Supporting Information

Li et al. 10.1073/pnas.1406383111

SI Materials and Methods

Plant Materials and Growth. Plants of the reference maize genotype, B73, were grown under greenhouse conditions (16-h day) at the University of Arizona during August–September 2009. Kernel and endosperm samples were collected from self-pollinated ovules. Self-pollinations were performed essentially as described previously (1). The 0-, 2-, 3-, 4-, and 6-days after pollination (DAP) kernels were collected by pushing down the glumes and pinching off the kernel at its base using a pair of forceps. The 8-, 10-, and 12-DAP endosperm samples were harvested by removing all tissues of the kernel from the pedicel up to the hilar region (2) and cutting open the pericarp along the edge of the hilar region to expose and extract the endosperm free of the embryo and the pericarp using a surgical blade. In all cases, the staged kernels and endosperm samples were harvested beginning at the same time during the day, frozen immediately in liquid nitrogen, and stored at −70 °C before RNA extractions were carried out. Of the three individual biological replicates of each developmental stage collected for this work, one [biological replicate 1 (Bio-rep 1), except for 0 DAP where two technical replicates were sequenced] was used for RNA-Seq, and all three were used for quantitative RT-PCR (qRT-PCR) analyses.

Defining Maize Early Kernel Developmental Stages. The 0- to 12-DAP kernels of B73 were dissected using a vibratome (Vibratome Series 1000, Technical Products International Inc.). Sagittal sections (200–300 μm) of developing kernels were stained using 0.1% Toluidine Blue O (TBO; Sigma-Aldrich) or iodine–potassium iodide (IKI) solutions as described previously (3, 4) and observed using light microscopy.

RNA Extraction and mRNA Purification. Total RNA for RNA-Seq and qRT-PCR analyses was isolated using a SDS-phenol method essentially as described previously (5), except that 50-mL Phase-Lock-Gel Light Tubes (5Prime, Inc.) were used to facilitate the phenol/chloroform extractions. Genomic DNA was removed using TURBO DNase I (Ambion), and RNA was further purified using RNeasy columns (Qiagen). For RNA-Seq analysis, two rounds of poly(A)+ mRNA selection were carried out on 500 μg of total RNA using the Poly(A)Purist MAG kit (Ambion). RNA was quantified using a Nanodrop 1000 Spectrophotometer (Thermo Scientific, Inc.), and RNA quality was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.) and found to be 9.3–10 in RNA integrity number.

Library Construction and SOLiD Sequencing. Standard protocols for SOLiD 4 System sequencing (Applied Biosystems, Inc.) were used for construction of the maize mRNA libraries. About 100 ng of input poly(A)+ mRNA was used to create the libraries with the SOLiD Whole Transcriptome Kit. Samples were sequenced in two batches: a 6-DAP kernel as a single sample in one lane and the rest (including the two technical replicates of 0-DAP RNA preparations) as pooled samples in a separate lane. The latter were pool-balanced via Bioanalyzer and qRT-PCR using the SOLiD TaqMan expression kit. Emulsion PCR was followed by Work Flow Analysis to verify bead quality, noise to signal, and quantity. SOLiD 4 System sequencing was performed according to the manufacturer's protocol (Applied Biosystems SOLiD 4 System Library Preparation Guide) at the Waksman Genomics Core Facility, Waksman Institute of Microbiology (Piscataway, NJ). An examination of the sequences suggested contamination of the pooled samples by RNA from the 10- or 12-DAP samples; the 0- to 4-DAP samples contained a basal level of mRNAs that are highly prevalent in the 10- or 12-DAP samples (e.g., α- and β-zeins) (Dataset S4, *Table S4*). qRT-PCR analysis indicated these sequences are not present in RNA from 0 to 4 DAP (Dataset S4, *Table S5*) and, therefore, that the contamination occurred in the process of library construction and not during RNA extraction and purification. These sequences were removed from 0- to 4-DAP data in the analysis of up- and down-regulated genes.

Bioinformatic Analyses. FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) was run initially to examine the quality of RNA-Seq reads. About 5.9% of total reads per sample that were deemed low quality were removed. RNA-Seq reads were aligned to the maize B73 reference genome using the TopHat pipeline (6) with the built-in Bowtie mapping program. The compiled gene set (39,656 genes in the filtered gene set and 8,701 expressed working gene set genes with combined read counts from the eight samples higher than 100 reads) included 2,605 maize TFs classified into 55 families, based on data available at GRASSIUS (7). To normalize the RNA-Seq data across the eight samples, we used the scaling normalization method provided in the edgeR package, based on a trimmed mean of M-values algorithm to compute the scaling factors according to the library size of each sample. After edgeR normalization, the two replicates of 0 DAP data were averaged and used in the analysis. We used StepMiner (8) to identify the genes showing up- or down-regulation patterns at specific developmental stages based on time-course expression data. The StepMiner analysis was performed on the $\log_2$ transformed read counts of the 46,274 genes (≥20 reads in at least one sample). Identification of gene ontology (GO) categories significantly (P value ≤0.05) enriched with up- and down-regulated genes was done using the agriGO program (9).

Validation of mRNA-Seq Data Using Quantitative RT-PCR. Three biological replicates (Bio-reps 1–3), including Bio-rep 1 used for RNA-Seq analysis (*Plant Materials and Growth*), were tested using quantitative RT-PCR. In no case did we detect any major variation in mRNA levels obtained from the biological replicates of individual stages indicating that our original sampling for mRNA-Seq was developmentally reliable. In addition, three biological replicates of endosperm isolated from 4- and 6-DAP kernels, and 6-DAP kernels from which endosperm had been removed, were also tested. Reverse transcription and qRT-PCR were performed as previously described (10). The gene-specific primer pair sequences of 13 transcription factor (TF) genes are listed in Dataset S4, *Table S6*, with amplicon sizes ranging between 150 and 200 bp. Three candidate housekeeping genes (Dataset S4, *Table S7*) were tested using qRT-PCR for variation during maize kernel and endosperm development; thioredoxin (*ZmTXN*, gene ID: GRMZM2G066612) was found to display the least level of variation and, therefore, was used as an internal reference gene. The amount of cDNA template for a specific TF gene was considered negligible when the Ct value was at or above 36 (10). Therefore, the normalized Ct values were manually cut off at 36. Hierarchical clustering analysis of qRT-PCR results was performed using a Cluster 3.0 (11) for detecting the dynamics of accumulation of transcription factors during kernel and endosperm development. The dendrogram tree was displayed by Java TreeView (12).

Determining a Threshold for Identifying Expressed Genes in RNA-Seq Data. To determine an empirical threshold of read count to define a gene as expressed, we performed qRT-PCR analysis of 13 candidate transcription-factor genes representing a range of RNA reads (0–17,523 reads) and found that mRNAs for genes with more than 10 normalized RNA reads were consistently and reproducibly detected in our biological material (Dataset S1 and Dataset S4, *Table S1*). Conversely, we found that mRNAs with fewer than 10 normalized reads did not produce reproducible qRT-PCR results. Therefore, we used 20 normalized reads as a cutoff to determine the number of expressed genes in all eight developmental samples.

In situ Hybridization. Staged kernels (as described above) were obtained from B73 plants grown at the University of Utah and fixed as previously described (13), except that whole kernels were fixed overnight in 4% (wt/vol) paraformaldehyde in 0.1 M phosphate buffer. The Paraplast-embedded kernels were then sectioned at 8–10 μm, and the tissue sections were placed onto glass slides. The in situ hybridization probes were generated from clones of B73 genomic DNA by PCR amplification using the primers listed in Dataset S4, *Table S8*. The resulting PCR products were cloned into pCRII-TOPO (Invitrogen). To generate the single-stranded RNA probes, the clones were first linearized using the appropriate restriction enzyme, and RNA was synthesized using either SP6 or T7 polymerase following the instructions in the Roche RNA Labeling Kit (Roche Applied Science). Before hybridization, kernel sections were soaked in xylene to remove the paraffin, hydrated with an ethanol series (100, 95, 85, 70, 50, 30, 15, and 0%, vol/vol), incubated with 1 μg/mL proteinase K for 30 min at 37 °C, dehydrated with an ethanol series (15, 30, 50, 70, 85, 95, and 100%, vol/vol), and then allowed to air dry. Hybridization solution, applied directly to the dried sections, contained the following: 1 μg/mL probe, 50% (vol/vol) formamide, 10 mM Tris (pH 7.5), 1 mM EDTA, 300 mM NaCl, 1% blocking reagent, 10% (wt/vol) dextran sulfate, and 500 μg/mL tRNA. The sections were covered with a coverslip, placed in a humidified box, and incubated at 50 °C overnight. Following hybridization, the sections were soaked in 2× SSC to remove the coverslips, incubated in 0.2× SSC at 55 °C for 30 min, incubated with 10 μg/mL ribonuclease A at 37 °C for 30 min, and then incubated in 0.2× SSC at 55 °C for 30 min. The sections were then incubated in 1% blocking reagent at room temperature for 30 min, incubated with the anti-digoxigenin antibody (diluted 1/1,500) in 1% blocking reagent at room temperature for 60 min, and incubated in 1% blocking reagent at room temperature for 30 min. The sections were maintained overnight at room temperature in 100 mM Tris (pH 9.5), 100 mM NaCl, 1.875 mg/mL nitro blue tetrazolium chloride, and 0.94 mg/mL 5-bromo-4-chloro-3-indolyl-phosphate. Finally, the sections were rinsed in Tris-EDTA buffer, mounted using Clear-Mount, and observed using differential interference contrast optics.

1. Poehlman JM (1977) , *Breeding Field Crops* (Avi Publishing Co., Westport, CT), 427 pp.
2. Felker FC, Shannon JC (1980) Movement of C-labeled assimilates into kernels of Zea mays L: III. An anatomical examination and microautoradiographic study of assimilate transfer. *Plant Physiol* 65(5):864–870.
3. Lux A, Morita S, Abe J, Ito K (2005) An improved method for clearing and staining free-hand sections and whole-mount samples. *Ann Bot* 96(6):989–996.
4. Yeung EC (1998) A beginner's guide to the study of plant structure. *Tested Studies for Laboratory Teaching*, Vol. 19. *Proceedings of the 19th Workshop/Conference of the Association for Biology Laboratory Education (ABLE)*, ed Karcher SJ (Purdue University, Lafayette, IN), pp 125–141.
5. Shirzadegan M, Christie P, Seemann JR (1991) An efficient method for isolation of RNA from tissue cultured plant cells. *Nucleic Acids Res* 19(21):6055.
6. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.
7. Yilmaz A, et al. (2009) GRASSIUS: A platform for comparative regulatory genomics across the grasses. *Plant Physiol* 149(1):171–180.
8. Sahoo D, Dill DL, Tibshirani R, Plevritis SK (2007) Extracting binary signals from microarray time-course data. *Nucleic Acids Res* 35(11):3705–3712.
9. Du Z, Zhou X, Ling Y, Zhang ZH, Su Z (2010) agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38(Web Server issue):W64–W70.
10. Wang D, et al. (2010) Identification of transcription-factor genes expressed in the Arabidopsis female gametophyte. *BMC Plant Biol* 10:110.
11. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95(25):14863–14868.
12. Saldanha AJ (2004) Java Treeview: Extensible visualization of microarray data. *Bioinformatics* 20(17):3246–3248.
13. Drews GN (1998) In situ hybridization. *Methods Mol Biol* 82:353–371.

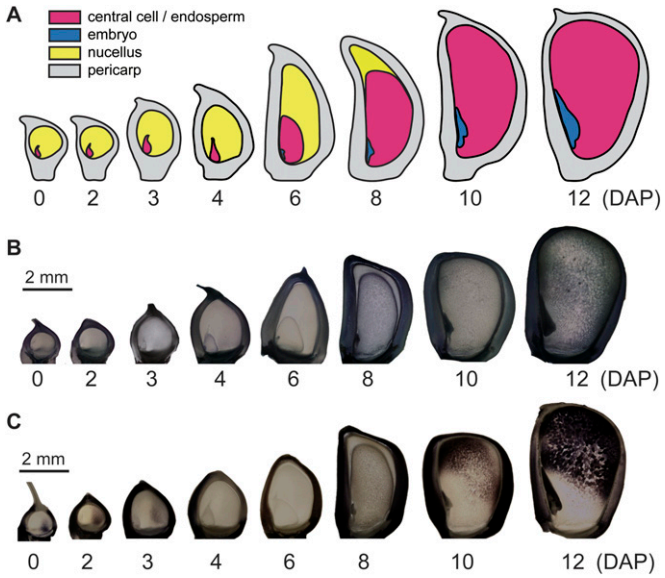

Fig. S1. Developmental stages of maize kernel and endosperm material used in this study. (*A*) Graphic representation of early development of maize kernel. The figure shows the extent of development for the central cell and endosperm (red), the egg cell and embryo (blue), and the nucellus (yellow) from immediately before fertilization (0 DAP) to 12 DAP in the inbred maize B73. (*B*) Sections of early kernel stained with TBO to show the contrast of different tissue structures. (*C*) Sections of early kernel stained with IKI to show starch accumulation.

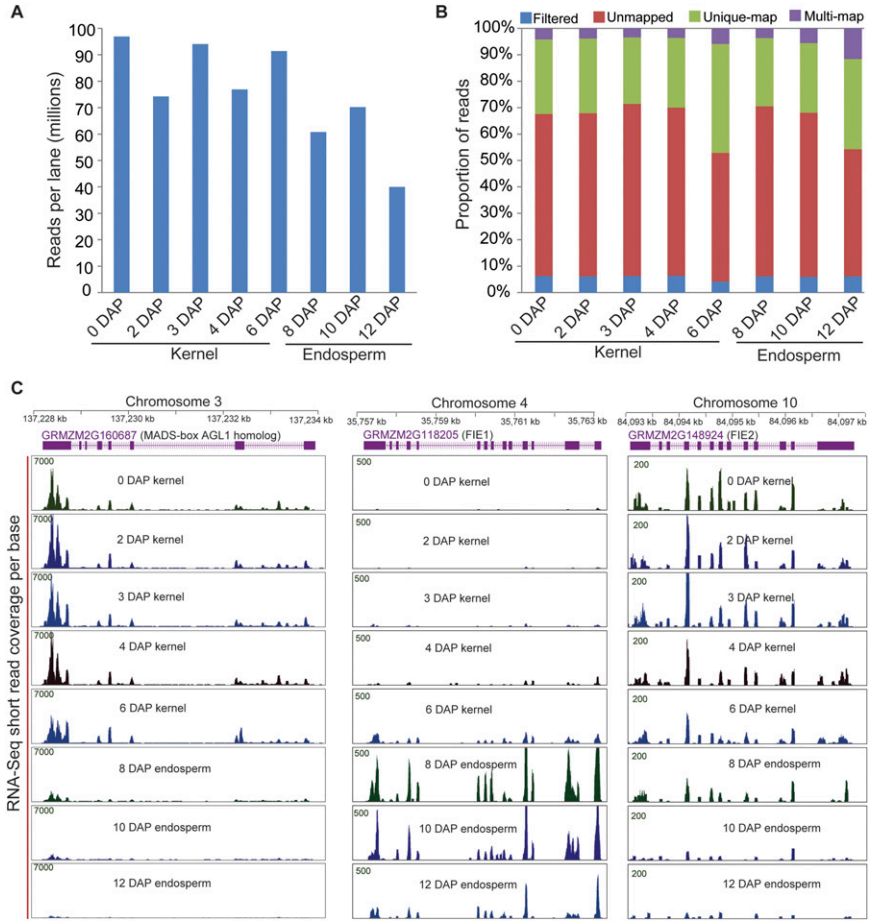

Fig. S2. Mapping and visualization of RNA-Seq reads. (*A*) Counts of raw reads produced from the SOLiD4 sequencing platform in each sample. (*B*) Mapping statistics of the reads aligned against the B73 reference genome. (*C*) Visualization in the University of California at Santa Cruz genome browser of RNA-Seq signals for three genes expressed in the kernel and endosperm based on previously published data. GRMZM2G160687 is a MADS-box TF homologous to the *Arabidopsis AGL1*, GRMZM2G118205 is *fie1*, and GRMZM2G148924 is *fie2*. As previously described (1), GRMZM2G160687 showed higher mRNA reads in 0- to 6-DAP kernel stages, whereas GRMZM2G118205 mRNA was primarily restricted to later-stage endosperm samples. This contrasts with the read pattern for GRMZM2G148924, which is expressed before fertilization in the embryo sac and predominantly in the embryo after fertilization, with much lower levels in the endosperm (1).

1. Danilevskaya ON, et al. (2003) Duplicated fie genes in maize: Expression pattern and imprinting suggest distinct functions. *Plant Cell* 15(2):425–438.

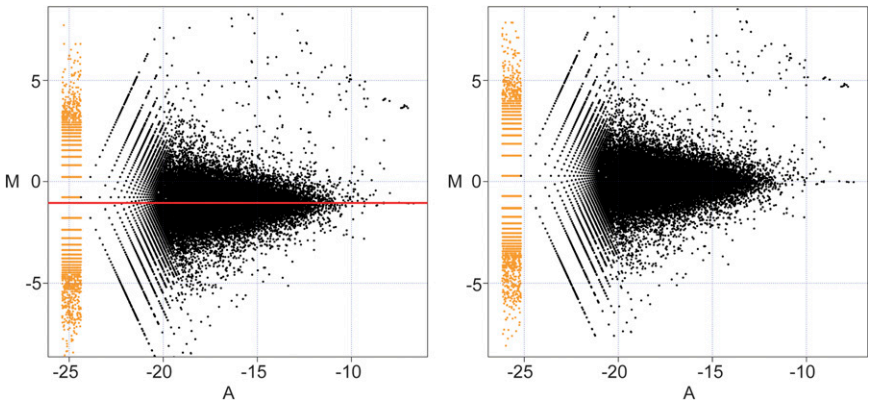

Fig. S3. Scaling normalization of 8-DAP endosperm sample by edgeR. The graphs show "MA" plots (log ratio versus abundance) before normalization (*Left*) and after normalization (*Right*). M (*y* axis) = $\log_2$ (8 DAP) − $\log_2$ (0 DAP). A (*x* axis) = ($\log_2$ (8 DAP) + $\log_2$ (0 DAP))/2.

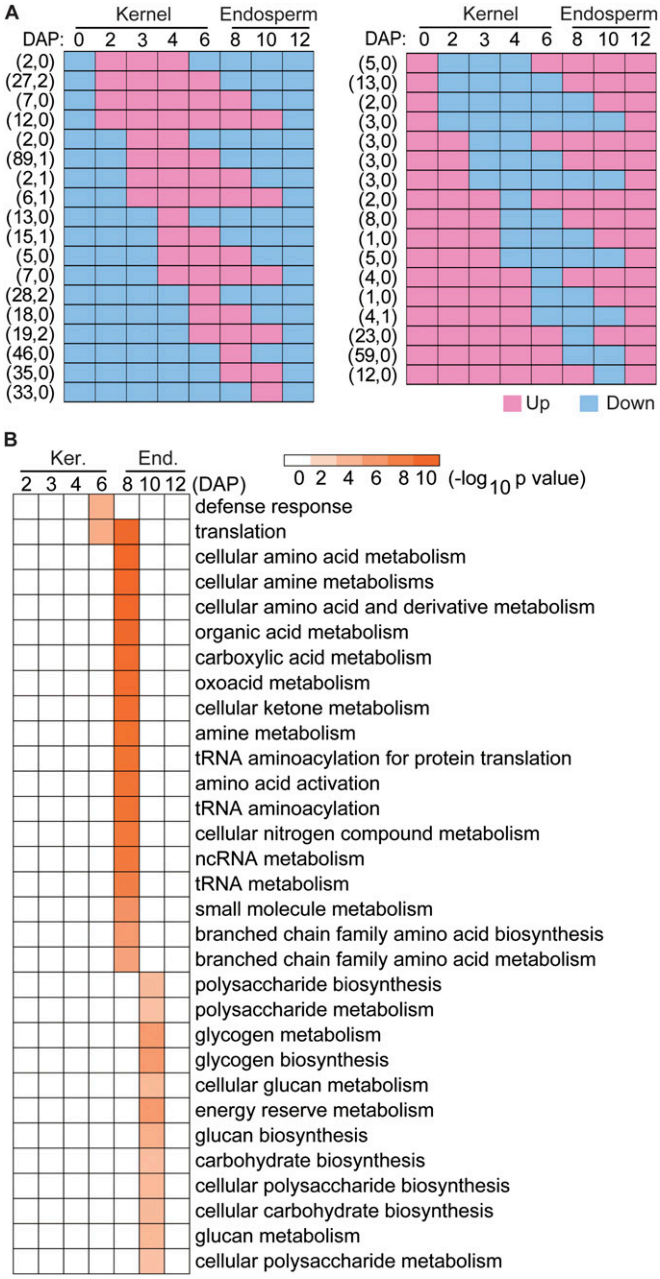

Fig. S4. Genes showing two-step patterns of expression and GO-term analysis for one-step-up genes. (*A*) StepMiner analysis showing mRNA level transitions from low to high and back down (two-step-up-down, *Left*) or from high to low and back up (two-step-down-up, *Right*) in a series of developmental stages. The total number of all genes (*Left*) and the coexpressed transcription-factor genes (*Right*) are indicated in parentheses for each expression pattern. (*B*) GO categories enriched for one-step up-regulated genes. End., endosperm; Ker., kernel.

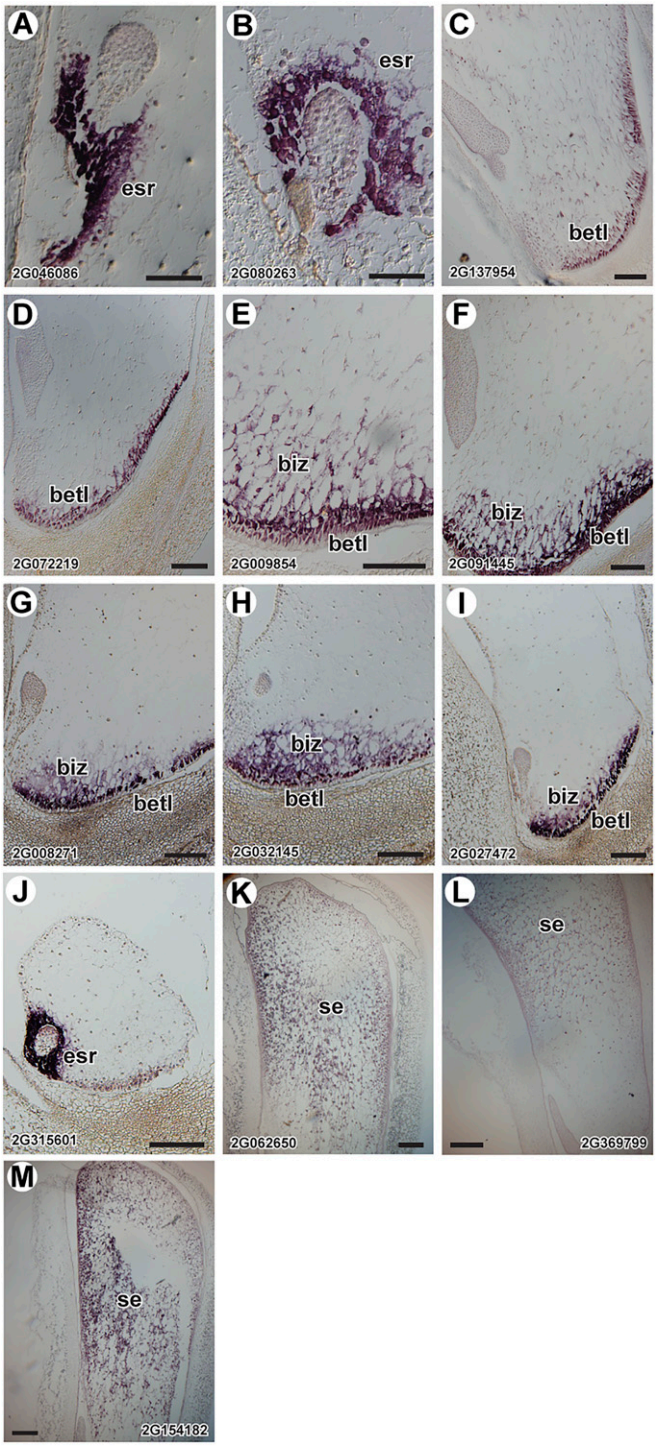

Fig. S5. Cell-specific expression of additional endosperm genes as determined by in situ hybridization. For mRNA patterns shown here and in Fig. 4, we focused on genes likely to be involved in cell differentiation events or markers of such events during endosperm development. We selected these genes based on expression pattern: low (fewer than 300 reads) at 0–3 DAP and strong up-regulation at 4–8 DAP (i.e., genes within the up-at-4-DAP, up-at-6-DAP, and up-at-8-DAP groups), which spans the periods of cell differentiation and the initial stages of storage deposition (Figs. 4 and 5; Figs. S1 and S5; Dataset S4, *Table S2*). (*A*) Expression in the embryo-surrounding region (ESR) only. Section of a 6-DAP kernel hybridized with a probe for GRMZM2G046086 (ESR1). (*B*) Expression in the ESR only. Section of a 6-DAP kernel hybridized with a probe for GRMZM2G080263 (glutelin-like). (*C*) Expression in the basal endosperm transfer layer (BETL) only. Section of a 10-DAP kernel hybridized with a probe for GRMZM2G137954 (uncharacterized). (*D*) Expression in the BETL only. Section of an 8-DAP kernel hybridized with a probe for GRMZM2G072219 (protease inhibitor, LTP family). (*E*) Expression in the BETL and basal-intermediate zone (BIZ). Section of an 8-DAP kernel hybridized with a probe for GRMZM2G009854 (protease inhibitor, LTP family). (*F*) Expression in the BETL and BIZ. Section of an 8-DAP kernel hybridized with a probe for GRMZM2G091445 (putative defensin). (*G*) Expression in the BETL and BIZ. Section of an 8-DAP kernel hybridized with a probe for GRMZM2G008271 [basal layer antifungal peptide (BAP)]. (*H*) Expression in the BETL and BIZ. Section of an 8-DAP kernel hybridized with a probe for GRMZM2G032145 (BURP domain-containing). (*I*) Expression in the BETL and BIZ. Section of an 8-DAP kernel hybridized with a probe for GRMZM2G027472 (BAP). (*J*) Expression in the ESR only. Section of a 10-DAP kernel hybridized with a probe for GRMZM2G315601 (ESR2). (*K*) Expression in the starchy endosperm

Legend continued on following page

(SE) only. Section of a 10-DAP kernel hybridized with a probe for GRMZM2G062650 (ZmNAC128). (*L*) Expression in the SE only. Section of a 10-DAP kernel hybridized with a probe for GRMZM2G369799 (ZmMYB127). (*M*) Expression in the SE only. Section of a 10-DAP kernel hybridized with a probe for GRMZM2G154182 (ZmNAC130). (Scale bars, 2 mm.) betl, basal endosperm transfer layer; biz, basal intermediate zone; esr, embryo surrounding region; se, starchy endosperm.

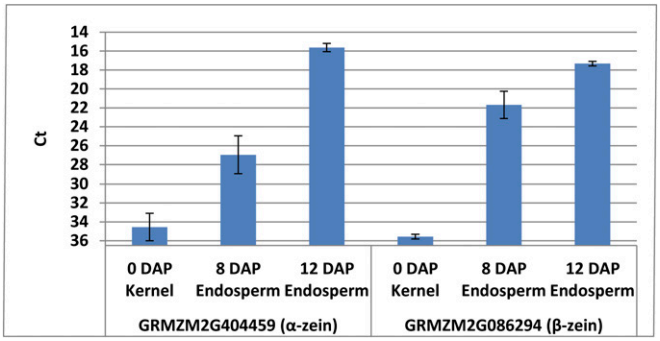

Fig. S6. Quantitative RT-PCR analysis of two zein genes expressed during endosperm development. The mRNA levels of an α-zein gene (*Left*) and a β-zein gene (*Right*) in unfertilized kernels (0 DAP) and 8- and 12-DAP hand-dissected endosperm measured using qRT-PCR.

Dataset S1.

Dataset S1

Dataset S2.

Dataset S2

(*Table S1*) Genes identified using StepMiner analysis. (*Table S2*) Transcription factor genes among one-step-up gene sets identified using StepMiner analysis.

Dataset S3.

Dataset S3

Dataset S4.

Dataset S4

(*Table S1*) qRT-PCR analysis of 13 TF genes shown as Ct values. (*Table S2*) Genes analyzed using in situ hybridization. (*Table S3*) Developmental stages and the associated gene numbers from this study compared with the data from Sekhon et al. (1) and Walley et al. (2). (*Table S4*) Normalized RNA-Seq reads of select zein genes in this study. (*Table S5*) qRT-PCR analysis of selected zein genes. (*Table S6*) Primer pair sequences of selected TF genes analyzed using qRT-PCR. (*Table S7*) Identification of the proper housekeeping gene for qRT-PCR validation of RNA-Seq data (using qRT-PCR, expressed in Ct values). (*Table S8*) Primers used to generate the clones for the in situ hybridization probes.

1. Sekhon RS, et al. (2013) Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. *PLoS ONE* 8(4): e61005.
2. Walley JW, et al. (2013) Reconstruction of protein networks from an atlas of maize seed proteotypes. *Proc Natl Acad Sci USA* 110(49):E4808–E4817.