Genome-Wide Profiling of Soybean WRINKLED1 Transcription Factor Binding Sites Provides Insight into the Regulation of Fatty Acid and Triacylglycerol Biosynthesis Program in Seeds

Authors:

Leonardo Jo[1†*], Julie Pelletier[†], John J Harada[*]

Affiliations:

Department of Plant Biology, College of Biological Sciences, University of California, Davis, CA 95616

Footnotes:

[1] Current address, Experimental and Computational Plant Development, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

† These authors contributed equally to this work

* To whom correspondences may be addressed. Email: jjharada@ucdavis.edu and l.jo@uu.nl

ABSTRACT

Understanding the regulatory mechanisms controlling seed lipid accumulation provides opportunities for the development of new crops with enhanced oil levels. The WRINKLED1 transcription factor (WRI1 TF) is a central regulator of lipid biosynthesis in many plant species. We characterized the genome-wide binding profile of soybean (Gm)WRI1 and showed that the TF directly regulates many genes required for the fatty acid and triacylglycerol biosynthesis pathway. We showed that GmWRI1 binds primarily to regions downstream of target gene transcription start sites. Regions bound by GmWRI1 were enriched for its canonical DNA binding element, the AW Box (CNTNGNNNNNNNCG), and another DNA element, the CNC Box (CNCCNCC), that shows some similarity to other characterized DNA motifs. Functional assays showed that both DNA elements are required to mediate transcriptional activation by GmWRI1. Comparison of genes targeted directly by GmWRI1 and by GmLEC1, a central regulator of the maturation phase of seed development, revealed that the two TFs act in a positive feedback subcircuit to control fatty acid and triacylglycerol biosynthesis. Our results provide new insights into the genetic circuitry by which GmWRI1 regulates fatty acid and triacylglycerol biosynthetic genes, and provides evidence that GmWRI1 contributes to establishing the regulatory state of the soybean embryo to promote oil accumulation.

Introduction

Seeds are the major source of plant oils that are used primarily for food-based applications but also have broad uses in industrial applications, including biofuels (1). Due to its growing importance, it is expected that the demand for plant oils will double in the following decades (2). These demands can be met by developing crops with increased seed oil content. To achieve this goal, it is crucial that we understand the regulatory mechanisms that underlie seed fatty acid and triacylglycerol biosynthesis and accumulation. Understanding how seed lipid accumulation is regulated can contribute to the development of biotechnological tools that improve the quality and quantity of these important metabolites.

In seeds, fatty acids (FAs) are mainly stored as triacylglycerol (TAG) in oil bodies, and the FA metabolic pathway in seeds has been extensively studied in many crop species (1, 3). The biosynthesis of FAs initiates in plastids with the synthesis of malonyl-CoA by the enzymatic complex acetyl-CoA carboxylase (ACCase, Fig. 1A). Biotin carboxyl carrier protein2 (BCCP2) and acetyl CoA carboxylase biotin carboxylase subunit (CAC2) are important protein subunits present in the ACCase complex. Malonyl-CoA is transferred to an acyl-carrier protein (ACP) to form malonyl-ACP, which is the main carbon donor for all subsequent cycles of FA elongation. The first four-carbon long acyl-ACP is formed by the condensation reaction between a malonyl-ACP and an acetyl-CoA catalyzed by the 3-ketoacyl-acyl carrier protein synthase III (KASIII) enzyme. Cyclic additions of two carbon units from malonyl-ACP donors into the acyl-ACP is catalyzed by beta-ketoacyl-ACP synthase I (KASI) to generate 16 carbon long acyl-ACPs. Further elongation can be achieved by KASII to catalyze an extra condensation reaction to form an 18-carbon long acyl-ACP. The release of FAs from ACPs is catalyzed by acyl-ACP thioesterases (FATA and FATB) in the plastids. In addition to the FA biosynthetic enzymes, several other enzymes responsible for the elongation and desaturation of FAs account for the great diversity of FAs and TAGs found in the seeds. The activity of these enzymes in the plastids contribute to the increase in the acyl-CoA pool that is used for the *de novo* formation of membrane lipids and storage TAGs. The acyl-CoA produced in the plastids, along with the glycerol-3-phosphate (G3P) are the main substrates for TAG assembly in the endoplasmic reticulum. The acylation of G3P in the endoplasmic reticulum requires the action of distinct enzymes such as glycerol-3-phosphate acyltransferase (GPAT) and diacylglycerol acyltransferase (DGAT). Upon synthesis, TAGs are largely stored in discrete subcellular organelles called oil bodies (3).

Thus, lipid accumulation in seeds requires the coordinated activity of several enzymes in distinct organelles. Oilseed crops, such as soybean, are particularly efficient in FA synthesis and TAG storage in the seed. In some oilseed crops, lipid storage molecules can account for more

than 40% of seed dry mass, whereas in soybean, lipids correspond to 20% of the seed dry mass (4, 5). This massive accumulation of TAGs suggests that FA and TAG biosynthesis during seed development is precisely regulated, and several reports have shown that the regulation FA biosynthesis occurs primarily at the transcriptional level (6–9). The expression level of many FA biosynthetic genes show a similar pattern of expression that peaks during the onset of the maturation program and declines during the later stages of maturation, suggesting a precise transcriptional coordination of gene expression for FA biosynthesis during seed development (10).

The WRINKLED1 (WRI1) transcription factor (TF), a member of the APETALA2 TF family, is characterized as a master regulator of seed FA biosynthesis in many plant species (11). WRI1 is both necessary and sufficient for FA accumulation in many oilseed crops (11–13). Arabidopsis (*At*)*wri1* loss-of-function mutants exhibit an 80% reduction in seed oil content (14). Ectopic overexpression of *WRI1* results in significant increases of FA-related gene expression and, consequently, increased oil content in several plant species (12, 13, 15–18). WRI1 affects seed oil levels by regulating the expression of genes involved in FA biosynthesis. Experimental modulation of *WRI1* expression alters seed oil levels and causes corresponding changes in FA-related gene expression, suggesting that WRI1 is a transcriptional activator of FA biosynthetic genes, including those encoding subunits of ACCase, ACP and KAS (16, 19). Consistent with this hypothesis, the AW Box DNA motif (CNTNGNNNNNNNCG) is enriched in genes responsive to alterations in WRI1 levels, and this element was shown to be bound by WRI1 *in vitro* (20).

Despite its importance for seed oil accumulation, a detailed, mechanistic understanding of how WRI1 regulates FA biosynthesis in seeds is lacking. In particular, a comprehensive, genome-wide description of all genes regulated by WRI1 has not been reported, making the exact mechanism by which this important TF acts to regulate FA synthesis unclear. To address this gap in knowledge, we analyzed the genome-wide binding profile of soybean WRI1 (GmWRI1) during seed development, the first comprehensive report of WRI1 binding sites in plants. Our results establish the extent to which WRI1 acts as a central regulator of lipid biosynthesis and provide new insights into the mechanisms by which WRI1 regulates FA and TAG accumulation in seeds.

Results

GmWRI1 Binds and Regulates Genes Involved in Fatty Acid Biosynthesis

We identified genes directly regulated by GmWRI1-A (Glyma.08G227700) and GmWRI1-B (Glyma.15G221600). Although there are several WRI1 homologs in soybean, *GmWRI1-A* and *GmWRI1-B* showed the highest degree of sequence similarity with *AtWRI1,* and other have demonstrated their ability to rescue the *wri1* seed oil mutant phenotype in Arabidopsis (21). Using

existing RNA-Sequencing (RNA-Seq) databases (GEO accession GSE99571), we showed that both mRNAs are present at high levels in embryos at the cotyledon (cot) and early maturation (em) stages and declines at later stages of embryo development, preceding the onset of FA accumulation in soybean seeds (Fig. 1B). Analysis of the spatial distribution of both *GmWRI1* mRNAs at the em stage (GEO accession GSE46096) showed that they accumulated predominately in embryo subregions of the seed (Fig. 1C). Both temporally and spatially, expression of *GmWRI1-A* and *GmWRI1-B* was consistent with their role in the control of the FA and TAG biosynthetic programs in embryos.

To identify genes that are directly regulated by GmWRI1 during seed development, we first identified genes bound by the TFs. We performed chromatin immunoprecipitation followed by DNA sequencing (ChIP-Seq) experiments using soybean embryos at the em stage with antibodies that recognize both GmWRI1-A and GmWRI1-B. Defining bound genes as those with a reproducible GmWRI1 ChIP-Seq peak within a genomic region 1 kp upstream the gene transcription start site (TSS) to the transcription termination site, we identified 11,203 genes that were bound by GmWRI1 (Fig. 1C, Dataset S1). Because *GmWRI1* expression is high in embryos at the cot stage (Fig. 1B), we also performed GmWRI1 ChIP-Seq experiments with cot stage embryos. However, unlike the experiments at em stage, the cot stage data did not pass ENCODE quality thresholds for ChIP-Seq data (22)(Dataset S2), which accounts for the lower number of bound genes detected at the cot stage compared to the em stage (Fig. S1A, Dataset S1). Nevertheless, a major overlap between the GmWRI1 bound genes at the cot and em stages was observed (Fig. S1A). Furthermore, several genes involved in FA biosynthesis that were shown by others to be upregulated in soybean seeds and roots in response to *GmWRI1* overexpression were bound by GmWRI1 in soybean embryos (12, 21)(Fig. S2). Specifically, 13 of 18 and 13 of 15 genes, respectively, that are upregulated by ectopic *GmWRI1* expression in soybean roots (21) and in transgenic soybean seeds (12) were bound by GmWRI1 (Fig. S2). Together, these results suggest strongly that the ChIP-seq data accurately profiles GmWRI1 binding sites in soybean embryos.

Because not all TF binding interactions represent transcriptional regulatory events (23), we filtered GmWRI1 bound genes for those that are co-expressed with the TF. The rationale is that the TF and the genes that it directly regulates must be expressed in the same cells. Given *GmWRI1* expression patterns, we defined co-expressed genes as those whose mRNA levels are at least 2-fold or higher in embryo cotyledon parenchymal tissues compared to seed coat parenchyma and seed coat hilum tissues, with a false-discovery rate (FDR) threshold of 0.05 (Harada–Goldberg Soybean Seed Development LCM RNA-Seq Dataset, GEO accession

GSE99109; Dataset S1). We identified 2,155 genes that were both bound by GmWRI1 and co-expressed with *GmWRI1* and designated these as GmWRI1 direct target (DT) genes (Fig. 1C). The overlap (P < 3.4e-58) between GmWRI1 bound and co-expressed genes was significant (Fig. 1D).

GmWRI1 DT genes were involved in FA and TAG biosynthesis. First, Gene Ontology (GO) enrichment analysis revealed that GmWRI1 DT genes were overrepresented for genes involved with FA and TAG biosynthesis, such as fatty acid biosynthetic process and acetyl-CoA metabolic process (Fig. 1F, Dataset S1). Surprisingly, other biological processes appeared to be highly enriched, such as cell population proliferation and cytokinesis by cell plate formation, suggesting that GmWRI1 might be involved with biological process other than lipid accumulation in seeds (Fig. 1F, Dataset S1). Second, because FA biosynthesis and TAG assembly are complex processes that involve many different enzymes in different organelles, we manually curated a list of FA and TAG biosynthesis (FAT) genes based on the Plant Metabolic Network (PMN) (https://plantcyc.org/databases/soycyc/8.0) and used the Arabidopsis Acyl-Lipid Metabolism (Aralip) (http://aralip.plantbiology.msu.edu/) for the annotation of those genes (Dataset S3). Of the 826 soybean FAT genes that encode different steps in the FA and TAG biosynthetic pathways (Dataset S3), 111 were GmWRI1 direct target genes at the em stage (P < 1.82e-30, Fig. 1E). Our results showed that that GmWRI1 regulates genes that encode enzymes involved in the initiation of FA biosynthesis in the chloroplast, such as the ACCase subunits, CAC2 and BCCP2, and KASIII (Fig. 1A). We also showed that genes encoding TAG assembly enzymes in the endoplasmic reticulum, such as GPDHC, GPAT and DGAT, are bound by or direct targets of GmWRI1 (Fig. 1A). Together, our results confirm that GmWRI1 directly regulates most genes required for lipid biosynthesis in soybean embryos (Fig. 1F, Dataset 2).

WRI1 Binds Intragenic and Intergenic Regions that are Enriched for Specific DNA Motifs

To gain insight into the mechanisms by which GmWRI1 transcriptionally regulates the FA biosynthesis program in soybean embryos, we analyzed regions bound by GmWRI1 in direct target genes. Surprisingly, we showed that the ChIP-Seq peak summits which presumably represents the GmWRI1 binding site in cot and em stage embryos were mostly located downstream the TSS in 5` untranslated (UTR) regions, exons, and introns (Figs. 2, A and B and Fig. S1B). These results differ from our previously characterized binding profiles of soybean (Gm)LEC1, GmAREB3, GmbZIP67, and Gm ABI3 TFs in embryos, in which ChIP-Seq peak summits were mostly located in region upstream the TSS (24, 25).

AW Box DNA motif is highly enriched in GmWRI1 binding sites

We asked if DNA motifs are enriched in a 100 bp region around the ChIP-Seq peak summit, designated as GmWRI1 binding sites, and showed that one motif identified with *de novo* discovery analysis in all GmWRI1 DT and in FAT GmWRI1 DT genes was the AW box, a DNA sequence previously shown to be bound by AtWRI1 in vitro (Fig. 2C, Dataset S4). All GmWRI1 DT and FAT GmWRI1 DT genes were significantly enriched for the AW Box (Bonferroni-adjusted P values < 0.01), with 35.2% and 56.3% of binding sites, respectively, containing full AW Boxes (Fig. 2D). These results suggest that the AW Box underlies the direct interaction of GmWRI1 with many, but not all, of its binding sites.

Given that most WRI1 binding sites were downstream of the TSS, we determined the distribution of AW Boxes in GmWRI1 DT and FAT GmWRI1 DT genes. AW Boxes were predominately found in regions downstream the TSS (Fig. S3), consistent with GmWRI1 binding site locations (Fig. 2A and Figs. S3, A and B). We note, however, that AW Box enrichment may be influenced by the higher GC content of region downstream the TSS. Nevertheless, AW Box distribution and enrichment in GmWRI1 binding sites provide strong evidence that this DNA motif is likely to play an important role in specifying GmWRI1 occupancy in regions downstream of the TSS.

Additional DNA motifs are enriched in WRI1 binding sites

De novo discovery analysis identified another enriched DNA motif, the CNC Box (CNCCNCC), in the binding sites of all GmWRI1 DT and FAT GmWRI1 DT genes (Fig. 2D, Dataset S4). Although the CNC Box has not yet been shown to be bound by WRI1, its enrichment in GmWRI1 binding sites suggests its role in mediating transcription of GmWRI1 DT genes. Consistent with this hypothesis, this element was previously shown to be enriched in the 5' UTR of FA-related genes in numerous plant species (26). Additionally, a GAGA rich DNA motif was discovered *de novo* in GmWRI1 binding sites (Figs. 2, B and C). This motif is similar to one which we previously identified in the binding sites of GmLEC1, GmAREB3, GmbZIP67 and GmABI3 (24, 25). Altogether, these results suggests that transcriptional regulation of GmWRI1 DT genes involve DNA elements other than the well characterized AW Box.

Because FAs and TAGs are synthesized during the maturation phase of seed development, we determined whether the CCAAT Box (CCAAT), G Box (CACGTG) and RY (CATGCA) DNA elements that are known to be bound by the major maturation TF regulators GmLEC1, GmABI3, GmFUS3, and GmLEC2, were enriched in GmWRI1 binding sites (27). These motifs were not identified by *de novo* discovery analysis, except for the G Box motif which was observed in only 3.3% of GmWRI1 binding sites. Aside from the G box, these motifs were not

significantly enriched in GmWRI1 binding sites of GmWRI1 DT and FAT GmWRI1 DT genes (Fig. 2D). This result suggests that GmWRI1-mediated transcriptional control of FA biosynthesis involves distinct cis-regulatory modules from those utilized by these other putative maturation TFs.

Identification of Arabidopsis Genes Directly-Regulated by WRI1

Most AW Box and CNC Box motifs are located downstream of the TSS in GmWRI1 DT genes (Fig. S3). To determine if these motifs were similarly located in Arabidopsis genes that are regulated by WRI1, we identified genes that were misregulated in *atwri1-1* mutant embryos and in Arabidopsis leaf cells ectopically expressing *GmWRI1* (Figs. 3, A and B and Fig. S4). We identified 301 and 1,213 differentially expressed (DE) genes, respectively, that were up and downregulated (two fold, FDR < 0.05) in *atwri1-1* compared with wild-type embryos at the linear-cotyledon stage (Fig. S5, Dataset S5). A previous microarray analysis of Arabidopsis seeds identified 25 downregulated genes in *atwri1* mutant when compared to wild-type (WT) seed (28), 11 of which overlapped with the 1213 list of downregulated genes in *atwri1* mutant embryos identified here. The low number of previously identified downregulated genes in *atwri1* mutants may reflect the lower sensitivity of this assay in that the microarray only queried ~3,500 genes and whole seed rather than embryo mRNA was assayed (28). Because *AtWRI1* is primarily expressed in embryonic tissues, our list of downregulated genes in *atwri1* mutant embryos provides a more comprehensive representation on the mRNA changes resulting from the *atwri1* mutation.

We also overexpressed soybean *GmWRI1* in Arabidopsis leaf protoplasts to identify DE genes (Fig. 3B). Protoplasts were transfected with the *GmWRI1-A* gene driven by the *CaMV35S* promoter (*p35S:GmWRI1*), and their mRNA transcriptome was compared to protoplasts transfected with the *mCHERRY* gene driven by the *35S* promoter (*p35S:mCHERRY,* Fig. 3B). We identified 1,490 and 245 genes, respectively, that were up and downregulated (two fold, FDR < 0.05) in cells overexpressing *GmWRI1* compared to the *MCHERRY* control (Fig. S5, Dataset S5). To our surprise, only 19 of 1,490 genes upregulated by overexpressing *GmWRI1* in Arabidopsis leaf protoplasts were previously reported as upregulated genes in two-week-old transgenic Arabidopsis seedlings overexpressing *AtWRI1* (20), suggesting that overexpressing *WRI1* in different tissues (leaf protoplasts versus seedlings) results in distinct transcriptional outputs. Nevertheless, FA and TAG biosynthesis GO terms were highly enriched in the genes upregulated by overexpressing *GmWRI1* in Arabidopsis leaf protoplasts, suggesting that

GmWRI1 overexpression is sufficient to promote the activation of the seed FA and TAG program in leaf cells (Fig. S5, Dataset S5).

A significant overlap of 401 genes ($p < 1.7e-214$) was observed between genes downregulated in *atwri1* mutant embryos and upregulated by *35S:GmWRI1* overexpression in leaf cells (Fig. 3C). We designated these as Arabidopsis WRI1 directly regulated (AtWRI DR) genes. GO enrichment analysis revealed that *atwri1* downregulated genes, *GmWRI1* OX upregulated genes, and AtWRI DR genes were highly enriched for GO terms related to FA and TAG biosynthetic processes, such as fatty acid biosynthetic process and acetyl-CoA biosynthetic process from pyruvate (Fig. 3D). These results indicate that WRI1 acts as a central regulator that activates the FA and TAG biosynthesis programs in seeds.

DNA motif enrichment analysis of AtWRI1 DR genes showed that the AW Box and CNC Box were overrepresented primarily in regions downstream the TSS (Fig. 3E, Dataset S5). The AW Boxes and CNC Boxes were detected downstream of the TSS in 70.3% and 59.6%, respectively, of AtWRI DR genes (Fig. 3E, Dataset S5). By contrast, the AW Boxes and CNC Boxes were detected in only 41% and 48.9%, respectively, of the downstream regions of 1000 sets of randomly matched sites (Dataset S5). In regions upstream of the TSS, the AW Box was enriched in 33.2% of AtWRI1 DR genes as compared with 18.5% in background regions (Fig. 3E, Dataset S5). No significant enrichment of the CNC Box was observed in upstream regions. These results are consistent with the observation that the AW Box and CNC Box are overrepresented in binding sites downstream of the TSS of GmWRI1 DT genes, suggesting that these elements contribute to the transcriptional regulation of these genes.

Functional Analysis of AW Boxes and CNC Boxes in GmWRI1 Direct Target Gene Binding Regions

We conducted transient assays with soybean to determine whether enriched AW Box and CNC Box motifs are required for the GmWRI1-mediated transcription FAT genes. We cloned WT versions of FAT gene GmWRI1 binding sites as well as mutant variants in which all AW Box (mAW), CNC Box (mCCNCC) or both (mAWmCNC) DNA motifs were mutated (Figs. 4, A and B, Dataset S6). Binding sites for the *ACP4* (Glyma.15G098500), *BCCP2* (Glyma.13G0 57400) and *CAC2* (Glyma.05G221100) genes contained at least one full AW Box motif (Fig. 4B, Dataset S6). We also isolated the binding site for *ACP3* (Glyma.19G240100) that contains a shortened version of the AW Box (CNTNGNNNNNNCG) which was shown to be bound by GmWRI1 *in vitro* (29). With the exception of *BCCP2,* these binding sites all contained at least one CNC Box (Fig. 4B). Binding sites were fused upstream of a firefly luciferase gene with a *35S* minimal promoter in a

plasmid that also contained a *35S:Renilla Luciferase* gene as a normalization control. Constructs were assayed by transfecting the plasmids into soybean embryo protoplasts and by co-transfecting the plasmid into tobacco (*Nicotiana benthamiana*) leaf protoplasts with and without a *35S:GmWRI1* gene (Figs. 4, C and D).

We showed that all WT GmWRI1 binding sites promoted reporter gene activity in soybean embryo protoplasts, and all except *ACP3* activated reporter genes in tobacco leaf protoplasts co-transfected with GmWRI1 (Figs. 4, C and D). These results suggest that elements within these binding sites, even those located downstream of the TSS, acted as enhancers to recruit GmWRI1 and activate gene transcription. Interestingly, binding sites with mutated DNA motifs caused variable responses in soybean embryo and tobacco leaf protoplasts (Figs. 4, C and D). Mutations in the AW Box of *ACP4* and *BCCP2* binding sites resulted in severe reductions in their enhancer activity in both soybean embryo cells and tobacco leaf cells (Figs. 4, C and D). By contrast, mutations in the CNC Box did not compromise enhancer activity (Figs. 4, C and D). This result suggests that GmWRI1 function depends primarily on AW Box elements in *ACP4* and *BCCP2* genes. By contrast, mutations in the AW Boxes and CNC Boxes of *CAC2* and *ACP3* binding sites did not compromise their enhancer activity in soybean embryos. However, mutation of either motif negatively affected GmWRI1-mediated enhancer activity of *CAC2* binding sites in tobacco leaf protoplasts (Figs. 4,C and D). Because there is positional overlap between one of the two CNC Boxes and the AW Box in the *CAC2* binding site (Fig. 4B), it was necessary to establish that the negative effect of the CNC Box mutations on enhancer activity did not result from alteration of the AW Box. We showed that mutation of only the upstream CNC Box (mCNCv2) reduced the binding site's enhancer activity in transactivation experiments with GmWRI1 in tobacco leaf protoplasts (Fig. 4D). Furthermore, *CAC2* binding sites with mutations in both CNC Boxes and in the AW Box (mAWmCNC) resulted in a greater reduction in reporter gene activity than those resulting from mutations in either motif separately (Fig. 4D). This suggests that the two motifs may act additively or synergistically in transcriptional regulation.

The *ACP3* binding site did not serve as an enhancer of GmWRI1 activity in tobacco leaf protoplasts (Fig. 4D), and AW Box and CNC Box mutations in this binding site did not compromise its enhancer activity in soybean embryo protoplast (Fig. 4C). These results open the possibility that GmWRI1 occupancy and function might require other TFs. We created a series of 100 bp overlapping deletions in the *ACP3* binding site (Fig. S6) and assayed these deletion constructs in soybean embryo protoplasts to identify regions with elements that may serve as enhancers. We showed that all deletions caused some reduction in reporter gene activity compared to the complete region (Fig. S6). However, deletions in regions containing the GmWRI1 ChIP-Seq peak

summit resulted in the greatest reduction in enhancer activity (Fig. S6), suggesting that DNA elements other than the AW Box or the CNC Box may be important to specify their enhancer activity in soybean embryo cells.

WRI1 and LEC1 Control Fatty Acid-Related Genes through Distinct Binding Loci

We and others showed that Arabidopsis AtLEC1 and soybean GmLEC1 are major regulators of FA and TAG biosynthesis in seeds (27, 30). We previously identified *GmWRI1-A* and *GmWRI1-B* as a direct target of GmLEC1 in soybean embryos (25) and show here that *GmLEC1-1* and *GmLEC1-2* are GmWRI1 DT genes (Dataset S1). These results suggest that GmWRI1 and GmLEC1 act in a transcriptional positive feedback circuit in the control of FA and TAG biosynthesis in seeds.

To understand how these two TFs regulate FA and TAG biosynthesis programs, we identified FAT GmWRI1 DT genes that are bound by GmLEC1 in em-stage embryos (25). A total of 69 (62% of) FAT GmWRI1 DT genes were also bound by GmLEC1, a significant overlap ($p < 1.49e-07$, Fig. 5A). The GmWRI1 DT/GmLEC1 bound genes encoded the vast majority of enzymes required for FA and TAG biosynthesis, although some key steps in the pathways are uniquely targeted by each TF (Fig. S7). Specifically, the KASIII enzyme which catalyzes the initial steps of FA biosynthesis in the chloroplasts (Fig. 1A) is a GmWRI1 DT that is not bound by GmLEC1 (Fig. S7, Dataset S3). By contrast, genes encoding FATB and GPATs were previously shown to be direct target of GmLEC1 (25) and are not GmWRI1 DT genes. These results suggests that although complementary, the two TFs serve some specific roles in FA and TAG biosynthesis in embryos. Moreover, comparison of FAT mRNA levels in soybean embryos at the em stage showed that GmWRI1 DT/GmLEC1 bound mRNA levels are higher than those of GmWRI1 DT mRNAs only (Fig. 5B). Our results suggest that GmWRI1 and GmLEC1 have additive or synergistic functions in FAT gene regulation in embryos.

We next evaluated the positional relationship between GmWRI1 and GmLEC1 binding sites in FAT GmWRI1 DT genes. The median distance between GmWRI1 and GmLEC1 ChIP-Seq summit positions was 142 bp, suggesting that these TFs bind with sites that are close but distinct in target genes (Fig. 5C). The findings that GmWRI1 and GmLEC1 binding sites were enriched for different DNA motifs (Fig. 5D) provides support for this conclusion (25). Because the distance between TF binding sites is short, we determined the enrichment of DNA motifs in a 50 bp window around ChIP-Seq peak summits. As observed previously, binding sites for GmWRI1 were strongly enriched for the presence of AW Box and CNC Box in the FAT genes targeted by GmWRI1 only and in the FAT GmWRI1 DT genes that are also bound by GmLEC1 (Fig. 5D).

GmLEC1 binding sites were enriched for G Box-type DNA motifs (ACGTG and G Box) and AW Box motifs, but not the CNC Box (Fig. 5D). Our results suggests that GmLEC1 occupancy in FAT GmWRI1 DT genes could be anchored by a bZIP type TF that can physically bind with GmLEC1 and G Box elements. Additionally, the enrichment of AW Box in GmLEC1 binding sites could indicate that GmLEC1 and GmWRI1 can form a transcriptional complex that regulates FAT target genes. However, the ability of GmLEC1 to physically interact with GmWRI1 has yet to be confirmed.

Discussion

We report for the first time in any plant species the genome-wide profile of WRI1 binding sites and a comprehensive list of genes directly regulated by the TF. Our results confirm GmWRI1 as a major regulator of FA biosynthesis in soybean embryos and provide new insights into the mechanisms by which this TF regulates FA and TAG biosynthetic programs in embryos.

GmWRI1 is a Central Regulator of Lipid Biosynthesis in Seeds

Because FA and TAG accumulation in embryos requires the activity of several enzymes that operate in different organelles (Fig. 1A)(1, 3), the expression of genes encoding these enzymes must be highly coordinated to ensure the rapid and efficient incorporation and storage of carbon as FAs and TAGs during embryo development. The WRI1 TF has been characterized extensively as a positive regulator of the FA biosynthetic program in seeds, and we provide several lines of evidence that confirms and expands the role of GmWRI1 in FA and TAG biosynthesis in soybean. First, genome-wide identification of GmWRI1 DT genes showed that most enzymes in the FA and TAG biosynthetic pathways are directly regulated by GmWRI1, from the synthesis of malonyl-CoA in the chloroplast to the assembly of TAG in the endoplasmic reticulum (Fig. 1A, Dataset S3). Second, GmWRI1 binds to several FAT genes previously reported as being upregulated by ectopic overexpression of GmWRI1 (Fig. S2)(12, 21). Third, we showed that the ectopic overexpression of GmWRI1 resulted in the upregulation of several FA biosynthesis-related genes in Arabidopsis leaf protoplasts (Dataset S4). Fourth, GmWRI1 binding sites of FAT genes act as positive enhancers in soybean embryo cells, and mutations in the canonical AW Box element that is bound by GmWRI1 significantly decreased the enhancer activity of these binding sites (Fig. 4C). Fifth, we showed that some GmWRI1 binding sites are transactivated by GmWRI1 in tobacco leaf cells (Fig. 4D), suggesting strongly that interactions between the TF and binding sites activates gene transcription. Together, our results provide strong supporting evidence that GmWRI1 acts as a positive regulator of the FA and TAG biosynthesis program in soybean embryos.

Biological Relevance of GmWRI1 Binding Preference for Intragenic Regions

Our genome-wide analysis of binding sites revealed that GmWRI1 preferentially binds DNA regions downstream of the TSS (Fig. 2A). This binding profile is consistent with the enrichment of AW Boxes and CNC Boxes in regions downstream the TSS (Fig. S3). A large-scale analysis of ChIP-Seq and ChIP-Chip data for 27 Arabidopsis TFs revealed a group of TFs with a binding preference for intragenic regions, although the implications of intragenic TF binding in transcriptional regulation remained unclear (31). Here, we show that intragenic GmWRI1 binding sites activate transcription when placed upstream of a minimal promoter (Figs. 4, C and D), suggesting that these intragenic binding sites act as *cis*-acting enhancers. A recent study revealed that intragenic binding events are involved in chromatin remodeling (32). Similar to GmWRI1, the cucumber *TENTRIL IDENTITY* (*CsTEN*) gene, a TCP TF of the CYC/TB1 clade, binds primarily to target gene intragenic regions resulting in chromatin disassembly and transcription activation (32). CsTEN possesses an N-terminal intrinsically disordered region domain (IDR) that was shown to have histone acetyltransferase activity that facilitates chromatin accessibility during target gene transcription. Interestingly, the AtWRI1 TF also has an N-terminal IDR domain (33). Similarities in the binding site position and presence of an IDR domain in CsTEN and AtWRI1 suggest the potential for mechanistic similarities. However, further experiments are needed to demonstrate such associations. Nevertheless, our results provided new and exciting insights into the mechanisms by which GmWRI1 functions in the control of FA and TAG biosynthesis in embryos.

GmWRI1 Function is Dependent on AW Boxes and CNC Boxes

In addition to confirming that WRI1 activates target gene transcription by binding with the canonical AW Box, our results revealed that other DNA elements are involved in GmWRI1 transcriptional activation. First, although AW Boxes in *ACP4* and *BCCP2* binding sites were required for their enhancer activity, mutations of AW Boxes did not compromise the enhancer activity of *CAC2* and *ACP3* binding sites in soybean embryos (Fig. 4C). Second, only ~55% of binding sites for FAT GmWRI1 DT genes contain full AW Boxes (Fig. 2D). Third, we showed that other DNA elements, such as the CNC Box, are enriched in GmWRI1 binding sites (Figs. 2, C and D). Fourth, AW Box and CNC Box elements act additively or synergistically in the *CAC2* binding site to activate GmWRI1-mediated transcription in tobacco leaf cells (Fig. 4D). Lastly, we showed that this CNC Box is highly enriched in regions downstream of the TSS of AtWRI1 DR genes (Fig. 3E). The enrichment of CNC Box in GmWRI1 binding sites raises the possibility that GmWRI1 occupancy in these sites depends on this DNA element.

Questions remain as to how the CNC Box acts mechanistically in the regulation of GmWRI1 DT genes. Although WRI1 has not been shown to bind physically with DNA motifs other than the AW Box, AtWRI1 tertiary structure revealed that it possesses an AP2/EREPBP DNA binding domains similar to that of AtERF1 that bind the GCC Box (AGCCGCC) (34). A survey of Arabidopsis cistromes revealed that several TFs with an AP2/EREBP domain similar to GmWRI1 binds DNA motifs similar to the CNC Box (35). These finding open the possibility that GmWRI1 could potentially bind directly with the CNC Box, although this hypothesis remains to be verified. Another possibility is that GmWRI1 occupancy in regions enriched for CNC Box elements is mediated through interactions with other TFs that bind this element. The CsTEN TCP-like TF binds primarily with intragenically localized CTCCGCC elements which shares similarity with the CNC Box (32). Thus, it is possible that a TCP-like TF could coregulate GmWRI1 DT genes through a mechanism that is independent of the AW Box. Recently, it was shown that AtWRI1 interacts physically with AtTCP4, AtTCP10 and AtTCP24, although these TFs appear to repress seed oil accumulation (36). One potential model is that a TCP-like TF physically binds with the CNC box and regulates GmWRI1 occupancy and function in an AW Box independent manner (Fig. 6). Further experiments are needed to define the role of CNC Box elements in controlling the transcription of FAT genes regulated by GmWRI1 (Fig. 6).

GmWRI1 and GmLEC1 Act in a Positive-Feedback Gene Circuit to Control Fatty Acid and Triacylglycerol Biosynthesis in Embryos

In addition to controlling the expression of genes encoding the vast majority of enzymes in the FA and TAG biosynthetic pathway, our results also suggest that GmWRI1 acts to alter the regulatory state of soybean embryos to promote oil accumulation. We showed that GmWRI1 directly regulates several other well characterized transcriptional regulator of lipid biosynthesis in seeds, including GmLEC1, GmABI3, GmFUS3, and GmbZIP14 (27, 30, 37). These results suggests that GmWRI1 may also control FA and TAG biosynthesis through its regulation of other TFs that promote lipid accumulation.

Interactions between WRI1 and LEC1 are particularly prominent. Here, we showed that GmLEC1 is directly regulated by GmWRI1, that overexpression of GmWRI1 in Arabidopsis leaf protoplasts resulted in the upregulation of *AtLEC1* (AT1G21970), and that *AtLEC1-LIKE* (AT5G47670) is downregulated in *atwri1* mutant embryos (Dataset S5). We showed previously that GmWRI1 is directly regulated by GmLEC1 (24, 25), that overexpression of *AtLEC1* resulted in elevated expression of *AtWRI1* (24), and that *AtWRI1* is downregulated in *atlec1* mutant seeds (24). These findings indicate that WRI1 and LEC1 act in a positive feedback subcircuit in the

control of FA and TAG biosynthesis in seeds through reciprocal regulation of the other TF's expression and through their regulation of FA and TAG biosynthetic genes (Fig. 6). In gene regulatory networks, positive feedback gene circuits are important to trigger irreversible changes in transcriptional programs (38). The onset of the maturation program of the seed is hallmarked by the *de novo* initiation of many biological programs that were inactive during the morphogenesis phase of the embryo development, such as the lipid and protein accumulation and storage (27). Initiation of the FA biosynthetic program in the embryo requires a robust and stable regulatory state to ensure proper expression of all the genes required for the quick and efficient allocation of carbon into seed oil bodies. In this sense, a positive feedback subcircuit between GmLEC1 and GmWRI1 is likely to be required for the rapid onset of the FA and TAG accumulation programs during seed development.

Due the physical proximity of GmWRI1 and GmLEC1 binding sites (Fig. 5C), one could speculate that the involvement of GmLEC1 and GmWRI1 in the regulation of FA biosynthesis occurs through the formation of a TF complex, although interaction between LEC1 and WRI1 has not been demonstrated. Nevertheless, we show that GmLEC1 occupancy in GmWRI1 FAT DTs can be specified through the interaction of a bZIP type of TF (Figs. 5D and 6). It was previously shown that the soybean GmLEC1 TF can physically interact with bZIP TFs (25, 39). It is likely that the control of the FA biosynthetic program in soybean seeds occurs through physical interactions between GmLEC1 and other bZIP type of TFs. Interestingly, we observed that FAT GmWRI1 DT genes that are also bound by GmLEC1 are expressed at higher levels when compared to genes bound by each individual TF (Fig. 5C). This result suggests that GmWRI1 and GmLEC1 co-binding can synergistically affect the transcriptional activation of FAT target genes. These results provide new insights into the relationship between these two important regulators of FA and TAG biosynthesis and can further aid studies designed to improve lipid accumulation in seeds.

Materials and Methods

Chromatin Immunoprecipitation and DNA sequencing

Soybean cv Williams 82 (W82) plants were grown, and embryos were isolated from seeds staged as described by Pelletier et al. 2017 (24). GmWRI1 ChIP assays were done as described previously (24, 25), with minor modifications. After nuclei purification, nuclei were resuspended in shearing buffer (10mM Tris pH8, 1mM EDTA, 0.1% SDS, 1mM PMSF), and chromatin was sheared to ~250 bp fragments using a Covaris E220. After shearing, Triton X-100 and NaCl were added to chromatin to final concentrations of 1% and 150mM, respectively. Anti GmWRI1

(Glyma.08G227700 and Glyma.15G221600) antibodies for ChIP were raised in rabbits against the peptide FSTQPGDHETESDVN, corresponding to the amino acids 352-366 of GLYMA.08G227700. The reactivity of antibodies to the peptide was assessed through Western blots (Dataset S2). DNA libraries of ChIP and input samples were prepared using the NuGEN Ovation Ultralow System V2. Amplified sub-libraries were pooled into one multiplex library, size-selected (100 to 500 bp) through gel electrophoresis, and purified using the QIAquick Gel Extraction Kit, following manufacturer instructions. The multiplexed library was sequenced at 50-bp single-end reads using an Illumina HiSeq 4000 sequencing system.

RNA-Sequencing Experiments with Arabidopsis Leaf Protoplasts Overexpressing *GmWRI1* and with *atwri1* Mutant Embryos

For *GmWRI1* overexpression assays with Arabidopsis leaf protoplasts, three-week-old *Arabidopsis thaliana* Col-0 plants were grown as described (40). Arabidopsis mesophyll protoplasts were isolated and transfected, according to Jo et al. 2020 (25). For each biological replicate, 5 x 10^5 cells were transfected with 10 µg of a plasmid carrying the *GmWRI1*-A cDNA, synthesized using primers listed in Table S6, and driven by the *35S* promoter (*p35S:GmWRI1*) or a *35S:mCherry:NOSt* plasmid as a negative control. After protoplasts transfection, cells were incubated for 16 hours at room temperature.

Homozygous *atwri1-1* and wild-type Col-0 seeds were plated on GM media and grown as described in Le et al. (2010)(41). Flowers from six-week old plants were tagged at anthesis, and 8 days after pollination embryos dissected from seeds (25-30 embryos per biological replicate, three biological replicate per genotype).

Total RNA was isolated using the RNAqueous™-Micro Total RNA Isolation Kit, following manufacturer's instructions. RNA samples were then treated with DNAse for digestion of DNA using the TURBO DNA-free™ Kit, following manufacturer's instructions. RNA-Seq libraries were prepared using Tecan Ovation RNA-Seq System V2 using 50 ng of total RNA, and double-stranded cDNA was fragmented to ~200 bp using a Covaris E220. End repair, A-tailing, adapter ligation and PCR enrichment were performed according to the library preparation protocol described in Ravi et al. 2012 (42), using the NEXTflex ChIP-Seq Barcodes (BioScientific). Libraries were quantified using Quant-iT PicoGreen dsDNA Reagent (Grand Island, NY) and a Nanodrop ND-3300 instrument, and sequenced on a NovaSeq 6000 sequencer.

Cloning and Mutation of GmWRI1 Binding Sites

GmWRI1 binding sites (200 to 500 bp around the WRI1 ChIP-seq peak summit) were PCR amplified from the soybean genomic DNA and cloned into a SpeI digested pDLUC15 vector (25) using the In-Fusion® HD Cloning Kit, according to the manufacturer's instructions. Mutant derivatives and *ACP3* deletion series were synthesized by Twist Bioscience and inserted into pDLUC15 using the same cloning strategy. For the mutant derivatives of *ACP4*, we used PCR in-site directed mutagenesis by amplifying the *ACP4* WT in pDLUC15 vector using primers with the respective mutations. Primers and the full DNA sequences of WRI1 binding sites and mutant derivatives are listed in Table S6.

Transient Expression Assays

Transient assays and Firefly and Renilla luciferase activity in protoplasts were performed as described by Jo et al. 2020 (25). All trials were done in three biological replicates and one-way or two-way ANOVA with Tukey's post hoc test ($P < 0.05$) was used for the statical analysis in soybean embryo protoplasts and tobacco leaf protoplasts, respectively.

Data Analysis

WRI1 ChIP-Seq in cot and em stage soybean embryos. ChIP-Seq data were analyzed as described previously (25) with minor modifications. In summary, de-multiplexed reads were quality filtered and mapped to the Gmax 275 version of the soybean genome (Wm82.a2.v1) using bowtie2 (43). Redundant reads were filtered using samtools (44). Quality assessment of the ChIP-seq libraries were determined by evaluating the non-redundant fraction (NRF) and strand cross-correlation (CC), following ENCODE guidelines for ChIP-Seq data (22). Description of NRF, CC values are summarized in Dataset S2. ChIP-Seq peaks were identified using MACS3 (https://github.com/macs3-project/MACS) and reproducible peaks between two biological replicates were identified using the irreproducible discovery rate (IDR) pipeline (https://github.com/nboley/idr) with an IDR threshold of 0.01.

As GmWRI1 binding pattern occurs mainly in regions downstream the TSS (Figure 2B), we considered genes bound by GmWRI1 as those with a reproducible ChIP-Seq peak within a 1 kb window upstream of the genes' TSS or within the annotated gene body.

GmWRI1 DT genes are defined as those bound by GmWRI1 (ChIP-Seq) and coexpressed by GmWRI1 (Dataset S1). Because GmWRI1 expression is mainly in embryonic cells (Figure 1C), we used the Harada-Goldberg Soybean Seed Development LCM RNA-Seq Dataset (GEO accessions, GSE57606, GSE46096, and GSE99109) to identify genes that are two-fold upregulated with an FDR < 0.01 in embryo cotyledon abaxial or adaxial parenchyma versus seed

coat hilum and seed coat parenchyma subregions. We then determine WRI1 co-expressed genes those that are upregulated in at least two of the above comparisons (Dataset S1).

Identification of soybean FAT genes. To identify FAT genes, we used the soybean plant metabolic network (PMN) database Soycyc and extracted the soybean genes associated with the pathways listed in the Dataset S3. Additionally, we also included genes annotated with the GO category for fatty acid biosynthetic process (GO:0006633).

RNA-Seq data from Arabidopsis protoplasts and embryos. Sequenced reads were demultiplexed and reads corresponding to rRNA sequences were removed. The resulting filtered reads were mapped to Arabidopsis primary transcripts (TAIR10) using bowtie v0.12.7 with parameters -v 2 -5 10 -3 4 -m 1 --best –strata. We used the EdgeR package (v3.10.5) to obtain normalized expression values using the Trimmed Mean of M-values (TMM) method, and to identify differentially expressed genes between the different genotypes (FDR < 0.05).

GO enrichment analysis. GO term enrichment analyses were done as described previously (16, 17). GO functional annotation was extracted from TAIR10 and Soybase databases and used for Arabidopsis and Soybean datasets, respectively. A q value threshold of 0.01 was used to determine the enriched GO categories.

DNA motif analysis. DNA motif *de novo* discovery analysis in binding sites of GmWRI1 DT genes and FAT genes (Figure 2C) was performed using the MEME-ChIP tool from the MEME suite v5.3.3 (45) as described previously (25), with modifications. Motif maximum length was set to 15 and a E-value cutoff of 0.01 was used determine de novo discovered DNA motifs. Discovered motifs were compared to the Arabidopsis DAP-Seq TF and the top ranked DNA motifs identified are listed in Dataset S4.

DNA motif enrichment analysis was performed as described (25). DNA motifs AW-BOX (CNTNGNNNNNNNCG), CNC-Box (CNCCNCC), G-Box (CACGTG), RY (CATGCA), and CAAT-Box (CCAAT) were screened in the GmWRI1 and GmLEC1 binding sites (Figure 2D, Figure 5D) and 500 bps windows around the TSS of WRI1 directly regulated genes in Arabidopsis (Figure 4E) using the known function from HOMER (46). The significance of enrichment was determined when comparing the percentage of GmWRI1 binding where a motif was identified against the percentage in a set of “random binding sites”. These “random binding sites” were generated from random genes and had a comparable number, length, and position to the relevant WRI1 binding site. Each set of WRI1 binding site was compared against 1,000 sets of “random binding sites” to calculate the P value significance of enrichment using a hypergeometric test. P values were adjusted using the Bonferroni method, and a significance threshold of $P < 0.01$ was used to identify significantly enriched DNA motifs.

Data availability

Data are available at GEO, under the accessions: GSE152310 and GSE253104 (GmWRI1 ChIP-seq at em at cot stage, respectively), GSE253109 (*atwri1* mutant embryo RNA-seq), and GSE253107 (GmWRI1 overexpression).

Acknowledgments

We thank Makenzie Melby, Sana Mohammad, and Mary Irani for their technical assistance. This work was supported by a grant from the National Science Foundation Plant Genome Research Program.

References

1. P. D. Bates, S. Stymne, J. Ohlrogge, Biochemical pathways in seed oil synthesis. *Current opinion in plant biology* **16**, 358–364 (2013).

2. J. Bruinsma, *World agriculture: towards 2015/2030: an FAO study* (Routledge, 2017).

3. S. Baud, Seeds as oil factories. *Plant reproduction* **31**, 213–235 (2018).

4. J. B. Taylor, T. M. Richard, C. L. Wilheim, OILSEEDS, OVERVIEW (2004).

5. T. E. Clemente, E. B. Cahoon, Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiol* **151**, 1030–1040 (2009).

6. P. O'Hara, A. R. Slabas, T. Fawcett, Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. *Plant Physiology* **129**, 310–320 (2002).

7. S. Baud, I. A. Graham, A spatiotemporal analysis of enzymatic activities associated with carbon metabolism in wild-type and mutant embryos of Arabidopsis using in situ histochemistry. *The Plant Journal* **46**, 155–169 (2006).

8. S. Baud, L. Lepiniec, Physiological and developmental regulation of seed oil production. *Progress in lipid research* **49**, 235–249 (2010).

9. M. A. Troncoso-Ponce, *et al.*, Comparative deep transcriptional profiling of four developing oilseeds. *The Plant Journal* **68**, 1014–1027 (2011).

10. S. Baud, L. Lepiniec, Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. *Plant Physiology and Biochemistry* **47**, 448–455 (2009).

11. Q. Kong, L. Yuan, W. Ma, WRINKLED1, a "Master Regulator" in transcriptional control of plant oil biosynthesis. *Plants* **8**, 238 (2019).

12. L. Chen, *et al.*, Soybean (Glycine max) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation. *Molecular genetics and genomics* **293**, 401–415 (2018).

13. J. Liu, *et al.*, Increasing seed mass and oil content in transgenic Arabidopsis by the overexpression of wri1-like gene from Brassica napus. *Plant Physiology and Biochemistry* **48**, 9–15 (2010).

14. N. Focks, C. Benning, wrinkled1: a novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant physiology* **118**, 91–101 (1998).

15. D. An, M. C. Suh, Overexpression of Arabidopsis WRI1 enhanced seed mass and storage oil content in Camelina sativa. *Plant Biotechnology Reports* **9**, 137–148 (2015).

16. S. Baud, S. Wuillème, A. To, C. Rochat, L. Lepiniec, Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in Arabidopsis. *The Plant Journal* **60**, 933–947 (2009).

17. A. Cernac, C. Benning, WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *The Plant Journal* **40**, 575–585 (2004).

18. B. Shen, *et al.*, Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. *Plant physiology* **153**, 980–987 (2010).

19. A. To, *et al.*, WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in Arabidopsis. *The Plant Cell* **24**, 5007–5023 (2012).

20. K. Maeo, *et al.*, An AP2-type transcription factor, WRINKLED1, of Arabidopsis thaliana binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. *The Plant Journal* **60**, 476–487 (2009).

21. B. Chen, *et al.*, Multiple GmWRI1s are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (Glycine max). *Plant biotechnology journal* **18**, 155–171 (2020).

22. S. G. Landt, *et al.*, ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research* **22**, 1813–1831 (2012).

23. P. J. Farnham, Insights from genomic profiling of transcription factors. *Nature Reviews Genetics* **10**, 605–616 (2009).

24. J. M. Pelletier, *et al.*, LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proceedings of the National Academy of Sciences* **114**, E6710–E6719 (2017).

25. L. Jo, *et al.*, Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. *Proceedings of the National Academy of Sciences* **117**, 1223–1232 (2020).

26. G. Bonaventure, J. B. Ohlrogge, Differential regulation of mRNA levels of acyl carrier protein isoforms in Arabidopsis. *Plant physiology* **128**, 223–235 (2002).

27. L. Jo, J. M. Pelletier, J. J. Harada, Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *Journal of integrative plant biology* **61**, 564–580 (2019).

28. S. A. Ruuska, T. Girke, C. Benning, J. B. Ohlrogge, Contrapuntal networks of gene expression during Arabidopsis seed filling. *The Plant Cell* **14**, 1191–1206 (2002).

29. Q. Kong, *et al.*, The Arabidopsis WRINKLED1 transcription factor affects auxin homeostasis in roots. *Journal of experimental botany* **68**, 4627–4634 (2017).

30. J. Mu, *et al.*, LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. *Plant physiology* **148**, 1042–1054 (2008).

31. K. S. Heyndrickx, J. V. de Velde, C. Wang, D. Weigel, K. Vandepoele, A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. *The Plant Cell* **26**, 3894–3910 (2014).

32. X. Yang, *et al.*, Regulation of plant architecture by a new histone acetyltransferase targeting gene bodies. *Nature Plants* **6**, 809–822 (2020).

33. W. Ma, *et al.*, Deletion of a C–terminal intrinsically disordered region of WRINKLED 1 affects its stability and enhances oil accumulation in Arabidopsis. *The Plant Journal* **83**, 864–874 (2015).

34. Z. Qiao, *et al.*, Molecular basis of the key regulator WRINKLED1 in plant oil biosynthesis. *Science Advances* **8**, eabq1211 (2022).

35. R. C. O'Malley, *et al.*, Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**, 1280–1292 (2016).

36. Q. Kong, *et al.*, TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR4 interacts with WRINKLED1 to mediate seed oil biosynthesis. *Plant physiology* **184**, 658–665 (2020).

37. M. Matthijs, *et al.*, The transcription factor bZIP14 regulates the TCA cycle in the diatom Phaeodactylum tricornutum. *The EMBO journal* **36**, 1559–1576 (2017).

38. W. Xiong, J. E. Ferrell, A positive-feedback-based bistable 'memory module'that governs a cell fate decision. *Nature* **426**, 460–465 (2003).

39. Y. Yang, *et al.*, Transcription factor bZIP52 modulates Arabidopsis seed oil biosynthesis through interaction with WRINKLED1. *Plant Physiology*, kiad270 (2023).

40. M. A. West, *et al.*, LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. *The Plant Cell* **6**, 1731–1745 (1994).

41. B. H. Le, *et al.*, Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences* **107**, 8063–8070 (2010).

42. R. Kumar, *et al.*, A high-throughput method for Illumina RNA-Seq library preparation. *Frontiers in plant science* **3**, 202 (2012).

43. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357–359 (2012).

44. H. Li, *et al.*, The sequence alignment/map format and SAMtools. *bioinformatics* **25**, 2078–2079 (2009).

45. P. Machanick, T. L. Bailey, MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).

46. S. Heinz, *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576–589 (2010).

47. E. Collakova, *et al.*, Metabolic and transcriptional reprogramming in developing soybean (Glycine max) embryos. *Metabolites* **3**, 347–372 (2013).

Figure Legends

Figure 1. GmWRI1 acts as a central regulator of fatty acid biosynthesis in soybean seeds. (A) Simplified schematic diagram of FA and TAG biosynthesis from (16). Reactions are catalyzed by enzymes (labeled in red) encoded by genes that are bound by (teal) and/or co-expressed with (grey) GmWRI1. (B) Temporal changes in relative fatty acid and protein levels (top) and of *GmWRI1-A* (*Glyma.08G227700*) and *GmWRI1-B* (*Glyma.15G221600*) mRNA levels (bottom) during soybean embryo development. The macromolecule graph was adapted from Collakova *et al*. (2013)(47), and mRNA transcriptome data was from the Harada Embryo mRNA-Seq Dataset (GEO Series GSE99571). Abbreviations: cotyledon (cot), early maturation (em), mid-maturation (mm), late maturation (lm), seedling (sdlg). (C) mRNA levels of *GmWRI1-A* and *GmWRI1-B* in seed subregions of the embryo axis, and cotyledons, endosperm, and seed coat at the early maturation stage (GEO Series GSE46096). (D) GmWRI1 DT genes in soybean embryos at the early maturation stage. Venn diagram shows the overlap of GmWRI1 bound and coexpressed genes. Statistical significance (hypergeometric distribution) of the overlap between bound genes and co-expressed genes is indicated. (E) Venn diagram showing the overlap of the GmWRI1 DT genes at the em stage and of FAT genes, with the statistical significance of the overlap indicated (hypergeometric distribution). (F) Heatmap showing the q value significance of GO term enrichment for GmWRI1 co-expressed, bound, and GmWRI1 DT genes. The GO terms listed are the top 20 most enriched biological process GO terms for GmWRI1 DT genes.

Figure 2. Binding profile and enriched cis-regulatory elements in GmWRI1 binding sites. (A) Position of GmWRI1 ChIP-seq peak summits in reference to the TSS of GmWRI1 DT genes. (B) Pie chart showing the fraction of GmWRI1 ChIP-Seq peaks that overlap with distinct genomic features (intergenic, 2 kb upstream of the TSS, 5' UTR, exons, introns, 3' UTR and 1 kb downstream the TTS). (C) Position weight matrices of DNA motifs discovered *de novo* in GmWRI1 binding sites of GmWRI1 DT (top), and FAT GmWRI1 DT (bottom) genes, and their significance as indicated by their associated E values. (D) DNA motif enrichment in the binding sites of GmWRI1 DT (left) and FAT GmWRI1 DT (right) genes. Graphs show the percentage of bound regions (colored) and randomly selected, equivalently sized and spaced regions (gray) containing the indicated DNA motifs. A Bonferroni-adjusted *P* value threshold of 0.01 was used to indicate the statistical significance (*) of enrichment, and n.s. indicates no significant difference.

Figure 3. Identification of AtWRI1 directly regulated (DR) genes in Arabidopsis. (A) Overview of strategy used to identify Arabidopsis genes differentially expressed (A) in *atwri1* mutants compared with wild-type (Col-0) embryos at 8 days after pollination and (B) in leaf protoplasts transfected with *35S:GmWRI1* or *35S:mCHERRY*, as a negative control. RNA-Seq experiments identified differentially expressed mRNAs that differed at least two-fold (FDR < 0.05) when compared to the control. (C) Venn diagram showing the overlap between genes downregulated in *atwri1* mutant embryos and upregulated in leaf protoplasts overexpressing *GmWRI1* to identify the AtWRI1 DR genes. (D) Heatmap showing the q value significance of enriched GO terms for downregulated genes in *atwri1* mutant embryos, upregulated genes in the *35S:GmWRI1* leaf cells, and the AtWRI1 DR genes. GO terms listed are the top 10 most enriched biological process terms (q value < 0.01). (E) Bubble plot showing the motif enrichment in 500 bp regions upstream and downstream the TSS of AtWRI1 DR genes. Diameter of the circles depict the frequencies at which DNA motifs were identified in the isolated regions, and the intensity of their color indicates the statistical significance (Bonferroni-adjusted *P* values) of the enrichment relative to the normal distribution of a population of randomly generated regions.

Figure 4. Functional analysis of GmWRI1 binding sites. (A) Genome browser view showing the position of GmWRI1 ChIP-Seq peaks for *BCCP2* (Glyma.13G057400), *ACP4* (Glyma.15G098500), *CAC2* (Glyma.05G221100), and *ACP3* (Glyma.19G240100). ChIP-Seq Peak summits (dashed lines) and TSS positions (tick mark) are shown. Numbers next the summit positions depict summit distances from the TSS in bp. (B) Diagrams of GmWRI1 binding sites for genes in (A) used in the transient assay experiments. Yellow, pink and brown symbols indicate the positions of AW Boxes, CNC Boxes, and short AW Boxes, respectively. (C) Effect of WT GmWRI1 binding sites and of binding sites with mutations in all detectable AW Box (mAW), CNC Box (mCNC) or both (mAWmCNC) DNA motifs on reporter gene activity in soybean cotyledon embryo protoplasts. (D) Transactivation analysis of WT and mutant GmWRI1 binding sites in tobacco leaf protoplasts transfected with *p35S:GmWRI1* (teal) or *p35S:mCHERRY* (grey). For (C) and (D), average firefly to Renilla luciferase activity ratios were multiplied by 100 and plotted with the respective standard errors. Lower-case letters indicate the significance of comparisons based on ANOVA and post hoc Tukey test results.

Figure 5. GmWRI1 and GmLEC1 act independently to control FA biosynthesis in soybean seeds. (A) Bar graph shows the number of FAT GmWRI1 DT genes relative to those that are also bound by GmLEC1 . The *P* value significance of the overlap is indicated (hypergeometric

distribution). (B) Box plots comparing the distribution of mRNA levels in em embryos (GEO accession GSE99571) of FAT GmWRI1 DT genes (teal), GmWRI DT genes that are also bound by GmLEC1 (dark red), and gene not bound by either GmWRI1 or GmLEC1 (grey). Asterisk indicate significant differences between groups (Wilcoxon rank-sum test, with a significance threshold of $p < 0.01$). (C) Position of the GmWRI1 ChIP peak summit position relative to LEC1 ChIP peak summits in FAT GmWRI1 DT genes in in those that are also bound by GmLEC1. (D) DNA motif enrichment in GmWRI1 DT genes and in those that are also GmLEC1 bound. Circle diameters and color intensity are as discussed in Fig. 3E.

Figure 6. **Model for the regulation of FA biosynthesis and lipid accumulation in seeds.** GmWRI1 and GmLEC1 act in a transcriptional positive feedback subcircuit to regulate FA biosynthesis program in the seed. GmWRI1 regulates FA-related gene expression by binding with AW Boxes and, possibly, with CNC Boxes located downstream of the TSS. Alternatively, another TF may binds with the CNC Box. GmLEC1, another regulator of FA biosynthesis, regulates transcription through its interaction with bZIP TFs that bind with G Box DNA elements located most frequently in regions upstream of the TSS. Dashed arrows indicate transcriptional activation.

Supplemental Figure Legends

Figure S1. Comparison of GmWRI1 ChIP-Seq analyses with cotyledon and early maturation stage embryos. (A) Venn diagram comparing genes bound by GmWRI1 in embryos at the cot and em stages. (B) Position of cot and em GmWRI1 ChIP-seq peak summits of bound genes relative to the TSS.

Figure S2. Comparison of genes bound by GmWRI1 with genes previously reported to be upregulated by overexpression of GmWRI1 (12, 21).

Figure S3. Positional distribution of the GmWRI1 ChIP-Seq peak summits, AW Boxes, and CNC Boxes in GmWRI1 DT genes. Heatmaps show the frequency of GmWRI1 peak summits and AW and CNC Box motifs in regions 1000 bp upstream and downstream of the TSS of GmWRI1 DT (A) and the FAT GmWRI1 DT (B) genes.

Figure S4. Whole mount comparison of 8 DAP wild-type and *wri1* mutant embryos.

Figure S5. AtWRI1 directly regulated genes. (A) Number of DE genes (two-fold, FDR<0.05) identified in WT versus *atwri1* mutant embryos at 8 DAP and in Arabidopsis leaf cells transfected with *35S:WRI1* versus *35S:mCHERRY.* (B) Heatmap show enriched GO terms for each DE gene set (q value < 0.01).

Figure S6. Functional analyses of the *ACP3* binding site. (A) Diagrams illustrating the WT *ACP3* binding site and the deletion constructs tested. (B) Luciferase activity of *ACP3* WT and deletion constructs in soybean embryo protoplasts. The average firefly to Renilla luciferase activity ratio was normalized to the pDLUC15 negative control and is plotted with the standard errors (n=3). Lower-case letters indicate the significance of comparisons based on ANOVA and post hoc Tukey test results.

Figure S7. GmWRI1 and GmLEC1 control FA biosynthesis in soybean seeds. Simplified schematic illustrating FA and TAG biosynthetic pathways highlighting enzymes encoded by genes that are GmWRI1 DT (teal) or GmWRI1 DT and bound by GmLEC1 (red).

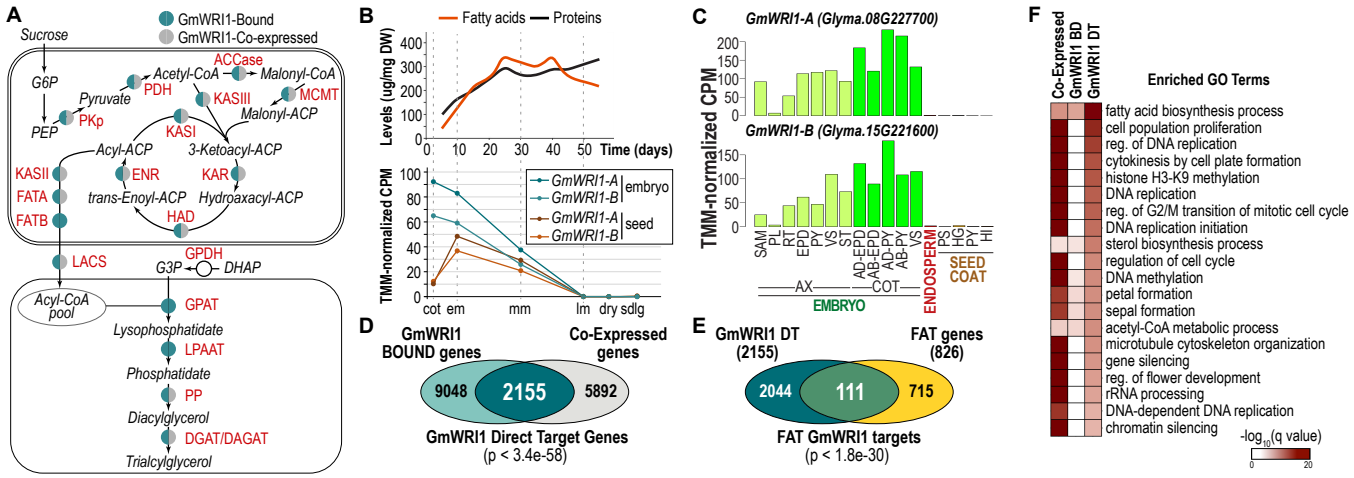

Figure 1. GmWRI1 acts as a central regulator of fatty acid biosynthesis in soybean seeds.
(A) Simplified schematic diagram of FA and TAG biosynthesis from (15). Reactions are catalyzed by enzymes (labeled in red) encoded by genes that are bound by (teal) and/or co-expressed with (grey) GmWRI1. (B) Temporal changes in relative fatty acid and protein levels (top) and of *GmWRI1-A (Glyma.08G227700)* and *GmWRI1-B (Glyma.15G221600)* mRNA levels (bottom) during soybean embryo development. The macromolecule graph was adapted from Collakova *et al*. (2013)(44), and mRNA transcriptome data was from the Harada Embryo mRNA-Seq Dataset (GEO Series GSE99571). Abbreviations: cotyledon (cot), early maturation (em), mid-maturation (mm), late maturation (lm), seedling (sdlg). (C) mRNA levels of *GmWRI1-A* and *GmWRI1-B* in seed subregions of the embryo axis, and cotyledons, endosperm, and seed coat at the early maturation stage (GEO Series GSE46096). (D) GmWRI1 DT genes in soybean embryos at the early maturation stage. Venn diagram shows the overlap of GmWRI1 bound and coexpressed genes. Statistical significance (hypergeometric distribution) of the overlap between bound genes and co-expressed genes is indicated. (E) Venn diagram showing the overlap of the GmWRI1 DT genes at the em stage and of FAT genes, with the statistical significance of the overlap indicated (hypergeometric distribution). (F) Heatmap showing the q value significance of GO term enrichment for GmWRI1 co-expressed, bound, and GmWRI1 DT genes. The GO terms listed are the top 20 most enriched biological process GO terms for GmWRI1 DT genes.

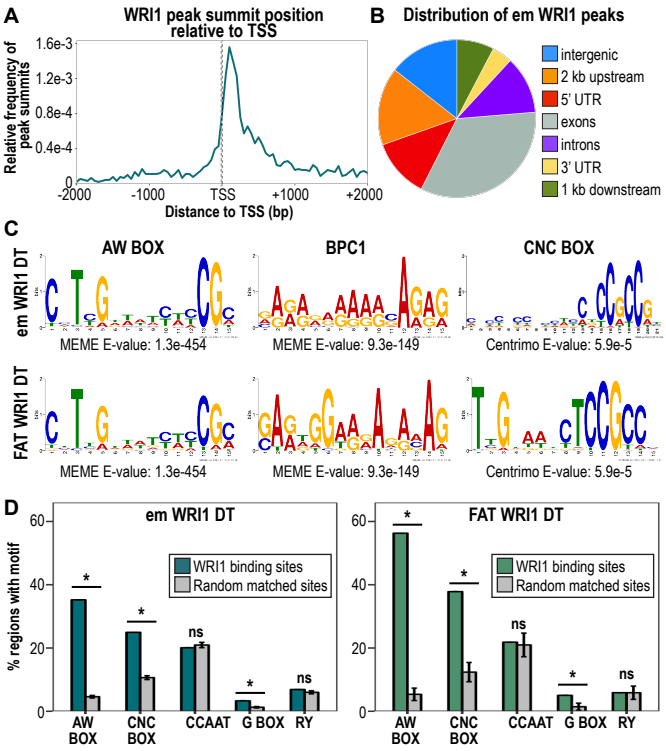

Figure 2. Binding profile and enriched cis-regulatory elements in GmWRI1 binding sites.
(A) Position of GmWRI1 ChIP-seq peak summits in reference to the TSS of GmWRI1 DT genes. (B) Pie chart showing the fraction of GmWRI1 ChIP-Seq peaks that overlap with distinct genomic features (intergenic, 2 kb upstream of the TSS, 5' UTR, exons, introns, 3' UTR and 1 kb downstream the TTS). (C) Position weight matrices of DNA motifs discovered de novo in GmWRI1 binding sites of GmWRI1 DT (top), and FAT GmWRI1 DT (bottom) genes, and their significance as indicated by their associated E values. (D) DNA motif enrichment in the binding sites of GmWRI1 DT (left) and FAT GmWRI1 DT (right) genes. Graphs show the percentage of bound regions (colored) and randomly selected, equivalently sized and spaced regions (gray) containing the indicated DNA motifs. A Bonferroni-adjusted P value threshold of 0.01 was used to indicate the statistical significance (*) of enrichment, and n.s. indicates no significant difference.

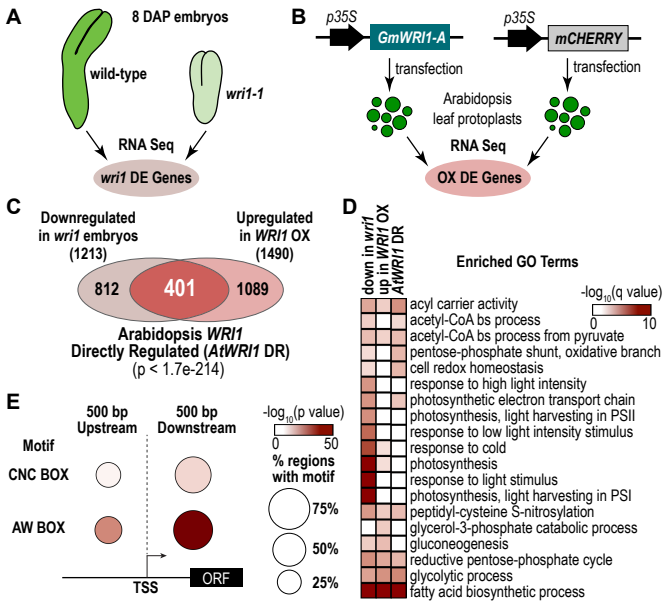

Figure 3. Identification of AtWRI1 directly regulated (DR) genes in Arabidopsis.
(A) Overview of strategy used to identify Arabidopsis genes differentially expressed (A) in *atwri1* mutants compared with wild-type (Col-0) embryos at 8 days after pollination and (B) in leaf protoplasts transfected with *35S:GmWRI1* or *35S:mCHERRY*, as a negative control. RNA-Seq experiments identified differentially expressed mRNAs that differed at least two-fold (FDR < 0.05) when compared to the control. (C) Venn diagram showing the overlap between genes downregulated in *atwri1* mutant embryos and upregulated in leaf protoplasts overexpressing *GmWRI1* to identify the AtWRI1 DR genes. (D) Heatmap showing the q value significance of enriched GO terms for downregulated genes in *atwri1* mutant embryos, upregulated genes in the *35S:GmWRI1* leaf cells, and the AtWRI1 DR genes. GO terms listed are the top 10 most enriched biological process terms (q value < 0.01). (E) Bubble plot showing the motif enrichment in 500 bp regions upstream and downstream the TSS of AtWRI1 DR genes. Diameter of the circles depict the frequencies at which DNA motifs were identified in the isolated regions, and the intensity of their color indicates the statistical significance (Bonferroni-adjusted P values) of the enrichment relative to the normal distribution of a population of randomly generated regions.

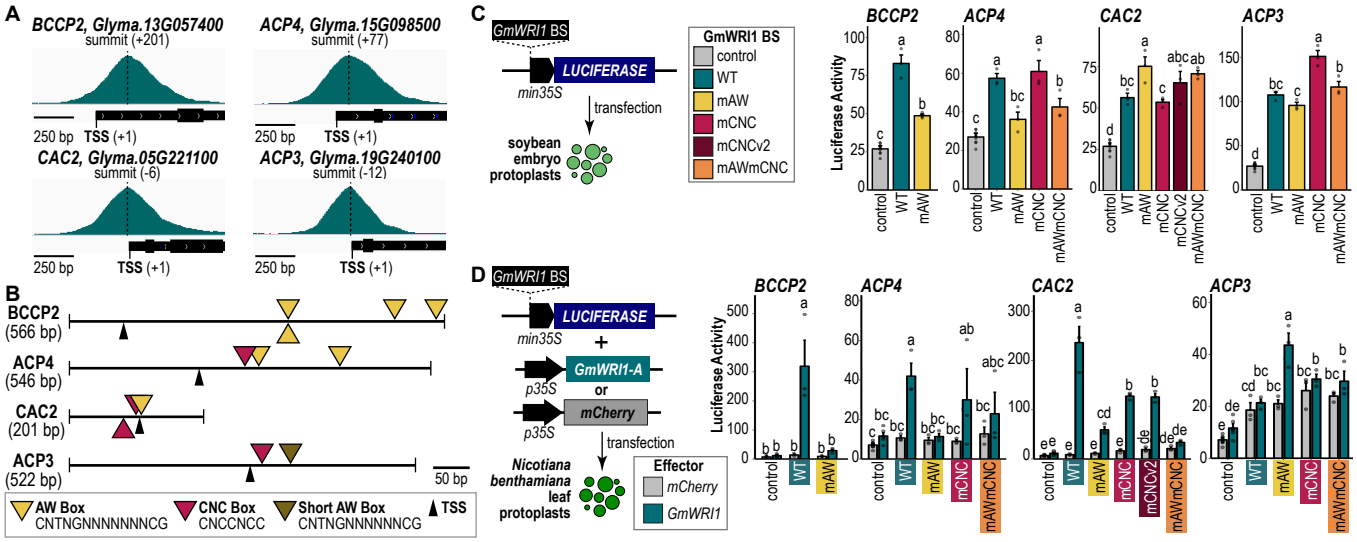

Figure 4. Functional analysis of GmWRI1 binding sites.
(A) Genome browser view showing the position of GmWRI1 ChIP-Seq peaks for *BCCP2 (Glyma.13G057400)*, *ACP4 (Glyma.15G098500)*, *CAC2 (Glyma.05G221100)*, and *ACP3 (Glyma.19G240100)*. ChIP-Seq Peak summits (dashed lines) and TSS positions (tick mark) are shown. Numbers next the summit positions depict summit distances from the TSS in bp. (B) Diagrams of GmWRI1 binding sites for genes in (A) used in the transient assay experiments. Yellow, pink and brown symbols indicate the positions of AW Boxes, CNC Boxes, and short AW Boxes, respectively. (C) Effect of WT GmWRI1 binding sites and of binding sites with mutations in all detectable AW Box (mAW), CNC Box (mCNC) or both (mAWmCNC) DNA motifs on reporter gene activity in soybean cotyledon embryo protoplasts. (D) Transactivation analysis of WT and mutant GmWRI1 binding sites in tobacco leaf protoplasts transfected with *p35S:GmWRI1* (teal) or *p35S:mCHERRY* (grey). For (C) and (D), average firefly to Renilla luciferase activity ratios were multiplied by 100 and plotted with the respective standard errors. Lower-case letters indicate the significance of comparisons based on ANOVA and post hoc Tukey test results.

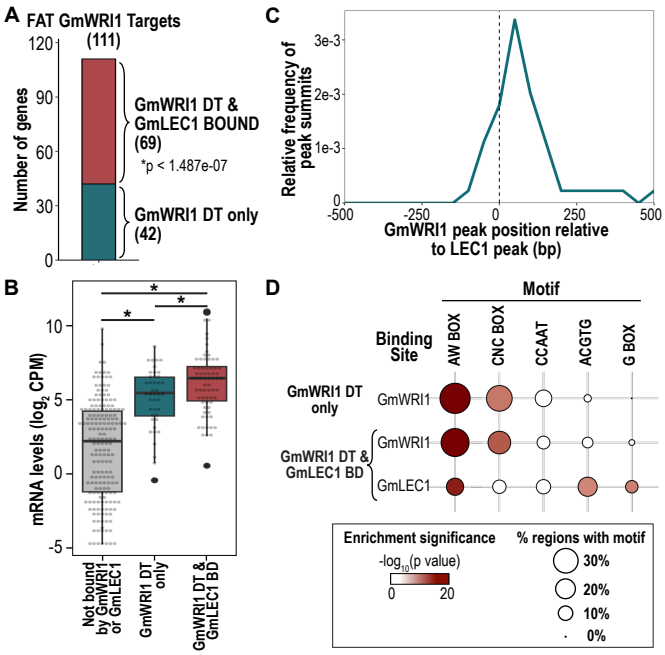

Figure 5. GmWRI1 and GmLEC1 act independently to control FA biosynthesis in soybean seeds. (A) Bar graph shows the number of FAT GmWRI1 DT genes relative to those that are also bound by GmLEC1 . The P value significance of the overlap is indicated (hypergeometric distribution). (B) Box plots comparing the distribution of mRNA levels in em embryos (GEO accession GSE99571) of FAT GmWRI1 DT genes (teal), GmWRI DT genes that are also bound by GmLEC1 (dark red), and gene not bound by either GmWRI1 or GmLEC1 (grey). Asterisks indicate significant differences between groups (Wilcoxon rank-sum test, with a significance threshold of $p < 0.01$). (C) Position of the GmWRI1 ChIP peak summit position relative to LEC1 ChIP peak summits in FAT GmWRI1 DT genes in in those that are also bound by GmLEC1. (D) DNA motif enrichment in GmWRI1 DT genes and in those that are also GmLEC1 bound. Circle diameters and color intensity are as discussed in Fig. 3E.

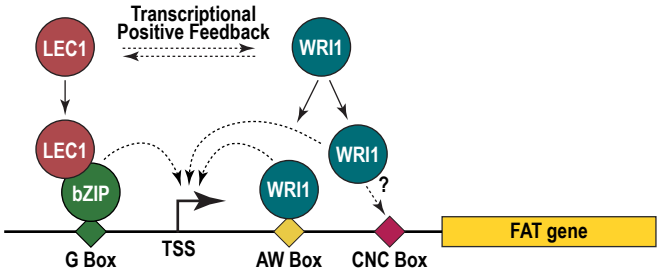

Figure 6. Model for the regulation of FA biosynthesis and lipid accumulation in seeds. GmWRI1 and GmLEC1 act in a transcriptional positive feedback subcircuit to regulate FA biosynthesis program in the seed. GmWRI1 regulates FA-related gene expression by binding with AW Boxes and, possibly, with CNC Boxes located downstream of the TSS. Alternatively, another TF may binds with the CNC Box. GmLEC1, another regulator of FA biosynthesis, regulates transcription through its interaction with bZIP TFs that bind with G Box DNA elements located most frequently in regions upstream of the TSS. Dashed arrows indicate transcriptional activation.

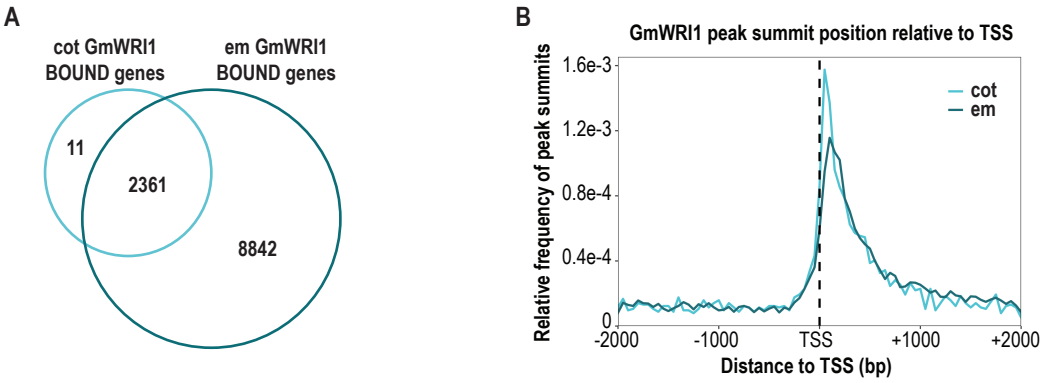

Figure S1. Comparison of GmWRI1 ChIP-Seq analyses with cotyledon and early maturation stage embryos.

(A) Venn diagram comparing genes bound by GmWRI1 in embryos at the cot and em stages. (B) Position of cot and em GmWRI1 ChIP-seq peak summits of bound genes relative to the TSS.

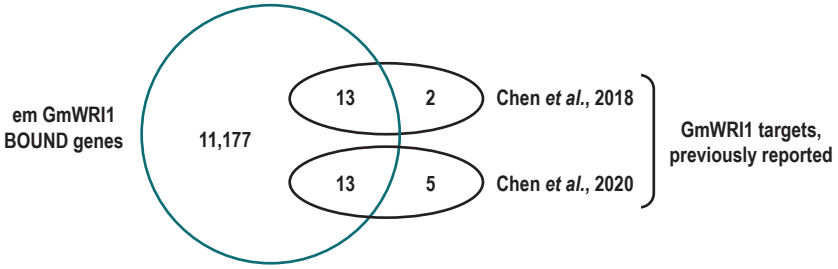

Figure S2. Comparison of genes bound by GmWRI1 with genes previously reported to be upregulated by overexpression of GmWRI1 (11, 20).

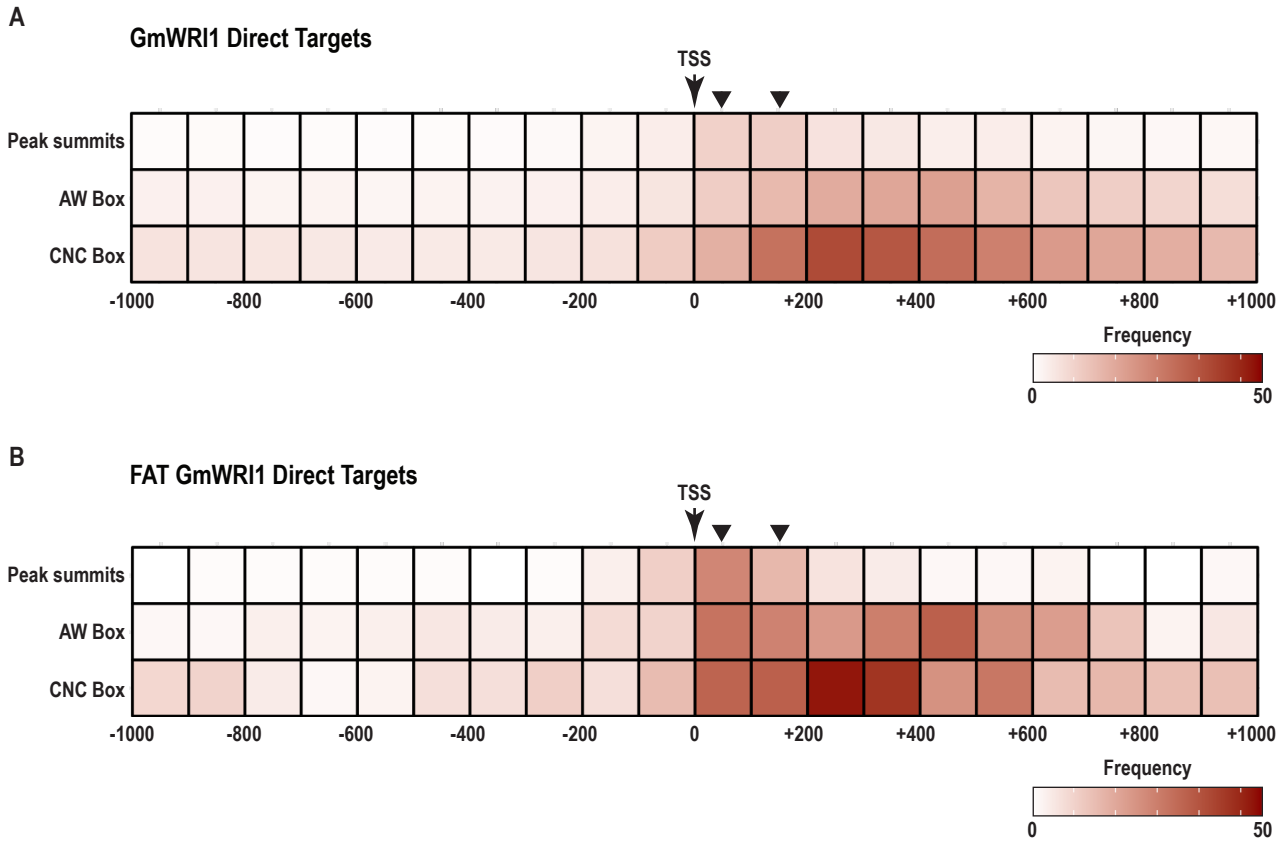

Figure S3. Positional distribution of the GmWRI1 ChIP-Seq peak summits, AW Boxes, and CNC Boxes in GmWRI1 DT genes.
Heatmaps show the frequency of GmWRI1 peak summits and AW and CNC Box motifs in regions 1000 bp upstream and downstream of the TSS of GmWRI1 DT (A) and the FAT GmWRI1 DT (B) genes.

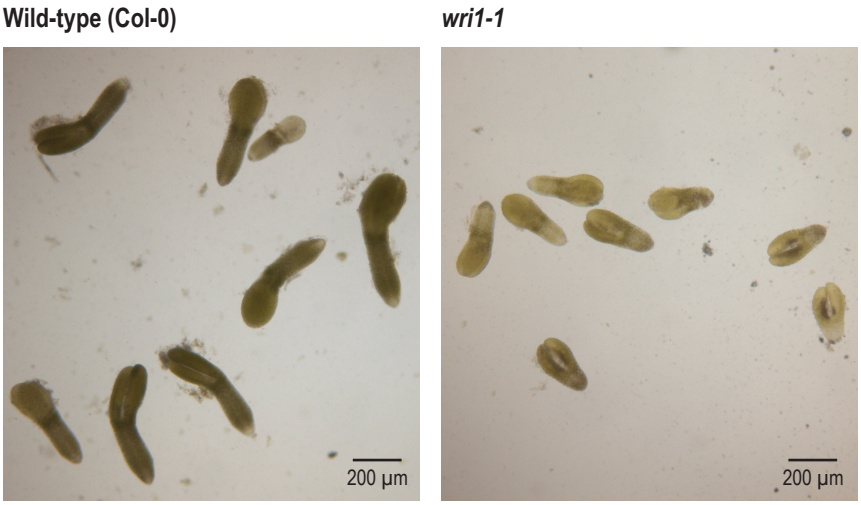

Figure S4. Whole mount comparison of 8 DAP wild-type and *wri1* mutant embryos.

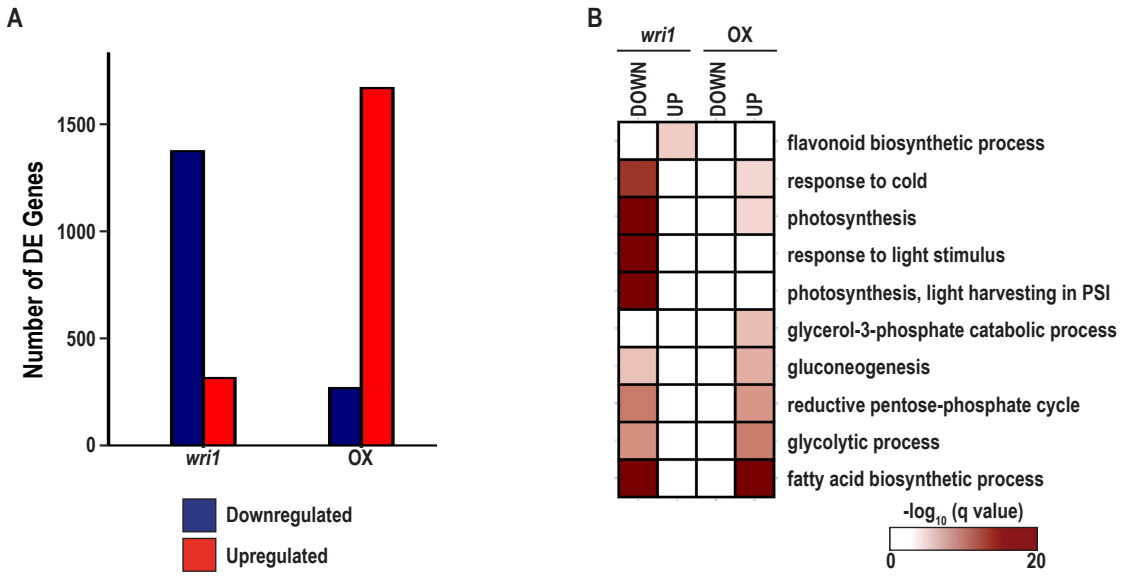

Figure S5. AtWRI1 directly regulated genes.
(A) Number of DE genes (two-fold, FDR<0.05) identified in WT versus *atwri1* mutant embryos at 8 DAP and in Arabidopsis leaf cells transfected with *35S:WRI1* versus *35S:mCHERRY*. (B) Heatmap show enriched GO terms for each DE gene set (q value < 0.01).

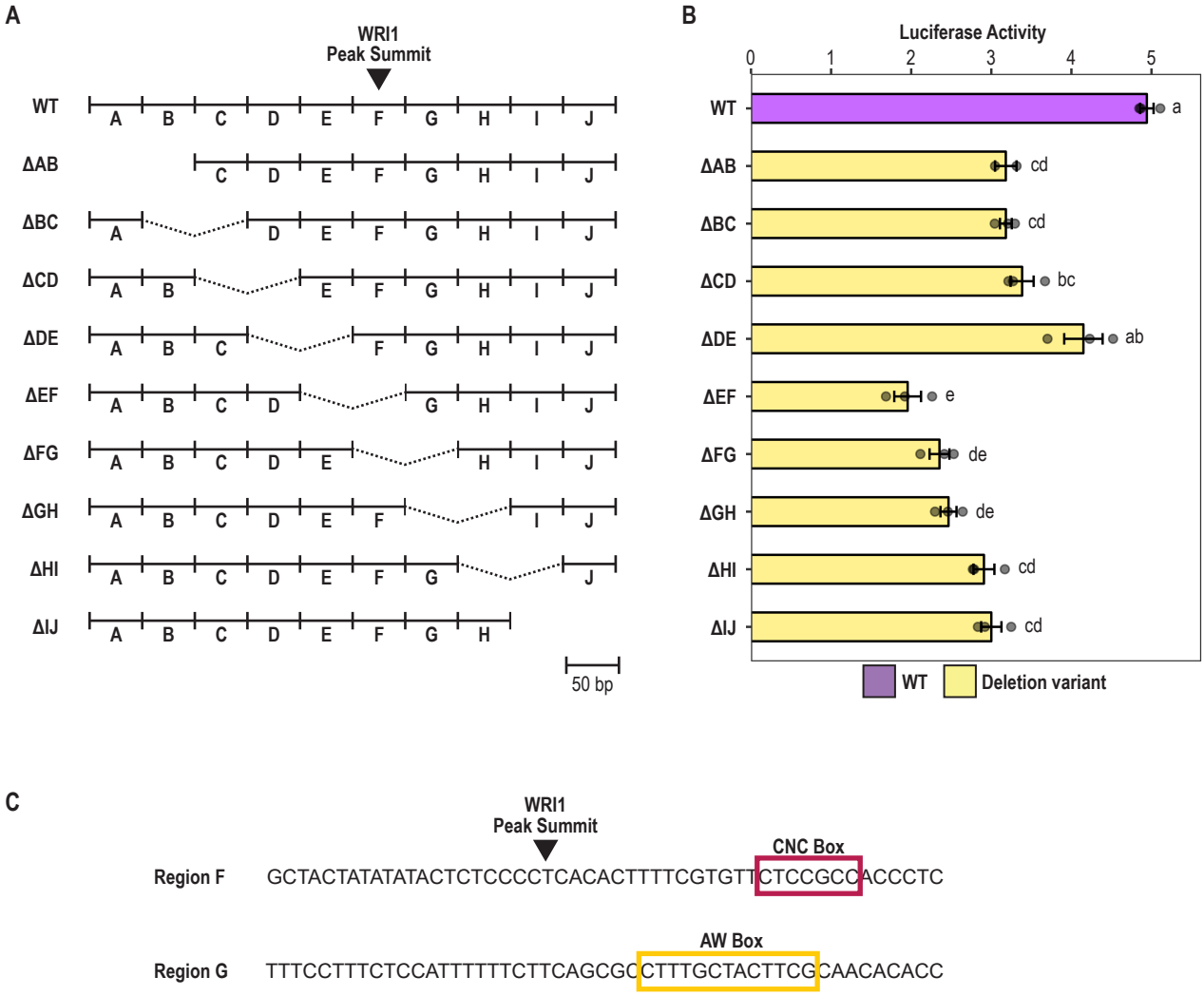

Figure S6. Functional analyses of the *ACP3* binding site.
(A) Diagrams illustrating the WT *ACP3* binding site and the deletion constructs tested. (B) Luciferase activity of *ACP3* WT and deletion constructs in soybean embryo protoplasts. The average firefly to Renilla luciferase activity ratio was normalized to the pDLUC15 negative control and is plotted with the standard errors (n=3). Lower-case letters indicate the significance of comparisons based on ANOVA and post hoc Tukey test results.

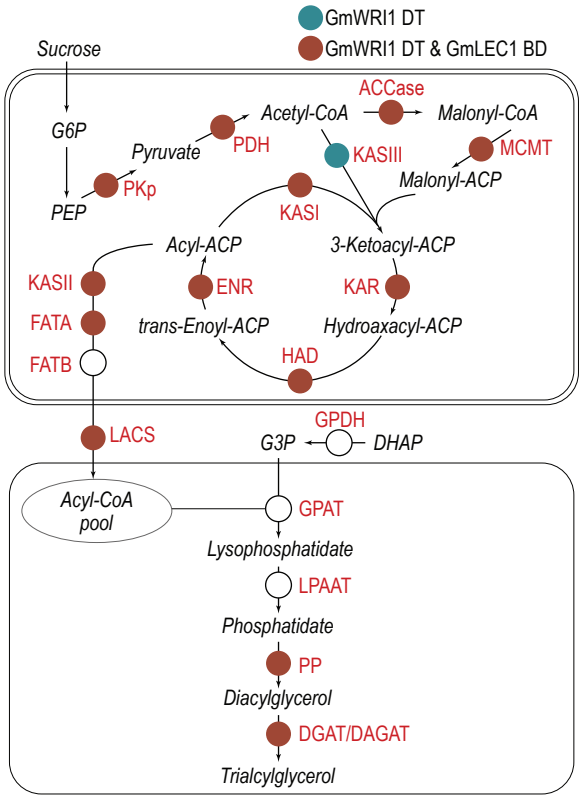

Figure S7. GmWRI1 and GmLEC1 control FA biosynthesis in soybean seeds.
Simplified schematic illustrating FA and TAG biosynthetic pathways highlighting enzymes encoded by genes that are GmWRI1 DT (teal) or GmWRI1 DT and bound by GmLEC1 (red).