LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development

Julie M. Pelletier[a], Raymond W. Kwong[a,1], Soomin Park[a,2], Brandon H. Le[b,3], Russell Baden[a,4], Alexandro Cagliari[a,5], Meryl Hashimoto[a,6], Matthew D. Munoz[a,7], Robert L. Fischer[c], Robert B. Goldberg[b,8], and John J. Harada[a,8]

[a]Department of Plant Biology, University of California, Davis, CA 95616; [b]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA 90095; and [c]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720

Contributed by Robert B. Goldberg, June 27, 2017 (sent for review May 13, 2017; reviewed by Seon-kap Hwang and Brian A. Larkins)

LEAFY COTYLEDON1 (LEC1), an atypical subunit of the nuclear transcription factor Y (NF-Y) CCAAT-binding transcription factor, is a central regulator that controls many aspects of seed development including the maturation phase during which seeds accumulate storage macromolecules and embryos acquire the ability to withstand desiccation. To define the gene networks and developmental processes controlled by LEC1, genes regulated directly by and downstream of LEC1 were identified. We compared the mRNA profiles of wild-type and *lec1*-null mutant seeds at several stages of development to define genes that are down-regulated or up-regulated by the *lec1* mutation. We used ChIP and differential gene-expression analyses in *Arabidopsis* seedlings overexpressing LEC1 and in developing *Arabidopsis* and soybean seeds to identify globally the target genes that are transcriptionally regulated by LEC1 *in planta*. Collectively, our results show that LEC1 controls distinct gene sets at different developmental stages, including those that mediate the temporal transition between photosynthesis and chloroplast biogenesis early in seed development and seed maturation late in development. Analyses of enriched DNA sequence motifs that may act as *cis*-regulatory elements in the promoters of LEC1 target genes suggest that LEC1 may interact with other transcription factors to regulate distinct gene sets at different stages of seed development. Moreover, our results demonstrate strong conservation in the developmental processes and gene networks regulated by LEC1 in two dicotyledonous plants that diverged ∼92 Mya.

maturation | photosynthesis | *Arabidopsis* | soybean

An unusual aspect of seed development is that it is temporally biphasic. After seed development is initiated with the double fertilization of the egg and central cells, giving rise to the zygote and endosperm mother cell, respectively, the embryo and endosperm undergo the morphogenesis phase. During this phase, the basic body plan of the embryo and endosperm are established through morphogenetic events that include cellular and nuclear proliferation, the specification and establishment of subregions and domains, and the differentiation of tissue and cell types (1, 2). Chloroplast biogenesis and photosynthesis are also initiated during this period in many angiosperm taxa (3). The maturation phase partially overlaps but largely follows the morphogenesis phase. During the maturation phase, cell proliferation and morphogenesis become arrested, storage macromolecules, such as lipids and proteins, accumulate to massive amounts and are sequestered in organelles, and the embryo acquires the ability to withstand desiccation (4, 5). At the end of seed development, the embryo and endosperm are arrested developmentally and quiescent metabolically, and they remain so until the seed germinates.

LEAFY COTYLEDON1 (LEC1), an unusual nuclear transcription factor YB (NF-YB) subunit of the NF-Y CCAAT-binding transcription factor (TF), is a central regulator of seed development (6). Loss-of-function *lec1* mutations cause defects in storage protein and lipid accumulation, acquisition of desiccation tolerance, and the suppression of germination and leaf primordia initiation (reviewed in refs. 5 and 7). The expression of many maturation genes encoding storage proteins, oil body proteins, and transcriptional regulators of the maturation phase is defective in *lec1* mutants. Moreover, ectopic expression of *LEC1* induces the activation of genes involved in maturation and in storage protein and lipid accumulation in vegetative organs (6, 8–10). These findings and others implicate LEC1 and the B3 domain TFs ABA INSENSITIVE3 (ABI3), FUSCA3 (FUS3), and LEC2 as master regulators of the maturation phase (reviewed in ref. 11). Analyses of interactions among these TFs suggest that *LEC1* acts at the highest level in the regulatory hierarchy controlling the maturation phase (4, 5, 9, 12). Despite its importance, knowledge of the gene-regulatory networks controlled by LEC1 is limited. LEC1 has been shown to bind to genes that are

Significance

Seed development is biphasic, consisting of the morphogenesis phase when the basic plant body plan is established and the maturation phase when the embryo accumulates storage reserves and becomes desiccation tolerant. Despite the importance of seeds as human food and animal feed, little is known about the gene-regulatory networks that operate during these phases. We identified genes that are regulated genetically and transcriptionally by a master regulator of seed development, LEAFY COTYLEDON1 (LEC1). We show that LEC1 transcriptionally regulates genes involved in photosynthesis and other developmental processes in early and maturation genes in late seed development. Our results suggest that LEC1 partners with different transcription factors to regulate distinct gene sets and that LEC1 function is conserved in *Arabidopsis* and soybean seed development.

Author contributions: J.M.P., R.W.K., S.P., B.H.L., R.B., A.C., M.H., R.L.F., R.B.G., and J.J.H. designed research; J.M.P., R.W.K., S.P., B.H.L., R.B., A.C., M.H., M.D.M., and J.J.H. performed research; J.M.P., R.W.K., S.P., B.H.L., R.B., and J.J.H. analyzed data; and J.M.P., R.B.G., and J.J.H. wrote the paper.

Reviewers: S.-k.H., Washington State University; and B.A.L., University of Nebraska.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession nos. GSE1051, GSE99528, GSE99529, GSE99587, and GSE99882).

[1]Present address: Beckman Coulter Inc., West Sacramento, CA 95691.

[2]Present address: Experiment Research Institute of National Agricultural Products Quality Management Service, Ministry of Agriculture, Food and Rural Affairs, Gimcheon, Korea.

[3]Present address: Department of Botany and Plant Sciences, University of California, Riverside, CA 92521.

[4]Present address: California Animal Health and Food Safety Laboratory, University of California, Davis, CA 95616.

[5]Present address: Universidade Estadual do Rio Grande do Sul, Santa Cruz do Sul, Rio Grande do Sul State, Brazil.

[6]Present address: Seminis, Inc., Woodland, CA 95695.

[7]Present address: Bionano Genomics, Inc., San Diego, CA 92121.

[8]To whom correspondence may be addressed. Email: jjharada@ucdavis.edu or bobg@ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707957114/-/DCSupplemental.

involved in lipid metabolism, hormone responses, and light signaling, and it appears to regulate transcriptionally genes involved in maturation in concert with two other TFs, NF-YC2 and basic LEUCINE ZIPPER TRANSCRIPTION FACTOR 67 (bZIP67) (8, 13, 14).

LEC1 is also required for other aspects of seed development. *lec1* mutants are defective in the maintenance of suspensor and cotyledon identity early in seed development, and ectopic *LEC1* expression results in somatic embryo formation on vegetative tissues (6, 15, 16). In addition to regulating maturation genes, ectopic *LEC1* expression up- and down-regulates genes involved in hormone responses and down-regulates genes that respond to light in seedlings (8). These findings suggest that LEC1 controls other aspects of seed development in addition to the maturation phase. However, the LEC1 gene networks that control these diverse sets of developmental processes remain to be identified.

We present studies that provide unexpected insights into the developmental processes and gene networks that are regulated by LEC1 during seed development. mRNA transcriptome analyses of *lec1*-null mutants were combined with the identification of genes directly regulated by the LEC1 TF in *Arabidopsis* seedlings ectopically expressing *LEC1* and in developing *Arabidopsis* seeds and soybean embryos. Together, our studies provide evidence that LEC1 regulates distinct developmental processes in seeds, including photosynthesis/chloroplast biogenesis and seed maturation. Moreover, our studies suggest that LEC1 may regulate distinct gene sets by working combinatorially with different TFs at different stages of seed development.

Results

mRNA Profiling of Developing *lec1*-Mutant *Arabidopsis* Seeds Indicates a Role for LEC1 in Several Developmental Processes. To obtain an overview of the developmental processes that are controlled by LEC1, we profiled mRNA populations in seeds homozygous for the *lec1-1*–null mutation at five different stages of seed development using Affymetrix ATH1 GeneChips. The 24 h after pollination (24H), globular (GLOB), and linear cotyledon (LCOT) stages and the mature green (MG) and postmature green (PMG) stages represent the morphogenesis and maturation phases, respectively.

Fig. 1*A* summarizes the number of diverse mRNAs that were detected as present in *lec1-1*–mutant seeds compared with previously determined values for wild-type seeds at the same stages [Gene Expression Omnibus (GEO) accession GSE680] (17). mRNA numbers in *lec1-1* seeds remained relatively constant throughout seed development ($P > 0.91$, ANOVA) in contrast to wild-type seeds in which mRNA numbers decreased significantly during seed maturation at the MG and PMG stages ($P < 0.001$) (17). This result is consistent with previous findings that *lec1-1*–mutant seeds, unlike wild-type seeds, do not become quiescent developmentally or metabolically at late seed-development stages (16).

We designated mRNAs regulated by LEC1 as those whose levels were at least twofold higher or lower in *lec1-1* mutant than in wild-type seeds at the same stage at a statistically significant level [false discovery rate (FDR) < 0.05] (Fig. 2*A* and Dataset S1). The *lec1-1* mutation prominently affected mRNA levels during the maturation phase. Ninety-five percent of the 2,624 *lec1*$^-$–down-regulated mRNAs that were at lower levels in *lec1-1* mutant than in wild-type seeds and 99% of the 3,256 *lec1*$^-$–up-regulated mRNAs that were at higher levels in *lec1-1* mutant than in wild-type seeds at any stage accumulated differentially at the MG and/or PMG stages (Fig. 1*B*). Similarly, pairwise comparisons of mRNA populations in wild-type and *lec1-1*–mutant seeds revealed strong similarities at the 24H, GLOB, and LCOT stages (Pearson correlation coefficients, 0.99, 0.98, and 0.98, respectively) but showed more substantial differences at the MG and PMG stages (Pearson correlation coefficients, 0.81 for both). Thus, relatively few differences in gene activity between WT and *lec1-1* seeds were detected early in development, but major differences were observed during the seed-maturation phase. The cause of this biased representation may be that *LEC1* and LEC1-regulated genes are expressed in the embryo and endosperm, and these seed regions constitute only a small part of the seed early in development.

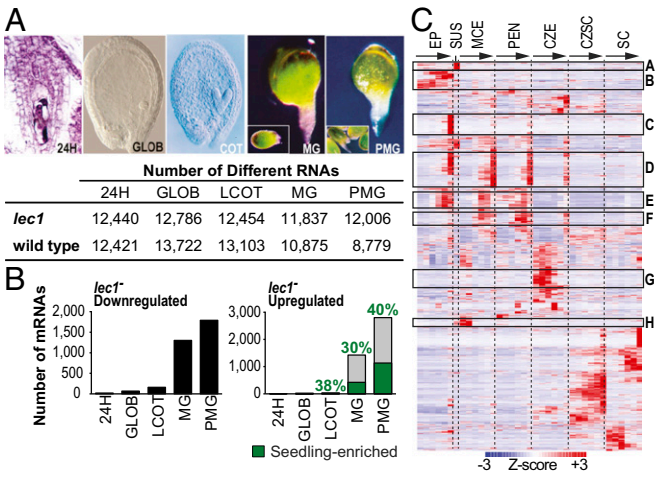

Fig. 1. mRNA profiling of *lec1* mutant seeds throughout development. (*A*) The number of diverse mRNAs detected in *lec1-1*–mutant seeds compared with wild-type seeds (17) at the indicated seed-development stages as determined in ATH1 GeneChip hybridization studies. Representative seeds and MG and PMG embryos as viewed by bright-field (24H), differential interference contrast (GLOB and LCOT), and dark-field whole-mount (MG and PMG) microscopy (*Insets*, MG and PMG seeds). (*B*) Numbers of mRNAs differentially expressed between *lec1-1* and wild-type seeds at the indicated stages define *lec1*$^-$–down-regulated (*Left*) and *lec1*$^-$–up-regulated (*Right*) mRNAs. The green shading and percentages denote *lec1*$^-$–up-regulated mRNAs that also are detected at significantly higher levels in seedlings than in seeds (seedling-enriched). Lists of the mRNAs and their levels that are present in *lec1-1* mutants, that are *lec1*$^-$ regulated, and that are seedling specific are given in Dataset S1. (*C*) Hierarchical clustering of *lec1*$^-$–down-regulated mRNAs. The heatmap shows relative mRNA levels in each subregion at the preglobular, GLOB, heart, LCOT, BCOT, and MG stages (left to right, as indicated by the arrow). SUS mRNAs are shown at the GLOB stage.

We obtained insight into LEC1-regulated processes by using hierarchical clustering to identify when and where *lec1*$^-$–down-regulated and –up-regulated mRNAs normally accumulate, taking advantage of our previously generated dataset of mRNA levels in the embryo proper (EP), suspensor (SUS), micropylar (MCE), peripheral (PEN), and chalazal (CZE) endosperm and the distal seed-coat (SC) and chalazal seed-coat (CZSC) subregions at six different stages of development: preglobular, GLOB, heart, LCOT, bent cotyledon (BCOT), and MG (18). mRNAs affected by the *lec1-1* mutation accumulated primarily in embryo and endosperm subregions in spatially and temporally controlled patterns (Fig. 1*C* and Fig. S1). Consistent with LEC1's role in controlling the maturation phase, one cluster (D) of *lec1*$^-$–down-regulated mRNAs accumulated late in development in embryo and endosperm subregions, and it was overrepresented ($P < 0.001$, hypergeometric distribution) for Gene Ontology (GO) terms associated with maturation processes, such as monolayer-surrounded lipid storage body, lipid storage, and seed oilbody biogenesis (Fig. 3 and Dataset S1). This cluster also contained TF mRNAs known to be involved in maturation, including ABI3, bZIP67, and ENHANCED EM LEVEL (EEL) (Dataset S1). Fig. 4 shows that 30 of 50 maturation (MAT) genes were *lec1*$^-$–down-regulated at the LCOT, MG, and/or PMG stages. MAT genes were shown previously in mRNA transcriptome studies to be expressed predominantly during the maturation phase and to encode proteins known or predicted to function in maturation processes (18). We also reanalyzed publically available datasets to identify MAT genes that were down-regulated by mutations in two

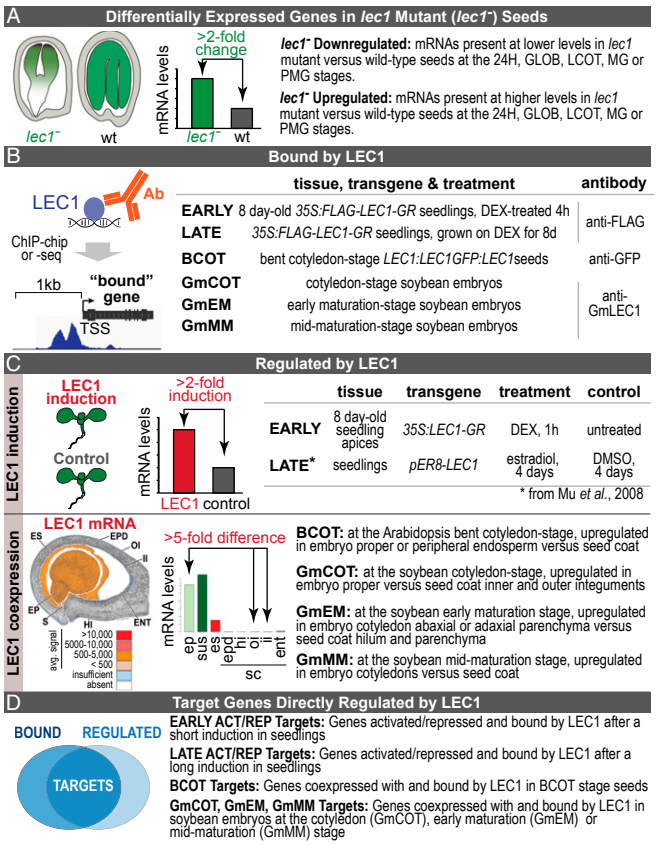

Fig. 2. Design of experiments to identify genes regulated genetically and directly by LEC1 in *Arabidopsis* and soybean. (*A*) *lec1*$^{-}$–down-regulated and *lec1*$^{-}$–up-regulated mRNAs accumulate to a level that is at least twofold lower or higher (FDR < 0.05), respectively, in *lec1-1*–mutant than in wild-type seeds. (*B*) LEC1-bound regions were identified with ChIP-chip (EARLY and LATE) or ChIP-seq (BCOT, GmCOT, GmEM, and GmMM) experiments. (*Left*) Bound genes have a LEC1-binding site within the 1-kb region upstream of the transcription start site (TSS). (*Right*) Plant materials and antibodies used for the ChIP experiments. (*C*, *Upper*) GeneChip experiments EARLY (1 h) and LATE (4 d) after LEC1 induction identified mRNAs whose levels increased (ACT) or decreased (REP) at least twofold (FDR < 0.05) relative to the controls. (LATE data are from ref. 10.) (*C*, *Lower Left*) *LEC1* is expressed in embryo subregions but not in seed-coat subregions. (*Lower Center*) LEC1-coexpressed mRNAs are present at fivefold or higher levels (FDR < 0.05) in embryo subregions than in seed-coat subregions. (*Lower Right*) The subregions compared at the indicated stages are listed. ent, endothelium; ep, embryo proper; epd, epidermis; es, endosperm; hi, hilum; ii, inner integuments; oi, outer integuments; sc, seed coat; sus, suspensor. (*D*, *Left*) LEC1 target genes are bound and regulated by LEC1. (*Right*) Target gene sets.

other maturation-phase regulators, *ABI3* and *FUS3* (GEO accession no. GSE61686) (Fig. 4) (19).

Analysis of other *lec1*$^{-}$–down-regulated mRNA clusters suggests that LEC1's role in seed development is not limited to the maturation phase. For example, one cluster (E) with mRNAs that accumulated in embryo and endosperm subregions primarily at the MG stage was overrepresented for GO terms related to photosynthesis and chloroplast biogenesis (abbreviated "PSN") (Figs. 1 and 3 and Dataset S1), suggesting that LEC1 regulates these processes directly or indirectly. Another cluster (B) of EP mRNAs was overrepresented for the GO terms organ morphogenesis and regulation of cell proliferation and contained TFs including BBM, PAN, and WOX2 that are known to be involved in embryo development. Other clusters contained mRNAs that accumulated primarily in a single subregion, including the SUS (cluster A), EP (clusters B and C), CZE (cluster G), and MCE (cluster H) (Fig. 1*C*), and none of these mRNA sets was overrepresented for GO terms related to maturation processes (Fig. 3 and Dataset S1).

By contrast, we found that many *lec1*$^{-}$–up-regulated mRNAs were normally expressed during seedling development. Approximately 30% and 40% of *lec1*$^{-}$–up-regulated mRNAs at the MG and PMG stages, respectively, overlapped with seedling-enriched mRNAs, i.e., mRNAs present at fivefold or higher levels in seedlings than in seeds at any stage (FDR < 0.05) (GEO accession no. GSE680) (Fig. 1*B*), and 20 of 55 and 49 of 86 overrepresented GO terms associated with *lec1*$^{-}$–up-regulated mRNAs at the MG and PMG stages, respectively, were also associated with seedling-specific mRNAs (Fig. S1 and Dataset S1). This finding is consistent with reports that LEC1 is required to inhibit postgerminative development in seeds (7). Many genes encoding PSN proteins were *lec1*$^{-}$–down-regulated at the MG stage and *lec1*$^{-}$–up-regulated at the PMG stage, suggesting that the *lec1* mutation compromised the activation of many PSN genes at or before the MG stage and their repression during the transition into metabolic quiescence (Dataset S1). Together, these results suggest that LEC1 directly or indirectly regulates a number of distinct cellular processes during seed development, including seed maturation and photosynthesis.

Fig. 3. Predicted biological functions of *lec1*$^{-}$–down-regulated and LEC1 target genes. Heatmaps show the *P* value (*Arabidopsis*, $P \leq 0.001$ cutoff) and *q* value (soybean, $q \leq 0.05$ cutoff) significance of GO terms for *lec1*$^{-}$–down-regulated gene clusters and LEC1 target gene sets. The GO terms listed represent the five most enriched GO terms for each gene set. The complete GO term lists, the corresponding genes, and their significance levels are given in Datasets S1, S2, and S6.

AGI	Gene	Downregulated in *lec1*⁻			Downregulated in *abi3*⁻		Downregulated in *fus3*⁻		Target Genes				
		LCOT	MG	PMG	12 DAP	16 DAP	8 DAP	12 DAP	LATE ACT	BCOT	GmCOT	GmEM	GmMM
AT1G03880	Cruciferin 2 (CRU2, CRB)												
AT1G32560	LEA family protein												
AT1G47540	Scorpion toxin-like knottin superfamily protein										NA	NA	NA
AT1G52690	LEA family protein												
AT2G18540	Cupin family protein												
AT2G23120	LEA family protein										NA	NA	NA
AT2G28490	Cupin family protein												
AT2G36640	Embryonic cell protein 63 (ECP63)												
AT2G41070	BZIP12 (DPBF4, ATBZIP12, EEL)												
AT3G01570	Oleosin family protein												
AT3G02480	LEA family protein												
AT3G44460	BZIP67 (DPBF2, ATBZIP67)												
AT4G25140	Oleosin 1 (OLEO1, OLE1)												
AT4G26740	Seed gene 1 (ATS1, ATPXG1, CLO1)												
AT4G27140	Seed storage albumin 1 (AT2S1, SESA1)										NA	NA	NA
AT4G27160	Seed storage albumin 3 (AT2S3, SESA3)										NA	NA	NA
AT4G27170	Seed storage albumin 4 (AT2S4, SESA4)										NA	NA	NA
AT4G28520	Cruciferin 3 (CRC, CRU3)										NA	NA	NA
AT4G34520	3-ketoacyl-CoA synthase 18 (KCS18, FAE1)										NA	NA	NA
AT5G40420	Oleosin 2 (OLEO2, OLE2)										NA	NA	NA
AT5G44120	Cruciferin 1 (CRU1, ATCRA1, CRA1)												
AT5G47670	Nuclear factor Y, subunit B6 (NF-YB6, L1L)												
AT5G49190	Sucrose synthase 2 (SSA, ATSUS2, SUS2)												
AT5G51210	Oleosin3 (OLEO3)										NA	NA	NA
AT5G54740	Seed storage albumin 5 (SESA5)										NA	NA	NA
AT1G01470	LEA protein (LSR3, LEA14)												
AT1G03890	RmlC-like cupins superfamily protein												
AT1G05510	Protein of unknown function (DUF1264)												
AT1G22710	Sucrose-proton symporter 2 (SUC2)												
AT1G62500	LTP/seed storage 2S albumin												
AT1G72100	LEA domain-containing protein										NA	NA	NA
AT2G21490	Dehydrin (LEA)												
AT2G25890	Oleosin family protein												
AT2G38530	Lipid transfer protein 2 (LP2, cdf3, LTP2)												
AT3G15670	LEA family protein										NA	NA	NA
AT3G17520	LEA family protein										NA	NA	NA
AT3G18570	Oleosin family protein												
AT3G22640	Cupin family protein (PAP85)												
AT3G27660	Oleosin 4 (OLE3,OLEO4)										NA	NA	NA
AT3G53040	LEA family protein, putative												
AT4G01410	LEA, hydroxyproline-rich glycoprotein family												
AT4G21020	LEA family protein												
AT4G27150	Seed storage albumin 2 (AT2S2, SESA2)										NA	NA	NA
AT4G36700	RmlC-like cupins superfamily protein										NA	NA	NA
AT4G38410	Dehydrin family protein										NA	NA	NA
AT5G07190	Seed gene 3 (ATS3)										NA	NA	NA
AT5G07500	PEI1										NA	NA	NA
AT5G44310	LEA family protein										NA	NA	NA
AT5G45830	Delay of germination 1 (ATDOG1)										NA	NA	NA
AT5G55240	ATPXG2												

Fig. 4. Maturation genes and their regulators. Filled squares indicate *Arabidopsis* MAT genes and their closest homologs in soybean that are *lec1*⁻– (black), *abi3*⁻– (dark gray), or *fus3*⁻– (light gray) down-regulated at the indicated seed-development stage or LEC1 target genes at the indicated stages (LATE ACT, red; BCOT, dark green; GmCOT, forest green; GmEM, light green; GmMM, gold). *Arabidopsis* MAT genes with no obvious ortholog in soybean are marked NA.

LEC1 Regulates Different Genes Early and Late After the Induction of LEC1 Activity in Seedlings. The *lec1*⁻ mRNA transcriptome analysis suggested that LEC1 directly and/or indirectly activates and represses genes involved in diverse developmental processes. To determine which processes are regulated transcriptionally by LEC1, we identified LEC1 target genes, defined as genes that are both bound and regulated by LEC1. LEC1 target genes were identified in seedlings containing an inducible form of LEC1. Experiments were done with seedlings, because mRNAs encoding LEC1 and other TFs with partially overlapping functions, such as ABI3, FUS3, and LEC2, are not normally present at appreciable levels in seedlings (5). Fig. 2 *B* and *C* summarizes the ChIP experiments used to identify LEC1-bound genes *in planta* and the differential expression analyses used to identify genes regulated following the induction of LEC1 activity. As shown in Fig. 5*A*, the induction of LEC1 activity caused the formation of embryo-like seedlings similar to those observed in *35S:LEC1* seedlings (6, 8, 10).

Genomic DNA regions bound* in planta *by LEC1. Because *LEC1* is expressed from the earliest stage of seed development through maturation (18), we conducted two sets of ChIP followed by DNA microarray (ChIP-chip) experiments to identify LEC1-binding sites: one 4 h after the induction of LEC1 activity (EARLY) in *35S:FLAG-LEC1-GR* seedlings 8 d after imbibition and another with *35S:FLAG-LEC1-GR* seedlings that were induced for 8 d beginning at germination (LATE). Following ChIP with an anti-FLAG antibody, the immunoprecipitated DNA was amplified and hybridized with the GeneChip *Arabidopsis* Tiling 1.0R Array containing probes for the complete nonrepetitive *Arabidopsis* genome.

As summarized in Fig. 5*B*, ChIP-chip analyses showed that LEC1 bound 2,753 genomic regions 4 h after induction of LEC1 activity (EARLY BD), and 4,297 genomic regions 8 d after induction of LEC1 activity (LATE BD). The bound regions were within the 1-kb upstream region of 1,252 and 2,539 genes, respectively, that are represented as single genes (singletons) on the ATH1 GeneChip (Dataset S2). Control experiments shown in Dataset S3 validated the ChIP experiments and provided strong evidence that the anti-FLAG antibody bound FLAG-LEC1-GR specifically. Thus, LEC1 bound genes both early and late following induction.

Identification of genes regulated by LEC1. Only a small fraction of the genes bound by a TF are regulated by that TF (20). For example, the chromosome browser view in Fig. 5*H* shows that several genes that were bound by LEC1 at a statistically significant level were not regulated by LEC1 in seedlings. Therefore, we identified genes that were regulated following a 1-h induction of LEC1 activity using ATH1 GeneChip hybridization experiments. We identified 382 EARLY ACTIVATED (EARLY ACT) mRNAs whose levels increased at least twofold and 193 EARLY REPRESSED (EARLY REP) mRNAs whose levels decreased to 50% or less of control levels following induction (FDR <0.05) (Fig. 5*C* and Dataset S2).

To identify genes regulated after a long-term (4-d) induction of LEC1 activity, we reanalyzed previously published data from ATH1 GeneChip hybridization experiments (10). Fig. 5*C* shows that 508 LATE ACTIVATED (LATE ACT) and 390 LATE REPRESSED (LATE REP) mRNAs, respectively, were up- and down-regulated relative to controls by this long-term induction of LEC1 activity. We tested mRNAs from *35S:LEC1-GR* seedlings grown for 8 d following LEC1 induction using qRT-PCR and showed that 15 of 16 up-regulated mRNAs tested were validated in the 8-d induction experiments (Dataset S3).

Target genes directly regulated by LEC1. We identified target genes that were bound by LEC1 and activated or repressed by the induction of LEC1 activity. Fig. 5 *D* and *F* shows that 16% of EARLY ACT and 14% of EARLY REP mRNAs were associated with genes that were bound by LEC1 at 4 h (Dataset S2). The overlaps between bound and regulated genes were statistically significant ($P < 1.4 \times 10^{-12}$ and $P < 7.4 \times 10^{-54}$, respectively, hypergeometric distribution). Similarly, 31% of LATE ACT ($P < 2.5 \times 10^{-29}$) and 12% of LATE REP ($P = 3.8 \times 10^{-1}$) mRNAs were bound by LEC1 at 8 d. The results suggest that LEC1 is involved in both the transcriptional activation and repression of genes early and late following induction of its activity.

Because the *lec1*⁻ transcriptome analyses suggested that LEC1 may regulate different developmental processes early and late in seed development, we compared EARLY and LATE target genes. Of the 63 EARLY ACT and 160 LATE ACT target genes, only one overlapped ($P = 0.40$) (Fig. 5*E*), and of the 27 EARLY REP and 47 LATE REP target genes only two overlapped ($P < 0.02$). Direct comparison of genes bound by LEC1 at 4 h and 8 d showed that there was substantial overlap in the genes that were bound at 4 h and 8 d after induction ($P < 0$) (Fig. 5*F*). By contrast, there was little overlap in the genes that were regulated by LEC1

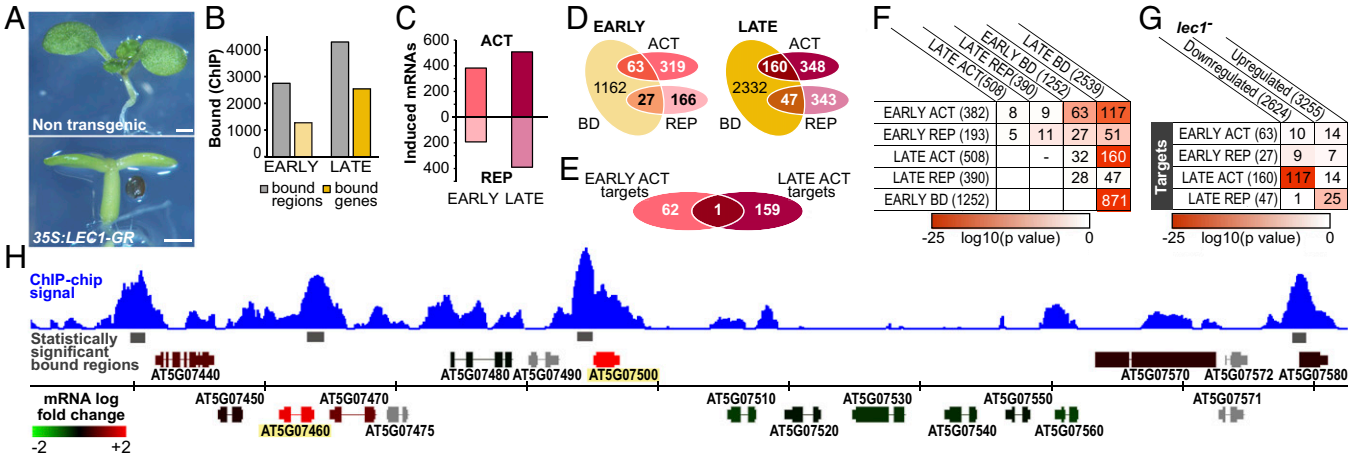

Fig. 5. LEC1 directly regulates different gene sets early and late following induction. (*A*) Effect of LEC1 activity on seedling development. *35S:LEC1-GR* and nontransgenic seedlings were grown on medium containing Dex for 12 d. (Scale bars: 5 mm.) (*B*) Numbers of genomic regions and genes bound by LEC1 in ChIP-chip experiments at 4 h (EARLY) or 8 d (LATE) after LEC1 induction in seedlings. (*C*) Numbers of mRNAs activated (ACT) and repressed (REP) following LEC1 induction for 1 h (EARLY) or 4 d (LATE) (10) at a 0.05 FDR significance level. (*D*) Target genes directly regulated by LEC1. Venn diagrams show the overlap between activated and repressed genes that are bound (BD) by LEC1 EARLY (*Left*) and LATE (*Right*) after LEC1 induction. (*E*) Overlap between activated target genes EARLY and LATE after induction. (*F*) Pairwise comparisons of the genes regulated and/or bound by LEC1 EARLY or LATE after induction. The number in each row and column intersection indicates the number of genes in both lists, and the heatmap shading represents the statistical significance of the overlap. (*G*) Comparison of EARLY and LATE LEC1 target genes and *lec1*$^{-}$-regulated genes. (*H*) Genome browser view of the chromosomal region surrounding the *PEI* gene (AT5G07500) showing enrichment of genomic regions bound by LEC1 (ChIP-chip signal relative to control, blue peaks), statistically significant LEC1-bound regions (gray bars), and gene models that are colored to indicate the mRNA fold-change following LEC1 induction (red, activated; green, repressed by LEC1; gray not present on the ATH1 gene chip). LEC1 target genes are highlighted in yellow. The axis is divided into 5-kb segments. Lists of bound genomic regions and genes, activated and repressed mRNAs, and target genes are given in Dataset S2.

early and late following induction. Only eight EARLY ACT and LATE ACT mRNAs overlapped ($P = 0.47$), and only 11 EARLY REP and LATE REP mRNAs overlapped ($P < 0.001$) (Fig. 5*F*). We did find, however, that many genes that were targets only early or late after LEC1 induction remained bound throughout the period tested. For example, 47 of 63 EARLY ACT target genes and 21 of 27 EARLY REP target genes remained bound at 8 d ($P < 2.3 \times 10^{-29}$ and $P < 1.9 \times 10^{-14}$, respectively), and 25 of 160 LATE ACT targets and 17 of 47 LATE REP targets were also bound at 4 h ($P < 2.1 \times 10^{-5}$ and $P < 1.4 \times 10^{-9}$, respectively). These results suggest that LEC1 binding alone is not sufficient to regulate the expression of these genes, opening the possibility that some other factor(s) contributes to the activation and repression of LEC1 target genes early and late after induction.

We compared the EARLY and LATE targets with genes that were affected by the *lec1-1* mutation and found that the most significant overlap occurred between LATE ACT targets and the *lec1*$^{-}$–down-regulated genes (Fig. 5*G*). Analysis of overrepresented GO terms showed that the LATE ACT targets had the greatest functional overlap with the *lec1*$^{-}$–down-regulated cluster D (Fig. 1*C*), in that they were overrepresented for the GO terms monolayer-surrounded lipid storage body, lipid storage, seed oilbody biogenesis, and seed germination, all of which are characteristic of maturation processes (Fig. 3 and Dataset S2). Moreover, of the 50 MAT genes listed in Fig. 4, 30 were LATE ACT target genes. In addition, genes encoding TFs known to play roles in controlling maturation, including *LEC1*, *FUS3*, *ABI3*, *bZIP67*, and *WRI1*, were LEC1 target genes (Dataset S2). *LEC2* is also an LEC1 target gene, because it is bound by LEC1 at 4 d and qRT-PCR experiments showed that LEC2 was up-regulated by LEC1 induction at 8 d (Dataset S3). By contrast, EARLY ACT target genes were most significantly enriched for the GO terms positive regulation of translation, kinase activity, response to abscisic acid stimulus, TF activity, trehalose-phosphatase activity, and biosynthetic process.

Together, our results indicate that LEC1 directly activates and represses different target genes at different times after induction. LEC1 binding alone does not appear to be sufficient to regulate gene expression, opening the possibility that other TFs participate in the activation and repression of LEC1 target genes early and late after induction.

LEC1 Transcriptionally Regulates Diverse Gene Sets in *Arabidopsis* Seeds. We identified LEC1 target genes in developing *Arabidopsis* seeds to determine if different target genes are activated at different stages of seed development as they are early and late following LEC1 induction in seedlings. We used transgenic *lec1-1*–mutant plants containing a *LEC1-GFP* chimeric gene that was fused with the endogenous *LEC1* 5′- and 3′-flanking regions (*LEC1:LEC1-GFP:LEC1*) (Fig. 2). As shown in Fig. 6 *A* and *B*, analysis of GFP activity confirmed that the transgene was active in embryo and endosperm subregions, as predicted from our previous analyses of *LEC1* mRNA levels (Fig. 6*C* and ref. 18).

As outlined in Fig. 2*B*, genes bound by LEC1 in BCOT-stage seeds were identified using ChIP experiments with an anti-GFP antibody followed by DNA sequencing analysis (ChIP-seq). We analyzed BCOT-stage seeds 8–9 d after pollination because the maturation phase is initiated at approximately this stage, and *LEC1* mRNA was prevalent in the embryo and endosperm at this stage (Fig. 6*C*). As summarized in Fig. 6*D*, we identified 3,703 singleton genes that were bound by LEC1 (Dataset S2). Control experiments validated the analysis and provided strong evidence that the anti-GFP antibody specifically immunoprecipitated LEC1-GFP (Dataset S3).

To identify genes that are activated by LEC1, we reasoned that their expression should be significantly higher in seed subregions containing *LEC1* mRNA than in those lacking *LEC1* mRNA. We profiled the mRNA transcriptomes of five seed subregions at the BCOT stage: EP, MCE, PEN, CZE, SC, and CZSC, and showed that similar numbers of distinct mRNAs accumulated in each subregion, as observed previously at other stages (Fig. S2, Dataset S4, and ref. 18). Because *LEC1* mRNA was present at high levels in the embryo proper and endosperm subregions at the BCOT stage and at extremely low levels in seed-coat subregions (Fig. 6*C* and Dataset S4), mRNAs coexpressed with LEC1 were defined as

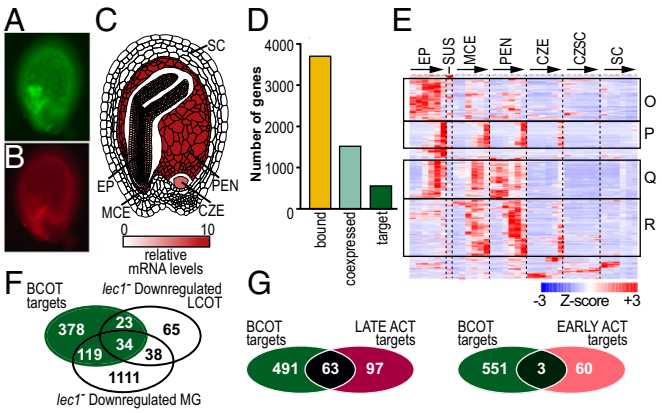

Fig. 6. LEC1 target genes in *Arabidopsis* BCOT-stage seeds. (*A* and *B*) GFP fluorescence in the embryo and endosperm of a heart stage *LEC1:LEC1-GFP: LEC1* seed (*A*) compared with the autofluorescence signal (*B*). (*C*) Relative *LEC1* mRNA levels in subregions of the BCOT seed. (*D*) Numbers of LEC1-bound (singletons), LEC1-coexpressed, and LEC1 target genes at the BCOT stage. (*E*) Hierarchical clustering of mRNAs corresponding to BCOT target genes. The heatmap organization of seed subregions and developmental stages is as in Fig. 1. (*F* and *G*) Venn diagrams showing the overlap between BCOT LEC1 target genes and *lec1*$^{-}$–down-regulated mRNAs at the LCOT and MG stages (*F*), LATE ACT targets (*G*, *Left*), and EARLY ACT targets (*G*, *Right*).

those that are present at a fivefold or higher level in the embryo proper and/or peripheral endosperm than in the distal seed coat (FDR <0.05) (Fig. 2). We identified 1,515 genes that were coexpressed and potentially activated by LEC1 (Fig. 6*D* and Dataset S2).

We identified 554 LEC1 target genes that represented a significant overlap between LEC1- bound and -coexpressed genes ($P < 2.0 \times 10^{-67}$) (Fig. 6*D* and Dataset S2). Of the LEC1 targets, 176 overlapped with the 1,390 genes that were *lec1*$^{-}$–down-regulated at the LCOT and MG stages ($P < 1.1 \times 10^{-70}$) (Fig. 6*F*), confirming their biological significance. Moreover, the BCOT target genes overlapped significantly with the LATE ACT target genes (63 of 554, $P < 4.8 \times 10^{-56}$) in seedlings but showed little similarity with EARLY ACT target genes (3 of 554, $P = 0.25$) (Fig. 6*G*).

We clustered the BCOT target mRNAs to obtain clues about LEC1-regulated processes in seeds and identified at least four mRNA sets with distinct spatial and temporal accumulation patterns (Fig. 6*E*). One cluster (O) with mRNAs that accumulated primarily in the EP at the earliest stages of seed development was enriched for GO terms related to growth and morphogenesis, including microtubule motor activity, phragmoplast, polarity specification of adaxial/abaxial axis, regulation of meristem structural organization, and asymmetric cell division, and contained TFs that play roles in morphogenetic processes in the embryo such as PHV, PHB, AS1, and SCR (Fig. 3 and Dataset S2). Another cluster (Q) contained mRNAs that accumulated in the EP, MCE, and PEN from middle to late developmental stages and had representatives of most gene families encoding the light-reaction components of photosystems I and II (Fig. S3). The great majority of these PSN target genes were also *lec1*$^{-}$–down-regulated (Fig. S4). This mRNA set was overrepresented for the GO terms chloroplast thylakoid membrane, chloroplast, chloroplast envelope, thylakoid, and photosynthesis (Fig. 3 and Dataset S2). Additional LEC1 target genes that were both related to chloroplast function and *lec1*$^{-}$–down-regulated were also identified (Dataset S5, Table S1), suggesting that LEC1 has an integral role in regulating photosynthesis and chloroplast functions in seeds. A maturation cluster (P) of mRNAs that accumulated at the latest stages of development in the EP and all three endosperm domains contained TFs known to regulate maturation processes, including EEL, ABI3, bZIP67, L1L, and 25 of the 50 MAT genes, although the mRNA levels of only 12 of these target genes were significantly affected by the *lec1-1* mutation (Fig. 4 and Dataset S2). This maturation mRNA set was overrepresented for the GO terms nutrient reservoir activity, monolayer-surrounded lipid storage body, lipid storage, endomembrane system, and seed oilbody biogenesis. A final cluster (R) contained mRNAs that accumulated primarily in all three endosperm domains and contained TFs known to regulate maturation, including LEC1, FUS3, and WRI1, although the overrepresented GO terms were not typical of maturation (Fig. 3). Together, these results suggest that LEC1 directly regulates distinct gene sets that mediate morphogenetic processes, photosynthesis, and maturation among other cellular processes during seed development.

Analyses of LEC1 Target Genes in Developing Soybean Seeds Indicate Different Roles for LEC1 Early and Late in Seed Development. Our results strongly suggested that LEC1 regulates different gene sets at different stages of seed development. To verify this conclusion and to determine if LEC1's diverse functions in seed development are conserved, we identified soybean LEC1 (GmLEC1) target genes at several stages of soybean seed development. Four *LEC1* paralogs were identified in soybean, *GmLEC1-1* (Glyma.07G268100), *GmLEC1-2* (Glyma.17G005600), *GmLEC1-3*, (Glyma.03G080700), and *GmLEC1-4* (Glyma.20G000600), with the first two displaying mRNA accumulation patterns most closely related to *Arabidopsis LEC1* (Fig. S5).

GmLEC1-bound genes were identified in ChIP-seq experiments using anti-GmLEC1 antibodies and embryos at the cotyledon [GmCOT, 15 d after pollination (DAP)], early maturation (GmEM, 23 DAP), and mid-maturation (GmMM, 40–45 DAP) stages that correspond to the morphogenesis phase, transition to maturation, and the maturation phase, respectively (Fig. 2*B*). As summarized in Fig. 7*A*, we identified 16,945, 16,657, and 18,749 genes that were bound by GmLEC1 at the GmCOT, GmEM, and GmMM stages, respectively (Dataset S6), and control experiments validated the ChIP-seq results (Dataset S3). We defined genes potentially regulated by GmLEC1 at the three stages using the strategy employed to identify LEC1-coexpressed genes in *Arabidopsis* BCOT-stage seeds and the Harada–Goldberg Soybean Seed Development LCM RNA-Seq Dataset (GEO accessions GSE57606, GSE46096, and GSE99109; https://www.ncbi.nlm.nih.gov/geo) (Fig. 2*C*). Potentially regulated genes numbered 3,337, 2,751, and 3,529 at the GmCOT, GmEM, and GmMM stages, respectively, (Fig. 7*A* and Dataset S6).

We identified 1,699 ($P < 2.2 \times 10^{-146}$), 1,450 ($P < 6.5 \times 10^{-154}$), and 1,983 ($P < 1.5 \times 10^{-180}$) LEC1 target genes that represented a significant overlap between bound and coexpressed genes at the GmCOT, GmEM, and GmMM stages, respectively (Fig. 7*A* and Dataset S6). The GmLEC1 target genes at the three stages exhibited significant overlap with their orthologous LEC1 target genes identified in BCOT-stage *Arabidopsis* seeds. Of the 432 *Arabidopsis* BCOT target genes with annotated soybean homologs, 32% ($P < 2.4 \times 10^{-50}$), 29% ($P < 1.8 \times 10^{-44}$), and 28% ($P < 2.5 \times 10^{-29}$) corresponded with GmLEC1 target genes at the GmCOT, GmEM, and GmMM stages, respectively (Fig. 7*B*). The results suggest that LEC1 plays similar roles in *Arabidopsis* and soybean seed development.

There was significant overlap in the GmLEC1 target genes at the three stages (Fig. 7*C*). Target genes at the GmEM and GmMM stages displayed the greatest overlap (43 and 58% of GmEM- and GmMM-stage target genes, respectively), followed by GmCOT and GmEM stages (41 and 48%, respectively). The largest numbers of stage-specific target genes were observed at the GmCOT and GmMM stages (814 and 945, respectively), suggesting that GmLEC1 regulates transitions in gene-expression programs from early to late seed development.

Hierarchical clustering of GmLEC1 target mRNA levels in embryos at the three stages (Harada Embryo mRNA-Seq Dataset, GEO accession no. GSE99571; https://www.ncbi.nlm.nih.gov/geo) provided additional support that GmLEC1 regulates distinct gene

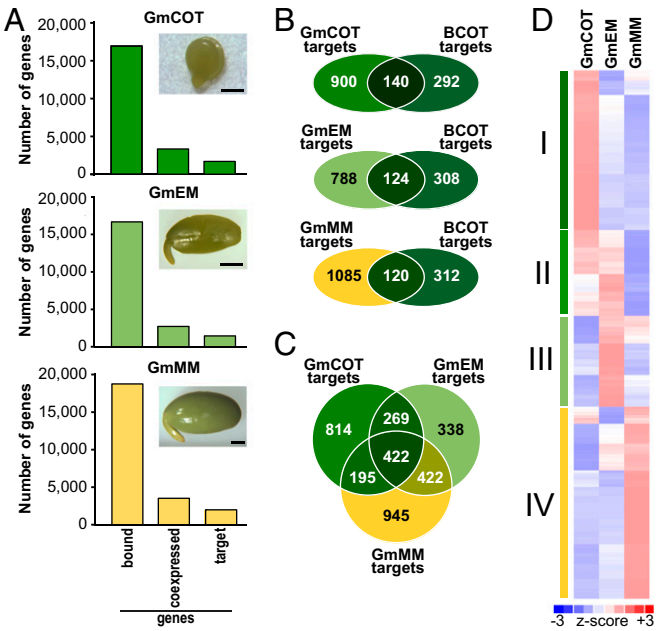

Fig. 7. Soybean LEC1 target genes during seed development. (*A*) Numbers of GmLEC1-bound, -coexpressed, and target genes at the GmCOT, GmEM, and GmMM stages. *Insets* show representative embryos at each stage. (Scale bars: GmCOT, 0.5 mm; GmEM and GmMM, 2 mm.) (*B*) Venn diagrams showing the overlap of the GmLEC1 target gene sets with the most closely related *Arabidopsis* BCOT target genes. (*C*) Venn diagram showing the overlap of the GmLEC1 target genes at the three stages. (*D*) Hierarchical clustering of embryo mRNA levels for the GmLEC1 target genes at the GmCOT, GmEM, and GmMM stages.

sets at different stages. Fig. 7*D* shows that the GmLEC1 target genes clustered into at least four groups. Cluster I mRNAs accumulated at highest levels in GmCOT-stage embryos and were most highly overrepresented for GO terms related to growth and morphogenesis, such as sequence-specific DNA-binding TF activity, nucleosome assembly, polarity specification of adaxial/abaxial axis, and determination of bilateral symmetry (Fig. 3 and Dataset S6). Clusters II and III mRNAs accumulated at highest levels in both the GmCOT and GmEM stages or in the GmEM stage and were primarily overrepresented for PSN GO terms. Cluster IV mRNAs accumulated at highest levels at the GmMM stage and were enriched for GO terms related to maturation such as lipid storage, seed dormancy process, monolayer-surrounded lipid storage body, and nutrient reservoir activity. These results are consistent with the hypothesis generated from the analyses of *Arabidopsis* LEC1 target genes that GmLEC1 regulates different genes involved in distinct cellular processes at different stages of seed development. The results also suggest that LEC1 function is conserved during seed development in *Arabidopsis* and soybean.

DNA Sequence Motifs Associated with Bound Genomic Regions Upstream of LEC1 Target Genes. To obtain clues about the mechanisms that underlie LEC1's ability to regulate transcriptionally distinct gene sets at different developmental stages, we identified overrepresented DNA sequence motifs in bound regions upstream of LEC1 target genes. LEC1 is an atypical NF-YB subunit of the NF-Y TF that binds the CCAAT DNA motif in association with other NF-Y subunits (21), and it also has been shown to interact with NF-YC and bZIP67, a TF that binds G-box–like motifs (13, 14). Fig. 8*A* and Dataset S5, Table S2 show the DNA sequence motifs that were enriched in LEC1-bound genomic regions 1 kb upstream of target genes as identified by de novo motif-discovery analyses. These motifs most closely corresponded with the G-box, ABRE-like, CCAAT, RY, and BPC1 *cis*-regulatory elements that are known to be involved in the control of gene transcription.

The motif discovery analysis was validated by quantifying the occurrence of the DNA motifs in the bound regions upstream of LEC1 target genes. Fig. 8*B* summarizes the *P* values for motif enrichment in the upstream region of *Arabidopsis* and soybean target genes, and Fig. S6 shows the frequencies at which these motifs were detected in upstream regions of target genes compared with comparably sized and spaced regions upstream of randomly selected genes. The G-box–like motifs, G-box (CACGTG) and ABRE-like (C/G/T)ACGTG(G/T)(A/C), were the only DNA sequence motifs that were significantly overrepresented in all LEC1 target gene sets identified in *Arabidopsis* and soybean. The RY motif (CATGCA) that was originally identified in the upstream region of storage protein genes (22) was significantly overrepresented in *Arabidopsis* and soybean LEC1 target genes except for the EARLY ACT targets. The BPC1 sequence motif (A/G)GA(A/G)AG(A/G)(A/G)A was overrepresented in all target gene sets identified in *Arabidopsis* and soybean seeds but not in the EARLY ACT and LATE ACT target gene sets. The CCAAT-binding sequence motif bound by the NF-Y transcription complex was significantly overrepresented only in the LEC1 target genes of *Arabidopsis* BCOT-stage seeds and soybean GmCOT-stage and GmEM-stage embryos. These results suggest that DNA motifs associated with LEC1 function are conserved in *Arabidopsis* and soybean.

We asked if GmLEC1 target gene clusters that were differentially expressed temporally during soybean seed development were enriched for distinct DNA motifs (Fig. 7*D*). Fig. 8*B* and Fig. S7 show that all four GmLEC1 target gene clusters were enriched for the G-box–like motifs, although the enrichment was most significant

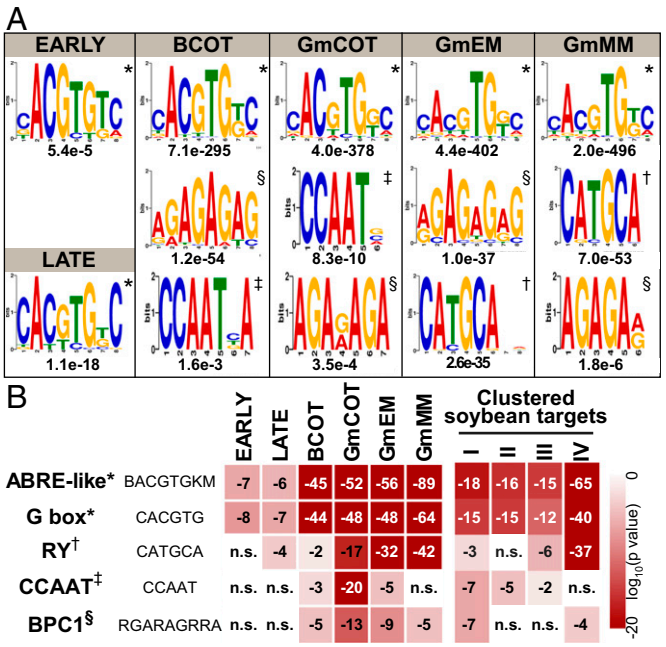

Fig. 8. DNA motifs bound by LEC1. (*A*) Sequence logos showing DNA motifs identified de novo that are enriched in the bound regions of LEC1 target genes at the indicated stages with their associated E values. Only significantly enriched (E values < 0.01) DNA motifs with homology to the known *cis*-regulatory elements in (*B*) are shown. All de novo-identified DNA motifs are shown in Dataset S5, Table S2. (*B*) Enrichment of annotated *cis*-regulatory elements homologous to the enriched de novo-identified DNA motifs in the bound promoter regions of LEC1 target genes. Heatmaps show the *P* value for DNA motif enrichment in LEC1-bound regions relative to randomly selected regions. Motif enrichment frequencies are shown in Fig. S6 and Dataset S5, Table S2.

for genes expressed at the latest stage (cluster IV). GmLEC1 target genes expressed at the earliest stages of seed development (clusters I and II) were enriched for the CCAAT motif, a known binding site of the LEC1 NF-Y complex. By contrast, genes expressed at the latest stage (cluster IV) were most strongly overrepresented for the RY motif. Similar results were obtained for *Arabidopsis* LEC1 BCOT target genes, with those expressed at early (cluster Q) and late (cluster P) stages being enriched for the CCAAT and RY motifs, respectively, and all target gene sets being overrepresented for the G-box–like motifs (Fig. S8). To determine if motif enrichment was associated with developmental function, we measured the frequencies with which motifs were linked with genes involved in (*i*) photosynthesis and chloroplast function (listed in Fig. S4) and (*ii*) maturation (listed in Fig. 4). PSN genes were significantly enriched for G-box–like and CCAAT motifs, whereas the MAT genes were significantly enriched for G-box–like and RY motifs (Fig. S9). Thus, these two functionally defined gene sets were distinguished by their enrichment for the CCAAT and RY motifs. The differential enrichment of DNA sequence motifs of genes expressed at different stages of development and of those involved in distinct physiological functions opens the possibility that LEC1 may operate in combination with different TFs to regulate distinct target gene sets.

Discussion

We profiled mRNA populations in *Arabidopsis lec1*-mutant seeds and identified LEC1 target genes in *Arabidopsis* seedlings ectopically expressing LEC1 and in developing *Arabidopsis* and soybean seeds to identify genes regulated directly by and downstream of LEC1. Our results demonstrate that LEC1 regulates distinct gene sets at different developmental stages, suggesting that LEC1 plays a more extensive role in controlling diverse aspects of seed development than appreciated previously.

LEC1 Transcriptionally Regulates Genes That Control Several Distinct Aspects of Seed Development. Our results confirmed a direct role for LEC1 in controlling the maturation phase of seed development. We showed that (*i*) the great majority of genes differentially expressed in wild-type and *lec1* mutant seeds were detected during the MG and PMG stages that encompass the maturation phase (Fig. 1); (*ii*) *lec1*$^{-}$–down-regulated genes were overrepresented for GO terms related to maturation (Fig. 3 and Dataset S1); (*iii*) target genes directly regulated by LEC1 in BCOT *Arabidopsis* seeds and GmEM and GmMM soybean embryos were overrepresented for maturation GO terms (Fig. 3 and Datasets S2 and S6); and (*iv*) 26 and 25 of 50 MAT genes were *lec1*$^{-}$–down-regulated and BCOT target genes, respectively (Fig. 4). These results are consistent with other reports showing that LEC1 is a master regulator of the maturation phase (15, 16, 23).

Comparison of genes that are directly vs. genetically regulated by LEC1 provides insight into the mechanism by which target gene transcription is controlled during seed maturation. Our finding that only 174 of 554 of BCOT LEC1 target genes were identified as *lec1*$^{-}$–down-regulated, including 13 of 25 MAT target genes, suggests that many target genes are not regulated solely by LEC1 (Figs. 4 and 6). These results implicate the involvement of other TFs in regulating LEC1 target genes. For example, our analyses of the mRNA transcriptomes of *abi3* and *fus3* mutants showed that of the 12 MAT genes that were LEC1 targets but were not *lec1*$^{-}$–down-regulated, eight were *abi3*$^{-}$–down-regulated, and six were *fus3*$^{-}$–down-regulated (Fig. 4). One interpretation of these results is that LEC1 may not be sufficient to activate some of its target genes completely and that other TFs are required to activate these genes fully. This interpretation is consistent with the findings that many maturation genes are regulated combinatorially by LEC1 and other TFs, including ABI3 and FUS3, which are both LEC1 target genes and are *lec1*$^{-}$–down-regulated (24–27). Together, these results are consistent with a model in which LEC1 activates *ABI3* and *FUS3* as well as other target genes (9). ABI3 and/or FUS3 may play major roles in fully activating some of these LEC1 target genes, whereas LEC1 may be predominately responsible for the activation of other target genes. Our results are consistent with the conclusions of other studies showing that LEC1 acts high in the regulatory hierarchy controlling maturation by activating ABI3 and FUS3 and that ABI3 and FUS3 are dominant regulators of many MAT genes (reviewed in refs. 4, 5, 11, and 12).

Several lines of evidence indicate that LEC1 is directly involved in regulating photosynthesis and chloroplast function during seed development. First, 19 of 32 BCOT LEC1 target genes encoding components of photosystem I and II, cyt b6f, and ATP synthase complexes were also *lec1*$^{-}$–down-regulated (Fig. S4 and Dataset S1). Second, BCOT LEC1 target genes and GmCOT and GmEM LEC1 targets were enriched for PSN genes (Figs. S3 and S4). Third, maturing *lec1* mutant embryos are a paler green than wild-type embryos, suggesting that LEC1 is necessary to activate PSN genes fully, although LEC1 must not be absolutely required for their expression, given that *lec1* mutants eventually become green (16). Fourth, our results are consistent with other studies that suggest a link between LEC1 and photosynthesis/chloroplast development. For example, LEC1 interacts with pirin to mediate blue light-induced expression of *LIGHT-HARVESTING CHLOROPHYLL A/B-BINDING PROTEIN* (*LHCB*) genes (28). Others have shown that LEC1 binds *CAB4/LHCA4*, *LHCB5*, and *LHCA1* promoters in seedlings ectopically expressing LEC1, although LEC1 binding was concluded to be involved in downregulating these genes (8).

LEC1's involvement in directly regulating genes required for photosynthesis and chloroplast biogenesis and the maturation phase is consistent with its role as a central regulator of seed development. Functional chloroplasts have been identified in *Arabidopsis* embryos and endosperm and soybean embryos (18, 29, 30), and we and others showed previously that photosynthesis and maturation are activated sequentially during *Arabidopsis* embryo and endosperm development (18, 31). Photosynthetic activity in oilseeds, such as *Arabidopsis* and soybean, serves a primary role in preventing anoxia through the generation of oxygen in internal tissues (29, 32–34) and enhancing carbon conversion efficiency by recycling CO_2 generated from fatty acid biosynthesis (35). Thus, LEC1 promotes photosynthesis and, therefore, fatty acid biosynthesis in oilseeds, the packaging of triacylglycerol into oil bodies, and storage protein accumulation that occurs during the maturation phase. LEC1 was first detected in land plant lineages in the lycophyte *Selaginella moellendorffii* (36–38). We showed previously that *SmLEC1* is expressed in structures that accumulate lipids and speculated that LEC1 may have arisen, in part, in non–seed-bearing land plants to promote fatty acid biosynthesis and storage. The dual role of LEC1 in promoting photosynthetic activity and maturation processes is consistent with this hypothesis.

Analysis of LEC1 target gene clusters suggests that LEC1 regulates several other aspects of seed development. For example, soybean cluster I suggests a role for LEC1 in controlling morphogenesis and cell growth early in seed development (Figs. 3 and 7), whereas *Arabidopsis* clusters O and R, respectively, suggest that LEC1 controls cell division in the EP and other processes in endosperm domains throughout development (Figs. 3 and 6). Together, these results support previous hypotheses about LEC1 function, based on analyses of mutant phenotypes, that LEC1 is a central regulator of seed development (5, 7).

LEC1 Regulates Transitions in Gene-Regulatory Programs During Seed Development. How does LEC1 directly activate different genes at different developmental stages? A potential explanation is that LEC1 may interact with different TFs to activate distinct gene sets, and the availability of these interacting TFs may be temporally regulated. LEC1 is a subunit of the NF-Y complex (21), and studies in animals and plants have shown that NF-Y complexes interact with a number of distinct TFs to regulate target gene transcription synergistically (39, 40; reviewed in ref. 41). Moreover, LEC1 has been

shown to interact with (*i*) NF-YC2 and bZIP67 to activate maturation genes (13, 14), (*ii*) PIF4 to coactivate genes involved in dark-induced hypocotyl elongation (42), (*iii*) TCL2 to activate genes that inhibit trichome formation (43), and (*iv*) pirin, a protein that enhances TF binding in mammals, to regulate *LHCB* genes (28).

We obtained support for this hypothesis by showing that target gene regions bound by LEC1 were enriched for different DNA motifs at different developmental stages. *Arabidopsis* PSN target genes at the BCOT stage were enriched for the CCAAT and G-box–like motifs, whereas MAT target genes were overrepresented for the RY and G-box–like motifs (Fig. S9). Similarly, GmLEC1 target gene clusters that were enriched for genes involved in photosynthesis and chloroplast function and in maturation, respectively, were overrepresented for the CCAAT and G-box–like DNA motifs and the RY and G-box–like motifs (Fig. 8). Differences in motif enrichment may reflect, in part, the binding specificities of the TFs with which LEC1 interacts. For example, LEC1 may interact with NF-YA and NF-YC subunits to form a NF-Y complex that binds a CCAAT motif to regulate PSN genes. This hypothesis is consistent with the reports that NF-Y complexes regulate genes involved in photosynthesis (28; reviewed in refs. 44 and 45). We also suggest that LEC1 is associated with RY motifs during the maturation phase, because it acts in concert with ABI3, an RY-binding TF, at *cis*-regulatory modules (25). LEC1 and ABI3 may interact indirectly through their mutual physical association with bZIP TFs (24, 27, 46). Although G-box–like motifs are enriched in both PSN and MAT target genes, it is unclear if the same or different G-box–like binding TFs work with LEC1 to activate these diverse gene sets. For example, bZIP67 interacts with LEC1 and NF-YC to activate genes involved in maturation, and we showed previously that bZIP67 is not detected until after LEC1 PSN target genes are activated, decreasing the possibility that bZIP67 interacts with these genes (18). Thus, it is possible that another bZIP TF that is expressed earlier in seed development than bZIP67, such as HY5, which regulates genes involved in chloroplast function (47), works with LEC1 to activate these target genes during seed development. Alternatively, it is possible that a basic helix–loop-helix (bHLH) TF that also binds G-box–like motifs interacts with LEC1 to regulate photosynthetic genes. For example, LEC1 was shown to interact with the bHLH TF PIF4 and to bind G-box–like motifs, although this combination of TFs represses genes involved in chloroplast development.

How does LEC1 act mechanistically to regulate different target gene sets during seed development? In animals, NF-Y complexes can act as pioneer TFs that facilitate the binding of other TFs (48). For example, NF-Y binds DNA motifs in nucleosomal DNA and promotes nucleosome repositioning and an open chromatin conformation that stabilizes the binding of colocalized master regulator TFs that govern mouse ES cell identity (49). The possibility that LEC1 serves as a pioneer TF could explain, in part, the observation that LEC1 remains bound with many genes early and late following induction in seedlings even though the corresponding genes are expressed at only one stage (Fig. 5). The influence of NF-Y on chromatin conformation may be mediated, in part, by its known effects on posttranslational histone modifications that are correlated with the activation or repression of gene transcription, both in animals (reviewed in ref. 41) and plants (40). Thus, LEC1 may bind DNA and create an open chromatin conformation that allows other TFs to bind and regulate target genes during seed development.

In conclusion, our study of genes regulated genetically and directly by LEC1 has demonstrated its role in regulating distinct gene sets at different stages of seed development. In addition to confirming LEC1's role in controlling the maturation phase, we revealed a direct role for LEC1 in controlling photosynthesis and chloroplast development and obtained evidence suggesting its involvement in other temporally and spatially regulated developmental processes, such as morphogenesis. Identification of overrepresented DNA motifs in target gene promoters suggests that LEC1 may regulate diverse target gene sets by interacting with different TFs. Moreover, our results provide strong evidence for the conservation of gene-regulatory networks that operate during seed development in two dicotyledonous plants, *Arabidopsis* and soybean, that diverged ~92 Mya. The role of LEC1 in controlling two developmental processes, photosynthesis/chloroplast function and maturation, is conserved in the two species, and there are strong similarities, although not complete identity, in the target genes of *Arabidopsis* and soybean LEC1. We note that similarities and differences are also seen in gene networks that operate in corresponding cell types in humans and mice that also diverged ~92 Mya (50, 51). Conservation of the developmental processes and gene regulatory networks controlled by LEC1 is consistent with the idea that LEC1 is a major regulator of seed development.

Materials and Methods

Plant Materials. *Arabidopsis* and soybean plants were grown as described in *SI Materials and Methods*.

35S:LEC1-GR and *35S:FLAG-LEC1-GR* were constructed using methods similar to those described in ref. 52; the details are provided in *SI Materials and Methods*. *LEC1:LEC1-GFP:LEC1* was created by using PCR to add a C-terminal $(Gly)_6$ linker to the *LEC1* cDNA followed by cloning in frame with *sGFP* (S65T) (53) and transferring the construct into the LEC1 expression cassette (54). Constructs were transferred into *Arabidopsis* Ws-0 and *lec1-1*–mutant plants as described (54).

lec1-1–mutant seeds were staged as described previously (17). Early LEC1-induction experiments with homozygous *35S:LEC1-GR* or *35S:FLAG-LEC1-GR* transgenic plants were performed as described (52). Shoot apices obtained by removing cotyledons and hypocotyls and whole seedlings were harvested. For the late-induction experiments, *35S:FLAG-LEC1-GR* seedlings were grown for 8 d on 30 μM dexamethasone (Dex). Embryos harvested from soybean GmCOT, GmEM, and GmMM seeds were staged as described (55).

RNA Analysis. Affymetrix *Arabidopsis* ATH1 GeneChips hybridization experiments were done as described (17). Laser-capture microdissection (LCM) experiments were performed as described (18).

ChIP. Antibodies used for the ChIP experiments are listed in Fig. 2 and described in *SI Materials and Methods*. ChIP assays were performed as described (56), with the modifications detailed in *SI Materials and Methods*. ChIP and input DNAs for ChIP-chip experiments were quantified and prepared as described (57) with modifications listed in *SI Materials and Methods* and were hybridized to the *Arabidopsis* GeneChip Tiling 1.0R Array. ChIP-seq libraries were prepared using the NuGEN Ovation Ultralow DR Multiplex System. Libraries were size-selected by electrophoresis, purified, and sequenced at 50-bp single-end reads using an Illumina HiSeq 2000 sequencing system. qPCR validation experiments were done in triplicate, with either 30 pg of unamplified chromatin or 1 ng of amplified DNA. Primers are listed in Dataset S5, Table S3.

Data Analysis. The mRNA profiling data were analyzed as described in refs. 18, 58, and 59 and as detailed in *SI Materials and Methods*. Methods used for hierarchical clustering (60) and GO term enrichment (18, 61) are described in *SI Materials and Methods*. ChIP-chip data were normalized using model-based analysis of tiling array (62), and significantly bound regions were identified using the CisGenome (v1.2) hidden Markov model (HMM) algorithm [posterior probability threshold 0.99999 (63)]. ChIP-seq data were analyzed using Bowtie v0.12.7 (64) and the PeakSeq algorithm of CisGenome (v2.0) as described in *SI Materials and Methods*. DNA sequence motifs were identified de novo using the MEME-ChIP suite (65) as described in *SI Materials and Methods*. Data are available at GEO under the following accessions: GSE1051 (*lec1-1*–mutant seed development), GSE99528 (LEC1-GR induction RNA series), GSE99529 (LEC1-GR ChIP-chip), GSE99587 (*Arabidopsis* BCOT ChIP-seq), and GSE99882 (soybean GmLEC1 ChIP-seq).

ACKNOWLEDGMENTS. We thank Dr. Jon Nield for allowing us to use the diagram in Fig. S3; Jiong Fei, Linda Kwong, Anhthu Bui, Min Chen, Alec Olson, and Mac Harada for technical assistance; and Siobhan Braybrook, Ryan Kirkbride, and Mark Belmonte for useful discussions. This work was supported by National Science Foundation grants (to J.J.H. and R.B.G.) and by Department of Energy grants (to J.J.H.).

1. Lau S, Slane D, Herud O, Kong J, Jürgens G (2012) Early embryogenesis in flowering plants: Setting up the basic body pattern. *Annu Rev Plant Biol* 63:483–506.
2. Li J, Berger F (2012) Endosperm: Food for humankind and fodder for scientific discoveries. *New Phytol* 195:290–305.
3. Puthur JT, Shackira AM, Saradhi PP, Bartels D (2013) Chloroembryos: A unique photosynthesis system. *J Plant Physiol* 170:1131–1138.
4. Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. *Trends Plant Sci* 13:624–630.
5. Santos-Mendoza M, et al. (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant J* 54:608–620.
6. Lotan T, et al. (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93:1195–1205.
7. Harada JJ (2001) Role of Arabidopsis LEAFY COTYLEDON genes in seed development. *J Plant Physiol* 158:405–409.
8. Junker A, et al. (2012) Elongation-related functions of LEAFY COTYLEDON1 during the development of *Arabidopsis thaliana*. *Plant J* 71:427–442.
9. Kagaya Y, et al. (2005) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol* 46:399–406.
10. Mu J, et al. (2008) LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. *Plant Physiol* 148:1042–1054.
11. Suzuki M, McCarty DR (2008) Functional symmetry of the B3 network controlling seed development. *Curr Opin Plant Biol* 11:548–553.
12. Junker A, Hartmann A, Schreiber F, Bäumlein H (2010) An engineer's view on regulation of seed development. *Trends Plant Sci* 15:303–307.
13. Mendes A, et al. (2013) bZIP67 regulates the omega-3 fatty acid content of Arabidopsis seed oil by activating fatty acid desaturase3. *Plant Cell* 25:3104–3116.
14. 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: 843–856.
15. Meinke DW, Franzmann LH, Nickle TC, Yeung EC (1994) Leafy cotyledon mutants of Arabidopsis. *Plant Cell* 6:1049–1064.
16. West MAL, et al. (1994) LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. *Plant Cell* 6:1731–1745.
17. 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* 107:8063–8070.
18. Belmonte MF, et al. (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci USA* 110: E435–E444.
19. Yamamoto A, et al. (2014) Cell-by-cell developmental transition from embryo to post-germination phase revealed by heterochronic gene expression and ER-body formation in Arabidopsis leafy cotyledon mutants. *Plant Cell Physiol* 55:2112–2125.
20. Farnham PJ (2009) Insights from genomic profiling of transcription factors. *Nat Rev Genet* 10:605–616.
21. Calvenzani V, et al. (2012) Interactions and CCAAT-binding of *Arabidopsis thaliana* NF-Y subunits. *PLoS One* 7:e42902.
22. Dickinson CD, Evans RP, Nielsen NC (1988) RY repeats are conserved in the 5′-flanking regions of legume seed-protein genes. *Nucleic Acids Res* 16:371.
23. Meinke DW (1992) A homoeotic mutant of *Arabidopsis thaliana* with leafy cotyledons. *Science* 258:1647–1650.
24. Alonso R, et al. (2009) A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. *Plant Cell* 21:1747–1761.
25. Baud S, et al. (2016) Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by master transcriptional regulators in Arabidopsis seed. *Plant Physiol* 171:1099–1112.
26. Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in Arabidopsis. *Development* 130:6065–6073.
27. Lara P, et al. (2003) Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. *J Biol Chem* 278: 21003–21011.
28. Warpeha KM, et al. (2007) The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in Arabidopsis. *Plant Physiol* 143:1590–1600.
29. Allorent G, et al. (2015) Adjustments of embryonic photosynthetic activity modulate seed fitness in *Arabidopsis thaliana*. *New Phytol* 205:707–719.
30. Saito GY, Chang YC, Walling LL, Thomson WW (1989) A correlation in plastid development and cytoplasmic ultrastructure with nuclear gene-expression during seed ripening in soybean. *New Phytol* 113:459–469.
31. Willmann MR, Mehalick AJ, Packer RL, Jenik PD (2011) MicroRNAs regulate the timing of embryo maturation in Arabidopsis. *Plant Physiol* 155:1871–1884.
32. Rolletschek H, Borisjuk L, Koschorreck M, Wobus U, Weber H (2002) Legume embryos develop in a hypoxic environment. *J Exp Bot* 53:1099–1107.
33. Rolletschek H, et al. (2005) Evidence of a key role for photosynthetic oxygen release in oil storage in developing soybean seeds. *New Phytol* 167:777–786.
34. Vigeolas H, van Dongen JT, Waldeck P, Huhn D, Geigenberger P (2003) Lipid storage metabolism is limited by the prevailing low oxygen concentrations within developing seeds of oilseed rape. *Plant Physiol* 133:2048–2060.
35. Allen DK, Ohlrogge JB, Shachar-Hill Y (2009) The role of light in soybean seed filling metabolism. *Plant J* 58:220–234.
36. Cagliari A, et al. (2014) New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants. *Genomics* 103:380–387.
37. Kirkbride RC, Fischer RL, Harada JJ (2013) LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of *Selaginella moellendorffii*. *PLoS One* 8:e67971.
38. Xie Z, et al. (2008) Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. *Mol Biol Evol* 25:1581–1592.
39. Liu JX, Howell SH (2010) bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in Arabidopsis. *Plant Cell* 22:782–796.
40. Hou X, et al. (2014) Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. *Nat Commun* 5:4601.
41. Dolfini D, Gatta R, Mantovani R (2012) NF-Y and the transcriptional activation of CCAAT promoters. *Crit Rev Biochem Mol Biol* 47:29–49.
42. Huang M, Hu Y, Liu X, Li Y, Hou X (2015) Arabidopsis LEAFY COTYLEDON1 mediates postembryonic development via interacting with PHYTOCHROME-INTERACTING FACTOR4. *Plant Cell* 27:3099–3111.
43. Huang M, Hu Y, Liu X, Li Y, Hou X (2015) Arabidopsis LEAFY COTYLEDON1 controls cell fate determination during post-embryonic development. *Front Plant Sci* 6:955.
44. Petroni K, et al. (2012) The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. *Plant Cell* 24:4777–4792.
45. Laloum T, De Mita S, Gamas P, Baudin M, Niebel A (2013) CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci* 18:157–166, and erratum (2013) 18:594–595.
46. Nakamura S, Lynch TJ, Finkelstein RR (2001) Physical interactions between ABA response loci of Arabidopsis. *Plant J* 26:627–635.
47. Lee J, et al. (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* 19:731–749.
48. Vernimmen D, Bickmore WA (2015) The hierarchy of transcriptional activation: From enhancer to promoter. *Trends Genet* 31:696–708.
49. Oldfield AJ, et al. (2014) Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. *Mol Cell* 55:708–722.
50. Cheng Y, et al.; Mouse ENCODE Consortium (2014) Principles of regulatory information conservation between mouse and human. *Nature* 515:371–375.
51. Stergachis AB, et al. (2014) Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 515:365–370.
52. Braybrook SA, et al. (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc Natl Acad Sci USA* 103:3468–3473.
53. Cava F, et al. (2008) Expression and use of superfolder green fluorescent protein at high temperatures in vivo: A tool to study extreme thermophile biology. *Environ Microbiol* 10:605–613.
54. Kwong RW, et al. (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* 15:5–18.
55. Goldberg RB, Hoschek G, Tam SH, Ditta GS, Breidenbach RW (1981) Abundance, diversity, and regulation of mRNA sequence sets in soybean embryogenesis. *Dev Biol* 83:201–217.
56. Gendrel AV, Lippman Z, Martienssen R, Colot V (2005) Profiling histone modification patterns in plants using genomic tiling microarrays. *Nat Methods* 2:213–218.
57. O'Geen H, Nicolet CM, Blahnik K, Green R, Farnham PJ (2006) Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques* 41:577–580.
58. Ritchie ME, et al. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47.
59. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
60. Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc Natl Acad Sci USA* 98:31–36.
61. Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol* 11:R14.
62. Johnson WE, et al. (2006) Model-based analysis of tiling-arrays for ChIP-chip. *Proc Natl Acad Sci USA* 103:12457–12462.
63. Ji H, et al. (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* 26:1293–1300.
64. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
65. Machanick P, Bailey TL (2011) MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* 27:1696–1697.
66. Gleave AP (1992) A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Mol Biol* 20:1203–1207.
67. Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 12:1360–1367.
68. Dahl JA, Collas P (2009) MicroChIP: Chromatin immunoprecipitation for small cell numbers. *Methods Mol Biol* 567:59–74.
69. Ji H (2010) Computational analysis of ChIP-seq data. *Methods Mol Biol* 674:143–159.
70. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
71. Kharchenko PV, Tolstorukov MY, Park PJ (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26:1351–1359.
72. Li Q, Brown JB, Huang H, Bickel PJ (2011) Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* 5:1752–1779.
73. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137.
74. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
75. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589.