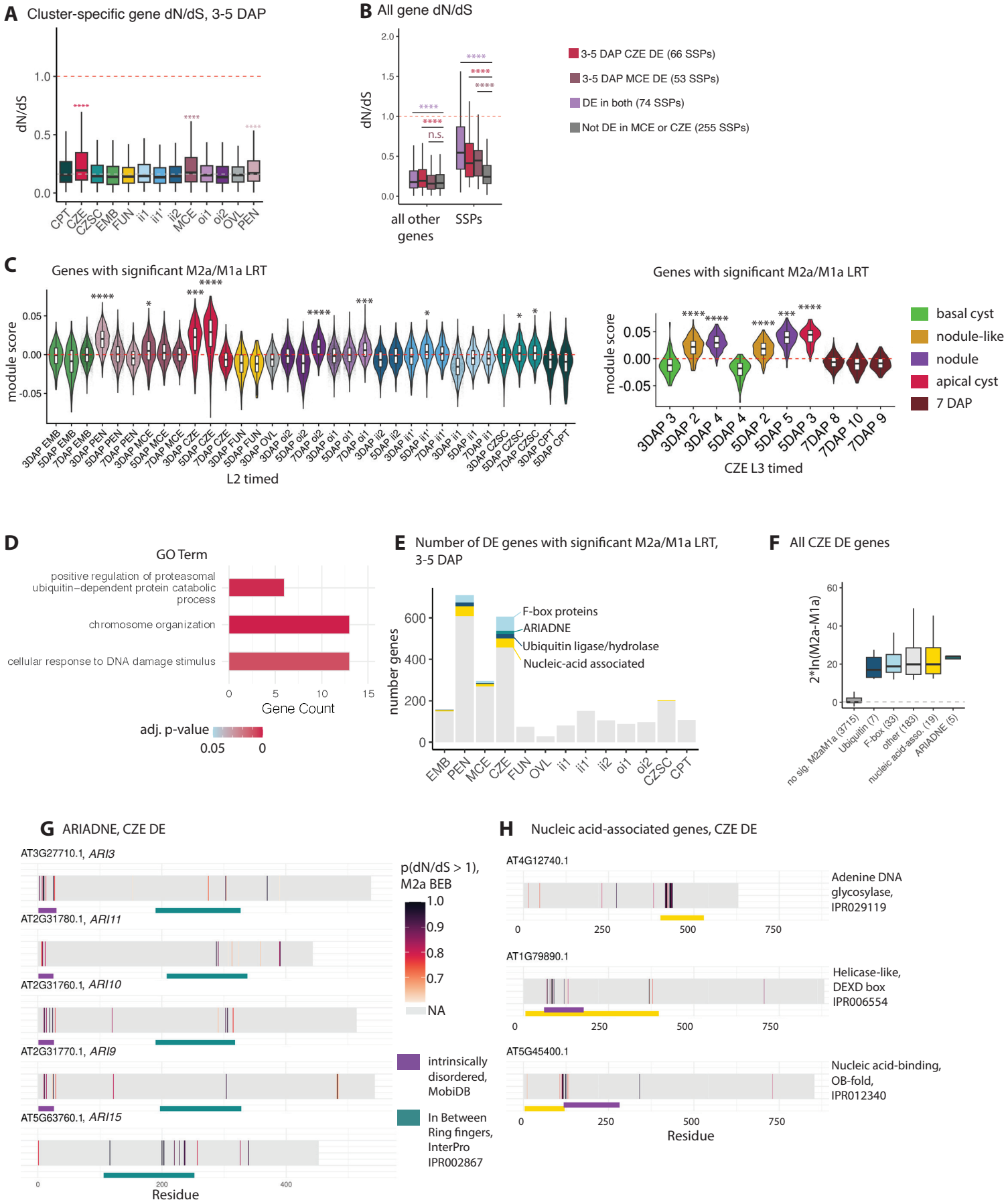
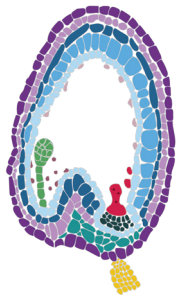

A Cluster-specific gene dN/dS, 3-5 DAP
dN/dS
CPT CZE CZSC EMB FUN ii1 ii1' ii2 MCE oi1 oi2 OVL PEN
B All gene dN/dS
dN/dS
all other genes
SSPs
n.s.
3-5 DAP CZE DE (66 SSPs)
3-5 DAP MCE DE (53 SSPs)
DE in both (74 SSPs)
Not DE in MCE or CZE (255 SSPs)
C Genes with significant M2a/M1a LRT
module score
L2 timed
Genes with significant M2a/M1a LRT
module score
CZE L3 timed
basal cyst
nodule-like
nodule
apical cyst
7 DAP
D GO Term
positive regulation of proteasomal ubiquitin-dependent protein catabolic process
chromosome organization
cellular response to DNA damage stimulus
Gene Count
adj. p-value
0.05
0
E Number of DE genes with significant M2a/M1a LRT, 3-5 DAP
number genes
F-box proteins
ARIADNE
Ubiquitin ligase/hydrolase
Nucleic-acid associated
EMB PEN MCE CZE FUN OVL ii1 ii1' ii2 oi1 oi2 CZSC CPT
F All CZE DE genes
2*ln(M2a-M1a)
no sig. M2aM1a (3715)
Ubiquitin (7)
F-box (33)
other (183)
nucleic acid-asso. (19)
ARIADNE (5)
G ARIADNE, CZE DE
AT3G27710.1, ARI3
AT2G31780.1, ARI11
AT2G31760.1, ARI10
AT2G31770.1, ARI9
AT5G63760.1, ARI15
Residue
p(dN/dS > 1), M2a BEB
NA
intrinsically disordered, MobiDB
In Between Ring fingers, InterPro IPR002867
H Nucleic acid-associated genes, CZE DE
AT4G12740.1
Adenine DNA glycosylase, IPR029119
AT1G79890.1
Helicase-like, DEXD box IPR006554
AT5G45400.1
Nucleic acid-binding, OB-fold, IPR012340
Residue

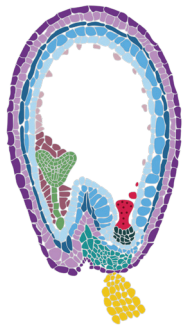
Figure 6: The CZE expresses genes under positive selection involved in protein degradation and nucleic acid interactions. A, Mean dN/dS for all L2 cluster DE genes at 3 and 5 DAP (p-value < 0.05, log2FC >1), outliers not shown. Significance is derived from a cluster vs. all other clusters Wilcoxon Ranked Sum test with Bonferroni correction (**** = p-value < 0.0001); significance is only shown for groups with significantly high dN/dS, although some groups are significantly low. Grey line indicates the median. **B**, Comparison of mean dN/dS for genes/SSPs DE (p-value < 0.05, log2FC >1) in the MCE, CZE, or both at 3 and/or 5 DAP versus all other genes in the Arabidopsis genome, outliers not shown. P-values from pairwise Student's T-test with Bonferroni correction (p-value **<= 0.01, ***<= 0.001, and ****<= 0.0001 for all figures). **C**, Left: GSEA of all genes with significant M2a/M1a LRTs for atlas L2 clusters. Right: the same GSEA from the left, partitioned by L3 annotation for all CZE nuclei. Statistical tests performed in an L3 cluster vs. all atlas cluster comparison (p-value **<= 0.01, ***<= 0.001, and ****<= 0.0001). **D**, A subset of the top GO terms for genes with significant M2a/M1a LRTs DE in the CZE. **E,** The proportionality of chromatin-related and protein degradation (F-box, ARIADNE, and ubiquitin ligases) genes for all genes that pass M2a/M1a LRTs significance and are DE in an L2 cluster between 3-5 DAP. **F**, The distribution of M2a/M1a LRTs for CZE DE genes split by the groups in E, shown with DE genes without significant M2a/M1a LRTs. **G**, Translated ARIADNE genes DE in the CZE, residues that are likely under positive selection are colored based on their Bayes Emprical Bayes (BEB) posterior probability of having a dN/dS > 1 under the M2a model. The majority of positively selected sites in CZE-DE ARIADNE genes fall into an N-terminal intrinsically disordered domain or an In Between Ring (IBR) fingers domain. **H**, The same as in G, but for a subset of nucleic acid-associated genes DE in the CZE. The yellow annotation corresponds to the domain listed to the right of the gene; purple corresponds to intrinsically disordered regions. All genes shown in G and H have statistically significant M2a/M1a LRTs. Domain coordinates were taken from InterPro for the primary isoform.

Level 3 annotations:



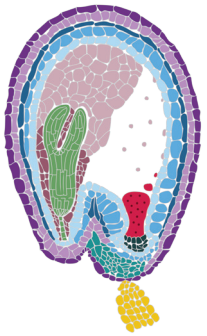
3 DAP

1. EMB (PDF1)
2. CZE nodule-like (MEA)
3. CZE basal cyst (RALFL3)
4. CZE nodule (CYCD4;2)
5. MCE (ARF13)
6. PEN general (AGL62)
7. PEN G2/M-phase (AGL62 and TOPII)
8. FUN hilum (ADPG1)
9. FUN epidermal (CER3)
10. FUN companion cells (SUC2)
11. OVL scenescing (BFN1)
12. CPT transient nucellus (PLL24)
13. CPT persistent nucellus (AGO3)
14. CZSC fluid transport (ARR22 and AAP6)
15. CZSC placentochalazal phosphate transport (ARR22, SWEET10, and UMAMIT14)
16. CZSC placentochalazal callose-synthesizing (ARR22, TET5, and CALS8)
17. ii1 endothelium (BAN and TT1)
18. ii1 basal (contains pigment strand) (BAN and SWEET4)
19. ii1' (DELTAVPE and UMAMIT29)
20. ii2 general (DELTAVPE and STK)
21. oi1 flavinol synthesizing (MYB11)
22. oi1 flavinol synthesizing high PCD (MYB11, SCOOP15, BAP1)
23. oi1 flavinol synthesizing S-phase (MYB11 and BARD1)
24. oi micropylar (SWEET12)
25. oi2 general (GLABRA2)
26. oi2 very long chain fatty acid-synthesizing (KCS6 and WRI4)
27. oi2 general (GLABRA2 and WRKY40)
28. oi2 S-phase (GLABRA2 and BARD1)



5 DAP

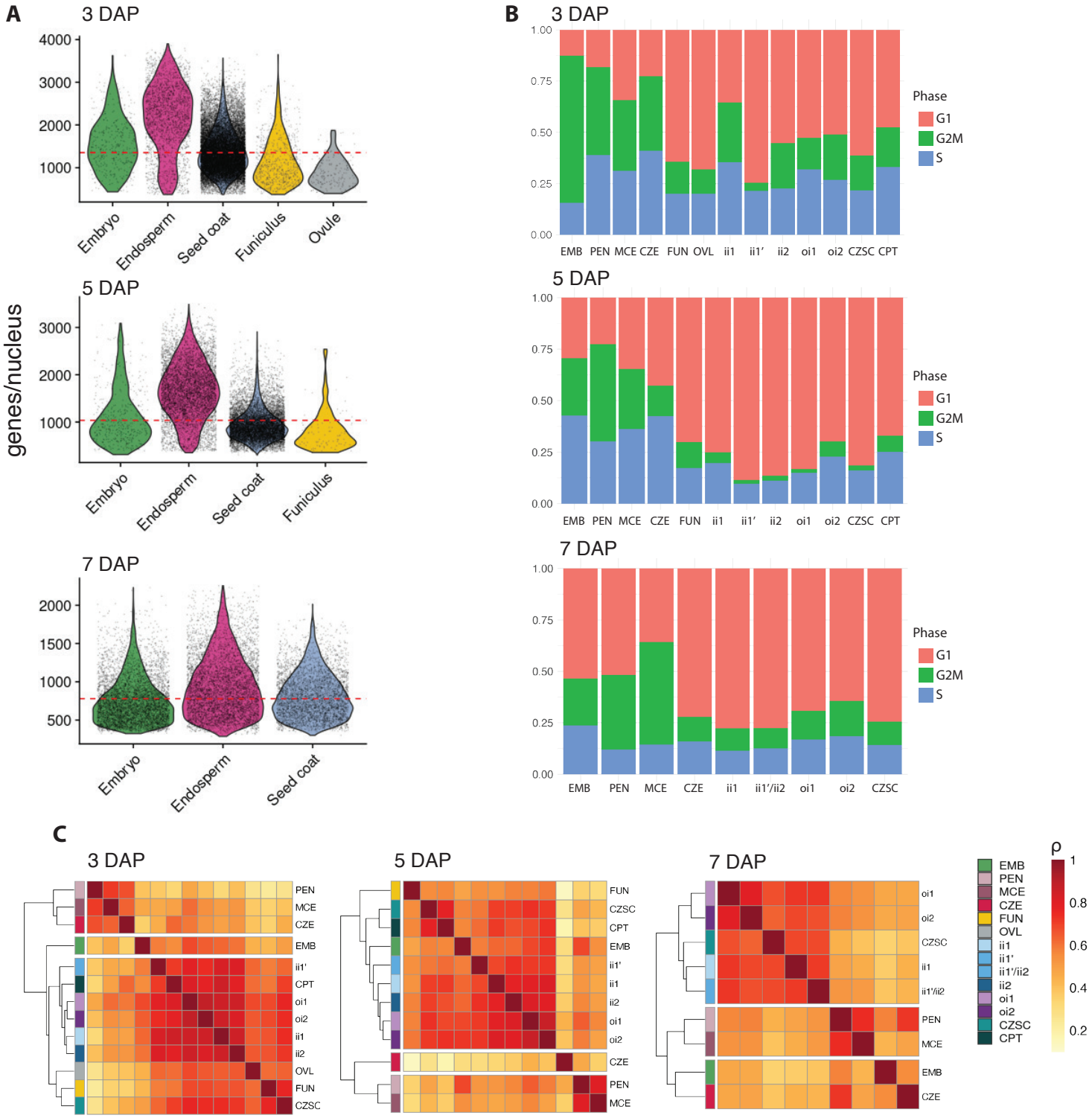
1. EMB (PDF1)
2. CZE nodule-like (MEA)
3. CZE apical cyst (MEE56)
4. CZE basal cyst (RALFL3)
5. CZE nodule (CYCD4;2)
6. MCE (ALE1)
7. PEN general I (IMK2)
8. PEN G2M-phase (MAD2)
9. PEN general II (NAC075)
10. PEN general V (PTR3)
11. PEN S-phase (HMGB6)
12. PEN general IV (TGA9)
13. FUN epidermal (CER3)
14. CPT persistent nucellus (AGO3)
15. CZSC placentochalazal phosphate transport (ARR22, SWEET10, and UMAMIT14)
16. CZSC placentochalazal callose-synthesizing (ARR22, TET5, and CALS8)
17. ii1 basal (contains pigment strand) (BAN and ARR22)
18. ii1 endothelium (BAN and TT1)
19. ii1' (DELTAVPE and UMAMIT29)
20. ii2 general (DELTAVPE and STK)
21. oi1 flavinol synthesizing (MYB11)
22. oi1 basal (MYB11 and PIN3)
23. oi1 flavinol synthesizing high PCD (MYB11, SCOOP15, BAP1)
24. oi micropylar (SWEET12)
25. oi1 S-phase (MYB11 and TSO2)
26. oi2 LTP-enriched (LTP1)
27. oi2 general (GLABRA2)
28. oi2 very long chain fatty acid-synthesizing (KCS6 and WRI4)
29. oi2 S-phase (GLABRA2, TSO2)



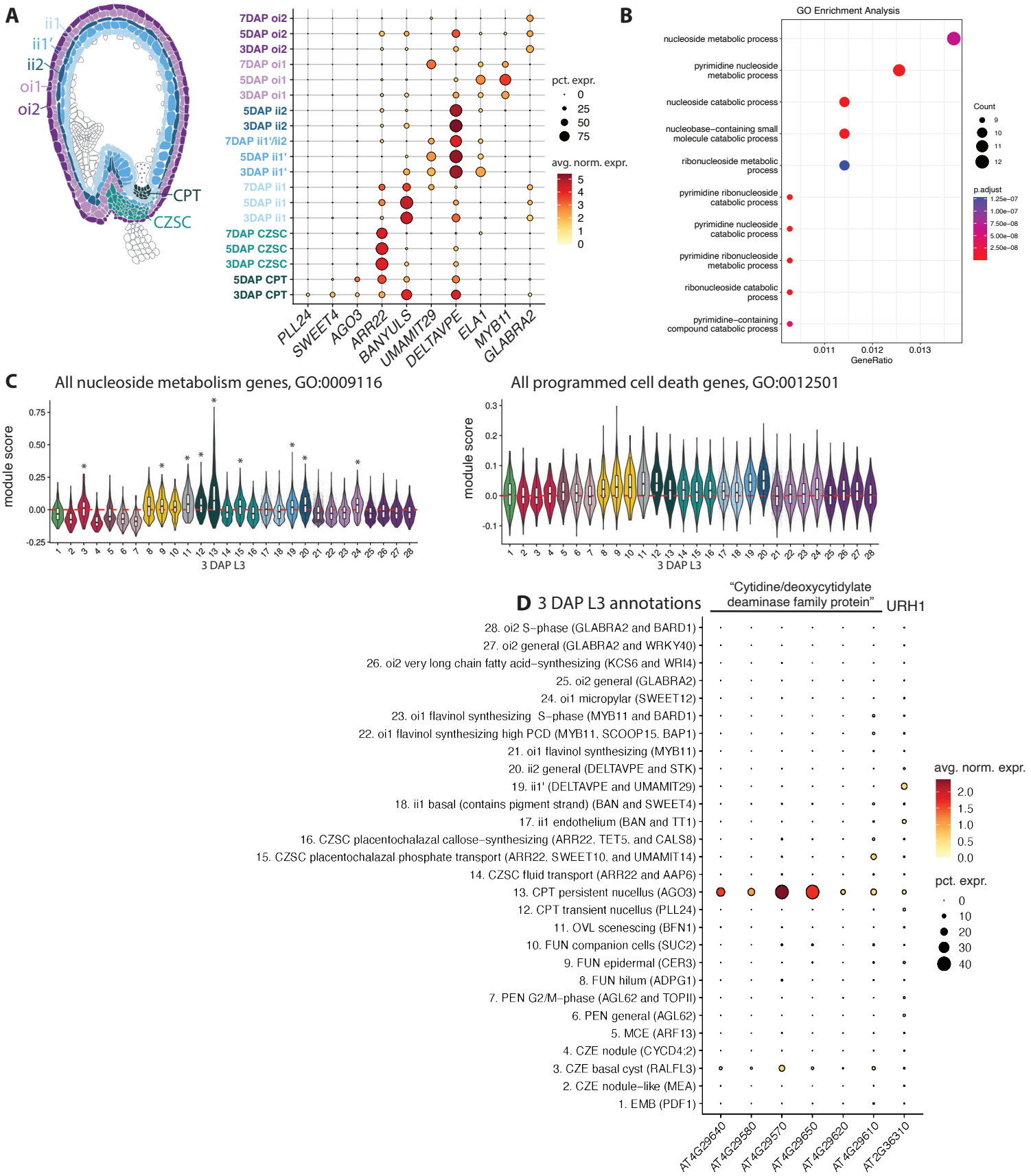
7 DAP

1. EMB inner cotyledon (XYL4)
2. EMB protoderm (PDF1)
3. EMB dividing (MAD2, HMGB6)
4. EMB vascular primordium (SHR)
5. EMB ground tissue initials (JKD, SCR)
6. EMB cortical initials (JKD)
7. EMB hypophysis (SMB)
8. CZE photosynthetic (GA3OX4)
9. CZE LEA-enriched (ABR)
10. CZE disaccharide metabolic (ARR21)
11. MCE_ESR high PCD (ZOU and NAC074)
12. MCE_ESR sheath promoting (ZOU, RGP3 and KRS)
13. PEN (ABS6)
14. PEN dehydrin enriched (ERD10)
15. CZSC general (ARR22, SWEET10, and TET5)
16. ii1 basal (contains pigment strand) (BAN and SWEET4)
17. ii1 endothelium (BAN and TT1)
18. ii1'/ii2 (DELTAVPE)
19. oi1 flavinol synthesizing (MYB11)
20. oi1 polyester monomer biosynthesis (MYB11 and CYP86B1)
21. oi2 general (GLABRA2)

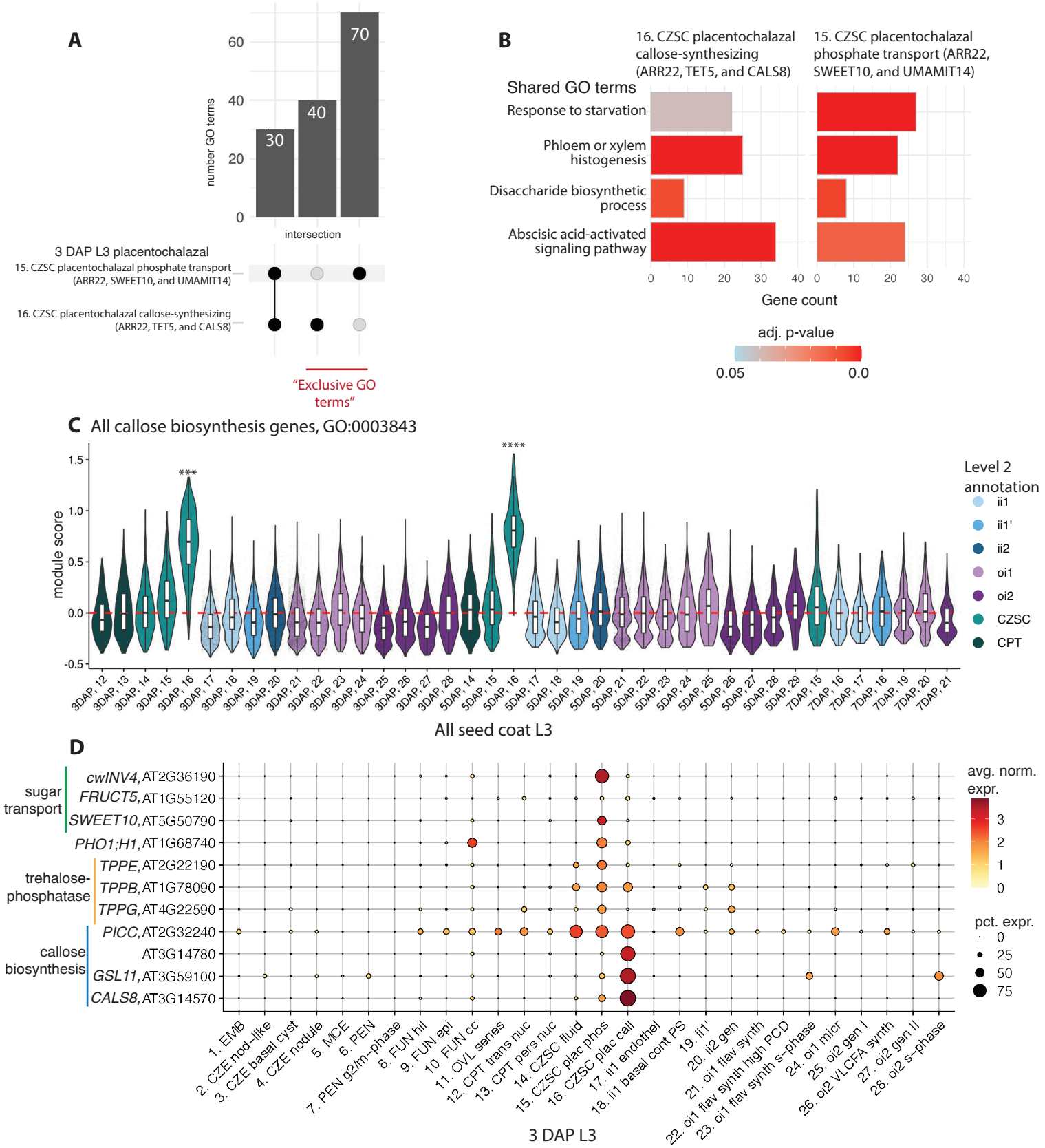
Extended Data 1. Level 3 annotation key. Full names for level 3 annotations for all timepoints, with defining marker genes in parentheses. EMB, embryo; CZE, chalazal endosperm; MCE, micropylar endosperm; PEN, peripheral endosperm; FUN, funiculus; OVL, ovule 5 days after emasculation; CPT, chalazal proliferating tissue; CZSC, chalazal seed coat; ii1, inner integument 1; ii1′, inner integument 1′; ii2; inner integument 2; oi1, outer integument 1; oi2, outer integument 2.



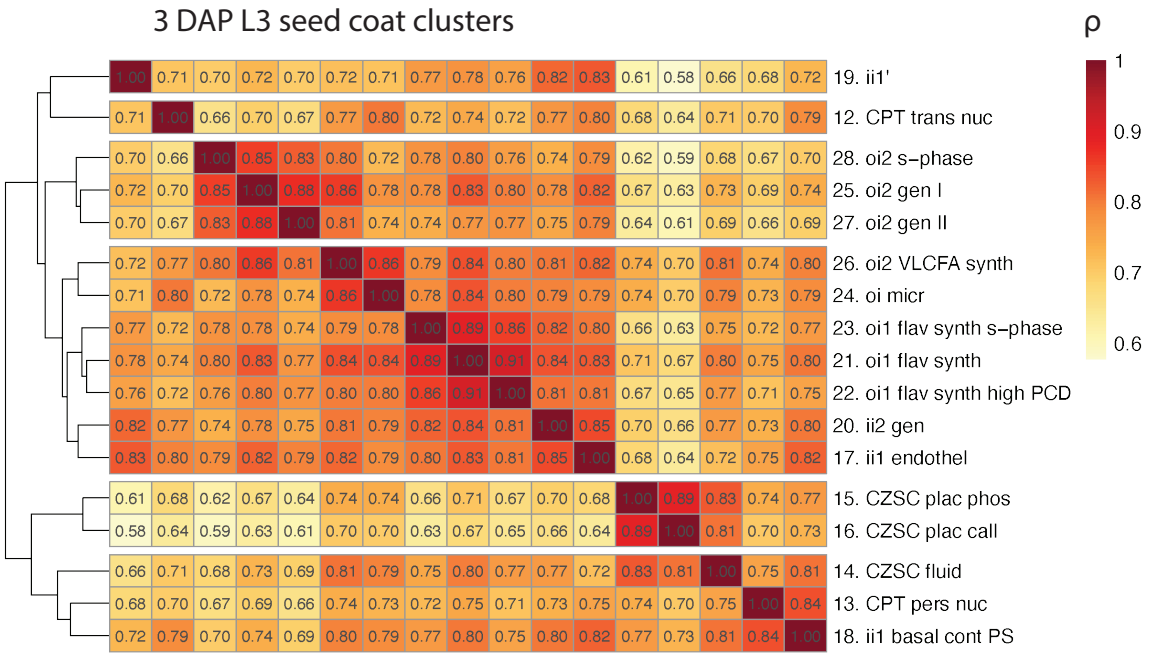
Extended Data 2. General cluster characteristics and similarity. A, Genes per nucleus distributions for L1 annotations across all timepoints. Red dotted line drawn at the median. **B**, cycle stage proportionality for L2 clusters across all timepoints. **C**, Clustered heatmaps of the Spearman correlation coefficients for the L2 cluster aggregated expression of highly variable genes across all timepoints.

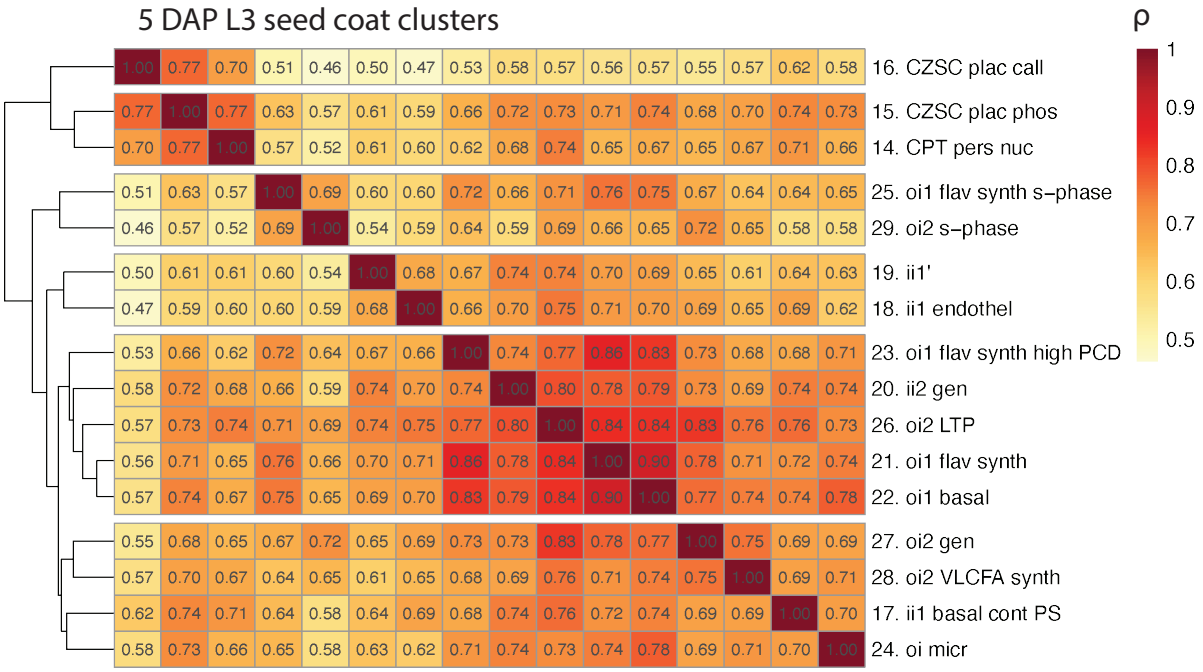


Extended Data 3. Informative markers for seed coat layers and nucleoside metabolism enrichment in the chalazal proliferating tissue. A, All markers except *AGO3* and *MYB11* have previously characterized layer-specific expression patterns. *AGO3* expression in the persistent CPT was validated in this study (Fig. 2), and *MYB11*, a regulator of flavanol biosynthesis, is the top marker for oi1 in this dataset. **B**, The most significant GO terms for the 3 DAP CPT, based on differentially expressed genes in a cluster vs. all 3 DAP clusters comparison (log2FC > 1, adj. p-value <0.05). **C**, GSEA results for all nucleoside metabolism (left) and programmed cell death (right) genes in 3 DAP clusters (p-value *<= 0.05,** <= 0.01, ***<= 0.001, and ****<= 0.0001 for all figures). **D**, the expression patterns of the top genes that underlie nucleoside metabolism enrichment in the 3 DAP persistent CPT.

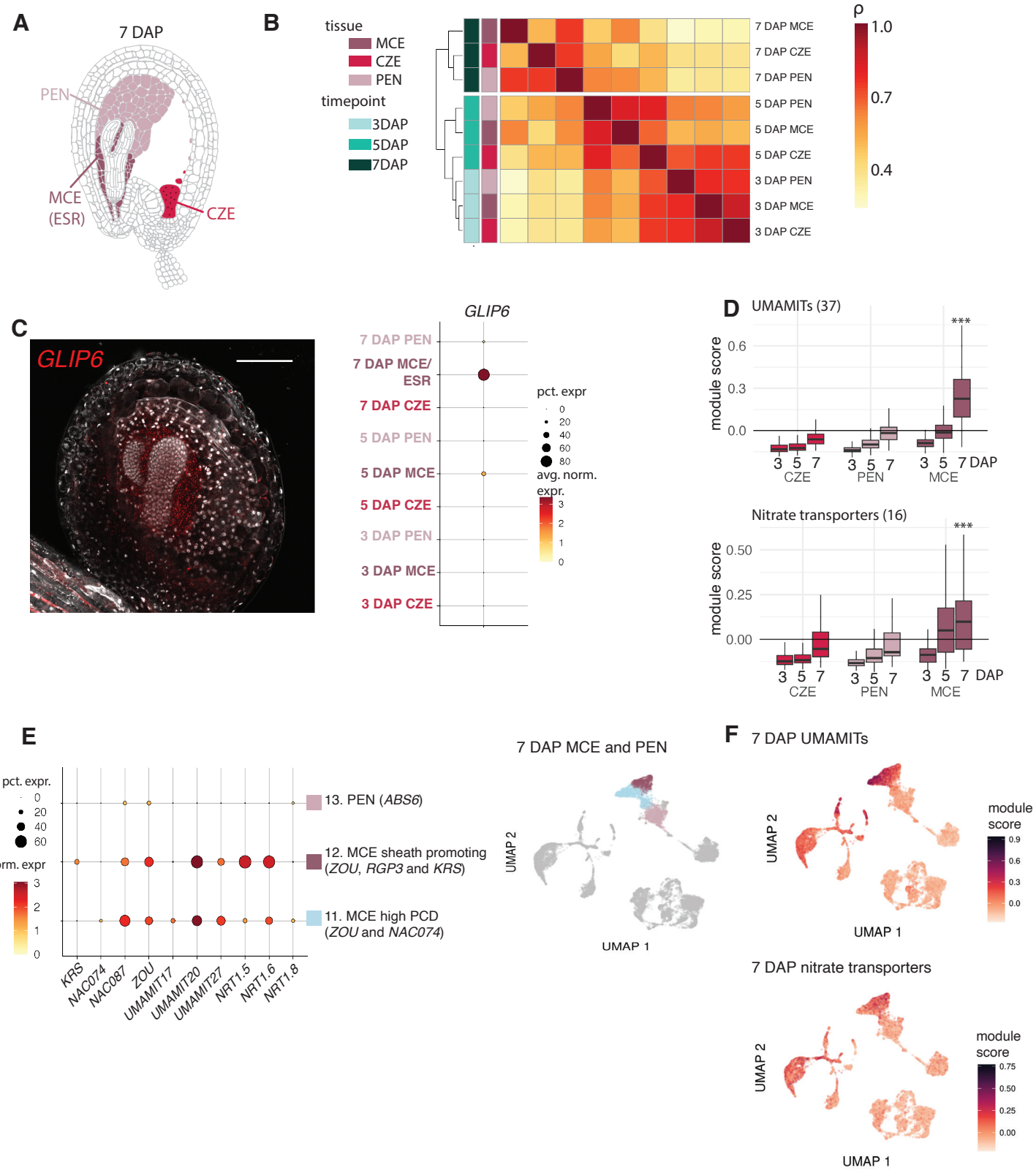


Extended Data 4. Cell types with complementary functions in the 3 DAP CZSC. A, GO term overlap for genes DE in two putative placentochalazal clusters. Exclusive GO terms are those associated with only one of the clusters. **B**, A subset of shared GO terms for the two putative placentochalazal clusters at 3 DAP. **C**, GSEA for all genes associated with the GO term "callose biosynthesis" shown across all L3 seed coat clusters in the atlas (p-value *<= 0.05,** <= 0.01, ***<= 0.001, and ****<= 0.0001 for all figures). **D,** Expression patterns of genes that underlie exclusive GO terms for the two putative CZSC clusters at 3 DAP. See Supplementary Table 3 to match abbreviated L3 names to their full descriptions.

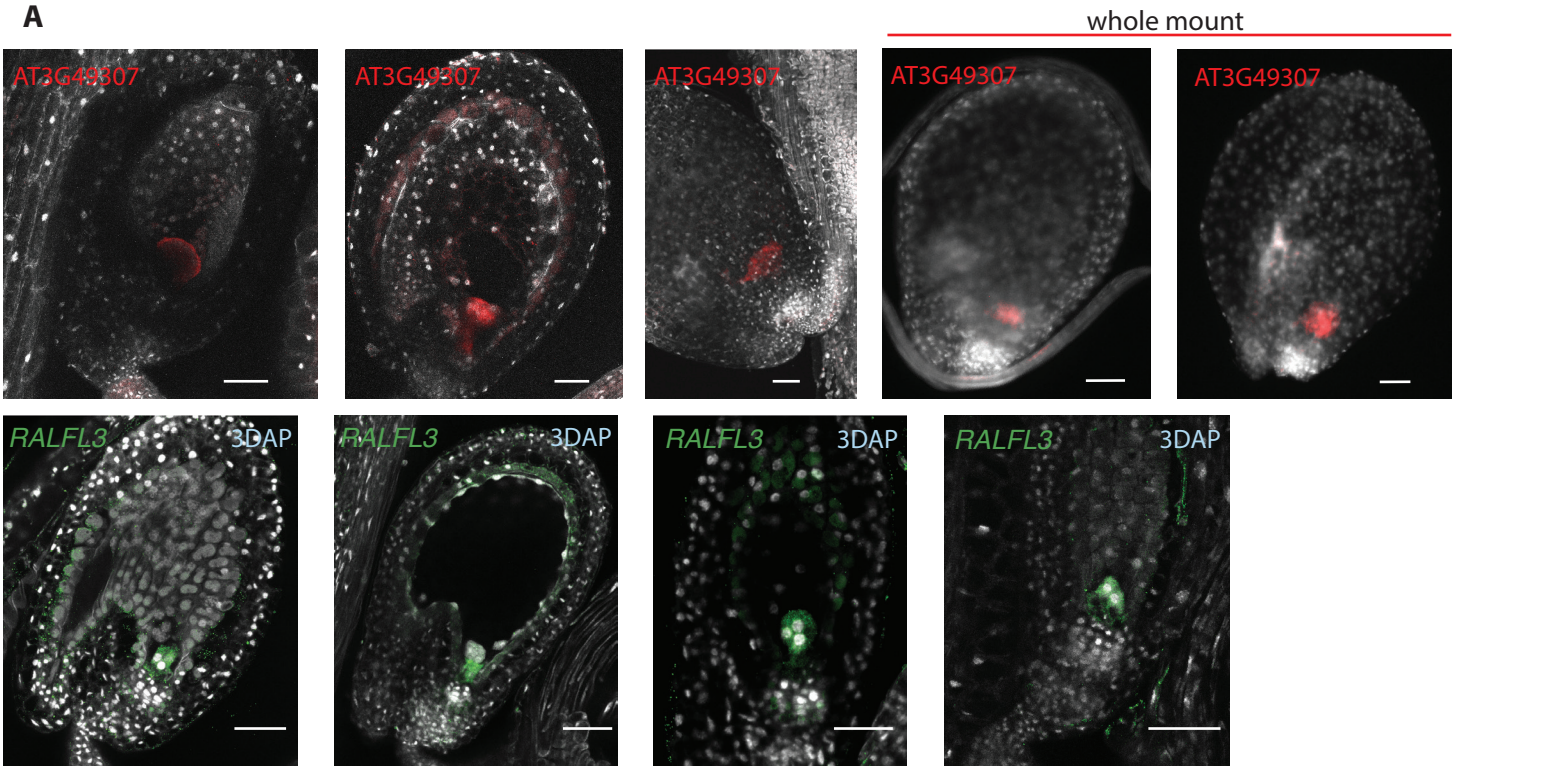


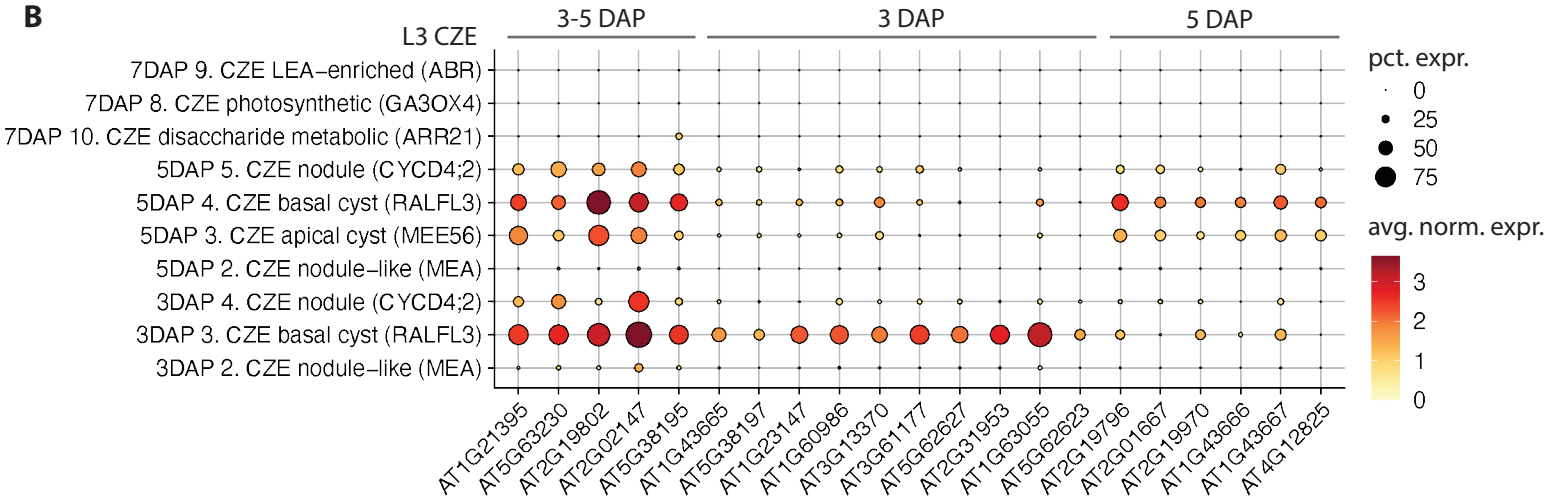


Extended Data 5. Level 3 seed coat cluster similarity at 3 and 5 DAP. Spearman correlation coefficient in pairwise comparisons of the aggregated expression of highly variable genes within L3 clusters. See Supplementary Table 3 to match abbreviated L3 names to their full descriptions.

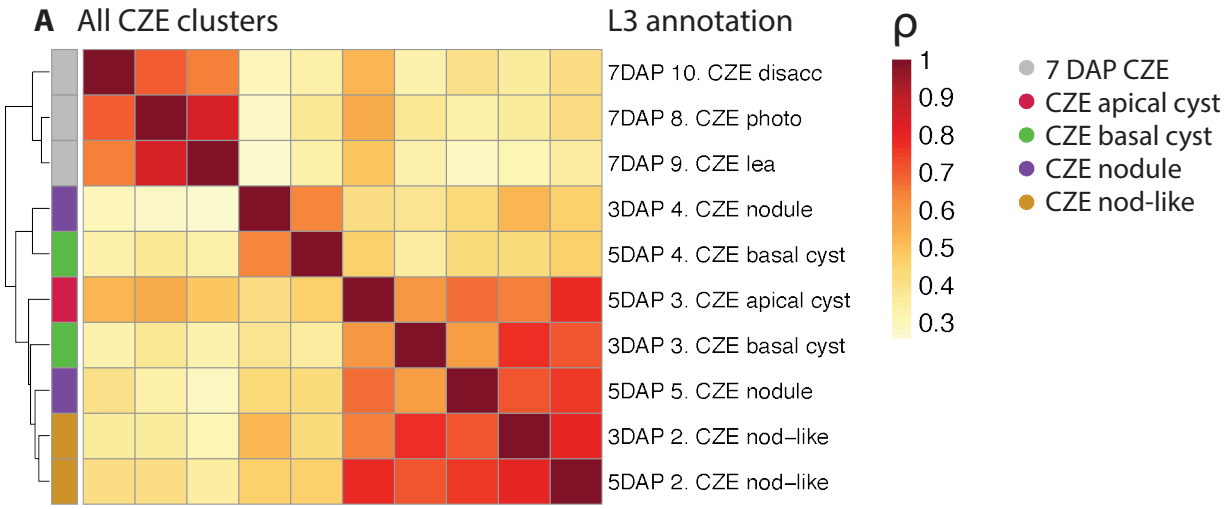


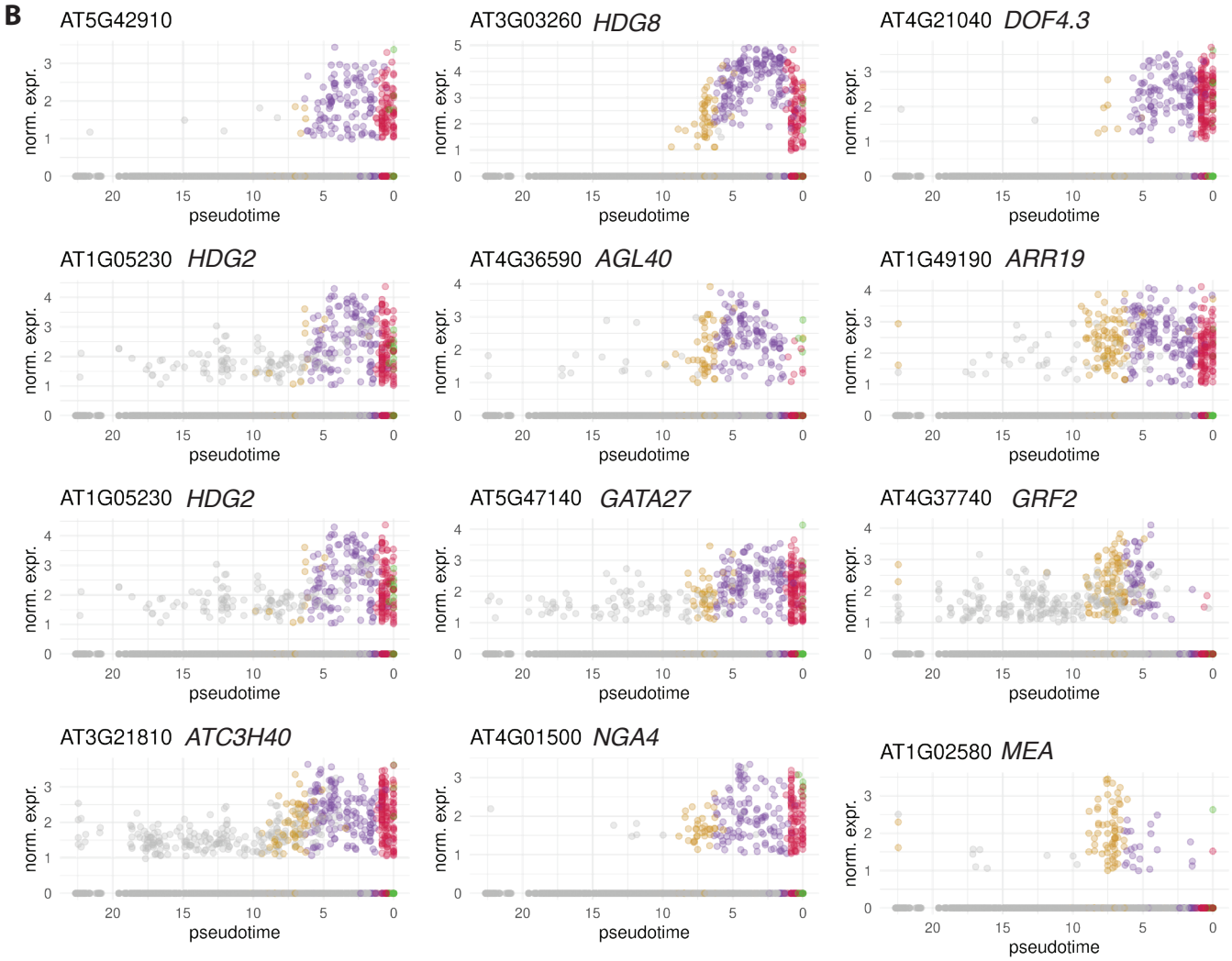
Extended Data 6. The 7 DAP MCE is the most distinctive endosperm subtype. A, L2 endosperm subtypes at 7 DAP. **B**, A clustered heatmap of the Spearman correlation coefficient (ρ) of aggregated expression for L2 endosperm clusters split by time. **C**, *GLIP6* is a highly specific marker for the ESR at 7 DAP, validated by HCR. Scale bar = 100µm. **D**, Gene set enrichment analysis for all UMAMITs (top) and nitrate transporters (bottom) detected in the atlas for all L2 endosperm clusters (p-value *<= 0.05,** <= 0.01, ***<= 0.001, and ****<= 0.0001 for all figures). **E**, 7 DAP L3 MCE clusters show differential expression of *KRS* and two NAC transcription factors. **F**, Nitrate transporters follow the *KRS-NAC074* transcriptional gradient while UMAMITs are more broadly expressed.



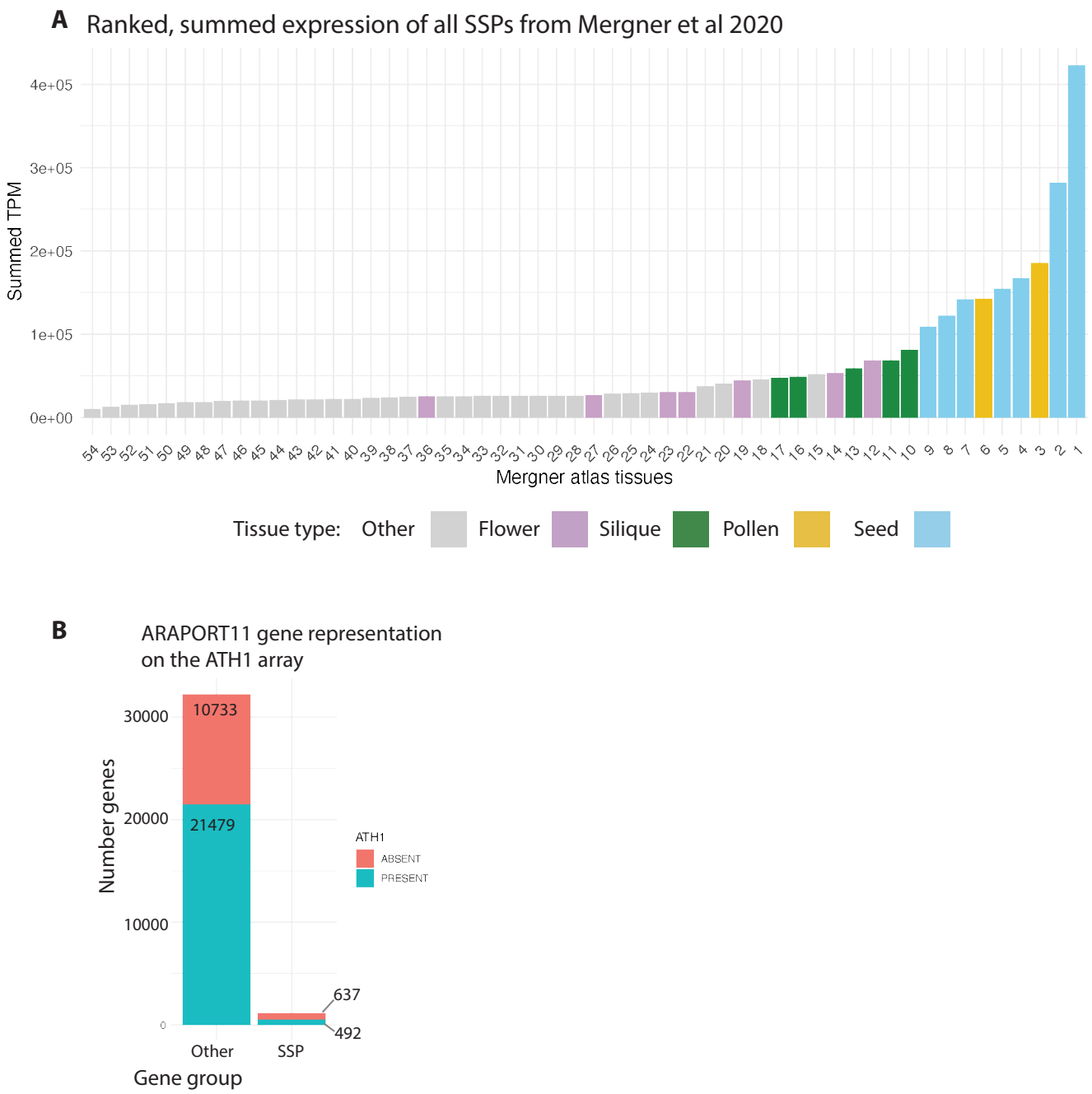


Extended Data 7. Differentially expressed genes along the apical/basal axis of the CZE. A, Top row: HCR for AT3G49307 (red), a gene enriched in the apical CZE during early to intermediate stages of development. These samples were not staged. The last two images are from whole mount seed preparations (Methods). Bottom row: HCR validation of *RALFL3* (green), a basal cyst marker, assayed on 3 DAP seeds. DAPI staining in white. Scale bar = 50µm. **B**, Expression patterns for the top basal cyst markers through development within L3 CZE subtypes.

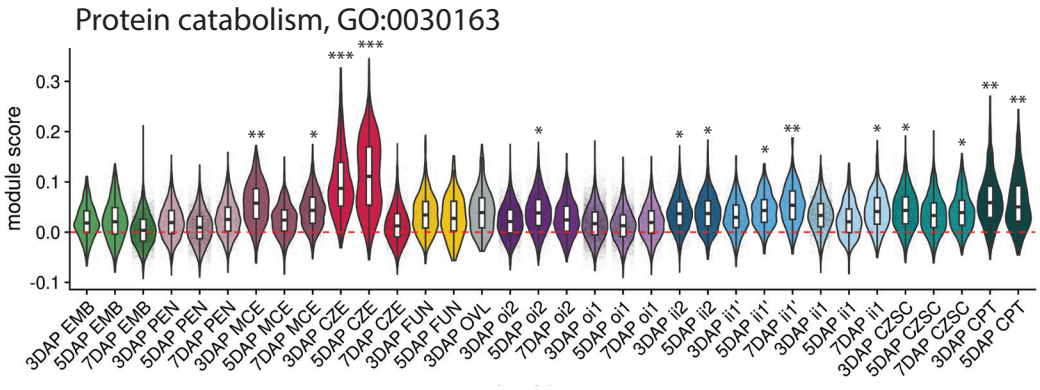


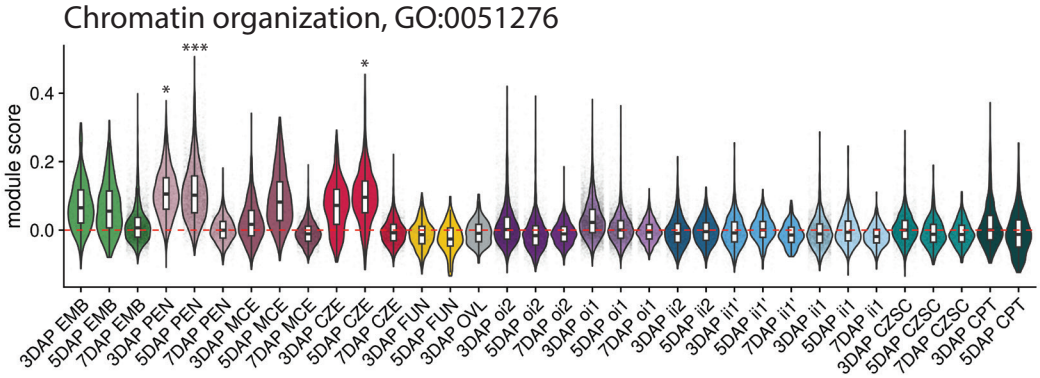


Extended Data 8. Genes that vary along the CZE nodule-to-cyst trajectory. A, Spearman correlation coefficients of the pairwise similarity between aggregated expression profiles of L3 CZE subtypes. 7 DAP L3 clusters could not be assigned apical/basal states. See Supplementary Table 3 to match abbreviated L3 names to their full descriptions. **B**, The normalized expression of a subset of genes that vary on 5 DAP PEN/CZE pseudotime is plotted for nuclei ordered by pseudotime.

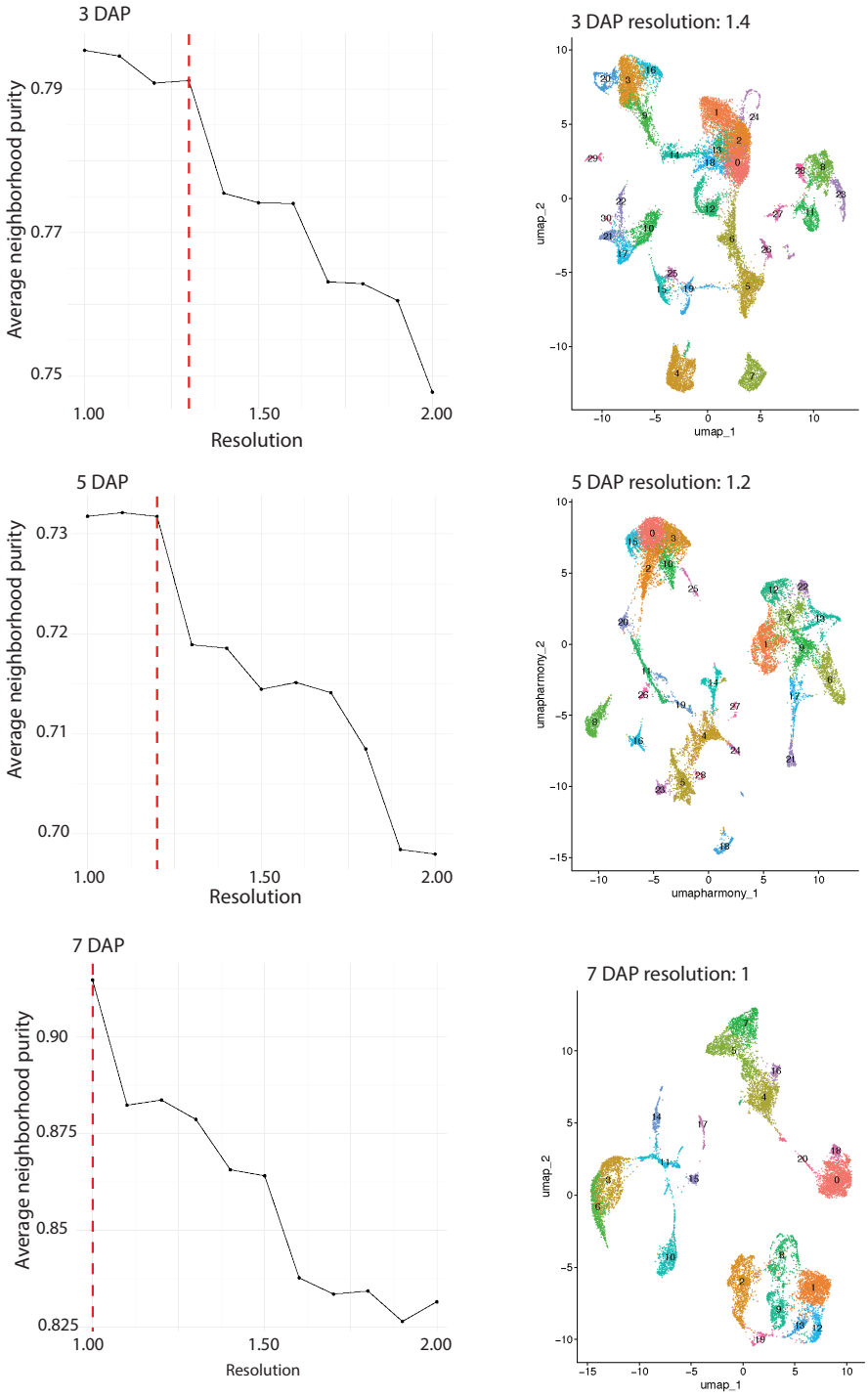


Extended Data 9. SSPs are highly expressed in seed tissues and were absent from existing transcriptional atlases of seed development. A, Ranked, summed TPM of all SSP genes used in this study using the bulk RNA-sequencing data of Mergner et al. 2020. See Supplementary Table 8 for sample names. **B**, ARAPORT11 gene annotation coverage of the Affymetrix ATH1 array.

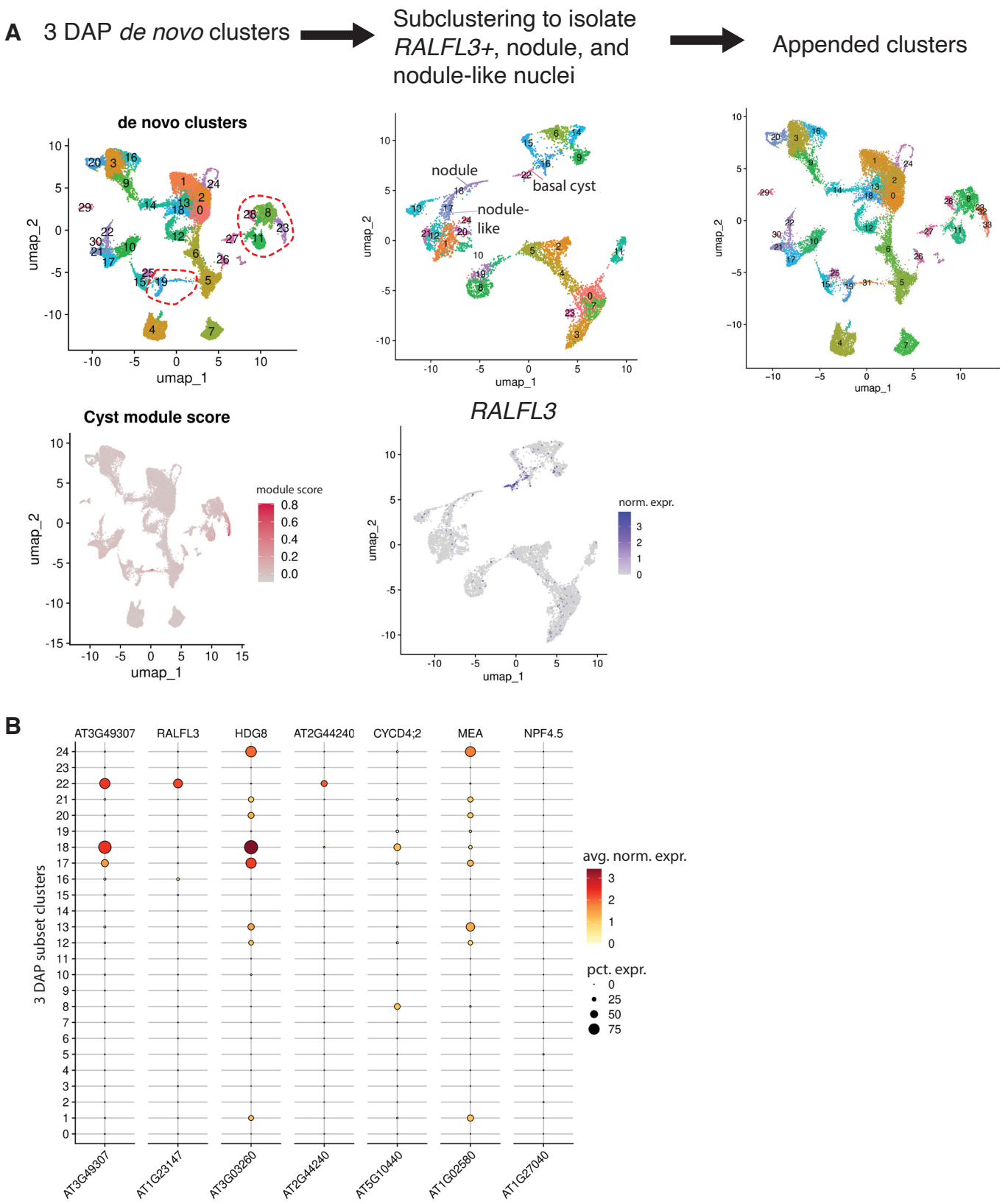




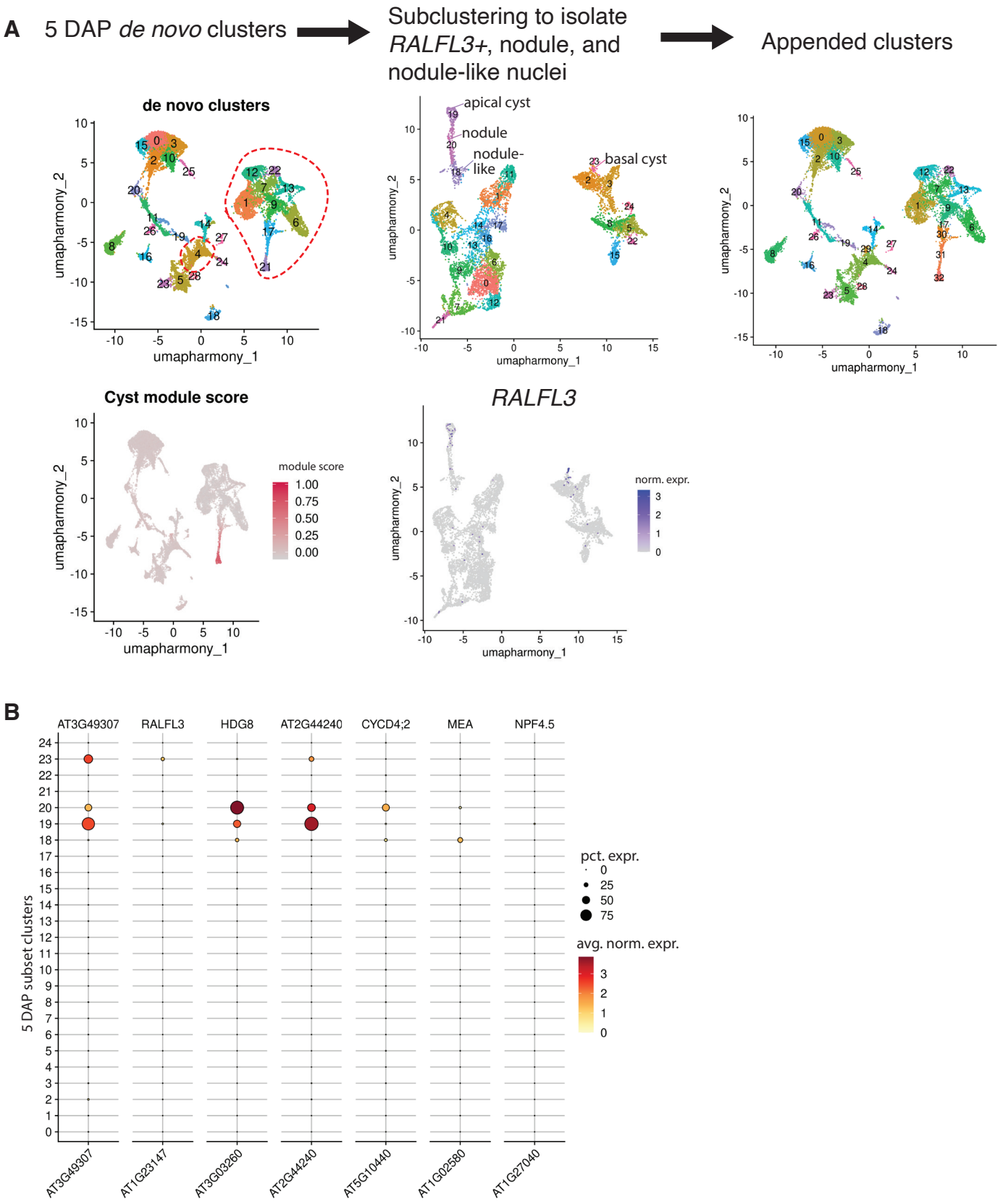
Extended Data 10. GO families containing rapidly evolving genes are enriched in the CZE and PEN. GSEA of all genes associated with protein catabolism and chromatin organization detected in the atlas across all L2 atlas clusters (p-value *<= 0.05,** <= 0.01, ***<= 0.001, and ****<= 0.0001 for all figures).



Supplementary Figure 1. Identifying an initial clustering resolution for level 3 annotations. Processed snRNA-seq libraries were merged within timepoints and subjected to a parameter sweep of clustering resolutions after initial dimensionality reduction. Each clustering resolution was evaluated by neighborhood purity, which is the percentage of neighbors that belong to the same cluster for each cell based on a weighted calculation. The left column contains the results of this parameter sweep for each timepoint, the red dotted line corresponds to the resolution that was selected for each timepoint. The right column displays the datasets with their initial clusters at the optimal clustering resolution.



Supplementary Figure 2. Isolating chalazal endosperm subtypes at 3 DAP through a subclustering analysis. **A**, Clusters expressing CZE markers from Picard et al. 2021 ("Cyst module") were subset from timepoint datasets and subject to an additional dimensionality reduction and subclustering. Published markers and those validated by HCR were used to annotated the subclusters (**B**), then the nuclei within endosperm clusters were manually labelled in the timepoint object.

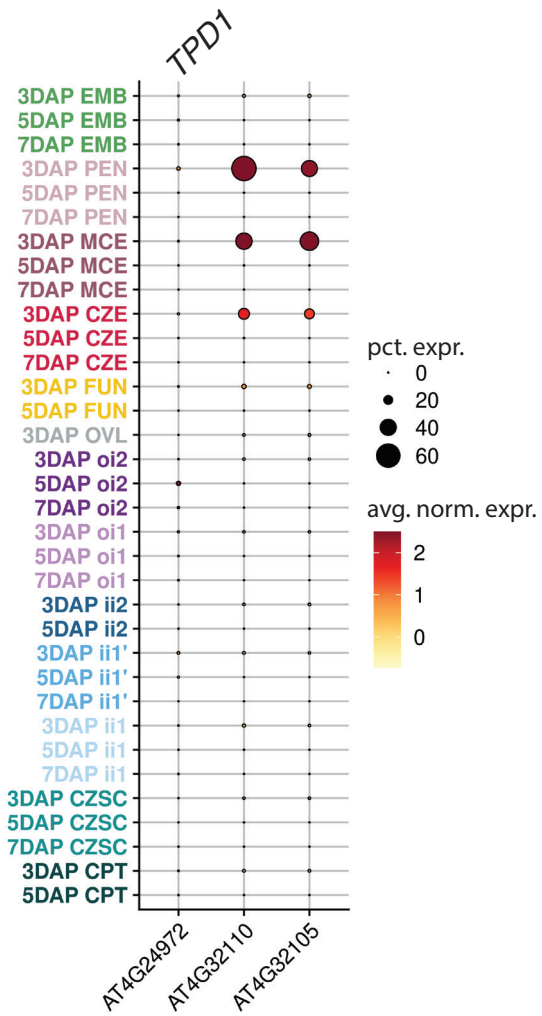


Supplementary Figure 3. Isolating chalazal endosperm subtypes at 5 DAP through a subclustering analysis. A, B, same procedure as described in Supplementary Figure 2, but for 5 DAP CZE nuclei.

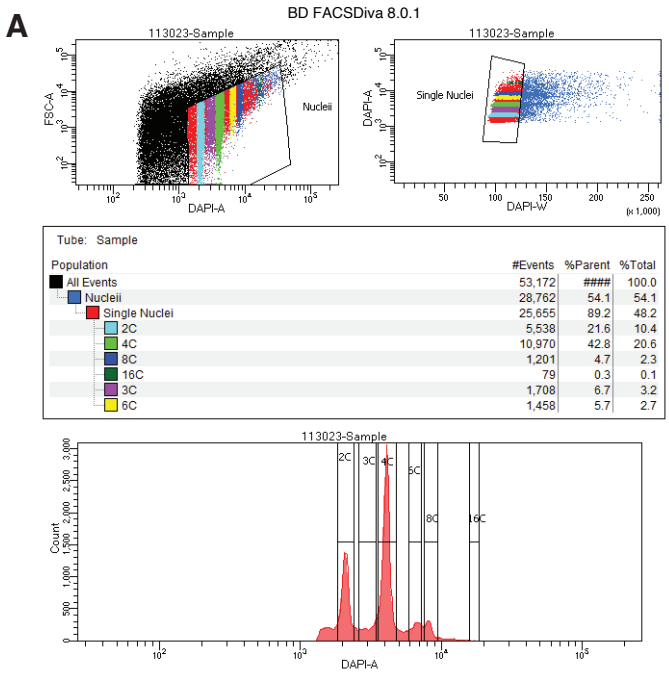
A

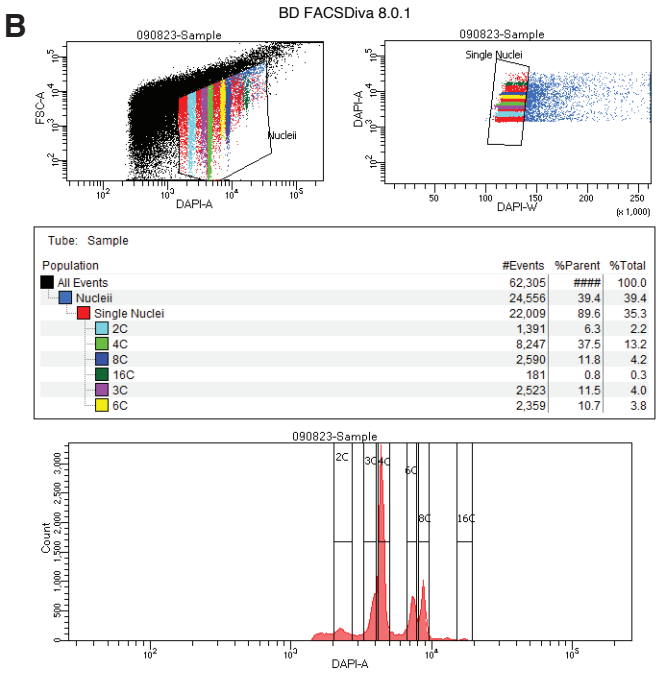
```
TPD1      --MNRRRLLVSATLLSYLLYGMALVSVEASGGEKLRDNLDLTKTTTSSPSISHRKML   55
AT4G32110 MASNACKLLCLVLFL-----------------------------------------A   16
AT4G32105 MASNACKVLCLVLFL-----------------------------------------A   16

TPD1      LLSPGTGKTERSVEPERIGEKCKSTDIVVNQAVTEPMPNGIPGYMVEITNQCMSGCI   112
AT4G32110 FVNQGY-------------GDCSLNSLSVKQSKTGKLVQNKPEWEVRVTNPCNN-CK   59
AT4G32105 FVNQGY-------------GDCHLNYLSVKQSKTGKLVQNKPEWEVRVTNPCNN-CK   59

TPD1      ISRIHINCGWFSSAKLINPRVFKRIHYDDCLVNNGKPLPFGSTLSFHYANTFPY--H   167
AT4G32110 FQNTELLCVGFNSVTPIDTSLLL-KSGDACLVNAGKFIVPHVDIVFKYVWDTSFDLK   115
AT4G32105 FQYTKLSCVGFQSVTPVATSLLS-KSGDICLLNAGNFIFPHVDFVFNYVWDTSFDLK   115

TPD1      LSVAFVTCA   176
AT4G32110 VIDGVMVCY   124
AT4G32105 VIDGVIVCP   124
```

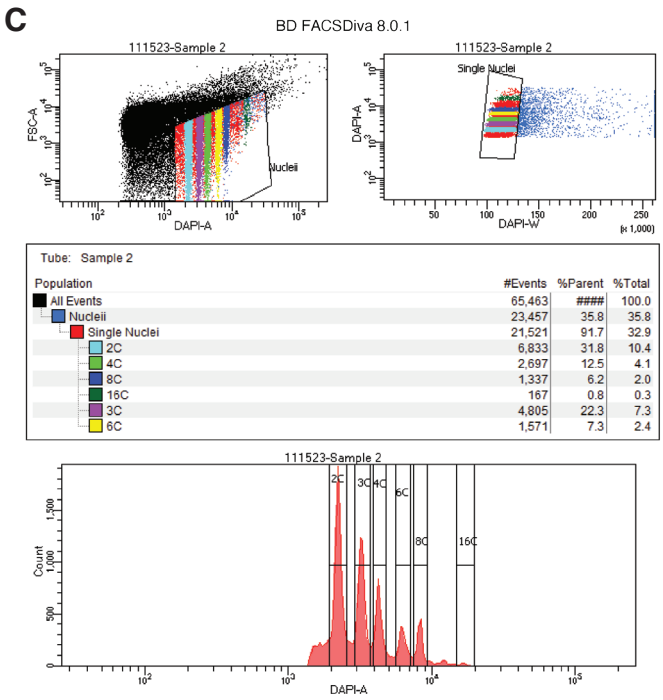
B



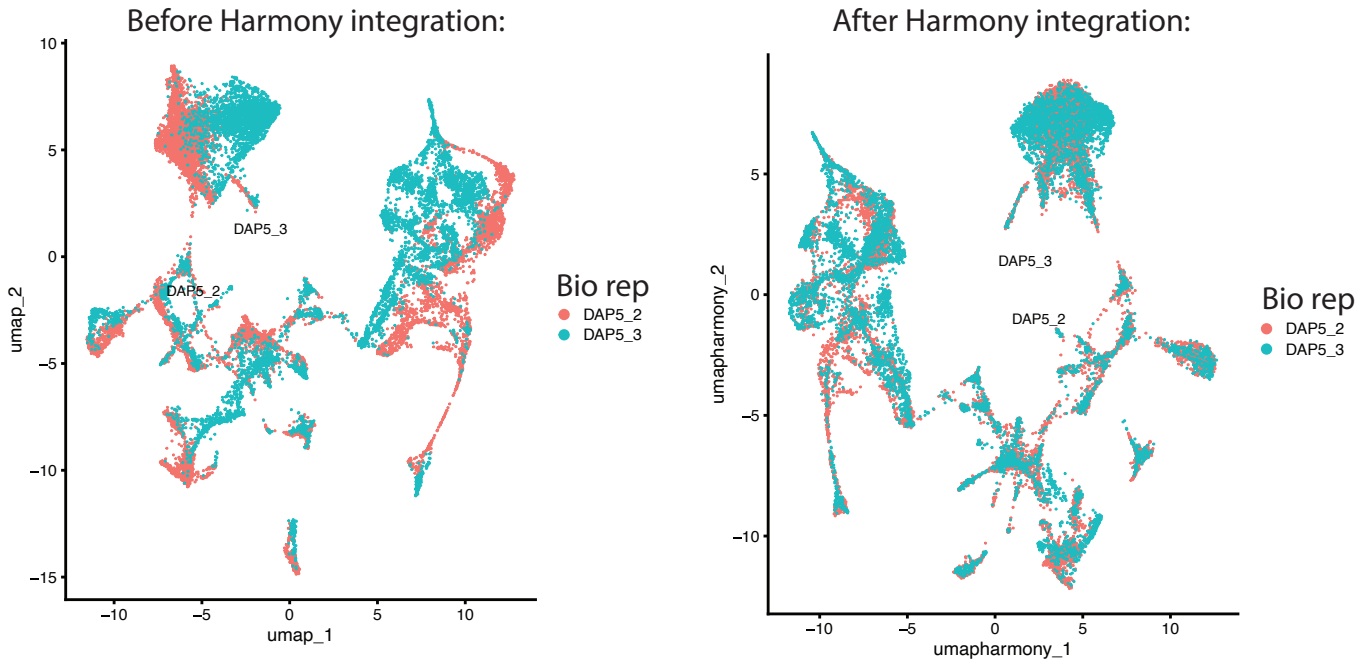
Supplementary Figure 4. TPD1- like peptides are expressed in the 3 DAP endosperm. A, Two TPD1-like peptides identified through a motif analysis in Ghorbani et al 2015, designated part of the "TPD tribe" in that study, in a protein alignment. A motif is conserved among all three peptides, including 6 cysteines, although overall protein % identity is low. **B**, Expression pattern of *TPD1* and TPD1-like peptides in all atlas L2 annotations.







Supplementary Figure 5. FANS gating scheme for purifying seed nuclei before snRNA-seq. A-C, Example gating parameters used for one of the 3-7 DAP replicates, respectively, all gates shown were collected.



Supplementary Figure 6. 5 DAP biological replicates before and after Harmony integration. Biological replicates showed improved mixing after Harmony integration.