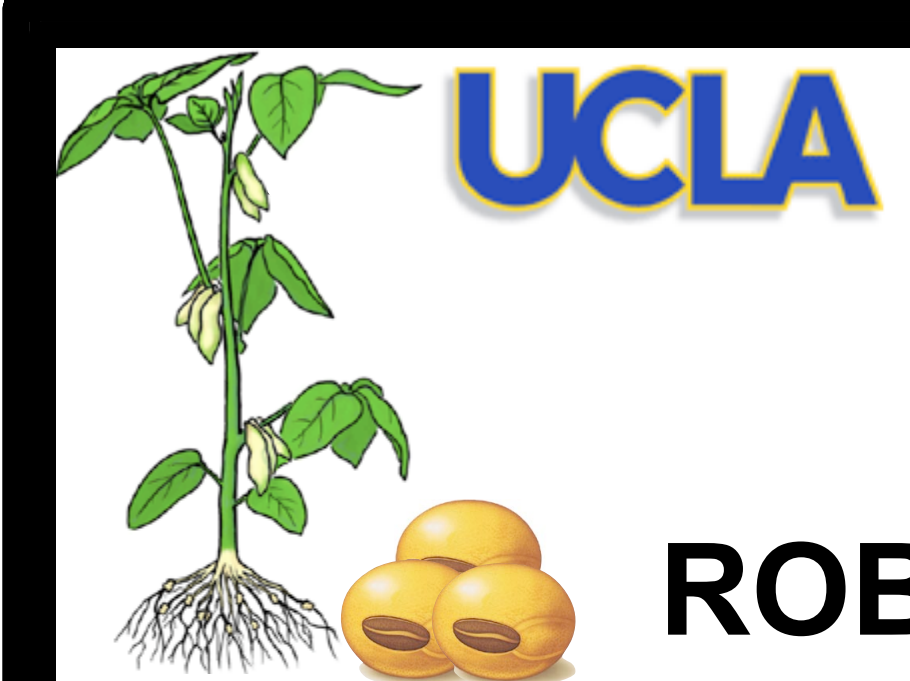


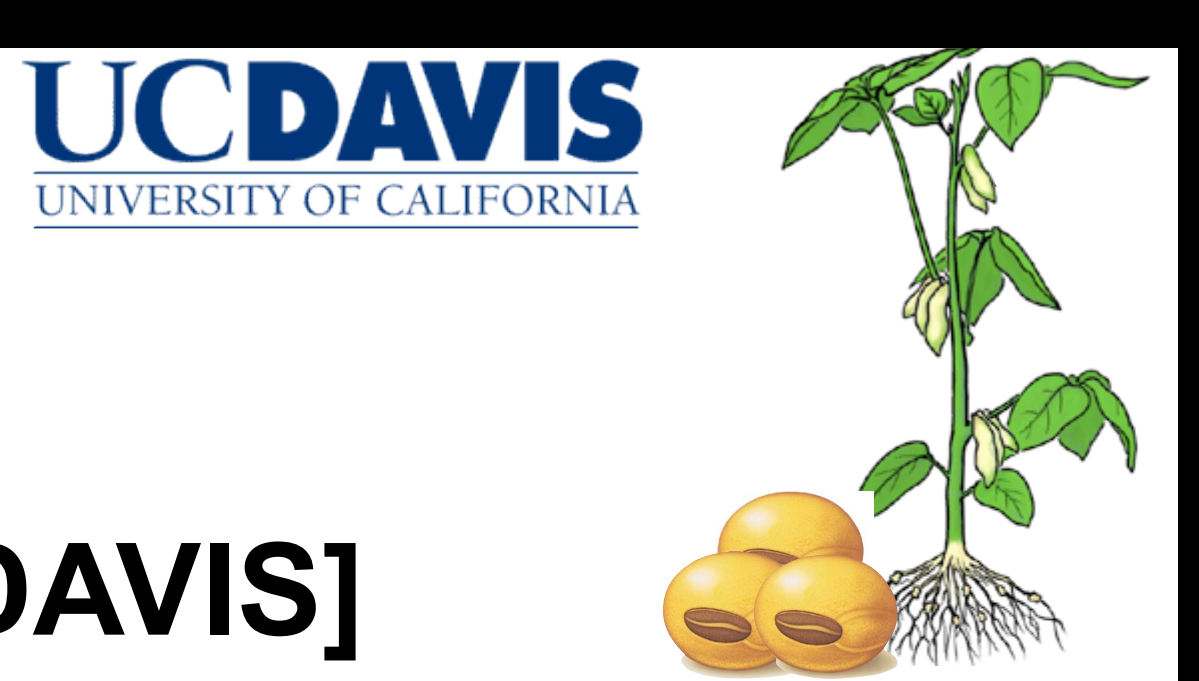


# WHAT ARE THE GENE REGULATORY PROCESSES REQUIRED TO MAKE A SOYBEAN SEED?

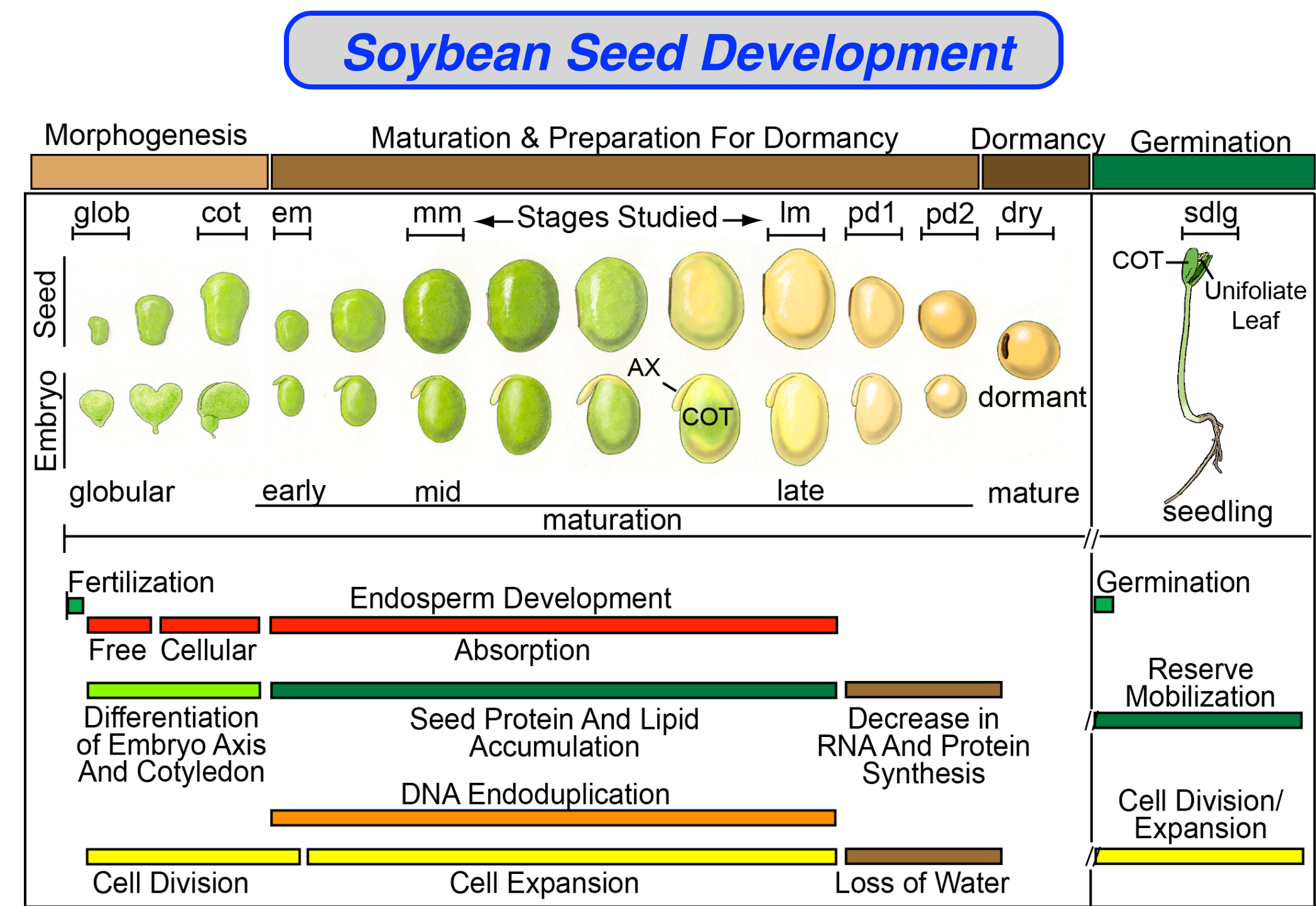
ROBERT B. GOLDBERG (PI) [UCLA]

MATTEO PELLEGRINI (CO-PI) [UCLA]

JOHN J. HARADA (CO-PI) [UC DAVIS]



## INTRODUCTION



**Figure 1. Major Events in Soybean Seed Development.** glob, hrt, cot, em, mm, lm, pd, and dry refer to globular, heart, cotyledon, early maturation, late maturation, pre-dormancy, and dormant seed stages, respectively. COT and sdg refer to cotyledon and seedling, respectively.

### Major Questions Being Addressed In Soybean Seed Development

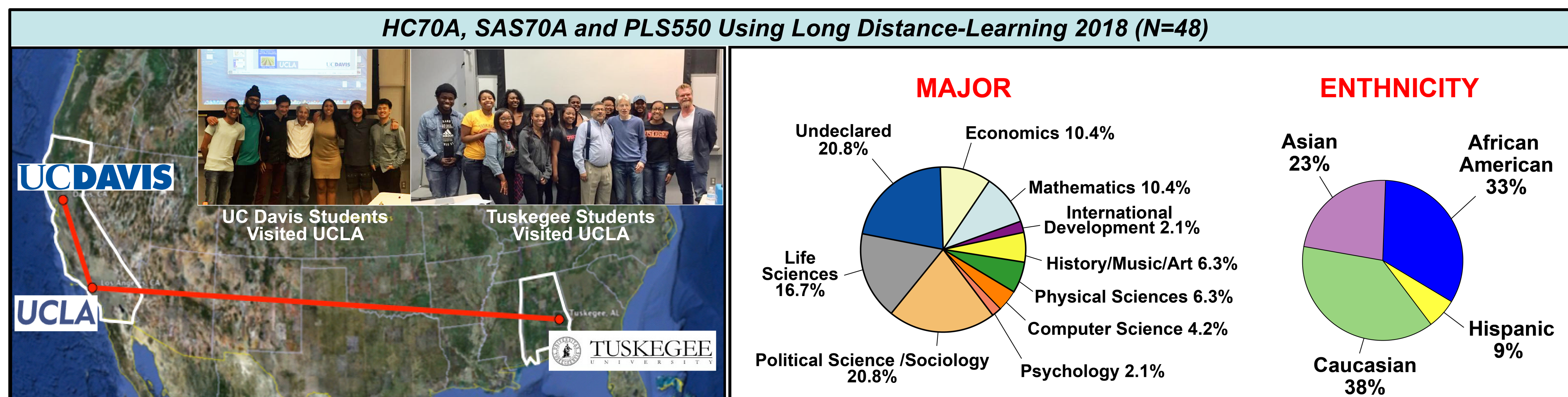
- What are the targets of seed region- and subregion-specific transcription factors?
- What processes do seed region- and subregion-specific transcription factors regulate?
- What cis-regulatory motifs do seed region- and subregion-specific transcription factors interact with?
- Are seed region- and subregion-specific transcription factor targets and interacting cis-regulatory modules conserved between different plant species?
- What are the gene networks required to program seed differentiation and maturation?

## RESEARCH PLAN & STRATEGY

- Identification of Soybean Seed Subregion-Specific Transcription Factor Targets and Control DNA Motifs.** ChIP-Seq experiments will be used to identify targets of transcription factors that are specific for all major seed regions and subregions throughout development.
- Characterization of Soybean Seed Subregion Regulatory Genome.** ATAC-Seq experiments will be used to identify transcription factor-binding domains in the chromatin of seed subregions throughout development.
- Determining the DNA Binding Motifs for Soybean Seed Subregion-Specific Transcription Factors in vitro.** SELEX experiments will be carried out to define DNA sequences bound by subregion-specific transcription factors.
- Functional Validation of Target Genes Regulated by Subregion-Specific Transcription Factors.** The activity of endogenous transcription factors will be down-regulated in seed protoplasts to identify genes regulated directly by the subregion-specific transcription factors.

## EDUCATION AND OUTREACH

### Broader Impacts - Using Long-Distance Learning To Teach Students Simultaneously At UCLA, UC Davis, And Tuskegee University



- "Genetic Engineering in Medicine, Agriculture, and the Law" was taught using a novel long-distance teaching approach that linked ~183 UCLA (HC70A), UC Davis (SAS70A) and Tuskegee (PLS550) students over the high speed internet-based video conferencing system into one "inter-campus" classroom. During the quarter UC Davis and Tuskegee students visited UCLA allowing in-person interactions with their UCLA counterparts.
- A lab class, *Gene Discovery Laboratory (HC70AL)* was taught to entering life science and non-science students to teach them about the "excitement of discovery."

## PROGRESS

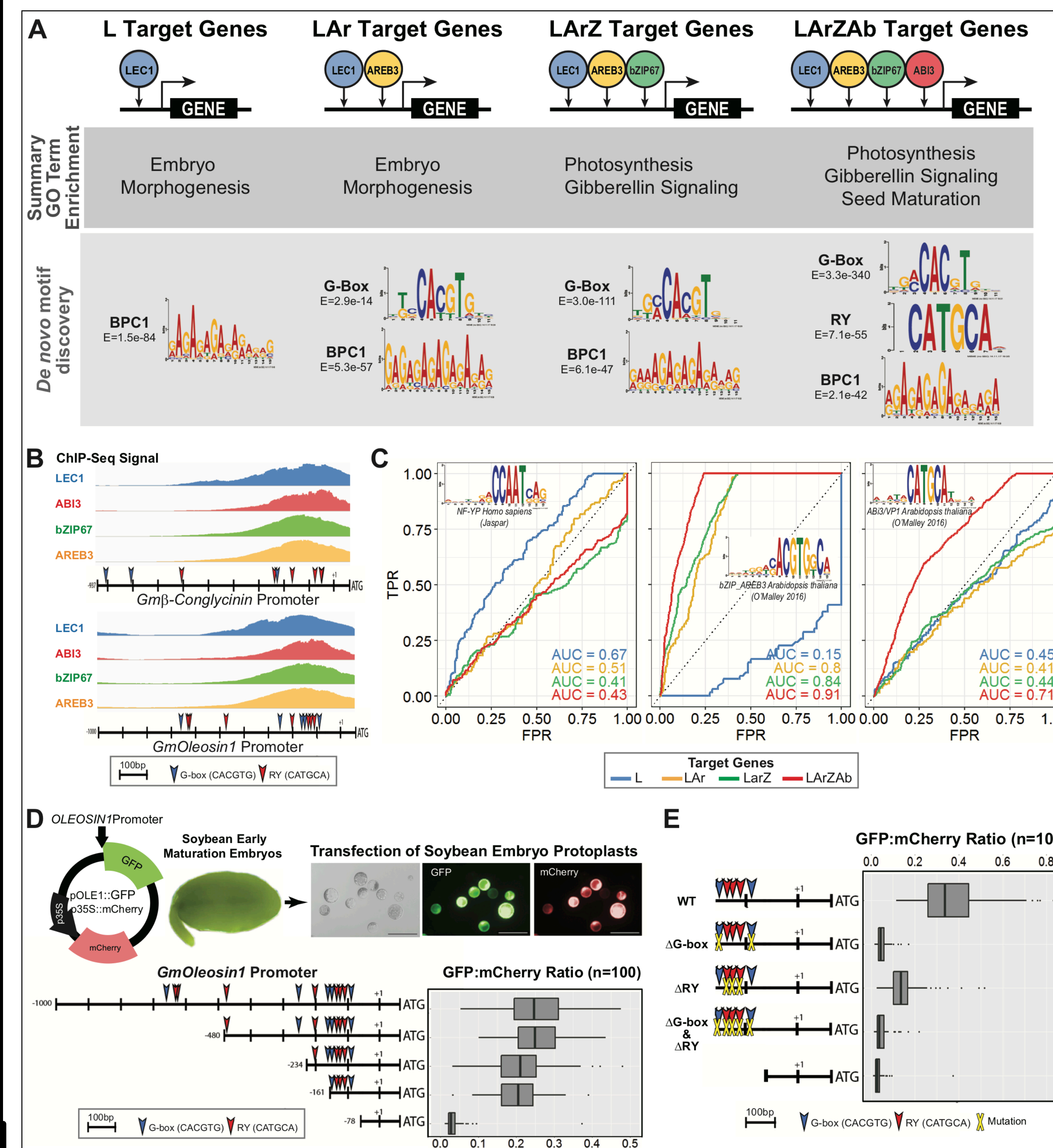
- Four soybean transcription factor genes – *LEC1*, *AREB3*, *bZIP67*, and *ABI3* have been shown to play important roles in controlling the maturation phase in Arabidopsis seeds. The different combinations of four transcription factor genes are recruited to their direct target genes by the presence of cis-regulatory modules with distinct sets of DNA motifs in their upstream regions. Both G-box and RY motifs are essential for the activity of a maturation gene that is regulated by all four transcription factor genes.
- Bimolecular fluorescence complementation (BiFC) assays has been performed to confirm physical interactions between *LEC1*, *AREB3*, *bZIP67*, and *ABI3*, which explain, in part, their concerted binding with cis regulatory modules.
- Target genes of *LEC1*, *bZIP67*, *ABI3*, and *AREB3* are being identified to expand the soybean seed maturation gene networks.
- Hypomethylated regions of the soybean genome, or DNA methylation valleys (DMVs), are highly enriched in transcription factor genes in seeds. Many DMV genes are regulated with respect to seed stage, region, and tissue – and contain H3K4me3, H3K27me3, or bivalent (H3K4me3 and H3K27me3) marks that fluctuate during development. Our results indicate that DMVs are a unique regulatory feature of both plant and animal genomes, and that a large number of seed genes are regulated in the absence of methylation changes during development – probably by the action of specific TFs and epigenetic events at the chromatin level.

## PUBLICATIONS

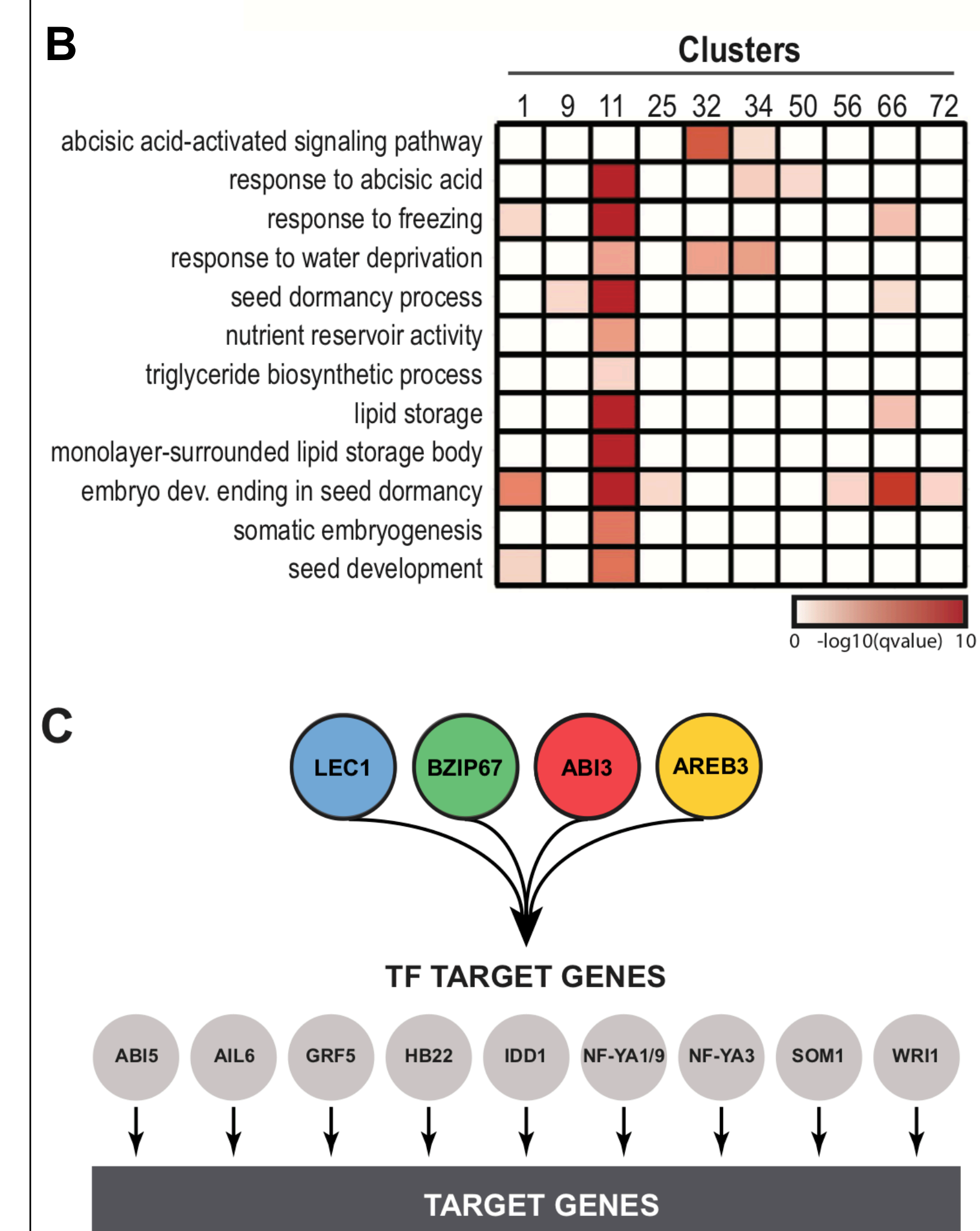
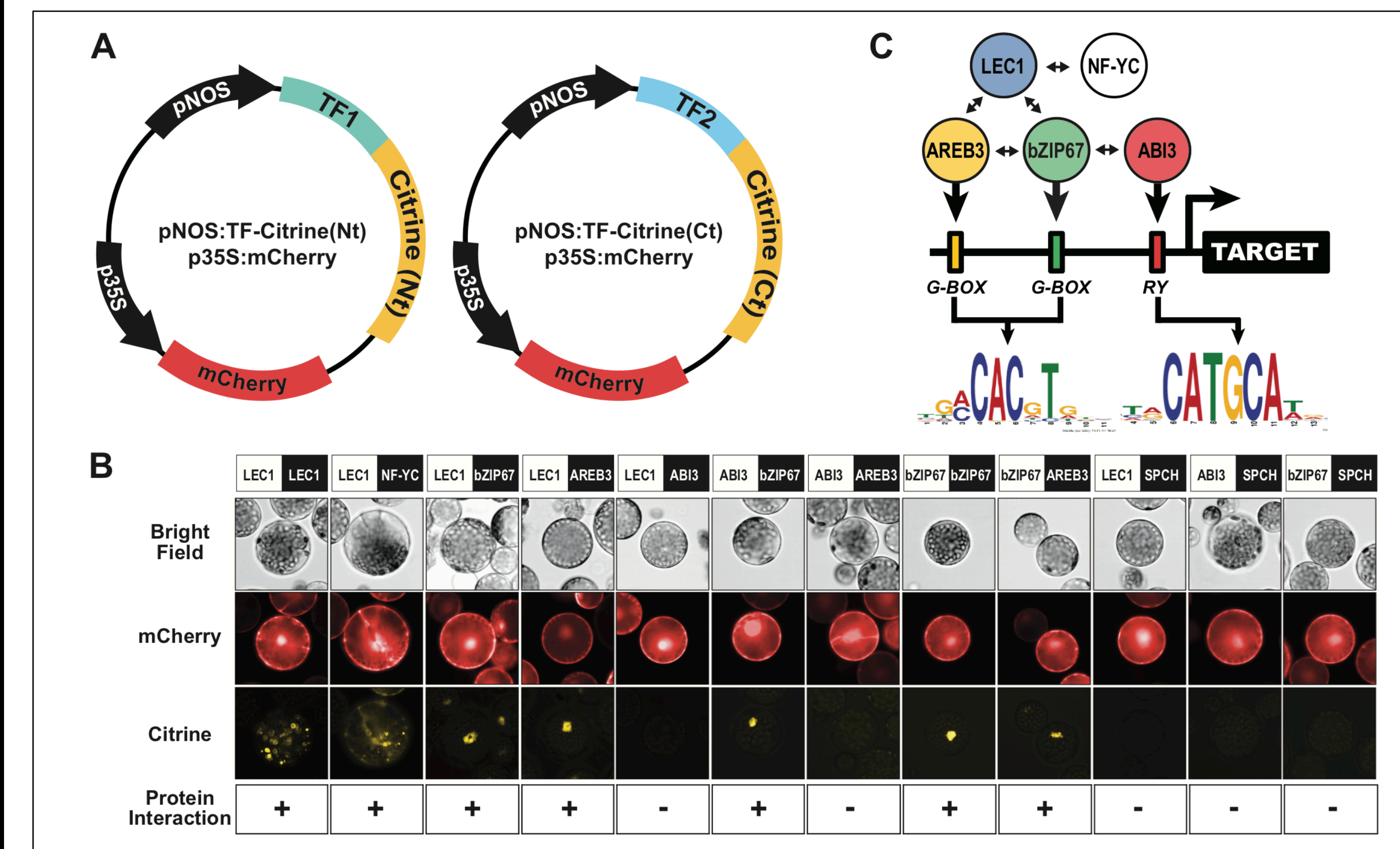
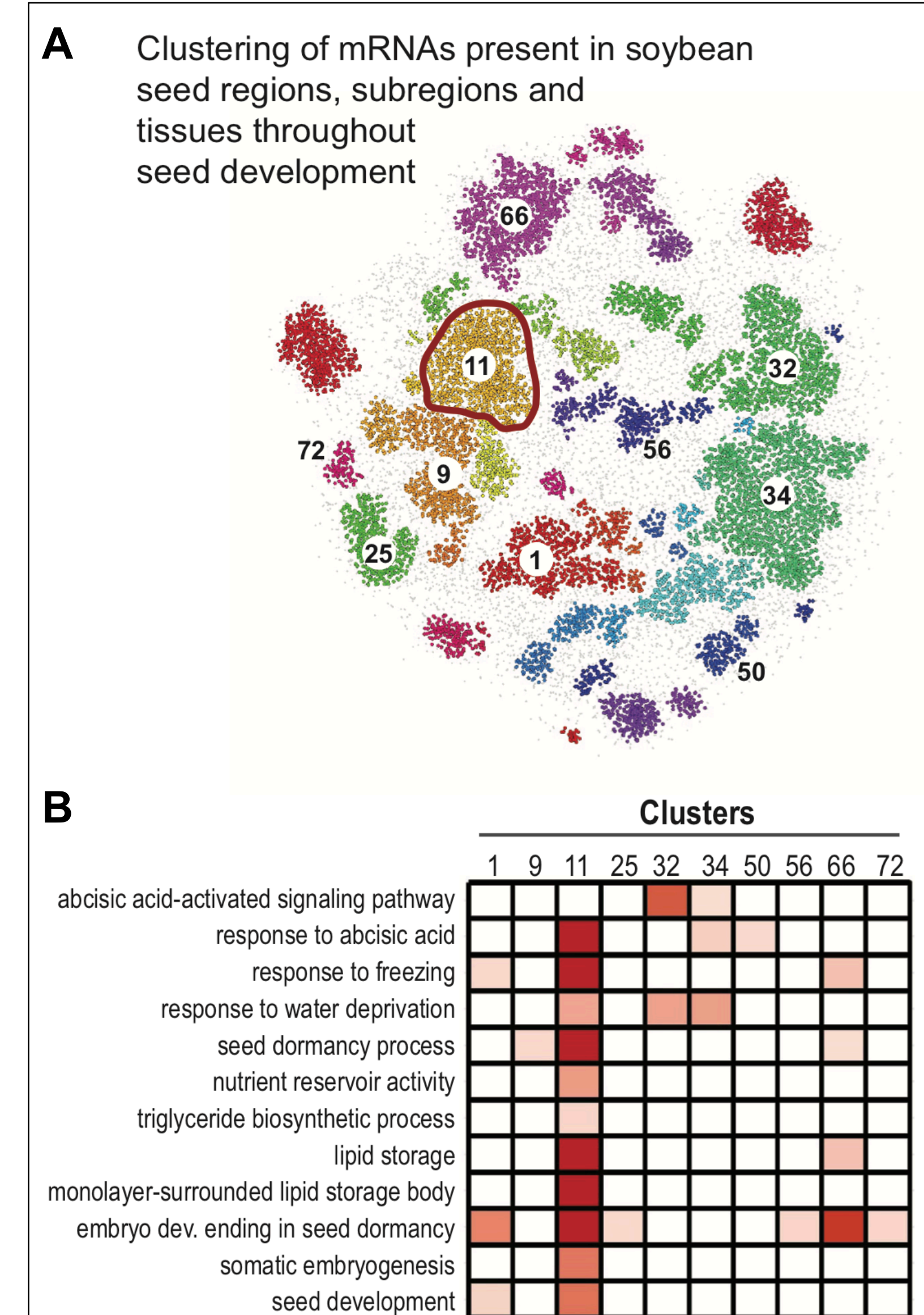
- Pelletier, J.M., Kwong, R. Park, S., Le, B.H., Baden, R., Cagliari, A., Hashimoto, M., Munoz, M., Fischer, R.L., Goldberg, R.B., Harada J.J. (2017) *LEC1* sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. Proc. Natl. Acad. Sci. USA, 114(32): E6710–E6719.
- Lin, J.Y., Le, B.H., Chen, M., Henry, K.F., Hur, J., Hsieh, T.F., Pellegrini, M., Fischer, R.L., Harada, J.J., Goldberg, R.B. (2017) Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. Proc. Natl. Acad. Sci. USA, 114(45): E9730–E9739.
- Henry, K.F., Bui, A.Q., Kawashima, T., Goldberg, R.B. (2018) A shared cis-regulatory module activates transcription in the suspensor of plant embryos. Proc. Natl. Acad. Sci. USA, 115(25): E5824–E5833.
- Chen, M., Lin, J.Y., Hur, J., Pelletier, J.M., Baden, R., Pellegrini, M., Harada, J., Goldberg, R.B. (2018) Seed genome hypomethylated regions are enriched in transcription factor genes. Proc. Natl. Acad. Sci. USA, 115(35): E8315–E8322.

## RESULTS

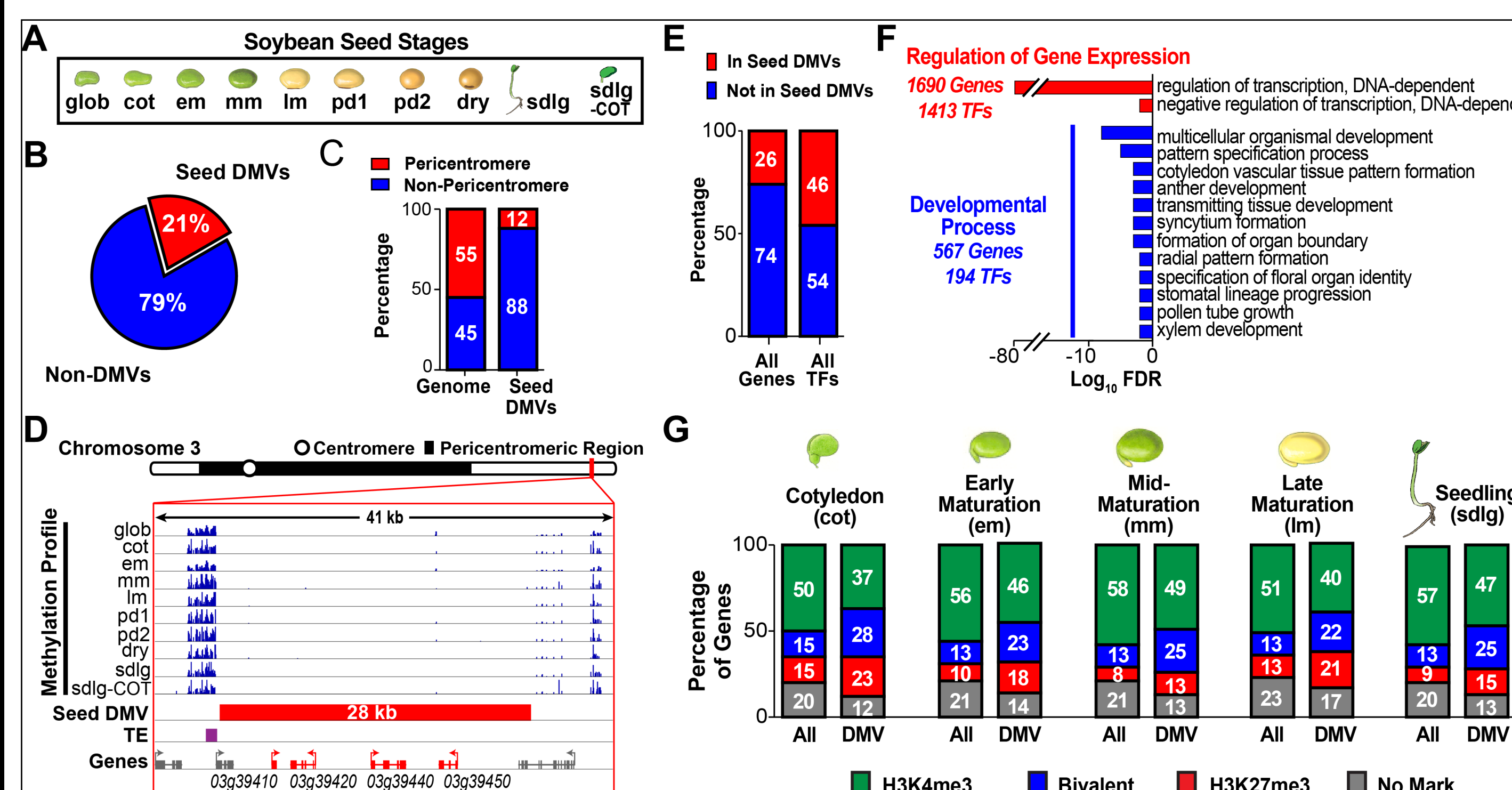
### Characterizing Regulatory Network That Operates During the Maturation Stage of Seed Development



**Figure 2. Cis-regulatory modules that operate in seed development networks.** (A) Representation of target genes directly regulated by different combinations of maturation TFs, Gene Ontology categories for which the target genes are overrepresented, and DNA motifs identified using de novo discovery analyses that are enriched in cis-regulatory modules bound by the TFs. (B) Cis-regulatory modules containing enriched DNA motifs in regions bound by all four transcription factors in the  $\beta$ -CONGLYCININ and OLE1 promoters. (C) Enrichment significance of the indicated DNA motifs in regions bound by the TFs in target genes. (D) Transient assays with a 5' deletion series of the OLE1 promoter in early maturation embryo cotyledon protoplasts show that the cis-regulatory module between -161 and -76 is necessary for promoter activity in embryo cotyledons. (E) Site-directed mutations in G-box and RY motifs in the OLE1 cis-regulatory module abolishes its ability to promote transcription.



**Figure 3. Four maturation TFs interact physically.** (A) Plasmids used for the BiFC assays to test for interactions between the TFs. (B) Interactions between TFs. mCherry signal indicates that the protoplast has been transfected, whereas the citrine signal indicates physical interactions between the two TFs. *SPEECHLESS* (*SPCH*) was used as a control TF that does not interact with any of other the four TFs. (C) Diagrammatic representation of the interactions between TFs indicated by the BiFC assays.



**Figure 5. DNA Methylation Valleys (DMVs) in soybean.** (A) Seed and post-germination methylomes used to identify DMVs (8). (B) Percentage of seed DMVs in the soybean genome. (C) Percentages of seed DMVs in chromosomal pericentromeric and non-pericentromeric regions. (D) Genome browser view of a 28 kb DMV located on chromosome 3. Genes in red color are located within this DMV, including 1 kb of 5' and 3' flanking regions. Genes in gray color are either located partially or outside this DMV region. (E) Proportion of seed DMV genes and transcription factor (TF) genes in the soybean genome. (F) Enriched GO terms with False Discovery Rate (FDR) < 0.05. (G) Percentage of DMV genes and all genes in the genome marked with H3K4me3, H3K27me3, H3K4me3 and H3K27me3 (bivalent mark), or no mark at each developmental stage

**Figure 4. Expanding the maturation gene network.** (A) Coexpressed gene clusters identified by clustering analysis of mRNAs present in specific seed subregions and tissues throughout development. (B) Overrepresented Gene Ontology terms for the indicated gene clusters. Cluster 11 was enriched for GO terms related to seed maturation processes. (C) Nine transcription factors that are direct targets of *LEC1*, *bZIP67*, *ABI3* and *AREB3* and members of cluster 11 were selected for further analyses in chromatin immunoprecipitation experiments. Antibodies were generated against two peptides specific for each TF, and their relative reactivities are shown.