Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transcriptome Profiling of CsAP2L1 Function in Cannabis sativa Cell Suspensions Identifies Putative Regulatory Targets in Flavonoid Biosynthesis

Adam T. Sumner
Virginia Tech
Keely M. Beard
Virginia Tech
Elyse A. Shoppell
Virginia Tech
Lucas Dumez
Virginia Tech
Joseph S. Taylor
Virginia Tech
Aida Shahabi
York University
Sarah Pullano
York University
Parmis Shariat Torbaghani
York University
Nikola Kovinich
Virginia Tech
David C. Haak
Virginia Tech
Bastiaan O. R. Bargmann
bastiaan@vt.edu
Virginia Tech

Article

Keywords:

Posted Date: October 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11064797/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Transcriptional control of specialized metabolism in *Cannabis sativa* remains poorly understood, limiting efforts to rationally engineer biosynthetic pathways. This study interrogates the AP2/ERF transcription factor CsAP2L1 using an estradiol-inducible overexpression system established in *C. sativa* cell suspensions. CsAP2L1 activation triggered broad transcriptional reprogramming, revealed by RNA-seq and differential expression analysis, including up-regulation of genes associated with flavonoid associated processes. Targeted qRT-PCR validated induction of candidate regulatory and metabolic genes, confirming the robustness of the RNA-seq inferences and linking the results to the valuable, cannabis-specific cannflavin compounds. Promoter motif analysis of CsAP2L1-responsive clusters highlighted *cis*-elements consistent with AP2/ERF binding preferences. Together, these results identify putative components of a transcriptional network linking CsAP2L1 activity to flavonoid biosynthesis, following a repeatable framework for dissecting and manipulating cannabis metabolic gene regulatory networks.

Introduction

Cannabis sativa has garnered significant interest in both scientific and commercial industries due to its diverse secondary metabolites, including cannabinoids and flavonoids. Many elite cannabis varieties are highly heterozygous and maintain outcrossing populations, resulting in a finely tuned but unstable balance of metabolite production [1]. Biotechnology could pave the way for a new ability to modulate the plant's secondary metabolite production in a precise manner and potentially unlock the ability to elevate levels of alternative metabolites that are typically generated in small amounts. While cannabinoid biosynthesis and their pharmaceutical properties have been extensively studied, the flavonoid biosynthetic pathway in cannabis remains underexplored, even though its products may have significant therapeutic potential. Flavonoids contribute to pigmentation, plant defense, and human health benefits, with particular interest in cannabis-specific flavonoids such as cannflavin A, B, and C, which exhibit strong anti-inflammatory properties [2]. However, the regulatory mechanisms governing flavonoid biosynthesis in cannabis are largely unknown.

Flavonoids are a diverse class of polyphenolic compounds synthesized via the phenylpropanoid pathway. In plants, flavonoids serve crucial functions, including UV protection, pathogen resistance, and pigmentation [3]. In *C. sativa*, unique flavonoids such as cannflavins have demonstrated significant medicinal potential, particularly in reducing inflammation through the inhibition of prostaglandin E2 production [4][5][6][7]. Similar to cannabinoids, these cannflavins have been found in isolated glandular trichomes [8]. Despite their importance, the transcriptional regulation of flavonoid biosynthesis genes in cannabis remains poorly characterized.

Several key steps have been characterized in the flavonoid biosynthetic pathway, including chalcone synthase (CHS), flavonoid 3'-hydroxylase (F3'H), flavonol synthase (FLS), O-methyltransferase (OMT), sulfotransferases (SOTs), and glutathione S-transferases (GSTs). F3'H catalyzes the hydroxylation of

flavonoids, playing a critical role in the synthesis of luteolin and quercetin derivatives, while OMT21 is essential for the conversion of luteolin to chrysoeriol, a precursor of cannflavins [2]. Sulfotransferases such as SOT5 and SOT15 contribute to flavonoid modification by adding sulfate groups, which can alter solubility and bioactivity [9]. Aminocyclopropane-1-carboxylate oxidase 1 (ACCO1), is involved in ethylene biosynthesis, a process known to influence flavonoid biosynthesis [10]. YABBY1, a transcription factor, is associated with leaf and trichome development, potentially affecting flavonoid distribution in cannabis tissues [11]. Lastly, GSTUJ19, a glutathione S-transferase, may be involved in the intracellular transport and sequestration of flavonoids, facilitating their accumulation in vacuoles [12].

Fundamental genetic research of metabolic pathways can be conducted with a gene regulatory network analysis approach, utilizing the ability of transcription factors to affect the expression of numerous genes in a pathway. Possibilities include enhancing or repressing the synthesis of certain secondary metabolite amounts, as well as directing metabolite flux into certain branches of the biosynthesis pathways. One such transcription factor has been suggested to have a role in the regulation of cannabis secondary metabolism, CsAP2L1. From the APETALA2/Ethylene Response Factor (AP2/ERF) family, this transcription factor has glandular trichome-specific expression [13]. Significantly, trichomes are the specialized structures in which cannabinoids are secreted at their highest level in cannabis [14]. CsAP2L1 was found to bind to the promoters of the Tetrahydrocannabinolic acid synthase (*THCAS*) gene and the Cannabichromenic acid synthase (*CBCAS*) gene in a yeast-one-hybrid screening, as well as transcriptionally activate the promoters in Arabidopsis protoplast assays [13]. These synthase genes encode terminal enzymes in the cannabinoid biosynthetic pathway. More recently, comparative transcriptome and yeast-hybrid analyses profiled 16 TFs and their relationship to the regulation of the terminal synthase genes, finding the MYB transcription factors CsMYB6 and CsMYB29 as direct regulators of *CBDAS*, underscoring that cannabinoid biosynthesis in *Cannabis sativa* is governed by an expanding, multi-TF regulatory network [15]. This previous research has suggested a potential role for CsAP2L1 in regulating cannabinoid biosynthesis, leading to our interest in further investigating the regulatory role of CsAP2L1 in cannabis secondary metabolism. In the present study, inducible expression of *CsAP2L1* in cannabis cell suspensions followed by RNA-seq analysis indicates that CsAP2L1 may also regulate flavonoid biosynthesis, as supported by differential gene expression, gene ontology enrichment analyses, and promoter motif analysis.

C. sativa is well known as recalcitrant to genetic transformation and regeneration, making *in vivo* studies challenging [16]. We have developed an Agrobacterium-mediated cannabis leaf disc transformation and subsequent transgenic callus culture, leading to the ability to produce transformed dedifferentiated cannabis material in the form of callus and cell suspensions. This method of culturing transgenic cannabis material is rapid and more feasible than the current state of cannabis stable whole plant transformation and regeneration. Our study utilizes this method to be able to study CsAP2L1 within the cannabis system, enabling plant-specific, genome-wide examination of its downstream transcriptional regulation. We validated the RNA-seq analysis using qRT-PCR to measure the relative quantification of select genes, *CsOMT21* and *CsACCO1*, which are implicated in flavonoid biosynthesis. These findings provide new insight into the transcriptional regulation of flavonoid biosynthesis in cannabis and offer a

foundation for future research aiming to enhance flavonoid production through targeted genetic and metabolic engineering.

Results

Establishment of estradiol-inducible CsAP2L1 cell suspensions

Functional characterization of transcription factors in *C. sativa* has historically been hindered by the plant's recalcitrance to transformation and the lack of reliable, inducible expression systems. To overcome these barriers and investigate the regulatory role of the CsAP2L1 transcription factor, estradiol-inducible cell suspensions were generated through Agrobacterium-mediated transformation of leaf-disc explants from a high-CBD/low-THC cannabis strain (Cherry Wine) followed by callus proliferation and subsequent fragmentation into liquid culture (Fig. 1). This approach enabled controlled overexpression of *CsAP2L1* within a cannabis cellular environment while circumventing challenges associated with stable whole-plant transformation and regeneration in this recalcitrant species. Inducible green fluorescent protein (GFP) control cell suspensions demonstrated successful transgene expression and induction capacity 24 h after estradiol treatment, with visible fluorescence confirming expression of the β-estradiol-inducible system (Fig. 2A). The estradiol-inducible *CsAP2L1* cell suspension showed robust induction of the transcript, not seen in the control cell suspension (Fig. 2B). The transformed cell suspensions maintained viability and responsiveness to hormonal induction through numerous subculture cycles, suggesting stable integration and providing a reproducible platform for transcriptomic analysis of CsAP2L1-regulated gene expression. These results demonstrate the establishment of a novel, robust cannabis gene expression method suitable for downstream metabolic and gene expression interrogation, as well as use in the cannabis gene regulation field as a reliable system for stable-transgenic studies.

Transcriptomic profiling of CsAP2L1 - regulated gene expression

We profiled transcriptional responses to *CsAP2L1* overexpression by generating RNA-seq libraries from three biological replicates of estradiol-induced and non-induced cell suspensions containing either the *CsAP2L1* or *GFP* overexpression construct, for a total of 12 samples (AP2L1 mock/induced and GFP mock/induced). High-quality reads from all samples were assembled *de novo* using Trinity to maximize transcript discovery independently of existing *C. sativa* annotations, yielding a set of 45,621 putative genes for downstream analysis.

Differential expression analysis was centered on the CsAP2L1 induced vs. CsAP2L1 non-induced comparison, with the GFP line serving primarily as a specificity filter for estradiol-responsive genes that are independent of CsAP2L1 activity. In the resulting multi-comparison, genes that are regulated only in CsAP2L1 cells separate clearly from those that also change in the GFP control background, providing a visual framework for prioritizing CsAP2L1-specific candidates for downstream functional analysis.

To identify broader transcriptional programs associated with CsAP2L1, clustering was performed directly on a variance-stabilized, size-factor–normalized expression matrix. After inspecting the variance distribution across all expressed genes in the CsAP2L1 comparison, the clustering gene set was defined as the top 3,000 most variable genes when variance was computed using only CsAP2L1 samples, to emphasize CsAP2L1-driven variance while down-weighting effects observed in the GFP control samples. The resulting matrix was z-score scaled per gene, a correlation distance was computed between gene expression profiles across all 12 sample libraries, and hierarchical clustering on cell line and treatment means was used to partition genes into five major clusters (Fig. 3). To visualize these clustered transcriptional programs, variance-stabilized, size-factor–normalized expression values for all clustered genes were averaged across biological replicates within each condition (GFP mock, GFP induced, CsAP2L1 mock, and CsAP2L1 induced), yielding a genes × 4 conditions matrix that was row-wise z-scaled and displayed as a heatmap with column annotations indicating vector (CsAP2L1 vs GFP) and induction status (mock vs induced), along with row annotation of the cluster groupings, as well as gene expression trend plots with centroid expression profiles for each of the five clusters (Fig. 3). Cluster 1, containing 310 genes, emerged as the CsAP2L1 induction module: its member genes showed mostly coordinated upregulation specifically in estradiol-induced CsAP2L1 samples, while remaining overall close to baseline in non-induced CsAP2L1 and in GFP control conditions. This cluster therefore represents a coregulated gene set whose activation depends on CsAP2L1 rather than on estradiol treatment or GFP based background alone. Cluster 2, containing 390 genes, represents the most coordinated CsAP2L1 downregulated module of genes, characterized by lower variance stabilized mean expression in the CsAP2L1 induced samples and relatively baseline or higher expression in the CsAP2L1 mock and GFP control groups. Clusters 3–5 contained genes with expression patterns that do not match expectations of specific regulation by CsAP2L1 and were not taken along for further characterization.

Characterization of CsAP2L1-regulated gene clusters

Functional analysis of the genes within Clusters 1 and 2 gave confidence in the selection of Cluster 1 as the main focus of interrogation. Specifically, gene ontology (GO) enrichment analysis of the Cluster 1 genes revealed a secondary metabolite biosynthetic process within the overall GO distribution (Fig. 4). Notably, and of particular interest to us, an enrichment of flavonoid biosynthetic processes was overrepresented among the genes in the cluster, leading to further analysis into the gene expression patterns of specific flavonoid pathway associated genes. Additional enriched terms included responses to UV-B, heat, sugars, and phytohormones, as well as regulation of stomatal closure and reproductive development, suggesting that CsAP2L1-driven metabolic reprogramming is embedded within a broader network of stress and developmental signaling. The enrichment of genes with a flavonoid-related GO term provides evidence for a potential role of CsAP2L1 in regulating flavonoid biosynthesis. The overrepresentation of this term suggests that CsAP2L1 not only regulates individual biosynthesis genes but also coordinates multiple steps related to cannabis flavonoid production, modification, and transport. Analysis of Cluster 2, which is downregulated in estradiol-induced CsAP2L1 samples, highlighted diverse GO categories associated with cell wall organization, stress and defense responses,

and symbiotic or pathogen interactions, indicating that CsAP2L1-mediated repression in this cluster may impact a broad set of stress and cell wall–related processes (Supplementary Tables 1–4).

Taking a deeper look into the GO correlation with flavonoid biosynthesis of Cluster 1, we examined the expression patterns of individual flavonoid associated genes within the cluster. Gene expression plots of transcripts matching several flavonoid biosynthetic genes demonstrated significant upregulation in the induced CsAP2L1-overexpression cell suspensions (Fig. 5A). Notably, we observed upregulation of *CsF3'H* and *CsOMT21*, two terminal biosynthetic genes in the cannflavin biosynthesis pathway. CsF3'H catalyzes a critical hydroxylation step in the cannflavin pathway, while CsOMT21 is responsible for O-methylation, a key modification in cannflavin production [2]. The significant induction of these genes provides evidence for CsAP2L1's role in regulating cannflavin biosynthesis, specifically. In addition to these key enzymes, we also observed upregulation of several other putative flavonoid biosynthesis pathway-related genes, including *CsSOT5, CsSOT15, CsACCO1, CsYABBY1*, and *CsGSTUJ19.* The induction of these genes suggests that CsAP2L1 may regulate multiple aspects of flavonoid metabolism, including modification (CsSOT5, CsSOT15), transport (CsGSTUJ19), and potentially, developmental processes related to flavonoid accumulation (CsYABBY1) [9] [12]. The coordinated upregulation of genes spanning these multiple functions implies that CsAP2L1 orchestrates a comprehensive metabolic program rather than regulating isolated enzymatic steps.

To validate our RNA-seq results and confirm that the observed expression changes were not artifacts of high-throughput sequencing, we performed qRT-PCR analysis on select flavonoid biosynthetic genes (Fig. 5B). This analysis of an independently generated and induced CsAP2L1-overexpression cell suspension confirmed the significant upregulation of key genes identified in our transcriptomic study: *CsOMT21* and *CsACCO1*. The qRT-PCR results not only corroborate our RNA-seq findings but also provide a more quantitative assessment of gene expression changes and establishes *CsOMT21* and *CsACCO1* as high-confidence targets of CsAP2L1 regulation.

Further investigating the regulatory mechanisms underlying CsAP2L1-mediated gene expression and determining if the regulation is direct or indirect, we performed a promoter motif analysis on 119 successfully mapped promoter regions of the 310 Cluster 1 genes identified in the hierarchical cluster. Using MEME Suite [17] for motif discovery, an unbiased search identified an AP2/ERF motif as the top enriched sequence within the extracted promoters, with Fisher’s exact test giving a p-value of 5.41 e-5 (Fig. 6). This GCCGCC motif matches the consensus binding site for AP2/ERF transcription factors [18], providing strong evidence for direct binding of genes within the cluster. *CsOMT21*, *CsACCO1*, *CsYABBY1*, *CsSOT5*, and *CsSOT15* were among the genes with the AP2/ERF motif within their promoter region. This supports the hypothesis that CsAP2L1 most likely functions as a direct transcriptional activator of many of the genes in Cluster 1, rather than acting through intermediate regulators.

Discussion

Our comprehensive analysis of CsAP2L1-mediated gene regulation in *C. sativa* has revealed important insights, particularly into the transcriptional control of flavonoid biosynthesis in this species. The identification of a co-expressed gene cluster enriched for flavonoid biosynthetic and associated genes, coupled with the discovery of AP2/ERF binding motifs in many of their promoters, provides strong evidence for CsAP2L1's role as a regulator of multiple areas of the flavonoid biosynthesis pathway. This finding is consistent with previous studies in other plant species, where transcription factors have been shown to regulate multiple genes within secondary metabolite pathways [19]. AP2/ERF transcription factors have been shown to play crucial roles in regulating secondary metabolite biosynthesis and we suggest CsAP2L1 as yet another example [20]. The significant upregulation of key flavonoid biosynthetic genes, particularly *CsF3'H* and *CsOMT21*, suggests that CsAP2L1 may play a crucial role in regulating the production of cannflavins. These cannabis-specific flavonoids have garnered interest due to their potential therapeutic properties, particularly their anti-inflammatory effects. Our findings provide a foundation for future efforts to enhance the production of these valuable compounds through genetic engineering or breeding strategies.

While our transcriptomic and qRT-PCR analyses provide strong evidence for CsAP2L1's role in regulating flavonoid biosynthesis, we did not detect any flavonoid accumulation in an LC-MS metabolite analysis of mock and estradiol-induced cell suspensions (data not shown). This discrepancy could be due to several factors. Cell suspension cultures may lack the necessary cellular compartmentalization or cofactors required for efficient flavonoid production. There also may be additional post-transcriptional regulatory mechanisms that influence flavonoid biosynthesis in cannabis, which are not captured in our current analysis.

A prior study identified CsAP2L1 as a putative positive regulator of the *CsTHCAS* promoter and the *CsCBCAS* promoter in yeast-one-hybrid assays and Arabidopsis protoplast assays [13]. However, the transcriptomic analyses conducted in the present study did not reveal a strong association between CsAP2L1 activity and cannabinoid biosynthetic gene expression in the Cherry Wine variety. It is important to note that Cherry Wine is a high-CBD hemp cultivar which lacks a functional *CsTHCAS* gene, a limitation that could inherently restrict discovery of such regulatory relationships [21–23]. The absence of clear CsAP2L1-mediated effects on cannabinoid genes in this system could reflect genotype-specific dynamics, differences in gene content, metabolite flux, or developmental stage of the cell suspensions compared to glandular trichomes in flower tissue. The results suggest that CsAP2L1's function may be context-dependent and influenced by cultivar background and cellular environment. Further studies leveraging cannabis varieties with active *CsTHCAS* loci and trichome-rich tissues will be necessary to fully resolve the regulatory scope of CsAP2L1 in cannabinoid biosynthesis.

The earlier characterization of CsAP2L1 by Liu et al. (2021) also concluded, based on amino acid sequence alignment, that CsAP2L1 lacks the conserved DLNxxP EAR-type repression motif present in other AP2/ERF proteins and therefore may not act as a transcriptional repressor. However, inspection of the *CsAP2L1* sequence analyzed in that study, together with our independently sequenced *CsAP2L1* clone and alignments against the wide range of available high-quality *C. sativa* genomes, indicates that

CsAP2L1 does in fact contain an intact DLNxxP motif in the expected position. This suggests that the previously reported absence of the motif likely reflects an annotation or alignment oversight, or potentially an uncharacterized allelic variant. Nonetheless, the concordance between our clone and multiple reference genomes supports the presence of an intact DLNxxP motif in CsAP2L1 in the cannabis varieties analyzed here.

Interestingly, our results also indicate that CsAP2L1 regulates genes involved in various aspects of flavonoid metabolism beyond just biosynthesis. The upregulation of sulfotransferases (*CsSOT5* and *CsSOT15*) suggests that CsAP2L1 may also control flavonoid modification, potentially influencing the diversity of flavonoid compounds produced in cannabis [9]. Similarly, the induction of *CsGSTUJ19*, a glutathione S-transferase, implies a role in flavonoid transport or sequestration, which could affect the accumulation and distribution of these compounds within plant tissues [12]. The involvement of CsYABBY1 in this regulatory network is likewise intriguing. YABBY transcription factors are typically associated with leaf development and polarity, and their connection to flavonoid biosynthesis is not immediately obvious [11]. This finding suggests that CsAP2L1 may coordinate flavonoid production with specific developmental processes or tissue types, a hypothesis that warrants further investigation.

Future extensions to this research should include whole-plant stable transformation studies to verify the role of CsAP2L1 and genes in the flavonoid biosynthesis pathway under physiological conditions. Additionally, protein-DNA binding assays could provide direct evidence of CsAP2L1 binding to the promoters of flavonoid biosynthetic genes to confirm direct regulation. Metabolic engineering studies, perhaps utilizing CRISPR/Cas9-mediated genome editing, could further elucidate the specific contributions of CsAP2L1 to cannflavin production and potentially lead to cannabis varieties with altered flavonoid profiles. In conclusion, our study provides a view of CsAP2L1-mediated regulation of flavonoid biosynthesis in *C. sativa*, offering valuable insights into the transcriptional control of these important secondary metabolites. These findings not only advance our understanding of flavonoid biosynthesis in cannabis but also provide a foundation for future efforts to modulate the production of these compounds for both scientific and potential therapeutic applications.

Methods

Plant Material

Cherry Wine variety of *C. sativa* was obtained from the Institute for Advanced Learning and Research (IALR; Danville, VA, USA) in the form of *in vitro* explants. Cherry Wine is a high CBD/low THC variety. Micropropagation of the explants was conducted according to [24]. Plants were incubated in a growth chamber (CMP6010, Conviron, USA) with a 16 h light and 8 h dark regime, 75 µmol m^{-2} s^{-1} PAR (Philips F39T5/841 HO fluorescent bulbs), at 25°C.

Estradiol-Inducible Vectors

The *CsAP2L1* coding sequence (LOC115712195) was cloned from Cherry Wine genomic DNA into pDONR221 using the following primers: CsAP2L1_AttB1 aaaaagcaggcttgATGGAAGGAGATGAGTGTTGT and CsAP2L1_AttB2 agaaagctgggtTCAATGATCTTGTTCATCGGAAT, and generic primers AttB1 ggggacaagtttgtacaaaaaagcaggct and AttB2 ggggaccactttgtacaagaaagctgggt, followed by a Gateway BP reaction. *CsAP2L1* was inserted into the pMDC7 binary vector downstream of a β-estradiol-inducible promoter [25] utilizing the Gateway LR Recombination Cloning Technology (Life Technologies). The pMDC7_eGFP control vector was published previously [26]. The binary vectors were transformed into *Agrobacterium tumefaciens* EHA105 to be used for explant transformations.

Explant Transformation and Generation of Cell Suspensions

Leaf discs were excised from young leaves of micropropagated Cherry Wine plants and incubated in 100 x 15 mm round Petri dishes with MB5D1K callus induction medium (4.4 g/l Murashige and Skoog basal salt mixture [Sigma, M5519], 1 mL/L Gamborg B5 vitamin solution [Sigma, G1019], 2% w/v sucrose, 1% w/v agar, 5 µM 2,4-Dichlorophenoxyacetic acid [2,4-D] and 1 µM kinetin, pH 5.8) in the same growth chamber as described above for two days. Agrobacterium cultures were grown overnight shaking at 28°C in 3 mL LB with selection antibiotics and resuspended the next day in liquid MB5D1K with 100 µM acetosyringone to an OD_{600} of 0.6. Agrobacterium-mediated transformation of the leaf discs was conducted via vacuum infiltration to transfect the leaf discs with the estradiol-inducible constructs. Transfected leaf discs were incubated on MB5D1K medium for three days and subsequently transferred to MB5D1K medium supplemented with 12.5 mg/L hygromycin and 50 mg/L carbenicillin. After three weeks, hygromycin-resistant callus was passaged for several rounds on fresh MBD1K medium with antibiotic selection. After adequate time for callus proliferation, an approximate 5–10 g clump of callus was transferred to 50 mL of liquid MB5D1K medium supplemented with 12.5 mg/L hygromycin and 50 mg/L carbenicillin in a 250 mL flask and subsequently vigorously fragmented before incubation in a shaker incubator with 150 rpm shaking under ambient light at 22°C. Resulting cell suspensions were refreshed weekly by passaging approximately 5 mL of settled cell volume to flasks with fresh liquid medium.

RNA Isolation and RNA-seq Library Construction

After a four-day subculture of a refreshed 50 mL cell suspension, 7 mL of cell suspension was moved to a 50 mL conical tube and treated with 1 µM β-estradiol (50 mM stock dissolved in DMSO) or a DMSO solvent control for 24 h in a shaker incubator. Medium was then removed using vacuum filtration and total RNA was extracted using the Qiagen RNeasy mini kit according to the manufacturer manual. cDNA libraries were prepared utilizing the Lexogen QuantSeq 3’ mRNA-SEQ FWD Library Generation Module in conjunction with the Lexogen PCR Add-on Kit for Illumina and the Lexogene UDI 12 nt Unique Dual Indexing Add-on Kits. Library samples were analyzed for integrity and quantity using a TapeStation

D5000 and a Qubit 3.0 Fluorometer (Agilent Technologies and Life Technologies) resulting in an estimated average fragment size of 250 bp.

Transcriptome Assembly

RNA-sequencing was performed with 150-bp paired-end reads using an Illumina NovaSeq 6000 S-Prime. Resulting sequenced reads were analyzed using both a reference-guided and *de novo* approach. Sequences were aligned to the cs10 (CBDRx) genome using STAR [27]. Sequences were *de novo* merged and assembled using Trinity, with the resulting assembly annotated using Trinotate and BLAST [28]. The *de novo* Trinity assembly contained 238,574 transcripts with mean transcript length 378 bp. Abundance of transcripts in the three biological replicates of each sample was achieved using Kallisto [26]. Raw data has been deposited in the NCBI Sequence Read Archive (SRA) under BioProject: PRJNA1511286

(reviewer link:

https://dataview.ncbi.nlm.nih.gov/object/PRJNA1511286?reviewer=6v9r7cq547in9o8kpi76t46dip).

Gene Expression Analysis

Differential gene expression data was obtained using the Bioconductor package DESeq2, version 1.42.1 [30]. Hierarchical clustering was performed in R on an CsAP2L1-centric gene set defined as the top 3,000 most variable genes in CsAP2L1 samples, using z-score–scaled variance-stabilized counts, Pearson correlation distance (1 − r), and Ward.D2 linkage. Functional annotation of cluster gene sets was first performed using the Blast2GO workflow, which combined BLASTX searches and InterProScan with GO term mapping and annotation [31]. The resulting GO annotations were then manually extended by cross-referencing BLASTX results with Arabidopsis UniProt descriptions, adding GO terms for well-annotated homologs that were not captured by the automated pipeline, to ensure more complete coverage of known GO terms. GO enrichment analysis was performed using the enrichGO R package to identify overrepresented biological processes among genes in the CsAP2L1-responsive hierarchical cluster [32]. The analysis tested for enrichment against a background of all expressed genes detected in the RNA-seq experiment. GO terms with adjusted p-values < 0.05 (Benjamini-Hochberg correction) were considered significantly enriched. Results were visualized using the dotplot function, displaying the top 20 most significant terms ranked by adjusted p-value. A promoter sequence motif analysis was performed using MEME Suite SEA after extracting the 1000 bp upstream of the transcripts (from cs10 genome annotation) within the CsAP2L1-responsive hierarchical Cluster 1 (119 promoters) and comparing to the promoter sequences of all expressed genes in the RNA-seq data (13,678 promoters) [17]. Standard parameters were used for the SEA analysis, selecting the options for Eukaryote DNA and the JASPAR CORE Plants and Arabidopsis motif database.

qRT-PCR Analysis

A Thermo Fisher Scientific QuantStudio 6 Pro Real-Time PCR System was utilized for qRT-PCR assays. Approximately 150 bp regions of the genes of interest, *CsOMT21* and *CsACCO1*, were amplified using the following primers: CsOMT21_Fw CTACAACGTGGTGACGTA and CsOMT21_Rv AAGTGATACCAACTCTCCA; CsACCO1_Fw GCAAGAAACAGAGGTGAA and CsACCO1_Rv TGTCTCTGCATATAAGGGG. *CsEF1a* (Elongation Factor 1-alpha) was included in the study as the housekeeping gene for the relative quantification qRT-PCR method using the primers: CsEF1a_Fw TGGTATGGTTGTCACCTTT and CsEF1a_Rv GATCATCCTTGGAGTTGGA. *CsAP2L1* was included in the study to confirm proper activation of the transcription factor in the estradiol-inducible cell suspensions using the primers: CsAP2L1_Fw CAAGAAGACCTTCACCTAG and CsAP2L1_Rv TTTAGGTCGGCCTTAAAC. RNA extraction was completed on the treated and mock cell suspensions with the same method as the RNA-sequencing analysis, and cDNA was synthesized using the Applied Biosystems High-Capacity cDNA Reverse Transcription Kit for three replicates of each treated (estradiol) and mock (DMSO solvent only). For each RNA preparation a no-RT control (identical reaction without reverse transcriptase) was included to assess genomic-DNA contamination. The RT-minus reactions were subjected to the same quantitative PCR conditions as the cDNA samples; no amplification was observed in any no-RT control. qRT-PCR was conducted with the PowerUp SYBR Green Master Mix. Resulting cycle threshold (Ct) values were normalized to the *CsEF1a* housekeeping gene and relative quantification was completed.

Declarations

Author contributions statement

BORB, DCH, and NK conceived the experiment(s) and acquired funding, ATS, EAS, KMB, LD, JST, BORB, AS, PST, and SP conducted the experiments, ATS, AS, PST, and SP analyzed the results, ATS wrote the manuscript draft. All authors reviewed and edited the manuscript.

Funding

This work is supported in part by USDA NIFA AFRI award 2023-67013-38996 to the Bargmann, Haak, and Kovinich labs.

Author Contribution

BORB, DCH, and NK conceived the experiment(s) and acquired funding, ATS, EAS, KMB, LD, JST, BORB, AS, PST, and SP conducted the experiments, ATS, AS, PST, and SP analyzed the results, ATS wrote the manuscript draft. All authors reviewed and edited the manuscript.

Data Availability

Raw RNA-seq data has been deposited in the NCBI Sequence Read Archive (SRA) under BioProject: PRJNA1511286 (reviewer link:https://dataview.ncbi.nlm.nih.gov/object/PRJNA1511286?reviewer=6v9r7cq547in9o8kpi76t46dip).

References

1. Johnson, M. S. & Wallace, J. G. Genomic and Chemical Diversity of Commercially Available High-CBD Industrial Hemp Accessions. *Front. Genet.* **12**, (2021).
2. Rea, K. A. *et al.* Biosynthesis of cannflavins A and B from Cannabis sativa L. *Phytochemistry* **164**, 162–171 (2019).
3. Grotewold, E. The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology* vol. 57 761–780 Preprint at https://doi.org/10.1146/annurev.arplant.57.032905.105248 (2006).
4. Barrett, M. L., Gordon, D. & Evmst, F. J. *ISOLATION FROM CANNABIS SATIVA L. OF CANNFLAVIN-A NOVEL INHIBITOR OF PROSTAGLANDIN PRODUCTION*. *Biochemical Pharmacology* vol. 34 (1985).
5. Chuanphongpanich, S., Racha, S., Saengsitthisak, B., Pirakitikulr, P. & Racha, K. Computational Assessment of Cannflavin A as a TAK1 Inhibitor: Implication as a Potential Therapeutic Target for Anti-Inflammation. *Sci. Pharm.* **91**, (2023).
6. Eggers, C., Fujitani, M., Kato, R. & Smid, S. Novel cannabis flavonoid, cannflavin A displays both a hormetic and neuroprotective profile against amyloid β-mediated neurotoxicity in PC12 cells: Comparison with geranylated flavonoids, mimulone and diplacone. *Biochem. Pharmacol.* **169**, 113609 (2019).
7. Werz, O. *et al.* Cannflavins from hemp sprouts, a novel cannabinoid-free hemp food product, target microsomal prostaglandin E2 synthase-1 and 5-lipoxygenase. *PharmaNutrition* **2**, 53–60 (2014).
8. Ferri, E. *et al.* Single trichome phytocannabinomics of two different cannabis varieties. *J. Pharm. Biomed. Anal.* **261**, 116836 (2025).
9. Hirschmann, F., Krause, F. & Papenbrock, J. The multi-protein family of sulfotransferases in plants: Composition, occurrence, substrate specificity, and functions. *Front. Plant Sci.* **5**, (2014).
10. Dilley, D. R. *et al.* 1-Aminocyclopropane-1-carboxylic acid oxidase reaction mechanism and putative post-translational activities of the ACCO protein. *AoB Plants* **5**, (2013).
11. Xiong, Y., Long, C., Liu, P. & Song, Y. Interactions between PtoYABBY5 and PtoMYB43 regulate flavonoid synthesis in Populus tomentosa. *Ind. Crops Prod.* **206**, (2023).
12. Dixon, D. P., Lapthorn, A. & Edwards, R. Protein family review Plant glutathione transferases. *Genome Biol.* **3**, (2002).
13. Liu, Y., Zhu, P., Cai, S., Haughn, G. & Page, J. E. Three novel transcription factors involved in cannabinoid biosynthesis in Cannabis sativa L. *Plant Mol. Biol.* **106**, 49–65 (2021).
14. Sirikantaramas, S. *et al.* The gene controlling marijuana psychoactivity. Molecular cloning and heterologous expression of Δ1-tetrahydrocannabinolic acid synthase from Cannabis sativa L. *Journal of Biological Chemistry* **279**, 39767–39774 (2004).

15. Li, Y. *et al.* Comparative transcriptome analysis reveals putative transcription factors components of the regulatory network for cannabidiol biosynthesis in Cannabis sativa L. *Front. Plant Sci.* **17**, (2026).
16. Monthony, A. S., Kyne, S. T., Grainger, C. M. & Jones, A. M. P. Recalcitrance of Cannabis sativa to de novo regeneration; a multi-genotype replication study. *PLoS One* **16**, (2021).
17. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
18. Ohme-Takagi ', M. & Shinshi, H. *Ethylene-Lnducible DNA Binding Proteins That Lnteract with an Ethylene-Responsive Element*. *The Plant Cell* vol. 7 (1995).
19. Yang, C. Q. *et al.* Transcriptional Regulation of Plant Secondary Metabolism. *J. Integr. Plant Biol.* **54**, 703–712 (2012).
20. Paul, P. *et al.* Mutually regulated AP2/ERF gene clusters modulate biosynthesis of specialized metabolites in PlantS1[open]. *Plant Physiol.* **182**, 840–856 (2020).
21. Laverty, K. U. *et al.* A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res.* **29**, 146–156 (2019).
22. Johnson, M. S. & Wallace, J. G. Genomic and Chemical Diversity of Commercially Available High-CBD Industrial Hemp Accessions. *Front. Genet.* **12**, (2021).
23. Onofri, C., De Meijer, E. P. M. & Mandolino, G. Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in Cannabis sativa L. and its relationship with chemical phenotype. *Phytochemistry* **116**, 57–68 (2015).
24. Beard, K. M., Boling, A. W. H. & Bargmann, B. O. R. Protoplast isolation, transient transformation, and flow-cytometric analysis of reporter-gene activation in Cannabis sativa L. *Ind. Crops Prod.* **164**, (2021).
25. Curtis, M. D. & Grossniklaus, U. A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. *Plant Physiol.* **133**, 462–469 (2003).
26. Gonzalez, J. H., Taylor, J. S., Reed, K. M., Wright, R. C. & Bargmann, B. O. R. Temporal control of morphogenic factor expression determines efficacy in enhancing regeneration. *Plants* **10**, (2021).
27. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
28. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
29. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
30. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).
31. Conesa, A. *et al.* Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
32. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, (2021).

Figures

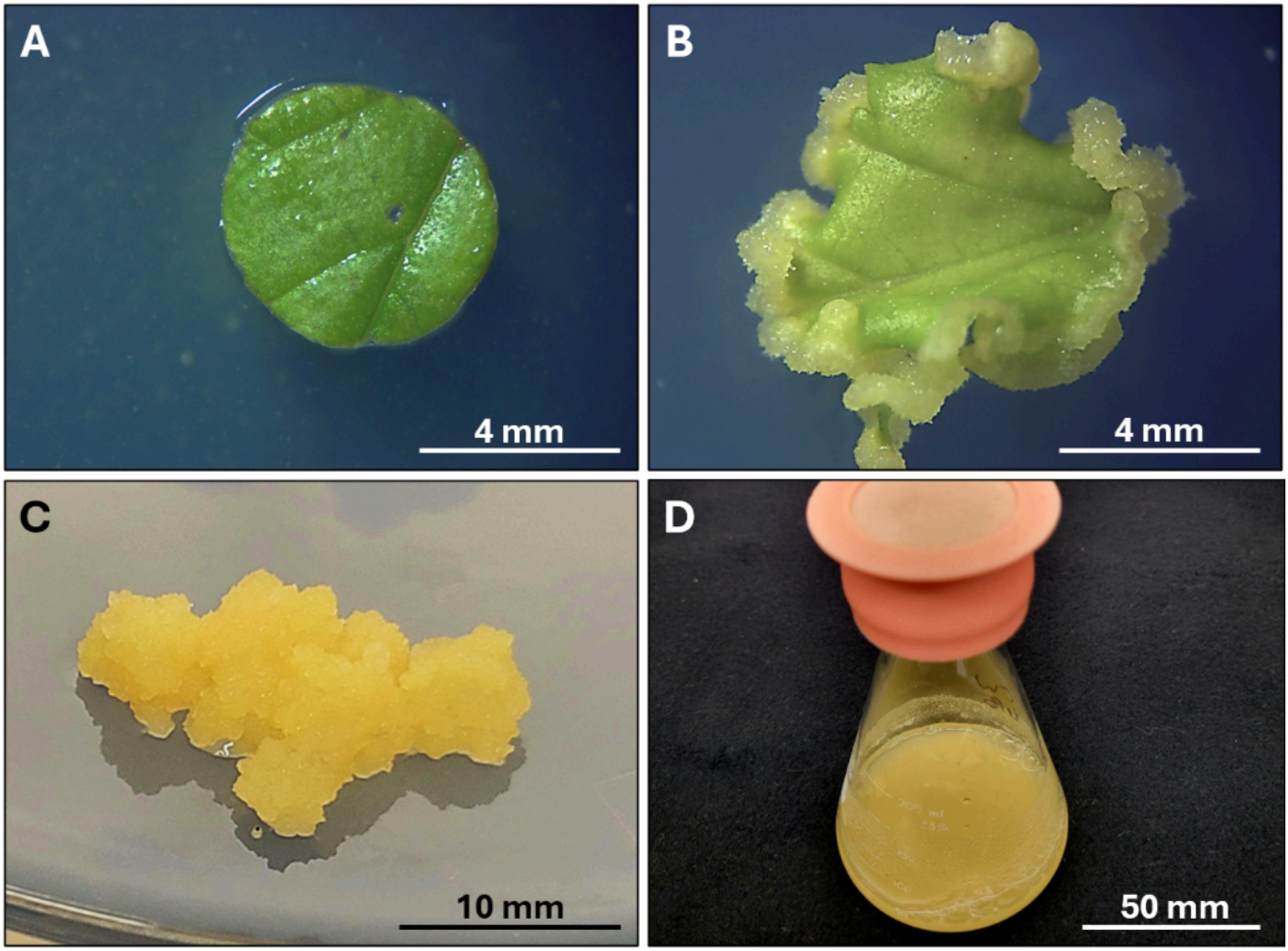


Figure 1

Establishment of transgenic cannabis cell suspensions. A) Cannabis leaf tissue is excised and subjected to Agrobacterium-mediated transformation (size bar = 4 mm). B) Transformed leaf explants are cultured on callus induction medium containing appropriate antibiotic to select for transgenic cell growth (size bar = 4 mm). C) Following callus formation, an individual section of callus originating from one transgenic event is excised and proliferated on callus induction media (size bar = 10 mm). D) Transgenic callus is transferred to liquid suspension culture medium in Erlenmeyer flasks and maintained under continuous shaking to establish stable transgenic cell suspension lines (size bar = 50 mm).

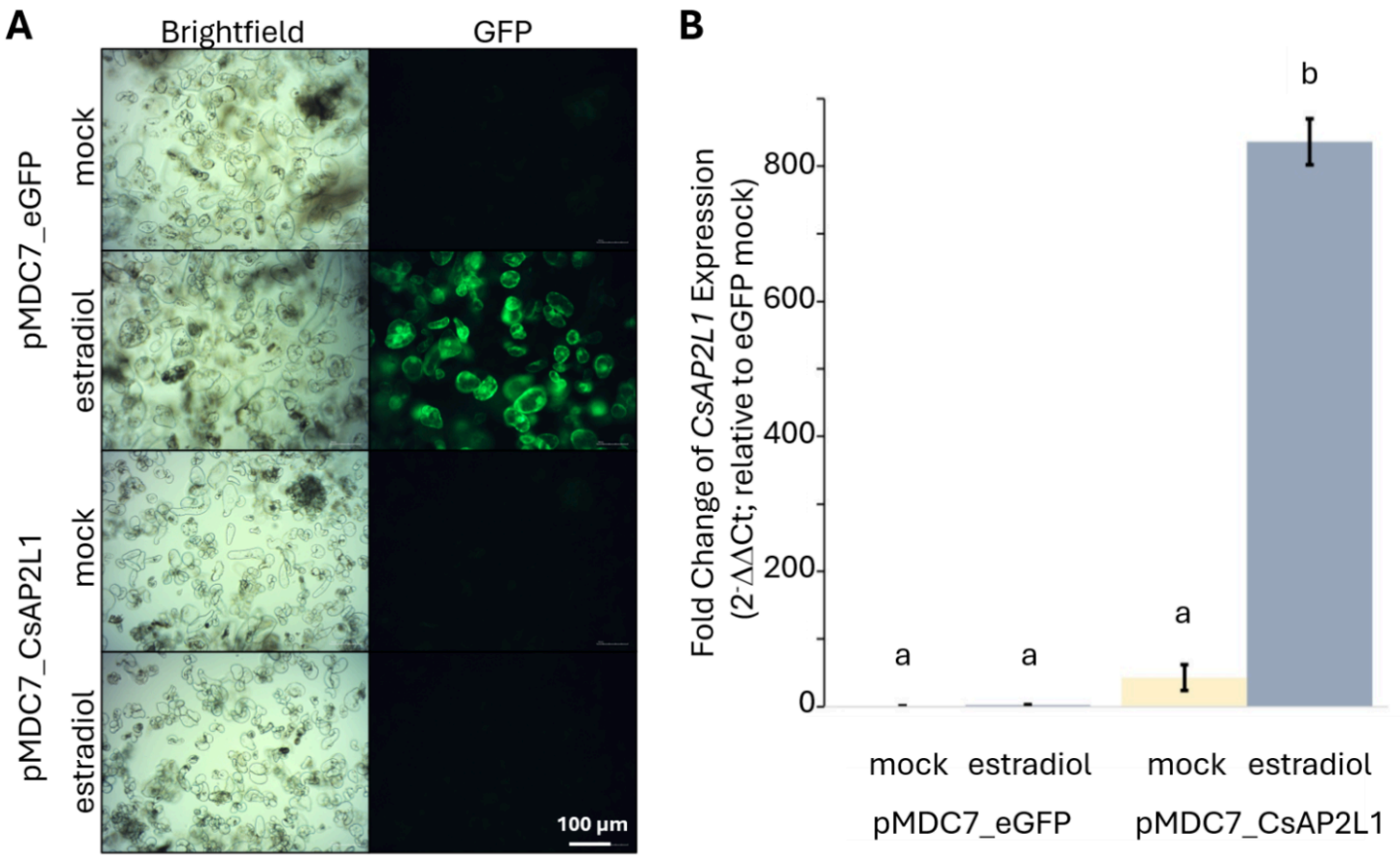


Figure 2

Estradiol-inducible gene expression in cannabis cell suspensions. A) Brightfield and GFP fluorescence microscopy of cannabis cell suspensions transformed with either the estradiol-inducible control construct pMDC7_eGFP or the estradiol-inducible *CsAP2L1*construct pMDC7_CsAP2L1. Cell suspensions were mock treated (DMSO solvent only) or treated with 1 µM estradiol for 24 h and imaged under identical acquisition settings (size bar = 100 µm). B) qRT-PCR analysis of *CsAP2L1* transcript levels in cannabis cell suspensions transformed with pMDC7_eGFP or pMDC7_CsAP2L1 and mock treated (DMSO solvent only) or treated with 1 µM estradiol for 24 h. Bars represent mean fold change in *CsAP2L1* expression (2 using the *CsEF1a* as a housekeeping transcript control) relative to estradiol-treated pMDC7_eGFP mock control, with error bars indicating standard error of the mean (n=3). Results assessed by one-way ANOVA with Tukey's HSD post-hoc test, $F(3, 8) = 441.98$, $p < 0.001$. Groups sharing the same letter are not significantly different ($p > 0.05$).

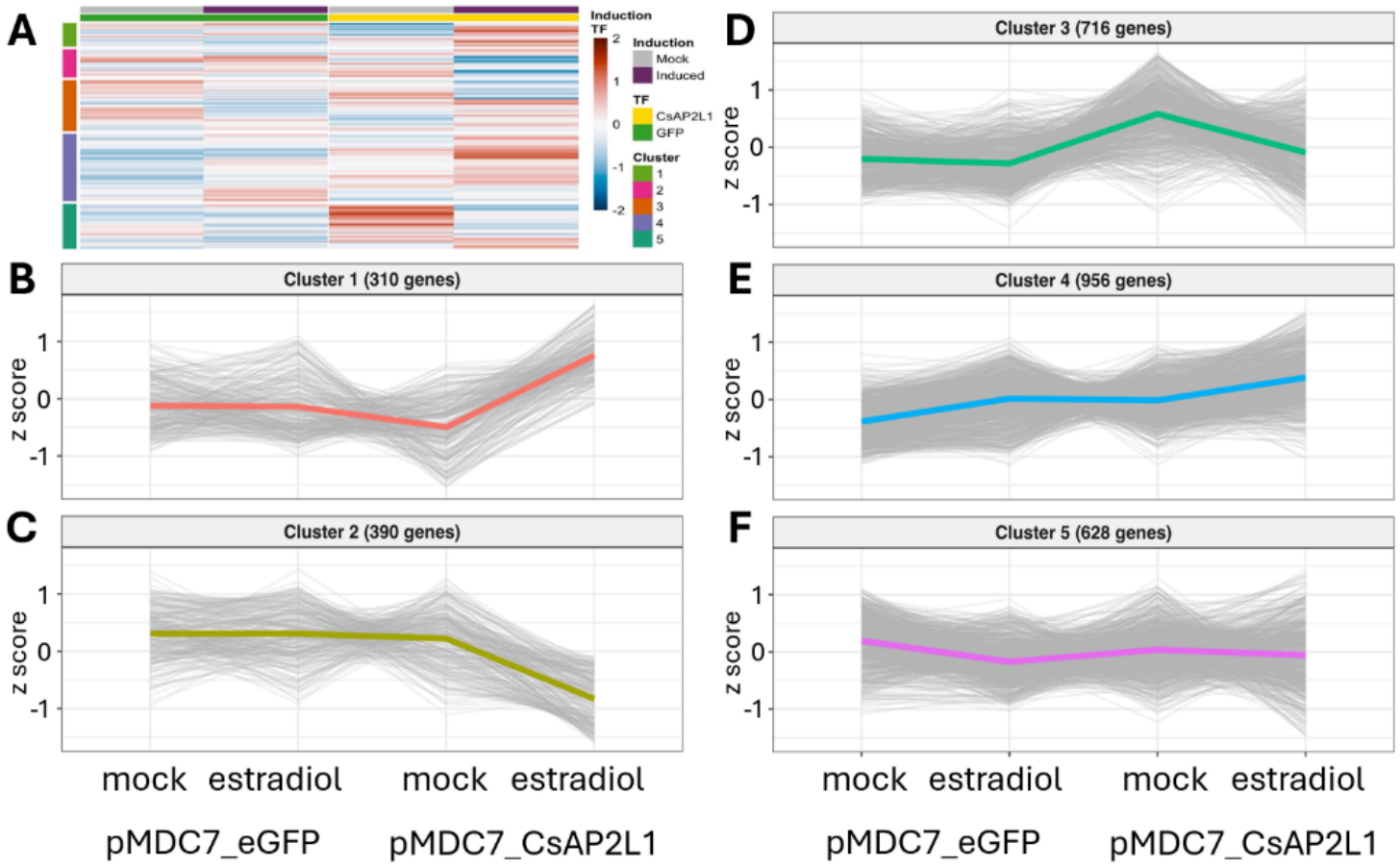


Figure 3

CsAP2L1 overexpression drives coordinated gene expression changes in select hierarchical clusters. A) Heatmap of all genes assigned to five hierarchical clusters, using variance-stabilized, z-scored expression values summarized as mean expression per condition (GFP mock, GFP induced, CsAP2L1 mock, CsAP2L1 induced). For each gene, vst-transformed expression values were averaged across biological replicates within each of the four conditions to generate a genes × 4 conditions matrix, which was then row-scaled prior to visualization. Genes were grouped into five hierarchical clusters based on their multi-condition expression profiles and displayed as contiguous blocks from top to bottom according to the specified cluster number. Columns are annotated by construct (CsAP2L1 vs GFP; top bar) and induction status (mock vs induced; second bar). The color scale represents gene-wise z-scores (−2 to +2) using a diverging blue–white–red palette, where blue indicates relatively lower and red indicates relatively higher expression compared to each gene's mean across the four conditions. B-F) Spaghetti plots of z-scaled expression for the five hierarchical clusters. Gray lines show the expression trajectory of each individual gene across the four experimental conditions (GFP mock, GFP induced, CsAP2L1 mock, CsAP2L1 induced). Overlaid on each panel is a bold colored line that represents the cluster centroid (mean expression of all genes in that cluster). The number of genes per cluster is shown in the plot title. The x-axis follows the biological order of treatment (mock to induced for each construct).

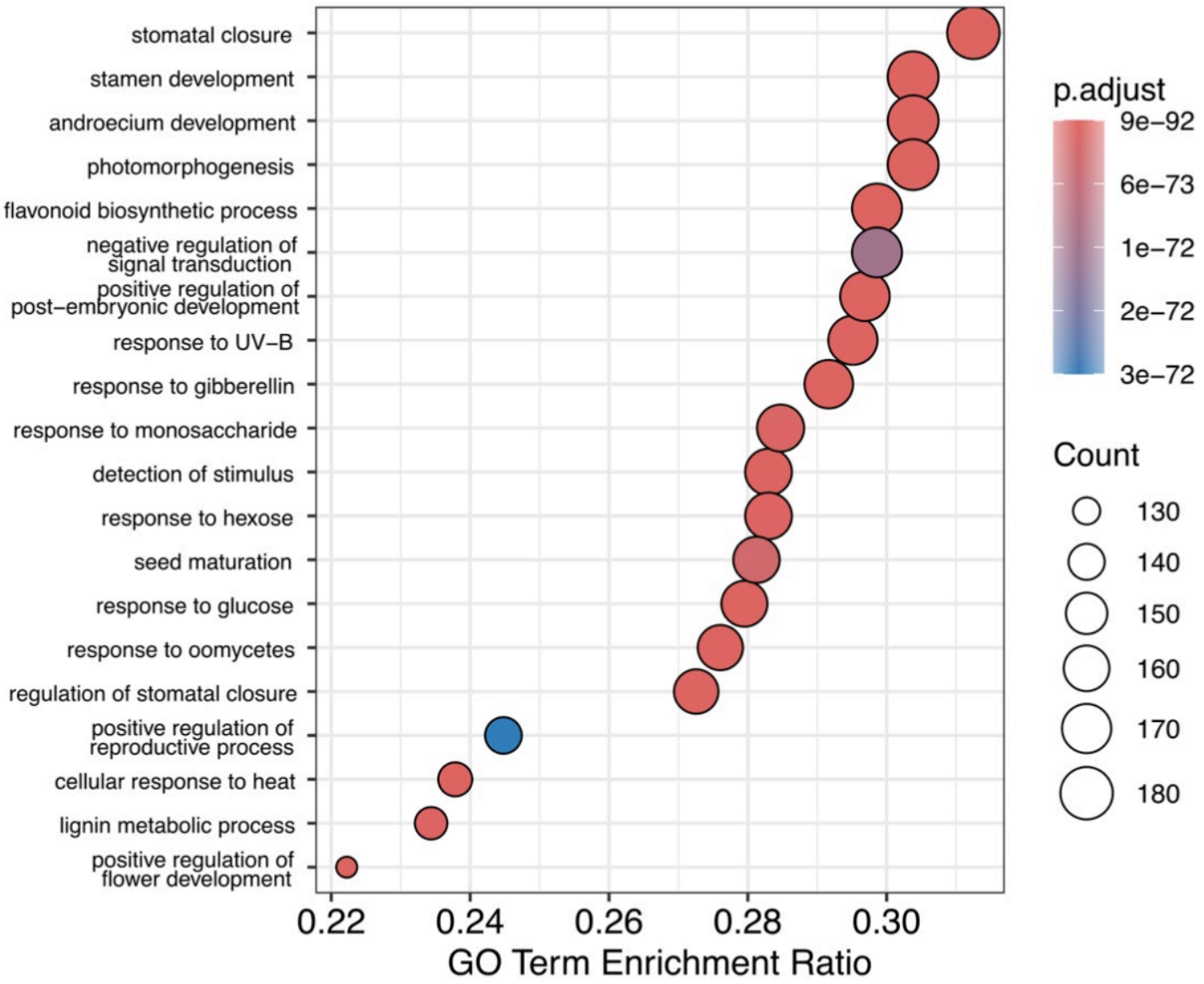


Figure 4

Gene Ontology enrichment analysis of biological processes in the CsAP2L1-upregulated hierarchical cluster. Dot plot showing the top 20 most significantly enriched Gene Ontology (GO) biological process terms for genes within the CsAP2L1-induced expression cluster. The x-axis represents the enrichment ratio (number of genes in the cluster annotated to each GO term divided by the expected number based on genome-wide frequency). Dot size indicates the number of genes annotated to each term (gene count), while dot color represents statistical significance (adjusted p-value). GO enrichment analysis was performed using the enrichGO R package.

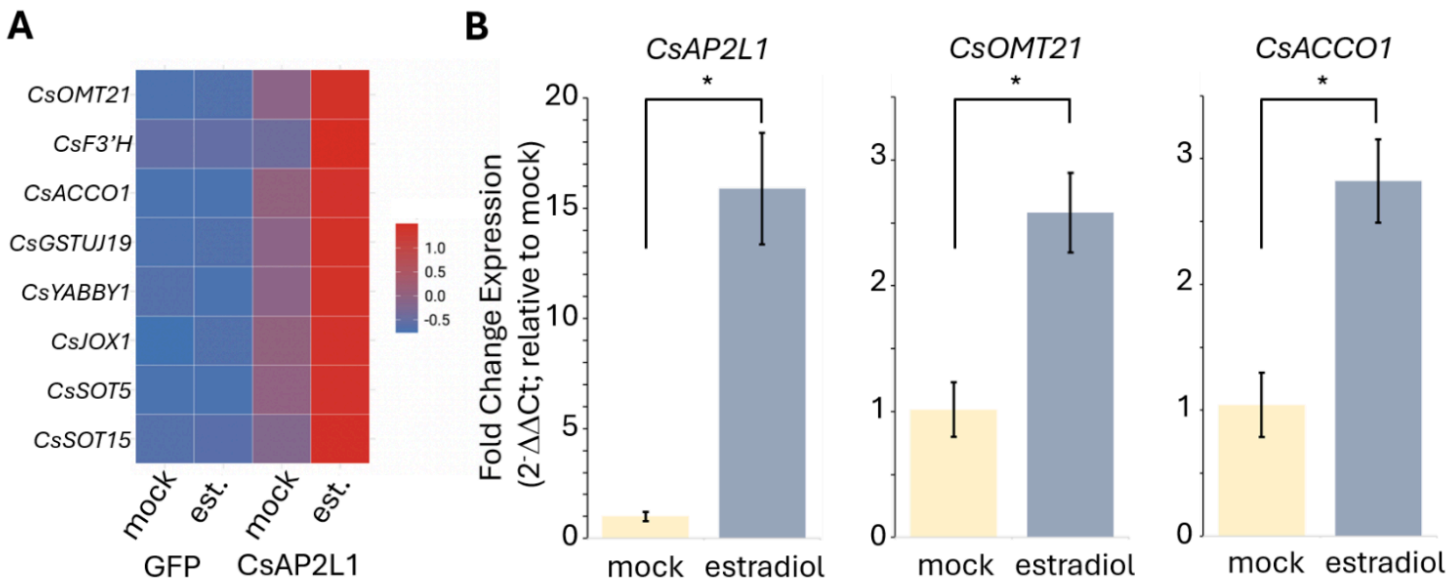


Figure 5

Expression profiles of flavonoid associated genes within the CsAP2L1-upregulated hierarchical cluster. A) Heatmap displaying row-normalized expression values from the RNA-seq analysis for eight flavonoid associated genes (*CsOMT21*, *CsF3'H*, *CsACCO1*, *CsGSTUJ19*, *CsYABBY1*, *CsJOX1*, *CsSOT5*, and *CsSOT15*) identified within the CsAP2L1-induced hierarchical cluster. Expression values represent mean row-normalized (z-score transformation) read counts from biological triplicates for each treatment condition (est. = estradiol). Color gradient represents relative expression levels. B) qRT-PCR validation of *CsAP2L1* and target gene expression changes in cell suspensions that were mock treated (DMSO solvent only) or treated with 1 µM estradiol for 24 h. Bars represent mean fold change in *CsAP2L1*, *CsOMT21*, and *CsACCO1* expression (2 using the *CsEF1a* as a housekeeping transcript control) relative to the mock control, with error bars indicating standard error of the mean (n=3). Asterisks denote statistically significant differences between samples (Welch's t-test $p<0.01$).

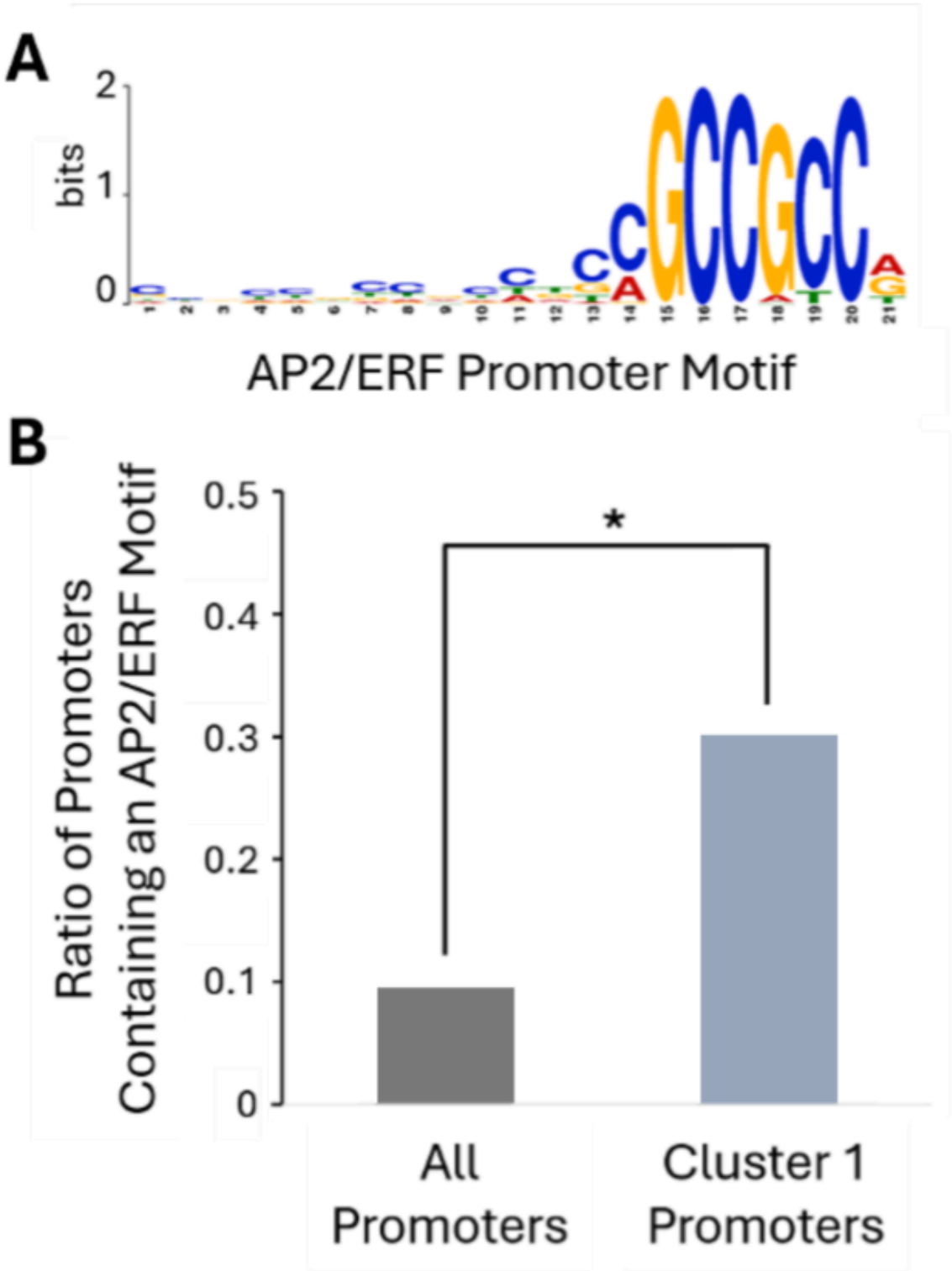


Figure 6

AP2/ERF motif enrichment in Cluster 1 promoters. (A) Sequence logo representing the AP2/ERF-class promoter motif enriched in CsAP2L1-upregulated genes, generated from promoter regions (1 kb upstream of translation start sites) of CsAP2L1 hierarchical Cluster 1 genes identified by MEME Suite SEA (Simple Enrichment Analysis) analysis. (B) Bar plot comparing the proportion of promoters containing AP2/ERF binding motifs between genes in the AP2L1-responsive hierarchical cluster (37/119

promoters, 31.1%) and the background set of extracted promoters of expressed genes from the RNA-seq analysis (1,342/13,678 promoters, 9.8%). Asterisk indicates significantly higher prevalence of the AP2/ERF motif in Cluster 1 promoters (Fisher's exact test as implemented in MEME Suite SEA: p=5.41e-5).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable2.csv
- SupplementaryTable3.csv
- SupplementaryTable4.csv