A modular system regulates specialized metabolite pathway branch choice in the medicinal plant *Catharanthus roseus*

Maite Colinas[1,2*], Jacob Pollier[1,3], Dries Vaneechoutte[1,2], Deniz G. Malat[1,2], Fabian Schweizer[1,2], Rebecca De Clercq[1,2], Joana G. Guedes[4,5,6], Teresa Martínez-Cortés[4□], Francisco J. Molina Hidalgo[1,2], Mariana Sottomayor[4,7], Klaas Vandepoele[1,2] and Alain Goossens[1,2]

[1] Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, B-9052 Gent, Belgium; [2] VIB Center for Plant Systems Biology, Technologiepark 71, B-9052 Gent, Belgium; [3] VIB Metabolomics Core, 9052 Ghent, Belgium; [4] CIBIO/InBIO-Centro de Investigac□ão em Biodiversidade e Recursos Genéticos, Universidade do Porto, 4485-661 Vairão, Portugal; [5] I3S-Instituto de Investigação e Inovação em Saúde, IBMC-Instituto de Biologia Molecular e Celular, Universidade do Porto, 4200-135 Porto, Portugal; [6] ICBAS – Instituto de Cie□ncias Biomédicas Abel Salazar, Universidade do Porto, 4050-313 Porto, Portugal; [7] Faculdade de Ciências da Universidade do Porto, 4169-007 Porto, Portugal; [*] Present address: Institute of Molecular Plant Biology, ETH Zurich, CH-8092 Zurich, Switzerland; □Present address: Universidade da Coruña, Facultad de Ciencias - Departamento de Biologia Animal, Biologia Vegetal y Ecologia, Centro de Investigaciones Científicas Avanzadas (CICA), A Zapateira s/n, 15071 A Coruña, Spain.

Authors for correspondence:

Maite Colinas

Email: maite.colinas@biol.ethz.ch

Alain Goossens

Tel: +32 9 331 38 51

Email: algoo@psb.vib-ugent.be

Total word count:	**6,349**
– Introduction	1,295
– Materials and Methods	583
– Results	2,910
– Discussion	1,561

Number of figures: 7 (all in color)

Number of tables: 0

Number of supporting information: 2 Methods, 10 Figures, 10 Tables, 2 Datasets

Summary

- *Catharanthus roseus* produces a diverse range of specialized metabolites of the monoterpenoid indole alkaloid (MIA) class in a heavily branched pathway. Recent great progress in identification of MIA biosynthesis genes revealed that the different pathway branch genes are expressed in a highly cell type- and organ-specific and stress-dependent manner. This implies a complex control by specific transcription factors (TFs), only partly revealed today.
- We generated and mined a comprehensive compendium of publicly available *C. roseus* transcriptome data for MIA pathway branch-specific TFs. Functional analysis was performed through extensive comparative gene expression analysis and profiling of over 40 MIA metabolites in the *C. roseus* flower petal expression system.
- We identified a MYB TF and additional members of the known BIS and ORCA regulators. Further detailed study of the ORCA TFs suggests subfunctionalization of ORCA paralogs in terms of branch-specific regulation and synergistic activity with the central jasmonate response regulator MYC2. Moreover, we identified specific amino acid residues within the ORCA DNA-binding domains that contribute to this branch-specific regulation.
- Our results advance our understanding of TF paralog specificity for which, despite the common occurrence of closely related paralogs in many species, comparative studies are scarce.

Key words: AP2/ERF transcription factor, bHLH transcription factor, *Catharanthus roseus*, Madagascar periwinkle, MYB transcription factor, Monoterpenoid indole alkaloids, transcription factor gene clusters

Introduction

Plant specialized metabolites are chemically diverse, typically species- or taxa-specific and have often evolved in adaptation to the ecological niche of the plant. Within the plant, their production often exclusively takes place in specific organs or cell types and may be up-regulated precisely in response to defined environmental conditions [for a review, see Colinas and Goossens (2018)]. The integration of internal cues, such as cell type environment and developmental stage, and external cues, such as the presence of pathogens or herbivores, is predominantly regulated at the transcriptional level by specific transcription factors (TFs). Different TFs might act independently or cooperatively, forming regulatory modules that ensure the optimal spatiotemporal expression of specific metabolic pathway genes.

The medicinal plant *Catharanthus roseus* is well-known as being the only source of the important anti-cancer compounds vinblastine and vincristine (van der Heijden *et al.*, 2004). These complex molecules belong to the class of monoterpenoid indole alkaloids (MIAs); around 150 different MIAs are estimated to occur in *C. roseus* (van der Heijden *et al.*, 2004). The identification of many of the involved biosynthetic enzymes has revealed that genes involved in the different MIA pathway steps and branches are expressed in a cell type-specific, organ-specific and stress-dependent manner, making this species an ideal model to study how a modular system of different TFs possibly regulates specialized metabolism (Courdavault *et al.*, 2014; Dugé de Bernonville *et al.*, 2015).

Different MIA pathway steps and branches can be organized into *subdivisions* with connected gene expression patterns (Fig. 1). The first of such *subdivisions* is the *iridoid pathway*, i.e. the biosynthesis of the precursor loganic acid from geranyl diphosphate (GPP) via seven enzymatic steps occurring in the specialized internal phloem-associated parenchyma (IPAP) cells (Geu-Flores *et al.*, 2012; Asada *et al.*, 2013; Simkin *et al.*, 2013; Miettinen *et al.*, 2014; Salim *et al.*, 2014). An iridoid intermediate, presumably loganic acid, is then transported to the epidermis, possibly by the recently characterized nitrate/peptide family (NPF) transporters (NPF2.4, NPF2.5, and NPF2.6) (Yamamoto *et al.*, 2016; Larsen *et al.*, 2017; Yamamoto *et al.*, 2019). The coupling of the then converted iridoid secologanin to the tryptophan-derived indole moiety tryptamine marks the formation of the first MIA strictosidine, from which, after deglycosylation, all MIAs derive (*strictosidine pathway*). A single

enzymatic step is needed to convert the strictosidine aglycone either into the heterohymbines ajmalicine or tetrahydroalstonine (and subsequently alstonine) (Stavrinides *et al.*, 2016; Dang *et al.*, 2018), or into vitrosamine (Stavrinides *et al.*, 2018). The branch leading to the precursors of the anti-cancer compounds has recently been resolved and involves seven enzymatic steps to yield the unstable intermediate dehydrosecodine (named *stemmadenine pathway* according to the name of one of the intermediates)(Tatsis *et al.*, 2017; Caputi *et al.*, 2018; Qu *et al.*, 2018; Qu *et al.*, 2019). From the latter either tabersonine or catharanthine are formed by two different enzymes (Caputi *et al.*, 2018; Qu *et al.*, 2018). Catharanthine forms one moiety of the MIA dimers vinblastine and vincristine and has been additionally suggested to have a repellent function as a monomer after secretion to the leaf surface by an epidermal ATP-binding cassette (ABC) transporter (Roepke *et al.*, 2010; Yu & De Luca, 2013). The second monomer, vindoline, is produced from tabersonine via seven enzymatic steps, from which the two last steps are thought to take place in specialized idioblasts or laticifers (St-Pierre & De Luca, 1995; St-Pierre *et al.*, 1999; Levac *et al.*, 2008; Liscombe *et al.*, 2010; Guirimand *et al.*, 2011; Liscombe & O'Connor, 2011; Besseau *et al.*, 2013; Qu *et al.*, 2015). Noteworthy, the *vindoline pathway* exclusively occurs in the green aerial parts of the plant (Liu *et al.*, 2019). By contrast, in roots, a different set of interconvertible *root-specific MIAs* (lochnericine, hörhammericine, minovincinine among others) is formed (Laflamme *et al.*, 2001; Giddings *et al.*, 2011; Carqueijeiro, I *et al.*, 2018; Carqueijeiro, IT *et al.*, 2018; Williams *et al.*, 2019). Finally, the assumed spatially separated monomers vindoline and catharanthine are dimerized to α-3',4'-anhydrovinblastine, the direct precursor of vinblastine and vincristine, presumably by a peroxidase (Costa *et al.*, 2008). Noteworthy, a very recent study suggests that catharanthine, as well as ajmalicine and its oxidation product serpentine, may also accumulate in idioblasts and laticifers in addition to the epidermis (Yamamoto *et al.*, 2019). Thus, more research is needed in order to clarify the hypothesis of spatial separation of the monomers and differences in the sites of biosynthesis and accumulation of MIA monomers.

There has been a substantial increase in knowledge about which specific TFs from different families control the different MIA pathway steps and branches, although not complete yet. The main focus has been on the jasmonate (JA)-dependent transcriptional activation, because many (but not all, see below) MIA pathway steps

are up-regulated upon exposure to this phytohormone that mainly mediates defense against necrotrophic pathogens and herbivores (Aerts *et al.*, 1994; Zhang *et al.*, 2011; De Geyter *et al.*, 2012; Goossens *et al.*, 2016; Zhou & Memelink, 2016; Goossens *et al.*, 2017). The primary response regulator of JA-dependent up-regulation of target genes is the basic helix-loop-helix (bHLH) TF MYC2, whose activity is post-translationally suppressed in the absence of JA (Chini *et al.*, 2007; Goossens *et al.*, 2017; Wasternack & Strnad, 2019). Recently, we have described how overexpression of the de-repressed *CrMYC2a*^{D126N} variant can boost the expression of the JA-responsive MIA biosynthesis genes (Schweizer *et al.*, 2018). This up-regulation may partly occur directly and/or via the up-regulation of branch-specific TFs described below (Schweizer *et al.*, 2018). For instance the regulators bHLH iridoid synthesis (BIS) 1 and BIS2 TFs exclusively up-regulate the *iridoid pathway* genes (Van Moerkercke *et al.*, 2015; Van Moerkercke *et al.*, 2016). The expression of *strictosidine pathway* genes is controlled by members of the octadecanoid derivative-responsive *catharanthus* apetala2-domain (ORCA) clade (Menke *et al.*, 1999; van der Fits & Memelink, 2000; van der Fits & Memelink, 2001; Paul *et al.*, 2016; Paul *et al.*, 2020; Singh *et al.*, 2020). In addition, overexpression of *ORCA3* has been shown to increase the expression of the *iridoid pathway* genes to some extent, part of the *stemmadenine pathway* genes and, when overexpressed in combination with *CrMYC2a*^{D126N}*, root specific MIA* genes in a highly synergistic manner (Schweizer *et al.*, 2018). While there is some knowledge about the target gene profiles of the other ORCAs (Paul *et al.*, 2016; Paul *et al.*, 2020; Singh *et al.*, 2020), a direct comparison of them under the same experimental condition is currently lacking. Moreover, it has not been clarified, which, if any, of the ORCAs, in addition to ORCA3, act synergistically with CrMYC2a^{D126N}. Noteworthy, the *vindoline pathway* does not appear to be up-regulated by JA and seems to be rather repressed upon overexpression of *CrMYC2a*^{D126N} (Schweizer *et al.*, 2018). Very recently it has been shown that, the *vindoline pathway* is at least in part up-regulated by a GATA TF and down-regulated by phytochrome-interacting factor TFs in a light-dependent manner (Liu *et al.*, 2019). However, the moderate changes in expression of the vindoline pathway genes and metabolite levels induced by these TFs suggest the presence of additional regulators. Other TFs, with a more limited impact on MIA biosynthesis have been described [for a review, see Liu *et al.* (2017)].

Together, the increasing identification of MIA biosynthesis genes over the recent years revealed a network of pathway branches (*subdivisions*) with differential gene expression profiles. This suggests the existence of specific regulatory factors acting independently or cooperatively in a branch-specific manner. Here, we sought out for regulators for pathway branches for which regulation was unclear or unknown by using co-expression analysis of publicly available and hitherto unpublished in-house generated RNA-Seq datasets. We identified additional members of well-known BIS, ORCA and MYB clades and revealed that subfunctionalization of the ORCA TFs leads to the control of specific pathway branches and to differential synergistic up-regulation with CrMYC2a^{D126N}.

Material and Methods

Generation of a RNAseq data compendium and co-expression network analysis to infer candidate regulators

The RNA-Seq compendium (Table S1) and co-expression network linking each gene to its top 500 most co-expressing genes, were generated with Prose from the Curse toolset (Vaneechoutte & Vandepoele 2019), while putative regulators were annotated using TRAPID (Van Bel et al., 2013). More details about RNA-Seq processing and mining of the RNA-Seq data compendium to identify candidate regulators are described in the Supporting Information.

Generation of DNA constructs

Coding sequences (CDSs) of TFs were amplified using cDNA from *C. roseus* var. “Little bright eyes” or *C. roseus* var. “Sunstorm apricot” as templates. Amplification was performed using Q5® High-Fidelity DNA Polymerase (New England BioLabs®) or iProof™ (Bio-Rad) and gene-specific oligonucleotides containing attB Gateway® recombination sites (Table S2). Fragments were gel-purified (GeneJET Gel Extraction Kit, ThermoFisher) and recombined into pDONR™207 using Gateway® BP clonase™ II enzyme mix (ThermoFisher), and the resulting clones sequence verified. ENTRY clones containing *CrMYC2a^{D126N}*, *BIS1*, *BIS2* were cloned previously (Van Moerkercke *et al.*, 2015; Van Moerkercke *et al.*, 2016; Schweizer *et*

al., 2018). ORCA3 protein mutants were constructed by site-directed mutagenesis using iProof™ (Bio-Rad) and pDONR207_CrORCA3 as a template and complementary primers. Chimeric CrORCA3-4 and CrORCA4-3 were constructed by overlap extension PCR using chimeric primers. ENTRY clones containing TFs were recombined using LR Clonase™ II enzyme mix (ThermoFisher) into pH7WG2D or pK7WG2D for overexpression in flower petals, or p2GW7 for promoter transactivation assays in tobacco protoplasts (Karimi *et al.*, 2002).

The promoter fragment of *STR* had been cloned previously (Vom Endt *et al.*, 2007). The promoter fragments of *GO*, *MAT*, *TAT*, *T3O*, *T3R*, *D4H* and *DAT* were amplified from *C. roseus* var. "Sunstorm apricot" genomic DNA using specific oligonucleotides (Table S3), recombined into pDONR™207 and sequence verified, as described for the TF CDSs. The resulting ENTRY clones containing promoter fragments were recombined into pGWL7 (Karimi *et al.*, 2002) for transient expression assays in tobacco protoplasts.

Transient expression assays in *Nicotiana tabacum* protoplasts

Promoter transactivation of *C. roseus* gene promoters by *C. roseus* TFs was measured via transient expression assays in *N. tabacum* 'Bright Yellow-2' protoplasts exactly as described previously (Vanden Bossche *et al.*, 2013; Schweizer *et al.*, 2018). Statistical analysis (ANOVA with Tukey's test for multiple comparisons) was performed using PRISM GraphPad 8 software.

Plant material and flower petal transformation

Catharanthus roseus var. "Little bright eyes" plants were used for flower petal infiltration with *A. tumefaciens* C58C1 harboring overexpression constructs as previously described (Schweizer *et al.*, 2018) with minor adaptations detailed in the supporting information.

Gene expression analysis by qPCR

RNA extraction and reverse transcription were performed as described (Schweizer *et al.*, 2018) with modifications detailed in the supporting information.

Metabolite profiling

For metabolite profiling, around 100 mg of infiltrated flower petal samples were ground in liquid nitrogen, lyophilized and extracted with 1 ml methanol containing 100 µM caffeine as an internal standard. Extraction took place exactly as described previously (Schweizer *et al.*, 2018). Detection was performed on an LC-MS and identifications of MIAs were done using FT-MS and MS^n data. More details about the experimental procedure can be found in the supporting information.

Gene IDs and data deposition

Gene IDs (*C. roseus* genome version 2) are listed with the corresponding primer sequences in Table S4. In some cases, the cloned sequences differed from publicly available genome and transcriptome data. All sequences cloned for this study have been deposited in the GenBank database; accession numbers are also listed in Table S4.

Results

TF candidate selection and screening impact on MIA biosynthesis

To enable comprehensive mining of *C. roseus* transcriptomes for novel MIA regulators, first a total of 82 RNA-Seq datasets were mapped to the latest *C. roseus* draft genome version (Kellner *et al.*, 2015) (http://medicinalplantgenomics.msu.edu/index.shtml). In addition to publicly available datasets, RNA-Seq data from hitherto unpublished in-house experiments were included, including data of cell type-enriched transcriptomes (see SI methods and Table S1). Across these datasets, branches of MIA pathway genes (Fig. 1b) showed a high level of co-expression.. Noteworthy, in particular the *vindoline* and catharanthine branches were not co-expressed with other MIA pathway branches (Fig. 1b). A k-nearest neighbor (KNN) network was created and for each MIA biosynthesis gene the top 500 most co-expressing genes were searched for genes with the GO term ‘regulation of gene expression’ (GO:0010468) (see method section for details)

yielding a list of 111 candidates (Fig. 2a). Manual sequence analysis of the candidates revealed that 15 were likely not unambiguously corresponding to genuine TFs and thus not considered for further analysis. Consistent with previous results, among the candidates we found the previously characterized TFs BIS1 and 2 and ORCA2, 4, 5 and 6, but not ORCA3. We further selected those TFs that showed co-expression with (i) multiple pathway genes and (ii) MIA pathway genes for which till now no transcriptional regulators are known, in particular genes belonging to the *stemmadenine* or *vindoline* branches or involved in catharanthine biosynthesis. Out of the 40 selected candidates, 28 could successfully be cloned and subjected to agroinfiltration in *C. roseus* flower petals. Four additional candidates that were not found through this co-expression analysis but through a preceding, preliminary, analysis were also tested in this screening (gene IDs in italic font in Fig. 2b). The latter preceding co-expression analysis was performed exactly as described above but on 72 RNA-Seq samples because one dataset (Pan *et al.*, 2018) had not yet been available at the start of this project (data not shown). While the candidate lists of the two analyses greatly overlapped those four candidates were exclusively found in the first analysis. Thus altogether 32 candidates were tested by bulk overexpression in *C. roseus* flower petals as described hereafter.

To enhance the throughput, up to four TF candidates were co-overexpressed under the control of the *CaMV35S* promoter (p35S). We specifically combined those candidates that were co-expressed with the same (or overlapping) subsets of MIA pathway genes. This resulted in nine overexpression groups of candidates (Fig. 2b). *A. tumefaciens* harboring a plasmid for *p35S::GUS* expression was infiltrated as a control. The overexpression level of the candidate TFs and the expression level of the MIA pathway genes were then measured by qPCR (Fig. S1). In case of successful overexpression (>5-fold) and observed increase(s) (>2-fold, P-value according to Student's *t*-test <0.05) in expression level(s) of potential target MIA pathway genes compared to the *p35S::GUS* control, the TF candidates from the respective group were overexpressed separately (Fig. 2b; Fig. S2). In the case of CRO_T132285 and CRO_T127367, for which the increase in target gene expression level was at the threshold level, additional promoter transactivation assays were carried out that did not confirm a (direct) up-regulation of MIA target genes (Fig. S2f). Altogether the results from the screening suggested that only the ORCA TFs (CRO_T110359, CRO_T110361, CRO_T110364) and clade IVa bHLH TFs (CRO_T123350,

CRO_T105905, CRO_T107533) reproducibly led to an increase in MIA biosynthesis gene expression levels (Fig. S1, S2 and below).

The *iridoid* regulatory module: Identification of an additional *BIS* paralog that clusters with *BIS1* and *2*

Overexpression of the three clade IVa bHLHs (CRO_T123350, CRO_T105905, CRO_T107533) led to enhanced expression of the iridoid genes, albeit to very different levels (Fig. 2b and S2a,c). Only overexpression of CRO_T107533 (referred to as *BIS3* hereafter) led to increases in *iridoid pathway* transcript levels (28- to 480-fold) similar to overexpression of *BIS1* and *2* in our previous study (Van Moerkercke *et al.*, 2016). Noteworthy, all three *BIS* TFs cluster within a 120-kb locus in the genome and are probably the result of recent tandem gene duplication events (Fig. 3a). This was previously not known but could now be detected because of the recent update (v2) of the *C. roseus* genome (http://medicinalplantgenomics.msu.edu). A ClustalW alignment revealed that the conservation of amino acid residues of the three BIS was high (54.2% pairwise identity), possibly explaining the redundant function in the up-regulation of the *iridoid* genes (Fig. 2c).

The ORCA cluster: Functional diversification into branch-specific regulatory modules

In addition to the long-known *ORCA2* and *3*, *ORCA4*, *5* and *6* have recently been found to be members of a gene cluster together with *ORCA3*, and their target gene profile has partly been elucidated (Paul *et al.*, 2016; Paul *et al.*, 2020). All ORCA TFs together form a cluster spreading over 70 kb, which can be subdivided into an ORCA3, 4 and 5 subcluster (around 25 kb) and an ORCA2 and 6 subcluster (around 10 kb) (Fig. 4a). A combination of different computational sequence analyses led us to hypothesize that the five ORCA TFs are possibly not functionally redundant. First, key amino acid residues (AARs) within the DNA-binding domain (DBD) differ between the ORCAs and are predicted to confer direct interaction with DNA differ (Fig. 5a). Second, it is generally accepted that the activation domains (ADs) of TFs are mostly poorly conserved at sequence level, but it is rather the structural properties defining the level of activation (Staby *et al.*, 2017). The combined information of

different prediction programs suggested that the ORCA TFs differ significantly in their structural organization, further hinting towards a certain degree of diversification (Fig. S3). Third, ORCA3-5 contain a serine-rich C-terminal domain that ORCA2 and 6 lack. Fourth, the five ORCA TFs showed a differential expression profile (Fig. 4b) and our co-expression analysis revealed that they co-express with different subsets of MIA pathway genes (Fig. 2a). Despite the occurrence of clusters of ORCA orthologs in other species, such as tomato and tobacco, it has never been systematically analyzed in any of these species if and to which level they are functionally redundant (Cárdenas *et al.*, 2016; Kajikawa *et al.*, 2017). Therefore, we sought out to decipher their specific regulatory roles by comparing their target gene and metabolite profile upon overexpression in flower petals. A PCA analysis of the MIA gene expression levels revealed that ORCA2-4 showed distinct, albeit partly overlapping target gene profiles (Fig. 4c). Particularly the expression of early and intermediate MIA pathway genes increased to similar levels upon overexpression of ORCA2-4 (Fig. 4d, for instance *SLS1*, *STR1*, *GS1*, *GS2*, *REDOX1*,*2*). However, in other cases, the fold changes differed considerably. For instance, *GO*, which encodes an enzyme producing the intermediate geissoschizine but also akuammicine, whose pathway branches of the *stemmadenine pathway* (see Fig. 1a), was considerably more highly up-regulated upon overexpression of *ORCA4*, and to a lesser extent of *ORCA2* compared to overexpression of the other *ORCAs* (Fig. 4d). Conversely, the *root MIA* genes *T19H* and *MAT* were more highly up-regulated when *ORCA3* or *ORCA5* were overexpressed. It should be noted that the overexpression level of the different *ORCA* paralogs differed considerably, with *ORCA3*, *4* and *6* being highly overexpressed (around 510-, 440-, and 960-fold, respectively) and with *ORCA2* and in particular *ORCA5* somewhat less (93-, and 15-fold, respectively) overexpressed (Fig. 4d). Thus, the overall lower MIA biosynthesis gene up-regulation levels observed for *ORCA5* overexpression might be partly due to the lower overexpression level of *ORCA5*. We did not observe any “intra-cluster” up-regulation between the different ORCA TFs. We next performed metabolite profiling on samples from this experiment (Figs 4g-f, S5, Table S6). A PCA analysis of the Mass spectrometry (MS) data revealed a (partially) differential profile (Fig. 4e), which was caused by differential accumulation of MIA metabolites (Fig. 4f). In addition to MIAs identified in a previous study (Schweizer *et al.*, 2018), the loadings plot further showed differentially accumulating compounds that we identified as MIAs (Fig. 4f (red),

Supporting Dataset S2). In particular, akuammicine and 16-hydroxytabersonine accumulation increased most pronouncedly when *ORCA4* was overexpressed and that of the root MIA derivative 16-methoxyhörhammericine when *ORCA3* was overexpressed (Fig. 4g). Overexpression of *ORCA6* had only mild effects on MIA pathway gene expression and no noticeable effects on MIA abundance. To confirm that the target genes were directly induced by the different ORCAs, we additionally performed promoter transactivation assays in tobacco protoplasts with promoter fragments of *STR*, *GO*, *MAT*, and *TAT* (Fig. S4). Similar as overexpression in flower petals, ORCA4 transactivated *pGO* to a much higher extent than the other ORCA TFs, whereas ORCA3, 4 and 5 similarly transactivated *pSTR*, *pMAT* and *pTAT*. Together, these data support both redundant and specific functions of the different ORCA TFs in the regulation of MIA biosynthesis.

Specific AARs within ORCA DBDs contribute to pathway branch specificity

Generally, the DBD is thought to define the specificity to the DNA motif, whereas the AD defines the level of up-regulation by recruiting the transcriptional machinery. In a previous study, the affinities of different orthologs of group IXa AP2/ERFs, i.e. ORCA3 and its orthologs from tobacco and *Arabidopsis thaliana*, to the DNA motif were compared (Shoji *et al.*, 2013). The results suggested that the specific binding affinity to different DNA motifs is indeed mediated by specific AARs in the DBD. However, specific AARs mediating target gene specificity among paralogs from the same species have never been characterized in detail to our knowledge. A protein sequence alignment of the DBDs of the ORCAs shows the expected overall high level of sequence conservation (Fig. 5a). Likewise, structural models of the ORCA DBDs are almost identical and highly align with the solved structure (PDB 1gcc) of the Arabidopsis ortholog AtERF1 (Allen *et al.*, 1998) (Fig. 5b). While the paralog-specific AARs corresponding to K101, P107, K110 and D117 in ORCA3 are not predicted to directly interact with DNA in the shown model, they are nevertheless predicted to be part of the beta-sheets in close proximity to the DNA (Fig. 4b). To compare a potential impact of these AARs on MIA target gene specificity, we performed site-directed mutagenesis on ORCA3 to obtain single (K101R, P107L, K110R, K110T and D117N) and double (K101R/P107L, K101R/K110R_K101R/K110T) mutants, whose sequence would correspond to that of

the other paralogs. The mutated versions were then overexpressed alongside native *ORCA3* in flower petals in two independent infiltration series (Fig. Fig. 5c-f and Fig. S6, respectively). Overall, our results pointed to two key amino acid residues leading to significant MIA target gene expression. Compared to overexpression of native *ORCA3*, overexpression of *ORCA3*^{K101R} (corresponding to an AAR change occurring in ORCA2, 4, 5 and 6) generally enhanced the expression of MIA pathway genes (Figs 5c, S5). By contrast, overexpression of *ORCA3*^{P107L} (corresponding to one of the AAR changes in ORCA4) led to reduced or abolished levels of MIA pathway genes that were partly recovered in the double mutant *ORCA3*^{K101RP107L} (corresponding to AARs in ORCA4) (Fig. 5c). Noteworthy, overexpression of *ORCA3*^{K101RP107L} still led to a lower up-regulation of *root MIA* genes than *ORCA3* or *ORCA3*^{K101R}, suggesting that the combination of these two AARs largely contributes to the differential target gene profile of ORCA3 and 4. The impact of the other AARs appeared to be less specific. Overexpression of the *ORCA5*-like *ORCA3*^{K101RK110R} led to an expression profile similar to that upon overexpression of *ORCA3*^{K101R}. Also, the much lower level of MIA target gene up-regulation by ORCA6 compared to the other ORCAs was not solely explained by the selected AARs (compare the overexpression of *ORCA3*^{K101RK110T} (Fig. 5c) and *ORCA3*^{D117N} (Fig. S6)). However, we did not produce a triple mutant in which all three AARs would correspond to those of ORCA6. Noteworthy, the observed differential target gene profiles upon overexpression of *ORCA3* mutants were also reflected at the metabolite level (Figs 5d-e, S7; Table S7). Noteworthy, in particular the levels of 16-hydroxytabersonine, akuammicine, 16-methoxyhörhammericine and 16-hydroxylochnericine (among others) upon overexpression of *ORCA3* versus the ORCA2-like *ORCA3*^{K101R} and the ORCA4-like *ORCA3*^{K101RP107L} were very similar to those upon overexpression of the actual *ORCA2*, *3* and *4* (Fig. 4g and S5).

Synergistic up-regulation in combination with de-repressed CrMYC2a^{D126N} is both ORCA- and target gene-dependent

Previously, we have shown that combinatorial overexpression of *ORCA3* and the de-repressed *CrMYC2a*^{D126N} variant leads to a synergistic up-regulation in particular of the *root MIA genes* compared to independent overexpression (Schweizer *et al.*, 2018). Based on this and our experiments above that suggest that the ORCA TFs to some

extent up-regulate different subsets of MIA pathway genes, we next investigated whether the combined overexpression of *CrMYC2a*^{D126N} with *ORCA2, 4, 5* or *6* leads to a synergistic up-regulation of target genes similar as with *ORCA3* and if different subsets of MIA pathway genes would be up-regulated (Fig. 6a). Indeed, a PCA depicted that in particular the combined overexpression of *CrMYC2a*^{D126N} with *ORCA2*, *3* or *4* led to differences in the MIA pathway gene expression profiles (Fig. 6b). Accordingly, PCA analysis of the MS data (Fig. 6c) revealed differential metabolite profiles for the combination of *CrMYC2a*^{D126N} with these three *ORCA* genes rather than with *ORCA5* or *6*, or upon overexpression of *CrMYC2a*^{D126N} alone; these changes were due to differential accumulation of MIAs (Fig. 6d). With regard to synergistic effects, the combined overexpression of *CrMYC2a*^{D126N} with *ORCA2* or *4*, synergistically up-regulated the *root MIA* genes, but to a lower extent than with *ORCA3* (Fig. 6a). Accordingly, at the metabolite level, the *root MIAs* 19-hydroxytabersonine and 16-hydroxy-19-O-acetyltabersonine accumulated to much higher levels upon *CrMYC2a*^{D126N}*/ORCA3* overexpression than upon overexpression of any of the other combinations (Fig. 6e). Notably, no synergistic, but rather an additive, up-regulation was observed for *GO*, although it was up-regulated upon overexpression of *CrMYC2a*^{D126N} and *ORCA4* alone (Figs 4d, 6a). Likewise, the level of the respective metabolic product akuammicine was only less than 2-fold higher in the samples with combined overexpression of *CrMYC2a*^{D126N} with *ORCA4*, compared to when *CrMYC2a*^{D126N} was overexpressed alone (Fig. 6e). These two observations suggest that the ORCA-CrMYC2a^{D126N} synergistic module is partly TF-dependent, i.e. more pronounced for ORCA3 and, moreover, target gene-dependent, i.e. specific to some of the *root MIA* genes.

To gain more insight into how a synergistic up-regulation could be mediated at the molecular level and to possibly shift the subset of target genes of this synergy, we next made chimeric constructs. Here, we focused on *ORCA3* and *4*, because their level of overexpression was most similar and the level of up-regulation of target genes was the strongest for these two paralogs. In these chimeric constructs, the N-terminal AD of ORCA3 was fused to the C-terminal DBD of ORCA4 (ORCA3-ORCA4) and *vice versa* (ORCA4-ORCA3) (Fig. 7a). Overexpression of these "*ORCA*-swaps" in combination with *CrMYC2a*^{D126N} led to shifts in the up-regulation of root MIA genes (Fig. 7b). Overexpression of *ORCA3-ORCA4* led to a 2 to 4 fold higher up-regulation of *root MIA genes* than overexpression of native *ORCA4*, whereas overexpression of

ORCA4-ORCA3 led to around 4-fold lower up-regulation than overexpression of native *ORCA3*. These results indicate that the AD of ORCA3 might at least in part be necessary for the synergistic up-regulation. On the other hand, overexpression of *CrMYC2a*^{D126N}/*ORCA3-ORCA4* did not lead to a higher up-regulation of *GO* than overexpression of *CrMYC2a*^{D126N}/*ORCA4*, further supporting the previous observation that the synergistic up-regulation is specific to *root MIA* genes, i.e. target gene specific.

Cell-specific regulation of the *vindoline* branch

Data from the cell type-enriched transcriptomes pointed towards a number of epidermis-specific TFs that might be candidate regulators of the epidermis-specific expression of some MIA pathway genes (*strictosidine* and *stemmadenine* branches and parts of the *vindoline* branch and catharanthine biosynthesis). Among these potential candidates, two MYB TFs were highly expressed in the stem epidermis (Fig. S9b). Both MYB TFs, hereafter named MYB96 and MYB96b, appeared to be potential orthologs of AtMYB96 and 94 (Fig. S9a), which amongst others positively regulate cuticular wax biosynthesis genes in Arabidopsis (Seo *et al.*, 2011; Lee & Suh, 2015; Lee *et al.*, 2016). In a preliminary screen, both MYB TFs were found to transactivate the *T3R* promoter in tobacco protoplasts, albeit not additively (Fig. S9c). Overexpression of *MYB96* in flower petals also led to an up-regulation of expression of *T3R* and of a putative ortholog of the *Arabidopsis* cuticular wax biosynthesis gene *CER1*, but this was not the case for *MYB96b* (Fig. S9d). Nonetheless, the combined overexpression of *MYB96* and *MYB96b* led to a very small increase in catharanthine levels, which was not observed upon overexpression of *MYB96* alone (Fig. S9g,h). Notably however, because we neither detected any up-regulation of other *vindoline pathway* genes by *MYB96* (Fig. S9d) nor did MYB96 transactivate promoters of other vindoline pathway genes in tobacco protoplasts (Fig.S9c), MYB96 appears to be specific to *T3R*. In an attempt to enhance the vindoline content we further overexpressed *MYB96* in combination with the above-described *MYC2a*^{D126N}*-ORCA4* module, which led to a higher availability of the vindoline precursor 16-hydroxytabersonine (Fig. S10). We observed that overexpression of *MYB96*, in addition to the observed up-regulation of *T3R* and *CER1*, also led to a slight but statistically significant up-regulation (around 2-fold) of *stemmadenine pathway* genes

and to a significant up-regulation of the catharanthine exporter gene *TPT2* (at least 4-fold) (Fig. S10). However, at the metabolite level, the overexpression of *MYB96* solely or in combination with *ORCA4* and/or *CrMYC2a*^{D126N} did not lead to increased vindoline or catharanthine content (Fig. S10d; Table S10). Furthermore, it should be noted that overexpression of *CrMYC2a*^{D126} appeared to lead to a down-regulation of most vindoline pathway genes except *T3R* (Fig. S10a), further suggesting that a synergistic/additive action between these TFs seems unlikely.

Discussion

While *C. roseus* is traditionally known as a medicinal plant, it is also increasingly being established as a model species for plant specialized metabolism in general. Thanks to the extensive work in the past decades by several labs, including ours, our knowledge about the network of MIA pathway branches represents one of the most detailed known biosynthesis networks of specialized metabolites, both at the enzyme and regulator level. In this study, we first aimed to identify new regulators for the different MIA branches, using the currently available pathway information. Thereby, we further underscore the importance of specific ORCA and bHLH TFs as modular regulatory hubs for different MIA pathway branches. While these TF families have been already generally implicated in the regulation of specialized metabolism, our study additionally provides evidence for the functional specialization of TF members of the same clades due to specific AAR changes in the DBD.

Functional redundancy and diversification among BIS and ORCA TF clusters

Clusters of TF paralogs appear among various species and families; the orthologs of *ORCA* genes occur for instance also as gene clusters in tobacco, tomato and potato (Shoji *et al.*, 2010; Cárdenas *et al.*, 2016; Kajikawa *et al.*, 2017) and so do the *Arabidopsis* BIS orthologs *bHLH18, 19* and *20* (Cui *et al.*, 2018). Our findings indicate that redundancy and neo/sub-functionalization may occur at the same time for different target genes. In case of the ORCA cluster, we provide evidence that they activate those pathway genes necessary to synthesize common precursors, i.e. “early pathway genes” in a redundant manner. Conversely, later MIA pathway branches are then controlled by individual members in a more branch-specific manner. ORCA3,

for instance, is more specific to some of the root MIAs and ORCA4 to the akuammicine branch. ORCA6 seems to have a rather limited impact on known MIA pathway genes, with the exception of a moderate up-regulation of the root MIA genes. The latter results are consistent with a recent study (Singh *et al.*, 2020). These observations are similar to what has been described for the *Medicago truncatula* BIS orthologs, which redundantly activate early triterpene saponin biosynthesis genes but differentially activate downstream branches in the triterpene biosynthesis pathways and/or in different organs (Mertens *et al.*, 2016; Ribeiro *et al.*, 2020).

Specific AARs in the DBDs of ORCA TFs modulate target gene specificity

Despite increasing knowledge about the DNA binding motifs of different TF families, much less is known about the specific binding affinities of TFs belonging to the same clade. A previous study compared the differential DNA binding affinities of ORCA3 and its orthologs from different species and defined some AARs conferring DNA binding specificity (Shoji *et al.*, 2013). Here, we specifically compare all members of clade XIa AP2/ERFs within the same species under the same experimental conditions and determine AARs conferring specificity that had not been identified previously. It will be interesting to further investigate whether similar AAR changes lead to pathway branch specialization in the other species with several paralogs and, moreover, determine paralog-specific DNA binding site sequences. Ultimately, further knowledge about such key AARs may offer more precise possibilities for metabolic engineering. For instance, the native version of a given TF might not be the most desirable one for flux into a specific pathway branch but an engineered version in which a key amino acid is substituted might do so. This smallest change for metabolic engineering purposes might be referred to as metabolic editing (Swinnen *et al.*, 2019). Similarly, once the respective DNA binding motifs of a TF are known, specific branches could be brought under control of that TF by promoter base editing (Swinnen *et al.*, 2016).

MYC2 – a long-known hub with remaining mechanistic mysteries

As described here and elsewhere, many MIA pathway genes are up-regulated by JA, which can be mimicked by overexpression of the de-repressed $CrMYC2a^{D126N}$

(Schweizer *et al.*, 2018). In addition, we have previously shown that the combinatorial overexpression with *ORCA3* leads to an up-regulation of the *root MIA pathway* genes and other target genes in a synergistic manner (Schweizer *et al.*, 2018). A similar synergistic up-regulation of target genes has previously been observed between the respective tomato and tobacco MYC2 and AP2/ERF orthologs, suggesting an evolutionary conservation of this combinatorial TF module (De Boer *et al.*, 2011; Cárdenas *et al.*, 2016). Here, we observe that in *C. roseus* the synergistic transactivation is to some extent paralog specific, as the level of synergy is higher in combination with ORCA3 than with ORCA4 for instance. Although we cannot rule out that this is (in part) invoked by different protein stabilities, our results further suggest that this is at least in part due to the AD as we could observe that the AD of ORCA3 significantly enhanced the level of target gene up-regulation when fused to the DBD of ORCA4, whereas the reverse swap rather led to a reduced up-regulation. Although ADs are generally little conserved at the sequence level, it has been shown that specific structural features or subdomains are essential for the TF activity, for instance for the recruitment of the transcriptional machinery (Staby *et al.*, 2017). Noteworthy, *in silico* analyses of the ORCA ADs predict structural differences, which further supports that they are functionally diverse. It remains to be uncovered if and which of these structural differences confer the ORCA3-specific synergistic up-regulation of target genes and if this is for instance mediated by direct physical interaction between the TFs or by other mechanisms. Likewise, more investigation is needed to clarify how the synergistic up-regulation is restricted to specific target genes, i.e. the *root MIA* genes, whereas the up-regulation of other target genes, for instance the BIS TFs, does not appear synergistic.

Outlook: touching the boundaries of co-expression analysis as a tool to find regulator candidates

Co-expression analysis, which is based on the assumption that the expression profiles of functionally related genes correlate, remains one of the most popular and successful strategies to identify both metabolic pathway genes and regulatory TFs (Movahedi *et al.*, 2012; Goossens, 2015; Wisecaver *et al.*, 2017). However, it has also been noted previously that a successful outcome is greatly dependent on the combination of datasets included in the analysis, the specific methodology used and

others (Goossens, 2015; Uygun *et al.*, 2016). In the case of our study, there are a number of conceivable reasons, technical or mechanistic, why we were not able to unambiguously uncover regulators for all MIA branches yet, e.g. the *vindoline pathway*. First, the number of available expression datasets for *C. roseus* is still limited to around 80. Hence, if the pathway branch of interest does not show a (highly) differential expression profile among these datasets, the correlation with expression profiles of regulator candidates might be low. More specifically, the induction of regulatory TFs identified in past studies on *C. roseus* and of their target MIA pathway genes upon JA exposure is highly correlated because of the renowned amplification loops in the JA signaling cascade (De Geyter *et al.*, 2012), which facilitated their discoveries. By contrast, the *vindoline pathway* is not induced upon JA exposure and, apart from a higher expression in green tissues, it is not differentially expressed otherwise. Hence, this pathway branch is either not differentially expressed in response to a specific environmental condition, or a discriminating RNA-Seq dataset for this hypothetical condition is not yet available. Further, transcriptional regulation may involve independently acting TFs individually mediating the specific responses to environment, development, cell type and others. The expression of such a condition-related TF might thus only be strongly correlated to its MIA target genes in a subselection of datasets and not across all datasets. Mining of selected dataset combinations or application of a bi-clustering approach to identify significant co-expression in a subset of samples, could improve this and could be recommended for future co-expression analyses, but would possibly also yield a much longer list of candidates. Next, at the experimental level, it is possible that the candidate list indeed contains positive regulators but that (i) their cloning failed, (ii) they were not sufficiently overexpressed, (iii) their target gene induction required the combined action of TFs (analogous to the ORCA3-CrMYC2a^{D126N} module) but the respective TFs did not happen to be in the same co-overexpression group, (iv) they were not active in flower petals due to epigenetic or post-translational mechanisms.

Our study, together with the vast work previously published by different research groups extensively illustrates that MYC2 TFs globally up-regulate MIA biosynthesis together with the branch-specific ORCA and BIS TFs with the exception of the *vindoline* branch. This raises questions about the biological function of vindoline and/or the intermediates in this pathway branch. While MIA biosynthesis has been generally connected to defense against for instance caterpillars to some

extent (Dugé de Bernonville *et al.*, 2017), the chemical ecology of the different specific MIAs is largely unknown. The decoupled expression profile of the *vindoline pathway* from other MIA branches and the absence of up-regulation by CrMYC2a^{D126N} (rather a potential down-regulation) concomitant with a light-induced regulation raise the possibility of other or additional biological functions of vindoline. Nevertheless, the biosynthesis (and/or putative transport) of precursors must be assured, a process that appears not to be light regulated. It is clear that additional studies, likely not based on transcriptome mining solely, are therefore necessary to further investigate how these different signals are integrated to regulate flux through the distinct MIA pathway branches.

Acknowledgements

We thank Robin Vanden Bossche for excellent technical assistance, and Alex Van Moerkercke for generating the RNA-Seq data of transformed *C. roseus* hairy root lines. M.C. and F.S. are indebted to the Swiss National Science Foundation (P300PA_177831 and P300PA_167772). This work was supported by European Union's Horizon 2020 research and innovation program under Grant Agreement No 760331 (Newcotiana). This work was supported by the Agency for Innovation by Science and Technology (IWT) in Flanders (predoctoral fellowship to D.V.). J.G.G. is indebted to the Portuguese Fundação para a Ciência e Tecnologia (FCT) and FEDER Funds (SFRH/BD/97590/2013).

Author contributions

M.C. and A.G. designed the research. M.C. performed the research and data analysis with contributions from D.M., F.S, R.D.C, J.G.G, T.M.-C. F.J.M.H and M.S.. D.V. and K.V. performed RNA-Seq data and co-expression analysis and inferred candidate regulators.. J.P. collected and analyzed metabolite data and performed compound identification. M.C. and A.G. wrote the manuscript with consent from all other authors.

References

Aerts RJ, Gisi D, Decarolis E, Deluca V, Baumann TW. 1994. Methyl jasmonate vapor increases the developmentally controlled synthesis of alkaloids in *Catharanthus* and *Cinchona* seedlings. *Plant Journal* **5**(5): 635-643.

Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M. 1998. A novel mode of DNA recognition by a β-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO Journal* **17**(18): 5484-5496.

Asada K, Salim V, Masada-Atsumi S, Edmunds E, Nagatoshi M, Terasaka K, Mizukami H, De Luca V. 2013. A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *Plant Cell* **25**(10): 4123-4134.

Besseau S, Kellner F, Lanoue A, Thamm AMK, Salim V, Schneider B, Geu-Flores F, Höfer R, Guirimand G, Guihur A, et al. 2013. A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus. Plant Physiology* **163**(4): 1792-1803.

Caputi L, Franke J, Farrow SC, Chung K, Payne RME, Nguyen T-D, Dang T-T, Soares Teto Carqueijeiro I, Koudounas K, Dugé de Bernonville T, et al. 2018. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* **360**(6394): 1235-1239.

Cárdenas PD, Sonawane PD, Pollier J, Vanden Bossche R, Dewangan V, Weithorn E, Tal L, Meir S, Rogachev I, Malitsky S, et al. 2016. GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nature Communications* **7**: 10654.

Carqueijeiro I, Dugé de Bernonville T, Lanoue A, Dang T-T, Teijaro CN, Paetz C, Billet K, Mosquera A, Oudin A, Besseau S, et al. 2018. A BAHD acyltransferase catalyzing 19-*O*-acetylation of tabersonine derivatives in roots of *Catharanthus roseus* enables combinatorial synthesis of monoterpene indole alkaloids. *Plant Journal* **84**: 469-484.

Carqueijeiro IT, Brown S, Chung K, Dang T-T, Walia M, Besseau S, Dugé de Bernonville T, Oudin A, Lanoue A, Billet K, et al. 2018. Two tabersonine 6,7-epoxidases start synthesis of lochnericine-type alkaloids in *Catharanthus roseus*. *Plant Physiology* **177**: 1473-1486.

Chini A, Fonseca S, Fernández G, Adie B, Chico JM, Lorenzo O, García-Casado G, López-Vidriero I, Lozano FM, Ponce MR, et al. 2007. The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448**(7154): 666-671.

Colinas M, Goossens A. 2018. Combinatorial transcriptional control of plant specialized metabolism. *Trends in Plant Science* **23**(4): 324-336.

Costa MM, Hilliou F, Duarte P, Pereira LG, Almeida I, Leech M, Memelink J, Barcelo AR, Sottomayor M. 2008. Molecular cloning and characterization of a vacuolar class III peroxidase involved in the metabolism of anticancer alkaloids in *Catharanthus roseus. Plant Physiology* **146**(2): 403-417.

Courdavault V, Papon N, Clastre M, Giglioli-Guivarc'h N, St-Pierre B, Burlat V. 2014. A look inside an alkaloid multisite plant: the *Catharanthus* logistics. *Current Opinion in Plant Biology* **19**: 43-50.

Cui Y, Chen C-L, Cui M, Zhou W-J, Wu H-L, Ling H-Q. 2018. Four IVa bHLH Transcription factors are novel interactors of fit and mediate JA inhibition of iron uptake in *Arabidopsis*. *Molecular Plant* **11**(9): 1166-1183.

Dang T-TT, Franke J, Carqueijeiro IST, Langley C, Courdavault V, O'Connor SE. 2018. Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nature Chemical Biology* **14**: 760-763.

De Boer K, Tilleman S, Pauwels L, Vanden Bossche R, De Sutter V, Vanderhaeghen R, Hilson P, Hamill JD, Goossens A. 2011. APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. *Plant Journal* **66**(6): 1053-1065.

De Geyter N, Gholami A, Goormachtig S, Goossens A. 2012. Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends in Plant Science* **17**(6): 349-359.

Dugé de Bernonville T, Carqueijeiro I, Lanoue A, Lafontaine F, Sánchez Bel P, Liesecke F, Musset K, Oudin A, Glévarec G, Pichon O, et al. 2017. Folivory elicits a strong defense reaction in *Catharanthus roseus*: metabolomic and transcriptomic analyses reveal distinct local and systemic responses. *Scientific Reports* **7**: 40453.

Dugé de Bernonville T, Clastre M, Besseau S, Oudin A, Burlat V, Glévarec G, Lanoue A, Papon N, Giglioli-Guivarc'h N, St-Pierre B, et al. 2015. Phytochemical genomics of the Madagascar periwinkle: Unravelling the last twists of the alkaloid engine. *Phytochemistry* **113**: 9-23.

Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, Wu C, Nims E, Cui Y, O'Connor SE. 2012. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**(7427): 138-142.

Giddings L-A, Liscombe DK, Hamilton JP, Childs KL, DellaPenna D, Buell CR, O'Connor SE. 2011. A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus. Journal of biological chemistry* **286**(19): 16751-16757.

Goossens A. 2015. It is easy to get huge candidate gene lists for plant metabolism now, but how to get beyond? *Molecular plant* **8**(1): 2-5.

Goossens J, Fernández-Calvo P, Schweizer F, Goossens A. 2016. Jasmonates: signal transduction components and their roles in environmental stress responses. *Plant Molecular Biology* **91**(6): 673-689.

Goossens J, Mertens J, Goossens A. 2017. Role and functioning of bHLH transcription factors in jasmonate signalling. *Journal of Experimental Botany* **68**(6): 1333-1347.

Guirimand G, Guihur A, Poutrain P, Héricourt F, Mahroug S, St-Pierre B, Burlat V, Courdavault V. 2011. Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus. Journal of Plant Physiology* **168**(6): 549-557.

Kajikawa M, Sierro N, Kawaguchi H, Bakaher N, Ivanov NV, Hashimoto T, Shoji T. 2017. Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. *Plant Physiology* **174**(2): 999-1011.

Karimi M, Inzé D, Depicker A. 2002. GATEWAY^{TM} vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science* **7**(5): 193-195.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**(6): 845-858.

Kellner F, Kim J, Clavijo BJ, Hamilton JP, Childs KL, Vaillancourt B, Cepela J, Habermann M, Steuernagel B, Clissold L, et al. 2015. Genome-guided investigation of plant natural product biosynthesis. *Plant Journal* **82**(4): 680-692.

Laflamme P, St-Pierre B, De Luca V. 2001. Molecular and biochemical analysis of a Madagascar periwinkle root-specific minovincinine-19-hydroxy-O-acetyltransferase. *Plant Physiology* **125**(1): 189-198.

Larsen B, Fuller VL, Pollier J, Van Moerkercke A, Schweizer F, Payne R, Colinas M, O'Connor SE, Goossens A, Halkier BA. 2017. Identification of iridoid glucoside transporters in *Catharanthus roseus*. *Plant & Cell Physiology* **58**(9): 1507-1518.

Lee SB, Kim HU, Suh MC. 2016. MYB94 and MYB96 additively activate cuticular wax biosynthesis in Arabidopsis. *Plant & Cell Physiology* **57**(11): 2300-2311.

Lee SB, Suh MC. 2015. Cuticular wax biosynthesis is up-regulated by the MYB94 transcription factor in Arabidopsis. *Plant & Cell Physiology* **56**(1): 48-60.

Levac D, Murata J, Kim WS, De Luca V. 2008. Application of carborundum abrasion for investigating the leaf epidermis: molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-*O*-methyltransferase. *Plant Journal* **53**(2): 225-236.

Liscombe DK, O'Connor SE. 2011. A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. *Phytochemistry* **72**(16): 1969-1977.

Liscombe DK, Usera AR, O'Connor SE. 2010. Homolog of tocopherol C methyltransferases catalyzes *N* methylation in anticancer alkaloid biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **107**(44): 18793-18798.

Liu J, Cai J, Wang R, Yang S. 2017. Transcriptional regulation and transport of terpenoid indole alkaloid in *Catharanthus roseus*: exploration of new research directions. *International Journal of Molecular Sciences* **18**(1): 53.

Liu Y, Patra B, Pattanaik S, Wang Y, Yuan L. 2019. GATA and phytochrome interacting factor transcription factors regulate light-induced vindoline biosynthesis in *Catharanthus roseus. Plant Physiology* **180**(3): 1336-1350.

Menke FLH, Champion A, Kijne JW, Memelink J. 1999. A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. *EMBO Journal* **18**(16): 4455-4463.

Mertens J, Pollier J, Vanden Bossche R, Lopez-Vidriero I, Franco-Zorrilla JM, Goossens A. 2016. The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiology* **170**(1): 194-210.

Metsalu T, Vilo J. 2015. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research* **43**(W1): W566-W570.

Miettinen K, Dong L, Navrot N, Schneider T, Burlat V, Pollier J, Woittiez L, van der Krol S, Lugan R, Ilc T, et al. 2014. The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications* **5**: 3606.

Movahedi S, Van Bel M, Heyndrickx KS, Vandepoele K. 2012. Comparative co-expression analysis in plant biology. *Plant, Cell & Environment* **35**(10): 1787-1798.

Pan Y-j, Lin Y-c, Yu B-f, Zu Y-g, Yu F, Tang Z-H. 2018. Transcriptomics comparison reveals the diversity of ethylene and methyl-jasmonate in roles of TIA metabolism in *Catharanthus roseus. BMC Genomics* **19**(1): 508.

Paul P, Singh SK, Patra B, Liu X, Pattanaik S, Yuan L. 2020. Mutually regulated AP2/ERF gene clusters modulate biosynthesis of specialized metabolites in plants. *Plant Physiology* **182**: 840-856.

Paul P, Singh SK, Patra B, Sui X, Pattanaik S, Yuan L. 2016. A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus. New Phytologist* **213**(3): 1107-1123.

Qu Y, Easson MEAM, Simionescu R, Hajicek J, Thamm AMK, Salim V, De Luca V. 2018. Solution of the multistep pathway for assembly of corynanthean, strychnos, iboga, and aspidosperma monoterpenoid indole alkaloids from 19*E*-geissoschizine. *Proceedings of the National Academy of Sciences of the United States of America* **115**(12): 3180-3185.

Qu Y, Easson MLAE, Froese J, Simionescu R, Hudlicky T, De Luca V. 2015. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **112**(19): 6224-6229.

Qu Y, Safonova O, Luca V. 2019. Completion of the canonical pathway for assembly of anticancer drugs vincristine/vinblastine in *Catharanthus roseus. Plant Journal* **97**(2): 257-266.

Ribeiro B, Lacchini E, Bicalho KU, Mertens J, Arendt P, Vanden Bossche R, Calegario G, Gryffroy L, Ceulemans E, Buitink J, et al. 2020. A seed-specific regulator of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell* DOI: 10.1105/tpc.19.00609

Roepke J, Salim V, Wu M, Thamm AMK, Murata J, Ploss K, Boland W, De Luca V. 2010. Vinca drug components accumulate exclusively in leaf exudates of Madagascar

periwinkle. *Proceedings of the National Academy of Sciences of the United States of America* **107**(34): 15287-15292.

Salim V, Wiens B, Masada-Atsumi S, Yu F, De Luca V. 2014. 7-deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in *Catharanthus roseus* iridoid biosynthesis. *Phytochemistry* **101**: 23-31.

Schweizer F, Colinas M, Pollier J, Van Moerkercke A, Vanden Bossche R, de Clercq R, Goossens A. 2018. An engineered combinatorial module of transcription factors boosts production of monoterpenoid indole alkaloids in *Catharanthus roseus*. *Metabolic Engineering* **48**: 150-162.

Seo PJ, Lee SB, Suh MC, Park M-J, Go YS, Park C-M. 2011. The MYB96 Transcription factor regulates cuticular wax biosynthesis under drought conditions in *Arabidopsis*. *Plant Cell* **23**(3): 1138-1152.

Shoji T, Kajikawa M, Hashimoto T. 2010. Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell* **22**(10): 3390-3409.

Shoji T, Mishima M, Hashimoto T. 2013. Divergent DNA-binding specificities of a group of ETHYLENE RESPONSE FACTOR transcription factors involved in plant defense. *Plant Physiology* **162**(2): 977-990.

Simkin AJ, Miettinen K, Claudel P, Burlat V, Guirimand G, Courdavault V, Papon N, Meyer S, Godet S, St-Pierre B, et al. 2013. Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry* **85**: 36-43.

Singh SK, Patra B, Paul P, Liu Y, Pattanaik S, Yuan L. 2020. Revisiting the ORCA gene cluster that regulates terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Science* **293**: 110408.

St-Pierre B, De Luca V. 1995. A cytochrome P-450 monooxygenase catalyzes the first step in the conversion of tabersonine to vindoline in *Catharanthus roseus. Plant Physiology* **109**(1): 131-139.

St-Pierre B, Vazquez-Flota FA, De Luca V. 1999. Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. *Plant Cell* **11**(5): 887-900.

Staby L, O'Shea C, Willemoës M, Theisen F, Kragelund BB, Skriver K. 2017. Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochemical Journal* **474**(15): 2509-2532.

Stavrinides A, Tatsis EC, Caputi L, Foureau E, Stevenson CEM, Lawson DM, Courdavault V, O'Connor SE. 2016. Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nature Communications* **7**: 12116.

Stavrinides AK, Tatsis EC, Dang T-T, Caputi L, Stevenson CEM, Lawson DM, Schneider B, O'Connor SE. 2018. Discovery of a short-chain dehydrogenase from *Catharanthus roseus* that produces a new monoterpene indole alkaloid. *ChemBioChem* **19**(9): 940-948.

Swinnen G, Goossens A, Colinas M. 2019. Metabolic editing: small measures, great impact. *Current Opinion in Biotechnology* **59**: 16-23.

Swinnen G, Goossens A, Pauwels L. 2016. Lessons from domestication: targeting *cis*-regulatory elements for crop improvement. *Trends in Plant Science* **21**(6): 506-515.

Tatsis EC, Carqueijeiro I, Dugé de Bernonville T, Franke J, Dang T-T, Oudin A, Lanoue A, Lafontaine F, Stavrinides AK, Clastre M, et al. 2017. A three enzyme system to generate the *Strychnos* alkaloid scaffold from a central biosynthetic intermediate. *Nature Communications* **8**(1): 316.

Uygun S, Peng C, Lehti-Shiu MD, Last RL, Shiu S-H. 2016. Utility and limitations of using gene expression data to identify functional associations. *PLoS Computational Biology* **12**(12): e1005244.

van der Fits L, Memelink J. 2000. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* **289**(5477): 295-297.

van der Fits L, Memelink J. 2001. The jasmonate-inducible AP2/ERF-domain transcription factor ORCA3 activates gene expression via interaction with a jasmonate-responsive promoter element. *Plant Journal* **25**(1): 43-53.

van der Heijden R, Jacobs DI, Snoeijer W, Hallard D, Verpoorte R. 2004. The *Catharanthus* alkaloids: pharmacognosy and biotechnology. *Current medicinal chemistry* **11**(5): 607-628.

Van Moerkercke A, Steensma P, Gariboldi I, Espoz J, Purnama PC, Schweizer F, Miettinen K, Vanden Bossche R, De Clercq R, Memelink J, et al. 2016. The basic helix-loop-helix transcription factor BIS2 is essential for monoterpenoid indole alkaloid production in the medicinal plant *Catharanthus roseus. Plant Journal* **88**(1): 3-12.

Van Moerkercke A, Steensma P, Schweizer F, Pollier J, Gariboldi I, Payne R, Vanden Bossche R, Miettinen K, Espoz J, Purnama PC, et al. 2015. The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences of the United States of America* **112**(26): 8130-8135.

Vanden Bossche R, Demedts B, Vanderhaeghen R, Goossens A. 2013. Transient expression assays in tobacco protoplasts. *Methods in Molecular Biology* **1011**: 227-239.

Vom Endt D, Soares e Silva M, Kijne JW, Pasquali G, Memelink J. 2007. Identification of a bipartite jasmonate-responsive promoter element in the *Catharanthus roseus ORCA3* transcription factor gene that interacts specifically with AT-hook DNA-binding proteins. *Plant Physiology* **144**(3): 1680-1689.

Wasternack C, Strnad M. 2019. Jasmonates are signals in the biosynthesis of secondary metabolites — Pathways, transcription factors and applied aspects — A brief review. *New Biotechnology* **48**: 1-11.

Williams D, Qu Y, Simionescu R, De Luca V. 2019. The assembly of (+)-vincadifformine- and (−)-tabersonine-derived monoterpenoid indole alkaloids in *Catharanthus roseus* involves separate branch pathways. *Plant Journal* **99**(4): 626-636.

Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. 2017. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* **29**(5): 944-959.

Yamamoto K, Takahashi K, Caputi L, Mizuno H, Rodriguez-Lopez CE, Iwasaki T, Ishizaki K, Fukaki H, Ohnishi M, Yamazaki M, et al. 2019. The complexity of intercellular localisation of alkaloids revealed by single-cell metabolomics. *New Phytologist* **224**(2): 848-859.

Yamamoto K, Takahashi K, Mizuno H, Anegawa A, Ishizaki K, Fukaki H, Ohnishi M, Yamazaki M, Masujima T, Mimura T. 2016. Cell-specific localization of alkaloids in *Catharanthus roseus* stem tissue measured with Imaging MS and Single-cell MS. *Proceedings of the National Academy of Sciences of the United States of America* **113**(14): 3891-3896.

Yu F, De Luca V. 2013. ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences of the United States of America* **110**(39): 15830-15835.

Zhang H, Hedhili S, Montiel G, Zhang Y, Chatel G, Pré M, Gantet P, Memelink J. 2011. The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the *ORCA* genes that regulate alkaloid biosynthesis in *Catharanthus roseus. Plant Journal* **67**(1): 61-71.

Zhou M, Memelink J. 2016. Jasmonate-responsive transcription factors regulating plant secondary metabolism. *Biotechnology Advances* **34**(4): 441-449.

Supporting information

Methods S1 Generation and mining of the RNA-Seq data compendium.

Methods S2 Flower petal transformation, gene expression analysis and metabolite profiling.

Fig. S1 MIA gene expression levels upon bulk overexpression of TF candidates.

Fig. S2 Separate overexpression of positive TF candidates.

Fig. S3 Comparison of structural features and motifs of ORCA activation domains (ADs).

Fig. S4 Trans-activation of *STR, GO, TAT* and *MAT* promoter fragments by ORCAs.

Fig. S5 MIA levels upon overexpression of *ORCA*s.

Fig. S6 MIA gene expression levels upon overexpression of *ORCA3* single mutants.

Fig. S7 MIA levels upon overexpression of *ORCA3* mutants.

Fig. S8 MIA levels upon combinatorial overexpression of $CrMYC2a^{D126N}$ and *ORCAs*.

Fig. S9 Orthologs of cuticular wax regulators up-regulate the vindoline pathway gene *T3R.*

Fig. S10 Combinatorial overexpression of *ORCA4, CrMYB96,* $CrMYC2a^{D126N}$.

Table S1 List of RNA-Seq datasets used.

Table S2 Oligonucleotide primers used for cloning of coding sequences.

Table S3 Oligonucleotide primers used for cloning of promoter fragments.

Table S4 List of all cloned TFs and promoter fragments and Genebank IDs.

Table S5 Oligonucleotide primers used for qPCR.

Table S6 MIA levels upon overexpression of *ORCAs*.

Table S7 MIA levels upon overexpression of *ORCA3* mutants.

Table S8 MIA levels upon combinatorial overexpression of *CrMYC2a*^{D126N} and *ORCAs*.

Table S9 MIA levels upon overexpression of *MYB96/b*.

Table S10 MIA levels upon overexpression of *MYB96/b*, *CrMYC2a*^{D126N} and *ORCA4*.

Supporting Dataset I *C. roseus* expression atlas.

Supporting Dataset II MIAs identified in this study.

FIGURE LEGENDS

Fig. 1 The MIA pathway consists of different branches with corresponding co-expression networks. (a) The MIA network and known or presumed cellular and organ-specific expression of the encoding MIA biosynthesis genes based on the current knowledge and available data. MIA biosynthesis is compartmentalized into the three different cell types shown on the left, although a cell type-specific expression has not been experimentally verified for all MIA biosynthesis genes such as the root MIA biosynthesis genes. Available organ-specific transcriptome data revealed that in particular two MIA branches are either root- or leaf-specific as shown on the right. In our study, transient overexpression of TFs in *C. roseus* flower petals is used to assess the impact on MIA biosynthesis at the gene expression level (by qPCR) and MIA metabolite level (by LC-MS analysis). Some MIA derivatives are detected in this flower petal expression system in this study, but it is not known where they occur in wild-type plants (upper right). Arrows indicate known characterized enzymatic reactions and corresponding enzymes; dashed black arrows indicate assumed enzymatic or chemical reactions for which no specific enzyme has been identified. Dashed grey arrows indicate transport and corresponding transporters. Note that NPF2.9 is an intracellular transporter conferring vacuolar export of strictosidine and TPT2 is an epidermis-specific transporter exporting catharanthine to the leaf surface. For clarity, only selected MIA metabolites are shown. (b) Co-expression network of MIA biosynthesis genes. From an expression atlas of 82 samples k-nearest neighbor (KNN) networks were created. A line between genes indicates co-expression. Opacity of the lines is scaled according to the mutual rank of co-expression between two genes in their KNN 500 clusters (darker = stronger co-expression). To simplify the complex MIA biosynthesis network, we have colored the different pathway branches. Abbreviations: GES, geraniol synthase; G8O, geraniol-8-oxidase; 8HGO, 8-hydroxygeraniol oxidoreductase; IS, iridoid synthase; IO, iridoid oxidase; 7DLGT, 7-deoxyloganetic acid glucosyl transferase; 7DLH, 7-deoxyloganic acid hydroxylase; LAMT, loganic acid *O*-methyltransferase; SLS, secologanin synthase; TDC, tryptophan decarboxylase; STR, strictosidine synthase; NPF, nitrate/peptide family; SGD, strictosidine β-glucosidase; VAS, vitrosamine synthase; THAS, tetrahydroalstonine synthase; AS, alstonine synthase: HYS, heteroyohimbine synthase; GS, geissoschizine synthase; GO, geissoschizine oxidase; SAT,

stemmadenine-*O*-acetyl- transferase; PAS, precondylocarpine acetate synthase; DPAS, dihydroprecondylocarpine synthase; CS, catharanthine synthase; TPT, catharanthine transporter; T16H, tabersonine 16-hydroxylase; 16OMT, 16-hydroxytabersonine *O*-methyltransferase; T3O, tabersonine 3-oxygenase; T3R, tabersonine 3-reductase; NMT, 2,3-dihydro-3-hydroxytabersonine-N-methyltransferase; D4H, desacetoxyvindoline 4-hydroxylase; DAT, deacetylvindoline 4-O-acetyltransferase; PRX, peroxidase; HL, hydrolase; V19H, vincadifformine-19-hydroxylase; MAT, minovincinine-19-*O*-acetyltransferase; T19H, tabersonine 19-hydroxylase; TAT, 19-hydroxytabersonine 19-*O*-acetyltransferase; TEX, tabersonine epoxidase.

Fig. 2 MIA regulator candidate selection and screening procedure. (a) Pathway genes (left) are plotted against regulator candidates (below); the relative rank of a regulator in the top 500 most co-expressed genes is depicted in red, with dark red indicating low ranks and strong associations between regulator and pathway gene. TF gene IDs are color-coded, indicating which genes by manual sequence analyses are likely not specific TFs and were thus not further considered (gray), which were considered for screening but cloning failed (yellow), and which were tested (green). Previously characterized TFs are labeled with their names in addition to "BIS3", identified in this study. (b) Overview about the screening procedure. To enhance the output, TF candidates were bulk-overexpressed ("Screening groups") by mixing *A. tumefaciens* strains containing the respective constructs. The screening was performed in two independent rounds of infiltration ("Screening infiltration 1 and 2"). In case of a detected up-regulation of target genes >2-fold and $P<0.05$ (tested by Student's *t*-test) (Fig. S1), the TF candidates from the respective group were overexpressed separately (Fig. S2). Gene IDs in italics were not picked in this co-expression analysis but in a previous one.

Fig. 3 BIS3 up-regulates iridoid pathway genes. (a) The latest version of the *C. roseus* genome revealed that the previously published BIS1 and 2, and BIS3 identified in this study, are all located in close proximity and probably arose from recent gene duplications. (b) A MUSCLE alignment of BIS1, 2 and 3 protein sequences showing

a high degree of sequence identity (37.6% of identical sites between the three BIS). The sequence identity between BIS2 and 3 is much higher (76.2%). Sites that are identical in all three sequences are colored in black; sites identical in two out of three sequences are colored in gray. (c) Overexpression of *BIS3* leads to increased expression levels of iridoid pathway genes similar to those previously observed for *BIS1* and *2* previously. *BIS1* and *2* themselves are not up-regulated by BIS3, suggesting that the iridoid pathway up-regulation is directly caused by BIS3 and that there is no cross up-regulation between the three BIS TFs. Different letters indicate statistically significant differences in expression ($P < 0.05$) as calculated by Student's *t*-test for each gene.

Fig. 4 Members of the ORCA TF cluster differentially up-regulate MIA biosynthesis genes. **(**a) Localization of the *ORCA* genes within the cluster. (b) Expression pattern of *ORCA* paralogs extracted from publicly available RNA-Seq data from indicated tissues and conditions. TPM, Transcripts Per Kilobase Million. (c) A PCA of MIA pathway genes measured by qPCR upon *ORCA* overexpression (shown in d) suggests overlapping but also differential MIA target gene expression patterns in particular between *ORCA2, 3* and *4*. PCA was performed with ClustVis (Metsalu & Vilo, 2015). (d) Expression of MIA biosynthesis genes measured by qPCR upon *ORCA* overexpression shown as log2 fold changes (FC) compared to the overexpression of *GUS* as a control. For each pathway gene, cells labeled with different letters represent statistically significant differences ($P<0.05$, ANOVA with Tukey's correction for multiple comparisons). Note that the overexpression levels of *ORCA3, 4* and *6* were higher than those of *ORCA2* and *5*. (e) PCA of metabolite profiling data from *ORCA* overexpression samples. (f) Loading plot of the PCA shown in (e), indicating that the differences found in the PCA are indeed predominantly caused by MIA metabolite levels. Selected known MIAs are labeled in blue; MIAs identified in this study (see Supporting Dataset S1) are labeled in red. (g) Levels of selected MIA metabolites expressed in average total ion current (TIC). The levels of all measured MIA metabolites can be found in the supporting information (Fig. S5; Table S6). Columns labeled with different letters represent statistically significant differences ($P<0.05$, ANOVA with Tukey's correction for multiple comparisons). Error bars are standard error of the mean (SEM) of four biological replicates.

Fig. 5 Specific amino acid residues (AARs) in the DNA-binding domain (DBD) of ORCA TFs influence target gene specificity. (a) Alignment of the DBDs of ORCAs. AARs labeled with an asterisk were identified previously as directly interacting with the DNA based on the solved crystal structure of the *Arabidopsis* AP2/ERF member AtERF1 DBD (Allen *et al*., 1998; Shoji *et al*., 2013). ORCA3 was used as a template to exchange AARs so that they would correspond to AARs in other ORCA TFs. The respective mutated AARs are indicated below the alignment. (b) Homology models of ORCA DBDs. The models (made by Phyre2 "intensive mode" (Kelley *et al.*, 2015); colored as in panel a) are overlaid with the AtERF1 structure bound to DNA (PDB 1GCC) colored in gray. AARs identified as directly interacting with the DNA (asterisks in panel a) are shown as lines of AtERF1. AARs differing in the ORCA DBDs and selected for site-directed mutagenesis are shown as sticks for all ORCA models, while the AAR numbering refers to ORCA3. The models suggest that the differing AARs are unlikely to interact directly with the DNA, but are in close proximity to the DNA and might thereby modulate DNA binding affinity. (c) Overexpression of *ORCA3* mutants in *C. roseus* flower petals and expression level of selected MIA target genes as measured by qPCR. Two AAR changes, K101R and P107L, seem to significantly modulate the MIA target gene profile. While K101R appears to increase the level of up-regulation, up-regulation is greatly reduced and in some cases nearly abolished when overexpressing the $ORCA3^{P107L}$ mutant. The double mutant $ORCA3^{K101RP107L}$, corresponding to the native ORCA4 sequence, recovers some activation potential, and reaches similar levels than observed upon overexpression of *ORCA4* (Fig.4d). This suggests that the combination of these AARs may account for the differences between ORCA3 and 4. For each pathway gene, cells labeled with different letters represent statistically significant differences ($P<0.05$, ANOVA with Tukey's correction for multiple comparisons). Results from overexpression of all *ORCA3* single AAR mutants (including $ORCA3^{D117N}$) can be found in Fig. S6. (d) PCA of metabolite profiling data. (e) Loading plot indicating that sample separation shown in the PCA plot is largely due to changes in MIA metabolite levels. Selected known MIAs are labeled in blue; MIAs identified in this study are labeled in red. (f) Levels of selected MIA metabolites expressed in average total ion current (TIC). The levels of all measured MIA metabolites can be found in the

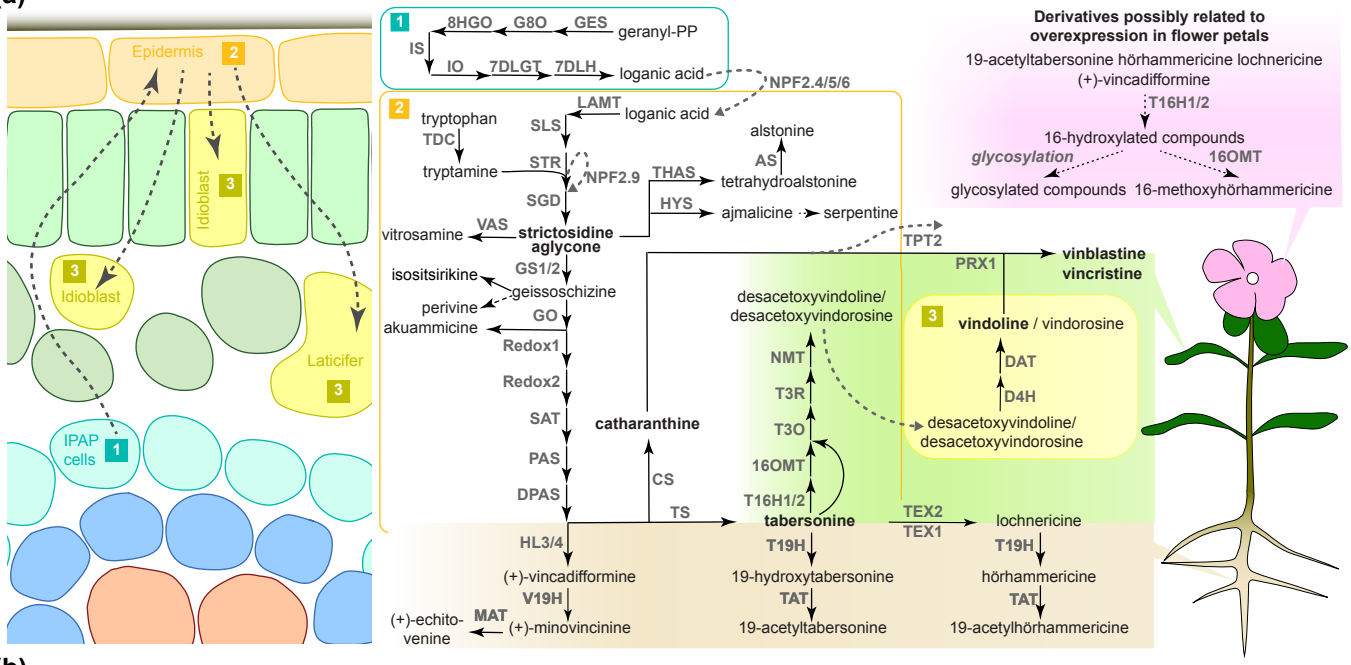
supporting information (Fig. S7; Table S7). Columns labeled with different letters represent statistically significant differences (P<0.05, ANOVA with Tukey's correction for multiple comparisons). Error bars are standard error of the mean (SEM) of four biological replicates.

Fig. 6 Combinatorial overexpression of *CrMYC2a*^{D216N} and *ORCAs*. (a) Expression profile of MIA pathway genes and overexpression levels. For each pathway gene, cells labeled with different letters represent statistically significant differences (P<0.05, ANOVA with Tukey's correction for multiple comparisons). Note that synergistic up-regulation of particularly root MIA genes is most pronounced by combinatorial overexpression of *CrMYC2a*^{D126N} and *ORCA3*. (b) PCA of the MIA gene expression data is suggesting that in particular the combinatorial overexpression of *CrMYC2a*^{D126N} and *ORCA2*, *3* and *4* lead to differential changes in the MIA gene expression pattern, whereas combinatorial overexpression with *ORCA5* or *6* does not differ substantially from overexpression of *CrMYC2a*^{D126N} alone. PCA was performed with ClustVis (Metsalu & Vilo, 2015). (c) PCA of metabolite profiling data. (d) Loading plot of PCA with selected known MIAs (blue) and MIAs identified in this study (red). (e) Levels of selected MIA metabolites expressed in average total ion current (TIC). The levels of all measured MIA metabolites can be found in the supporting information (Fig. S7; Table S8). Columns labeled with different letters represent statistically significant differences (P<0.05, ANOVA with Tukey's correction for multiple comparisons). Error bars are standard error of the mean (SEM).

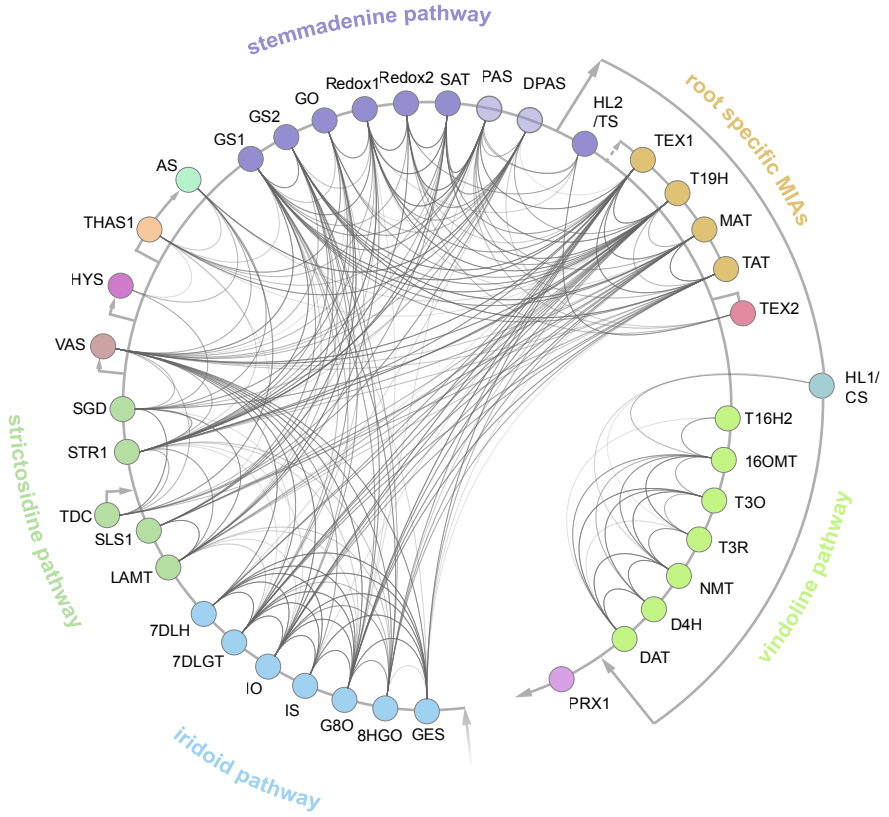
Fig. 7 Swapping of DNA-binding domain (DBD) and activation domain (AD) of ORCA3 and 4. In order to get insight into the role of the AD in the synergistic up-regulation of MIA target genes, the respective DBDs and ADs of ORCA3 and 4 were swapped. (a) Cartoon showing the domain organization of ORCA3 and 4 and the respective chimeric proteins. (b) Combinatorial overexpression of *CrMYC2a*^{D126N} with *ORCA3*, *4* or chimeric proteins *ORCA3-ORCA4* or *ORCA4-ORCA3*. These results suggest that the AD of ORCA3 improves the synergistic up-regulation of

target genes. Cells labeled with different letters represent statistically significant differences ($P<0.05$, ANOVA with Tukey's correction for multiple comparisons).

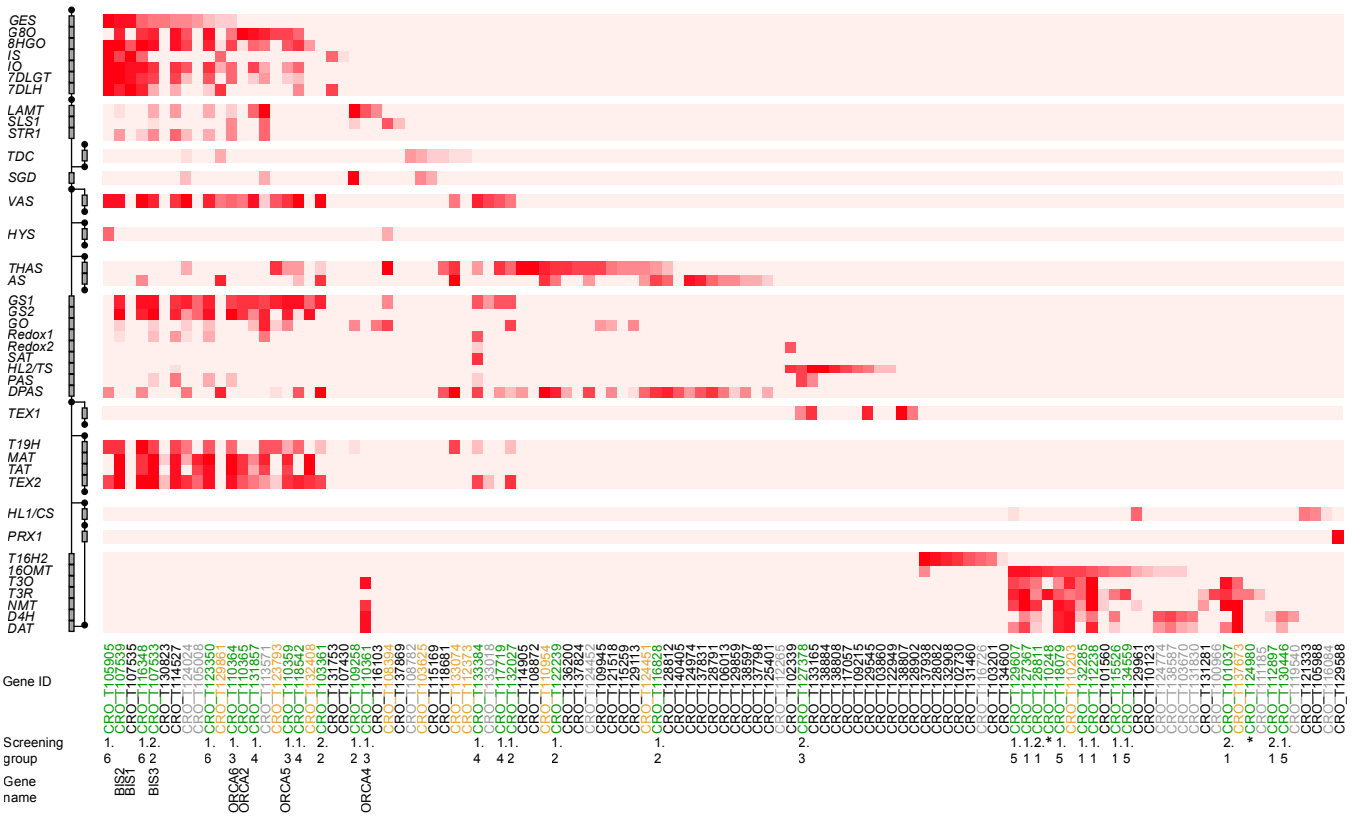
(a)


Epidermis
2
Idioblast
3
Idioblast
3
Laticifer
3
IPAP cells
1
1
8HGO
G8O
GES
geranyl-PP
IS
IO
7DLGT
7DLH
loganic acid
NPF2.4/5/6
2
LAMT
loganic acid
tryptophan
TDC
SLS
alstonine
AS
tryptamine
STR
NPF2.9
THAS
tetrahydroalstonine
SGD
HYS
ajmalicine
serpentine
vitrosamine
VAS
strictosidine aglycone
TPT2
PRX1
vinblastine
vincristine
isositsirikine
GS1/2
geissoschizine
perivine
GO
akuammicine
Redox1
Redox2
SAT
catharanthine
PAS
CS
DPAS
TS
desacetoxyvindoline/ desacetoxyvindorosine
NMT
T3R
T3O
16OMT
T16H1/2
tabersonine
TEX2
TEX1
lochnericine
T19H
HL3/4
(+)-vincadifformine
V19H
(+)-echito-venine
MAT
(+)-minovincinine
19-hydroxytabersonine
TAT
19-acetyltabersonine
hörhammericine
19-acetylhörhammericine
3
vindoline / vindorosine
DAT
D4H
desacetoxyvindoline/ desacetoxyvindorosine
Derivatives possibly related to overexpression in flower petals
19-acetyltabersonine hörhammericine lochnericine
(+)-vincadifformine
T16H1/2
16-hydroxylated compounds
glycosylation
16OMT
glycosylated compounds
16-methoxyhörhammericine

(b)


stemmadenine pathway
root specific MIAs
strictosidine pathway
vindoline pathway
iridoid pathway
Redox1
Redox2
SAT
PAS
DPAS
GO
GS2
GS1
AS
THAS1
HYS
VAS
SGD
STR1
TDC
SLS1
LAMT
7DLH
7DLGT
IO
IS
G8O
8HGO
GES
HL2 /TS
TEX1
T19H
MAT
TAT
TEX2
HL1/ CS
T16H2
16OMT
T3O
T3R
NMT
D4H
DAT
PRX1

(a)



(b)

Bulk overexpression in flower petals

Screening infiltration 1

Group 1.1	Group 1.2	Group 1.3	Group 1.4	Group 1.5	Group 1.6
CRO_T112238	CRO_T132027	CRO_T110359	CRO_T118542	CRO_T129607	CRO_T116348
CRO_T132285	CRO_T109258	CRO_T110361	CRO_T117719	CRO_T134559	CRO_T123350
CRO_T127367	CRO_T116828	CRO_T110364	CRO_T133384	CRO_T130446	CRO_T105905
CRO_T115526	CRO_T122239		CRO_T131857	CRO_T118079	

Screening infiltration 2

Group 2.1	Group 2.2	Group 2.3
CRO_T112897	CRO_T107533	*CRO_T110868*
CRO_T128811	CRO_T103361	*CRO_T126752*
CRO_T100680		*CRO_T134504*
CRO_T101037		CRO_T127378

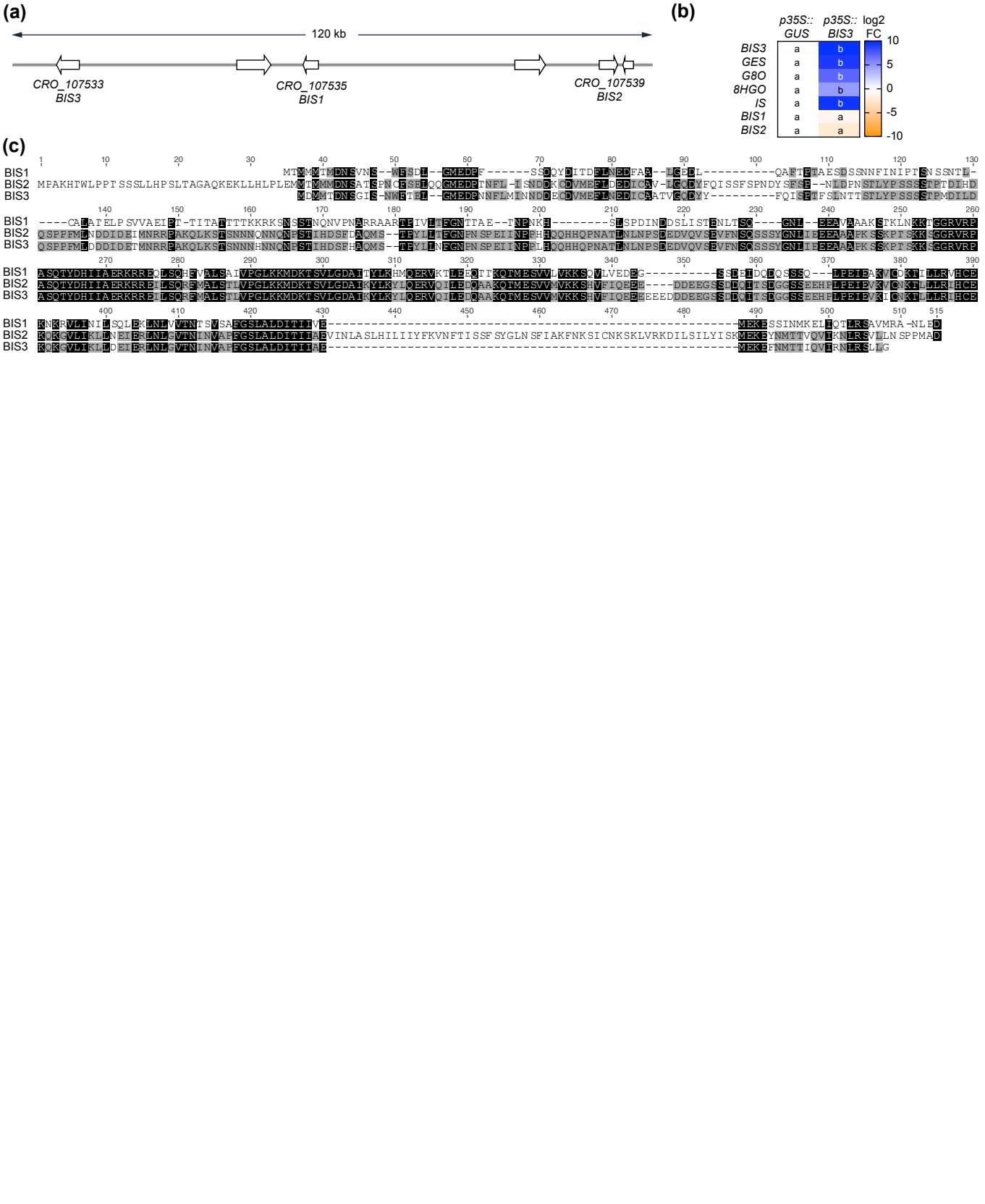
Measure overexpression level and target gene expression levels compared to *GUS* overexpression by qPCR (Figure S1)

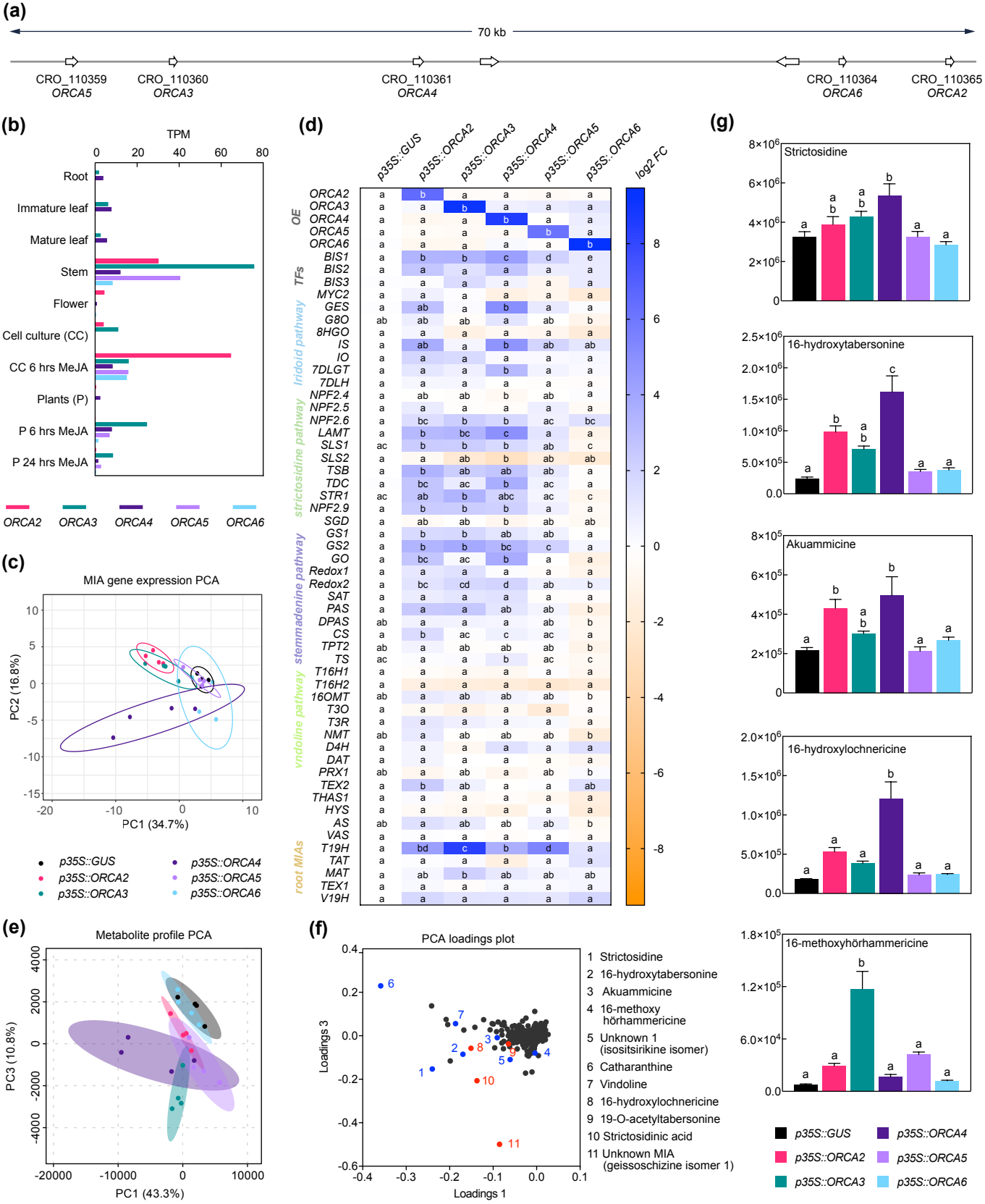
Overexpression of TF candidates of positive groups

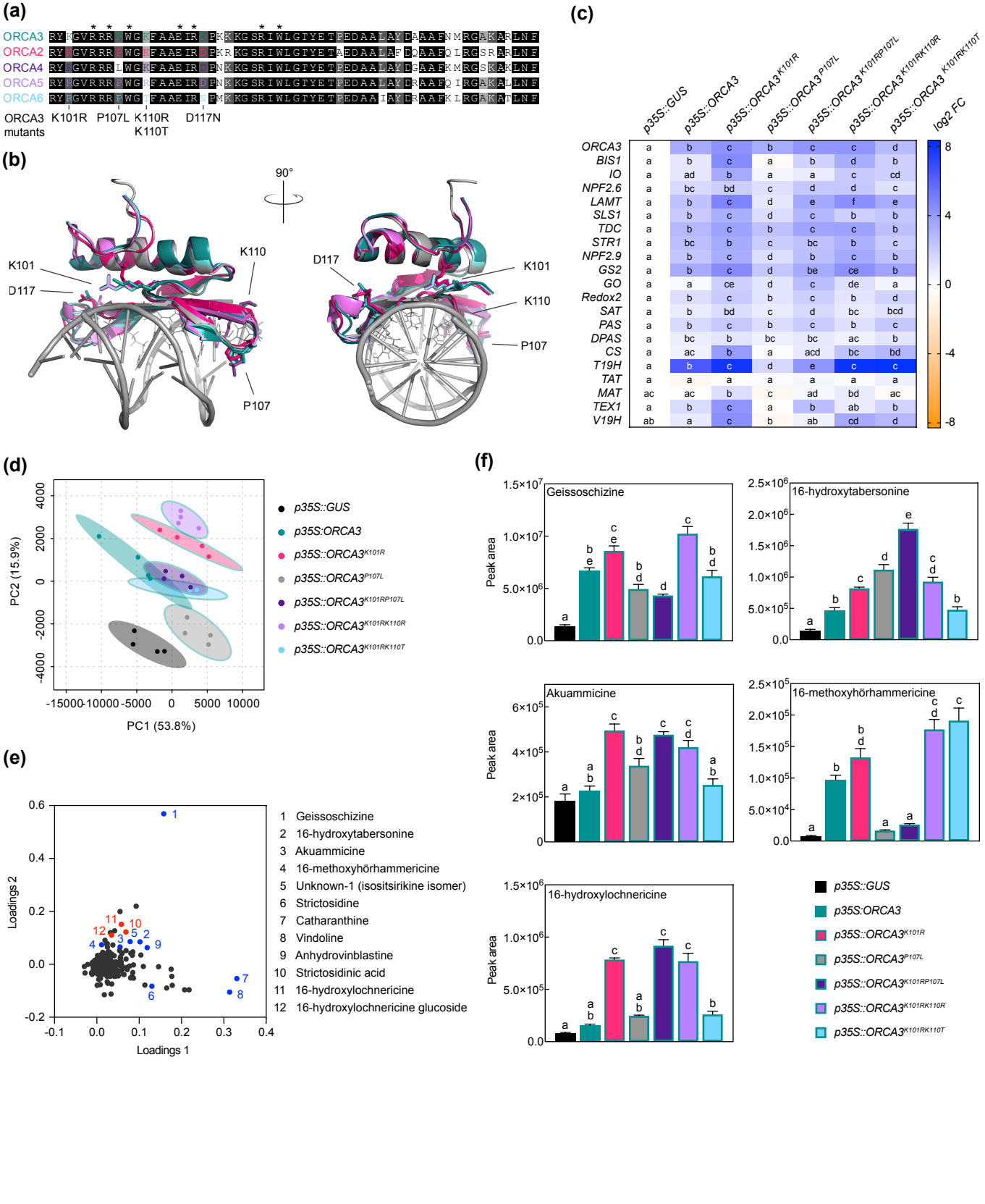
Separate infiltration 1	CRO_T112238	CRO_T132285	CRO_T127367	CRO_T115526	CRO_T123350	CRO_T105905
Separate infiltration 2	CRO_T132027	CRO_T109258	CRO_T116828	CRO_T122239		
Separate infiltration 3	CRO_T112897	CRO_T128811	*CRO_T100680*	CRO_T101037	CRO_T107533†	CRO_T103361
Separate infiltration 4*	CRO_T110248*					
Separate infiltration 5*	CRO_T124980*					

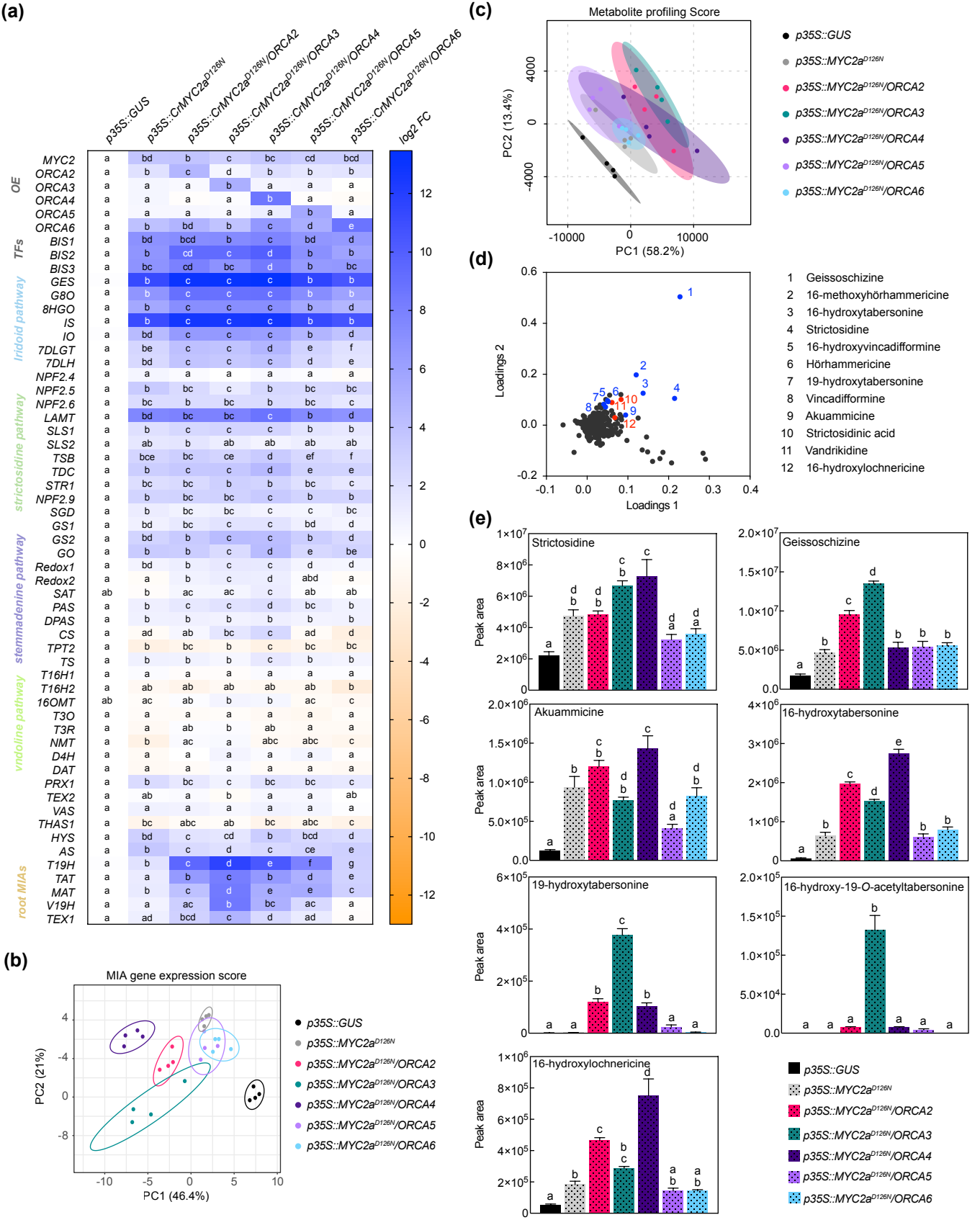
Measure overexpression level and target gene expression levels compared to *GUS* overexpression by qPCR (Figure S2)

Additional promoter transactivation assays TFs with subtle effect in flower petal overexpression (Figure S2f)


(a)
120 kb
CRO_107533
BIS3
CRO_107535
BIS1
CRO_107539
BIS2
(b)
p35S::GUS
p35S::BIS3
log2 FC
BIS3 a b
GES a b
G8O a b
8HGO a b
IS a b
BIS1 a a
BIS2 a a
10
5
0
-5
-10
(c)
BIS1
BIS2
BIS3


(a)
70 kb
CRO_110359
ORCA5
CRO_110360
ORCA3
CRO_110361
ORCA4
CRO_110364
ORCA6
CRO_110365
ORCA2
(b)
TPM
0
20
40
60
80
Root
Immature leaf
Mature leaf
Stem
Flower
Cell culture (CC)
CC 6 hrs MeJA
Plants (P)
P 6 hrs MeJA
P 24 hrs MeJA
ORCA2
ORCA3
ORCA4
ORCA5
ORCA6
(c)
MIA gene expression PCA
PC2 (16.8%)
PC1 (34.7%)
p35S::GUS
p35S::ORCA2
p35S::ORCA3
p35S::ORCA4
p35S::ORCA5
p35S::ORCA6
(d)
p35S::GUS
p35S::ORCA2
p35S::ORCA3
p35S::ORCA4
p35S::ORCA5
p35S::ORCA6
log2 FC
OE
TFs
Iridoid pathway
strictosidine pathway
stemmadenine pathway
vindoline pathway
root MIAs
ORCA2
ORCA3
ORCA4
ORCA5
ORCA6
BIS1
BIS2
BIS3
MYC2
GES
G8O
8HGO
IS
IO
7DLGT
7DLH
NPF2.4
NPF2.5
NPF2.6
LAMT
SLS1
SLS2
TSB
TDC
STR1
NPF2.9
SGD
GS1
GS2
GO
Redox1
Redox2
SAT
PAS
DPAS
CS
TPT2
TS
T16H1
T16H2
16OMT
T3O
T3R
NMT
D4H
DAT
PRX1
TEX2
THAS1
HYS
AS
VAS
T19H
TAT
MAT
TEX1
V19H
(e)
Metabolite profile PCA
PC3 (10.8%)
PC1 (43.3%)
(f)
PCA loadings plot
Loadings 3
Loadings 1
1 Strictosidine
2 16-hydroxytabersonine
3 Akuammicine
4 16-methoxy hörhammericine
5 Unknown 1 (isositsirikine isomer)
6 Catharanthine
7 Vindoline
8 16-hydroxylochnericine
9 19-O-acetyltabersonine
10 Strictosidinic acid
11 Unknown MIA (geissoschizine isomer 1)
(g)
Strictosidine
16-hydroxytabersonine
Akuammicine
16-hydroxylochnericine
16-methoxyhörhammericine
p35S::GUS
p35S::ORCA2
p35S::ORCA3
p35S::ORCA4
p35S::ORCA5
p35S::ORCA6

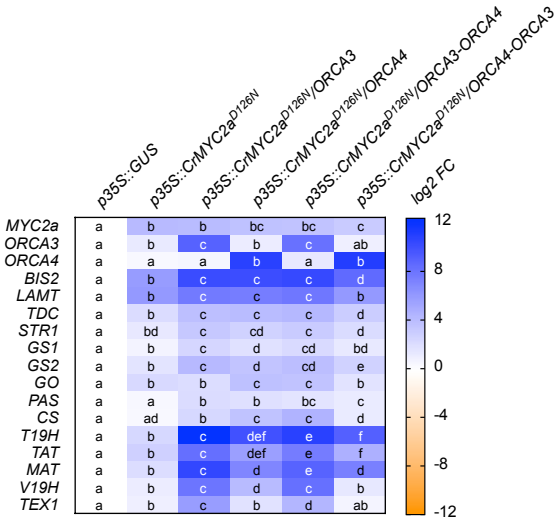

(a)
ORCA3
ORCA2
ORCA4
ORCA5
ORCA6
ORCA3 mutants
K101R
P107L
K110R
K110T
D117N
(b)
90°
K101
D117
K110
P107
(c)
p35S::GUS
p35S::ORCA3
p35S::ORCA3K101R
p35S::ORCA3P107L
p35S::ORCA3K101RP107L
p35S::ORCA3K101RK110R
p35S::ORCA3K101RK110T
log2 FC
ORCA3
BIS1
IO
NPF2.6
LAMT
SLS1
TDC
STR1
NPF2.9
GS2
GO
Redox2
SAT
PAS
DPAS
CS
T19H
TAT
MAT
TEX1
V19H
(d)
PC2 (15.9%)
PC1 (53.8%)
p35S::GUS
p35S:ORCA3
p35S::ORCA3K101R
p35S::ORCA3P107L
p35S::ORCA3K101RP107L
p35S::ORCA3K101RK110R
p35S::ORCA3K101RK110T
(e)
Loadings 2
Loadings 1
1 Geissoschizine
2 16-hydroxytabersonine
3 Akuammicine
4 16-methoxyhörhammericine
5 Unknown-1 (isositsirikine isomer)
6 Strictosidine
7 Catharanthine
8 Vindoline
9 Anhydrovinblastine
10 Strictosidinic acid
11 16-hydroxylochnericine
12 16-hydroxylochnericine glucoside
(f)
Geissoschizine
16-hydroxytabersonine
Akuammicine
16-methoxyhörhammericine
16-hydroxylochnericine
Peak area
p35S::GUS
p35S:ORCA3
p35S::ORCA3K101R
p35S::ORCA3P107L
p35S::ORCA3K101RP107L
p35S::ORCA3K101RK110R
p35S::ORCA3K101RK110T


(a)
p35S::GUS
p35S::CrMYC2aD126N
p35S::CrMYC2aD126N/ORCA2
p35S::CrMYC2aD126N/ORCA3
p35S::CrMYC2aD126N/ORCA4
p35S::CrMYC2aD126N/ORCA5
p35S::CrMYC2aD126N/ORCA6
log2 FC
OE
TFs
Iridoid pathway
strictosidine pathway
stemmadenine pathway
vindoline pathway
root MIAs
MYC2 a bd b c bc cd bcd
ORCA2 a b c d b b b
ORCA3 a a a b a a a
ORCA4 a a a a b a a
ORCA5 a a a a a b a
ORCA6 a b bd b c d e
BIS1 a bd bcd b c bd d
BIS2 a b cd c d b b
BIS3 a bc cd bc d b bc
GES a b c c c b b
G8O a b c c c b b
8HGO a b c c c b d
IS a b c c c b b
IO a bd c c c b d
7DLGT a be c c d e f
7DLH a bd c c c d e
NPF2.4 a a a a a a a
NPF2.5 a b bc c b bc bc
NPF2.6 a bc b bc b bc c
LAMT a bd bc bc c b d
SLS1 a b c c c b b
SLS2 a ab b ab ab ab ab
TSB a bce bc ce d ef f
TDC a b c c d e e
STR1 a bc bc b b bc c
NPF2.9 a b bc bc c b b
SGD a b bc bc c c bc
GS1 a bd bc c c bd d
GS2 a bd b c c b d
GO a b b c d e be
Redox1 a bd b c c d d
Redox2 a a b c d abd a
SAT ab b ac ac c ab ab
PAS a b c c d bc b
DPAS a b b b b b b
CS a ad ab bc c ad d
TPT2 a b bc b c bc bc
TS a b b b c b b
T16H1 a a a a a a a
T16H2 a ab ab ab ab ab b
16OMT ab ac ab b b ac c
T3O a a a a a a a
T3R a a ab b ab a a
NMT a b ac a abc abc c
D4H a a a a a a a
DAT a a a a a a a
PRX1 a b bc c bc bc c
TEX2 a ab a b a a ab
VAS a a a a a a a
THAS1 a bc abc ab bc abc c
HYS a bd c cd b bcd d
AS a b c d c ce e
T19H a b c d e f g
TAT a a b c b d e
MAT a b c d e e c
V19H a a ac b bc ac a
TEX1 a ad bcd c d ad a
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
(b)
MIA gene expression score
PC2 (21%)
PC1 (46.4%)
4
-4
0
-8
-10
-5
0
5
10
p35S::GUS
p35S::MYC2aD126N
p35S::MYC2aD126N/ORCA2
p35S::MYC2aD126N/ORCA3
p35S::MYC2aD126N/ORCA4
p35S::MYC2aD126N/ORCA5
p35S::MYC2aD126N/ORCA6
(c)
Metabolite profiling Score
PC2 (13.4%)
PC1 (58.2%)
4000
0
-4000
-10000
0
10000
p35S::GUS
p35S::MYC2aD126N
p35S::MYC2aD126N/ORCA2
p35S::MYC2aD126N/ORCA3
p35S::MYC2aD126N/ORCA4
p35S::MYC2aD126N/ORCA5
p35S::MYC2aD126N/ORCA6
(d)
Loadings 2
Loadings 1
0.6
0.4
0.2
0.0
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
1 Geissoschizine
2 16-methoxyhörhammericine
3 16-hydroxytabersonine
4 Strictosidine
5 16-hydroxyvincadifformine
6 Hörhammericine
7 19-hydroxytabersonine
8 Vincadifformine
9 Akuammicine
10 Strictosidinic acid
11 Vandrikidine
12 16-hydroxylochnericine
(e)
Strictosidine
Geissoschizine
Akuammicine
16-hydroxytabersonine
19-hydroxytabersonine
16-hydroxy-19-O-acetyltabersonine
16-hydroxylochnericine
Peak area
p35S::GUS
p35S::MYC2aD126N
p35S::MYC2aD126N/ORCA2
p35S::MYC2aD126N/ORCA3
p35S::MYC2aD126N/ORCA4
p35S::MYC2aD126N/ORCA5
p35S::MYC2aD126N/ORCA6

(a)

ORCA3
AD
DBD
ORCA4
AD
DBD
ORCA3-ORCA4
AD
DBD
ORCA4-ORCA3
AD
DBD

(b)


p35S::GUS
p35S::CrMYC2a^D126N
p35S::CrMYC2a^D126N/ORCA3
p35S::CrMYC2a^D126N/ORCA4
p35S::CrMYC2a^D126N/ORCA3-ORCA4
p35S::CrMYC2a^D126N/ORCA4-ORCA3
log2 FC
MYC2a a b b bc bc c
ORCA3 a b c b c ab
ORCA4 a a a b a b
BIS2 a b c c c d
LAMT a b c c c b
TDC a b c c c d
STR1 a bd c cd c d
GS1 a b c d cd bd
GS2 a b c d cd e
GO a b b c c b
PAS a a b b bc c
CS a ad b c c d
T19H a b c def e f
TAT a b c def e f
MAT a b c d e d
V19H a b c d c b
TEX1 a b c b d ab
12
8
4
0
-4
-8
-12