A redundant transcription factor network steers spatiotemporal Arabidopsis triterpene synthesis

Trang Hieu Nguyen[1,2], Louis Thiers[1,2,3], Alex Van Moerkercke[1,2], Yuechen Bai[1,2,11], Patricia Fernández-Calvo[1,2,12], Max Minne[1,2], Thomas Depuydt[1,2], Kevin Verstaen[4,5], Gert Van Isterdael[6,7], Hans-Wilhelm Nützmann[8,9], Anne Osbourn[8], Yvan Saeys[4,5], Bert De Rybel[1,2], Klaas Vandepoele[1,2,10], Andrés Ritter[1,2,13] & Alain Goossens[1,2,13*]

[1] Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium.

[2] VIB Center for Plant Systems Biology, Ghent, Belgium.

[3] Plant Sciences Unit, Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Melle, Belgium.

[4] Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium.

[5] VIB Center for Inflammation Research, Data Mining and Modelling for Biomedicine, Ghent, Belgium.

[6] VIB Flow Core, VIB Center for Inflammation Research, Ghent, Belgium.

[7] Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium.

[8] Department of Biochemistry and Metabolism, John Innes Centre, NR4 7UH Norwich, United Kingdom.

[9] Department of Biology and Biochemistry, The Milner Centre for Evolution, University of Bath, BA2 7AY Bath, United Kingdom.

[10] Bioinformatics Institute Ghent, Ghent University, Technologiepark 71, 9052 Ghent, Belgium.

[11] Present address: School of Life Sciences, Fudan University, Shanghai 200438, P.R. China

[12] Present address: Misión Biolóxica de Galicia, CSIC, Pontevedra, Spain and Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM)-Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Campus Montegancedo UPM, 28223 Pozuelo de Alarcón (Madrid), Spain.

[13] These authors jointly supervised this work: Andrés Ritter, Alain Goossens.

✉e-mail: alain.goossens@psb.vib-ugent.be

Plant specialized metabolites modulate developmental and ecological functions and comprise many therapeutic and other high-value compounds. The mechanisms determining their cell-specific expression remain however unknown. Here, we describe the transcriptional regulatory network that underlies cell-specific biosynthesis of triterpenes in *Arabidopsis thaliana* root tips. Expression of thalianol and marneral biosynthesis pathway genes depends on the phytohormone jasmonate and is limited to outer tissues. We show that this is promoted by the activity of redundant bHLH-type transcription factors from two distinct clades and coactivated by homeodomain factors. Conversely, the DOF-type transcription factor DAG1 and other regulators prevent expression of the triterpene pathway genes in inner tissues. We thus show how precise expression of triterpene biosynthesis genes is determined by a robust network of transactivators, coactivators and counteracting repressors.

One-Sentence Summary: A robust transcriptional network steers cell-specific and jasmonate-inducible triterpene biosynthesis in Arabidopsis root tips.

Plants have the ability to synthesize a wide range of specialized metabolites with important roles in growth, development, defense, and/or interactions with the environment[1,2]. Accordingly, together with the structural variety of the metabolites, complex regulatory networks have coevolved to ensure correct spatiotemporal activation of the corresponding biosynthetic pathway[2,3]. Although, in the past, several regulators involved in the hormone- or stress-mediated induction of specialized metabolism have been identified[4], the cell-specific regulation of specialized metabolic pathways remains elusive. For instance, some terpenes, one of the major class of plant specialized metabolites[5], can accumulate constitutively, while the production of others is (transiently) induced by a variety of developmental or environmental signals, often in association with biotic interactions and in a tissue-specific manner[6-8]. The model plant *Arabidopsis thaliana* (Arabidopsis) synthesizes several types of terpene compounds[5,9,10], such as triterpenes that are built from the general precursor oxidosqualene through the action of triterpene synthases (TTSs) as the first committed step. The Arabidopsis genome possesses five TTSs organized in "operon-like" biosynthetic gene clusters (BGCs), from which the thalianol and marneral BGCs have been characterized best (Supplementary Fig. 1)[11-14]. These BGCs are predominantly expressed in roots[7,11,12], and their expression results in the accumulation and/or secretion of root triterpenes, such as various thalianol derivatives that selectively modulate Arabidopsis root growth/development and the recruitment of the Arabidopsis root microbiota[7,15]. The marneral and thalianol BGCs are present in transcriptionally active chromosomal sites in roots and localized to heterochromatic chromosomal domains when silenced in leaves[13], and the involvement of localized chromatin modifications in their regulation has been reported[13,14]. We recently revealed that jasmonate (JA) activates gene expression and increases metabolite accumulation of triterpenes from the thalianol pathway[15]. JA is a conserved phytohormone, widely renowned for its elicitor action in the production of a wide range of plant specialized metabolites[4,16,17]. In the absence of JA, the signaling pathway is repressed by a protein corepressor complex, composed of the JASMONATE ZIM-DOMAIN (JAZ), the NOVEL INTERACTOR OF JAZ and TOPLESS proteins, which binds to specific transcription factors (TFs) to impede the JA response[18-21]. Upon stress, JA is synthesized under its bioactive form, JA-Ile, which is perceived by a coreceptor complex composed of the CORONATINE INSENSITIVE 1 (COI1) F-box protein that associates with CUL1, Rbx1, and Skp1-like proteins to assemble the SCF^{COI1} ubiquitin-ligase complex[21-23]. The SCF^{COI1} complex will subsequently ubiquitinate the JAZ proteins, resulting in their proteasomal degradation and thereby releasing the inhibited TFs[17,19,21,24]. Despite this knowledge and the central role of the Arabidopsis triterpenes in metabolic regulation, TFs acting on the Arabidopsis triterpene BGCs have not been identified yet. Here, we reveal a robust transcriptional gene regulatory network that determines the cell-specific and JA-inducible expression of triterpene biosynthesis genes in Arabidopsis root tips, thereby exposing a fundamental regulatory mechanism of specialized metabolite production in plants.

Results

JA triggers root cell-specific expression of triterpene biosynthesis in a COI1-dependent manner

Our previous data, together with available transcriptome datasets of the JA repressor mutants *ninja* and *jazQ*, indicated that JA activates expression of marneral and thalianol BGCs[15,25,26]. We first corroborated these data in roots of 12-day-old Arabidopsis Col-0 wild-type (wt) and *coi1-1* mutant seedlings treated either with mock or 50 µM JA for 6 h (Fig. 1a). Expression of genes from both the thalianol BGC, i.e. *THAS, THAH* and *THAO*, and the marneral BGC, i.e. *MRN1, CYP705A12*

and *MRO*, were strongly induced by JA treatment in wt roots, whereas this effect was abolished or significantly lowered in the *coi1-1* background (Fig. 1a). Correspondingly, roots of JA-treated wt seedlings accumulated significantly higher amounts of thalianol compared to mock-treated plants (Fig. 1b). In the *coi1-16* mutant, the JA-mediated increase in thalianol content was largely reduced; no significant differences in thalianol levels were observed in mock conditions in the *coi1-16* mutant versus the wt (Fig. 1b). Together, these data support the control of root triterpene biosynthesis by COI1-dependent JA signaling.

To further map the spatiotemporal regulation of the marneral and thalianol BGCs, we profiled the transcriptional changes in root tips treated for 2 h with mock or 50 µM JA at single-cell resolution (Fig. 1c and Supplementary Fig. 2-3). Though JA induced marked transcriptional changes in most sampled cells, the UMAP representations of mock and JA transcriptomes suggest a differential responsiveness of different cell types to JA (Fig. 1c and Supplementary Fig. 3-4). The analysis of Euclidean distances between transcript profiles of the two treatments indicated that cells belonging to the lateral root cap (LRC) or the metaxylem were more responsive, and others, such as vascular initial or pericycle cells, had a reduced responsiveness to JA treatment (Supplementary Fig. 4). Genes in the thalianol BGC showed strong expression upon JA treatment in cells of the root cap, epidermis, cortex, columella, atrichoblast and trichoblast (Fig. 1d and Extended Data Fig. 1). We corroborated these results using a promoter reporter line expressing NLS-GFP under the control of the *THAS* promoter[15], which largely recapitulated the predicted single-cell RNA sequencing (scRNAseq) expression pattern for the *THAS* gene, with low expression in mock conditions and a strong NLS-GFP signal after JA treatment in the root cap and epidermis (Fig. 1e). The genes in the marneral BGC showed similar expression patterns as compared to those in the thalianol BGC, with a strong expression pattern upon JA treatment located at the LRC, epidermis and atrichoblast (Fig. 1d and Extended Data Fig. 2). A *ProMRO:GFP-GUS* reporter line confirmed this predicted scRNAseq expression pattern (Fig. 1e). Overall, these results indicate that thalianol and marneral BGCs are specifically expressed in the outer cell layers of the root tips, and their expression is strongly activated by JA.

MYC TFs partially regulate thalianol and marneral gene expression in the root tip

The main regulator of the JA-triggered stress response in Arabidopsis is the clade IIIe basic helix-loop-helix (bHLH) TF MYC2, which is predominantly expressed in roots[20,25,27] and acts redundantly with MYC3 and MYC4 for many JA-dependent responses[20,28]. To assess whether thalianol and marneral BGC expression depends on these TFs, we first analyzed BGC transcript accumulation by qPCR in the single *myc2* and the triple *myc2 myc3 myc4* (*mycT*) mutants (Fig. 2a). No differences in transcript accumulation between wt and *myc* mutant lines could be observed for any of the BGC genes in mock conditions. However, after 6 h of JA treatment, JA-induced expression of most genes in the thalianol and marneral BGCs was significantly reduced in roots of the *myc2* and *mycT* mutants compared to the wt (Fig. 2a). Thalianol metabolite profiling of *myc2* and *mycT* roots indicated that though *myc2* roots show a decrease in JA-induced thalianol levels as compared to the wt (Fig. 1b), the difference is not significant yet. In contrast, the thalianol content is significantly reduced compared to the wt after JA treatment of the *mycT* mutants (Fig. 1b). In accordance with the transcript data (Fig. 2a), and as observed for the *coi1-16* mutant, no significant differences were observed in thalianol levels in the *myc* mutants in mock conditions. Given that thalianol and marneral BGC expression was found to be confined to the outer cell layers of the root tips (Fig. 1d,e), we next evaluated the expression patterns of *MYC2*, *MYC3* and *MYC4* in our scRNAseq dataset. Surprisingly, *MYC2* was predicted to show a strong and ubiquitous expression in the scRNAseq dataset, independent of JA treatment (Extended Data Fig. 3a). As we

expected *MYC2* expression levels to be modified by stresses such as protoplast generation, we next analyzed *MYC2* transcript levels in bulk RNA extracted from intact roots or root protoplasts. This demonstrated that protoplast generation strongly increased the accumulation of *MYC2* transcripts as also already reported previously[29], even overruling the JA effect on *MYC2* levels in intact roots (Extended Data Fig. 3b). Notably, this effect was not observed for *MYC3* or *MYC4*, or any of the genes from the thalianol and marneral BGCs (Extended Data Fig. 3b). As such, expression of *MYC2* cannot be reliably predicted by the scRNAseq data. To overcome this limitation, we used reporter constructs that express *NLS-VENUS* driven by the corresponding *MYC* gene promoters[25]. The expression of *ProMYC2:NLS-VENUS*, but not that of *ProMYC3:NLS-VENUS* or *ProMYC4:NLS-VENUS,* correlated with that of the triterpene BGCs at the epidermis, cortex, LRC and columella (Fig. 2b). Notably, though JA treatment increased expression of the *ProMYC2:NLS-VENUS* construct, it did not alter the spatial expression pattern (Fig. 2b). Chromatin immunoprecipitation followed by sequencing (ChIPseq) has shown that both thalianol and marneral BGC promoters could be physically bound by MYC2 and MYC3[30]. To verify whether MYC2 can transactivate these genes, we made *fLUC* reporter constructs driven by 2-kb promoter fragments of the *THAS* (*ProTHAS*), *THAO* (*ProTHAO*) and *MRO* (*ProMRO*) genes and performed transactivation assays in tobacco protoplasts. MYC2 could activate all promoters by at least five-fold compared to the negative control (Fig. 2c), thereby further corroborating its direct role in the JA-mediated activation of triterpene biosynthesis. Taken together, our data support a role for the three MYC TFs in the regulation of triterpene biosynthesis, with MYC2 likely being the most important contributor in this cellular process.

The JA-induced expression of TFs from the bHLH clade IVa specifically activates triterpene gene expression in the outer layers of the root tip

While the above-described results demonstrate a role for MYC2 as transcriptional activator for both thalianol and marneral BGCs, the qPCR analyses of the single and triple *myc* mutants suggest the existence of additional factors controlling the JA response. Likewise, the fact that ubiquitous expression of *MYC2* following protoplasting in the scRNAseq data did not coincide with ubiquitous expression of the BGCs suggests the existence of additional factors controlling their spatiotemporal expression (Fig. 1d,e and Extended Data Fig. 1-3). To identify these additional factors, we used our scRNAseq dataset to generate a gene co-expression matrix for *THAS* across all root tip cell types (Extended Data Fig. 4 and Supplementary Table 1). As expected, a list of the 100 most co-expressed genes included the thalianol pathway genes *THAO*, *THAH*, *THAR1*, *THAR2* and *THAA1,* and the marneral pathway genes *MRN*, *MRO* and *CYP75A12*. Furthermore, many known JA-responsive genes were found to be co-expressed, such as JA and glucosinolate biosynthesis genes, confirming the robustness of this dataset. This analysis also identified *bHLH19* and *bHLH20* (*NAI1*), both bHLH clade IVa TFs, which is in agreement with previously reported functions of such clade IVa bHLH TFs in *C. roseus* and *M. truncatula* as regulators of terpene biosynthesis[31,32]. The Arabidopsis bHLH clade IVa also includes *bHLH18* and *bHLH25*, yet the latter two showed overall reduced expression levels in our dataset (Extended Data Fig. 5a). Further predictions of expression by scRNAseq and validation by qPCR indicated that JA induces expression of these clade IVa TFs in similar cell types as those of the triterpene BGCs, i.e. in the epidermis, atrichoblast, cortex, columella and LRC (Fig. 3a and Extended Data Fig. 5a). A reporter line carrying the *ProbHLH19:GFP-GUS* gene confirmed expression in the outer cell layers of wt root tips, in which *bHLH19* expression was highly JA induced in the tissues of the stele, LRC, epidermis and cortex and, to a lesser extent, in columella cells (Extended Data Fig. 5b). Moreover, the JA inducibility of *bHLH19* and *bHLH20* was dependent on COI1 (Extended Data Fig. 5c). We

then assessed whether JA-induced expression of the clade IVa bHLH genes involves the MYCs. Compared to the wt, JA-induced expression was significantly reduced for all clade IVa genes in the *mycT* line (Extended Data Fig. 6a). The promoter regions of Arabidopsis bHLH clade IVa genes contain a G-box (Supplementary Table 2) and have been shown to be bound by MYC2[30,33]. We accordingly tested whether MYC2 could transactivate these promoters and if they can be transactivated by themselves. Because *bHLH19* and *bHLH20* showed a strong and significant upregulation upon JA treatment and were predicted to be the most highly expressed in root tips (Fig. 3a and Extended Data Fig. 5-6), we cloned 1926-bp (*ProHLH19*) and 2000-bp (*ProbHLH20*) promoter fragments and performed transactivation assays using wt MYC2, desensitized MYC2^{D105N}, bHLH19 and bHLH20. The MYC2^{D105N} mutation abolishes the interaction between MYC2 and most JAZ proteins, by which transactivation becomes typically more pronounced[34]. Strong transactivation of both *ProHLH19* and *ProHLH20* was triggered by MYC2^{D105N}, whereas wt MYC2 had a moderate effect (Extended Data Fig. 6b). Conversely, bHLH19 and bHLH20 only had minor transactivation effects on their respective promoters. Taken together, our data indicate that, like the genes of the thalianol and marneral BGCs, the expression of the Arabidopsis clade IVa bHLH TF genes is induced by JA in outer root tip cell layers through the canonical signaling module that involves COI1 and MYCs.

To verify whether the bHLH clade IVa TFs could also directly modulate the expression of thalianol and/or marneral biosynthesis genes, we first performed transient promoter transactivation assays in tobacco protoplasts (Fig. 3b and Supplementary Fig. 5). This analysis suggested both redundant and additive activities for these TFs. For instance, when tested alone, bHLH19, bHLH20 and bHLH25 showed similar capacities and preferences for transactivating BGC promoters. Combining bHLH19 with bHLH18, bHLH20 or bHLH25 increased transactivation of the thalianol BGC promoters, as compared to bHLH19 alone, reflecting the additive contributions of these TFs in the regulation of triterpene BGC expression (Fig. 3b). Similarly, bHLH20 with either bHLH18 or bHLH19 resulted in an additive transactivation of the marneral promoters. These results support the concerted actions of MYC and bHLH clade IVa TFs to activate the expression of triterpene BGCs. We next generated mutant lines for the four loci encoding the bHLH clade IVa TFs (Supplementary Fig. 6). These mutations were introduced in both the wt and the *mycT* background, resulting in two independent quadruple mutant lines hereafter named *bHLHIVq-1* and *bHLHIVq-2,* and a septuple mutant line named *mycT;bHLHIVq*. We inquired whether the expression of triterpene BGCs was compromised in these lines after JA induction. Transcript levels of all triterpene genes were already reduced in the absence of JA in the *bHLHIVq* lines compared to the wt (Fig. 3c). Furthermore, the *bHLHIVq* lines showed an overall significantly reduced JA induction of both marneral and thalianol BGCs that was more pronounced than in the *mycT* mutant for the thalianol genes, further supporting a major role for the bHLH clade IVa TFs in the regulation of this pathway. Accordingly, the *mycT;bHLHIVq* septuple mutant showed even less JA induction for most tested triterpene BGCs compared to either the *bHLHIVq* or the *mycT* lines (Fig. 3c). Remarkably, in most cases, like for *CYP705A12, THAS, THAO* and *THAH*, the JA induction of triterpene gene expression in the *mycT;bHLHIVq* mutant was attenuated to levels comparable to those in the *coi1-16* mutant, which is a weak *coi1* allele. In accordance with the transcript data, also the thalianol content of the *bHLHIVq* mutant was significantly reduced compared to wt (Fig. 1b). The *mycT;bHLHIVq* mutants showed an additional reduction (though not significant) compared to the *mycT* or *bHLHIVq* mutants after JA treatment. Together, these data confirm the redundant roles of bHLH clade IVa and MYC TFs in the regulation of triterpene biosynthesis in Arabidopsis roots.

Cell-specific expression of triterpene BGCs is set by a network of redundant coactivators and repressors

Although the transcription of the genes encoding the bHLH clade IVa TFs and of the triterpene BGCs is activated by MYC2, these genes maintain cell-specific expression patterns in our scRNAseq datasets (Fig. 1d,e and Extended Data Fig. 1-2), contrary to MYC2 (Extended Data Fig. 3a). Hence, we questioned whether additional regulatory components may confine spatial expression of these genes, and overrule JA or other elicitor effects such as protoplast generation (Extended Data Fig. 3b). Triterpene BGCs have been shown to be located at transcriptionally active chromosomal sites in root tissues[13]. We therefore postulated that cell-specific expression of triterpene BGCs is controlled either by (*i*) 'repressive' TFs that are constitutively and specifically expressed in inner root tip tissues to thereby suppress the expression of triterpene biosynthesis genes, or (*ii*) additional 'activating' factors that are specifically expressed in outer root tip tissues and whose presence is essential to coactivate the transcription of triterpene pathway genes. We searched for such TFs using a motif collection that consists of 1,699 position weight matrices, representing the DNA-binding sites for 1,143 Arabidopsis TFs. We accordingly mapped TFs potentially binding to every gene in the thalianol and marneral BGCs, as well as those encoding the MYC and bHLH clade IVa TFs. This analysis identified a set of 21 TFs, of which eleven belonged to the DNA-binding one finger (DOF) family and seven to the homeodomain glabrous (HDG) family (Fig. 4a, Supplementary Table 3). Most of the DOF TFs were predicted to be expressed in the inner root tip cell layers, whereas HDG TFs were rather predicted to be expressed in outer root tip cell layers (Extended Data Fig. 7). Markedly, expression of all of these DOF and HDG TFs was predicted insensitive to JA treatment in the outer root tip cell layers, which was further confirmed by analysis of root tips of a mock- and JA-treated *ProDAG1:GUS* reporter line (Fig. 4b,c and Extended Data Fig. 7).

Members of the HDG family have been shown to be involved in root terpene biosynthesis and epidermis development[35]. For this reason, we selected GL2, HDG2 and HDG5 for further characterization and performed transient promoter transactivation assays in tobacco protoplasts. Transfection with either GL2, HDG2 or HDG5 alone did not markedly affect the activity of the *ProTHAS* reporter (Extended Data Fig. 8). However, cotransfection of GL2 with either MYC2^{D105N} or bHLH19 and bHLH20 had a pronounced synergistic effect on BGC reporter activities (Fig. 4d, Extended Data Fig. 8 and Supplementary Fig. 7b). The effect of the other HDGs was less pronounced or consistent, but nonetheless additive trends could be observed in some combinations (Extended Data Fig. 8). Overall, this supports a redundant role for these HDG TFs as coactivators of the root triterpene BGCs. We next carried out a qPCR-based expression analysis in two available *gl2* mutant alleles[36,37]. The effects of loss-of-GL2-function in the two examined alleles resulted in a consistent decrease in both the basal expression and JA induction of root triterpene BGCs, being most pronounced for the marneral BCG (Fig. 4e).

For the DOF TFs, we selected DAG1 and vDOF1 (DOF4.6), given their reported opposed expression patterns compared to the triterpene BGC genes (Fig. 4b and Extended Data Fig. 7)[38,39] and their transcriptional repressor activities[39,40]. DAG1, but not vDOF1, showed strong repression activities on promoters in both the thalianol and marneral BGCs, as well as on *ProbHLH19* in protoplast assays (Fig. 4d, Extended Data Fig. 9 and Supplementary Fig. 7a). Furthermore, cotransfection of DAG1 with either MYC2^{D105N} or bHLH19 and bHLH20 counteracted the transactivation capacities of the transcriptional activators (Fig. 4d and Supplementary Fig. 7a). These results support a role for DAG1 as a cell-specific repressor of triterpene pathway genes. Hence, we subsequently analyzed whether the available *dag1* mutant or *ProCaMV35S:DAG1* (*DAG1-OE*) lines[41,42] show altered triterpene gene expression. Expression of *DAG1* by the

ProCaMV35S promoter, which drives ectopic expression in the outer root cell layers (Supplementary Fig. 8), indeed resulted in a significantly decreased expression of most genes in the triterpene BGCs under JA treatment as compared to wt plants (Fig. 4e), corroborating an *in planta* function for this TF as triterpene BGC repressor. No significant differences were observed between the *dag1* and wt lines, however, either in mock or JA conditions, likely pointing to redundant functions with yet undetermined additional repressors, plausibly other DOF TFs, to confine the expression of triterpene biosynthesis to specific root cells.

Discussion

Together with the vast variety in structures and bioactivities of specialized metabolites, complex regulatory networks co-evolved to ensure correct expression of the corresponding biosynthetic pathways. Whereas regulators involved in the stress-mediated induction of specialized metabolism have been identified repeatedly, knowledge of the regulators driving cell specificity of specialized metabolite pathways is scarce. Here, via scRNAseq analysis, we identified a transcriptional regulatory network that underlies both the cell-specific and stress-inducible expression of Arabidopsis triterpene biosynthesis in the root tip (Extended Data Fig. 10). Expression of the genes in the thalianol and marneral BGCs is enhanced by JA via the canonical signaling cascade, and limited to the outer layers of root tip tissues. The latter is consistent with the function played by these triterpenes in microbiome recruitment and modulation of root growth[7,15]. Previous work has shown that the spatial expression of the thalianol and marneral BGCs differs in the leaves and roots, where these clusters are silenced and transcriptionally active, respectively[13]. Here, we provide evidence for the mechanism of transcriptional regulation of these BGCs in root tips, through a complex interplay between cell-specific activator and repressor TFs. A redundant set of bHLH-type TFs from two distinct clades, assisted by HDG-type coactivator TFs, promotes triterpene biosynthesis in the outer cell layers. Interestingly, the expression of the genes in the thalianol and marneral BGCs is rescued in the *jazQmycT* mutant line, suggesting that, in addition to the MYC TFs, other yet unidentified JAZ-interacting TFs may activate the expression of triterpene BGCs and/or bHLH clade IVa TFs[26]. This is also corroborated by the thalianol transcript and metabolite profiling of the *mycT*, *bHLHIVq* and *mycT;bHLHIVq* mutant roots in this study, in which the final reduction in thalianol accumulation after JA treatment is not as drastic as that in the *coi1-16* line, suggesting that other as yet unidentified COI1-dependent TFs may contribute to the activation of triterpene biosynthesis genes. Therefore, the model of the regulatory network that we propose here does not reflect the full complexity yet (Extended Data Fig. 10). Conversely, other regulatory TFs, hallmarked by the DOF-type TF DAG1, prevent the expression of the triterpene pathway genes in inner root tip tissues. Notably, in its function as a repressor, DAG1 is capable of overruling the activity of MYC2 and the other bHLH TFs in root tips. The observed redundancy within this model implicates the emergence of a very robust regulatory network, wherein both cell-specific activation and repression were subjected to strong selective pressure, to ensure that bioactive metabolites are produced at the right place at the right time to promote survival and interaction with the environment, and avoid self-toxicity or unnecessary metabolic investments. This regulatory network may also involve so-called super enhancers (SEs), a cluster of cis-regulatory elements that can increase the transcription of their cognate genes. Very recently, such an SE has been identified within the thalianol BGC, and its deletion resulted in the transcriptional repression of the thalianol BGC genes[43]. How specific such SEs may work, in terms of target genes and/or hormone-responsive and/or cell-specific expression, remains to be determined. We postulate that analogous regulatory frameworks are applicable to many other cell-specific plant specialized metabolite pathways, in Arabidopsis and beyond, and during evolution-

involved recruitment of identical, similar and/or distinct sets of redundant activators and repressors.

Methods

Plant material and treatments

The *coi1-1*, *coi1-16*, *myc2 (jin1-2)* and *myc2 myc3 myc4* mutants, and the *ProTHAS:NLS-GFP-GUS, ProMYC2:NLS-VENUS, ProMYC3:NLS-VENUS* and *ProMYC4:NLS-VENUS* reporter lines are in the *Arabidopsis thaliana* Col-0 background and were described before [20,25,27,44-46]. The *dag1*, *dag1 DAG1*-OE and *ProDAG1-GUS* lines are on the Wassilewskija (Ws) background and were described before[38,41]. The *ProMRO:GFP-GUS* and *ProbHLH19:GFP-GUS* lines were generated by floral dip of *A. thaliana* in the Col-0 ecotype using *Agrobacterium tumefaciens* strain C58C1. Knock-out mutants of the *bHLH IVa* genes were generated in the Col-0 and *mycT* backgrounds using CRISPR-Cas9 genome editing technology. Single guide RNAs (sgRNAs) targeting the first or second exons of *bHLH18, bHLH19, bHLH20* and *bHLH25* were designed by CRISP-OR (http://crispor.tefor.net/)[47]. For each target gene, two specific sgRNAs were selected and synthesized as indicated in Supplementary Fig. 6, using the primers listed in Supplementary Table 4. A single CRISPR/Cas9 vector containing the *Cas9* gene and eight sgRNAs targeting all four *bHLH IVa* genes simultaneously was generated using GreenGate technology as previously described[48]. Briefly, for a pair of two gRNA target sites, two oligonucleotides with 16-bp overhangs were ordered (primers P25-P32; Supplementary Table 4). The primers were annealed on the template pEN-2x*AtU6*, then *BbsI*-ligated into the Golden Gate entry vectors pGG-A-*AtU6ccdB*-B, pGG-B-*AtU6ccdB*-C, pGG-C-*AtU6ccdB*-D or pGG-D-*AtU6ccdB*-E. Four Golden Gate entry modules, with each of them containing two sgRNAs were assembled into the binary vector pFASTGK-*AtCas9*-A-*ccdB*-G. CRISPR constructs were transfected into *A. tumefaciens* C58C1 (pMP90) and used to transform Arabidopsis Col-0 and *mycT* using floral dip. Primary transformants (T1) were selected on kanamycin (50 µg/mL) and the genomic region spanning the predicted Cas9 cut site was sequenced using primers P35-P42 (Supplementary Table 4). T2 plants were monitored for segregation of the T-DNA locus using kanamycin resistance and were genotyped using Cas9-specific primers (P33/34; Supplementary Table 4) to identify null segregants. In these plants, all four bHLH loci were reanalyzed by PCR to identify genotypes that were now non-chimeric and either homo- or hetero-allelic. The most promising plants were then propagated to T3, in which the absence of Cas9 and the homogenous state of the edited allele were confirmed by PCR and sequencing. The AGI identifiers for the genes used in this study are listed in Supplementary Table 5.

Arabidopsis seeds were surface-sterilized for 5 min in 70% ethanol to which 1% Tween-20 was added, followed by a 5-min incubation in 95% ethanol. Seeds were dried for 3 h before plating. For treatments, seeds were placed on nylon meshes (Sefar, 03-20/14) on 0.5x Murashige and Skoog (MS) (including vitamins), 0.5% MES (pH 5.8), 0.7% phytoagar plates. After stratification for two days at 4°C, the plates were transferred to standard growth conditions (21°C, 16-h/8-h light/dark regime), and grown vertically for eight days. For transcript analyses, the meshes with seedlings were then transferred to plates containing 50 µM JA in ethanol (Sigma-Aldrich) or mock (ethanol) and roots were collected 6 h after treatment. For confocal imaging and GUS histochemical analyses seedlings were transferred to plates containing 50 µM JA in ethanol or mock for 24 h. For comparative expression analysis in Col-0 and the *coi1-1* mutant, seedlings were first grown for 5 days on mock plates or plates containing 50 µM JA, respectively, to enable selection of homozygous *coi1-1* plants, and subsequently transferred to plates with ½ MS

(including vitamins), 0.5% MES (pH 5.8), 1% phytoagar, without JA and grown vertically on a nylon mesh for seven days. Prior to sampling the roots, the mesh was transferred to plates containing 50 µM JA or mock for 6 h. For metabolite analyses, 5-day-old seedlings were transferred on 6-well culture plates containing liquid ½ MS (including vitamins), 0.5% MES (pH 5.8), 1% sucrose and grown in standard light/dark growth conditions with shaking (90 rpm) for eight days as previously described[15]. For treatment, the medium was replaced with fresh liquid medium containing 50 µM JA in ethanol or mock, then roots were collected 24 h after treatment, flash-frozen and freeze-dried.

Subcellular protein localization

Fluorescence (GFP or Venus) was monitored in primary root cell tips of 5-day-old seedlings grown vertically on ½ MS agar medium. Before imaging, root tips were stained with propidium iodide and mounted on slides. Fluorescence was followed with a Zeiss LSM 710 confocal microscope using 40x magnification. Images were then processed using Zeiss Zen 2 software. All optical cross sections were made in the middle of the meristem, determined by the quiescent center on one end and the first elongating cortex cell on the other end.

GUS histochemical analysis

Histochemical GUS staining was performed in 5-day-old seedlings grown vertically on ½ MS agar medium. The plant material was incubated at 37°C, in the dark, for 2 h in a staining buffer containing 1 mM 5-bromo-4-chloro-3-indolyl ß-D-glucopyranoside sodium salt (X-Gluc), 0.5 % Triton X-100, 1 mM 5ethylenediaminotetraacetic acid (EDTA) pH 8.0, 0.5 mM potassium ferricyanide (K3Fe(CN)6), 0.5 potassium ferricyanide (K4Fe(CN)6) and 500 mM sodium phosphate buffer pH 7.0. The reaction was terminated by replacing the staining buffer with 70 % ethanol. The material was mounted in 25% lactic acid; 50% glycerol and analyzed with an Olympus BX51 microscope equipped with a Hitachi DK-H32 camera.

Promoter transactivation assays in tobacco protoplasts

For transactivation assays, the promoters of *bHLH19* (*ProHLH19*), *bHLH20* (*ProbHLH20*) were isolated using primer pairs P1/2 and P3/4 (Supplementary Table 4), respectively and cloned into pGG-AB before transferring to pGGIB-U1U2 using Golden Gate cloning technology (New England Biolabs). Promoters of *THAS* (*ProTHAS*), *THAH* (*ProTHAH*), *THAO* (*ProTHAO*), *MRO* (*ProMRO*) and *CYP705A12* (*ProCYP705A12*) were isolated using primer pairs P5/6, P7/8, P9/10, P11/12 and P13/P14 (Supplementary Table 4), respectively, and BP-recombined into pDONR221 (Invitrogen). The coding sequences (CDSs) of *bHLH18*, *bHLH19*, *bHLH20* and *bHLH25* were isolated using primers P15/16, P17/18, P19/20, and P21/22 (Supplementary Table 4), respectively, and cloned into pDONR207 (Invitrogen). The entry plasmids were sequence-verified and subsequently LR-recombined into pGWL7 for the promoters and p2GW7 for the CDSs using Gateway technology (Invitrogen®). The CDS of *DAG1*, vDOF1, GL2, HDG2 and HDG5 were isolated using primers P23/24, P69/70, P71/72, P73/74 and P75/76 (Supplementary Table 4), respectively, and cloned into pGG-CD before transferring to pGGIB-U1U2. We cloned MYC2 in $MYC2^{D105N}$ and p2GW7 as previously described[34].

Transient promoter transactivation assays in tobacco protoplasts were performed essentially as described by Vanden Bossche *et al.*[49]. In all cases, a *ProCaMV35S*-driven intron-containing β-glucuronidase (*GUS*) construct was used as control transformation. Normalization of firefly luciferase activity was achieved using a *ProCaMV35S*-driven *Renilla* luciferase cotransfection. Measurements were performed using the Dual-Luciferase® Reporter Assay System (Promega).

RNA extraction and qRT-PCR analysis
Total RNA was isolated from roots using the RNeasy Plant Mini Kit (Qiagen) and 1 µg was used for cDNA synthesis using the qScript cDNA SuperMix Kit (Quanta-Bio). Quantitative RT-PCR (qPCR) was performed as previously described[32], using primers P43-P64 (Supplementary Table 4). For normalization, housekeeping genes *PROTEIN PHOSPHATASE 2A SUBUNIT A3* (*PP2A; At1g13320*, primer pair P65/66) and *MONENSIN SENSITIVITY1* (*MON1; AT2G28390*, primer pair P67/68) were used.

Single-cell RNA sequencing (scRNAseq)
Single-cell RNA sequencing was performed as described in Wendrich *et al.*[50] with following modifications.

Protoplasting conditions and FACS
Six-day-old seedlings were transferred to 50 µM JA-supplemented medium or an equal volume of DMSO as mock treatment and allowed to grow for 2 h. Note that the DMSO mock sample was identical to the one described in Yang *et al.*[51], as these experiments were performed together. The root tips were then cut and incubated in protoplasting Solution B [1.5% (wt/vol) cellulysin and 0.1% (wt/vol) pectolyase in Solution A (600 mM mannitol, 2 mM $MgCl_2$, 0.1% (wt/vol) bovine serum albumin, 2 mM $CaCl_2$, 2 mM MES, 10 mM KCl, pH 5.5)] for approximately 1 h at room temperature. Cells were filtered through a 70-µm cell strainer and spun down at 200 g for 6 min, resuspended in Solution A, filtered through a 40-µm cell strainer and stained for live/dead using 4′,6-diamidino-2-phenylindole (DAPI) at a 14-µM final concentration. Cells were sorted on a BD Aria II and protoplasts without the DAPI signal were selected for further analysis.

10X Genomics sample preparation, library construction and sequencing
Sorted cells were centrifuged at 400 g at 4°C and resuspended in Solution A to yield an estimated concentration of 1,000 cells per µL. Cellular suspensions were loaded on a Chromium Single Cell 3′ GEM, Library & Gel Bead Kit (V3 chemistry, 10X Genomics) according to the manufacturer's instructions. Sequencing libraries were loaded on an Illumina HiSeq4000 and sequenced following recommendations of 10X Genomics at the VIB Nucleomics Core (VIB, Leuven).

Raw data processing and generation of gene expression matrix
Demultiplexing of the raw sequencing data was done by the 10x CellRanger version 3.1.0 software 'cellranger mkfastq', which wraps Illumina's bcl2fastq. The fastq files obtained after demultiplexing were used as input for 'cellranger count', which aligns the reads to the *Arabidopsis thaliana* reference genome (EnsembleTAIR10.40) using STAR and collapses them to unique molecular identifier (UMI) counts. The result is a large digital expression matrix, with cell barcodes as rows and gene identities as columns. Initial filtering in CellRanger recovered 11,569 cells for the mock-treated and 9,660 cells for the JA-treated sample. To ensure that we only used high-quality cells for further analysis, we used the filtered data provided by cell ranger. This corresponds to 26,756 mean reads per cell in the mock-treated sample and 34,275 mean reads per cell in the JA-treated sample.

Data analysis (clustering, identity assignment, differential gene expression, quality control and co-expression)
All analyses were performed in R (version 3.6.0). Preprocessing of the data was done by the scater package (version 1.10.1) according to the workflow proposed by the Marioni lab[52]. Outlier cells were identified on the basis of 2 metrics (library size and number of expressed genes) and were

tagged as outliers when they were 4 median absolute deviations away from the median value of these metrics across all cells. Normalizing the raw counts, detecting highly variable genes, finding clusters and creating t-distributed stochastic neighbor embedding (t-SNE) plots were done using the Seurat pipeline (version 3.2.3). Differential expression analysis for marker gene identification per subpopulation was based on the non-parametric Wilcoxon rank sum test implemented within the Seurat pipeline. Clusters with the same cell annotation based on gene expression analysis were combined to generate a more comprehensible dataset. Potential doublets were identified using the DoubletFinder algorithm (version 2.0.0)[53]. The number of high-confidence doublets was below 1% (113 out of 11,313 cells for the mock-treated sample and 96 out of 9,572 cells for the JA-treated sample). For comparing mock to JA treatments, averaged expression levels for each cell type were calculated on the log-normalized UMI counts of the scRNAseq data to build an expression matrix. Pairwise Euclidean distances were calculated between the control and JA-treated versions of the same cell type, as an estimator of the effect of JA treatment on the expression profile of each cell type. Genes co-expressed with the *THAS* gene were identified by computing the Pearson correlation coefficient on the log-normalized UMI counts from the scRNAseq dataset. Genes were next sorted from high to low correlation and the top 100 genes were extracted for further analysis.

Transcription factor mapping to promoters of the triterpene BGCs

TF motifs modeled as position weight matrices (PWMs) were retrieved from the CisBP 2.00 (downloaded on Dec 2019)[54] and JASPAR2020[55] databases and combined with a collection of manually curated motifs[56]. A total of 1,699 motifs, corresponding to 1,143 TFs, were collected. TF motifs were mapped to a gene's regulatory regions using Cluster Buster (version compiled on Sept 2017[57]) and FIMO (version 4.11.3[58]). Cluster Buster was run with the -c parameter set to 0 and PWMs were scaled to 0-100. FIMO was run with default parameters and PWMs were scaled to 0-1. The regulatory regions used for motif mapping were extracted from Arabidopsis Araport11[59], only retaining the longest splicing variants, downloaded from PLAZA Dicots 4.5[60]. These regions were defined as 5000 bp upstream and 1000 bp downstream from translation start/stop sites. Introns were included in the regulatory regions, while coding exons were excluded. If another gene was present within the 5000 bp-1000 bp window, this region was cut where the other gene started. Finally, mappings from Cluster Buster and FIMO were combined into an ensemble motif mapping gene regulatory network (GRN), following the protocol described by Kulkarni *et al.*[61]. Next, the ensemble GRN was queried for TFs that were directly connected to each of the triterpene BGCs, and were selected for further analysis.

Thalianol profiling

For thalianol measurements, approximately 100 mg of fresh root material was extracted with 1 mL of methanol. Metabolite extraction was performed for 1 h at room temperature, followed by centrifugation for 10 min at 20,8000g. 50 µL of the supernatant was collected and evaporated to dryness under vacuum. The remaining plant material was lyophilized for dry weight determination. The residue obtained from metabolite extraction was trimethylsilylated using 50 µL derivatization mixture (N-methyl-N-(trimethylsilyl)trifluoroacetamide:pyridine in a 5:1 ratio). The derivatized samples were shaken for 1 h at room temperature. GC-MS analyses were performed on an Agilent 7250 QTOF-MS equipped with an Agilent 7890B GC system. One µL of derivatized sample was injected in splitless mode with the injector port set to 280°C. All biological samples were analyzed at random and pooled samples were included for system reproducibility. Separation was achieved with a VF-5ms column (40 m × 0.25 mm, 0.25 µm; Varian CP9013; Agilent) with helium carrier gas at a constant flow rate of 1.2 mL/min. The oven was held at 80°C for 1 min, ramped to 320°C

at a rate of 10°C/min and then held at 320°C for 5 min. MS transfer line, MS ion source and quadrupole were held at 280°C, 230°C and 150°C, respectively. The MS detector was operated in EI mode at 70 eV. Full EI-MS spectra were recorded by scanning the m/z range of 50 to 800 with a solvent delay of 6 min. Identification of thalianol was conducted by comparison with the retention time and mass spectrum of a pure standard[7]. Peak areas were integrated using Masshunter Qualitative Analysis software (Agilent) and normalized against the dry weight of the samples.

Data availability

Upon acceptance, the scRNAseq data will be made accessible via an online browser tool (http://bioit3.irc.ugent.be/plant-sc-atlas/) and raw data can be accessed at NCBI with GEO numbers GSE179820 and GSE212826, for the mock- and JA-treated root tips, respectively. All other data are either in the main paper or the Supplementary Information. Material requests should be directed to the corresponding author.

References

1. Erb, M. & Kliebenstein, D. J. Plant secondary metabolites as defenses, regulators, and primary metabolites: the blurred functional trichotomy. *Plant Physiol.* **184,** 39-52 (2020).
2. Lacchini, E. & Goossens, A. Combinatorial control of plant specialized metabolism: mechanisms, functions, and consequences. *Annu. Rev. Cell Dev. Biol.* **36,** 291-313 (2020).
3. Colinas, M. & Goossens, A. Combinatorial transcriptional control of plant specialized metabolism. *Trends Plant Sci.* **23,** 324-336 (2018).
4. Wasternack, C. & Strnad, M. Jasmonates are signals in the biosynthesis of secondary metabolites — Pathways, transcription factors and applied aspects — A brief review. *New Biotechnol.* **48,** 1-11 (2019).
5. Tholl, D. Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng./Biotechnol.* **148,** 63-106 (2015).
6. Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P. & Osbourn, A. Triterpene biosynthesis in plants. *Annu. Rev. Plant Biol.* **65,** 225-257 (2014).
7. Huang, A. C. *et al.* A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science* **364,** eaau6389 (2019).
8. Pichersky, E. & Raguso, R. A. Why do plants produce so many terpenoid compounds? *New Phytol.* **220,** 692-702 (2018).
9. Sohrabi, R. *et al.* In planta variation of volatile biosynthesis: an alternative biosynthetic route to the formation of the pathogen-induced volatile homoterpene DMNT via triterpene degradation in Arabidopsis roots. *Plant Cell* **27,** 874-890 (2015).
10. Tholl, D. & Lee, S. Terpene specialized metabolism in *Arabidopsis thaliana*. *Arabidopsis Book* **9,** e0143 (2011).
11. Field, B. *et al.* Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci. USA* **108,** 16116-16121 (2011).
12. Field, B. & Osbourn, A. E. Metabolic diversification—independent assembly of operon-like gene clusters in different plants. *Science* **320,** 543-547 (2008).
13. Nützmann, H.-W. *et al.* Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc. Natl. Acad. Sci. USA* **117,** 13800-13809 (2020).
14. Nützmann, H.-W., Huang, A. & Osbourn, A. Plant metabolic clusters - from genetics to genomics. *New Phytol.* **211,** 771-789 (2016).

15. Bai, Y. *et al.* Modulation of *Arabidopsis* root growth by specialized triterpenes. *New Phytol.* **230,** 228-243 (2021).
16. Goossens, J., Fernández-Calvo, P., Schweizer, F. & Goossens, A. Jasmonates: signal transduction components and their roles in environmental stress responses. *Plant Mol. Biol.* **91,** 673-689 (2016).
17. Chini, A., Gimenez-Ibanez, S., Goossens, A. & Solano, R. Redundancy and specificity in jasmonate signalling. *Curr. Opin. Plant Biol.* **33,** 147-156 (2016).
18. Pauwels, L. *et al.* NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* **464,** 788-791 (2010).
19. Chini, A. *et al.* The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448,** 666-671 (2007).
20. Fernández-Calvo, P. *et al.* The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* **23,** 701-715 (2011).
21. Thines, B. *et al.* JAZ repressor proteins are targets of the SCF^{COI1} complex during jasmonate signalling. *Nature* **448,** 661-665 (2007).
22. Fonseca, S. *et al.* (+)-7-*iso*-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. *Nat. Chem. Biol.* **5,** 344-350 (2009).
23. Sheard, L. B. *et al.* Jasmonate perception by inositol-phosphate-potentiated COI1–JAZ co-receptor. *Nature* **468,** 400–405 (2010).
24. Yan, Y. *et al.* A downstream mediator in the growth repression limb of the jasmonate pathway. *Plant Cell* **19,** 2470-2483 (2007).
25. Gasperini, D. *et al.* Multilayered organization of jasmonate signalling in the regulation of root growth. *PLoS Genet.* **11,** e1005300 (2015).
26. Major, I. T. *et al.* Regulation of growth-defense balance by the JASMONATE ZIM-DOMAIN (JAZ)-MYC transcriptional module. *New Phytol.* **215,** 1533-1547 (2017).
27. Lorenzo, O., Chico, J. M., Sánchez-Serrano, J. J. & Solano, R. *JASMONATE-INSENSITIVE1* encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis. *Plant Cell* **16,** 1938-1950 (2004).
28. Schweizer, F. *et al. Arabidopsis* basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell* **25,** 3117-3132 (2013).
29. Birnbaum, K. *et al.* A gene expression map of the *Arabidopsis* root. *Science* **302,** 1956-1960 (2003).
30. Zander, M. *et al.* Integrated multi-omics framework of the plant response to jasmonic acid. *Nat. Plants* **6,** 290-302 (2020).
31. Mertens, J. *et al.* The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.* **170,** 194-210 (2016).
32. Van Moerkercke, A. *et al.* The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proc. Natl. Acad. Sci. USA* **112,** 8130-8135 (2015).
33. Van Moerkercke, A. *et al.* A MYC2/MYC3/MYC4-dependent transcription factor network regulates water spray-responsive gene expression and jasmonate levels. *Proc. Natl. Acad. Sci. USA* **116,** 23345-23356 (2019).
34. Goossens, J., Swinnen, G., Vanden Bossche, R., Pauwels, L. & Goossens, A. Change of a conserved amino acid in the MYC2 and MYC3 transcription factors leads to release of JAZ repression and increased activity. *New Phytol.* **206,** 1229-1237 (2015).

35. Kemen, A. C. *et al.* Investigation of triterpene synthesis and regulation in oats reveals a role for β-amyrin in determining root epidermal cell patterning. *Proc. Natl. Acad. Sci. USA* **111,** 8679-8684 (2014).
36. Ohashi, Y. *et al.* Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. *Science* **300,** 1427-1430 (2003).
37. Wang, S., Barron, C., Schiefelbein, J. & Chen, J. G. Distinct relationships between GLABRA2 and single-repeat R3 MYB transcription factors in the regulation of trichome and root hair patterning in Arabidopsis. *New Phytol.* **185,** 387-400 (2010).
38. Gualberti, G. *et al.* Mutations in the Dof zinc finger genes *DAG2* and *DAG1* influence with opposite effects the germination of Arabidopsis seeds. *Plant Cell* **14,** 1253-1263 (2002).
39. Ramachandran, V. *et al.* Plant-specific Dof transcription factors VASCULAR-RELATED DOF1 and VASCULAR-RELATED DOF2 regulate vascular cell differentiation and lignin biosynthesis in Arabidopsis. *Plant Mol. Biol.* **104,** 263-281 (2020).
40. Boccaccini, A. *et al.* The DOF protein DAG1 and the DELLA protein GAI cooperate in negatively regulating the *AtGA3ox1* gene. *Mol. Plant* **7,** 1486-1489 (2014).
41. Papi, M. *et al.* Identification and disruption of an *Arabidopsis* zinc finger gene controlling seed germination. *Genes Dev.* **14,** 28-33 (2000).
42. Gabriele, S. *et al.* The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene *AtGA3ox1*. *Plant J.* **61,** 312-323 (2010).
43. Zhao, H. *et al.* Identification and functional validation of super-enhancers in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **119,** e2215328119 (2022).
44. Baudry, A., Caboche, M. & Lepiniec, L. TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. *Plant J.* **46,** 768-779 (2006).
45. Ellis, C. & Turner, J. G. A conditionally fertile *coi1* allele indicates cross-talk between plant hormone signalling pathways in *Arabidopsis thaliana* seeds and young seedlings. *Planta* **215,** 549-556 (2002).
46. Xie, D.-X., Feys, B. F., James, S., Nieto-Rostro, M. & Turner, J. G. *COI1*: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280,** 1091-1094 (1998).
47. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17,** 148 (2016).
48. Decaestecker, W. *et al.* CRISPR-TSKO: a technique for efficient mutagenesis in specific cell types, tissues, or organs in Arabidopsis. *Plant Cell* **31,** 2868-2887 (2019).
49. Vanden Bossche, R., Demedts, B., Vanderhaeghen, R. & Goossens, A. Transient expression assays in tobacco protoplasts. *Methods Mol. Biol.* **1011,** 227-239 (2013).
50. Wendrich, J. R. *et al.* Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* **370,** eaay4970 (2020).
51. Yang, B. *et al.* Non-cell autonomous and spatiotemporal signalling from a tissue organizer orchestrates root vascular development. *Nat. Plants* **7,** 1485-1494 (2021).
52. Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17,** 75 (2016).
53. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Systems* **8,** 329-337 (2019).

54. Lambert, S. A. *et al.* Similarity regression predicts evolution of transcription factor sequence specificity. *Nat. Genet.* **51,** 981-989 (2019).
55. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48,** D87-D92 (2020).
56. Kulkarni, S. R., Vaneechoutte, D., Van de Velde, J. & Vandepoele, K. TF2Network: predicting transcription factor regulators and gene regulatory networks in Arabidopsis using publicly available binding site information. *Nucleic Acids Res.* **46,** e31 (2017).
57. Frith, M. C., Li, M. C. & Weng, Z. Cluster-Buster: finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res.* **31,** 3666-3668 (2003).
58. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27,** 1017-1018 (2011).
59. Cheng, C.-Y. *et al.* Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89,** 789-804 (2017).
60. Van Bel, M. *et al.* PLAZA 5.0: extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Res.* **50,** D1468-D1474 (2022).
61. Kulkarni, S. R., Jones, D. M. & Vandepoele, K. Enhanced maps of transcription factor binding sites improve regulatory networks learned from accessible chromatin data. *Plant Physiol.* **181,** 412-425 (2019).

Acknowledgments

This article is written in loving memory of Alex Van Moerkercke (1979-2021). The authors wish to thank Annick Bleys for critically reading the manuscript. We thank Deborah Gasperini for kindly sharing the *ProMYCs:NLS-VENUS* reporter lines, and Paola Vittorioso for the *dag1* mutant, *ProDAG1:GUS* and *DAG1* over-expressing lines. We wish to thank Jos R. Wendrich and Thomas Eekhout for their assistance in the launching and analysis of the scRNAseq experiment, and Sandrien Desmet and Geert Goeminne from the VIB Metabolomics Core - Ghent for the thalianol profiling. This work has been supported by the European Community's Seventh Framework Program (FP7/2007–2013) under grant agreement 613692-TriForC and H2020 Program under grant agreement 760331-Newcotiana to A.G., the Special Research Fund from Ghent University to A.G. and A.R. (project BOF18/GOA/013), M.M. (project BOF20/GOA/012), the Flemish Government (AI Research program) to Y.S., the Research Foundation Flanders with research project grants to AG (G004515N and G008417N) and a postdoctoral fellowship to P.F.-C., and the China Scholarship Council for a PhD scholarship to Y.B. A.O. would like to acknowledge funding support from the John Innes Foundation and the BBSRC Institute Strategic Program Grant 'Molecules from Nature – Products and Pathways' (BBS/E/J/000PR9790).

Author contributions

Conceptualization: A.G., H.N.T., A.V.M., P.F.C., A.R.; Methodology: H.N.T., L.T., M.M., T.D., K.V., G.V.I., Y.S., B.D.R., K.V., A.R., A.G.; Investigation: H.N.T., Y.B., A.V.M., P.F.C., L.T., T.D., G.V.I., A.R., A.G.; Visualization: L.T., A.R., K.V., T.D.; Funding acquisition: A.G, B.D.R., K.V.; Project administration: A.G., A.R.; Supervision: A.G., A.R., B.D.R., K.V.P., Y.S.; Writing – original draft: A.V.M., A.R., A.G.; Writing – review & editing: H.N.T., B.D.R.

Competing interests

The authors declare no competing interests.

Additional Information

Extended data is available for this paper at https://doi.org/

Supplementary information The online version contains supplementary material available at https://doi.org/

Correspondence and requests for materials should be addressed to Alain Goossens.

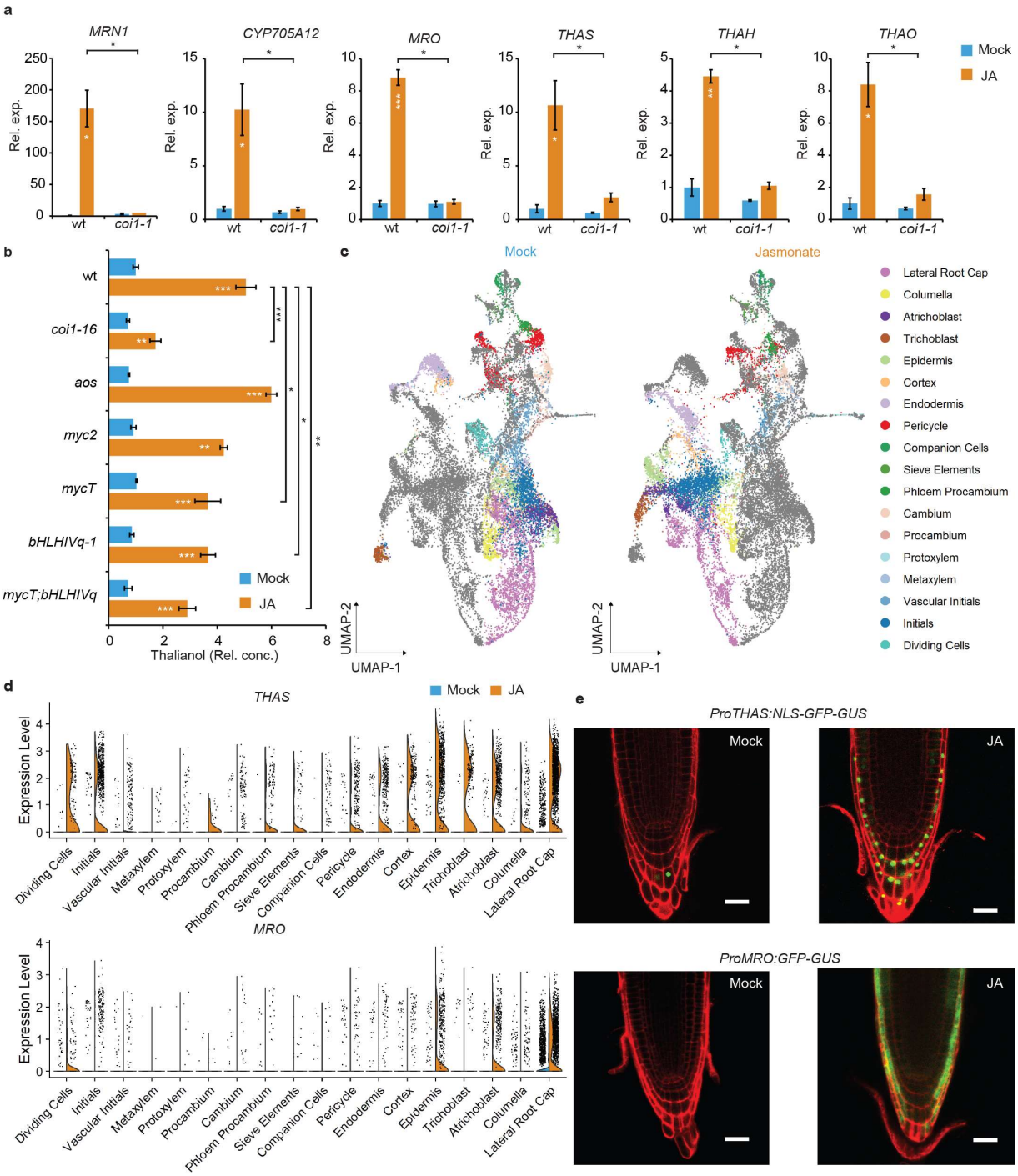

Fig. 1| Single-cell transcript patterns of genes in the triterpene BGC in root meristem cells upon jasmonate (JA) treatment. a, qPCR showing steady-state transcript levels of genes in the marneral and thalianol BGCs in roots of Col-0 (wt) and *coi1-1* seedlings after a 6-h mock or 50-µM JA treatment. Values on the Y-axis represent fold-induction compared to mock-treated wt (set to 1). The error bars designate SE of the mean (n = 3). Statistical significance between genotypes and treatments was determined using the Student's *t*-test (*P < 0.05, **P < 0.005, ***P < 0.0005). **b,** Thalianol metabolite levels in wt, *coi1-16, aos, myc2, mycT, bHLHIVq-1* and *mycT;bHLHIVq-1*

seedlings after a 24-h mock or 50-µM JA treatment. Values on the X-axis represent fold-induction compared to mock-treated wt (set to 1). The error bars designate the SE of the mean (n = 5). Statistical significance between genotypes and treatments was determined using the Student's *t*-test (*P < 0.05, **P < 0.005, ***P < 0.0005). **c,** Uniform Mainfold Approximation and Projection (UMAP) representation of the mock- and JA-treated scRNAseq datasets (2-h mock or 50-µM JA treatment). Cell identities are shown by clusters of different colors. **d,** Violin plots showing cell-specific expression of *THAS* and *MRO* upon 2-h mock and 50-µM JA treatment of wt root tips. **e,** Expression profiles of *ProTHAS:NLS-GFP-GUS* and *ProMRO:GFP-GUS* in wt root tips grown on mock or 50-µM JA for 24 h. Scale bars = 20 µm.

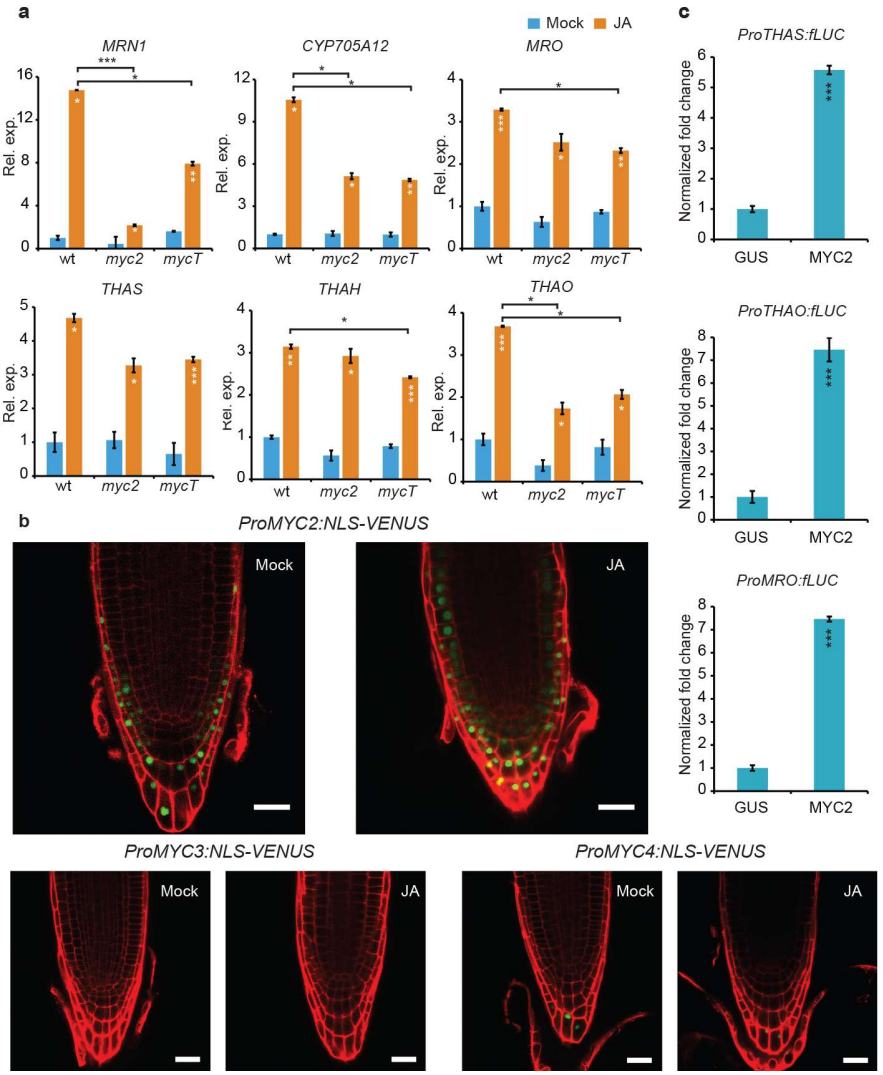

Fig. 2| Triterpene BGC expression depends partially on MYC2. a, qPCR showing steady-state transcript levels of genes in the marneral and thalianol BGCs in roots of Col-0 (wt), single *myc2* and triple *myc2 myc3 myc4* (*mycT*) seedlings after a 6-h mock or 50-µM JA treatment. Values on the Y-axis represent fold-induction compared to mock-treated wt (set to 1). The error bars designate the SE of the mean (n = 3). Statistical significance between genotypes and treatments was determined using the Student's *t*-test (*P < 0.05, **P < 0.005, ***P < 0.0005). **b,** Expression profiles of *ProMYC2:NLS-VENUS*, *ProMYC3:NLS-VENUS* and *ProMYC4:NLS-VENUS* in wt root tips grown on mock or 50-µM JA for 24 h. Scale bars = 20 µm. **c,** Transactivation in transfected *N. tabacum* protoplasts of the *ProTHAS*-, *ProTHAO*- and *ProMRO*-driven *fLUC* reporters by MYC2. Values on the Y-axis are normalized fold changes relative to protoplasts cotransfected with the reporter constructs and a *pCaMV35S:GUS* (GUS) control plasmid (set to 1). The error bars designate the SE of the mean (n = 8). Statistical significance was determined using the Student's *t*-test (***P < 0.0005).

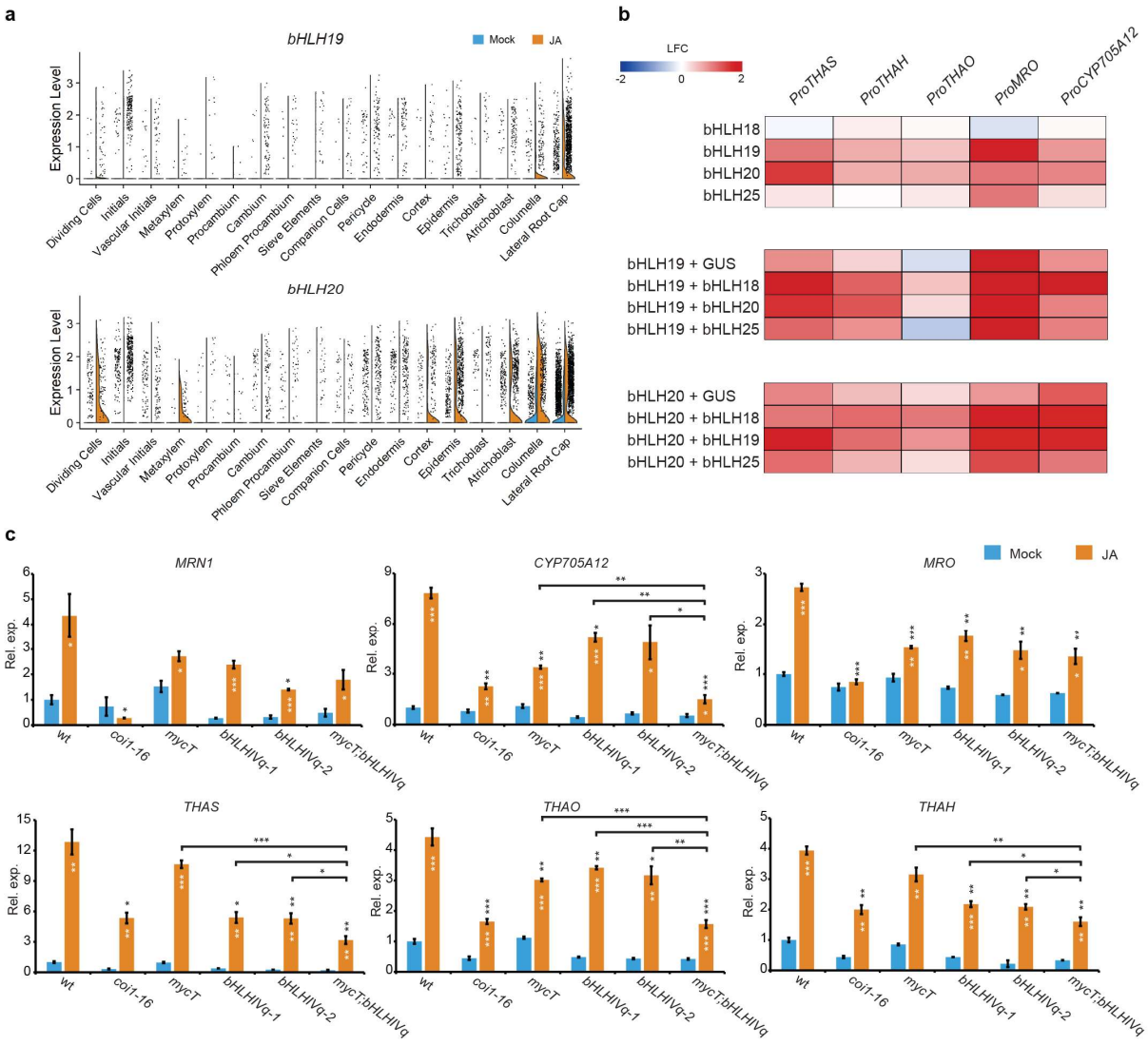

Fig. 3| Cell-specific activation of genes in the triterpene BGC is codetermined by MYC2 and bHLH clade IVa TFs. a, Violin plots showing cell-specific expression of *bHLH19* and *bHLH20* in 2-h mock- and JA-treated wt root tips. **b,** Heatmaps representing transactivation assays of triterpene BGC *fLUC* reporters (indicated on top) by individual clade IVa bHLH TFs, as well as possible combinations of bHLH19 or bHLH20 with the other clade IVa bHLH TFs. Values are represented as log fold change (LFC) relative to protoplasts cotransfected with the reporter constructs and a *pCaMV35S:GUS* control plasmid. **c,** qPCR showing transcript levels of the marneral and thalianol BGCs in root tips of Col-0 (wt), *coi1-16*, *mycT*, *bHLHIVq-1*, *bHLHIVq-2* and *mycT;bHLHIVq* seedlings after a 6-h mock or 50-µM JA treatment. Values on the Y-axis represent fold-induction compared to mock-treated wt (set to 1). The error bars designate the SE of the mean (n = 3). Statistical significance between genotypes and treatments was determined using the Student's *t*-test (*P < 0.05, **P < 0.005; ***P < 0.0005).

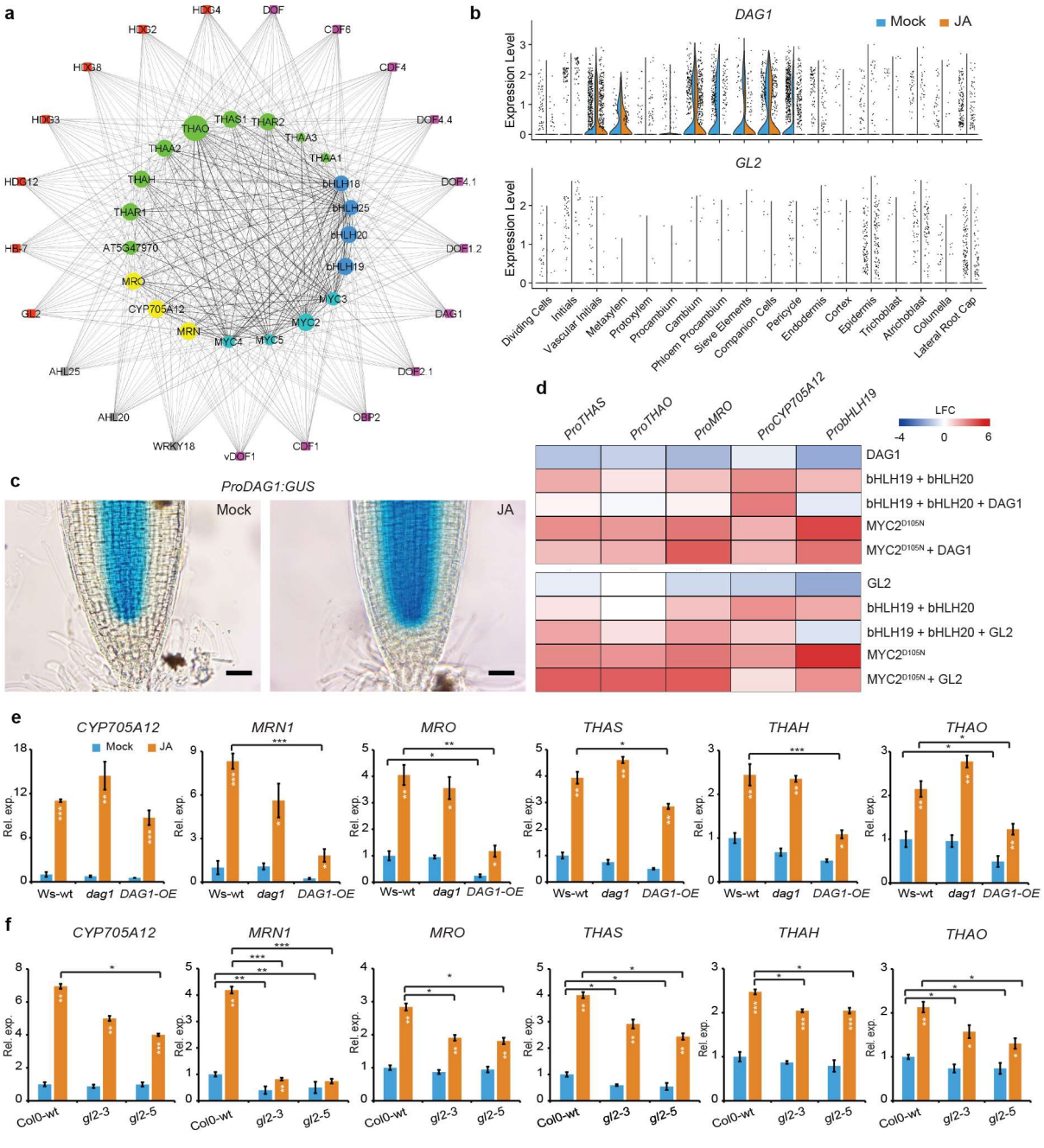

Fig. 4| Expression of triterpene genes in the BGCs is coactivated by GL2 and repressed in inner root tissues by the DOF factor DAG1. a, Motif mapping-based network of triterpene biosynthesis genes and their predicted TF regulators. **b,** Violin plot showing the tissue-specific expression profile of *DAG1* and *GL2* in mock- and JA-treated conditions. **c,** Histochemical analysis of the spatial expression of *ProDAG1:GUS* in mock- and JA-treated conditions. Scale bars = 20 µm. **d,** Heatmap representing transactivation assays of triterpene BGC promoters driving *fLUC* reporters (indicated on top) by individual and combinations of clade IVa bHLH TFs, MYC2^{D105N}, DAG1 and GL2. Values are represented as LFC relative to protoplasts cotransfected with the reporter constructs and a *pCaMV35S:GUS* control plasmid. **e,** qPCR showing transcript accumulation levels of the indicated genes in the marneral and thalianol BGCs in roots of Ws-4

wt, *dag1* and *35S:DAG1-HA;dag1* (*DAG1-OE*) seedlings after a 2-h mock or JA treatment. **f,** qPCR showing transcript accumulation levels of the indicated genes in the marneral and thalianol BGCs in roots of Col-0 wt, *gl2-5* and *gl2-3* seedlings after a 2-h mock or JA treatment. Values on the Y-axis represent fold-induction compared to mock-treated wt (set to 1). The error bars designate the SE of the mean (n = 3). Statistical significance between genotypes and treatments was determined using the Student's *t*-test (*P < 0.05, **P < 0.005, ***P < 0.0005).