the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2022) **110,** 589–606 doi: 10.1111/tpj.15682

RESOURCE

Exploring *Camelina sativa* lipid metabolism regulation by combining gene co-expression and DNA affinity purification analyses

Fabio Gomez-Cano[1], Yi-Hsuan Chu[1], Mariel Cruz-Gomez[1], Hesham M. Abdullah[2,3], Yun Sun Lee[1], Danny J. Schnell[2] and Erich Grotewold[1,*]

[1]*Department of Biochemistry and Molecular Biology, Michigan State University, 603 Wilson Road, Room 212, Biochemistry Building, East Lansing, MI 48824-6473, USA,*

[2]*Department of Plant Biology, Michigan State University, 612 Wilson Road, Room 166, East Lansing, MI 48824-1312, USA, and*

[3]*Biotechnology Department, Faculty of Agriculture, Al-Azhar University, Cairo 11651, Egypt*

Received 17 September 2021; revised 18 January 2022; accepted 20 January 2022; published online 22 January 2022.
*For correspondence (e-mail grotewol@msu.edu).

SUMMARY

Camelina (*Camelina sativa*) is an annual oilseed plant that is gaining momentum as a biofuel cover crop. Understanding gene regulatory networks is essential to deciphering plant metabolic pathways, including lipid metabolism. Here, we take advantage of a growing collection of gene expression datasets to predict transcription factors (TFs) associated with the control of Camelina lipid metabolism. We identified approximately 350 TFs highly co-expressed with lipid-related genes (LRGs). These TFs are highly represented in the MYB, AP2/ERF, bZIP, and bHLH families, including a significant number of homologs of well-known Arabidopsis lipid and seed developmental regulators. After prioritizing the top 22 TFs for further validation, we identified DNA-binding sites and predicted target genes for 16 out of the 22 TFs tested using DNA affinity purification followed by sequencing (DAP-seq). Enrichment analyses of targets supported the co-expression prediction for most TF candidates, and the comparison to Arabidopsis revealed some common themes, but also aspects unique to Camelina. Within the top potential lipid regulators, we identified CsaMYB1, CsaABI3AVP1-2, CsaHB1, CsaNAC2, CsaMYB3, and CsaNAC1 as likely involved in the control of seed fatty acid elongation and CsaABI3AVP1-2 and CsabZIP1 as potential regulators of the synthesis and degradation of triacylglycerols (TAGs), respectively. Altogether, the integration of co-expression data and DNA-binding assays permitted us to generate a high-confidence and short list of Camelina TFs involved in the control of lipid metabolism during seed development.

Keywords: transcription factor, RNA-seq, co-expression, DAP-seq, fatty acid, triacylglycerol biosynthesis.

INTRODUCTION

The Brassicaceae Camelina (*Camelina sativa* L. Crantz) annual plant is gaining increasing attention as a potential oilseed crop with characteristics that make it alluring as a renewable feedstock for biofuels and biobased products, among many other applications (Carlsson, 2009; Iskandarov et al., 2014). Camelina has a hexaploid genome that harbors approximately 90 000 genes organized into 20 chromosomes (Kagale et al., 2014; Liang et al., 2013). When compared with Arabidopsis (*Arabidopsis thaliana*), Camelina genes are classified into three types, including syntenic orthologs (syntelogs, approximately 70% of all genes), tandem duplicates (approximately 12%), and non-syntenic genes (approximately 18%) (Kagale et al., 2014). Of the syntelogs (a.k.a. paralogs), 10% are defined as fractionated because not all three copies are conserved (Kagale et al., 2014). Remarkably, in addition to having a low rate of fractionation, the majority of Camelina's paralogs (in the case of triplicated genes) display no significant differences in expression levels (Kagale et al., 2014). Despite the challenges imposed by its polyploid genome,

extensive gene expression analyses performed on developing Camelina seeds provided a transcriptome reference for this emerging crop, which includes 26 different datasets obtained at 13 different time points during seed development and one immediately after germination, expression data that are available at CamRegBase (Gomez-Cano et al., 2020). Yet, despite the growing collection of mRNA accumulation data, expression information from early time points during seed development, a critical stage for lipid biosynthesis (Rodríguez-Rodríguez et al., 2013; Pollard et al., 2015), is largely missing. Another important available resource in Camelina, given its biotechnological implications, is the growing list of genes associated with fatty acid (FA) and oil biosynthesis (Abdullah et al., 2016; Gomez-Cano et al., 2020; Mudalkar et al., 2014; Nguyen et al., 2013). This was to a large extent possible thanks to the close phylogenetic relationship of Camelina with Arabidopsis, reflected in the high sequence similarity of their genomes (Mandáková et al., 2019; Nikolov et al., 2019).

Camelina seeds are approximately 50 times larger than those of Arabidopsis, and they are rich in triacylglycerols (TAGs) containing mainly long unsaturated FAs, including linoleic acid (C18:3), which are excellent sources of omega-3 FAs (Berti et al., 2016; Gugel and Falk, 2006). Depending on the ecotype, Camelina oil may represent up to 40% of the total seed dry weight, which also contains high levels of vitamin E and antioxidants responsible for extending the lifetime of Camelina oil-containing products (Berti et al., 2016; Chaturvedi et al., 2019; Malik et al., 2018). As in other plants, Camelina TAG synthesis starts with the synthesis of FAs in plastids (Voelker and Kinney, 2001). In Camelina embryos, the maximum rate of oil synthesis is at mid-maturation ('green cotyledon'), i.e., between 14–20 days post-anthesis (DPA), while the mid-point for oil deposition is around 17–18 DPA. Consistently, C18:3 reaches its highest accumulation rate at 22 DPA (Pollard et al., 2015). In addition to C18:3, Camelina TAGs also contain significant amounts of very long-chain FAs (VLCFA) (C20–C24) with similar accumulation rates to C18:3 (maximum rate approximately 22–24 DPA), and detected as early as 11 DPA (Pollard et al., 2015).

A growing number of Camelina genes involved in FA and TAG biosynthesis are being identified (Abdullah et al., 2016; Morineau et al., 2017; Neumann et al., 2021; Nguyen et al., 2013; Ozseyhan et al., 2018). However, Camelina transcription factors (TFs) that control the expression of the corresponding enzymatic genes remain largely unknown. In higher plants, the synthesis of FA and TAG in seeds is tightly coordinated with development. In Arabidopsis, there is a growing number of TFs involved in seed development with direct or indirect effects on FA/TAG synthesis (Baud and Lepiniec, 2010; Le et al., 2010; Leprince et al., 2016; Tian et al., 2020). Major regulators include the ABI3VP1 proteins LEAFY COTYLEDON 2 (LEC2), ABSCISIC ACID INSENSITIVE 3 (ABI3), and FUSCA3 (FUS3), which besides controlling seed development-related processes are also positive regulators of FA/TAG synthesis (Bäumlein et al., 1994; Giraudat et al., 1992; Stone et al., 2001). Other important regulators include LEAFY COTYLEDON 1 (LEC1) and LEAFY COTYLEDON1-LIKE (L1L), which are CCAAT-HAP3 proteins (Kwong et al., 2003; Lotan et al., 1998), the basic leucine zipper 53 (bZIP53) (Alonso et al., 2009), AGAMOUS-Like15 (AGL15) (Zheng et al., 2009), the MYB proteins MYB115, MYB118, MYB107, and MYB9 (Lashbrooke et al., 2016; Troncoso-Ponce et al., 2016; Wang et al., 2009), and the homeobox GLABRA2 (Shen et al., 2006). Also, VP1/ABSCISIC ACID INSENSITIVE3-LIKE1, 2, and 3 (VAL1, 2, 3), all members of the ABI3VP1 family, are known for their roles in repressing the seed maturation program before germination (Guerriero et al., 2009; Suzuki and McCarty, 2008; Tsukagoshi et al., 2007). Downstream of some of these developmental regulators are several TFs that modulate specific aspects of lipid metabolism, including WRINKLED1 (WRI1), which controls carbon flux from sucrose to FA biosynthesis (Cernac and Benning, 2004). WRI1 is regulated at the transcriptional level by LEC1 and MYB89 and at the post-translational level by KIN10 and TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR 4 (TCP4) (Kong, Singh, et al., 2020; Li et al., 2017; Pelletier et al., 2017; Zhai et al., 2017). Some of these regulatory associations are conserved between species (Devic and Roscoe, 2016; Kong and Ma, 2018; Kong, Yang, et al., 2020). In Camelina, the overexpression of Arabidopsis MYB96 led to a significant increase in epicuticular and total wax (Lee et al., 2014), resembling the functions that MYB96 has in Arabidopsis under drought conditions (Seo et al., 2011). To what extent these regulatory networks are conserved between Arabidopsis and Camelina remains unknown.

Despite the close phylogenetic relationship of Arabidopsis and Camelina, they accumulate different quantities and types of seed oils (Li et al., 2006). Thus, understanding the regulatory processes associated with these differences provides opportunities for further enhancing seed oil production. Here, we describe the use of gene co-expression analyses to identify several TF candidates associated with the regulation of lipid biosynthesis in Camelina. These predictions were confirmed using DNA affinity purification followed by sequencing (DAP-seq) analysis of the corresponding TF candidates. Altogether, we identify and associate different TF candidates with specific aspects of lipid-related processes, including key players in regulating lipid accumulation during seed development in Camelina.

RESULTS

Expression analysis of genes involved in lipid accumulation during Camelina seed development

To complement the sparse gene expression information available for early stages of Camelina seed development

(Table S1), we collected seeds from Suneson Camelina plants at 5, 8, and 11 DPA. The sampling was performed for three biological replicates and RNA was extracted from seeds at the corresponding developmental stages and then used to perform RNA-seq analyses (see Methods). To characterize the expression of genes involved in lipid metabolism, we first collated lipid-related genes (LRGs) from CamRegBase (https://camregbase.org/) (Gomez-Cano et al., 2020) and classified them according to the information provided by AraLip (http://aralip.plantbiology.msu.edu/) (Li-Beisson et al., 2013). In accordance with these criteria, a total of 2765 Camelina LRGs were identified, which were then classified into 25 different groups according to their role in different aspects of lipid metabolism (Figure S1a and Table S2) and based on their homology to well-described Arabidopsis lipid regulators (Table S3).

We used publicly available developing-seed gene expression datasets and the RNA-seq information generated here from 5–11 DPA seeds (Table S1) to analyze mRNA accumulation patterns of the annotated LRGs. Overall, we identified four major types of genes based on their mRNA abundance during seed development. The smallest group (121 genes) corresponded to genes expressed at high levels (on average, 380 transcripts per million [TPM]) across all the developing-seed stages tested. These genes were largely associated with functions such as TAG and FA synthesis (Figure 1a,b). The second highest expressed group (on average, 20 TPM) consisted of 553 genes with predominant functions associated with lipid synthesis, desaturation, and export from plastids. Most of the genes were associated with two groups with medium-low (on average, 4.5 TPM; 5847 genes) and low expression (on average, 0.5 TPM, 1244 genes), primarily related to functions associated with the biosynthesis of membrane lipids, waxes, and suberins (Figure 1a,b).

Genes highly expressed in developing seeds corresponded to functions associated with (i) TAG synthesis, (ii) FA synthesis, and (iii) FA elongation & desaturation (Figure 1b), and this is why we analyzed the mRNA accumulation dynamics of these major groups of lipid metabolic genes across the various developmental stages. TAG synthesis genes peak at 18–29 DPA with expression values (TPM) several times (>10 times) higher than those of genes involved in the other two processes (Figure 1c). Partially consistent with metabolite data (Pollard et al., 2015), FA synthesis, elongation, and desaturation genes peak at 10–11 DPA (Figure 1c). The value of the newly added RNA-seq datasets (indicated with red in Figure 1a), particularly for 5 and 8 DPA, is evident from the high level of expression of several LRGs early during seed development (a few examples are indicated with asterisks in Figure 1a). Taken together, our analyses provide a comprehensive overview of the expression of LRGs during Camelina seed development, featuring specific gene sets with potential major lipid metabolism roles, providing an opportunity to uncover key regulators.

Identification of candidate lipid transcriptional regulators by co-expression analysis

To identify candidate genes encoding TFs potentially associated with the regulation of Camelina LRGs, we estimated the mutual information (MI) between each of the 5590 TFs annotated in CamRegBase and each gene in the genome using all the available Camelina gene expression data (Table S1). For each TF, we extracted the highest 200 genes (average MI $\geq$ 1) as corresponding to the co-expressed genes of the corresponding TF. We then evaluated whether LRGs were statistically overrepresented (false discovery rate [FDR] < 0.05, Fisher's exact test) within these 200 genes (Figure S1b). From the 5590 TFs analyzed, we identified 350 TFs that met the criteria. The 350 TFs belonged to 52 different TF families and those with the highest representation corresponded to MYB, AP2/ERF, bZIP, and bHLH families (Figure 1d; Table S4).

We compared our list of TF candidates with 36 Arabidopsis TFs known to participate in the regulation of lipid and/or seed development (Table S3). The 36 Arabidopsis TF corresponded to 105 Camelina homologous genes, as reported in CamRegBase (Gomez-Cano et al., 2020), consistent with the hexaploid nature of the Camelina genome. We excluded 10 out of the 105 Camelina TFs because of the absence of evidence for expression in the available Camelina expression data. We found a significant overlap between the TFs annotated by homology as Arabidopsis lipid regulators and those TFs predicted by our analysis (28 TFs overlapped, $P < 0.05$, hypergeometric test), providing confidence in our approach. These 28 TFs included homologs of WRI1, WRI4, ABI3, FUS3, LEC2, MYB9, MYB41, MYB107, MYB94, AGL15, VAL2, EEL, and DEWAX (Tables S3 and S4) (Bensmihen et al., 2002; Braybrook and Harada, 2008; Cernac and Benning, 2004; Focks and Benning, 1998; Go et al., 2014; Kosma et al., 2014; Lashbrooke et al., 2016; Lee et al., 2016; Lee and Suh, 2015; Meinke et al., 1994; Pouvreau et al., 2020; To et al., 2012; Tsukagoshi et al., 2007; Zhang et al., 2016; Zheng et al., 2009). Of note, not all Camelina paralogs were co-expressed with the same number of LRGs. For example, one of the three Camelina homologs of Arabidopsis AtMYB94, AtMYB41, AtVAL2, AtWRI4, and AtDEWAX were not co-expressed significantly with LRGs. Similarly, only one of the three Camelina paralogs of AtAGL15 and AtLEC2 were present in the list of 350 Camelina TFs (Figure S2a).

To prioritize Camelina TF candidates for functional studies, we ranked the 350 identified TFs based on the number of co-expressed LRGs (Table S4). Notably, the top candidates also showed preferential expression in seeds, as indicated by the seed Z-scores (Table S4; Figure S1c) (Kryuchkova-Mostacci and Robinson-Rechavi, 2017) (see

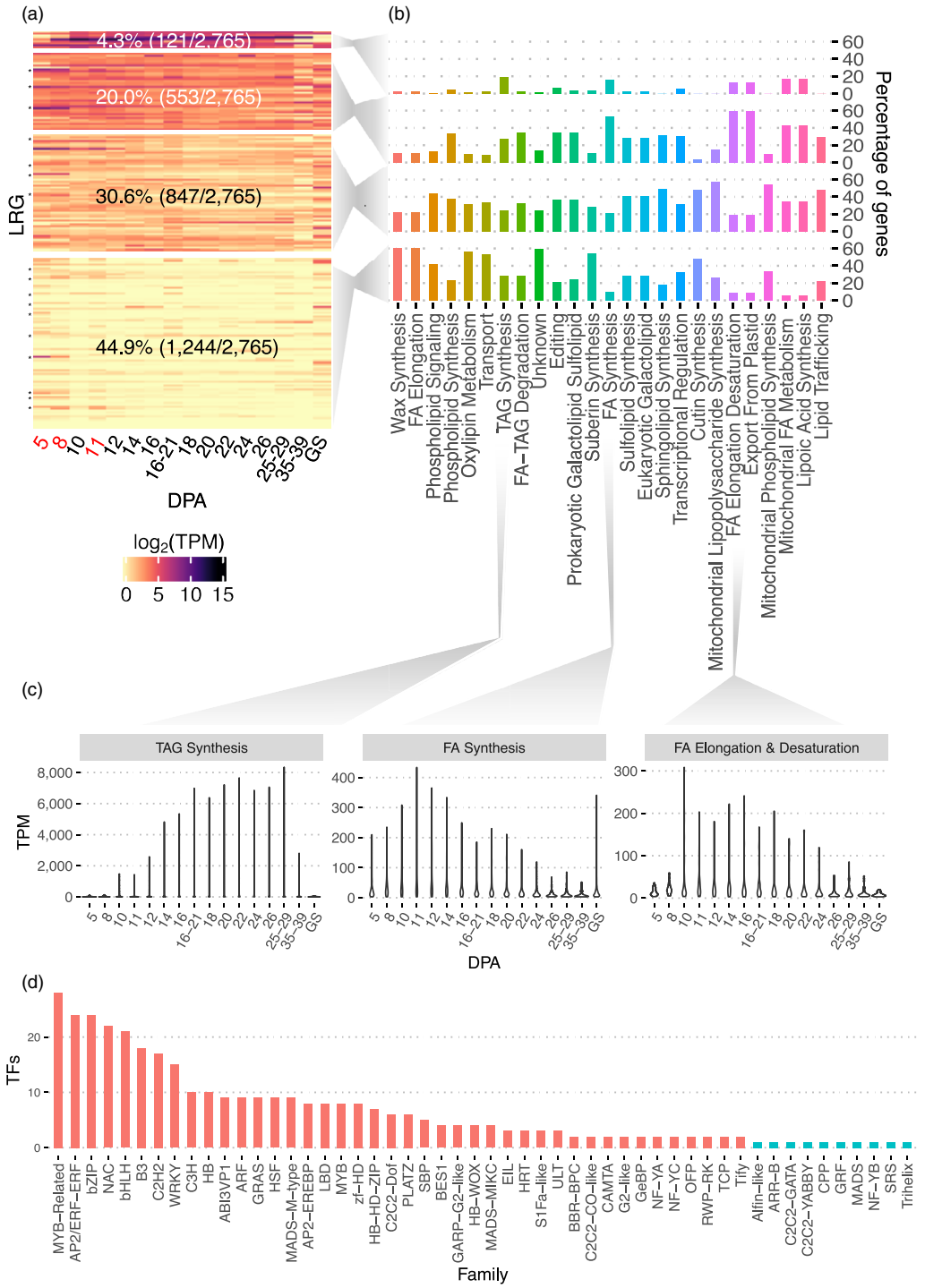


Figure 1. Expression dynamics of LRGs during seed development. (a) Heatmap representing mRNA accumulation information data highlighting four LRG clusters (rows). The clusters were generated based on the expression level of the corresponding genes at 16 timepoints across Camelina seed development including samples immediately after germination (columns). In total we analyzed 2765 LRGs collected from CamRegBase. DAP, days after pollination; GS, germinated seed. (b) Bar graph indicating the percentage of LRGs assigned to different lipid-related processes by each of the clusters of expression presented in (a). The lipid-related processes were defined based on homology with Arabidopsis and following the AraLip classification. (c) Expression variation across seed development of three major LRG groups. (d) Bar graph indicating the number of TFs classified by families identified as potential lipid metabolism regulators in Camelina. Red color indicates significantly enriched TF families (FDR < 0.05, Fisher's exact test).

Methods S1). From the ranked list (Table S4), we selected the top 35 TFs, which included 13 pairs of paralogs. From the paralog pairs, we selected only the TFs with the largest number of co-expressed LRGs and the highest expression, resulting in a final list of 22 TFs (Table 1) that were subjected to further analyses. Four of these TFs were homologs of known seed development and/or lipid metabolism regulators in Arabidopsis, corresponding to ABI3, FUS3, MYB9, and MYB107 (Table 1) (Giraudat et al., 1992; Keith et al., 1994; Lashbrooke et al., 2016).

To further characterize the TF candidates, we evaluated the conservation of the predicted TF–LRG associations between Camelina and Arabidopsis. For this, we re-analyzed >250 publicly available Arabidopsis RNA-seq experiments using identical pipeline and metrics as for Camelina, selecting datasets similar to the samples used for the Camelina co-expression analyses (Table S5). We focused specifically on our list of 22 Camelina TFs. Arabidopsis homologs of CsaMYB1 and CsaMYB3 were not expressed in the analyzed data and therefore excluded from this analysis. In total, within the remaining 20 TFs, 10 showed a conserved significant co-expression with LRGs (Figure S2b). Substantiating our analyses, the three well-described Arabidopsis lipid regulators AtABI3 (CsaABI3VP1-1), AtFUS3 (CsaABI3VP1-2), and AtMYB9 (CsaMYB2) were identified as part of the conserved co-expression associations. This co-expression analysis identified seven Camelina TFs (and their Arabidopsis homologs) that had not been previously associated with lipid metabolism, including CsaNAC1, CsaNAC2, Csazf-HD1, CsaB3-1, CsaAP2/B3-like-1, CsaULT1, and CsaLBD1 (Figure S2b). The remaining 10 Camelina TFs that did not show conserved co-expression with Arabidopsis LRGs are likely to correspond to Camelina-specific lipid regulators, or alternatively they are not involved in the control of lipid metabolism.

Establishing the DNA-binding landscape of the candidate transcription factors

To further characterize the 22 TFs and to identify potential target genes, we applied DAP-seq (O'Malley et al., 2016). We synthesized and cloned the corresponding open reading frames (ORFs) for the 22 TFs in a vector that permitted expression of the protein fused at the N-terminus to a Halo-tag (Bartlett et al., 2017). We also generated a Camelina unmethylated DAP-seq DNA library (ampDAP-seq) from green tissues of mature plants (see Methods). We reasoned that unmethylated DNA better captures the majority of the protein–DNA interactions (PDIs) in which these TFs are likely to participate (O'Malley et al., 2016) and eliminates variations in methylation patterns between cell types or tissues. We performed DAP-seq in duplicate for each Halo-TF, and with the Halo-tag alone as the control. We obtained on average 25.5 million reads per sample, out of which about half mapped uniquely to the available Camelina genome (v2, cv. DH55) (Kagale et al., 2014) (Table S6). To assess the variance and reproducibility of the experiments, we performed a principal component analysis (PCA) using uniquely mapped reads. The first two principal components (PCs) showed all TFs

Table 1 List of the top candidate TFs identified

TF gene ID	TF name in CamRegBase	At homolog gene ID	At homolog gene name	LRGs in top 200	TF family
Csa08g011090	CsaC3H2	At5g18550	–	55	C3H
Csa18g003140	CsaHB2	At5g46880	*AtHB7*	49	HB
Csa13g011780	CsabZIP1	At5g10030	*TGA4*	45	bZIP
Csa02g060070	CsaHRT1	At5g56780	*ET2*	44	HRT
Csa11g072230	Csazf-HD1	At5g42780	*ZHD13/RHD1*	43	zf-HD
Csa13g021450	CsaC3H1	At5g18550	–	40	C3H
Csa19g036630	CsaABI3VP1-1	At3g24650	*ABI3*	37	ABI3VP1
Csa04g012470	CsaABI3VP1-2	At3g26790	*FUS3*	35	ABI3VP1
Csa10g008930	CsaAP2/B3-like-1	At4g33280	*REM16*	35	AP2/B3-like
Csa11g015700	CsaULT1	At4g28190	*ULT1*	35	ULT
Csa02g073390	CsaC2C2-Dof1	At5g65590	*SCAP1*	33	C2C2-Dof
Csa17g034220	CsaB3-1	At1g26680	–	33	B3
Csa04g050690	CsaS1Fa-like-1	At2g37120	–	31	S1Fa-like
Csa09g088620	CsaNAC1	At2g24430	*NAC38*	31	NAC
Csa13g019570	CsaMYB1	At5g16770	*MYB9*	30	MYB
Csa06g004330	CsaWRKY1	At2g03340	*WRKY3*	30	WRKY
Csa04g023980	CsaNAC2	At3g44290	*NAC60*	29	NAC
Csa19g005080	CsaMYB2	At3g02940	*MYB107*	29	MYB
Csa11g037200	CsaTify1	At4g14720	*PPD2*	29	Tify
Csa01g013850	CsaMYB3	At3g12720	*MYB67*	29	MYB
Csa06g051300	CsaHB1	At2g44910	*HB4*	28	HB
Csa05g079950	CsaLBD1	At1g67100	*LBD40*	27	LBD

well separated from the control (Halo) (Figure S3a). However, we also observed five TFs with strikingly different replicates, indicating low reproducibility between them. For each TF, we also analyzed the similarity of the uniquely mapped reads between each pair of replicates (Figure S3b), which confirmed the differences observed in the PCA for replicates of the five TFs (Figure S3a). Based on these observations, we discarded the DAP-seq results obtained for CsaABI3VP1-1 and CsaB3-1 (because of its high correlation with the Halo control) and settled on analyzing the replicates of CsaMYB2, CsaULT1, and CsaTify1 independently (replicates with Pearson correlation coefficient [PCC] < 0.7) (Figure S3b). For the remaining 17 TFs, DNA-binding regions (peaks) were called using both replicates. Thus, in total, 20 TFs were tested for the presence of peaks. The number of identified peaks varied greatly between the TFs, with CsaC2C2-Dof1 showing >100 000 peaks and four TFs having <500 peaks (CsaTify1, CsaS1Falike-1, CsaMYB2, and CsaULT1) (Figure S4a), which were not further used. In consequence, a total of 16 TFs were kept for further analyses.

The analysis of the distance between the peak summit and the closest annotated transcription start sites (TSSs) indicated that, on average, 63% of the total peak summits are within 3 kbs of the TSSs (Figure S4b). Thus, our results are in agreement with the peak genomic distribution patterns previously observed in DAP-seq experiments for Arabidopsis and maize (*Zea mays*) (Galli et al., 2018; O'Malley et al., 2016). We compared, in terms of successful identification of TF-binding motifs (TFBMs), all our DAP-seq results (including those which failed to pass the quality controls) with those performed in Arabidopsis (O'Malley et al., 2016) and determined that 17 common TFs were tested (TF homologs). To note, 3/17 TFs did not work in either plant, 7/17 TFs worked in Camelina but not in Arabidopsis, and 6/17 TFs worked in both plants (Figure S5a). The remaining TF (AtMYB107, homolog of CsaMYB2) worked only in Arabidopsis, likely related to the lack of MYB domains on the Camelina-annotated transcript (Figure S5b). Finally, the corresponding genes for CsaMYB1, CsaNAC2, and CsaC3H2 were not previously tested in Arabidopsis. In summary, we provide here high-confidence DNA-binding data for 16 TFs, of which 10 were previously unknown in Arabidopsis.

To evaluate the quality of the predicted DAP-seq peaks of the corresponding 16 TFs, we determined the $\log_2$ fold change of the binding ($\log_2$FC, Methods S1). We defined high-confidence peaks for further analyses as those showing $\log_2$FC > 0.5 in both replicates, which represented approximately 32.5% of the total peaks called (Figure S6). One additional criterion that we applied to decide whether DAP-seq provided meaningful information or not was the enrichment for particular TFBMs within the recovered peaks, a widely accepted characteristic of the DNA fragments recognized by TFs (Lambert et al., 2018). To identify the TFBMs associated with each TF, we ranked all the high-quality peaks based on their $\log_2$FC, selected the top 1000 peaks for each TF, and identified the motif consensus using MEME-ChIP (Machanick & Bailey, 2011). To evaluate the relevance of the predicted TFBMs in the context of the identified peaks, we searched each TFBM across the full set of peaks for each TF, focusing on two specific aspects: (i) the fraction of peaks that harbored the motif and (ii) the localization of the motif within the peak (distance to the summit). We carried out this analysis by extending each peak 50 bps around the summit (Figure S7). The most significant motifs identified for each of the 16 TFs corresponded to those with the largest abundance and which displayed a clear accumulation close to the summit of each peak (Motif 1 in Figure S7). Thus, for the rest of this study, we considered as high-confidence peaks those that harbored such a motif, corresponding to approximately 92% of all the peaks evaluated.

We compared the DNA-binding specificities provided by DAP-seq between the corresponding six Camelina and Arabidopsis homologs (Figure S5; Table S7). We re-analyzed all six Arabidopsis DAP-seq using the same pipeline employed in the current study. Five TF pairs (AtNAC38 and CsaNAC1; AtFUS3 and CsaABI3VP1-2; AtMYB67 and CsaMYB3; AtTGA4 and CsabZIP1; AtWRKY3 and CsaWRKY1) showed almost identical DNA-binding preferences (Figure S8), suggesting that the amino acid residues that distinguish the Arabidopsis and Camelina homologs are not significantly affecting *in vitro* DNA-binding specificities. The only exception was CsaAP2/B3-like-1, for which none of the top motifs identified (Figure S8d) matched the TTTGGCGGGAA sequence consensus predicted for AtREM1 (Figure S8d). This result puzzled us, hence we decided to re-check if the Arabidopsis and Camelina genes were properly annotated. Indeed, we determined that one of the B3 domains that characterizes the DNA-binding domain of AtREM1 (Romanel et al., 2009) was absent in the cloned CsaAP2/B3-like-1 ORF, because of a likely error in the current Camelina genome annotation (Figure S9). Taken together, we identified the DNA-binding patterns for 16 Camelina TFs and determined a similar correspondence with the Arabidopsis homolog, when available.

Predicting gene targets for the selected TFs

To identify potential gene targets for the 16 TF candidates, we determined which genes were located within 3 kbps of each high-confidence peak summit, since 3 kbps capture many of the biologically relevant TF–target gene interactions (Springer et al., 2019). We identified a total of 31 898 potential targets for these 16 TFs, with CsaMYB1 and CsaHRT1 showing the largest (6816 genes) and lowest (nine genes) number of target genes (Table S8), respectively. As a first step towards assessing the biological

significance of the DAP-seq results and its concordance with the co-expression prediction, we tested if the predicted targets were enriched in LRGs and/or in TFs associated with the control of LRGs and seed development (TF–LRG/development) (Table S3). Tellingly, 4/16 and 6/16 sets of targets showed significant enrichment ($P \leq 0.05$, Fisher's exact test) on LRGs and TF–LRG/development targets, respectively (Figure 2a). Moreover, CsaABI3VP1-2 (Camelina homolog of AtFUS3) showed enrichment on both sets of genes, suggesting an important role in lipid metabolism. Thus, in total, nine out of the 16 TFs tested showed a significant enrichment for target genes associated with lipid metabolism in Camelina.

Previously, we showed that half of the candidate TF homologs in Arabidopsis were enriched in LRGs by co-expression (Figure S2b). Thus, we tested if they were also enriched in target genes annotated as LRGs, as we found for the Camelina TFs (Figure 2a). We performed the analysis with the six Arabidopsis TFs for which we previously evaluated TFBMs (Figure S8). The analysis of the Arabidopsis DAP-seq data was performed using the same pipeline and controls as we used for the Camelina data. Out of the six Arabidopsis TFs, five showed enrichment for target genes annotated as LRGs and TF–LRG/development (Figure 2b). This finding, along with the conservation of the corresponding TFBMs (Figure S8), suggests conservation of the corresponding regulatory functions. Curiously, the five Arabidopsis TFs showed target enrichment for both types of genes: LRGs and TF–LRGs (Figure 2b). This contrasts with what we found for the corresponding Camelina homologs which showed in all cases but one either enrichment for LRGs or TF–LRGs but not in both (Figure 2a). Finally, neither CsaAP2/B3-like-1 nor AtREM16 showed targets enriched in LRGs or TF–LRG/development. While this is consistent with the possibility that we used a truncated protein for CsaAP2/B3-like-1, our results suggest that AtREM16 plays a secondary role as a lipid metabolism regulator.

To characterize other functional roles of the set of predicted target genes associated with the corresponding TFs, we investigated enrichment for Gene Ontology (GO) terms. All the TFs tested have at least one GO term enriched that is lipid-related (Table S9). After removing redundant and general terms (Figure S10, selected GO terms within dashed gray line), we clustered all the TFs based on the top 10 GOs for each (based on *P*-values), allowing us to separate them into two main clusters. One cluster (indicated in green, Figure S11) was associated with a wide range of GO terms, including regulation of development and several metabolic processes, particularly lipid metabolism-related functions, as well as phenylpropanoid and carboxylic acid biosynthesis processes. Members of the other cluster (indicated in orange, Figure S11) have in common the terms signal transduction, defense responses, regulation of gene expression, and regulation of nitrogen compounds. When all the data are considered together, these analyses provide additional evidence that the DAP-seq results bore biologically meaningful targets and support the initial co-expression predictions, including the discovery of previously unrecognized candidate regulators of LRGs.

Similarities in the functional annotation of target genes among TFs may indicate that the corresponding TFs share common targets. Alternatively, the TFs could regulate different genes in the same process/pathway. To distinguish between these two possibilities, we evaluated the overlap in targets among the 16 TFs. Almost half of the comparisons showed significant target overlaps ($P < 0.05$, Fisher's exact test) (Figure 2c, darker colors indicate smaller *P*-values). As anticipated, TFs from the same family (CsaMYB1 and CsaMYB3; CsaNAC1 and CsaNAC2) had the largest number of shared target genes, likely driven by the very similar *in vitro* DNA-binding consensus of the corresponding TFs. Of note, while significant, the overlap comprises only a subset of all the targets for each of these TFs, suggesting that outside the shared core motif, each TF has specific DNA-binding preferences (Figure S7). Many of the TF pairs have overlapping targets (e.g., CsaMYB1 and CsaMYB3; CsaHB1, CsaABI3VP1-2, and CsaC2C2-Dof1; CsaNAC1, CsaNAC2, and CsaAP2/B3-like-1), indicating that they function in the control of related biological processes. We explored this hypothesis by comparing two of the non-homologous TF pairs with the highest number of common targets, corresponding to CsaMYB1-CsaWRKY1 and CsabZIP1-CsaHB1, which had 922 and 468 common targets, respectively. For the CsaMYB1–CsaWRKY1 pair, we found that shared targets were enriched in multiple lipid-related GO terms at several levels of the GO hierarchy, including carboxylic acid biosynthesis and VLCFA biosynthesis (Figure S12a). Contrary to the pattern observed for CsaMYB1–CsaWRKY1, common targets of CsabZIP1 and CsaHB1 were enriched in a more diverse list of biological processes not observed on the corresponding individual list of enriched GO terms, including flavone biosynthesis, regulation of transcription, activation of protein kinase activity, root hair cell tip growth, and leaf senescence (Figure S12b), suggesting that their role in lipid metabolism control is not linked to common target genes in the pathway.

To further understand the potential participation of CsaMYB1 and CsaWRKY1 in gene co-regulation of their common targets, we evaluated the distribution of binding sites in the 922 shared targets. For most of them, the binding sites were within a few hundred base pairs apart from each other (the average distance was 320 bps; Figure 2d), highlighting a possible cooperative work at the DNA level (post-DNA binding) (Reiter et al., 2017). The proximity and potential significance for transcriptional regulation is

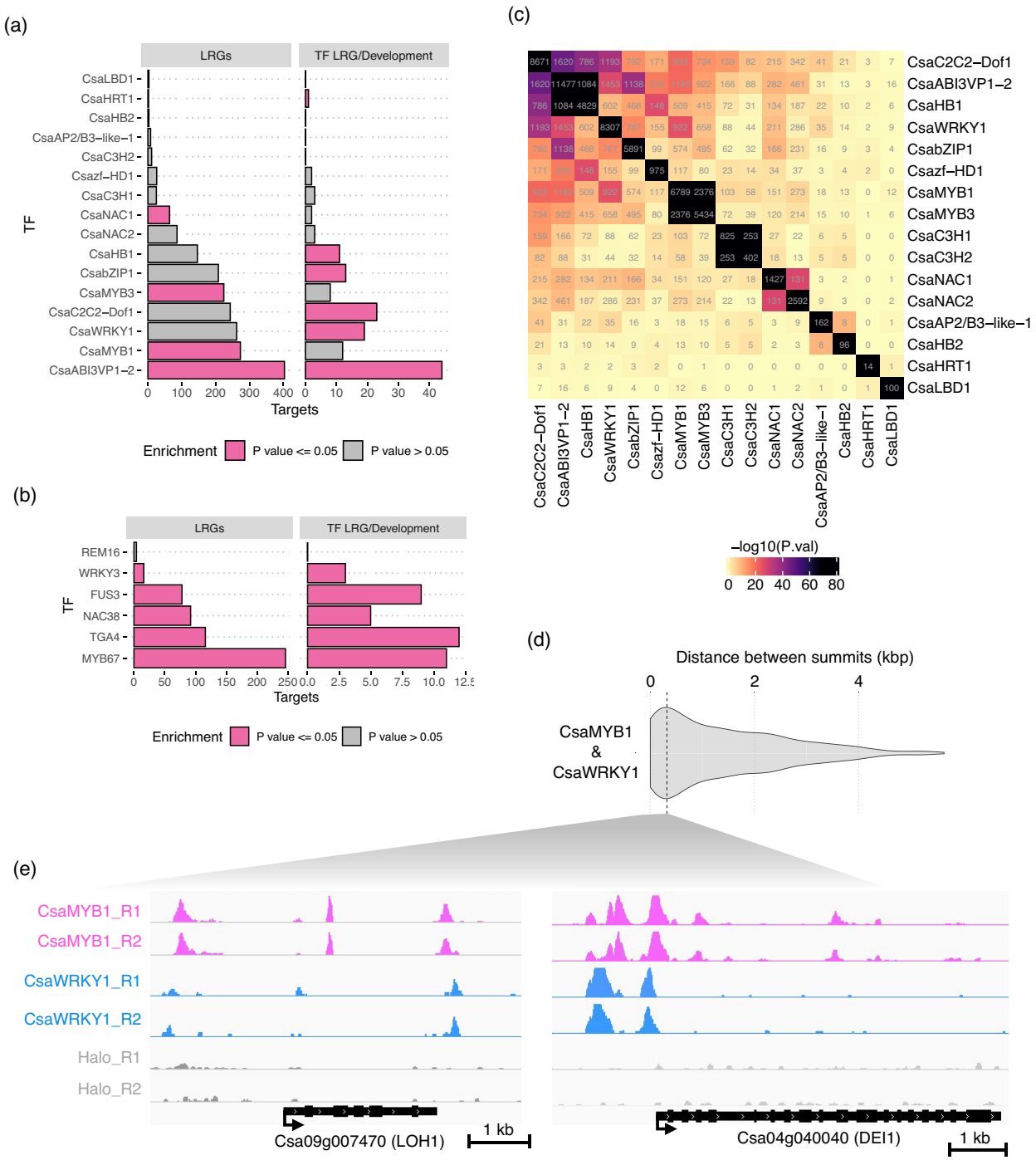


Figure 2. Regulatory landscape of predicted lipid-related regulators based on DAP-seq. The bar graph indicates the number of predicted target genes annotated as (a) Camelina or (b) Arabidopsis (b) LRGs. Arabidopsis TFs correspond to homologs of the Camelina predicted candidates. The red color indicates the significance of the overlap of target genes annotated as LRG versus the total number of predicted target genes for each of the tested TFs ($P \leq 0.05$, Fisher's exact test). (c) Heatmap indicating the number of common targets between pairs of TFs. Color scale indicates the *P*-value associated with the corresponding number of common targets (Fisher's exact test). (d) Violin plot showing the distribution of distances (in bps) between summits (peak centers) of CsaMYB1 and CsaWRKY1 mapped to common targets. The vertical dashed line indicates the most frequent distance between summits. (e) IGV plots with co-binding profiles (peak) generated from the DAP-seq experiments of CsaMYB1 and CsaWRKY1 highlighting two shared targets with the respective gene models obtained from Camelina V2.0 at the bottom. Peak heights correspond to the number of reads by bins (10 bp) per million mapped reads.

exemplified by the two shared targets Csa03g002110 and Csa04g040040 (Figure 2e), Arabidopsis homologs of 3-KETOACYL-COA SYNTHASE (KCS1, At1g01120) and PASTICCINO 1 (PAS1/DEI1, At3g54010), which are involved in FA and VLCFA synthesis (Roudier et al., 2010; Shang et al., 2016), respectively, further underscoring the potential regulatory role of CsaMYB1 and CsaWRKY1 in lipid metabolism.

Identified TFs associate with distinct aspects of lipid metabolism

To better understand the specific aspects of lipid metabolism that each of the identified TFs might be involved in, we scored how many targets of each TF corresponded to each of the lipid pathway categories (as presented in Figure 1b). In total, 11/16 TFs showed significant enrichment for targets annotated across several lipid-related processes ($P < 0.05$, Fisher's exact test) (Table S10). As examples, CsaMYB3, CsaMYB1, CsaWRKY1, and CsaABI3VP1-2 were enriched in more than four different processes, with their top target processes being suberin synthesis (18.2%), cutin synthesis (25.3%), and transcriptional regulation (18.1% and 41.9%), respectively (Table S10). Remarkably, several combinations of TFs showed significant enrichment for the same processes. Finally, we also observed that the targets for CsaABI3VP1-2 and CsaWRKY1 were significantly enriched in genes associated with TAG synthesis (22.6%) and FA and TAG degradation (15.7%), respectively, which are core processes in the accumulation of seed oil.

In parallel, to evaluate the biological significance of the regulatory interactions predicted at the pathway co-expression level, we applied the gene set enrichment analysis (GSEA) algorithm (Subramanian et al., 2005) using the PCC as the scoring metric. Thus, significant positive and negative enrichment values indicate association of the corresponding TF with a metabolic pathway in a positive or negative fashion, respectively. Also, under these conditions, GSEA permits the identification of TF–process relationships that have significant co-expression signals at the pathway level rather than as individual target genes (Subramanian et al., 2005). Eight out of the 16 TFs tested showed significant enrichment ($P < 0.05$) for at least one of the processes tested (Figure S13). CsaABI3VP1-2 showed the largest number of significant associations (up to 10), including FA elongation and desaturation, FA and TAG synthesis, and transcriptional regulation (Figure S13a). The second and third TFs with most enriched processes were CsaWRYK1 and CsaMYB1, with seven each (Figure S13b, c). We also observed 11 TF–process associations with negative enrichment scores, indicating enrichment for negative co-expression values, within which CsaWRKY1, CsaMYB1, CsaMYB3, and CsabZIP1 are included (Figure S13b–e). The former showed enrichment for negative scores on its corresponding targets annotated under FA synthesis, transcriptional regulation, and transport (Figure S13b) and the latter with targets annotated under cutin synthesis, wax synthesis, and FA elongation (Figure S13e). These results suggest major roles of these TFs as negative regulators of the mentioned pathways.

Finally, we combined both sets of results (target enrichment and GSEA results) to identify high-confidence TF–process associations. Six of the 11 TFs analyzed showed significant associations in both tests with at least six different processes, to a total of 10 TF–pathway associations (pink edges in Figure 3). Transcriptional regulation was the process with the largest number of connections. CsaNAC1, CsaWRKY1, and CsaABI3VP1-2 were the three TFs with the largest number of associations (two for each of them, Figure 3). Three out of the 10 TF–pathway associations showed significant negative enrichment (Figure 3 and S13), indicating transcriptional repression roles of the corresponding TFs on the respective pathways. Also, it is worth noting that one of the main processes enriched for the targets of CsaMYB1 and CsaNAC2 was cutin synthesis (Figure 3; Table S10). CsaABI3VP1-2 was the only TF significantly enriched in TAG synthesis- and transcriptional regulation-related targets (Figure 3), and remarkably we found that the large majority of the targets that we predicted for CsaABI3VP1-2 were also TF targets previously identified for AtFUS3 by either chromatin immunoprecipitation-DNA microarray (ChIP-chip) (Wang & Perry, 2013) or DAP-seq assays (O'Malley et al., 2016) (Figure S14), uncovering potential Camelina-specific interactions as well as unreported Arabidopsis targets. Altogether, these analyses underscore CsaABI3VP1-2 as a good candidate playing a major role in lipid metabolism in Camelina, similar to AtFUS3 (Yamamoto et al., 2010; Wang & Perry, 2013; Zhang et al., 2016).

Dynamic behavior of the predicted networks during seed development

To gain further insights into the regulatory effect of the identified TF–target interactions in Camelina seeds, we performed a second co-expression analysis with GENIE3 (Huynh-Thu et al., 2010) using only expression data from seeds. GENIE3 uses a regression tree and random forest algorithm to make regulatory prediction implying causality (Huynh-Thu et al., 2010). Thus, we assumed that predictions identified by GENIE3 and supported by DAP-seq are highly confident regulatory interactions occurring specifically in seeds (Table S11). The significance of the predicted score was assayed using a permutations test (FDR ≤ 0.001, 1000 permutations). Overall, 35% of the targets identified by DAP-seq were also predicted as targets of the corresponding TFs by GENEI3 (Figure 4a). The highest percentage of DAP-seq seed co-expressed targets was observed for CsaNAC2 and CsabZIP1 (approximately 54% each, Figure 4a). These results suggest that many of the predicted

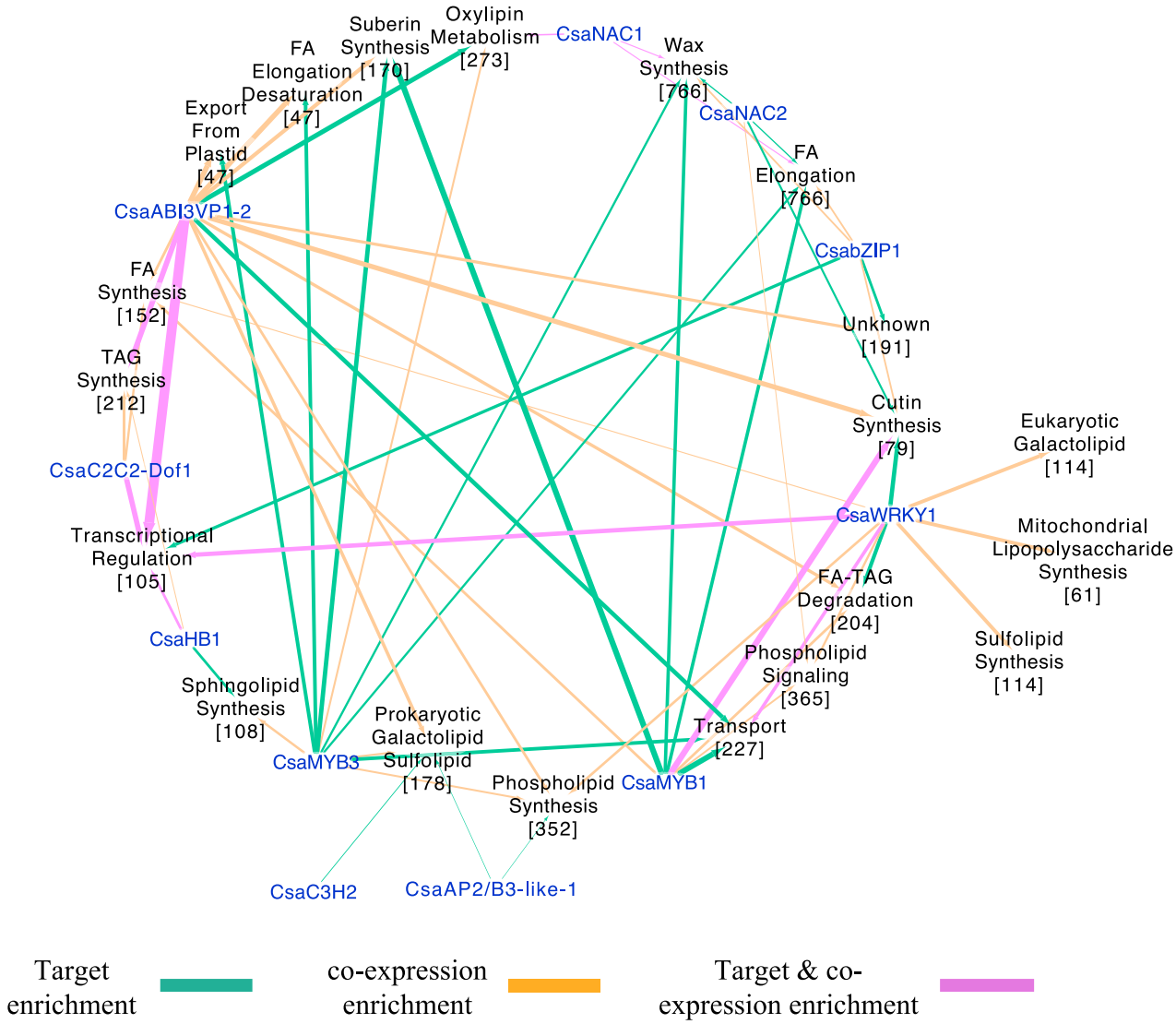
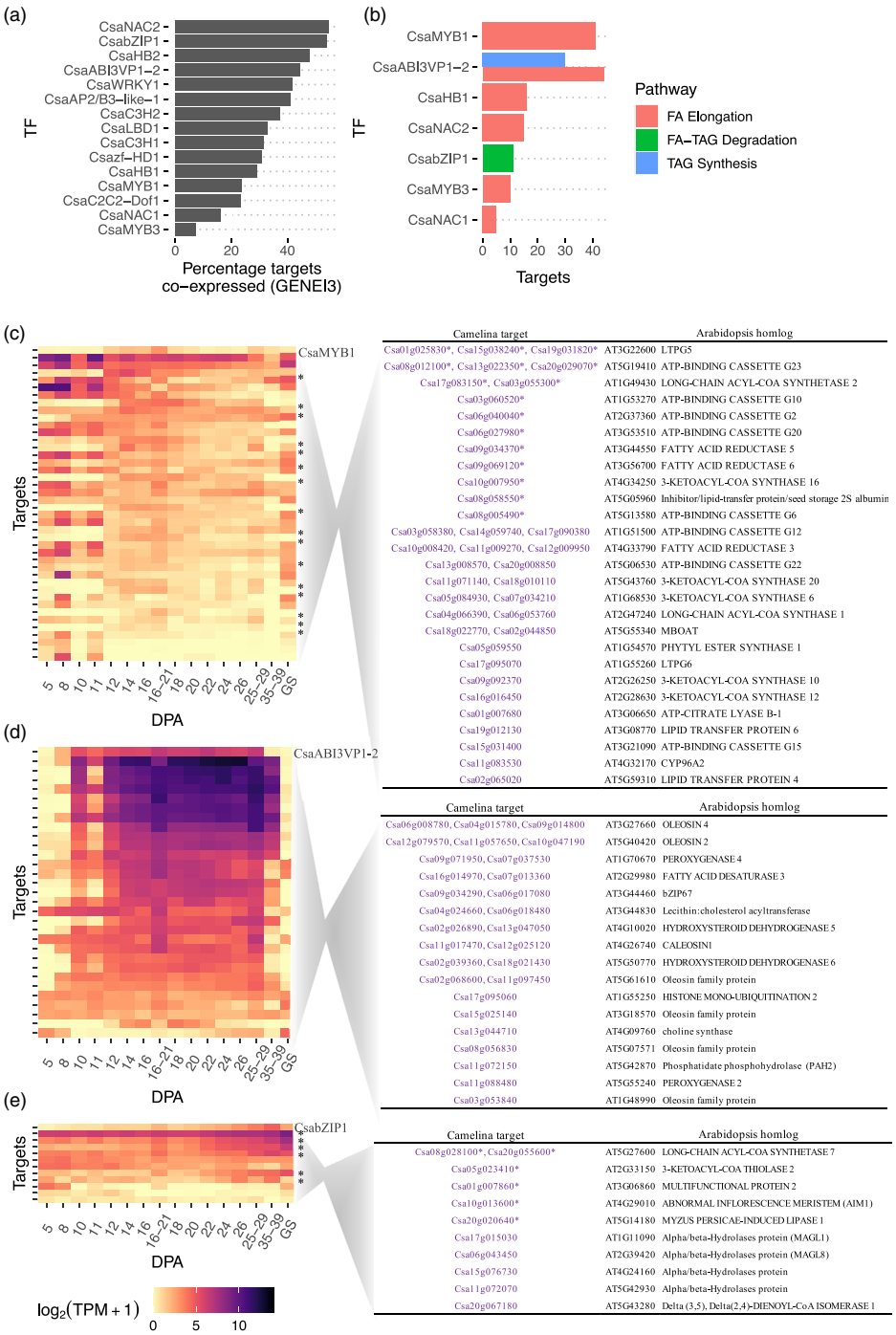


Figure 3. High-confidence TF–process network. Associations predicted based on target enrichment and GSEA using the TF–target PCC as score metric are indicated by lines joining TFs (blue) and specific processes associated with lipid metabolism (black). The thickness of the edges represents the fraction of lipid-related genes in the pathway that is being targeted by the corresponding TF. The total number of genes annotated for each of the corresponding lipid-related processes is indicated inside square brackets.

TF–target associations have a regulatory effect in the context of seed development.

To parse TFs involved in controlling FA- and TAG-related genes in seeds, we combined the target enrichment and the GSEA results (Figure 3) to select TFs associated with the corresponding pathways. Consequently, we reduced the TF–target DAP-seq network to only targets co-expressed in seeds (as predicted by GENIE3) (Figure 4a). With this subset of TF–target interactions, we tested the enrichment for targets on the corresponding pathways once again to determine if the reduced TF–target network still had a significant number of targets associated with FA- and TAG-related processes. Seven TFs showed enrichment for seed co-expressed targets associated with at least three different pathways (FDR < 0.05, Fisher's exact test) (Figure 4b). FA elongation was the pathway most frequently targeted, with six different TFs associated with it (Figure 4b). CsaABI3VP1-2 and CsaMYB1 were the two TFs with most seed co-expressed targets annotated under FA elongation. However, TAG synthesis and FA and TAG degradation were significantly targeted by just one TF each, CsaABI3VP1-2 and CsabZIP1, respectively (Figure 4b).

We selected three of seven TF–pathway interactions (CsaMYB1 & FA elongation, CsaABI3VP1 & TAG synthesis, and CsabZIP1 & FA/TAG degradation, Figure 4b) to analyze the expression dynamics of the corresponding TFs and targets during seed development. CsaMYB1 showed two



Camelina target	Arabidopsis homlog	
Csa01g025830*, Csa15g038240*, Csa19g031820*	AT3G22600	LTPG5
Csa08g012100*, Csa13g022350*, Csa20g029070*	AT5G19410	ATP-BINDING CASSETTE G23
Csa17g083150*, Csa03g055300*	AT1G49430	LONG-CHAIN ACYL-COA SYNTHETASE 2
Csa03g060520*	AT1G53270	ATP-BINDING CASSETTE G10
Csa06g040040*	AT2G37360	ATP-BINDING CASSETTE G2
Csa06g027980*	AT3G53510	ATP-BINDING CASSETTE G20
Csa09g034370*	AT3G44550	FATTY ACID REDUCTASE 5
Csa09g069120*	AT3G56700	FATTY ACID REDUCTASE 6
Csa10g007950*	AT4G34250	3-KETOACYL-COA SYNTHASE 16
Csa08g058550*	AT5G05960	Inhibitor/lipid-transfer protein/seed storage 2S albumin
Csa08g005490*	AT5G13580	ATP-BINDING CASSETTE G6
Csa03g058380, Csa14g059740, Csa17g090380	AT1G51500	ATP-BINDING CASSETTE G12
Csa10g008420, Csa11g009270, Csa12g009950	AT4G33790	FATTY ACID REDUCTASE 3
Csa13g008570, Csa20g008850	AT5G06530	ATP-BINDING CASSETTE G22
Csa11g071140, Csa18g010110	AT5G43760	3-KETOACYL-COA SYNTHASE 20
Csa05g084930, Csa07g034210	AT1G68530	3-KETOACYL-COA SYNTHASE 6
Csa04g066390, Csa06g053760	AT2G47240	LONG-CHAIN ACYL-COA SYNTHASE 1
Csa18g022770, Csa02g044850	AT5G55340	MBOAT
Csa05g059550	AT1G54570	PHYTYL ESTER SYNTHASE 1
Csa17g095070	AT1G55260	LTPG6
Csa09g092370	AT2G26250	3-KETOACYL-COA SYNTHASE 10
Csa16g016450	AT2G28630	3-KETOACYL-COA SYNTHASE 12
Csa01g007680	AT3G06650	ATP-CITRATE LYASE B-1
Csa19g012130	AT3G08770	LIPID TRANSFER PROTEIN 6
Csa15g031400	AT3G21090	ATP-BINDING CASSETTE G15
Csa11g083530	AT4G32170	CYP96A2
Csa02g065020	AT5G59310	LIPID TRANSFER PROTEIN 4

Camelina target	Arabidopsis homlog	
Csa06g008780, Csa04g015780, Csa09g014800	AT3G27660	OLEOSIN 4
Csa12g079570, Csa11g057650, Csa10g047190	AT5G40420	OLEOSIN 2
Csa09g071950, Csa07g037530	AT1G70670	PEROXYGENASE 4
Csa16g014970, Csa07g013360	AT2G29980	FATTY ACID DESATURASE 3
Csa09g034290, Csa06g017080	AT3G44460	bZIP67
Csa04g024660, Csa06g018480	AT3G44830	Lecithin:cholesterol acyltransferase
Csa02g026890, Csa13g047050	AT4G10020	HYDROXYSTEROID DEHYDROGENASE 5
Csa11g017470, Csa12g025120	AT4G26740	CALEOSIN1
Csa02g039360, Csa18g021430	AT5G50770	HYDROXYSTEROID DEHYDROGENASE 6
Csa02g068600, Csa11g097450	AT5G61610	Oleosin family protein
Csa17g095060	AT1G55250	HISTONE MONO-UBIQUITINATION 2
Csa15g025140	AT3G18570	Oleosin family protein
Csa13g044710	AT4G09760	choline synthase
Csa08g056830	AT5G07571	Oleosin family protein
Csa11g072150	AT5G42870	Phosphatidate phosphohydrolase (PAH2)
Csa11g088480	AT5G55240	PEROXYGENASE 2
Csa03g053840	AT1G48990	Oleosin family protein

Camelina target	Arabidopsis homlog	
Csa08g028100*, Csa20g055600*	AT5G27600	LONG-CHAIN ACYL-COA SYNTHETASE 7
Csa05g023410*	AT2G33150	3-KETOACYL-COA THIOLASE 2
Csa01g007860*	AT3G06860	MULTIFUNCTIONAL PROTEIN 2
Csa10g013600*	AT4G29010	ABNORMAL INFLORESCENCE MERISTEM (AIM1)
Csa20g020640*	AT5G14180	MYZUS PERSICAE-INDUCED LIPASE 1
Csa17g015030	AT1G11090	Alpha/beta-Hydrolases protein (MAGL1)
Csa06g043450	AT2G39420	Alpha/beta-Hydrolases protein (MAGL8)
Csa15g076730	AT4G24160	Alpha/beta-Hydrolases protein
Csa11g072070	AT5G42930	Alpha/beta-Hydrolases protein
Csa20g067180	AT5G43280	Delta (3,5), Delta(2,4)-DIENOYL-CoA ISOMERASE 1

Figure 4. Integration of seed co-expression and DNA-binding information. (a) Bar graph indicating the percentage of DAP-seq targets supported by the co-expression associations predicted with GENEI3 using seed expression data. (b) Bar graph of the seven most significant TF–lipid-related process interactions that passed the enrichment test after incorporation of seed co-expressed targets. For each TF, the total number of target genes annotated for the corresponding FA- and TAG-related processes is indicated. (c) Heatmap representing the expression dynamics of targets of CsaMYB1 associated with FA elongation during seed development. (d) Heatmap representing the expression dynamics of targets of CsaABI3VP1 associated with TAG synthesis during seed development. (e) Heatmap representing the expression dynamics of targets of CsabZIP1 associated with FA and TAG degradation during seed development. The right panel lists the gene IDs for the Camelina genes represented in the heatmaps and the gene IDs for the corresponding Arabidopsis homologs.

expression windows, one at 12–21 DPA and a second at 25–29 DPA (Figure 4c). These expression profiles are in concordance with the reported peaks of FA synthesis and TAG accumulation (11–24 DPA) (Pollard et al., 2015). CsaABI3VP1-2 showed a broader expression window, starting at 8 DPA with constant expression until 25–29 DPA

(Figure 4d). Finally, CsabZIP1 is mainly expressed during the later stages of seed development (expression peak approximately 35–39 DPA) (Figure 4e), consistent with the expected pattern for controlling FA and TAG degradation right before seed germination.

As for the corresponding target genes, we observed several expression patterns consistent with activation or repression by the respective TFs, as exemplified for the targets of CsaMYB1 and CsabZIP1 (Figure 4c,e). Within the set of the CsabZIP1 targets, it is worth mentioning that multiple Arabidopsis homologs are involved in FA beta-oxidation during seed germination (Footitt et al., 2006; Fulda et al., 2004; Jiang et al., 2011; Richmond and Bleecker, 1999) and homeostasis of phospholipid and neutral lipids (Ghosh et al., 2009) (Figure 4e). Finally, most (28/30) of the CsaABI3VP1-2 targets showed a similar expression to the corresponding TF (Figure 4d). Within these targets, it is worth noting the Csa16g014970/Csa07g013360 and Csa09g034290/Csa06g017080 gene pairs, which are homologs of Arabidopsis FATTY ACID DESATURASE 3 (AtFAD3) (At2g29980) and AtbZIP67 (At3g44460), respectively. The former is involved in linolenic acid synthesis (O'Neill et al., 2011), while the latter is a known regulator of AtFAD3 (Mendes et al., 2013), highlighting a potential feedforward loop between CsaABI3VP1-2, Csa09g034290/Csa06g017080, and Csa16g014970/Csa07g013360 in Camelina.

DISCUSSION

Camelina is an oilseed crop that has emerged as a prominent feedstock for biofuels and industrial oils during the past decade. Its polyploid genome makes it challenging to identify genes involved in the biosynthesis or regulation of seed oils by classical loss-of-function approaches. The homology to Arabidopsis has permitted to translate knowledge gained in this model plant to Camelina, exemplified in the manipulation of epicuticular and total wax production by the overexpression of AtMYB96 (Lee et al., 2014) or in the increase of seed oil by the overexpression of Arabidopsis WRI1 (An & Suh 2015). However, homology-based approaches are unlikely to reveal the regulators that make Camelina such a good oil producer. Moreover, techniques such as ChIP-seq, classically used to discover TF targets, can be challenging to implement because of the difficulties associated with developing antibodies that recognize a single homolog in a polyploid, and the use of an epitope-tagged version of the TF for ChIP experiments is questionable because the function of the epitope-tagged TF cannot be tested unless a mutant is available (which again is difficult to obtain in a polyploid).

We present here a co-expression-guided approach to identify candidate Camelina TFs involved in the control of seed oils, followed by the evaluation of TF target genes based on DAP-seq. While not perfect, this strategy overcomes many of the limitations imposed by a polyploid genome, providing a small set of candidate TFs that can be used for metabolic engineering efforts (Grotewold, 2008). Co-expression analyses identified 22 TFs strongly co-expressed with LRGs, which were further reduced to 16 after several quality control steps. Furthermore, co-expression analyses with seed expression data allowed us to identify specific metabolic processes targeted by our regulators, including the control of FA- and TAG-related genes during Camelina seed development.

Evidence of the robustness of our co-expression analysis is provided by the inclusion in our list of candidate TFs homologs of well-known regulators of lipid-related metabolism in Arabidopsis, including the Camelina homologs of ABI3, FUS3, MYB9, and MYB107 (Giraudat et al., 1992; Keith et al., 1994; Lashbrooke et al., 2016). Our list of Camelina TFs also includes homologs of TFs indirectly associated with lipid metabolism in Arabidopsis, such as the Arabidopsis homolog of CsaNAC2 (AtNAC60). AtNAC60 was shown to play a role in sugar sensing (Li et al., 2014), as a negative regulator of AtABI5 (Yu et al., 2020), and is a target of AtABI4 (Li et al., 2014). Both AtABI5 and AtABI4 are known regulators of sugar-responsive expression, seed germination, and lipid metabolism (Chandrasekaran et al., 2020; Skubacz et al., 2016). While AtULT1 (homolog of CsaULT1) has been implicated in various Arabidopsis plant developmental processes (Fletcher, 2001; Ornelas-Ayala et al., 2020; Pires et al., 2015), a recent transcriptome analysis of loss of AtUTL1 function (Tyler et al., 2019) showed a significant enrichment ($P < 0.05$, Figure S15) for LRGs among the differentially expressed genes, indicating a participation of AtUTL1 in the control of lipids. We could not test the co-expression of AtMYB67 (homolog of CsaMYB3) with Arabidopsis LRGs because it is expressed at very low levels. However, supporting a potential role of CsaMYB3 in lipid-related metabolism, AtMYB67 physically interacts with the known negative regulator of cuticular wax biosynthesis AtDEWAX (Trigg et al., 2017), which is a target of AtAGL15, a regulator of embryogenesis and gibberellic acid catabolism (Zheng et al., 2013). However, our analysis also identified 13 Camelina TFs (CsaC3H2, CsaHB2, CsazfHD1, CsaC3H1, CsaAP2/B3-like-1, CsaC2C2-Dof1, CsaB3-1, CsaS1Fa-like-1, CsaNAC1, CsaWRKY1, CsaTify1, CsaHB1, and CsaLBD1; Table 1) that were previously not associated with the regulation of lipids.

To further elucidate the role of these 22 Camelina TFs in lipid regulation and to identify potential targets of these TFs, we applied DAP-seq to them. DAP-seq has many limitations; chief among them is the fact that it is performed in a chromatin-free context, resulting in the identification of binding sites and potential targets that might not be accessible *in vivo*. However, it is easy to implement, and it is not affected by a polyploid genome, as ChIP techniques

are. DAP-seq permitted us to identify TFBMs and potential target genes for 16 out of the 22 TFs identified. When we compared our DAP-seq results with those derived from a large-scale analysis conducted for Arabidopsis TFs (O'Malley et al., 2016), we determined that DNA-binding properties for 10 out of the 16 TFs are not available for the corresponding Arabidopsis orthologs (Figure S5a), either because were not tested (homologs of CsaMYB1, CsaNAC2, and CsaC3H2), or because the corresponding experiments for Arabidopsis TFs did not result in meaningful results (homologs of CsaC2C2-Dof1, CsaC3H1, CsaHB1, CsaHB2, CsaHRT1, CsaLBD1, and Csazf-HD1).

The analysis of TF target enrichment for LRGs and GO terms within the sets of predicted targets provided additional validation for the results obtained from the co-expression analyses for nine TFs (Figure 2a and Figures S10 and S11). Interestingly, in several instances, the GO enrichment analysis also exposed biological processes previously reported for the corresponding Arabidopsis homologs. For example, the DAP-seq results for CsaMYB1 and CsaWRKY1 showed enrichment for targets associated with suberin biosynthesis and defense responses, which are known functions of AtMYB9 and AtWRKY3, respectively (Birkenbihl et al., 2018; Lai et al., 2008; Lashbrooke et al., 2016). The targets of CsaNAC2 were also enriched in genes associated with fruit ripening, hormone biosynthesis processes, exit from dormancy, inositol lipid-mediated signaling, and lipid homeostasis terms (Figure S11, Table S9), which are in good agreement with the functions attributed to AtANAC60, the Arabidopsis homolog (Li et al., 2014; Yu et al., 2020). The targets of Csazf-HD1 showed enrichment for several GO terms, including cell death, cell wall organization, cellular response to endogenous stimulus, and cellular response to hormone stimulus, matching the predicted functions of AtZHD13/RHD1 (Liu et al., 2021). CsaHB1's targets were enriched in GO terms related to leaf development and light responses, response to auxin, response to abscisic acid, and post-embryonic plant organ development, among others (Table S9), similar to the known functions of AtHB4 (Bou-Torrent et al., 2012; Carabelli et al., 1993; Sorin et al., 2009). Finally, the targets of CsaABI3VP1-2 showed enrichment for GO terms that cover the full spectrum of the functions known for its Arabidopsis homolog, AtFUS3 (Curaba et al., 2004; Kagaya et al., 2005; Lumba et al., 2012; Tang et al., 2017; Tiedemann et al., 2008). When considered together, these results not only provide evidence of the biological significance of the associations identified here, but also reveal the intertwined connections between lipid metabolism and other biological processes in Camelina.

Regulators of plant metabolism often regulate multiple genes in a pathway, making them attractive for metabolic engineering (Broun 2004; Grotewold 2008). We took advantage of this characteristic of metabolic regulators to identify TF–process relationships with significant co-expression signals at the pathway level, rather than as individual target genes, by applying GSEA (Subramanian et al., 2005). This permitted us to identify previously unreported associations that further support several of the identified TFs as important lipid regulators (Figure 3 and Figure S13). Finally, we took advantage of a computational method (GENIE3) that involves causality (rather than simply correlation) to further support the role of several of the identified TFs in controlling particular aspects of lipid metabolism in seeds. When taken together with the results from the other methods applied in this study, our results suggest that CsabZIP1 is involved in controlling FA and TAG degradation just before seed germination, CsaMYB1 regulates FA elongation, and CsaABI3VP1 controls the synthesis of TAGs (Figures 3 and 4). Our results also imply CsaMYB1 and CsaNAC2 participate in the regulation of cutin biosynthesis (Figure 3).

CONCLUSION

Gene regulation is at the core of many important agronomic attributes and TFs have a large potential to modify complex traits (Century et al. 2008; Springer et al. 2019). Identifying TFs that control specific metabolic or developmental processes in polyploids is challenging because the effect of mutations is often masked by redundancy, and traditional approaches to investigate TF function are limited by high sequence identity between homologs. Our strategy to identify Camelina candidate TFs involved in the regulation of lipid metabolism was based on a combination of co-expression analyses and target identification using DAP-seq. These resulted in the identification of a set of 16 TFs. The presence among these 16 TFs of several that were previously shown in Arabidopsis to participate in different aspects of lipid accumulation furnished a validation for the approach. Incorporating into our pipeline co-expression analyses that imply causality and that take into consideration that TFs often control multiple genes in a metabolic pathway further provided a better picture of the regulatory events involving the identified TFs in seed oil accumulation in Camelina. Similar combination of approaches could significantly contribute to identify key regulators for important agronomic traits in other polyploids.

EXPERIMENTAL PROCEDURES

Plant materials and growth conditions

Camelina sativa cultivar Suneson was grown in the plant biology greenhouse at Michigan State University. RNA-seq and DAP-seq experiments were performed on plants grown for 1 month at 22°C and under 16/8-h light/dark cycles. For seed RNA-seq, total RNA was extracted from seedpods harvested at 5, 8, and 11 DPA. For DAP-seq, a pool of 10 leaves from six mature (two-month-old) plants was collected.

Cloning and expression of transcription factors for DAP-seq

A set of 22 full-length Camelina TF ORFs were annotated using the *C. sativa* cultivar DH55 reference genome V2.0 (http://camelinadb.ca). Coding regions were assembled (when required) using expression data available for the Camelina cultivar Suneson (Table S12). Gateway-compatible TF ORFs were synthesized by Genewiz (https://www.genewiz.com/Public/Services/Gene-Synthesis/Standard). Clones were recombined using LR clonase II (Life Technologies, Carlsbad, CA, USA) into the pIX-Halo expression vector containing both T7 and SP6 promoters (pIX-Halo:ccdB), a 6×His-tag at the C-terminus, and no stop codon but a T7 terminator.

RNA-seq library preparation

Total RNA from fresh seed, after removing pod covers, was extracted using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's protocol. The total RNA was prepared with three biological replicates, each with approximately 100 mg seeds. The quality of total RNA was determined by a TapeStation4200 (Agilent, Santa Clara, CA, USA), and cDNA library was generated with 1 μg of total RNA using TruSeq stranded mRNA (Illumina). The pooled libraries were sequenced with a paired-end read length of 150 bp by the Illumina HiSeq 4000 platform at the Research Technology Support Facility Genomics Core at Michigan State University.

DAP-seq library preparation

Camelina sativa genomic DNA (gDNA) was extracted using urea buffer (7 M urea, 350 mM NaCl, 50 mM Tris–HCl [pH 8], 20 mM EDTA, 1% *N*-lauroyl sarcosine) and mixed with phenol:chloroform:isoamyl alcohol 25:24:1. The supernatants containing DNA were further precipitated using 3 M NaOAc (pH 5.2) and isopropanol followed by 70% ethanol wash. The DNA pellet was resuspended in UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen, Waltham, MA, USA) followed by RNase A (Roche, Basel, Switzerland) treatment and ethanol precipitation. DAP-seq gDNA libraries were constructed following the protocol of Bartlett et al. (2017) with minor modifications. Extracted gDNA was fragmented to the size range between 200–400 bp using a Diagenode Bioruptor® 300 for 40 cycles with 30 sec on/off at high energy. The fragmented DNA was further used for end repair and adapter ligation. To create modification-free DNA, 11 additional PCR cycles were performed using the adapter-ligated libraries, followed by ethanol precipitation. Finally, the amplified gDNA libraries (ampDAP) were used for all PDI procedures. The Halo-tagged TFs were expressed in the wheat (*Triticum aestivum*) germ *in vitro* transcription/translation SP6 promoter system (Promega, Madison, WI, USA). All the buffers and procedures for PDI were as published (Bartlett et al., 2017), except for the input gDNA library amount and the final step of library size selection. About 200 ng of ampDAP gDNA library was added as an input to mix with each pIX-Halo-TF protein. Finally, to perform double-size selection targeting 300–400-bp fragments, 0.7 volumes of Agencourt AMPure XP beads (35 μl) and 1 volume of sample (50 μl) were mixed for 5 min and the bead was discarded to remove fragments with size larger than 400 bp. Next, the supernatant (85 μl) containing <400-bp fragments was added to 0.2 volumes of Agencourt AMPure XP beads (10 μl) and mixed for 15 min. The bound fragments were eluted from the beads by adding 18 ml UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen, Waltham, MA, USA). The concentrations of eluted DNA were measured using the Qubit HS dsDNA assay kit, and final concentrations of approximately 5–20 ng μl^{-1} were obtained. The fragment size and binding capacity to the flow cell were further examined on the agarose gel by six PCR cycles using 2 μl of eluted ampDAP-seq library with Illumina P5 and P7 primers. Twelve libraries were pooled in one lane and sequenced by the Illumina HiSeq 4000 platform with single-end 50-bp reads at the RTSF Genomics core at Michigan State University.

Data processing, quantification, and statistical analyses

Detailed descriptions of RNA-seq, DAP-seq, quality control analyses, and the different co-expression analysis are provided in Methods S1.

DATA AVAILABILITY AND ACCESSION NUMBERS

The data supporting the findings of this work are available in the supplementary files. Raw sequencing data generated can be found in the NCBI SRA database under accession number PRJNA763897. Processed data have been deposited in the NCBI GEO database under accession number GSE184283.

CODE AVAILABILITY

Codes associated with descriptions and figures of the RNA-seq, co-expression, and DAP-seq results presented here are available at https://github.com/gomezcan/TransRegLipMet_CamelinaSeed.git.

ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0018269.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Building a co-expression-based LRG-regulatory network. (a) Total LRGs were collected from CamRegBase as input for the construction of the co-expression network. The classification of the different lipid-related processes/pathways was according to AraLip (http://aralip.plantbiology.msu.edu/). (b) Flowchart indicating the major steps and data used to predict transcriptional regulators for the LRGs described in (a). (c) Bar graph indicating the Z-score corresponding to seed expression for the top 19 TF candidates with the largest number of co-expressed LRGs.

Figure S2. Co-expression of known lipid/seed development regulators and LRGs in Camelina and Arabidopsis. The bar graphs show the total number of LRGs co-expressed with (a) Camelina homologs of each Arabidopsis TF (note that there are three bars for each Arabidopsis regulator because of the hexaploid nature of the Camelina genome) or (b) Arabidopsis homologs (names in square brackets) for the Camelina top TFs. The color of the bar indicates the significance of the number of LRGs co-expressed (light red, FDR ≤ 0.05; turquoise, FDR > 0.05).

Figure S3. Quality control analysis of Camelina DAP-seq experiments. (a) Principal component analysis (PCA) of total mapped reads for all top 22 TF candidates and the respective Halo control. Dots with the same color represent biological replicates (indicated as R1 and R2). Inset bar graph represents the total variance explained by the first 10 principal components (PCs). The two main PCs explained approximately 54% of the observed variance, with PC1 showing that all TFs are well separated from the Halo

control. The majority (approximately 16/20) of the TFs are separated by PC2. (b) Heatmap representing the Spearman correlation coefficients (lower coefficient indicated with darker colors) of all DAP-seq experiments against each other. The red triangle highlights samples highly correlated with the Halo control. Gray arrows indicate biological replicates with low correlations. The majority (40/44) of the replicates showed a high correlation with each other (Spearman coefficient correlation [ρ] > 0.7). However, consistent with the PCAs (Figure S3a), five TFs (CsaMYB2, CsaULT1, CsaTify1, CsaABI3VP1-1, and CsaB3-1) showed either a low correlation between the replicates ($0.5 < \rho < 0.7$) (gray arrows) or a comparable correlation to the Halo control (CsaABI3VP1-1, CsaB3-1) (red triangles).

Figure S4. DNA-binding events (peaks) by TFs. (a) Total predicted peaks by TFs. (b) Frequency of the distance between the final set of peak summits and annotated Camelina TSSs.

Figure S5. Comparison of successful and unsuccessful Camelina and Arabidopsis DAP-seq experiments. The Camelina DAP-seq is part of this study, while the Arabidopsis DAP-seq derives from O'Malley et al. (2016). (a) Venn diagrams indicating the number of TFs tested by DAP-seq in Arabidopsis versus those in Camelina. Arabidopsis categories we created comparing the list of TFs tested versus the list of TFs with available DAP-seq data as reported by O'Malley et al. (2016) in Table S1b,d, respectively. (b) CsaMYB2, unlike its corresponding Arabidopsis homolog, failed to pass the DAP-seq quality filters, which may be explained by the absence of a MYB domain in the Camelina protein used.

Figure S6. Reproducibility analysis between TF replicates based on DNA-binding fold changes. We calculated the $\log_2$ of the binding fold change ($\log_2$FC) for the total predicted peaks for each TF dividing the number of reads obtained for each peak with Halo-TF by the number of reads obtained for the same peak for the Halo control. Peaks with $\log_2 FC \geq 0.5$ in both replicates were defined as highly reproducible peaks.

Figure S7. Distribution of predicted TF-binding motifs (TFBMs) in the predicted peaks. The prediction of TFBMs resulted in up to three different motifs for some of the TFs. To identify the main motif, we counted the frequency and location of each of the predicted TFBMs in all the predicted peaks. The frequency of each TFBM is presented as the motif Z-score in a heatmap indicating the start position of the TFBMs on the peak.

Figure S8. Predicted Camelina TFBMs are conserved for the corresponding Arabidopsis homologs. Alignment of TFBMs predicted for (a) CsaNAC1, (b) CsaABI3VP1-2, (c) CsaMYB3, (d) CsaAP2/B3-like-1, (e) CsabZIP1, and (f) CsaWRKY1 with the TFBMs of their corresponding Arabidopsis homologs.

Figure S9. Protein sequence alignment of CsaAP2/B3-like-1 and its corresponding Arabidopsis homolog (REM16). The light blue lines show the B3-2 domain that is absent in the annotated Camelina protein used for the DAP-seq experiments.

Figure S10. Top GO terms enriched in the predicted target genes. Top 20 GO terms enriched presented as a hierarchical GO graph for (a) CsaMYB, (b) CsaNAC2, (c) CsaNAC1, (d) CsaHRT1, (e) CsaLBD1, (f) CsaC2C2-Dof1, (g) CsaWRKssY1, (h) CsaHB1, (i) CsaAP2/B3-like-1, (j) Csazf-HD1, (k) CsabZIP1, (i) CsaMYB1, (m) CsaC3H1, (n) CsaC3H2, (o) CsaHB2, and (p) CsaABI3VP1-2. Gray dashed box highlights the top 10 GO terms used in further analysis after discarding general terms. GO terms significantly enriched are highlighted in square boxes ($P \leq 0.05$, Fisher's exact test). The colors of the boxes indicate the statistical significance (with lower *P*-values indicated by darker colors).

Figure S11. Heatmap and hierarchical clustering of TF candidates based on the top 10 GO significantly enriched terms. GO terms not significantly enriched are shown in white. The color indicates the percentage of targets annotated within the corresponding GO term.

Figure S12. Top GO terms enriched in common targets of (a) CsaMYB1 & CsaWRKY1 and (b) CsabZIP1 & CsaHB1. GO terms significantly enriched are highlighted in square boxes ($P \leq 0.05$, Fisher's exact test). The colors of the boxes indicate the statistical significance (with lower *P*-values indicated by darker colors).

Figure S13. GSEA of lipid-related processes (pathways). The summary plot represents the position of the corresponding target genes (each gene represented by a vertical bar; if positively correlated the bar is above the central axis, if negatively regulated it is below the central axis) in the rank of all Pearson correlation coefficient (PCC) values with the corresponding TF for each metabolic pathway for (a) CsaABI3VP1-2, (b) CsaWRKY1, (c) CsaMYB1, (d) CsaMYB3, (e) CsabZIP1, (f) CsaC2C2-Dof1, (g) CsaNAC1, and (h) CsaHB1. Only pathways with a significant normalized enrichment score (NES) are presented.

Figure S14. Analysis of targets of CsaABI3VP1-2 and corresponding Arabidopsis homologs. The large majority of the lipid/seed development regulators that are targets of CsaABI3VP1-2 are also targets of Arabidopsis FUS3. Some of these regulators are also differentially expressed in the corresponding TF mutant (labeled with asterisks) (Yamamoto et al., 2010).

Figure S15. Venn diagrams comparing UTL1-regulated genes and annotated Arabidopsis LRGs. The lower part of the figure shows the list of differentially expressed LRGs in Arabidopsis plants with gain or loss of UTL1 function and the corresponding *P*-values for the comparisons. Gene expression information was obtained from Tyler et al. (2019).

Methods S1 Supporting Experimental procedures.

Table S1. Camelina RNA-seq samples collected from CamRegBase and generated as part of this study.

Table S2. Lipid-related genes collected and used in this study. Annotation and classification follow Aralip nomenclature (http://aralip.plantbiology.msu.edu/pathways/pathways).

Table S3. Transcription factors associated with the control of lipid metabolism and seed development annotated in Arabidopsis and their corresponding Camelina homologs.

Table S4. List of 350 TFs co-expressed with LRG.

Table S5. Arabidopsis RNA-seq datasets used to calculate co-expression between the Arabidopsis homologs of Camelina TF candidates.

Table S6. Summary of the sequenced and mapped Camelina DAP-seq reads and total peaks predicted by TF.

Table S7. Comparison of total peaks for seven pairs of TF homologs between Camelina and Arabidopsis.

Table S8. Camelina TF–target network predicted based on high-quality and reproducible DAP-seq peaks.

Table S9. Total list of GO terms significantly enriched in predicted targets for each of the TFs tested.

Table S10. Total list of TF targets annotated by lipid-related pathway.

Table S11. GENEI3 co-expression network. Predicted targets were filtered by a permutation analysis and corrected for multiple testing.

Table S12. TF coding sequences used in the *in vitro* transcription/translation and subsequent DAP-seq assays.

REFERENCES

Abdullah, H.M., Akbari, P., Paulose, B., Schnell, D., Qi, W., Park, Y. *et al.* (2016) Transcriptome profiling of *Camelina sativa* to identify genes

involved in triacylglycerol biosynthesis and accumulation in the developing seeds. *Biotechnology for Biofuels*, **9**, 136.

Alonso, R., Oñate-Sánchez, L., Weltmeier, F., Ehlert, A., Diaz, I., Dietrich, K. *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.

An, D. & Suh, M.C. (2015) Overexpression of ArabidopsisWRI1 enhanced seed mass and storage oil content in *Camelina sativa*. *Plant Biotechnology Report*, **9**, 137–148.

Bartlett, A., O'Malley, R.C., Huang, S.-S.C., Galli, M., Nery, J.R., Gallavotti, A. *et al.* (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature Protocols*, **12**, 1659–1672.

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

Bäumlein, H., Miséra, S., Luerßen, H., Kölle, K., Horstmann, C., Wobus, U. *et al.* (1994) The FUS3 gene of Arabidopsis thaliana is a regulator of gene expression during late embryogenesis. *The Plant Journal*, **6**, 379–387.

Bensmihen, S., Rippa, S., Lambert, G., Jublot, D., Pautot, V., Granier, F. *et al.* (2002) The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. *Plant Cell*, **14**, 1391–1403.

Berti, M., Gesch, R., Eynck, C., Anderson, J. & Cermak, S. (2016) Camelina uses, genetics, genomics, production, and management. *Industrial Crops and Products*, **94**, 690–710.

Birkenbihl, R.P., Kracher, B., Ross, A., Kramer, K., Finkemeier, I. & Somssich, I.E. (2018) Principles and characteristics of the Arabidopsis WRKY regulatory network during early MAMP-triggered immunity. *The Plant Journal*, **96**, 487–502.

Bou-Torrent, J., Salla-Martret, M., Brandt, R., Musielak, T., Palauqui, J.-C., Martínez-García, J.F. *et al.* (2012) ATHB4 and HAT3, two class II HD-ZIP transcription factors, control leaf development in Arabidopsis. *Plant Signaling & Behavior*, **7**, 1382–1387.

Braybrook, S.A. & Harada, J.J. (2008) LECs go crazy in embryo development. *Trends in Plant Science*, **13**, 624–630.

Broun, P. (2004) Transcription factors as tools for metabolic engineering in plants. *Current Opinion in Plant Biology*, **7**, 202–209.

Carabelli, M., Sessa, G., Baima, S., Morelli, G. & Ruberti, I. (1993) The Arabidopsis Athb-2 and -4 genes are strongly induced by far-red-rich light. *The Plant Journal*, **4**, 469–479.

Carlsson, A.S. (2009) Plant oils as feedstock alternatives to petroleum – a short survey of potential oil crop platforms. *Biochimie***91**, 665–670. https://doi.org/10.1016/j.biochi.2009.03.021

Century, K., Reuber, T.L. & Ratcliffe, O.J. (2008) Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. *Plant Physiology*, **147**, 20–29.

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

Chandrasekaran, U., Luo, X., Zhou, W. & Shu, K. (2020) Multifaceted Signaling networks mediated by abscisic acid insensitive 4. *Plant Commun*, **1**, 100040.

Chaturvedi, S., Bhattacharya, A., Khare, S.K. & Kaushik, G. (2019) *Camelina sativa*: an emerging biofuel crop. In: Hussain, C.M. (Ed.) *Handbook of environmental materials management*. Cham: Springer International Publishing, pp. 2889–2925.

Curaba, J., Moritz, T., Blervaque, R., Parcy, F., Raz, V., Herzog, M. *et al.* (2004) AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis. *Plant Physiology*, **136**, 3660–3669.

Devic, M. & Roscoe, T. (2016) Seed maturation: simplification of control networks in plants. *Plant Science*, **252**, 335–346.

Fletcher, J.C. (2001) The ULTRAPETALA gene controls shoot and floral meristem size in Arabidopsis. *Development*, **128**, 1323–1333.

Focks, N. & Benning, C. (1998) 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.

Footitt, S., Marquez, J., Schmuths, H., Baker, A., Theodoulou, F.L. & Holdsworth, M. (2006) Analysis of the role of COMATOSE and peroxisomal beta-oxidation in the determination of germination potential in Arabidopsis. *Journal of Experimental Botany*, **57**, 2805–2814.

Fulda, M., Schnurr, J., Abbadi, A., Heinz, E. & Browse, J. (2004) Peroxisomal acyl-CoA synthetase activity is essential for seedling development in *Arabidopsis thaliana*. *Plant Cell*, **16**, 394–405.

Galli, M., Khakhar, A., Lu, Z., Chen, Z., Sen, S., Joshi, T. *et al.* (2018) The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nature Communications*, **9**, 4526.

Ghosh, A.K., Chauhan, N., Rajakumari, S., Daum, G. & Rajasekharan, R. (2009) At4g24160, a soluble acyl-coenzyme A-dependent lysophosphatidic acid acyltransferase. *Plant Physiology*, **151**, 869–881.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. & Goodman, H.M. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell*, **4**, 1251–1261.

Go, Y.S., Kim, H., Kim, H.J. & Suh, M.C. (2014) Arabidopsis cuticular wax biosynthesis is negatively regulated by the DEWAX gene encoding an AP2/ERF-type transcription factor. *Plant Cell*, **26**, 1666–1680.

Gomez-Cano, F., Carey, L., Lucas, K., García Navarrete, T., Mukundi, E., Lundback, S., Schnell, D. and Grotewold, E. (2020) CamRegBase: a gene regulation database for the biofuel crop, *Camelina sativa*, *Database*, **2020**, baaa075. https://doi.org/10.1093/database/baaa075

Grotewold, E. (2008) Transcription factors for predictive plant metabolic engineering: are we there yet? *Current Opinion in Biotechnology*, **19**, 138–144.

Guerriero, G., Martin, N., Golovko, A., Sundström, J.F., Rask, L. & Ezcurra, I. (2009) The RY/Sph element mediates transcriptional repression of maturation genes from late maturation to early seedling growth. *The New Phytologist*, **184**, 552–565.

Gugel, R.K. & Falk, K.C. (2006) Agronomic and seed quality evaluation of *Camelina sativa* in western Canada. *Canadian Journal of Plant Science*, **86**, 1047–1058.

Huynh-Thu, V.A., Irrthum, A., Wehenkel, L. and Geurts, P. (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, **5**, e12776. https://doi.org/10.1371/journal.pone.0012776

Iskandarov, U., Kim, H.J. & Cahoon, E.B. (2014) Camelina: an emerging oilseed platform for advanced biofuels and bio-based materials. In: McCann, M.C., Buckeridge, M.S. & Carpita, N.C. (Eds.) *Plants and bioenergy*. New York: Springer, pp. 131–140.

Jiang, T., Zhang, X.-F., Wang, X.-F. & Zhang, D.-P. (2011) Arabidopsis 3-ketoacyl-CoA thiolase-2 (KAT2), an enzyme of fatty acid β-oxidation, is involved in ABA signal transduction. *Plant & Cell Physiology*, **52**, 528–538.

Kagale, S., Koh, C., Nixon, J. *et al* (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nature Communications*, **5**, 1–11.

Kagaya, Y., Okuda, R., Ban, A., Toyoshima, R., Tsutsumida, K., Usui, H. *et al.* (2005) Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. *Plant & Cell Physiology*, **46**, 300–311.

Keith, K., Kraml, M., Dengler, N.G. & McCourt, P. (1994) fusca3: a heterochronic mutation affecting late embryo development in Arabidopsis. *Plant Cell*, **6**, 589–600.

Kong, Q. and Ma, W. (2018) WRINKLED1 transcription factor: how much do we know about its regulatory mechanism? *Plant Science*, **272**, 153–156. https://doi.org/10.1016/j.plantsci.2018.04.013

Kong, Q., Singh, S.K., Mantyla, J.J., Pattanaik, S., Guo, L., Yuan, L. *et al.* (2020) TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR4 interacts with WRINKLED1 to mediate seed oil biosynthesis. *Plant Physiology*, **184**, 658–665.

Kong, Q., Yang, Y., Guo, L., Yuan, L. & Ma, W. (2020) Molecular basis of plant oil biosynthesis: insights gained from studying the WRINKLED1 transcription factor. *Frontiers in Plant Science*, **11**, 24.

Kosma, D.K., Murmu, J., Razeq, F.M., Santos, P., Bourgault, R., Molina, I. *et al.* (2014) AtMYB41 activates ectopic suberin synthesis and assembly in multiple plant species and cell types. *The Plant Journal*, **80**, 216–229.

Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. (2017) A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics*, **18**, 205–214.

Kwong, R.W., Bui, A.Q., Lee, H., Kwong, L.W., Fischer, R.L., Goldberg, R.B. *et al.* (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell*, **15**, 5–18.

Lai, Z., Vinod, K., Zheng, Z., Fan, B. & Chen, Z. (2008) Roles of Arabidopsis WRKY3 and WRKY4 transcription factors in plant responses to pathogens. *BMC Plant Biology*, **8**, 68.

Lambert, S.A., Jolma, A., Campitelli, L.F. *et al.* (2018) The human transcription factors. *Cell*, **175**, 598–599.

Lashbrooke, J., Cohen, H., Levy-Samocha, D. *et al.* (2016) MYB107 and MYB9 homologs regulate suberin deposition in angiosperms. *Plant Cell*, **28**, 2097–2116.

Le, B.H., Cheng, C., Bui, A.Q. *et al.* (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 8063–8070.

Lee, S.B., Kim, H., Kim, R.J. & Suh, M.C. (2014) Overexpression of Arabidopsis MYB96 confers drought resistance in *Camelina sativa* via cuticular wax accumulation. *Plant Cell Reports*, **33**, 1535–1546.

Lee, S.B., Kim, H.U. & Suh, M.C. (2016) MYB94 and MYB96 additively activate cuticular wax biosynthesis in Arabidopsis. *Plant & Cell Physiology*, **57**, 2300–2311.

Lee, S.B. & Suh, M.C. (2015) Cuticular wax biosynthesis is up-regulated by the MYB94 transcription factor in Arabidopsis. *Plant & Cell Physiology*, **56**, 48–60.

Leprince, O., Pellizzaro, A., Berriri, S. & Buitink, J. (2016) Late seed maturation: drying without dying. *Journal of Experimental Botany*, **68**, 827–841.

Li, D., Jin, C., Duan, S. *et al* (2017) MYB89 transcription factor represses seed oil accumulation. *Plant Physiology*, **173**, 1211–1225.

Li, P., Zhou, H., Shi, X. *et al.* (2014) The ABI4-induced Arabidopsis ANAC060 transcription factor attenuates ABA signaling and renders seedlings sugar insensitive when present in the nucleus. *PLoS Genetics*, **10**, e1004213.

Li, Y., Beisson, F., Pollard, M. & Ohlrogge, J. (2006) Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. *Phytochemistry*, **67**, 904–915.

Liang, C., Liu, X., Yiu, S.-M. & Lim, B.L. (2013) De novo assembly and characterization of *Camelina sativa* transcriptome by paired-end sequencing. *BMC Genomics*, **14**, 146.

Li-Beisson, Y., Shorrosh, B., Beisson, F. *et al.* (2013) Acyl-lipid metabolism. *Arabidopsis Book*, **11**, e0161.

Liu, P., Nie, W.-F., Xiong, X. *et al.* (2021) A novel protein complex that regulates active DNA demethylation in Arabidopsis. *Journal of Integrative Plant Biology*, **63**, 772–786.

Lotan, T., Ohto, M., Yee, K.M. *et al.* (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell*, **93**, 1195–1205.

Lumba, S., Tsuchiya, Y., Delmas, F., Hezky, J., Provart, N.J., Qing Shi, L. *et al.* (2012) The embryonic leaf identity gene FUSCA3 regulates vegetative phase transitions by negatively modulating ethylene-regulated gene expression in Arabidopsis. *BMC Biology*, **10**, 8.

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

Malik, M.R., Tang, J., Sharma, N., Burkitt, C., Ji, Y., Mykytyshyn, M. *et al.* (2018) *Camelina sativa*, an oilseed at the nexus between model system and commercial crop. *Plant Cell Reports*, **37**, 1367–1381.

Mandáková, T., Pouch, M., Brock, J.R., Al-Shehbaz, I.A. & Lysak, M.A. (2019) Origin and evolution of diploid and allopolyploid camelina genomes were accompanied by chromosome shattering. *Plant Cell*, **31**, 2596–2612.

Meinke, D.W., Franzmann, L.H., Nickle, T.C. & Yeung, E.C. (1994) Leafy cotyledon mutants of Arabidopsis. *Plant Cell*, **6**, 1049–1064.

Mendes, A., Kelly, A.A., van Erp, H., Shaw, E., Powers, S.J., Kurup, S. *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.

Morineau, C., Bellec, Y., Tellier, F., Gissot, L., Kelemen, Z., Nogué, F. *et al.* (2017) Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. *Plant Biotechnology Journal*, **15**, 729–739.

Mudalkar, S., Golla, R., Ghatty, S. & Reddy, A.R. (2014) De novo transcriptome analysis of an imminent biofuel crop, *Camelina sativa* L. using Illumina GAIIX sequencing platform and identification of SSR markers. *Plant Molecular Biology*, **84**, 159–171.

Neumann, N.G., Nazarenus, T.J., Aznar-Moreno, J.A., Rodriguez-Aponte, S.A., Mejias Veintidos, V.A., Comai, L. *et al.* (2021) Generation of camelina mid-oleic acid seed oil by identification and stacking of fatty acid biosynthetic mutants. *Industrial Crops and Products*, **159**, 113074.

Nguyen, H.T., Silva, J.E., Podicheti, R. *et al.* (2013) Camelina seed transcriptome: a tool for meal and oil improvement and translational research. *Plant Biotechnology Journal*, **11**, 759–769.

Nikolov, L.A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I.A., Filatov, D. *et al.* (2019) Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *The New Phytologist*, **222**, 1638–1651.

O'Malley, R.C., Huang, S.S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R. *et al.* (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, **165**, 1280–1292.

O'Neill, C.M., Baker, D., Bennett, G., Clarke, J. & Bancroft, I. (2011) Two high linolenic mutants of *Arabidopsis thaliana* contain megabase-scale genome duplications encompassing the FAD3 locus. *The Plant Journal*, **68**, 912–918.

Ornelas-Ayala, D., Vega-León, R., Petrone-Mendoza, E., Garay-Arroyo, A., García-Ponce, B., Álvarez-Buylla, E.R. *et al.* (2020) ULTRAPETALA1 maintains Arabidopsis root stem cell niche independently of ARABIDOPSIS TRITHORAX1. *The New Phytologist*, **225**, 1261–1272.

Ozseyhan, M.E., Kang, J., Mu, X. & Lu, C. (2018) Mutagenesis of the FAE1 genes significantly changes fatty acid composition in seeds of *Camelina sativa*. *Plant Physiology and Biochemistry*, **123**, 1–7.

Pelletier, J.M., Kwong, R.W., Park, S. *et al.* (2017) LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E6710–E6719.

Pires, H.R., Shemyakina, E.A. & Fletcher, J.C. (2015) The ULTRAPETALA1 trxG factor contributes to patterning the Arabidopsis adaxial-abaxial leaf polarity axis. *Plant Signaling & Behavior*, **10**, e1034422.

Pollard, M., Martin, T.M. & Shachar-Hill, Y. (2015) Lipid analysis of developing *Camelina sativa* seeds and cultured embryos. *Phytochemistry*, **118**, 23–32.

Pouvreau, B., Blundell, C., Vohra, H., Zwart, A.B., Arndell, T., Singh, S. *et al.* (2020) A versatile high throughput screening platform for plant metabolic engineering highlights the major role of ABI3 in lipid metabolism regulation. *Frontiers in Plant Science*, **11**, 288.

Reiter, F., Wienerroither, S. & Stark, A. (2017) Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development*, **43**, 73–81.

Richmond, T.A. & Bleecker, A.B. (1999) A defect in beta-oxidation causes abnormal inflorescence development in Arabidopsis. *Plant Cell*, **11**, 1911–1924.

Rodríguez-Rodríguez, M.F., Sánchez-García, A., Salas, J.J., Garcés, R. & Martínez-Force, E. (2013) Characterization of the morphological changes and fatty acid profile of developing *Camelina sativa* seeds. *Industrial Crops and Products*, **50**, 673–679.

Romanel, E.A.C., Schrago, C.G., Couñago, R.M., Russo, C.A.M. & Alves-Ferreira, M. (2009) Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PLoS One*, **4**, e5791.

Roudier, F., Gissot, L., Beaudoin, F. *et al.* (2010) Very-long-chain fatty acids are involved in polar auxin transport and developmental patterning in Arabidopsis. *Plant Cell*, **22**, 364–375.

Seo, P.J., Lee, S.B., Suh, M.C., Park, M.-J., Go, Y.S. & Park, C.-M. (2011) The MYB96 transcription factor regulates cuticular wax biosynthesis under drought conditions in Arabidopsis. *Plant Cell*, **23**, 1138–1152.

Shang, B., Xu, C., Zhang, X., Cao, H., Xin, W. & Hu, Y. (2016) Very-long-chain fatty acids restrict regeneration capacity by confining pericycle competence for callus formation in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 5101–5106.

Shen, B., Sinkevicius, K.W., Selinger, D.A. & Tarczynski, M.C. (2006) The homeobox gene GLABRA2 affects seed oil content in Arabidopsis. *Plant Molecular Biology*, **60**, 377–387.

Skubacz, A., Daszkowska-Golec, A. & Szarejko, I. (2016) The role and regulation of ABI5 (ABA-insensitive 5) in plant development, abiotic stress responses and phytohormone crosstalk. *Frontiers in Plant Science*, **7**, 1884.

Sorin, C., Salla-Martret, M., Bou-Torrent, J., Roig-Villanova, I. & Martínez-García, J.F. (2009) ATHB4, a regulator of shade avoidance, modulates hormone response in Arabidopsis seedlings. *The Plant Journal*, **59**, 266–277.

Springer, N., de León, N. & Grotewold, E. (2019) Challenges of translating gene regulatory information into agronomic improvements. *Trends in Plant Science*, **24**, 1075–1082.

Stone, S.L., Kwong, L.W., Yee, K.M., Pelletier, J., Lepiniec, L., Fischer, R.L. *et al.* (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 11806–11811.

Subramanian, A., Tamayo, P., Mootha, V.K. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 15545–15550.

Suzuki, M. & McCarty, D.R. (2008) Functional symmetry of the B3 network controlling seed development. *Current Opinion in Plant Biology*, **11**, 548–553.

Tang, L.P., Zhou, C., Wang, S.S., Yuan, J., Zhang, X.S. & Su, Y.H. (2017) FUSCA3 interacting with LEAFY COTYLEDON2 controls lateral root formation through regulating YUCCA4 gene expression in Arabidopsis thaliana. *The New Phytologist*, **213**, 1740–1754.

Tian, R., Paul, P., Joshi, S. & Perry, S.E. (2020) Genetic activity during early plant embryogenesis. *The Biochemical Journal*, **477**, 3743–3767.

Tiedemann, J., Rutten, T., Mönke, G. *et al.* (2008) Dissection of a complex seed phenotype: novel insights of FUSCA3 regulated developmental processes. *Developmental Biology*, **317**, 1–12.

To, A., Joubès, J., Barthole, G., Lécureuil, A., Scagnelli, A., Jasinski, S., Lepiniec, L. and Baud, S. (2012) WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in Arabidopsis. *The Plant Cell*, **24**, 5007–5023. https://doi.org/10.1105/tpc.112.106120

Trigg, S.A., Garza, R.M., MacWilliams, A. *et al.* (2017) CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nature Methods*, **14**, 819–825.

Troncoso-Ponce, M.A., Barthole, G., Tremblais, G., To, A., Miquel, M., Lepiniec, L. *et al.* (2016) Transcriptional activation of two Delta-9 palmitoyl-ACP Desaturase genes by MYB115 and MYB118 is critical for biosynthesis of Omega-7 monounsaturated fatty acids in the endosperm of Arabidopsis seeds. *Plant Cell*, **28**, 2666–2682.

Tsukagoshi, H., Morikami, A. & Nakamura, K. (2007) Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2543–2547.

Tyler, L., Miller, M.J. & Fletcher, J.C. (2019) The Trithorax Group Factor ULTRAPETALA1 regulates developmental as well as biotic and abiotic stress response genes in Arabidopsis. *G3*, **9**, 4029–4043.

Voelker, T. & Kinney, A.J. (2001) Variations in the biosynthesis of SEED-storage lipids. *Annual Review of Plant Physiology and Plant Molecular Biology*, **52**, 335–361.

Wang, F. & Perry, S.E. (2013) Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. *Plant Physiology*, **161**, 1251–1264.

Wang, X., Niu, Q.-W., Teng, C., Li, C., Mu, J., Chua, N.-H. *et al.* (2009) Over-expression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in Arabidopsis. *Cell Research*, **19**, 224–235.

Yamamoto, A., Kagaya, Y., Usui, H., Hobo, T., Takeda, S. & Hattori, T. (2010) Diverse roles and mechanisms of gene regulation by the Arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis. *Plant & Cell Physiology*, **51**, 2031–2046.

Yu, B., Wang, Y., Zhou, H., Li, P., Liu, C., Chen, S. *et al.* (2020) Genome-wide binding analysis reveals that ANAC060 directly represses sugar-induced transcription of ABI5 in Arabidopsis. *The Plant Journal*, **103**, 965–979.

Zhai, Z., Liu, H. & Shanklin, J. (2017) Phosphorylation of WRINKLED1 by KIN10 results in its proteasomal degradation, providing a link between energy homeostasis and lipid biosynthesis. *Plant Cell*, **29**, 871–889.

Zhang, M., Cao, X., Jia, Q. & Ohlrogge, J. (2016) FUSCA3 activates triacylglycerol accumulation in Arabidopsis seedlings and tobacco BY2 cells. *The Plant Journal*, **88**, 95–107.

Zheng, Q., Zheng, Y. & Perry, S.E. (2013) AGAMOUS-Like15 promotes somatic embryogenesis in Arabidopsis and soybean in part by the control of ethylene biosynthesis and response. *Plant Physiology*, **161**, 2113–2127.

Zheng, Y., Ren, N., Wang, H., Stromberg, A.J. & Perry, S.E. (2009) Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15. *Plant Cell*, **21**, 2563–2577.