Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed

Mark F. Belmonte[a,1,2], Ryan C. Kirkbride[a,1], Sandra L. Stone[a,3], Julie M. Pelletier[a], Anhthu Q. Bui[b,4], Edward C. Yeung[c], Meryl Hashimoto[a], Jiong Fei[a], Corey M. Harada[a], Matthew D. Munoz[a,5], Brandon H. Le[b], Gary N. Drews[d], Siobhan M. Brady[a,e], Robert B. Goldberg[b,6], and John J. Harada[a,6]

[a]Department of Plant Biology and [e]Genome Center, University of California, Davis, CA 95616; [b]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA 90095; [c]Department of Biological Sciences, University of Calgary, Calgary, AB, Canada T2N 1N4; and [d]Department of Biology, University of Utah, Salt Lake City, UT 84112

Contributed by Robert B. Goldberg, December 20, 2012 (sent for review December 7, 2012)

Seeds are complex structures that consist of the embryo, endosperm, and seed-coat regions that are of different ontogenetic origins, and each region can be further divided into morphologically distinct subregions. Despite the importance of seeds for food, fiber, and fuel globally, little is known of the cellular processes that characterize each subregion or how these processes are integrated to permit the coordinated development of the seed. We profiled gene activity genome-wide in every organ, tissue, and cell type of *Arabidopsis* seeds from fertilization through maturity. The resulting mRNA datasets offer the most comprehensive description of gene activity in seeds with high spatial and temporal resolution, providing unique insights into the function of understudied seed regions. Global comparisons of mRNA populations reveal unexpected overlaps in the functional identities of seed subregions. Analyses of coexpressed gene sets suggest that processes that regulate seed size and filling are coordinated across several subregions. Predictions of gene regulatory networks based on the association of transcription factors with enriched DNA sequence motifs upstream of coexpressed genes identify regulators of seed development. These studies emphasize the utility of these datasets as an essential resource for the study of seed biology.

laser-capture microdissection | mRNA localization | transcriptome

The significance of seeds is reflected by their relevance to diverse biological areas. Evolutionarily, the ability of flowering plants to make seeds has conferred significant selective advantages, accounting, in part, for their dominance among the Plantae. The seed habit facilitates fertilization in nonaqueous environments, provides protection and nutrients for the developing embryo, and permits the embryo to remain quiescent until conditions are favorable for seedling development (1). Seeds are a key to global food security, because they account for the large majority of calories consumed by humans. An estimated 70–100% more food will need to be produced worldwide by 2050 without an appreciable increase in arable land and despite global climate change (2). A detailed understanding of seed development may enable the design of cogent strategies to enhance seed quality and yield.

The developmental significance of seeds is that they are complex yet elegant structures, consisting of embryo, endosperm, and seed-coat regions that are each divided into subregions (3). The complexity of the seed originates with its precursor, the ovule, which consists of the female gametophyte embedded within integument layers. Seed development is initiated with the fusion of the egg and central cells of the female gametophyte with two sperm cells from the pollen tube. This double fertilization, unique to flowering plants, produces the progenitors of the embryo and endosperm regions of the seed, respectively. Patterning and morphological differentiation occur in the embryo and endosperm regions early in seed development, during the morphogenesis phase. In many plants, including *Arabidopsis*, the embryo undergoes stereotypic cell-division patterns, differentiating into the embryo proper that becomes the body of the vegetative plant and the suspensor, an ephemeral structure that serves as a conduit between the embryo proper and the seed coat (Figs. 1 *A*–*F*). The primary endosperm cell undergoes nuclear but not cell divisions, and nuclei migrate to form three subregions: micropylar, which is nearest the young embryo; peripheral, in the center of the endosperm region; and chalazal, at the pole opposite to the embryo. Cellularization of the endosperm proceeds in a wave-like manner from the micropylar to chalazal end (4). Ovule integument cells divide and differentiate into the distinct cell types of the seed coat that envelope the embryo and endosperm. Late in seed development during the maturation phase, the embryo accumulates storage macromolecules and becomes tolerant of desiccation. Although development of these subregions has been well-characterized morphologically, little is known of the cellular processes that occur in these subregions or how the development of the subregions is coordinated within the context of seed development.

A key to dissecting seed development is to obtain an integrated understanding of gene activity, and therefore the cellular processes that occur in seed regions throughout development.

Significance

Seeds are complex structures that are comprised of the embryo, endosperm, and seed coat. Despite their importance for food, fiber, and fuel, the cellular processes that characterize different regions of the seed are not known. We profiled gene activity genome-wide in every organ, tissue, and cell type of *Arabidopsis* seeds from fertilization through maturity. The resulting mRNA datasets provide unique insights into the cellular processes that occur in understudied seed regions, revealing unexpected overlaps in the functional identities of seed regions and enabling predictions of gene regulatory networks. This dataset is an essential resource for studies of seed biology.

Author contributions: M.F.B., R.C.K., S.L.S., J.M.P., E.C.Y., R.B.G., and J.J.H. designed research; M.F.B., R.C.K., S.L.S., J.M.P., A.Q.B., E.C.Y., M.H., J.F., and M.D.M. performed research; C.M.H., B.H.L., G.N.D., and S.M.B. contributed new reagents/analytic tools; M.F.B., R.C.K., S.L.S., J.M.P., E.C.Y., S.M.B., and J.J.H. analyzed data; and M.F.B., R.C.K., and J.J.H. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE12404).

[1]M.F.B. and R.C.K. contributed equally to the manuscript.

[2]Present address: Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada R3T 2N2.

[3]Present address: Research Services, University of Saskatchewan, Saskatoon, SK, Canada S7N 4J8.

[4]Present address: BASF Plant Sciences, Research Triangle Park, NC 27709.

[5]Present address: Bioinformatics and Medical Informatics Graduate Program, San Diego State University, San Diego, CA 92182.

[6]To whom correspondence may be addressed. E-mail: jjharada@ucdavis.edu or bobg@ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1222061110/-/DCSupplemental.

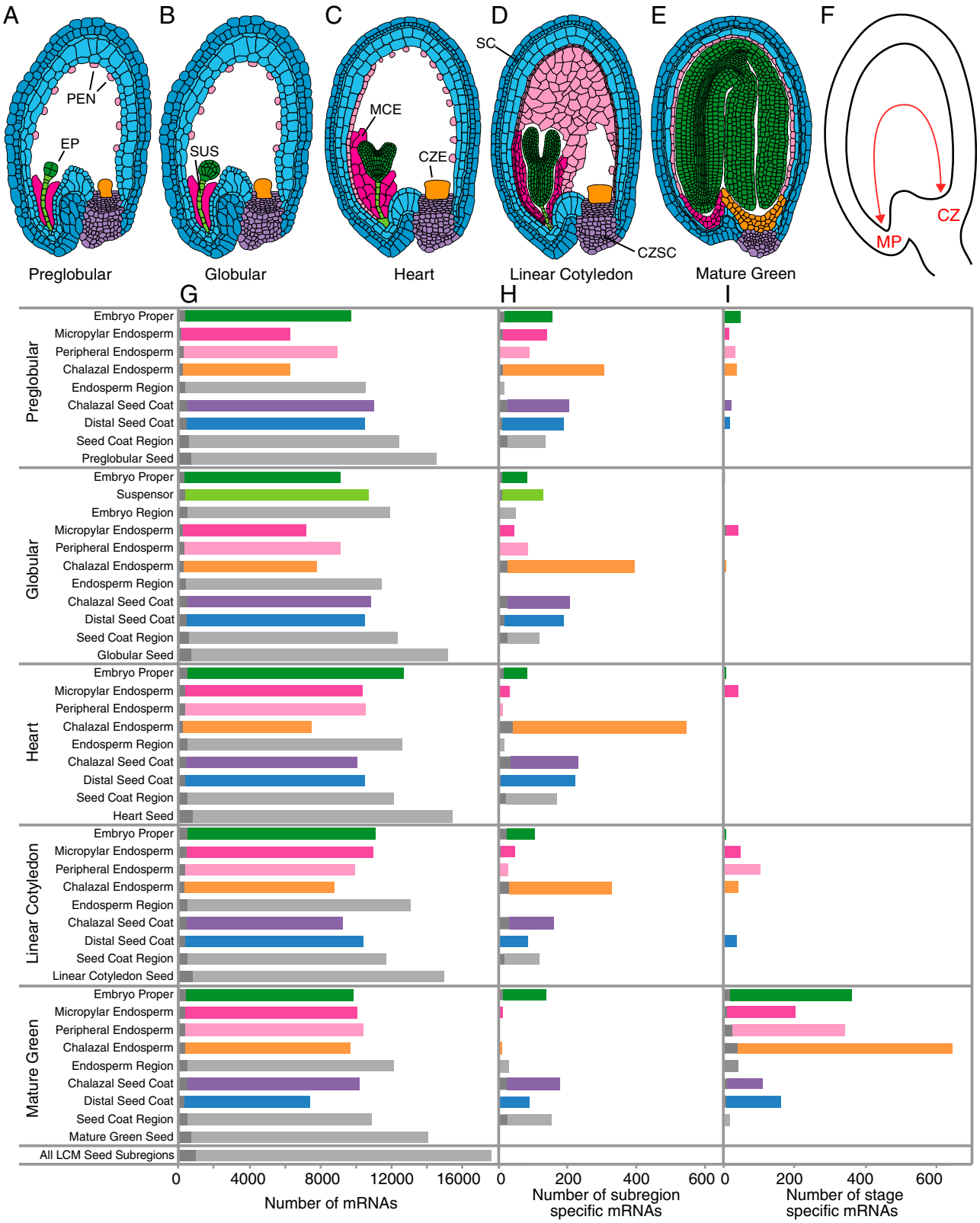

Fig. 1. Gene activity in *Arabidopsis* seed regions and subregions throughout Development. (*A–F*) Representation of *Arabidopsis* seeds at the (*A*) preglobular stage, (*B*) globular stage, (*C*) heart stage, (*D*) linear cotyledon stage, and (*E*) mature green stage. (*F*) Diagram of a seed showing micropylar (MP) and chalazal (CZ) poles. (*G*) Number of distinct mRNAs detected in seed subregions (colored bars), regions, and seeds (light gray bars) at different stages. Dark gray bars indicate the number of distinct transcription factor mRNAs. Lists of mRNAs and their levels are in Dataset S2. (*H*) Number of distinct mRNAs that accumulate specifically in a subregion or region at a given stage. Subregion and region-specific mRNAs are listed in Dataset S3. (*I*) Number of distinct mRNAs that accumulate at a specific stage in a subregion or region. Stage-specific mRNAs are listed in Dataset S3. Abbreviations are given in Table 1.

We previously analyzed genes expressed in developing whole seeds at several developmental stages and identified seed-specific genes and transcription factors, and these data provided clues about temporally regulated processes that occur during seed development (5). Questions remain, however, about the processes that occur specifically in a subregion and the interactions among different regions. A number of recent studies have reported gene activity in specific seed regions at the whole-genome level, such as the embryo, endosperm, and seed coat (reviewed by ref. 6). However, these studies do not enable an integrated understanding of seed development, because most focused on a specific stage of seed development and few reported gene activity in more than one region. Here, we describe gene activity genome-wide in all subregions and regions of seeds of the model plant *Arabidopsis*, from fertilization through maturity. The temporal and spatial integration of cellular and physiological processes in multiple subregions and stages permits seminal insights into the developmental processes that characterize specific seed regions and the gene regulatory programs that underlie seed development. Use of uniform platforms of subregion isolation and RNA analyses permit direct comparisons of mRNA levels in different subregions and stages, enabling an integrated understanding of seed development.

Results

Spatial and Temporal Resolution of mRNA Profiles During Seed Development. We profiled mRNA populations from six to seven seed subregions at five stages of development (Fig. 1 *A–E* and Fig. S1 *A–E*) to obtain the most comprehensive description of gene activity in seed development, representing 31 combinations of subregions and stages. Laser-capture microdissection (LCM) (*Materials and Methods*) was used to isolate the embryo proper (EP) and suspensor (SUS) of the embryo region, micropylar (MCE), peripheral (PEN), and chalazal (CZE) subregions of the endosperm region, and the chalazal (CZSC) and distal (SC) seed coat (Fig. S1 *F–Q* and Dataset S1, Table S1). The subregions were isolated in replicate at the preglobular, globular, and heart stages that collectively represent the morphogenesis phase (Fig. 1 *A–C*). Subregions isolated at the mature green stage correspond to the maturation phase (Fig. 1*E*), whereas the linear-cotyledon stage (Fig. 1*D*) is a transition between the two phases. All subregions and stages and their abbreviations are listed in Table 1. The ephemeral SUS was isolated only at the globular stage because an average of 1,700 captured subregions were needed for each biological replicate (Dataset S1, Table S2). mRNAs in each subregion were detected and quantified using stringent analyses of Affymetrix ATH1 GeneChip hybridization data (*Materials and Methods* and Dataset S2). These data are available at the Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo) as series GSE12404 and in an interactive form at http://seedgenenetwork.net. Exhaustive control experiments, including validation of relative mRNA levels using quantitative RT-PCR (qRT-PCR), comparisons of mRNA accumulation patterns with promoter activities, and confirmation that mRNA sequence amplification was unbiased, established that the LCM seed dataset is representative of subregion RNA populations, qualitatively and quantitatively (Fig. S2 and Dataset S1, Tables S3 and S4).

Fig. 1*G* summarizes gene activity in subregions, regions, and whole seeds at each developmental stage. We detected between ~6,000 and 13,000 distinct mRNAs in each subregion. The number of mRNAs detected in a region, calculated as the union of mRNAs present in each of its constituent subregions, was not appreciably higher than that of individual subregions. Similarly, the union of mRNAs present in embryo, endosperm, and seed-coat regions, representing whole-seed mRNA number, was not appreciably higher than that of a single region. These results indicate that there is substantial overlap in the genes expressed in regions and subregions of a seed. An average of ~14,800 distinct mRNAs was detected throughout the seed at each stage, and collectively a minimum of 17,594 distinct mRNAs were detected in at least one subregion and stage of seed development. These are minimum estimates given the stringency of the analyses, the absence of probes for ~17.5% of *Arabidopsis* genes on the ATH1 GeneChip, and the lack of consistent detection of mRNAs present at the lowest prevalence levels (fraction of mRNA $< 10^{-5}$) in GeneChip experiments (5).

Table 1. Abbreviations for developmental stages and seed subregions

Abbreviation	Region
Stage identifiers	
pg	Preglobular
g	Globular
h	Heart
lc	Linear cotyledon
mg	Mature green
Subregion identifiers	
EP	Embryo proper
SUS	Suspensor
MCE	Micropylar endosperm
PEN	Peripheral endosperm
CZE	Chalazal endosperm
CZSC	Chalazal seed coat
SC	Distal seed coat

Global Comparisons of mRNA Populations Reveal Functional Relationships Among Seed Subregions. To provide a framework to understand how seed development is coordinated, we compared mRNA populations from all 31 combinations of subregions and stages globally using principal component analysis (PCA). The rationale was that subregions identified as being most closely associated in the analysis were likely to share the greatest similarity in overall gene expression and, therefore, cellular functions. Fig. 2 shows that four groups of subregions were identified by the analysis: (*i*) the EP, SUS, MCE, and PEN subregions at the preglobular to heart/linear cotyledon stage (green); (*ii*) CZE subregions at these early stages (yellow); (*iii*) all embryo and endosperm subregions late in development (red); and (*iv*) all CZSC and SC subregions at all stages (blue).

The finding that the MCE and PEN early in development shared greater similarity with both embryo subregions than with the CZE was surprising, because the embryo and endosperm arise from separate fertilization events. Late in development, all embryo and endosperm subregions, including the CZE, formed a group that was distinct from the same subregions early in development, suggesting that a major shift in gene expression occurs during the transition from early to late stages. The CZSC and SC at all stages grouped, suggesting that these subregions of the seed-coat region share greater similarity with each other than with embryo and endosperm subregions. Identical results were obtained using hierarchical clustering of mRNA populations from all subregions and stages (Fig. S3*A*), further supporting the biological significance of the groupings. The same relationships among subregions were obtained when mRNAs at each stage were clustered hierarchically (Fig. S3 *B–F*). Taken together, the results suggest that the maternally derived seed coat differs fundamentally from the embryo and endosperm that both arise from fertilization events and that embryo and endosperm subregions share a complex relationship that is dependent on spatial and temporal cues.

Diverse Sets of Coexpressed Genes Underlie Seed Development. ***Subregion-specific gene sets.*** We identified mRNAs that accumulate specifically in a subregion to begin to discover the gene-expression programs that underlie the complex relationship among subregions of the seed. We defined subregion-specific mRNAs as those that accumulated at a statistically significant ($q < 0.001$, mixed-model ANOVA), fivefold or higher level in one subregion relative to all others at a given stage. The fivefold cutoff value was based on the finding that fold-change values for

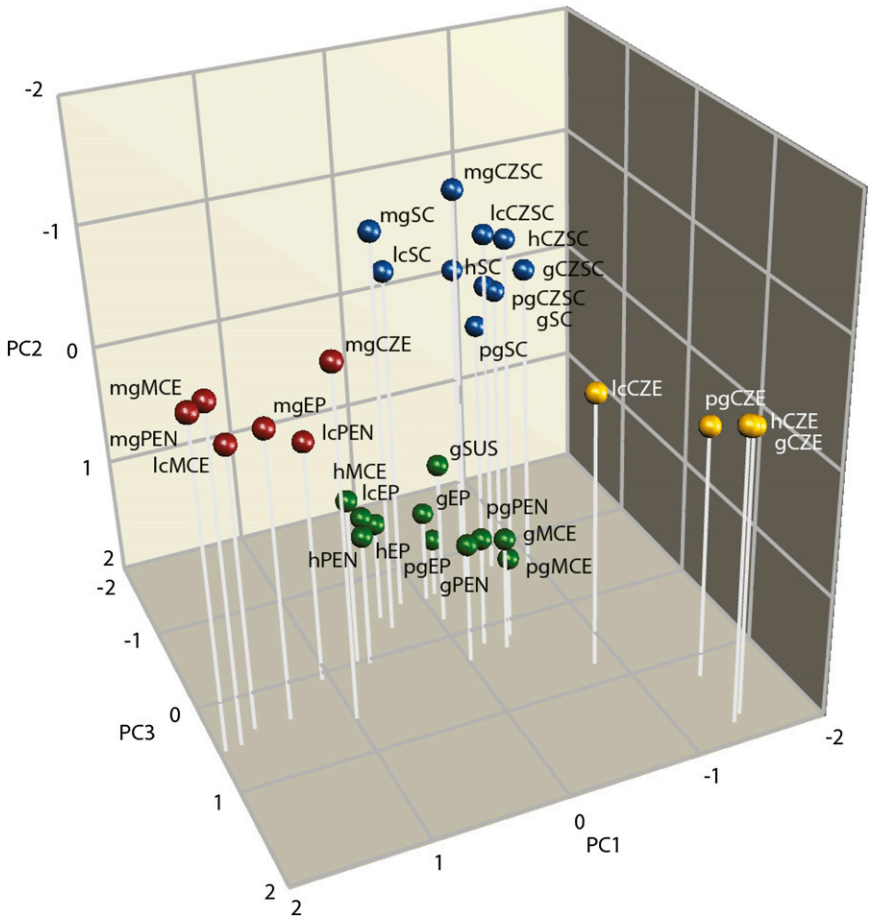

Fig. 2. PCA of seed subregion mRNA populations. PCA plot shows four distinct groups of subregion mRNA populations: subregions of the seed coat region at all stages (blue), EP, SUS, MCE, and PEN subregions at early stages (green), CZE subregions at early stages (yellow), and the EP and all endosperm subregions at the maturation phase (red). Principal components one through three collectively represent 55.6% of the variance in the dataset. Abbreviations are given in Table 1.

mRNAs significantly higher in one subregion versus all others ranged from 1.01 to 210, with a median of 4.7. Fig. 1*H* shows that between 0 (mature-green PEN) and 545 (heart CZE) subregion-specific mRNAs were identified (Dataset S3). Thus, few mRNAs accumulated specifically at the cell or tissue level relative to the total number of mRNAs in each subregion.

To determine how subregion-specific mRNAs changed over time, we clustered mRNAs from all subregions and stages to define 47 dominant expression patterns (DPs) (Fig. 3*A* and Fig. S4) (7) and assigned mRNAs to these patterns on the basis of correlation (Pearson's correlation > 0.8) (Dataset S4). Several of the coexpressed gene sets consisted of mRNAs that accumulated primarily in one subregion (Fig. 3*A* and Fig. S4) (DPs 15, 18, 20, 21, 24, 29, and 37), and an average of 70% of mRNAs in these gene sets were subregion-specific (≥fivefold enrichment, $q < 0.001$). The accumulation patterns show that some mRNAs accumulated predominately in one subregion at several stages. In contrast, each subregion, with the exception of the mature-green PEN, contained between 6 and 135 mRNAs that accumulated subregion-specifically at only one stage. Thus, some genes were expressed subregion-specifically at a single stage, whereas others were expressed specifically over several stages. These temporal variations of subregion-specific expression patterns add complexity to the gene regulatory networks that operate during seed development.

Stage-specific gene sets. Because global comparisons of mRNA populations suggested a concerted temporal shift in gene expression during development (Fig. 2), we identified mRNAs in each subregion that accumulated specifically at a single stage. We defined stage-specific mRNAs as those present at a statistically significant ($q < 0.001$, mixed-model ANOVA), fivefold or higher level at one stage relative to all others in a given subregion (Dataset S3). Fig. 1*I* shows that relatively few genes were expressed stage-specifically early during seed development. Rather, an average of 71% of the stage-specific mRNAs of each subregion accumulated at the mature-green stage. These results indicate that a major transition in gene activity occurs at the mature-green stage.

Many of the same mature-green stage-specific mRNAs accumulated in all embryo and endosperm subregions. Two coexpressed gene sets, DP 12 and DP 19 (Fig. 3*A*), consisted of mRNAs that accumulated primarily at the mature-green stage in all embryo and endosperm subregions and to a lesser extent in seed-coat subregions. An average of 58% and 66% of mRNAs in DP 12 and DP 19, respectively, were also designated as mature-green stage-specific mRNAs (≥fivefold enrichment, $q < 0.001$) in the EP, MCE, PEN, and CZE. By contrast, averages for the CZSC and SC were 16% and 14%, respectively. Taken together, these results suggest a common set of genes is coordinately up-regulated in embryo and endosperm subregions late in seed development.

Roles of Subregion-Specific Genes in Seed Development. We obtained insight into the cellular processes that characterize each subregion by identifying Gene Ontology (GO) terms and metabolic pathways that were significantly overrepresented ($P < 0.001$, hypergeometric distribution) (Dataset S3) among subregion-specific mRNAs. Fig. 4*A* lists GO terms and metabolic pathways for subregion-specific mRNAs that were overrepresented at two or more stages and/or for the DP gene set that exhibits subregion specificity. The analysis confirmed known functions for some subregions and provided a glimpse into the specific functions of other subregions whose roles in seed development were not known.

Embryo proper. EP-specific mRNAs were significantly enriched for GO terms known to be associated with patterning events that occur during embryo development, such as determination of bilateral symmetry and abaxial cell fate specification (8).

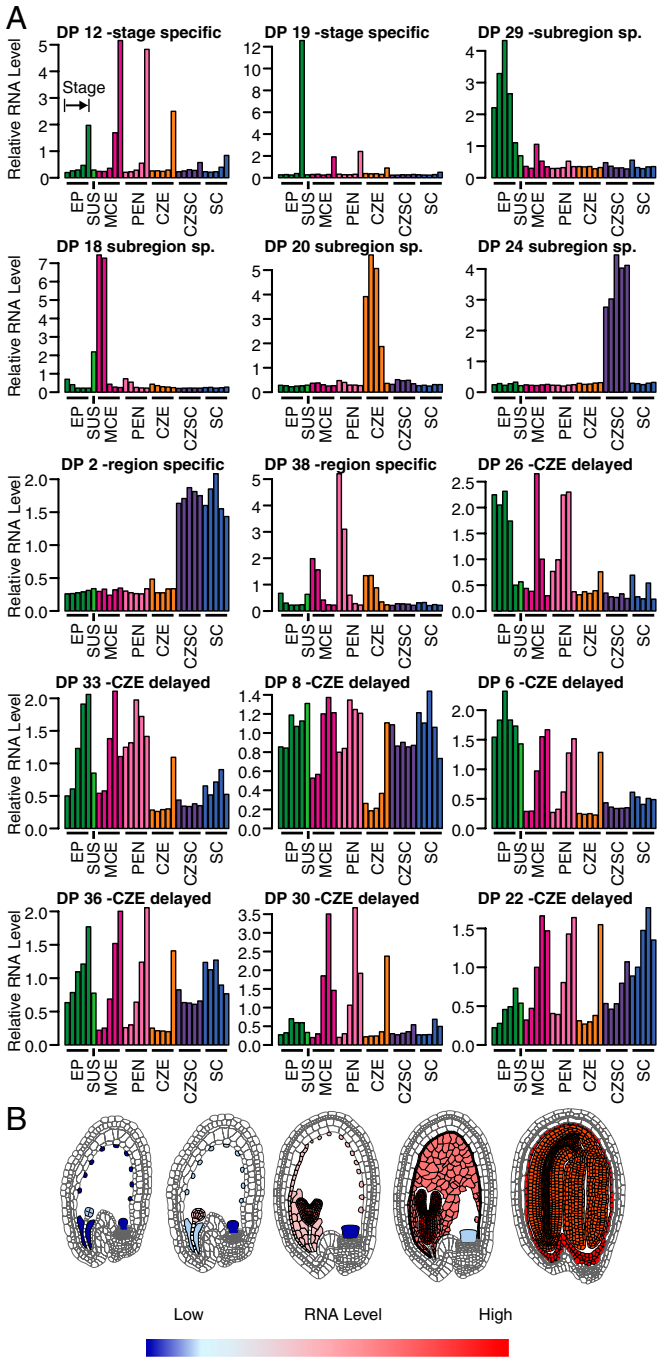

Fig. 3. Dominant patterns of gene expression during seed development. (*A*) Forty-seven DPs were identified using Fuzzy *K* means clustering of the 50% most variant mRNAs in all seed subregions and stages. Bar graphs depict median mRNA levels in each subregion (colored bars) at each stage (left to right, preglobular to mature-green stage). DPs representing the indicated stage-specific, subregion-specific, region-specific, and CZE-delayed coexpressed gene sets are shown. Remaining DPs are shown in Fig. S4, and mRNAs in all DPs are listed in Dataset S4. The average number of mRNAs in each DP gene set was 103. (*B*) Heat map of conceptualized CZE-delayed mRNA accumulation patterns in embryo and endosperm subregions. mRNA accumulation in seed coat subregions is not shown.

Seed coat. SC-specific mRNAs were overrepresented for processes associated with the synthesis of flavonoids that serve to provide protection for seeds against biotic and abiotic stresses (9).

Peripheral endosperm. At the preglobular and globular stages, PEN-specific mRNAs were overrepresented for GO terms for chloroplast compartments and metabolic pathways related to photosynthesis. These processes were not known to occur in the PEN, but their occurrence was validated functionally as discussed below.

Chalazal seed coat. mRNAs associated with trehalose and cytidine metabolism were overrepresented in the CZSC. Although trehalose plays an essential role in seed development (10), localization of key enzymes required for its synthesis to the CZSC was not known previously.

Suspensor. The SUS is an embryonic structure of 8–10 cells, and little is known about its cellular functions (11). Because SUS-specific mRNAs in DP 15 (Fig. S4) that were overrepresented for the GO term, pattern-specification process, encode efflux transporters for the hormone auxin, we analyzed a number of mRNAs involved in auxin signaling. Fig. 4 *C–E* shows that mRNAs encoding an auxin biosynthetic enzyme, YUC3, an auxin efflux carrier, PIN7, and a transcription factor responsive to auxin, ARF16, were more prevalent in the SUS than in the EP. These results are consistent with the SUS serving as a site of perception of the auxin gradient across the SUS and EP early in seed development that is essential for patterning of the embryo

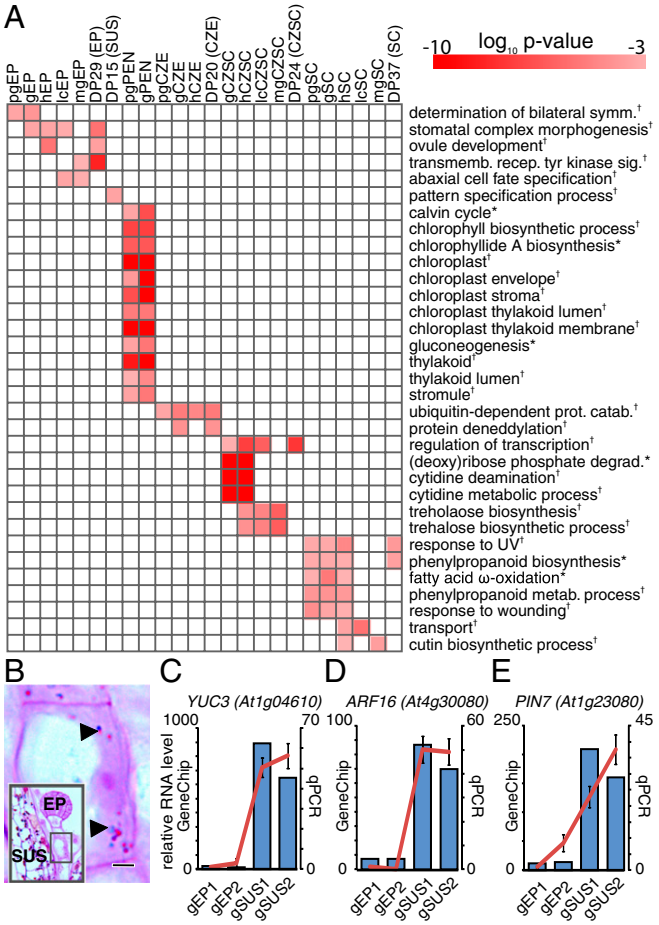

Fig. 4. Functions of subregion-specific genes. (*A*) Heat map showing the *P* value significance of enrichment of GO terms (†) or metabolic pathways (*) for subregion-specific mRNAs (Dataset S3). The listed GO terms are for biological processes or cellular components and metabolic pathways that were overrepresented at two or more stages and/or for the DP gene set that exhibit subregion specificity. (*B*) Histochemical staining of starch granules (arrowheads) in the suspensor. (Scale bar, 3 μm.) (*Inset*) The location of the enlarged area relative to the embryo proper and suspensor. (*C–E*) Relative mRNA levels determined in GeneChip experiments (bar plots) and qRT-PCR experiments (line plots) for the indicated genes involved in (*C*) auxin biosynthesis, (*D*) auxin response, and (*E*) polar auxin transport.

(12, 13). Additionally, they open the possibility that the SUS may serve as an auxin source for the gradient. We also discovered that mRNAs encoding all enzymes involved in starch biosynthesis were detected in the SUS (Dataset S1, Table S5) and demonstrated the presence of a functional pathway by showing that starch grains accumulate in SUS cells (Fig. 4*B*).

Chalazal endosperm. The CZE is a unique subregion developmentally. Early in seed development, the CZE possessed the largest number of subregion-specific mRNAs (Fig. 1*H*). Moreover, the CZE was highly enriched for mRNAs that were only detected during seed development. Seed-specific mRNAs were identified as those present in at least one subregion and stage of seed development but that were not detected in GeneChip experiments in seedlings, leaves, stems, or roots of vegetatively growing plants and flower buds or ovules of reproductively growing plants, as described previously (5). Of at least 1,316 seed-specific mRNAs (Dataset S2), the largest fraction accumulated predominately in the CZE, as shown in Fig. 5*A*. Additionally, 244 of 788 CZE-specific mRNAs accumulated seed specifically. These results are consistent with our previous report that the promoters of several seed-specific transcription factor genes are active specifically in the CZE (5). Together, these gene expression patterns support the conclusion that the CZE differs fundamentally from other subregions early in seed development, suggesting that novel processes occur in the CZE.

Given the uniqueness of the CZE, we were interested to understand its role in seed development. CZE-specific mRNAs were overrepresented for the GO term ubiquitin-dependent protein catabolism (Fig. 4*A*), suggesting a potential regulatory role for the CZE. We also showed that rate-limiting enzymes for the biosynthesis of the hormones gibberellic acid, abscisic acid, and cytokinin, accumulated primarily, but not exclusively, in the CZE (Fig. 5*B*), consistent with other reports showing that hormone metabolism genes are expressed in the CZE (14–16). Because of the importance of these hormones for seed development, these results suggest that the CZE may serve as a communication hub that integrates developmental processes within the seed.

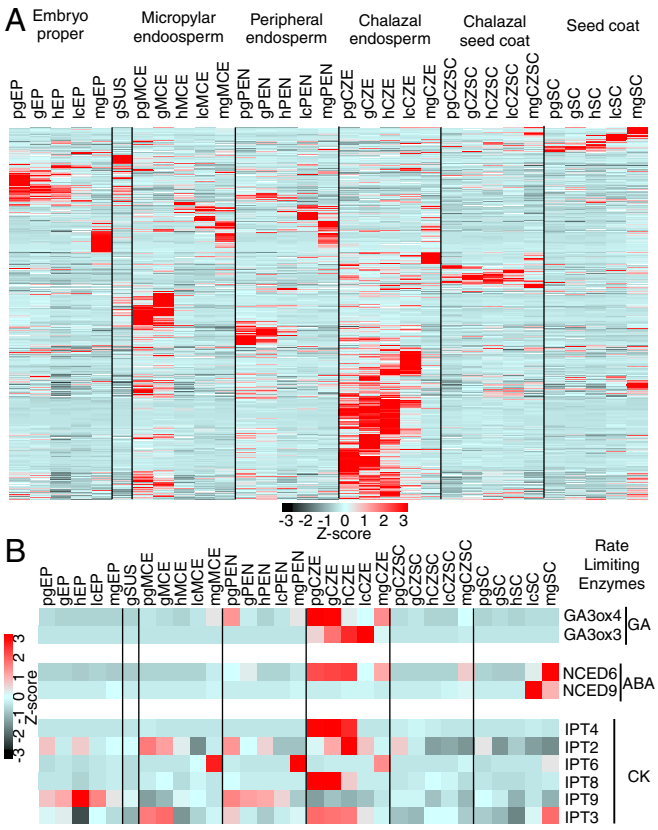

Fig. 5. CZE is a unique seed subregion developmentally. (*A*) Hierachical clustering of seed-specific mRNAs. The largest number of seed-specific mRNAs accumulate primarily in the CZE. (*B*) Heat map depicting relative levels of mRNAs encoding rate-limiting enzymes for gibberellic acid (GA; GA3ox), abscisic acid (ABA; NCED), and cytokinin (CK; IPT) biosynthesis.

Integration of Gene Activity and Cellular Function Across Subregions and Stages. ***Gene sets temporally regulated in embryo and endosperm subregions.*** Although many gene sets expressed subregion- and stage-specifically were identified, we were interested to know the extent to which gene expression was coordinated across distinct subregions and stages during seed development. We identified gene sets that were coexpressed in several subregions and stages. The most prominent coexpression pattern, representing 11 DPs (Fig. 3*A* and Fig. S4) (DP 1, 6, 7, 8, 9, 14, 22, 26, 30, 33, and 36), involved mRNAs that accumulated in the EP and all endosperm subregions but whose accumulation in the CZE was delayed relative to the other subregions (Fig. 3*B*).

Functions of CZE-delayed gene sets. The expression patterns of CZE-delayed gene sets (Fig. 3*B*) suggest that specific cellular processes that occur in all embryo and endosperm subregions are delayed in the CZE. We identified significantly enriched ($P < 0.001$, hypergeometric distribution) GO terms and metabolic pathways for the CZE-delayed gene sets and showed that several gene sets were implicated to play important roles in seed development (Dataset S4). For example, the DP 26 gene set was overrepresented for GO terms related to cytokinesis and the phragmoplast, a cytoskeletal structure specific to dividing plant cells (Figs. 3*A* and 6*E*). Cytokinesis occurs in the embryo throughout the morphogenesis phase early in development. By contrast, the endosperm initially undergoes nuclear but not cell divisions, with subsequent cellularization and cell division occurring sequentially from the micropylar to the chalazal end (4). Thus, the DP 26 coexpression pattern is coincident with the patterns of cytokinesis during seed development.

Another CZE-delayed gene set, DP 33, was significantly enriched for GO terms and metabolic pathways related to photosynthesis and carbon metabolism, including chloroplast structure and function, glycolysis, gluconeogenesis, starch biosynthesis, and fatty acid biosynthesis (Figs. 3*A* and 6*E*, and Dataset S4). Similarly, DP 8 was associated with glycolysis and gluconeogenesis. These associations were surprising, because photosynthesis and carbon metabolism are known to occur in the embryo, but much less was known about these processes in the endosperm (17). In support of the prediction that starch and lipid biosynthesis occur in the endosperm, qRT-PCR experiments showed that mRNAs encoding all enzymes required for starch and fatty acid biosynthesis were detected in endosperm subregions (Dataset S1, Table S5). The enriched GO terms were predictive of cellular function, because differentiated chloroplasts, starch grains, and lipids were detected in endosperm cells (Fig. 6 *A*–*D*). Thus, processes involved in photosynthesis and carbon metabolism that are known to characterize the embryo also occur in all three endosperm subregions, although these processes are delayed in the CZE.

Two other CZE-delayed gene sets, DP 6 and DP 36, were significantly enriched for the DNA methylation GO term (Figs. 3*A* and 6*E*, and Dataset S4). DNA methylation in plants is mediated primarily through three pathways involving METHYLTRANSFERASE1 (MET1), CHROMOMETHYLASE3 (CMT3), and the RNA-directed DNA methylation (RdDM) enzymes (18). Of the three, only mRNAs involved in RdDM exhibited a CZE-delayed accumulation pattern (Fig. 6*F*, *Upper*). DNA methylation functions primarily to silence transposon activity (19), and the endosperm is notable because transposons become hypomethylated in the central cell of the female gametophyte, the precursor of the endosperm (20). Consistent with these observations, transposon activity was high in all three endosperm subregions early in seed development but decreased late in seed development coincident with the increase

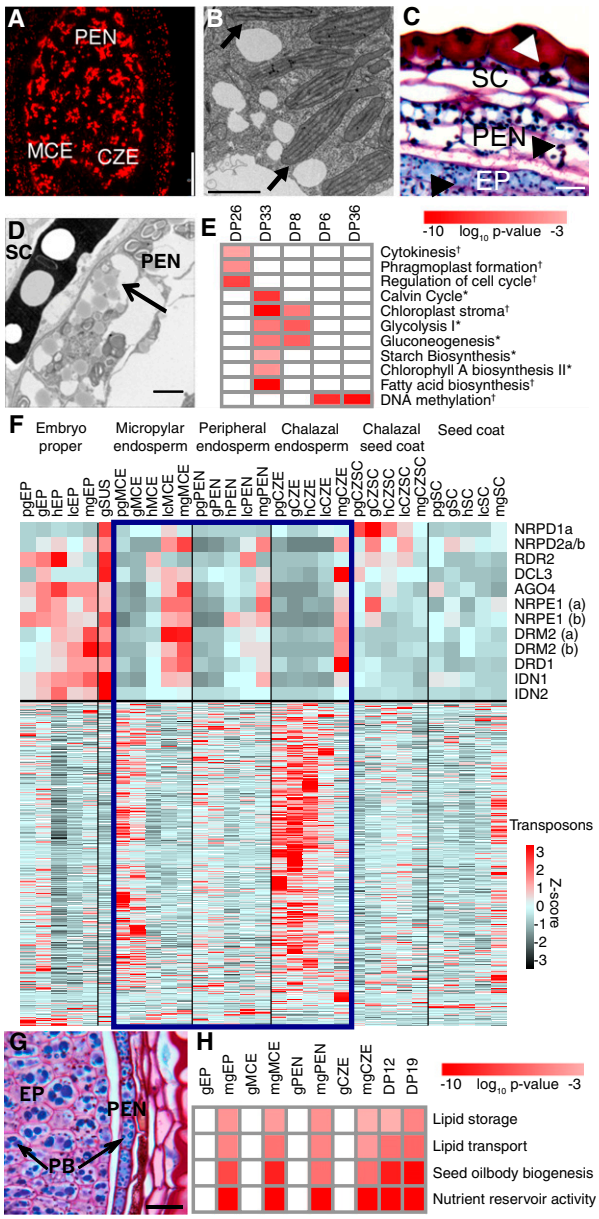

Fig. 6. Functions of CZE-delayed coexpressed gene sets. (*A*) Autofluorescent chloroplasts in the endosperm of a globular-stage seed. (*B*) Transmission electron microscopy of chloroplasts (arrows) in the PEN. (*C*) Histochemical staining of starch granules (arrowheads) in the SC, PEN, and EP. (*D*) Transmission electron micrograph of oil bodies (arrows) in cellularized PEN. (*E*) Heat map showing *P* value significance of enrichment of selected GO terms (†) or metabolic pathways (*) associated with the indicated CZE-delayed gene sets. (*F*) Heat maps showing mRNAs involved in the RNA-dependent DNA methylation pathway (*Upper*) and 1,155 probesets corresponding to transposons (*Lower*) (21). (*G*) Detection of protein bodies (PB) in the EP and PEN. (*H*) Heat map showing *P* values for GO-term enrichment of mature-green stage-specific mRNAs and the indicated DPs at the globular and mature-green stages. (Scale bars: 25 μm in *A*, 0.5 μm in *B*, 10 μm in *C*, 3 μm in *D*, and 10 μm in *G*.)

in RdDM mRNA levels (Fig. 6*F*, *Lower*) (21). The anticorrelation between RdDM mRNAs and transposon activity opens the possibility that DNA methylation is required to silence transposons late in endosperm development.

Reprogramming of Seed Development During the Maturation Phase.

Maturation occurs in embryo and endosperm subregions. The reprogramming of gene expression that occurs late in seed development appears to be associated with the onset of the maturation phase. Several gene sets comprised of mRNAs that accumulated primarily at the mature-green stage, including mature-green stage-specific mRNAs in the EP, MCE, PEN, and CZE (Fig. 1*I*) and DP 12 and 19 (Fig. 3*A*), were all significantly enriched ($P < 0.001$, hypergeometric distribution) for GO terms characteristic of seed maturation, including nutrient reservoir activity, lipid storage, and seed oil body biogenesis, among others (Fig. 6*H* and Datasets S3 and S4).

The finding that the same sets of maturation-related mRNAs accumulated in the embryo and in all three endosperm subregions was unexpected. Although the embryo undergoes maturation, and lipids are known to accumulate in the endosperm (22), the extent to which the maturation program occurs in endosperm subregions was not known. We showed that cellular structures that accumulate in maturation-phase embryos, such as storage protein bodies (Fig. 6*G*) and oil bodies (Fig. 6*D*), were detected in endosperm subregions. These results provide compelling evidence that maturation processes associated with seed filling occur in the embryo and all three endosperm subregions and involve many of the same genes.

Predicted regulatory circuitry controlling genes expressed during the maturation phase. One key to improving seeds as food is to define the gene regulatory networks that control the accumulation and composition of storage products during the maturation phase. We developed a framework to predict transcriptional modules that link transcription factors with their potential coexpressed target genes (*Materials and Methods*). The strategy associates DNA sequence motifs that are significantly enriched in the upstream regions of coexpressed genes ($P < 0.001$, hypergeometric distribution) with coexpressed transcription factors known or predicted to bind the overrepresented motifs. We showed that several transcriptional modules were predicted for seed-coat region-specific genes linking enriched MYB, HD-ZIP IV, and AG DNA sequence motifs with transcription factors known to be involved in seed coat and ovule development, such as MYB5, GL2, SHP2, STK, SHP1, and SEP (Fig. S5*A* and Dataset S1, Table S6) (23, 24). Thus, the approach identified known developmental regulators.

Fig. 7*E* and Dataset S1, Table S6 show that the DP 12 gene set, consisting of mRNAs that accumulate predominately in embryo and endosperm subregions during the maturation phase, defined a transcriptional module in which significantly enriched DNA motifs with a G-box core, including ABRE, ABRE-like, and DPBF1 and -2, were associated with bZIP transcription factors. Many of these overrepresented DNA sequence motifs are known to characterize the promoters of maturation expressed genes (25), and two of the associated transcription factors, EEL and bZIP67, play roles in regulating maturation genes (26, 27). Furthermore, transcriptional modules derived from coexpressed genes associated with a specific GO term were also identified. These submodules identify potential regulatory circuits that control processes associated with the GO term during the maturation phase. We also generated transcriptional modules for mature-green stage-specific genes expressed in the EP, MCE, PEN, and CZE and showed that there was substantial overlap in the enriched DNA motifs and associated transcription factors identified in each subregion (Fig. 7 *A*–*D*). The results suggest that maturation processes are regulated similarly but not identically in the embryo and endosperm subregions.

Analyses of other coexpressed gene sets identified transcription factors known to play critical roles in seed development. For example, genes expressed primarily in the MCE [Fig. 3*A* (DP 18), Fig. S5*D*, and Dataset S1, Table S6] were enriched for the W-box DNA sequence motif that is predicted to associate with MINISEED3, a WRKY transcription factor that is expressed primarily in the MCE and is a regulator of seed size (28). Similarly, CCA1, a transcription factor involved in controlling seed dormancy as a central circadian clock regulator, was associated with the CCA1 binding-site motif that is enriched in the promoter of genes expressed early in endosperm development [Fig. 3*A* (DP 27), Fig. S5*C*, and Dataset S1, Table S6] (29). Thus, the transcriptional

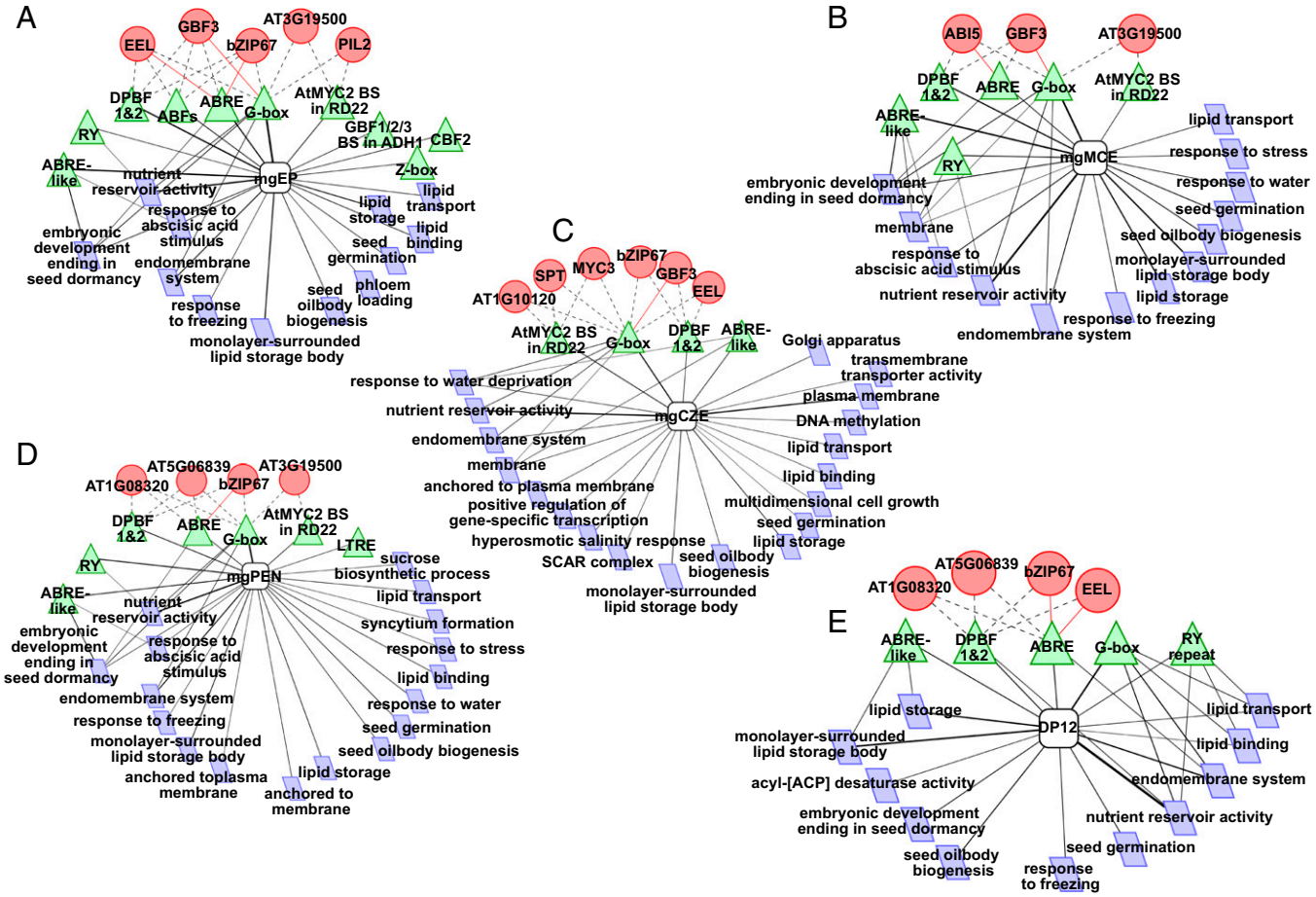

Fig. 7. Predicted transcriptional modules regulating maturation in seeds. DNA motifs (green triangles) and GO terms (blue parallelogram) that are significantly overrepresented ($P < 0.001$, hypergeometric distribution) within the coexpressed gene set (open squircle). Coexpressed transcription factors are represented as circles. Transcriptional modules were predicted for mature-green stage-specific genes in the (*A*) EP, (*B*) MCE, (*C*) CZE, and (*D*) PEN, and for (*E*) DP 12. All four mature green-stage subregions possess transcriptional modules in which bZIP transcription factors known to regulate maturation genes such as bZIP67 (AT3G44460), EEL (AT2G41070), or ABI5 (AT2G36270) are associated with overrepresented G box-like DNA motifs such as ABRE and DPBF1 and -2. Edges in red indicate known interactions between transcription factors and DNA motifs, whereas dashed lines represent predicted interactions. All enriched DNA motifs and GO terms are listed in Dataset S1, Table S6.

modules identified key regulators of seed development, suggesting their utility as predictive tools to provide insight into gene regulatory networks controlling seed development.

Discussion

We profiled RNA populations in every cell type, tissue, subregion, and region of *Arabidopsis* seeds throughout development to obtain an integrated understanding of the processes that underlie seed development. A minimum of 17,594 distinct mRNAs were detected in at least one subregion and stage, indicating that at least 60% of the *Arabidopsis* genome is expressed during seed development. The use of LCM facilitated gene discovery. Compared with our previous analyses of mRNA populations in whole-mount seeds (5), the LCM profiling experiments detected more mRNAs throughout seed development (17,594 vs. 15,577), a higher average number of mRNAs present in seeds at each stage (14,800 vs. 11,780), and a higher number of seed-specific mRNAs (1,316 vs. 289). The dataset provides the most comprehensive description of gene activity during seed development.

Coexpressed Gene Sets Inform the Cellular Processes that Underlie Seed Development. The LCM profiling experiments describe global gene activity in seed subregions that were previously inaccessible to such analyses. Identification of both region-specific mRNAs and subregion-specific mRNAs (Fig. 1*H*) suggests that subregions within the same region have both shared and distinct functions. For example, mRNAs that accumulate specifically in the seed-coat region are overrepresented for the GO terms flavonoid biosynthetic process and proanthocyanidin biosynthetic process (Dataset S3), suggesting that processes associated with responses to biotic and abiotic stresses occur in both the SC and CZSC (9). By contrast, the CZSC alone is overrepresented for subregion-specific mRNAs associated with trehalose and cytidine metabolism, but these mRNAs are not detected at substantial levels in the SC (Fig. 4*A* and Dataset S3). Thus, distinct gene sets are involved in controlling region-specific and subregion-specific functions.

Our data suggest that the functional differentiation of subregions within a region occur through at least two distinct sets of processes. First, genes expressed specifically within a subregion appear to play a significant role in specifying its function. GO terms and metabolic pathways enriched for mRNAs specifically expressed in the EP, SC, and PEN accurately predict functions that are known or that we have shown to occur in these subregions (Figs. 4 and 6). Many subregion-specific genes are active at the preglobular stage, suggesting that subregion identity is specified at the earliest stage of seed development (Figs. 1*H* and 3*A*, and Fig. S4). For example, consistent with our finding that many genes are expressed specifically in each endosperm subregion at the preglobular stage, others have shown that the

MCE, PEN, and CZE can be differentiated morphologically at the 16-nuclei stage that corresponds to the zygote-stage of seed development (30). Second, subregion function is also influenced by temporal differences in the expression of gene sets. CZE-delayed gene sets consist of mRNAs that accumulate later in the CZE than in the other embryo and endosperm subregions (Fig. 3*B*). Delayed accumulation of these mRNAs accounts, in part, for the finding that the CZE differs from the EP, SUS, MCE, and PEN early in seed development (Fig. 2). Together, these results define gene sets with diverse coexpression patterns that contribute to the overall complexity of seed development.

Gene Sets Associated with the Control of Seed Mass. Seed mass is positively correlated with seedling survival and, therefore, is a determinant of plant fitness (31). The ability to modulate seed mass has important implications for altering crop yield. Several CZE-delayed gene sets are associated with processes that control seed mass (Figs. 3*A* and 6). For example, DP 26 is overrepresented for mRNAs involved in cytokinesis. Because the timing of endosperm cellularization is correlated with seed mass (32, 33), with smaller seeds undergoing early cellularization, mechanisms that regulate this gene set are likely to be involved in controlling seed mass. Similarly, a second CZE-delayed gene set, DP 33, is enriched for mRNAs involved in photosynthesis and carbon metabolism, and photosynthetic activity in seeds is correlated with seed biomass (34).

Two other CZE-delayed gene sets, DP 6 and DP 36, which are associated with DNA methylation via the RdDM pathway, may also be related to the control of seed size (Figs. 3*A* and 6). A potential tie between the RdDM pathway and seed mass is that DNA methylation is implicated to control the expression of many imprinted genes in the endosperm, genes that are expressed specifically or preferentially from either maternal or paternal alleles (20, 35). Imprinted genes are predicted and, in one case, shown to be involved in controlling resource allocation to the embryo, a process that is critical in determining seed mass (36, 37). Moreover, imprinted genes are often flanked by transposons, and the methylation status of the transposable element is thought to determine the activity of many imprinted genes (38, 39). Consistent with the interpretation that transposons affect the activity of neighboring genes, NRPD1a, a component of the RdDM pathway, is required to silence genes encoding endosperm-specific transcription factors that are adjacent to transposons (40). Thus, accumulation of RdDM mRNAs late in endosperm development that appears to correlate with transposon silencing and, presumably, DNA methylation may also result in the silencing of imprinted genes (Fig. 6*F*). We estimate that 35 of 47 imprinted genes (39) are down-regulated in the endosperm coincident with the activation of RdDM genes. Because imprinted genes are thought to enable the endosperm to promote early embryo development, silencing of these genes late in seed development may be required to allow induction of the maturation phase in the endosperm.

Coordinated Gene Expression in the Embryo and Endosperm and Its Relevance to the Origin of the Endosperm. An overriding theme that emerges from this comprehensive developmental profile of mRNA populations is that there is extensive overlap in the gene-expression programs that characterize embryo and endosperm subregions. Although each subregion possesses mRNAs that accumulate specifically in that subregion (Fig. 1*H*), a global comparison of mRNA populations demonstrated unexpected similarities between embryo and endosperm subregions (Fig. 2). These similarities result, at least in part, from the large number of CZE-delayed genes that are coexpressed in EP, SUS, MCE, and PEN subregions early in seed development (Fig. 3 and Fig. S4). Thus, the same sets of genes that are associated with photosynthesis, carbon metabolism, cytokinesis, and DNA methylation are active in all embryo and endosperm subregions early in seed development, although their activity is delayed in the CZE. Consistent with the extensive overlap in embryo and endosperm gene activity, a large set of genes is up-regulated coordinately in the embryo proper and all endosperm subregions during the transition from the morphogenesis to the maturation phase, and many of the same putative regulators operate in the two seed regions [Fig. 3 (DP 12 and 19) and Fig. 7]. These results suggest that there is substantial coordination of the biological processes that occur in embryo and endosperm regions.

Parallels in embryo and endosperm expression programs have implications for resolving longstanding questions about the evolution of seeds. The endosperm region is unique to angiosperms, and two major hypotheses have been advanced to explain its evolutionary origin (41, 42). One hypothesis is that the endosperm is a modified supernumerary embryo resulting from a second fertilization event that acquired embryo-nourishing functions. The second hypothesis proposes that the endosperm is homologous with the gymnosperm female gametophyte, the development of which is promoted by a second fertilization event. Morphological analysis of endosperm development in basal angiosperm taxa suggests that there are shared features of early embryo and endosperm development, including unequal division and polarization of the initial cell and differential development at the micropylar and chalazal poles (43). Our results demonstrating strong overlap in embryo and endosperm gene activity are consistent with an embryo-based evolutionary origin of the endosperm, although we cannot exclude the possibility of homology between the endosperm and the female gametophyte.

Conclusions

The LCM seed dataset represents a robust resource to support studies of seed biology. We have demonstrated that the dataset can be used to identify sets of genes that are expressed in specific subregions and stages and other gene sets the expression patterns of which are integrated across multiple subregions and stages. Thus, these data define coexpressed gene sets with extremely high spatial and temporal resolution. Analysis of these coexpressed genes can accurately predict the biological function of seed subregions, providing fresh insights into the cellular processes that underlie seed development.

A key to understanding seed development is to define the regulatory circuitry that governs these diverse coexpressed gene sets. The dataset serves as a platform to identify the gene regulatory networks that operate during seed development, in part, by identifying the transcription factors that accumulate in spatially restricted locations within the seed at specific stages. We have shown that known regulators of seed development can be identified by the association of overrepresented DNA sequence motifs with coexpressed transcription factors.

The biological stories that we have presented demonstrate the utility of the dataset in uncovering new and significant information about the processes that underlie seed development. Although much remains to be learned about seed biology to obtain the basic information needed for the design of strategies to improve crops for agriculture and enhanced food security, we anticipate that this dataset will be a critical tool in enabling these discoveries.

Materials and Methods

Profiling Subregion mRNA Populations. Siliques containing seeds of *Arabidopsis thaliana* (L.) Heynh, ecotype Wassilewskija (Ws-0) were staged according to criteria described previously (5) and detailed in Dataset S1, Table S1. Details about seed collection, histological protocols, and microdissection using a Leica LMD6000 Laser Microdissection System (Leica Microsystems) are given in *SI Materials and Methods*.

Total RNA was extracted, purified from captured microdissected subregions, analyzed, and amplified as described in *SI Materials and Methods*. The number of captured subregions per biological replicate and total RNA yields are summarized in Dataset S1, Table S2. Amplifed cDNA was hybridized with the Affymetrix GeneChip ATH1 *Arabidopsis* Genome Array as previously described (5). The effects of amplification on relative RNA levels were determined using qRT-PCR experiments on cDNA amplified from 2 ng of total RNA and cDNA synthesized from 1 μg of total RNA (ThermoScript RT-PCR Systems). Dataset S1, Table S4 shows that linear cDNA amplification did not appreciably alter the representation of mRNAs in the population.

Data Analysis. GeneChip hybridization data were analyzed and detection calls for mRNAs (present, absent, or marginal) were made as previously described (5). For quantitative comparisons of mRNA levels, signal intensities from all 75 GeneChip experiments were normalized using RMA (44). Correlation between RMA normalized biological replicates averaged 0.96 (Pearson's correlation, Dataset S2) and ranged between 0.93 and 0.98, which was higher than that obtained using other normalization methods. Relative RNA levels were validated with qRT-PCR experiments as previously described (45). DNA sequences and efficiencies of primer pairs used for qRT-PCR experiments and comparison of relative mRNA levels determined in GeneChip and qRT-PCR experiments are given in Dataset S1, Table S3.

Mixed-model linear ANOVA, used to assess the significance of mRNA level comparisons in different samples, and identification of dominant expression patterns was done as previously described (7). Other analyses, including hierarchical clustering and bootstrapping analysis and PCA, are described in *SI Materials and Methods*.

Identification of Transcriptional Modules. The software package, ChipEnrich (46), was modified to identify significantly enriched DNA sequence motifs upstream of coexpressed genes that were associated with transcription factors known or predicted to bind the motifs as described in *SI Materials and Methods*. Files used to generate the transcriptional modules are in Dataset S1, Table S6.

Microscopy. Procedures for light and fluorescence microscopy, transmission electron microscopy, and confocal laser scanning microscopy are described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Samantha Duong, Maichi Phan, Emilia Madejska, Xiaohua Lu, Alexander Olson, and Chen Cheng for technical assistance, and Bob Fischer for comments about the manuscript. This work was supported by grants from the National Science Foundation Plant Genome Program (to R.B.G. and J.J.H.) and a postdoctoral fellowship from the Natural Sciences and Engineering Research Council (to M.F.B.).

1. Steeves TA (1983) The evolution and biological significance of seeds. *Can J Bot* 61(12): 3550–3560.
2. Godfray HCJ, et al. (2010) Food security: The challenge of feeding 9 billion people. *Science* 327(5967):812–818.
3. Ohto M, Stone SL, Harada JJ (2007) Genetic control of seed development and seed mass. *Seed Development, Dormancy, and Germination*, eds Bradford KJ, Nonogaki H (Blackwell, Oxford), pp 1–24.
4. Brown RC, Lemmon BE, Nguyen H, Olsen O-A (1999) Development of endosperm in *Arabidopsis thaliana*. *Sex Plant Reprod* 12(1):32–42.
5. Le BH, et al. (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci USA* 107(18):8063–8070.
6. Harada JJ, Pelletier J (2012) Genome-wide analyses of gene activity during seed development. *Seed Sci Res* 22(Suppl S1):S15–S22.
7. Brady SM, et al. (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318(5851):801–806.
8. Capron A, Chatfield S, Provart N, Berleth T (2009) Embryogenesis: Pattern formation from a single cell. *Arabidopsis Book* 7:e0126.
9. Pourcel L, Routaboul J-M, Cheynier V, Lepiniec L, Debeaujon I (2007) Flavonoid oxidation in plants: From biochemical properties to physiological functions. *Trends Plant Sci* 12(1):29–36.
10. Schluepmann H, Paul M (2009) Trehalose metabolites in *Arabidopsis*-elusive, active and central. *Arabidopsis Book* 7:e0122.
11. Kawashima T, Goldberg RB (2010) The suspensor: Not just suspending the embryo. *Trends Plant Sci* 15(1):23–30.
12. Wang JW, et al. (2005) Control of root cap formation by MicroRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* 17(8):2204–2216.
13. Weijers D, et al. (2006) Auxin triggers transient local signaling for cell specification in *Arabidopsis* embryogenesis. *Dev Cell* 10(2):265–270.
14. Hu JH, et al. (2008) Potential sites of bioactive gibberellin production during reproductive growth in *Arabidopsis*. *Plant Cell* 20(2):320–336.
15. Lefebvre V, et al. (2006) Functional analysis of *Arabidopsis* NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *Plant J* 45(3):309–319.
16. Miyawaki K, Matsumoto-Kitano M, Kakimoto T (2004) Expression of cytokinin biosynthetic isopentenyltransferase genes in *Arabidopsis*: Tissue specificity and regulation by auxin, cytokinin, and nitrate. *Plant J* 37(1):128–138.
17. Xiang DQ, et al. (2011) Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. *Plant Physiol* 156(1): 346–356.
18. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11(3):204–220.
19. Lisch D (2009) Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* 60(1):43–66.
20. Köhler C, Wolff P, Spillane C (2012) Epigenetic mechanisms underlying genomic imprinting in plants. *Annu Rev Plant Biol* 63(1):331–352.
21. Slotkin RK, et al. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136(3):461–472.
22. Penfield S, et al. (2004) Reserve mobilization in the *Arabidopsis* endosperm fuels hypocotyl elongation in the dark, is independent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1. *Plant Cell* 16(10):2705–2718.
23. Colombo L, Battaglia R, Kater MM (2008) *Arabidopsis* ovule development and its evolutionary conservation. *Trends Plant Sci* 13(8):444–450.
24. Li SF, et al. (2009) The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell* 21(1): 72–89.
25. Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C (2007) Combined networks regulating seed maturation. *Trends Plant Sci* 12(7):294–300.
26. Bensmihen S, et al. (2002) The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. *Plant Cell* 14(6):1391–1403.
27. Yamamoto A, et al. (2009) *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J* 58(5): 843–856.
28. Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005) MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. *Proc Natl Acad Sci USA* 102(48): 17531–17536.
29. Penfield S, Hall A (2009) A role for multiple circadian clock genes in the response to signals that break seed dormancy in *Arabidopsis*. *Plant Cell* 21(6):1722–1732.
30. Brown RC, Lemmon BE, Nguyen H (2003) Events during the first four rounds of mitosis establish three developmental domains in the syncytial endosperm of *Arabidopsis thaliana*. *Protoplasma* 222(3–4):167–174.
31. Westoby M, Jurado E, Leishman M (1992) Comparative evolutionary ecology of seed size. *Trends Ecol Evol* 7(11):368–372.
32. Garcia D, et al. (2003) *Arabidopsis* haiku mutants reveal new controls of seed size by endosperm. *Plant Physiol* 131(4):1661–1670.
33. Scott RJ, Spielman M, Bailey J, Dickinson HG (1998) Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125(17):3329–3341.
34. Goffman FD, Alonso AP, Schwender J, Shachar-Hill Y, Ohlrogge JB (2005) Light enables a very high efficiency of carbon storage in developing embryos of rapeseed. *Plant Physiol* 138(4):2269–2279.
35. Raissig MT, Baroux C, Grossniklaus U (2011) Regulation and flexibility of genomic imprinting during seed development. *Plant Cell* 23(1):16–26.
36. Costa LM, et al. (2012) Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Curr Biol* 22(2):160–165.
37. Haig D, Westoby M (1991) Genomic imprinting in the endosperm: Its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philos T R Soc B* 333(1266): 1–13.
38. Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 324(5933): 1447–1451.
39. Hsieh TF, et al. (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324(5933):1451–1454.
40. Lu J, Zhang C, Baulcombe DC, Chen ZJ (2012) Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proc Natl Acad Sci USA* 109(14):5529–5534.
41. Baroux C, Spillane C, Grossniklaus U (2002) Evolutionary origins of the endosperm in flowering plants. *Genome Biol* 3(9):reviews1026.
42. Friedman WE (2001) Developmental and evolutionary hypotheses for the origin of double fertilization and endosperm. *C R Acad Sci III* 324(6):559–567.
43. Floyd Sandra K, Friedman William E (2000) Evolution of endosperm developmental patterns among basal flowering plants. *Int J Plant Sci* 161(S6):S57–S81.
44. Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4(2):249–264.
45. Yamagishi K, et al. (2005) TANMEI/EMB2757 encodes a WD repeat protein required for embryo development in *Arabidopsis*. *Plant Physiol* 139(1):163–173.
46. Orlando DA, Brady SM, Koch JD, Dinneny JR, Benfey PN (2009) Manipulating large-scale *Arabidopsis* microarray expression data: identifying dominant expression patterns and biological process enrichment. *Methods Mol Biol* 553:57–77.

Supporting Information

Belmonte et al. 10.1073/pnas.1222061110

SI Materials and Methods

Plant Materials and Growth. Wild-type *Arabidopsis* were grown in a peat-based medium (Sunshine Mix #1) at 22 °C in 50–70% relative humidity under constant light (100–150 $\mu E/m^2 \cdot s^{-1}$). Plant material was harvested between 2:00 PM and 5:00 PM for consistency.

Laser-Capture Microdissection. Whole siliques or seeds dissected from siliques were collected and immediately fixed in 3:1 95% (vol/vol) ethanol:acetic acid at 4 °C under RNase-free conditions (1). Whole siliques were cut into ~3-mm segments before fixation to promote fixative penetration. The material was then vacuum-infiltrated for 30 min and fixed overnight at 4 °C. The plant material was rinsed three times with 70% (vol/vol) ethanol, dehydrated in a graded ethanol series (70%, 85%, 95%, 100%, 100% ethanol), and infiltrated with xylenes (1:3, 1:1, 3:1 xylenes: ethanol, followed by 100% xylenes twice). Samples were incubated with paraffin chips overnight at room temperature, at 42 °C for at least 1 h, and at 60 °C for 30 min. The xylenes-paraffin mixture was then replaced with 100% paraffin, and samples were infiltrated with changes in paraffin for 2–3 d at 60 °C.

Seeds or siliques were sectioned at either a 5- or 7-μm thickness using a Leica RM2125RT rotary microtome (Leica Microsystems) and mounted on RNase-free polyethylene napthalate (PEN)-membrane slides (Leica Microsystems). Use of 5- to 7-μm sections minimized contamination from adjacent and underlying cell types during microdissection. Slides were dried at room temperature and deparaffinized twice in 100% xylenes for 1 min.

Each seed subregion was microdissected independently to minimize contamination from adjacent cell and tissue types (Fig. S1 *F*–*Q*). Subregions were captured through sequential serial sections to obtain an accurate representation of all mRNAs within a subregion. For example, serial sections encompassing the entire embryo proper of mature green-stage seeds were captured. The one exception is that the embryo proper (EP) and suspensor (SUS) of globular-stage seeds were microdissected from medial sections to avoid endosperm contamination. We also captured serial sections of entire whole seeds at each developmental stage. Two biological replicates were captured for each subregion at each developmental stage, with the exception of the globular-stage chalazal endosperm (CZE), and heart-stage CZE and chalazal seed coat (CZSC) for which three biological replicates were obtained. Each biological replicate consisted of captured subregions from at least 10 seeds. All subregions were captured within 1 mo of fixation to maximize RNA quality.

Affymetrix GeneChip Hybridization Experiments. Microdissected subregions were harvested directly into RNA extraction buffer and total RNA was extracted (RNAqueous-Micro; Ambion). Following treatment of the samples on the RNA purification column with RNase-free DNase (1:4 dilution of DNase I in RDD buffer; Qiagen), RNA levels were quantified (Quant-iT RiboGreen RNA Assay Kit; Invitrogen) using a ND-3330 Fluorospectrometer (Nano-Drop). The numbers of captured sections per bioreplicate ranged from 53 to 2,167, and total RNA yields were between 5 and 66 ng depending on the seed subregion, as detailed in Dataset S1, Table S2. Total RNA was analyzed by microcapillary electrophoresis (RNA 6000 Pico Chip, Agilent 2100 BioAnalyzer; Agilent Technologies), using whole-mount silique RNA as a control.

Two to 6 ng of total RNA was converted to cDNA using a linear amplification method (WT-Ovation Pico RNA Amplification System; NuGEN Technologies), and the cDNA was fragmented and labeled (FL-Ovation cDNA Biotin Module V2; NuGEN Technologies). Five micrograms of amplifed cDNA was hybridized with the Affymetrix GeneChip ATH1 *Arabidopsis* Genome Array as described by Le et al. (2).

To determine the number of distinct mRNAs in a seed subregion, we normalized GeneChip hybridization data and assigned present, absent, and marginal signal detection calls using MAS 5.0 software (Affymetrix) (3). mRNAs were designated as detected in a population if their signal detection calls were present in either both biological replicates or in at least two of three of the replicates. Signal detection calls and relative levels for mRNAs in all GeneChip experiments are given in Dataset S2.

Global Comparisons of mRNA Populations. Hierarchical clustering and bootstrapping analysis was conducted with RMA-normalized data for all GeneChip biological replicates using the pvclust package with default settings (http://cran.r-project.org/web/packages/pvclust/pvclust.pdf) (4). Principal component analysis was carried out on RMA-normalized, replicate averaged data using the prcomp function in R (5).

Identification of Coexpressed Gene Sets. ***Subregion-specific and region-specific mRNAs.*** A mRNA specific to a seed subregion was defined as one whose relative level is at least fivefold higher and significantly different ($q < 0.001$, mixed-model ANOVA) than those detected in all other subregions at a given developmental stage. Lists of subregion-specific mRNAs and their overrepresentation for Gene Ontology (GO) terms, DNA motifs, and metabolic pathways are given in Dataset S3.

Seed regions are defined as the embryo, consisting of the EP and SUS, the endosperm, consisting of the micropylar (MCE), peripheral (PEN), and CZE, and the seed coat, consisting of the SC and CZSC. Region-specific mRNAs are present at a fivefold or higher level in all subregions of a region and significantly different ($q < 0.001$, mixed-model ANOVA) than in all other seed subregions. Embryo region-specific mRNAs were only identified at the globular stage because the SUS subregion was not analyzed at other developmental stages. Lists of region-specific mRNAs and their enrichment for GO terms, DNA motifs, and metabolic pathways are given in Dataset S3.

Stage-specific mRNAs. A mRNA specific to a developmental stage is one whose level is fivefold higher and significantly different ($q < 0.001$, mixed-model ANOVA) at one developmental stage relative to all other stages in a given seed subregion. Stage-specific mRNAs and their enriched GO terms, DNA motifs, and metabolic pathways are listed in Dataset S3.

Dominant expression pattern identification. Dominant expression patterns (DPs) were identified essentially as previously described (6). RMA-normalized and averaged data from all seed subregions and developmental stages, but not from whole seeds, were filtered to remove probesets that did not exceed a minimum expression cutoff (RMA value of 15, >75% of RNAs designated present by MAS 5.0 analysis exceeded this cutoff value) in at least one seed subregion and developmental stage. To identify DPs, the sample variance of the remaining probesets were ranked, and the 50% most variant, corresponding to 8,047 RNAs, were retained for clustering analysis. Filtered data were clustered using the FKM implementation FANNY (http://cran.r-project.org/web/packages/cluster/cluster.pdf) (7) with a K of 50 and a probability for cluster membership (m-value = 0.44) that resulted in most probesets being assigned to a single cluster.

Attempts to use smaller K values eliminated clusters with significant patterns, whereas use of larger K values did not generate additional clusters with novel RNA accumulation patterns. The median RNA accumulation pattern for all mRNAs in a cluster was determined, and the 50 patterns were subjected to hierarchical clustering using $1 - r$ (r = Pearson's correlation coefficient) as the distance metric. Members of distinct clusters that shared significant similarity (separated by a node with a tree height of less than 0.15 in the dendrogram) were combined, and a new median RNA accumulation pattern was determined and reanalyzed using hierarchical clustering. These procedures generated 47 different median RNA accumulation patterns. mRNAs were then reassigned to clusters based on correlation with the 47 median accumulation patterns. mRNAs whose levels were above the minimum expression cutoff, variance were among the top 75%, and accumulation pattern correlated strongly (Pearson's correlation > 0.8) with a median RNA accumulation pattern were assigned to that cluster. Coexpressed gene sets corresponding to DPs contained an average of 104 mRNAs, ranging between 3 and 508. Lists of mRNAs and their enrichment for GO terms, DNA motifs, and metabolic pathways are given in Dataset S4.

Seed-Specific mRNAs. Seed-specific mRNAs were those that are called as: (*i*) present in all or a majority of replicates for at least one seed subregion; and (*ii*) absent in all replicates of reproductive (ovules and floral buds) and vegetative organs (leaf, stem, root, and seedling) (2).

Quantitative RT-PCR Experiments. Results of GeneChip hybridization experiments were validated using quantitative RT-PCR (qRT-PCR) experiments. PCR amplification reactions and data analysis was done as described previously (8), except that 100 pg of amplifed cDNA derived from microdissected seed subregions were used and data were normalized to *PP2AA3* levels, (At1g13320) (9). PP2AA3 RNA levels were relatively constant in all seed subregions and developmental stages. Primer pairs for amplification of specific mRNAs were designed using Beacon Designer 3 (Premier Biosoft International). The DNA sequences and efficiencies of primer pairs used to validate GeneChip data and the corresponding relative mRNA levels are provided in Dataset S1, Table S3.

ChipEnrich. We modified the ChipEnrich software program (10) to identify GO terms, metabolic pathways, transcription-factor families, and DNA sequence motifs that are overrepresented in coexpressed gene sets and to discover potential transcriptional modules. This Java program was developed originally to identify significantly enriched GO terms (2009 download) and transcription factor families from gene lists. Significance of enrichment is reported as P values calculated from the hypergeometric distribution (11) using the Apache Commons Math library (http://jakarta.apache.org/commons/math). The following functions were added to ChipEnrich which is available at http://seedgenenetwork.net/presentation#software.

Metabolic pathway enrichment analysis. Genes represented on the ATH1 GeneChip were annotated according to metabolic pathways described in the PATHWAYS database from AraCyc (2008 download). Enrichment was defined as the ratio of: (*i*) the number of AGI locus identifiers in the query list annotated as belonging to a pathway to (*ii*) the number of AGI locus identifiers associated with the pathway on the GeneChip compared with the ratio of (*iii*) the total number of AGI locus identifiers present in the query list to (*iv*) the total number of AGI locus identifiers present on the GeneChip.

DNA motif enrichment analysis. Gene sets were analyzed to identify enriched DNA sequence motifs known to interact with transcription factors (*Arabidopsis* Gene Regulation Information Server, http://arabidopsis.med.ohio-state.edu/, August 2009) that are located in the region 1-kb upstream of the gene's transcription start site (TAIR9, www.arabidopsis.org) as described by others (12, 13). The background distribution was determined by identifying DNA motifs for all genes represented as singletons on the *Arabidopsis* ATH1 GeneChip (see Dataset S1, Table S6 for a list of all DNA motifs used in this study). Statistical enrichment (P value < 0.001) was determined for each gene list using the hypergeometric distribution. Enriched DNA sequence motifs were also identified among genes overrepresented for a GO term within a gene list.

Putative transcriptional modules. To discover putative transcriptional modules, we associated significantly enriched DNA sequence motifs with transcription factors known or predicted to bind the motifs. We used known interactions between transcription factors and DNA motifs specified in AtcisDB (14) and defined by others in the literature and assumed that transcription factors of a particular family bind to the same DNA motif (6). Two variations of this approach were used. In the first approach, we associated DNA motifs significantly enriched within a coexpressed gene set with their cognate transcription factors that were included in the coexpressed gene set. In the second variation, we identified DNA motifs that were significantly enriched for genes corresponding to an overrepresented GO term and associated coexpressed transcription factors known or predicted to bind the enriched DNA motifs. Overrepresented GO terms, DNA motifs, and their associated transcription factors were compiled into two Cytoscape compatible files that were used as network and node attribute files, and the modules were visualized with Cytoscape. All files, including the network and node attribute files, used to generate the transcriptional modules, are found in Dataset S1, Table S6.

Outputs are summarized in a text file, <significant.txt>, in which the gene set name is in the first column, enriched GO terms, DNA motifs, or transcription factor families are listed in the second column, and P values indicating the significance of enrichment are given in the third column. Each unique enriched category is set in a new row. If a DNA motif is significantly overrepresented within a gene list (P < 0.001), it is also determined if the motif is enriched among genes significantly overrepresented for a GO term (P < 0.001). In the <significant.txt> file, the overrepresented GO terms are listed in the first column, enriched DNA motifs are in the second column, and P values are in the third column. Transcription factos in the gene list (first column) that are predicted or known to bind with enriched DNA motifs (second column) are also listed in the <significant.txt> file. A separate node attribute file is also provided from ChipEnrich that describes whether a node (first column of <significant.txt> file) is a pattern, GO term, DNA motif, transcription factor family, or transcription factor.

The <significant.txt> file is designed to be used as the network file for the network graphing software, Cytoscape (version 2.6.3, www.cytoscape.org). The <node.txt> file is used as the attributes file (15). P values are also imported with the network file as edge attributes. For visualization purposes, a thicker line represents a lower P value, a dashed line represents a transcription factor with a predicted binding interaction, and a solid red edge is an experimentally determined transcription factor–DNA motif interaction. All files, including the network and attributes files, used to generate the transcriptional modules are found in Dataset S1, Table S6.

Analysis of GFP Activity. Activities of selected promoters were evaluated using promoter-GFP chimeric genes generated previously (16). Transgenic seeds with GFP reporter genes were analyzed using a Zeiss Axioskop 2 plus compound microscope equipped with a FS10 FITC filter set (exciter PB 450–490; Carl Zeiss) (16). An exposure time of 500 ms was used for all images.

Seeds were analyzed at all stages of development, although images shown in Fig. S2 are primarily of globular-stage seeds. Seeds from at least four plants were examined for each promoter-GFP line.

Histology. For light microscopy, samples were fixed in 2.5% (vol/vol) glutaraldehyde and 1.6% (wt/vol) paraformaldehyde buffered with 0.05 M phosphate buffer, pH 6.9, dehydrated with methyl cellosolve followed by two changes of absolute ethanol, and then infiltrated and embedded in Historesin (Leica) according to the methods of Yeung (17). Three micrometer-thick serial sections were stained with periodic acid-Schiff for total carbohydrates and counterstained with amido black 10B for protein or with Toluidine blue O for general observations.

Chloroplasts were localized by auto fluorescence using 633-nm excitation and a 650-nm filter using a Zeiss-700 confocal scanning laser microscope. Seeds from at least three different plants were used for observations. LSM Zen imaging software (Carl Zeiss) was used to construct 3D images from 40 optical slices in the *z*-dimension (*z*-stack). No other image enhancement was performed.

Oil bodies and chloroplast ultrastructure were visualized using transmission electron microscopy. Seeds were fixed in 2% (wt/vol) paraformaldehyde, 2.5% (vol/vol) glutaraldehyde and 0.1 M sodium phosphate buffer, pH 7.2 (Karnovsky's fixative), using a PELCO Biowave microwave (Ted Pella) under vacuum (20 psi) as follows: 5 min at 000 W, 10 s at 200 W, 20 s at 155 W, 10 s at 250 W. Tissue was stored in Karnovsky's fixative until it was processed further. Tissue was rinsed with 0.05 M sodium phosphate buffer, pH 6.9, and postfixed with 1% (vol/vol) osmium tetroxide in 0.05 M sodium phosphate buffer, pH 6.9 for 2 h followed by microwave fixation under vacuum for 40 s at 250 W. The tissue was then incubated in 0.1% aqueous tannic acid for 30 min and rinsed and stained with 2% (wt/vol) aqueous uranyl acetate. Dehydration was accomplished by immersing the tissue for 20 min each in three changes of 95% (vol/vol) acetone followed by two changes of 100% acetone. The tissue was infiltrated in 3:1 and 2:1 acetone:Spurr's resin for at least 1 h each, 1:1 acetone:resin overnight, 1:2 acetone:resin for 24 h, and in two changes of pure resin for 24 h each. Samples were embedded in capsules and polymerized overnight at 70 °C. Thick sections were cut on a Leica Ultracut UCT Ultramicrotome (Leica Microsystem) at 400 nm and stained with Methylene blue and Azure B. Thin sections were cut using a Diatome Diamond Knife (Diatome) at 60–90 nm, transferred to copper grids and double-stained with 4% (wt/vol) alcoholic uranyl acetate and lead citrate. The sections were viewed on a Philips CM120 Biotwin Lens (FEI) and images were captured using a Gatan BioScan camera.

1. Kerk NM, Ceserani T, Tausta SL, Sussex IM, Nelson TM (2003) Laser capture microdissection of cells from plant tissues. *Plant Physiol* 132(1):27–35.
2. Le BH, et al. (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci USA* 107(18):8063–8070.
3. Hubbell E, Liu WM, Mei R (2002) Robust estimators for expression analysis. *Bioinformatics* 18(12):1585–1592.
4. Suzuki R, Shimodaira H (2006) Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12):1540–1542.
5. Team RDC (2012) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
6. Brady SM, et al. (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318(5851):801–806.
7. Kaufman L, Rousseeuw PJ (1990) *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, New York).
8. Yamagishi K, et al. (2005) TANMEI/EMB2757 encodes a WD repeat protein required for embryo development in *Arabidopsis*. *Plant Physiol* 139(1):163–173.
9. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol* 139(1):5–17.
10. Orlando DA, Brady SM, Koch JD, Dinneny JR, Benfey PN (2009) Manipulating large-scale *Arabidopsis* microarray expression data: Identifying dominant expression patterns and biological process enrichment. *Methods Mol Biol* 553:57–77.
11. Gadbury GL, Garrett KA, Allison DB (2009) Challenges and approaches to statistical design and inference in high-dimensional investigations. *Methods Mol Biol* 553:181–206.
12. O'Connor TR, Dyreson C, Wyrick JJ (2005) Athena: A resource for rapid visualization and systematic analysis of *Arabidopsis* promoter sequences. *Bioinformatics* 21(24): 4411–4413.
13. Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, Van de Peer Y (2009) Unraveling transcriptional control in *Arabidopsis* using *cis*-regulatory elements and coexpression networks. *Plant Physiol* 150(2):535–546.
14. Davuluri RV, et al. (2003) AGRIS: *Arabidopsis* gene regulatory information server, an information resource of *Arabidopsis cis*-regulatory elements and transcription factors. *BMC Bioinformatics* 4:25.
15. Cline MS, et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2(10):2366–2382.
16. Steffen JG, Kang IH, Macfarlane J, Drews GN (2007) Identification of genes expressed in the *Arabidopsis* female gametophyte. *Plant J* 51(2):281–292.
17. Yeung EC (1999) The use of histology in the study of plant tissue culture systems: Some practical comments. *In Vitro Cell Dev Biol Plant* 35(2):137–143.

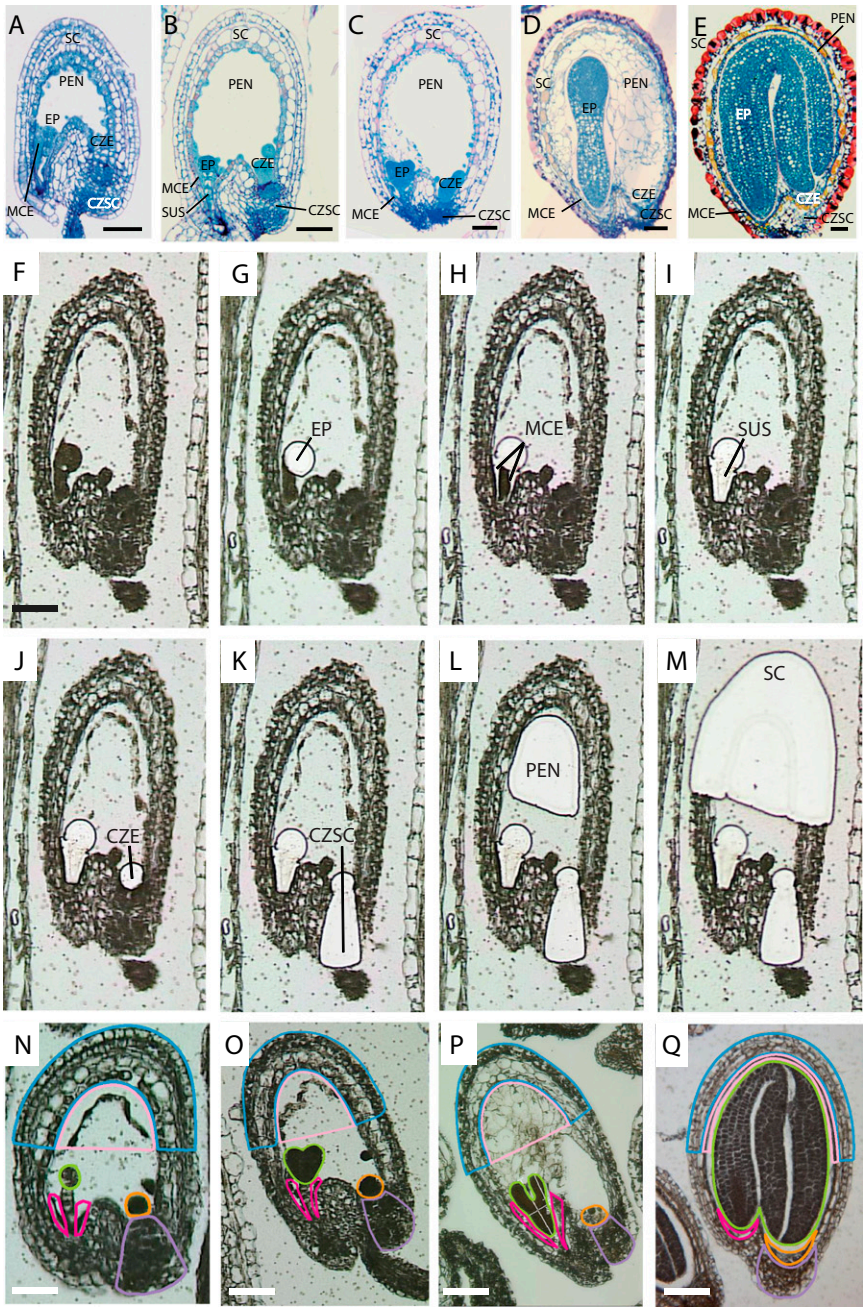

Fig. S1. Microdissection of seed subregions during *Arabidopsis* seed development. (*A–E*) Longitudinal sections of developing *Arabidopsis* seeds across five stages of development: (*A*) preglobular stage, (*B*) globular stage, (*C*) heart stage, (*D*) linear cotyledon stage, and (*E*) mature-green stage. (*F–M*) Order of subregion microdissection from a globular-stage seed. Medial longitudinal sections through the embryo proper and suspensor of a globular-stage embryo are shown. (*N–Q*) Outline of subregions captured at the (*N*) preglobular, (*O*) heart, (*P*) linear cotyledon, and (*Q*) mature-green stages of seed development. Subregions outlined are EP (green), MCE (dark pink), PEN (light pink), CZE (orange), CZSC (purple), and SC (blue). Abbreviations are given in Table 1. (Scale bars: *A–E*, 50 μm; *F–M*, 45 μm; *N*, 40 μm; *O*, 60 μm; *P*, 85 μm; *Q*, 120 μm.)

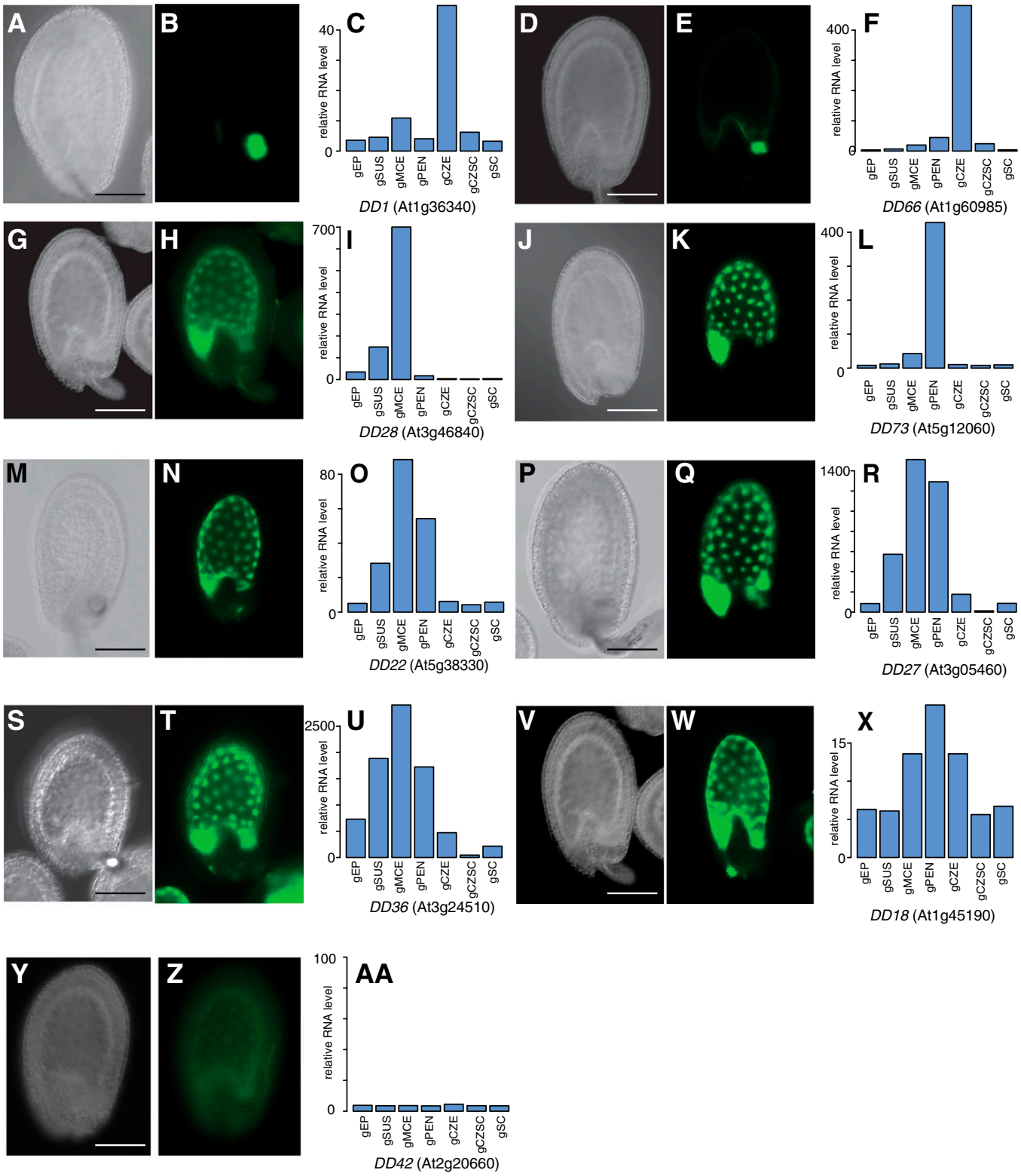

Fig. S2. Comparison of relative RNA levels and promoter activities of endosperm-expressed genes. (*B*, *E*, *H*, *K*, *N*, *Q*, *T*, *W*, *Z*) GFP activity of (*A*, *D*, *G*, *J*, *M*, *P*, *S*, *V*, *Y*) globular-stage seeds containing reporter genes fused with the indicated promoters. (*C*, *F*, *I*, *L*, *O*, *R*, *U*, *X*, *AA*) Relative RNA levels as determined by GeneChip hybridization experiments. (*A–C*) *DD1* (At1g36340) and (*D–F*) *DD66* (At1g60985) are active specifically in the CZE. (*G–I*) *DD28* (At3g46840) is active primarily in the MCE and SUS. (*J–L*) *DD73* (At5g12060), (*M–O*) *DD22* (At5g38330), and (*P–Q*) *DD27* (At3g05460) are active primarily in the MCE and PEN. (*S–U*) *DD36* (At3g24510) and (*V–X*) *DD18* (At1g45190) are active in all three endosperm subregions. (*Y–AA*) Promoter activity and mRNA accumulation for the negative control gene, *DD42* (At2g20660). Exposure times for fluorescence micrographs were 500 ms. (Scale bar, 100 μm.)

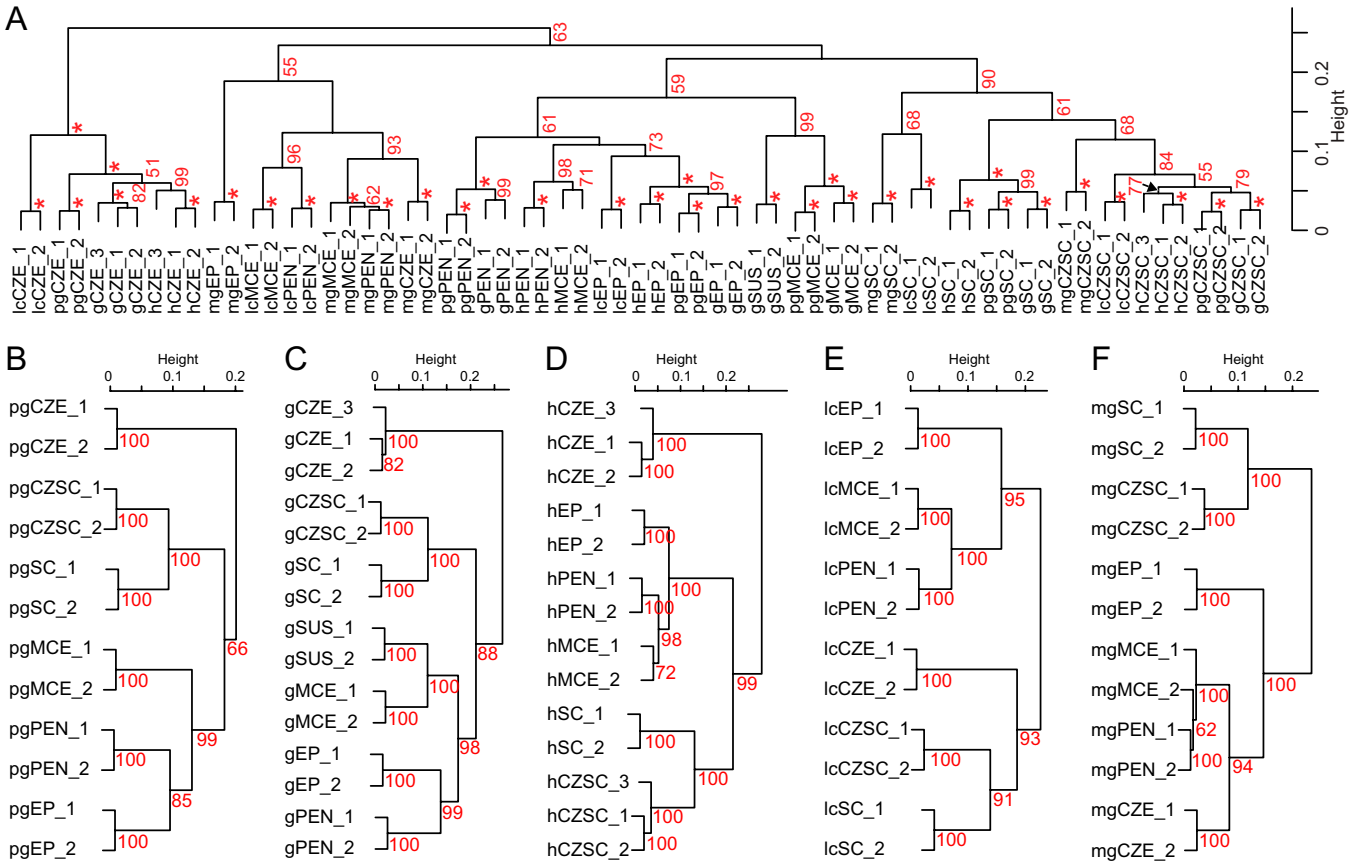

Fig. S3. Hierarchichal clustering of seed subregion mRNA populations. (*A*) Correlation-based hierarchical clustering of all biological replicates of mRNAs populations in seed subregions. (*B–F*) Hierarchical clustering of mRNA populations at the (*B*) preglobular, (*C*) globular, (*D*) heart, (*E*) linear-cotyledon, and (*F*) mature-green stages. Bootstrap values are shown in red. Asterisks indicate a bootstrap value of 100. Numbers in the sample name indicate biological replicates.

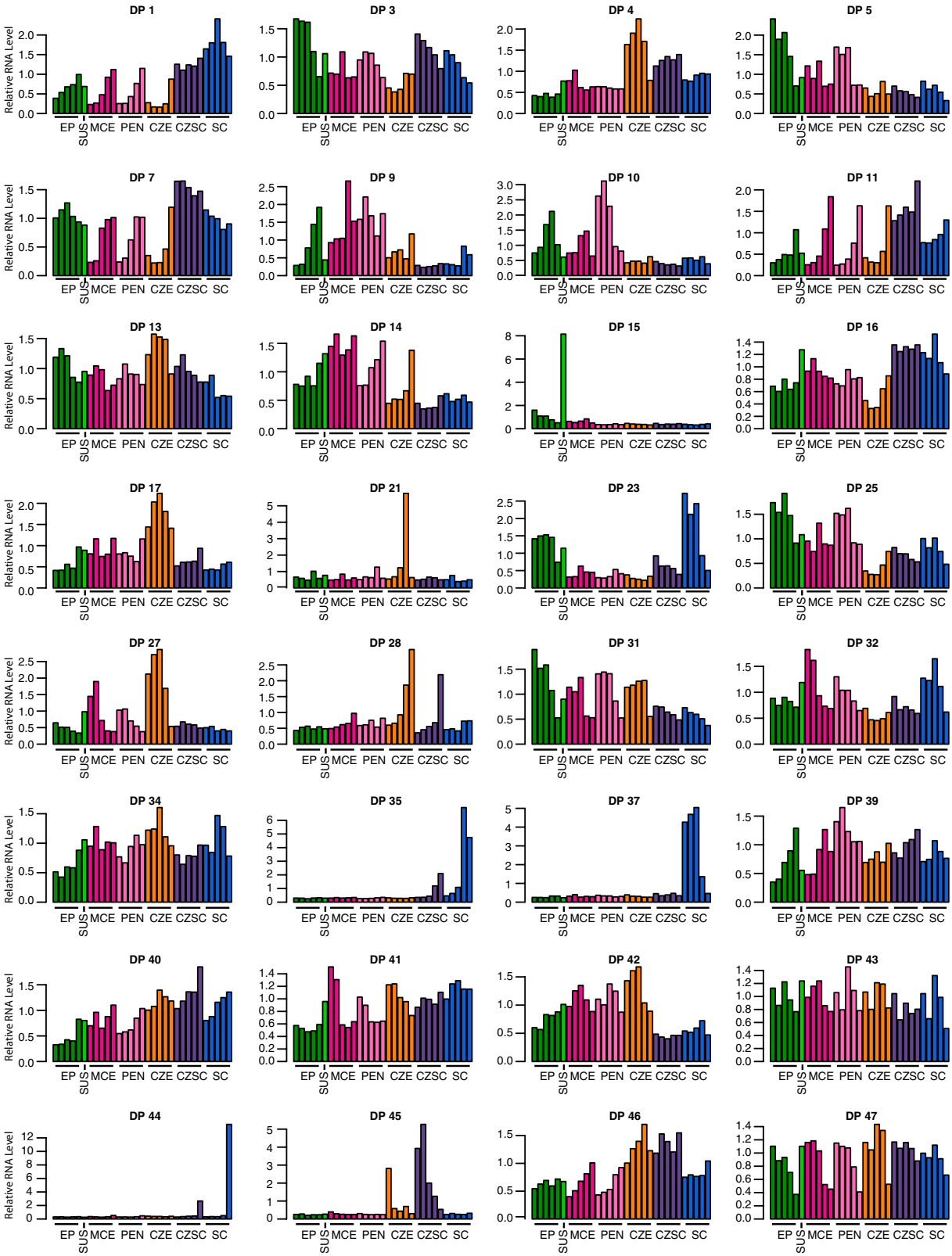

Fig. S4. DPs of gene expression during seed development. DPs that are not shown in Fig. 3. These patterns were defined with Fuzzy *K* means clustering of the 50% most variant mRNAs in all combinations of seed subregions and stages.

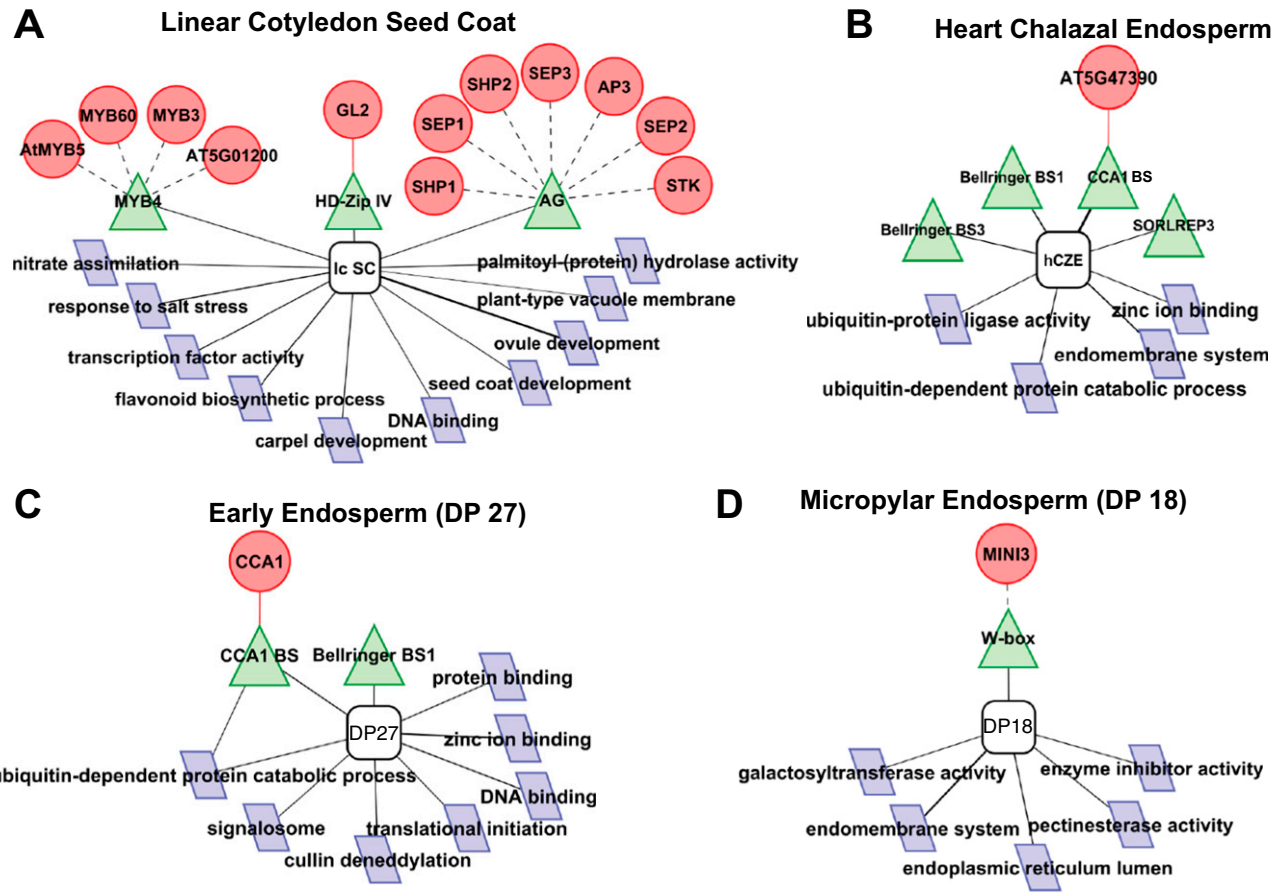

Fig. S5. Predicted transcriptional modules of coexpressed gene sets. Squircles represent sets of coexpressed genes, parallelograms and triangles depict significantly enriched ($P < 0.001$, hypergeometric distribution) GO terms and DNA motifs, respectively, and circles correspond to coexpressed transcription factors predicted or known to interact with DNA motifs. Abbreviated GO terms are given in Dataset S1, Table S6. (*A*) MYB, HD-ZIP, and MADS transcriptional modules based on mRNAs that accumulate specifically in the seed-coat region at the linear-cotyledon stage. (*B*) A CCA1-like transcription factor is associated with the CCA1 binding site among genes expressed specifically in the heart CZE. (*C*) Genes expressed in all endosperm subregions early in development (DP 27) are the basis for a transcriptional module with the CCA1 transcription factor associated with the CCA1 DNA motif. We previously reported overrepresentation of the CCA1 DNA motif for transcription factor mRNAs that accumulate specifically in the CZE (2). (*D*) A MINISEED1 module linking the transcription factor with the enriched W-box DNA motif in a MCE-specific gene set (DP 18).

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLSX)
Dataset S3 (XLSX)
Dataset S4 (XLSX)