Interplay between petal identity and cell layer identity in petunia flowers

Quentin Cavallini-Speisser[1,5], Emma Désert[1,5], Evelyne Duvernois-Berthet[1], Pierre Chambrier[1], Patrice Morel[1], Brice Letcher[2], Carine Rey[3], Jérémy Just[1], Suzanne Rodrigues Bento[1], Daniel Bouyer[1,*] and Marie Monniaux[1,4,*]

[1] Laboratoire de Reproduction et Développement des Plantes, Université de Lyon, ENS de Lyon, UCB Lyon 1, CNRS, INRAE, 69007 Lyon, France

[2] Laboratoire de Biologie et Modélisation de la Cellule, École Normale Supérieure de Lyon, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, Lyon, France

[3] CIRI, Centre International de Recherche en Infectiologie, Univ Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, F-69007, Lyon, France

[4] Current address: Univ. Lille, CNRS, UMR 8198, Evo-Eco-Paleo, F-59000 Lille, France

[5] These authors contributed equally

* Correspondence: marie.monniaux@univ-lille.fr, daniel.bouyer@ens-lyon.fr

ABSTRACT

In flowering plants, floral organ identity is specified by the combinatorial action of homeotic genes. While the role of these genes in the early specification of organ identity is well established, their late function throughout floral organ development and in specific cell types is much less characterized. In particular, since plant organs are structured in clonally-independent cell layers, whether and how homeotic identity interacts with cell layer identity is unknown. We have previously identified cell layer-specific mutants for the petal identity gene *PhDEF* in petunia flowers, resulting in drastically different petal phenotypes whether *PhDEF* is expressed in the petal epidermis or in the mesophyll. In this study, using a combination of single-cell RNA-Seq and chromatin immunoprecipitation on *phdef* cell layer-specific mutants, we find that PhDEF regulates a different set of target genes in the petal epidermis and mesophyll, with a major regulatory action in the epidermis. We uncover a high diversity of binding profiles in PhDEF target genes, with a complex combination of layer-specific or -aspecific binding sites, and a much more prominent binding of PhDEF in the epidermis than in the mesophyll. Our study highlights that floral homeotic genes like *PhDEF* can have different regulatory actions in different cell contexts, here different cell layers, and that cell layer identity indeed influences the regulatory processes underlying homeotic identity.

INTRODUCTION

Homeotic genes, in particular *Hox* genes and *MADS-box* genes, define the identity of body segments in the animal embryo and floral organs in flowering plants respectively[1,2]. They encode master transcription factors that regulate the expression of a wide set of genes, building fully developed organs. However, the action of homeotic genes is not restricted to an early specification of organ identity, since both *Hox* genes and *MADS-box* genes continue to be expressed throughout organ development, and evidence for late functions of homeotic genes, in specific differentiating cell types, has been accumulating in the litterature[3–5]. In contrast to their role in the early specification of organ identity, how homeotic genes participate to the acquisition of specific cell fates in mature organs has been under-explored, both in animals and plants. Additionally, in flowering plants all aerial organs are organized in layers deriving from the three clonally-distinct layers (L1, L2 and L3) from the shoot apical meristem[6–9]. In particular, epidermal identity is established from the embryonic stage onwards, as evidenced by the expression of a specific set of genes[10,11]; hence floral organ identity is superimposed on a pre-established layer identity, raising the question of whether homeotic function is dependent or influenced by layer identity.

The identity of floral organs is specified by the combinatorial action of homeotic genes, as outlined in the ABC model. Established in the 90s in *Arabidopsis thaliana* and *Antirrhinum majus*[12,13], this model proposes that sepal identity is specified by A-class genes, petal identity by a combination of A- and B-class genes, stamen identity by a combination of B- and C-class genes, and carpel identity by C-class genes. Most of those genes are part of the MADS-box gene family, which encode transcription factors that bind DNA as dimers and tetramers, with a combinatorial mode of action (e.g. B- and C-class proteins form protein complexes that define stamen identity)[14–17]. This model is widely valid for all flowering plants[1], although the A-function is not well supported[18–20]. In this study, we focus on petal identity in the model species *Petunia x hybrida* (petunia). Petals are often key players in pollinator attraction, and in the case of petunia, this is subtended by the bright pigmentation or strong scent emission from the conical cells of the limb epidermis, while the long tube observed in some petunia species limits nectar access to adapted pollinators[21,22]. In petunia, petal identity is defined by B class genes only: *PhDEF* (*Petunia x hybrida DEFICIENS*), *PhGLO1* (*GLOBOSA1*) and *PhGLO2*[20,23,24]. The encoded proteins form PhDEF/PhGLO1 and PhDEF/PhGLO2 obligate heterodimers that translocate to the nucleus, bind DNA and regulate gene transcription, as was reported for their homologs in Antirrhinum and Arabidopsis[25–27]. Importantly, these complexes activate the expression of the *PhDEF*, *PhGLO1* and *PhGLO2* genes themselves, resulting in an auto-activation loop[25,26,23]. This loop, that is conserved in all eudicots, buffers stochastic noise, leading to a sharp transition to petal identity, and ensures the maintenance of high expression level of these genes until petals are mature[28,29]. Other proteins, in

69 particular the E-class SEPALLATA MADS-box proteins[30,31], are part of higher-order complexes
70 with the B-class heterodimers to define petal identity[32–34].
71 B-class genes, after defining petal identity, also participate to their development and to the
72 differentiation of their cell types. Indeed, experiments of temporal inactivation of B-class genes in
73 *A. thaliana* have revealed that, at intermediate stages of flower development, petals easily revert to
74 a sepal-like identity[35], suggesting that expression of B-class genes needs to be maintained over the
75 course of petal development for normal petal identity specification. Similarly, the late decrease in
76 B-class gene expression in *A. majus* causes a reduction in petal size, conical cell size and scent
77 emission[36], and a reduction in scent production and emission was also recently reported in petunia
78 petals when *PhDEF* expression is repressed at anthesis[37]. This suggests that petal identity remains
79 partly uncommitted until intermediate-to-late stages of flower development, and that B-class genes
80 act to maintain this late specification of identity through the control of specific cell fates and traits.
81 In this study, we investigate the late role of the petal identity regulator PhDEF in the two cell
82 layers of the petunia petal, stemming from our previous identification of cell layer-specific *phdef*
83 mutants[38]. In petunia, petals derive from the L1 and L2 meristematic layers that will form the petal
84 epidermis and mesophyll respectively[39,38]. While the *phdef* mutant flowers display a homeotic
85 conversion of petals into sepals, the layer-specific *phdef* mutation results in two contrasted
86 phenotypes[38] (Figure 1A): flowers with a *phdef* mutant epidermis (that we named *star*) form a
87 normal tube but small, star-shaped and unpigmented limb. Meanwhile, flowers with a *phdef* mutant
88 mesophyll (that we named *wico* for *wide corolla*) have a much shorter tube than normal, but the
89 limb is normally-shaped and pigmented, although petals are often pink rather than red. These
90 flowers derive from the excision of the transposable element *dTph1* (inserted in the first exon of the
91 *PhDEF* gene and knocking-out its function) in a unique cell layer, as we previously described[38]. The
92 *star* and *wico* phenotypes show that PhDEF plays distinct roles in the two layers of the petal and
93 participates to the modular development of this organ, independently in the tube and the limb. Here,
94 to further understand how PhDEF achieves its cell layer-specific role, we identified the cell layer-
95 specific targets of PhDEF by characterizing layer-specific transcriptional profiles obtained by
96 single-cell RNA-Seq (scRNA-Seq) and layer-specific binding profiles obtained by chromatin
97 immunoprecipitation (ChIP-Seq), in wild-type (WT), *star* and *wico* petals. We found that, although
98 the WT petal epidermis and mesophyll express a similar number of genes overall, there are many
99 more genes whose expression is enriched (stronger and more specific) in the epidermis than in the
100 mesophyll. Moreover, although *PhDEF* is similarly expressed in the two layers of the petal, we
101 found that it binds to and regulates the expression of many more target genes in the epidermis than
102 in the mesophyll. The target genes of PhDEF display a combination of layer-specific and -aspecific
103 binding sites in their regulatory regions, suggesting a complex pattern of transcriptional control

among cell layers, and epidermal-specific binding sites are prominent. This study offers a novel cell layer-specific perspective on transcriptional regulation and shows that homeotic regulators such as PhDEF are indeed influenced by the pre-established cell layer identity.

RESULTS

Single-cell RNA-sequencing reveals separate clusters of epidermal and mesophyll cell types in the petunia petal

The *star* and *wico* phenotypes (Figure 1A), described in the Introduction, suggest that petal identity is specified by *PhDEF* independently in the two cell layers of the petal. The most likely mechanism for this cell layer-specific function is that PhDEF regulates a distinct set of target genes in the two cell layers. We first aimed to identify the genes that exhibited shared or specific expression in each petal layer, and for this we sequenced the transcriptome of single cells from wild-type (WT) petunia petals (at anthesis, Figure 1A). To this end, the petals were isolated by carefully removing the stamen filaments that are fused to the base of the tube, and we performed a 5h-long protoplast isolation of the whole corolla (tube and limb) by cell wall-digesting enzymes (Figure 1B). A visual inspection of protoplasts under the microscope allowed to evaluate the proportion of epidermal (pigmented with pink anthocyanins) and mesophyll (non pigmented and/or chloroplastic) cells recovered, which was not significantly different to the proportion of epidermal and mesophyll cells observed in petal cross-sections, indicating that protoplasting allowed to efficiently recover cells from the two petal layers (Figure S1). The transcriptomes from single cells were subjected to sequencing with the 10X Genomics Chromium technology for library preparation and Illumina short-read sequencing (Figure S2), and we obtained the transcriptome from 11,632 cells from WT mature petals in two biological replicates (Pearson's correlation coefficient r = 0.96 between pseudo-bulk datasets of the two replicates). We mapped the reads on the *Petunia axillaris* genome and transcriptome[38,40] and we applied an adapted Seurat[41] pipeline to obtain the Uniform Manifold Approximation and Projection (UMAP) plot of the transcriptomic identity of the cells (Figure 1C). We applied MultiK[42] for unsupervised clustering and chose a low clustering resolution yielding 8 distinct clusters.

The B-class petal identity genes *PhDEF*, *PhGLO1* and *PhGLO2* were broadly expressed in all clusters (Figure 1D), while the stamen and carpel identity genes *PhTM6*, *FBP6* and *pMADS3*[24,43] were hardly detected, indicating no contamination by stamen tissue. In order to assign identity to each cluster, we looked at known marker genes from the literature, identified their putative homologs in petunia (Table S1) and observed their expression in our clusters (Figure 1D, Figure S3). We also explored Gene Ontology (GO) term enrichment for the best marker genes of each cluster (Figure S4, Table S2) and we determined the limb or tube origin of each cluster by RT-qPCR on a chosen set of cluster markers (Figure S5). The specific expression of the sucrose transporter SWEET11, and several UmamiT amino acid transporters and sulfate transporters, clearly marked cluster 5 as a group of vascular cells[44] (Figure S3). Similarly, the strong expression of several anthocyanin biosynthesis genes allowed the identification of cluster 6 as pigmented cells, hence

cells from the upper limb epidermis[40]. Other known epidermal markers, such as ketoacyl-coA synthases (KCS, wax synthesis enzymes)[45] or members of the Homeodomain Glabrous (HDG) family[46], identified clusters 0 and 4 as two other epidermal clusters. Cluster 0 likely groups cells from the lower and upper limb epidermis; indeed, these cells faintly express pigmentation genes (Figure 1D); tested marker genes from this cluster are strongly expressed in limb tissue (Figure S5A) ; and *in situ* hybridization for the cluster marker *IAA14* clearly showed its expression in both upper and lower limb epidermis (Figure S5C). Consequently, this suggests that cells in cluster 6 represent only a fraction of the upper limb epidermal cells, *i.e.* the ones most strongly expressing pigmentation genes, perhaps the most distal cells from the limb that are the latest to differentiate. The highly specific expression of *TERPENE SYNTHASE1 (PhTPS1)*, a tube-specific gene involved in volatile emission inside the flower[47], suggested that cluster 4 contains cells from the tube epidermis (Figure S5A, B). The strong and specific expression of several histone genes and S-phase genes (Figure S3C, D) identified the smallest cluster (7) as a cell-cycle state cluster (*i.e.* dividing cells, independent of a specific cell identity). Finally, we assigned by default a mesophyll identity to the remaining clusters (1, 2 and 3), since they expressed few highly-specific genes, strongly associated with photosynthetic activity for cluster 3 and with water transport for cluster 2 (Figure S4, Table S2), and these clusters contained cells from either tube or limb origin (Figure S5A). Taken together, although there remains some uncertainty about the precise spatial location of each cluster, we are confident about their belonging to either the epidermis or the mesophyll layer.

Protoplasting is known to induce strong transcriptomic changes that could impact the clustering and the assignment of cell identities[48]. To estimate the effect of protoplasting on gene expression in our samples, we performed bulk RNA-Seq on protoplasted vs. crude WT petal tissue, and we observed a high correlation between the gene expression levels in the two conditions (r = 0.84). Still, a high number of differentially expressed genes (DEGs) were identified (5,571 and 5,891 genes respectively upregulated and downregulated by the protoplasting process, Table S3). We explored the effect of removing DEGs on the clustering from the scRNA-Seq data (Figure S6): the UMAP plots exhibited a high degree of similarity in their shape irrespective of the removal of the DEGs, and the same clusters were recovered except for one cluster that disappeared in the stringent case where we removed all DEGs with |log2FC|>1. This cluster corresponds to cluster 6, containing the strongly pigmented cells from the upper limb epidermis. Indeed, several pigmentation genes (21 out of the 42 manually-annotated genes involved in anthocyanin biosynthesis and its regulation[38]) were differentially expressed after the protoplasting process (Table S3), which might be linked with the role of anthocyanins in the response to abiotic stress[49], and logically affects the definition of the cluster of strongly pigmented cells. Since this cell type is

important for the comprehension of the cell layer-specific action of PhDEF, we decided to keep all genes and all cells in our analysis.

In summary, our assignment of cluster identities revealed that the three mesophyll clusters are not clearly distinct in the UMAP space and can be grouped together, in contrast to the three epidermal clusters that have distinct transcriptomic profiles. Importantly, there is a clear separation of epidermal and mesophyll cell types in the UMAP space, as previously observed in scRNA-Seq data from the petal of mature tobacco flowers[50], demonstrating that these layers have distinct transcriptomic signatures overall. Since our main interest was in the layer-specific functions of PhDEF, we next merged the three epidermal clusters together, and the three mesophyll clusters together, setting apart the vasculature.

The petal epidermis expresses more specific genes than the petal mesophyll

We first explored the difference in gene expression between WT petal layers. We computed gene differential expression between the epidermis and the mesophyll, and we examined the layer-enriched genes, *i.e.* with a higher expression level in one layer and with a high specificity of expression in one layer (higher percentage of cells expressing the gene in this layer). We noticed that there were consistently more genes whose expression is specifically enriched in the epidermis than in the mesophyll (Figure 1E). For instance, with log2FoldChange > 0.75 and layer-specificity of expression > 10 % (see Methods), there were 695 genes enriched in the epidermis compared to 151 genes enriched in the mesophyll (Table S4), and this asymmetry is observed at varying log2FoldChange thresholds (Figure 1E). Some of the GO terms enriched in layer-enriched genes were consistent with the function of these layers (Figure S7, Table S2): the epidermis-enriched genes were associated with GO terms related to the metabolism of surface lipids or the phenylpropanoid pathway (an extensive network that leads to the production of lignin, anthocyanins and flavonols), while the mesophyll-enriched genes were associated with several GO terms related to photosynthesis. Overall, this initial analysis of layer-enriched genes showed that there are more epidermis-enriched genes than mesophyll-enriched genes in the WT petal.

PhDEF regulates a different set of genes in the two petal layers

To explore how petal identity is defined by PhDEF in a layer-specific fashion, we generated scRNA-Seq data from 3,875 and 3,737 cells from *star* and *wico* mature petals respectively. Due to the difficulty in collecting high numbers of flowers from these genetic chimeras, we were only able to produce a single biologial replicate for each. We integrated the WT, *star* and *wico* scRNA-Seq data with Harmony[51], allowing the identification of epidermal and mesophyll clusters in all samples (Figure 2A, B). The pseudo-bulk transcriptomes displayed a higher correlation between WT

replicates (Pearson's correlation coefficient, r = 0.96) than with the other genotypes (r = 0.9 between WT and *wico*, r = 0.93 between WT and *star*). The levels of expression of the main petal identity genes *PhDEF* and *PhGLO1* were consistent with expected (Figure 2C, D): in WT petals, both genes were expressed evenly in all cell types, while in *star* petals we observed high expression in the mesophyll clusters and low expression in the epidermal clusters, and the converse was observed in *wico* petals. This is particularly apparent when looking at *PhGLO1* expression, that is higher than *PhDEF* expression at this stage.

When comparing the *star* and *wico* petals with WT petals, we did not uncover new clusters, which one might have expected from altered cell identity, but we observed changes in the proportions of cells in each cluster (Figure 2B). First, in *star*, a marginal number of upper limb epidermal cells were recovered, which corresponds to the small secondary L1-revertant sectors that are frequently observed in *star* petals[38] (Figure S8) and that were thus removed from the dataset. Apart from that, two major epidermal clusters (tube epidermis and limb epidermis) were strikingly depleted in cell numbers in *star* as compared to their WT counterparts. We interpret this as the result of two processes: (1) the lower proportion of limb tissue, that is mostly constituted of epidermal cells[52,38], in *star* petals than in WT petals; (2) the alteration of epidermal cell identity by the *phdef* mutation, which causes epidermal cells to cluster with mesophyll cells. Indeed, we observed that the epidermal and mesophyll clusters appeared less distinct in *star* than in WT on the UMAP (Figure 2A), and we found that the *star* epidermis and mesophyll transcriptomes were overall very similar (r = 0.97 when comparing their pseudo-bulk transcriptomes, while the WT and *wico* epidermis and mesophyll transcriptomes were more distinct, r = 0.94 and 0.93, respectively) (Figure 2E). Moreover, the *star* epidermis transcriptome exhibited a greater resemblance to the WT mesophyll than to the WT epidermis transcriptome (r = 0.94 and 0.88, respectively, Figure 2E), suggesting that the *PhDEF* mutation in the *star* epidermis caused the cells to adopt a mesophyll-like identity. This is reminiscent of our previous observation that *star* epidermal cells, appearing as large domed cells, had a morphology intermediate between the small conical cells of the WT petal epidermis and the large flat puzzle cells of the *phdef-151* sepal epidermis, also suggestive of a mixed identity[38].

In *wico* petals, the main epidermal clusters were recovered in proportions higher than in *star* petals, but still much lower than in WT petals (Figure 2B). Again, an overall depletion in epidermal cells could result from the lower size of the limb in *wico* petals as compared to WT, and from a slight alteration of epidermal identity[38], although this layer expresses *PhDEF* in *wico* flowers. In contrast to *star*, the *wico* epidermis and mesophyll transcriptomes remained quite distinct to each other ($r^2 = 0.93$), indicating that *PhDEF* mutation in the mesophyll did not cause the cells to adopt an epidermal-like identity. The *wico* mesophyll, that does not express *PhDEF*, had a transcriptome more similar to the WT mesophyll (r = 0.92) than to the WT epidermis (r = 0.84) but the correlation

remained low. Surprisingly, the small cluster of strongly pigmented limb epidermal cells, which has a very clear transcriptional signature in WT petals (cluster 6 in Figure 1C), was not found in *wico* flowers. We speculate that this is because the *wico* flowers have a reduced pigmentation as compared to WT petals (Figure 1A, Figure S8), which we previously found to be caused by the 6-bp insertion left by the transposon excision from the *PhDEF* locus, slightly altering the function of PhDEF in activating pigmentation[38]. Consistently, we found 13 genes, out of the 42 associated with anthocyanin biosynthesis and its regulation, differentially expressed in the *wico* epidermis as compared to the WT epidermis, and the correlation between the *wico* and WT epidermal transcriptomes remained relatively low (r = 0.9, Figure 2E), although the *wico* epidermis expresses *PhDEF*. In summary, the coarse inspection of our scRNA-Seq datasets suggests that the petal epidermis transcriptional identity is altered towards a mesophyll-like identity when *PhDEF* is mutated (*star* flowers); in contrast, the petal mesophyll transcriptional identity remains very distinct to the epidermal identity when *PhDEF* is mutated (*wico* flowers). This suggests a particularly prominent role of PhDEF in the petal epidermis to establish its distinctive transcriptional identity.

Next, we explored in details the effect of *phdef* mutation on differential gene expression in a given petal layer. Since the *star* flowers are only mutant for *PhDEF* in the epidermis, we defined PhDEF epidermal-specific targets (both direct and indirect) as DEGs in the *star* epidermis as compared to the WT epidermis. Conversely, PhDEF mesophyll-specific targets are DEGs in the *wico* mesophyll as compared to the WT mesophyll. With these definitions and manually chosen thresholds for differential expression, we found that 287 genes were targets of PhDEF in both layers, while 628 were exclusive targets in the epidermis and 364 were exclusive targets in the mesophyll (Figure 2F, Table S4). These genes were either activated (75%) or repressed (25%) by PhDEF (Figure 2F), in proportions comparable to the global effect of PhDEF on gene transcription that we previously observed from bulk RNA-Seq data (61.7% and 38.3% of genes activated and repressed by PhDEF, respectively)[38]. There was a slightly higher proportion of activated target genes in the epidermis (78.5%) than in the mesophyll (69.3%, Figure 2F). The vast majority (99.3%) of the genes commonly targeted by PhDEF in both layers, hereafter named common targets, were regulated in the same direction in the two layers, *i.e.* either activated (75.2% of genes) or repressed (24.8% of genes, Figure 2G). Overall, we conclude that the direction of regulation by PhDEF is not heavily influenced by the petal layer in which it acts. As expected, *PhDEF*, *PhGLO1* and *PhGLO2* were found among the 287 common targets, activated by PhDEF in the two petal layers (Figure 2G, Table S3). Gene Ontology (GO) terms enriched in epidermis-specific PhDEF targets were related to redox processes (glutathione metabolism, cell redox homeostasis), while those enriched in mesophyll-specific PhDEF targets were strongly associated with photosynthesis, and various different GO terms were enriched in PhDEF common targets (Figure 2H, Table S2).

In summary, our scRNA-Seq dataset from layer-specific *phdef* mutant flowers indicates that PhDEF regulates a fraction of its target genes in both petal layers (representing 31-44% of its targets) but mostly target genes specifically in the epidermis or in the mesophyll (representing 56-69% of its targets), as originally suggested by the drastically different *star* and *wico* phenotypes. We also found more epidermal-specific than mesophyll-specific targets of PhDEF, although the *PhDEF* transcript accumulates similarly in the two petal layers.

PhDEF has more direct target genes in the petal epidermis than in the petal mesophyll

The scRNA-Seq transcriptomic data confounds the effect of direct gene regulation by PhDEF, and the indirect gene regulation that results from it, which might be extremely different in terms of numbers of genes, direction and intensity of the transcriptional regulation among the pathways affected. To circumvent this limitation, we then attempted to identify PhDEF direct targets in the two petal layers. For this, we performed chromatin immunoprecipitation followed by sequencing (ChIP-Seq) in WT, *star* and *wico* petals, using a custom antibody directed against the endogenous PhDEF protein[38]. We performed the experiment on two biological replicates, on flowers at stage 8 when the tube is about half its final size[38]. After sequencing and mapping the reads on the *P. axillaris* genome[40], we called the peaks for each replicate separately, then computed the Irreproducible Discovery Rate (IDR) to identify reproducible peaks and removed peaks detected in the input chromatin, which resulted in 3,134, 4,276 and 2,216 peaks for WT, *wico* and *star* respectively (Table S6). Out of these peaks, 2,017, 2,771 and 1,136 were associated with genes, for WT, *wico* and *star* respectively (Table S6, Table S7). Most of the peaks were located in promoter regions (defined as 6 kb upstream the start codon, 34.5% for WT, as compared to a random 14.9% occupancy, $p = 2.06e^{-153}$, hypergeometric test) or terminator regions (defined as 6 kb downstream the stop codon, 20.1% for WT, $p = 6.44e^{-13}$) of genes, with a small fraction found in gene bodies (9.8% on average, non-significantly enriched as compared to a random 9.8% occupancy, $p = 0.51$) (Figure 3A, Table S6). These proportions were similar in WT and *wico*, but *star* displayed a slightly lower proportion of peaks located in the promoter regions (24.5% for *star* vs. 34.5% and 36.2% for WT and *wico*, respectively, Table S6), suggesting that binding of PhDEF in promoter regions is generally stronger in the epidermis than in the mesophyll. Most genes (91.8% on average) were associated with a single ChIP peak, with the rest displaying between 2 and 4 binding sites for PhDEF (Table S6). Out of the genes bound by PhDEF in WT, we found a substantial proportion (36.4%) also differentially expressed in the *phdef-151* mutant[38], which is consistent with the proportions that have been reported from previous ChIP-Seq assays on MADS-box proteins[35,53–55]. The most-enriched motif under the gene-associated ChIP peaks corresponded to a CArG box, the typical MADS-box binding site (Figure 3B, Figure S9), and we could detect clear binding of

PhDEF on its own promoter, and on the promoters from the other B-class genes *PhGLO1* and *PhGLO2* (Figure 3D-F). These results are consistent with the expected binding specificity of PhDEF, and its function in defining and maintaining petal identity through the PhDEG/PhGLO1/PhGLO2 auto-regulatory loop.

We observed that PhDEF bound to similar regions of the genome in WT, *wico* and *star*, with a generally higher read coverage in WT and *wico* samples than in *star* samples (Figure 3C, Table S5). This indicates that binding of PhDEF to the genome is stronger in the epidermis than in the mesophyll, but overall takes place in the same regions of the genome. To further explore differential binding of PhDEF across the petal cell layers, we examined the intersection of gene-associated peaks (including promoter, gene body and terminator regions) between WT, *wico* and *star* flowers (Figure 3G, Table S7). Strikingly, and consistently with the high number of peaks detected in *wico*, there were many more epidermal-specific peaks (*i.e.* found in WT+*wico* or *wico* alone, 2,042 peaks in 1,896 genes) than mesophyll-specific peaks (*i.e.* found in WT+*star* or *star* alone, 407 peaks in 398 genes). A large number of peaks (1,176 in 1,110 genes) were also shared between layers (*i.e.* found in WT, *star+wico* or *WT+star+wico*). When shared, most peaks were higher in *wico* than in WT than in *star* (Figure 3A, C-F, Figure 4B-H), supporting again that the interaction between PhDEF and its binding sites is stronger in the epidermis than in the mesophyll. Hence for most peaks, PhDEF binding in the WT sample can be interpreted as the average of strong PhDEF binding in the epidermis and weak or absent PhDEF binding in the mesophyll, explaining why some peaks are detected in *wico* but do not pass the IDR threshold in WT. The 317 peaks uniquely found in *star* were also weakly bound in the epidermis, and the 1,147 peaks uniquely found in *wico* were also weakly bound in the mesophyll (Figure S10), showing that differential binding between layers is a quantitative rather than qualitative process, but that its relative enrichment in the epidermis or in the mesophyll differs between genomic regions. The intersection of PhDEF binding sites with the layer-specific target genes of PhDEF that we identified by scRNA-Seq, as differentially expressed in a *phdef* mutant layer, was not very high (Figure S11). For instance, only 86 genes out of the 628 ones differentially expressed in the *star* epidermis (*i.e.* 13.7%) also had a PhDEF binding site in this layer. Unexpectedly, a similar proportion of genes (12.3%) were found to be differentially expressed in the *wico* mesophyll and bound by PhDEF preferentially in the epidermis, a situation that should not be observed if these genes were direct targets of PhDEF, which suggests that most of the DEGs examined here are actually indirect targets of PhDEF.

When examining the motifs enriched under gene-associated ChIP-Seq peaks for WT, *wico* and *star*, we found that all samples were enriched in CArG-boxes (Figure 3B, Figure S9), which likely reflects PhDEF direct binding but might also reflect other MADS-box proteins binding, together with PhDEF as dimers or tetramers. Indeed, PhDEF was shown to interact with the B-class

proteins PhGLO1 and PhGLO2[23], and with the E-class proteins FBP2 and FBP5 (orthologous to Arabidopsis SEPALLATA3 (SEP3) and SEP1/2, respectively)[33,34]. We examined the cell layer-specific expression of these genes in our WT scRNA-Seq dataset and found that the B-class genes and *FBP2* were overall homogeneously expressed between the two petal layers, while *FBP5* and its paralog *pMADS12* were slightly more expressed in the epidermis, and other SEP-like genes[34] were hardly detected (Figure S12). It is therefore possible that different floral quartets are more or less abundant in the epidermis and in the mesophyll, contributing to the cell layer-specific action of PhDEF. We also found enrichment for TCP (TEOSINTE BRANCHED 1 / CYCLOIDEA / PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR1) binding sites in WT and *wico* gene-associated peaks, but not in *star* (Figure 3B, Figure S9). TCP transcription factors are known to be partners of MADS-box proteins[56,57], and also to regulate the same genes as the ones regulated by MADS-box proteins[58], explaining why their motifs are occasionally found in co-occurrence under ChIP peaks[55,59]. The fact that this motif is not found enriched under *star* peaks could be due to two reasons: either the number of peaks in *star* was too low to actually detect such enrichment; or TCP proteins could be partners of PhDEF or coregulators of PhDEF target genes only in the petal epidermis. We examined the expression of petunia *TCP* genes[60] in our WT petal scRNA-Seq dataset and found overall low expression and no clear enrichment in the epidermis, except slightly for the gene *PaTCP4a* (Figure S12). Interestingly, it was reported that the expression of *PaTCP4a* increases during flower development in *Petunia axillaris*, and was higher in flowers with large petals than in flowers with small petals[60], making it an interesting candidate for an epidermal-enriched regulator that might dimerize with PhDEF or participate to the regulation of its target genes in the epidermis only. We did not detect a significantly enriched distance between CArG boxes and TCP binding sites, which would have suggested direct interaction and formation of a regulatory complex.

Different PhDEF binding profiles underlie layer-specific target gene expression

Combining our scRNA-Seq and ChIP-Seq assays on WT, *star* and *wico* flowers, we could explore further how different PhDEF binding profiles are associated with layer-specific or -aspecific expression of its target genes. This revealed a multiplicity of binding profiles, and the naive assumption that layer-specific peaks would associate with layer-specific expression was not necessarily true. Some genes displayed such an expected PhDEF binding profile; for instance, the gene *FLAVONONE 3-HYDROXYLASE* (*F3H*) is specifically expressed in the petal epidermis (Figure 4A), differentially expressed upon *phdef* mutation in the epidermis only, and displays an epidermal-specific PhDEF binding site (Figure 4B), strongly suggesting that this peak drives the specific activation of *F3H* expression in the epidermis. *F3H* encodes an enzyme from the general

flavonoid pathway resulting in the biosynthesis of anthocyanins[61], the red pigments that accumulate in the petal epidermis.

In order to broadly explore the binding and regulatory control behaviour of PhDEF, we defined all possible regulation profiles (Figure 4C, Table S8), depending on the layer-specific expression of the target gene (epidermis, mesophyll or both layers), its differential expression upon *phdef* mutation (only in the epidermis, only in the mesophyll or in both layers), and the presence of layer-specific binding sites in its regulatory regions (binding sites in the epidermis only, in the mesophyll only, in both layers only, some in the epidermis and some in both layers, some in the mesophyll and some in both layers). Remarkably, this analysis revealed that similar proportions of the different binding types were observed, whatever the gene expression profile and regulation by PhDEF considered (Figure 4C). The most prominent binding profile was epidermal peaks only (accounting on average for 51.0% of bound genes), with also a large proportion of genes displaying shared peaks only (*i.e.* peaks found in the two petal layers, 25.7% of bound genes). The rarest binding events were mesophyll peaks only, epidermis + shared peaks, and mesophyll + shared peaks, accounting for 12.8%, 9.8% and 0.4% of bound genes on average, respectively. This reflects the generally low level of PhDEF binding in the *star* samples, resulting in the identification of much fewer peaks in the mesophyll than those identified in the epidermis. Overall, this coarse analysis indicates that PhDEF binding profile is largely dominated by epidermis-specific peaks, and somehow independent from the patterns of layer-specific expression and regulation of its target genes.

Inspecting individual PhDEF binding profiles for selected target genes suggested a complex layer-dependent transcriptional control. For instance, the gene encoding the key regulator of anthocyanin biosynthesis ANTHOCYANIN1 (AN1)[62,61] displayed three binding sites for PhDEF in its regulatory regions, two of them being epidermal-specific and one of them being shared between petal layers (Figure 4D); a similar profile was found for the gene encoding the auxin importer LIKE-AUX1 (LAX1), that plays a particularly prominent role in the shoot epidermis to stabilize auxin accumulation foci[63,64] (Table S8). Both *AN1* and *LAX1* are specifically expressed in the petal epidermis, and differentially expressed in the *star* epidermis. A gene encoding a WRKY transcription factor displayed a clearly shared PhDEF binding site, while being expressed in the epidermis and differentially expressed in the *star* epidermis only (Figure 4E); a similar profile was observed for the gene encoding the regulator of anthocyanin accumulation and root hair formation ANTHOCYANINLESS2 (ANL2)[65] (Table S8), belonging to the HDG transcription factor family, typically involved in the specification of epidermal identity[10,11]. Among the 3 genes found to be epidermis-enriched but differentially expressed in both *phdef* mutant layers, we found the gene *EMISSION OF BENZENOIDS II* (*EOBII*), encoding an R2R3-MYB transcription factor described

to regulate late petal maturation traits (petal opening and scent emission, among others)[66–68], and to be directly regulated by PhDEF[37]. This gene displays one epidermal-specific and one shared binding site for PhDEF (Figure 4F). Interestingly, among the only 2 other genes in the same category was *PhMYB4* (Table S8), encoding another major regulator of petal scent emission and also a member of the R2R3-MYB family of transcription factors[69]. Therefore, our approach of classifying genes with their patterns of binding, layer-specific expression and regulation by PhDEF might occasionally reveal functional modules, here related to the regulation of scent emission.

Out of the few genes enriched in the petal mesophyll and differentially expressed in the *wico* petal mesophyll, the gene *AGP9* (encoding an arabinogalactan protein, typically found in plant cell walls) displays a clearly shared PhDEF binding site (Figure 4G). This suggests that PhDEF binding to *AGP9* in the epidermis might either play no role in its regulation, or repress its expression to maintain it mesophyll-specific. The gene *RGP3*, encoding an UDP-arabinose mutase, is also enriched in the mesophyll but down-regulated by PhDEF in the epidermis, and displays an epidermal PhDEF binding site (Figure 4H); therefore the expression of this gene might be repressed by PhDEF in the epidermis, resulting in mesophyll-enriched expression. Among the genes expressed in both petal layers and differentially expressed in both petal layers, we confirmed the presence of *PhDEF*, *PhGLO1* and *PhGLO2*, together with *C4Hb* encoding an enzyme from the core phenylpropanoid pathway[70] (eventually leading to the production of anthocyanins, flavonols and lignin), and *BPBT* encoding an enzyme necessary for the production of benzenoids, major compounds of the fragrance of petunia flowers[71] (Table S8). Overall, the intersection of our dataset allowed to identify high-confidence PhDEF target genes with putative roles in petal pigmentation, volatile production or petal development. Strikingly, PhDEF binding is much more prominent in the epidermis than in the mesophyll layer, but the link bewteen layer-specific binding and regulation remains to be established.

DISCUSSION

In this work, we investigated the cell layer-specific functions of the petal identity regulator PhDEF in petunia, based on the initial observation that epidermal- or mesophyll-specific expression of PhDEF yielded drastically different petal phenotypes. We found that PhDEF regulates a different set of target genes in the two petal layers, and our study uncovered an unexpectedly high complexity of binding profiles for PhDEF target genes, with epidermal-binding sites being the most prominent, often associated with shared binding sites in the two petal layers. We found no link between binding profiles and regulation profiles, suggesting that a PhDEF binding profile, even with layer-specific information, is a poor predictor of its target gene expression pattern. The reasons for this are manifold: PhDEF can act either as an activator or as a repressor of gene transcription, a given binding event might not always lead to gene regulation, or on the contrary, a binding site distant from a gene (and in particular from its promoter) might still participate in its regulation through long-distance loopings, as has been demonstrated for MADS-box proteins[17,72]. For instance, we found three PhDEF binding sites along the gene *AN1*, one of them being only bound in the epidermis. The two peaks bound in both layers might be useless in activating *AN1* expression, or on the contrary they might mediate long-range interactions with the epidermal-specific binding site, necessary for transcriptional activation of this gene only in the epidermis.

Globally, gene transcriptional regulation was consistently more prominent in the epidermis than in the mesophyll. First, we observed that the epidermis expresses more specific genes than the mesophyll, which, to our knowledge, had never been reported before. Second, we found that *phdef* mutation leads to the differential expression of more genes in the epidermis than in the mesophyll. Third, we observed that PhDEF strikingly binds to more binding sites and target genes in the epidermis than in the mesophyll. This strong effect in the epidermis might be partly due to the stage of development that we chose: petal development occurs in two partially overlapping phases, with tube growth being essentially finished when limb growth is fully occurring, continuing up to the mature stage that we have sampled for scRNA-Seq[52]. Since limb tissue is mostly composed of two stacks of epidermal tissue, with few interspersed mesophyll cells in between[38], we expect that the regulatory processes taking place in the mesophyll are limited in limb tissue. Therefore, we likely underestimated the gene regulation processes from the mesophyll by using mature petals for scRNA-Seq, in which tube development is essentially completed, but this effect is likely negligible in our ChIP-Seq assay for which petal samples were younger. However, the strong epidermal role of PhDEF might not only be artefactual. Indeed, the mesophyll might be regarded as a default identity for the petal cells, upon which the regulation of a large set of genes is appended for the specification of epidermal identity. This intuitive concept is subtended by several cell features specifically observed in the epidermis, not only in the petal but in all aerial organs, such as trichomes, stomatas,

481 glandular cells or conical cells, and the very specific functions that the epidermis can fulfil for
482 pigmentation, volatile emission, cuticle biosynthesis or defense against biotic and abiotic stress[11,79–
483 81]. In contrast, the mesophyll appears less differentiated, with cells visually quite similar and
484 fulfilling less diverse functions than the epidermis[79,82,83], although this might be biased by the low
485 accessibility of this tissue, and the reduced interest it has attracted. Several epidermal-specific genes
486 have been described over the years, in particular the early specifiers of embryonic epidermal
487 identity *MERISTEM LAYER1* (*ML1*) or *PROTODERMAL FACTOR2* (*PDF2*) from the previously
488 cited family of HDG transcription factors[84–86,10], or the genes acting in later epidermal-specific
489 functions, such as stomata[81] or trichome[87] formation, cuticle synthesis[88] or pigmentation[89]. In
490 contrast, a much smaller number of genes specific to the L2 layer have been described[86,90], and these
491 genes are all expressed post-embryonically[86], suggesting the possibility that the sub-epidermal
492 identity does not require specific regulators for its determination. Therefore, establishing epidermal
493 identity might truly require the regulation of several specific genes, starting from the canvas of a
494 default mesophyll identity. The fact that PhDEF, a homeotic transcription factor evenly expressed in
495 the two petal layers, also adopts the behaviour of binding to and regulating more genes in the
496 epidermis than in the mesophyll, demonstrates for the first time that the regulatory function of
497 homeotic factors is influenced by cell layer identity.

498 What might the physical nature of this influence be? We have two main hypotheses to
499 explain why PhDEF adopts a different binding and regulatory behaviour in the epidermis and in the
500 mesophyll: First, PhDEF might have access to different protein partners in the two petal layers, and
501 since binding to different protein partners entails changes in DNA-binding specificity, which has
502 been demonstrated for MADS-box proteins[16,91], this would lead to the regulation of different target
503 genes. Thus, different MADS-box protein complexes, or MADS-TCP complexes, might form in the
504 epidermis and in the mesophyll, and these complexes would contribute to the cell layer-specific
505 binding and regulatory action of PhDEF. Second, PhDEF might have access to different target genes
506 in the two petal layers, due to different chromatin accessibilities. Recently, single nuclei ATAC-Seq
507 (Assay for Transposase-Accessible Chromatin) assays in maize and sorghum leaves revealed that
508 epidermal and mesophyll chromatin accessibilities were indeed very different[92]. MADS-box
509 proteins have also been shown to interact with chromatin remodellers[93,94], which might participate in
510 the reinforcement of differential chromatin accessibility between layers. Apart from these two main
511 non-mutually exclusive scenarios, other processes could also influence PhDEF behaviour between
512 layers, such as the regulation of *PhDEF* translation or PhDEF protein movement between layers[95,96],
513 or PhDEF post-translational modifications such as phosphorylation[97]. Other studies than ours have
514 reported that regulators could have different functions in different cell layers[73–78]; which suggests
515 that cell layer identity-specific impact on transcription factors is likely a general rule.

ACKNOWLEDGMENTS

We thank Francesco Quattrocchio and Shuangjiang Li for sharing their protocol for protoplast isolation; Cyril Dégletagne and the Cancer Genomics Platform from the Centre de Recherche en Cancérologie de Lyon for support in the scRNA-Seq experiment; Nicolas Dalle and Annick Dubois for assistance in chromatin immunoprecipitation; Benjamin Gillet and Sandrine Hugues from the sequencing platform of the Institut de Génomique Fonctionnelle de Lyon for library preparation and sequencing of the transcriptomes of this study; the Master of Biology of the Ecole Normale Superieure de Lyon for their participation in funding and analyzing the ChIP-seq experiment; and Sergio Sarnataro from Spatial-Cell-ID for advice on scRNA-Seq analysis. This work was supported by a grant to QC-S and MM from the Agence Nationale de la Recherche (grant ANR-19-CE13-0019, FLOWER LAYER), by a grant to ED from the French Ministry of Higher Education and Research, by grants to DB from the Agence Nationale de la Recherche (grants ANR-21-CE12-0036-01 and ANR-21-CE20-0007-02) and by the EquipEx+ Spatial-Cell-ID under the "Investissements d'avenir" program (ANR-21-ESRE-00016). We gratefully acknowledge support from the CBPsmn (PSMN, Pôle Scientifique de Modélisation Numérique) of the ENS de Lyon for the computing resources. The platform operates the SIDUS solution[98] developed by Emmanuel Quemener.

AUTHOR CONTRIBUTIONS

QC-S, ED, DB and MM conceived the project and designed experiments. QC-S, ED, PC, PM, and SRB performed the experiments and QC-S, ED, ED-B, BL, CR, JJ and MM analyzed the data. QC-S, ED, DB and MM designed figures and wrote the manuscript, with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURES

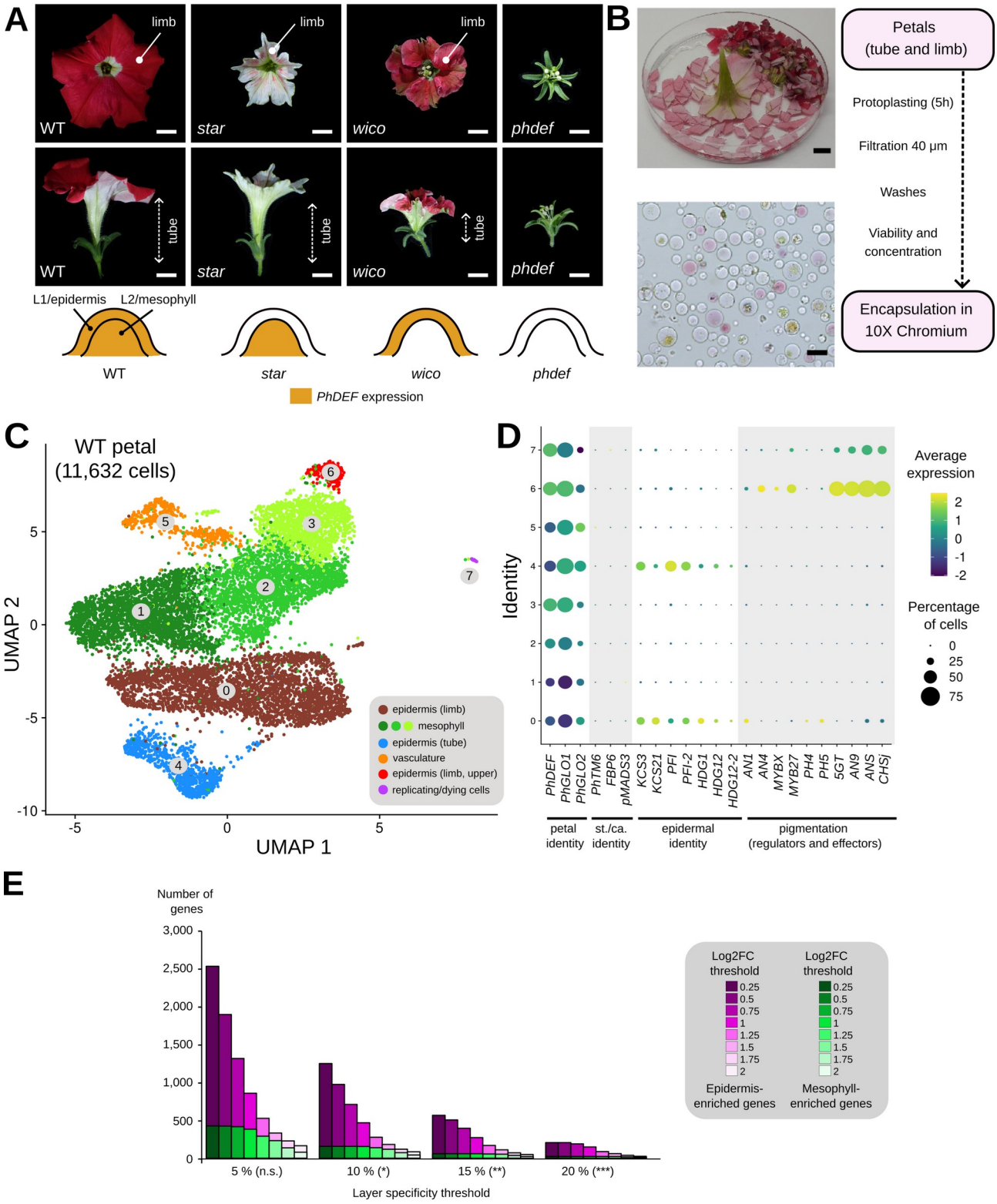

Figure 1. Single-cell RNA-sequencing separates epidermal and mesophyll cell types in the petunia petal.

(A) Representative petunia wild-type (WT), *star*, *wico* and *phdef* flowers from an upper (top picture) and side view (bottom picture). The limb and tube are indicated. Scale bar = 1 cm. Below

the flower pictures, a schematic petal primordium is depicted with L1 (future epidermis) and L2 (future mesophyll) cell layers, and *PhDEF* expression in the WT, *star*, *wico* and *phdef* primordia is indicated in orange. **(B)** Overview of the experimental protocol to produce petal protoplasts ready for isolation with the 10X Genomics Chromium device. Upper picture: a WT flower cut into small fragments in the cell wall-digesting solution (scale bar = 1 cm). Bottom picture: isolated protoplasts from a WT petal after a 5h-long digestion process, viewed by bright-light microscopy (scale bar = 50 µm). **(C)** Uniform Manifold Approximation and Projection (UMAP) plot of 11,632 WT petal cells sequenced for their transcriptome, after integration from two biological replicates. Clusters of cells are color-coded and are ordered from the biggest to the smallest, based on the number of cells. **(D)** Dotplot of key marker genes for petal identity, stamen and carpel identity, epidermal identity and pigmentation. The size of the dot represents the percentage of cells that express a given gene, and the color scale indicates the average expression level of the gene across all cells in a given cluster. **(E)** Barplot of the number of genes enriched in the epidermis (magenta) or in the mesophyll (green), as defined by varying the log2FoldChange (log2FC) threshold and the layer specificity threshold (% of cells expressing the gene in the epidermis - % of cells expressing the gene in the mesophyll). The Kolmogorov-Smirnov two-sided test (KS test) was applied to compare the gene number distributions between epidermis- and mesophyll-enriched genes at a given layer-specificity threshold, n.s. non significant, * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$.

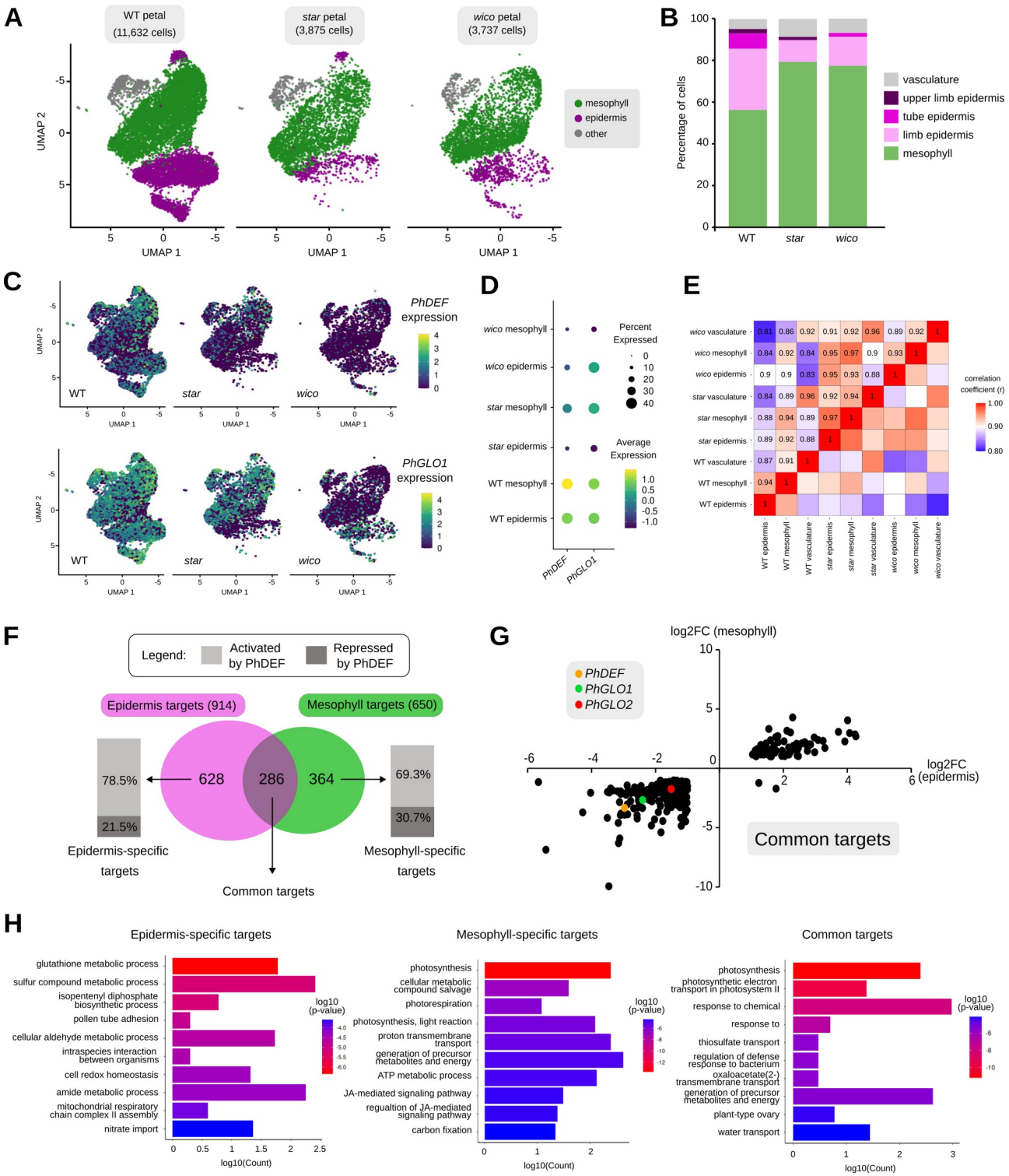

Figure 2. PhDEF regulates a different set of target genes in the petal epidermis and mesophyll. **(A)** Uniform Manifold Approximation and Projection (UMAP) plots of 11,632 WT, 3,875 *star* and 3,737 *wico* petal cells sequenced for their transcriptome, after integration and clustering. Here, epidermal clusters (purple) and mesophyll clusters (green, without the vasculature) were merged. **(B)** Percentage of cells from the different epidermal and mesophyll identities identified from WT, *star* and *wico* petal scRNA-Seq. **(C)** Per-cell expression levels of *PhDEF* and *PhGLO1* in WT, *star* and *wico* displayed on UMAP plots. **(D)** Dotplot of PhDEF and PhGLO1 expression level (color-coded) and prcentage of cells expressing the gene (coded in the size of the dot) in the epidermis and

mesophyll cells of WT, *star* and *wico* petals. **(E)** Pearson's correlation coefficient (r) plot between the pseudo-bulk transcriptomes from the mesophyll, epidermis and vasculature cells extracted from the WT, *star* and *wico* scRNA-Seq data. A Fisher r-to-z transformation test indicates that all correlation coefficients are significantly different ($p < 0.05$). **(F)** Venn diagram of the number of epidermis (purple) and mesophyll (green) targets of PhDEF, revealing the number of epidermis- and mesophyll-specific targets, as well as common targets (*i.e.* differentially expressed in both layers), based on WT, *star* and *wico* scRNA-Seq data. The percentage of activated (light grey) and repressed (dark grey) targets is also displayed. **(G)** Differential expression (log2FC = log2(FoldChange)) of the common targets of PhDEF in the epidermis or in the mesophyll, showing that almost all targets are either activated or repressed in both layers. Three genes of interest (*PhDEF*, *PhGLO1* and *PhGLO2*) are displayed as colour points. **(H)** Ten most enriched Gene Ontology (GO) terms for biological processes in epidermis-specific (left), mesophyll-specific (middle) and common (right) PhDEF targets, after GO term redundancy reduction and sorting by *p*-value.

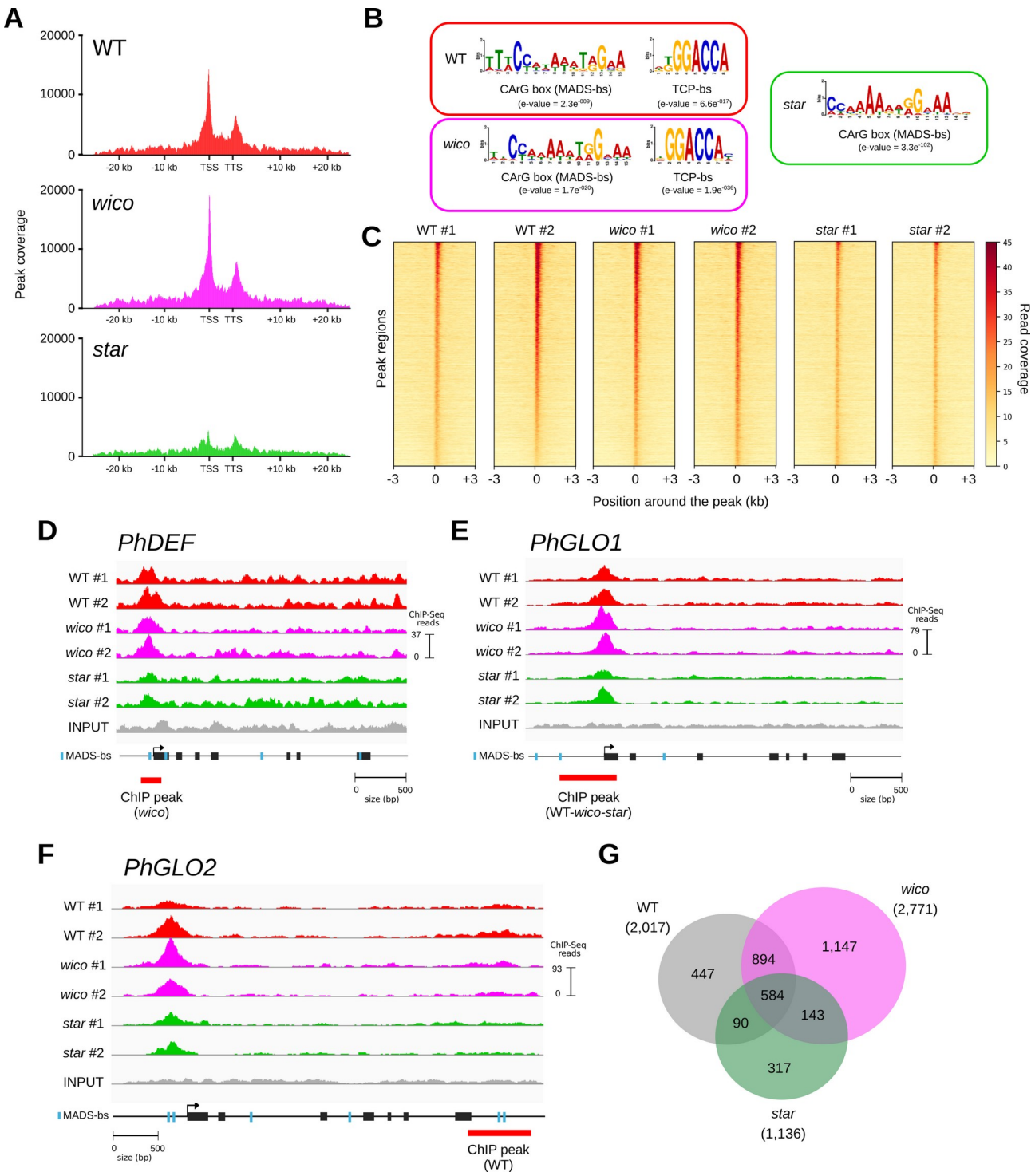

Figure 3. PhDEF regulates directly more target genes in the epidermis than in the mesophyll. **(A)** Metaplot of PhDEF binding sites in WT, *wico* and *star* petals, with the distance from the Transcription Start Site (TSS) and the Transcription Termination Site (TTS). **(B)** Selected motifs enriched under PhDEF ChIP-Seq peaks for WT, *wico* and *star* petals, bs = binding site. The full list of motifs detected is in Figure S9. **(C)** Heatmap of the read coverage of all peaks detected in WT samples, and of corresponding regions of the genome in *wico* and *star* samples. The peaks are sorted according to the highest to lowest read coverage in WT #2, so that the same regions of the genome are in the same horizontal line in all librairies. The read coverage is color-coded, and

position 0 represents the start of the peak. **(D-F)** PhDEF binding profiles in WT, *star* and *wico* petals in the genomic regions of *PhDEF* (D), *PhGLO1* (E) and *PhGLO2* (F). Peaks that pass our pipeline for reproducibility (see Methods) are displayed as thick red lines. Predicted MADS-binding sites (MADS-bs) are indicated by blue rectangles, while no TCP-bs were found in these sequences. Although the peak in the promoter region of *PhDEF* does not pass the IDR threshold, and the peak in the promoter region of *PhGLO2* is removed because it overlaps with a peak in the input, their binding profiles and the presence of MADS-bs strongly suggests that PhDEF binds to these positions. **(G)** Intersection of the number of gene-associated peaks from WT, *star* and *wico* ChIP-Seq data. The intersection occasionally resulted in the artifical duplication of peaks, and once corrected, this marginally changed the total number of peaks identified for each genotype in the intersection.

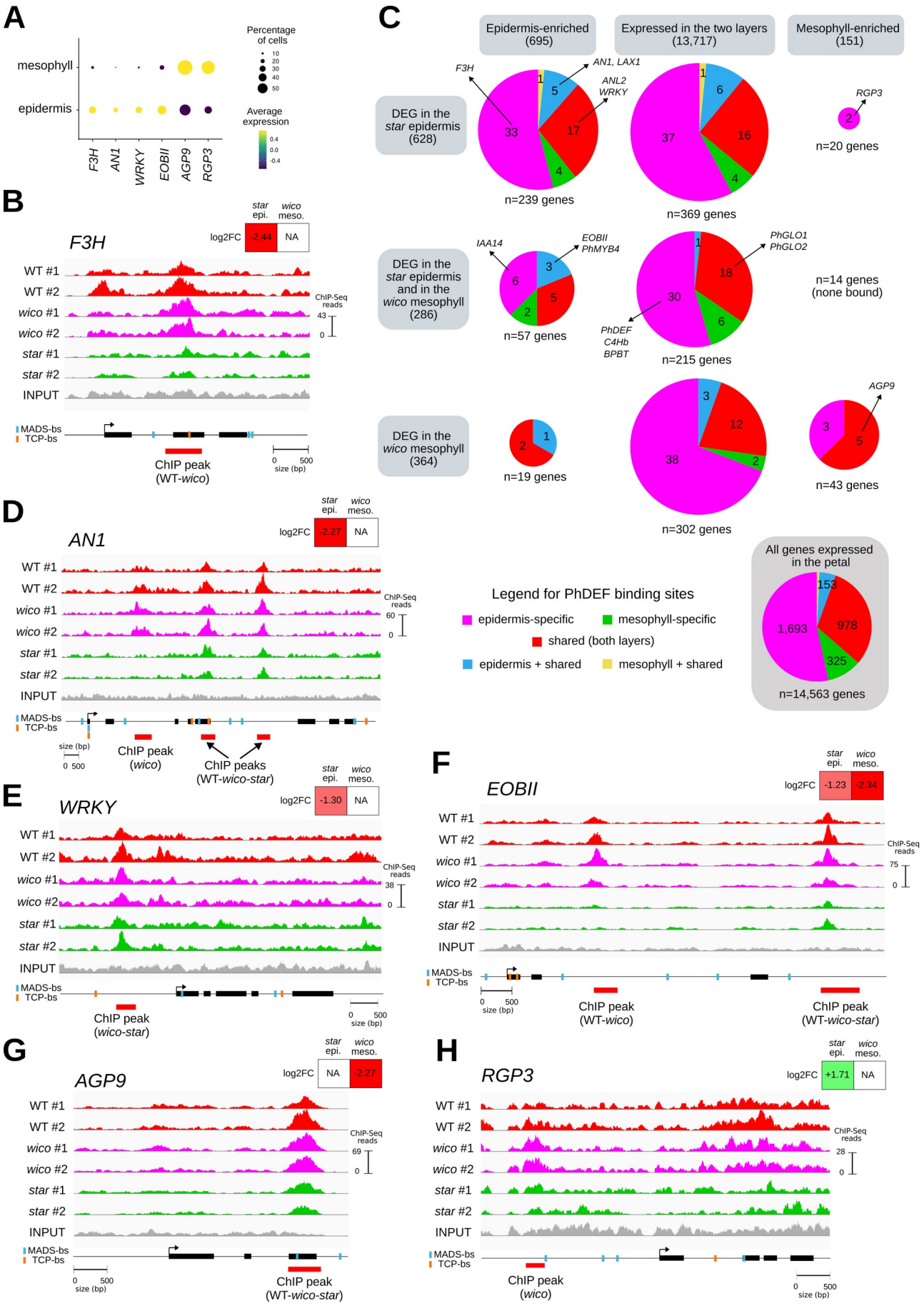
A
mesophyll
epidermis
F3H
AN1
WRKY
EOBII
AGP9
RGP3
Percentage of cells
Average expression
B
F3H
star epi.
wico meso.
log2FC
-2.44
NA
WT #1
WT #2
wico #1
wico #2
star #1
star #2
INPUT
ChIP-Seq reads
MADS-bs
TCP-bs
ChIP peak (WT-wico)
size (bp)
C
Epidermis-enriched (695)
Expressed in the two layers (13,717)
Mesophyll-enriched (151)
DEG in the star epidermis (628)
F3H
AN1, LAX1
ANL2
WRKY
n=239 genes
n=369 genes
RGP3
n=20 genes
DEG in the star epidermis and in the wico mesophyll (286)
IAA14
EOBII
PhMYB4
n=57 genes
PhGLO1
PhGLO2
PhDEF
C4Hb
BPBT
n=215 genes
n=14 genes (none bound)
DEG in the wico mesophyll (364)
n=19 genes
n=302 genes
AGP9
n=43 genes
Legend for PhDEF binding sites
epidermis-specific
mesophyll-specific
shared (both layers)
epidermis + shared
mesophyll + shared
All genes expressed in the petal
153
1,693
978
325
n=14,563 genes
D
AN1
log2FC
-2.27
NA
ChIP peak (wico)
ChIP peaks (WT-wico-star)
E
WRKY
log2FC
-1.30
NA
ChIP peak (wico-star)
F
EOBII
log2FC
-1.23
-2.34
ChIP peak (WT-wico)
ChIP peak (WT-wico-star)
G
AGP9
log2FC
NA
-2.27
ChIP peak (WT-wico-star)
H
RGP3
log2FC
+1.71
NA
ChIP peak (wico)

Figure 4. A diversity of PhDEF binding profiles underlie gene layer-specific expression and regulation by PhDEF.

(A) Dotplot of the average expression level and the percentage of cells expressing the genes *F3H*, *AN1*, *WRKY*, *EOBII*, *AGP9* and *RGP3*, from the WT scRNA-Seq data. **(B, D-H)** PhDEF binding profiles on the genes *F3H*, *AN1*, *WRKY*, *EOBII*, *AGP9* and *RGP3*, in WT, *star* and *wico* petal samples. For each gene, the read coverage of the ChIP-Seq experiment is displayed, scaled for each gene individually depending on the number of reads found under the peak. The gene model is represented below, with exons as black rectangles and an arrow indicating the transcription start site. Predicted MADS-binding sites (MADS-bs) are indicated by blue rectangles, while TCP-bs are indicated by orange rectangles, on the gene models or occasionally below for readibility (*AN1*). The ChIP peak detected by our pipeline is indicated with a red rectangle, and the sample or intersection of samples in which it was detected. Above the binding profile, the log2 Fold Change (log2FC) of the gene expression in the *star* epidermis or in the *wico* mesophyll is indicated, as compared to WT, from the scRNA-Seq data, with colors representing the direction of differential expression (red: down-regulated, green: up-regulated, NA: not applicable). **(C)** Pie charts representing the proportions of the different PhDEF binding profiles, as indicated in the legend, for different gene expression profiles from the WT scRNA-Seq data, and different gene regulation profiles by PhDEF from the *star* and *wico* scRNA-Seq data. The size of the slices represent the proportion of genes, as compared to all genes bound, with the corresponding profiles, and the number of genes is indicated in the slices. The total number of genes for which binding profiles have been explored are indicated below each pie, and the size of the pie is proportional to this total number. The name of some genes of interest are indicated, that are either the ones whose profiles are shown in this Figure or in Figure 3C, or mentioned in the text. The pie chart of PhDEF binding profiles for all genes expressed in the petal allows a direct comparison of proportions with the other pie charts, all differences being non-significant with a hypergeometric test.

MATERIAL AND METHODS

Plant material and culture conditions

Plants were grown in a culture room in long day conditions (16h day at 22°C, 8h night at 18°C, 75-Valoya NS12 LED bars, light intensity: 130 µE, 60% humidity). All plant material derives from the *Petunia x hybrida* R27 line containing the active *dTph1* transposable element. The *wico* and *star* flowers spontaneously arose from *phdef-151* homozygous mother plants, by the cell layer-specific excision of *dTph1*, resulting in branches carrying either *wico* or *star* flowers that were repeatedly obtained from several different *phdef-151* individuals, as described in[38]. These branches were subsequently maintained by cuttings.

Histology

Petal cross-sections and toluidine blue staining were performed as described in[38]. Cell types were counted as indicated in Figure S1, from 4 different cross-section images per limb and per tube (n = 1,038 limb cells and 2,736 tube cells counted in total).

Bulk RNA-Seq

From WT petals at stage 12[38,52], a fraction of the corollas was directly flash-frozen; the rest was protoplasted as described below and the protoplast pellet was flash-frozen. RNA extraction was performed with Sigma's Spectrum™ Plant Total RNA Kit with On-Column DNase I Digestion, following manufacturer's recommendations. RNA integrity and quantity were determined using a Bioanalyzer RNA 6000 Nano assay (Agilent), libraries were prepared with the CORALL RNA-Seq Library Prep (Lexogen) and sequenced with an Illumina NextSeq500 (single-end reads, 84 bp). Reads were mapped on the *P. axillaris* transcriptome[99] as described in[38]. Differential gene expression was computed with DESeq2[100] (1.34.0) in RStudio (R 4.4.1, RStudio 2024.04.2).

RT-qPCR

For RNA extraction, limb or tube tissue was collected from WT flowers at anthesis in 3 biological replicates. For limb samples, the limb from 1 petal from 3 different flowers constitutes one replicate; for tube tissue, the whole tube from 3 different flowers constitutes one replicate. RNA extraction, Reverse Transcription (RT), quantitative PCR (qPCR) and result analysis were performed as described previously[34]. Primer sequences are available in Table S4.

In situ hybridization

In situ hybridization was performed as previously described[38], using *f3h* mutant flowers with white petals[101], to avoid detecting an aspecific signal in pigmented epidermal cells. The primers used to synthesize the antisense probe for *IAA14* can be found in Table S5.

Single-cell RNA-Seq

Petal protoplast isolation

Protoplast isolation was performed in sterile conditions. WT, *star* and *wico* corollas (ca. 15, 30 and 30 flowers per replicate, respectively) were briefly soaked in 70% ethanol and in 0.5% bleach for 30’’ then rinsed 3 times with sterile water. Sterile corollas were transferred in a petri dish containing 2 mL of Digestion Mix (0.4% macerozyme R-10, 0.8% Cellulase Onozuka R-10, w/v in TEX Buffer (3.1 g/L Gamborg B5 salts, 500 mg/L MES, 750 mg/L $CaCl_2$*$2H_2O$, 250 mg/L NH_4NO_3, 136.9 g/L Sucrose, pH 5.7)) and cut in ca. 0.5 cm² pieces, using a new scalpel blade for each corolla to reduce tissue wounding. 10 mL of Digestion Mix were added and digestion was performed for 5 h at 26°C in the dark, with gentle orbital agitation (20 rpm) for the last 15’. After filtration through a 40 µm mesh, the volume was adjusted to 25 mL using 0.4 M Sucrose (492 mOsm.kg-1 H_2O) and the mixture was centrifuged for 10’ at 100 g with acceleration 2/9 and deceleration 0/9 with a swing-out rotor. The underlying buffer was removed using a peristaltic pump (Gilson MINIPULS™ Evolution) connected to a sterile Pasteur pipette without perturbing the protoplast layer, at a rate of ca. 100 µL/sec. After adjusting again the volume to 25 mL using 0.4 M Sucrose, the whole process was repeated twice. Protoplast concentration and viability was assessed using a Kova slide and staining with 1% Evans Blue dye solution (w/v in 0.4 M Sucrose) as described in[102]. Cell types were estimated as indicated in Figure S1, from 4 different pictures taken from a single protoplast isolation event (n = 3,338 cells counted in total).

From single cell transcriptome sequencing to clustering

Protoplast suspensions were adjusted to 345, 480, 590 and 560 cells/µL with Mannitol-BSA (0.44 M Mannitol, 0.1 % BSA (w/v), 498 mOsm.kg-1 H2O) for WT (first replicate), *star*, *wico* and WT (second replicate) respectively. Protoplasts were loaded in the 10x Genomics Chromium chip and libraries were prepared using the Chromium Next GEM Single Cell 3′ Reagent Kits v3.1 (dual index) kit following manufacturer's instructions. Libraries were sequenced on an Illumina NovaSeq 6000 SP or S1. Read quality was checked with FastQC[103] (v0.12.1) and reads were aligned on the *P. axillaris* v1.6.2 transcriptome (annotation from[40] transferred to a genome improved by HiC[38,104], which was the best available version of a Petunia genome at that time). CellRanger count (v7.0.1, 10X Genomics) was used to filter reads, count barcodes and UMIs, and generate the HDF5 matrix. The second WT replicate was split and sequenced in two technical replicates on the same Chromium chip, and read counts were aggregated with the Cell Ranger function aggr. Normalized gene expression, UMAP plots and cluster markers were computed with Seurat 5.1.0[41,105] in R[106] (v4.3.3) as follows, using default parameters unless stated otherwise: cells with less than 200 genes and genes dectected in less than 3 cells were removed; data was normalized with NormalizeData, highly variable features were identified with FindVariableFeatures and scaled with ScaleData; data dimension was reduced by Principal Component Analysis with RunPCA; Doublets were counted and removed with DoubletFinder[107]; the number of significant Principal Components (PC) was

determined with the function JackStraw[108]; Uniform Manifold Approximation and Projection (UMAP) dimension reduction[109] was performed with the function RunUMAP using the appropriate number of PCs. For the WT sample, MultiK[42] was used to explore the optimal number of clusters; a nearest-neighbor graph was constructed with the function FindNeighbors; clustering was computed with the function FindClusters, using the parameters determined by MultiK; cluster markers were identified with the function FindMarkers. For the identification of clusters, orthology to *Arabidopsis thaliana* genes was defined as best or second best reciprocal Blast hit. A similar pipeline, up to clustering, was applied to the *star* and *wico* scRNA-Seq samples.

Data integration

To estimate the effect of protoplasting on WT scRNA-Seq clusters, genes differentially expressed after protoplasting with different log2FoldChange values (absolute log2FoldChange > 1, 1.5, 2 or 2.5) were removed from the scRNA-Seq dataset, to generate 4 datasets. These were integrated with the original WT dataset with Harmony[51] (implemented in Seurat), and cell proportions per cluster were compared. To compare the WT, *star* and *wico* samples, the two biological replicates of WT scRNA-Seq were first integrated with Harmony; then the WT, *star* and *wico* samples (after computing the UMAP) were integrated together with Harmony, to ensure an unbiased representation of each genotype for cluster identification after integration.

Layer-specific gene differential expression

On the WT, *star* and *wico* datasets after integration, the 3 epidermal clusters and the 3 mesophyll clusters were merged together by redefining their identities, while the vasculature and replicating/dying cells were kept untouched. For the *star* sample only, the small cluster of limb epidermal cells (secondary L1-revertants) was removed from the epidermal cluster. Only genes with more than 50 total counts in the dataset were kept for further analysis. Layer-specific DEGs were computed with the function FindMarkers, and manually subsetted for an absolute log2FoldChange > 0.75 and pct.1 - pct.2 > 10% (pct.1 being the percentage of cells from group 1 expressing the gene, and pct.2 from group 2). DEGs in the *star* epidermis and *wico* mesophyll, as compared to the corresponding tissue in WT, were computed similarly, and subsetted with an absolute log2FoldChange > 1, adjusted *p*-value < 0.01 and pct.1 - pct.2 >10 %. Venn diagrams were generated with the help of Interactivenn[110].

Chromatin immunoprecipitation and sequencing (ChIP-Seq)

ChIP assay and sequencing

Chromatin immunoprecipitation on two biological replicates for WT, *star* and *wico* petals at stage 8 was performed as previously described[111,38], using polyclonal antibodies directed against PhDEF devoid of its highly conserved DNA-binding domain. The starting amount of sonicated chromatin for the IP was 25 µL, 100 µL and 100 µL for WT, *star* and *wico* respectively, following visual

inspection of sonicated chromatin amounts by gel electrophoresis and enrichment tests on control genes by ChIP-qPCR. The quality of IP and INPUT (WT chromatin) samples was assessed with Tapestation 4150 HS D5000 (Agilent). Libraries were prepared with MicroPlex Library Preparation Kit v3 (Diagenode) and around 1 ng of each library was PCR-amplified (12 cycles), following Diagenode's recommendations. Libraries were analyzed TapeStation 4150 HS D5000 and quantified with Qubit 4.0 with the Qubit dsDNA HS Assay Kit (Thermofisher), then mixed at equimolar ratio for sequencing with an Illumina NextSeq 500 (paired-end, 2×76 bp, dual indexing). Read quality was assessed with FastQC and MultiQC[112], and reads were trimmed with fastx-trimmer[113] to remove the first 6 nucleotides on the 5' end and the last 2 nucleotides on the 3' end of each read, resulting in 2×68 bp reads. Amplification bias was removed as previously described[114]. Reads were mapped on the *P. axillaris* genome (v1.6.2 superscaffolded by HiC[40,104,38]), which was the most recent version of the genome available at the time data analysis was initiated. Mapping was performed with Bowtie2[115] (v2.4.2) with default parameters, and only reads with mapping quality above 20 were kept. The effective genome size was estimated at 9.99e8 with khmer[116], with a kmer size of 68.

Peak detection and annotation

Peaks were called on the IP and INPUT samples separately with MACS2[117] with a False Discovery Rate (FDR) threshold of respectively 0.1 and 0.05. The Irreproducible Discovery Rate (IDR)[118] was computed on IP samples to identify reproducible peaks between replicates, and only peaks with an IDR value below 0.1 were kept. Peaks detected in the INPUT sample were removed from the reproducible IP peaks if they overlaped by at least 25 % reciprocally, or if INPUT peaks were included in IP peaks, which was computed with a custom script using a combination of bedtools[119] (v2.26.0) commands (intersect, window, substract). Peaks of WT, *star* and *wico* were intersected using bedtools, and shared peaks were defined by a reciprocal overlap of at least 25%, in order to identify marked regions in one, two or three genotypes. This occasionally resulted in the artefactual duplication of peaks, which was manually corrected. Annotation of genes close to peaks was performed with Python, and only peaks within 6 kb upstream of the TSS and 6 kb downstream of the TTS were further analyzed. This distance was chosen based on the metaplots, and because it encompasses 80% of the ChIP peaks. In cases when a peak was assigned to several possible genes, the gene with the shortest distance to the peak was retained. Motifs under the peaks were explored with MEME-ChIP (v5.5.7)[120] with a minimal size of 6 bp and a maximal size of 15 bp against the JASPAR 2024[121] CORE non-redundant v2 database, and selected motifs (WT motifs from Fig. 3B) were used to predict binding sites on individual genomic sequences with FIMO[122]. Peaks and gene models were visualized with the Integrative Genome Viewer[123]. For heatmaps of read coverage, coverage files (bigwig files) were calculated from BAM files using bamCoverage from the package

deepTools2[124] (v3.5.4). Heatmaps were built using computeMatrix and plotHeatmap functions using a window of 3 kb beginning at the start of the peak.

GO term database improvement and enrichment analysis

Protein sequences were predicted from the *P. axillaris* genome annotation transferred from the published version v1.6.2[99] to the Hi-C superscaffolded one using GFFread[125] and TransDecoder.LongOrfs[126] (v5.5.0) with parameter "-m50". A total of 56,351 predicted protein sequences were obtained. We then applied custom Perl scripts (refer to *Data availability* section) to select the most plausible CDS for each predicted mRNA, resulting in 31,626 predicted protein sequences. These sequences were then compared with UniProtKB protein database version 2024_06[127]. The complete Swiss-Prot database of curated proteins (containing 44,534 plant sequences and 571,609 non-plant sequences) and the plant subset of the noncurated database TrEMBL (containing 20,024,921 sequences) were used. The comparison was performed using BlastP[128] program (AB-Blast v3.0 release 2020-03-17) with parameters "W=3 Q=7 R=2 matrix=BLOSUM80 B=200 V=200 E=1e-6 hitdist=60 hspsepqmax=30 hspsepsmax=30 sump postsw". The BLAST output was filtered using custom Perl scripts to retain only those matches with a log10(*e*-value) no lower than 75% of the best log10(*e*-value). The GO annotations for proteins contained in the Uniprot KnowledgeBase were downloaded as a GPA file (v222, 2024-08-01) from the GOA database[129] hosted at EBI. The Gene Ontology database[130] was retrieved via the QuickGO web interface[131] in October 2024; GO term relationships and UniprotKB GO annotations were subsequently loaded into a custom PostgreSQL database (v13.13) for easy querying. Custom Perl scripts were written to associate each predicted petunia protein sequence to the GO terms associated with all its matched proteins or with at least with five matched proteins. The comprehensive list of GO annotations for the entire Petunia proteome is accessible in the project dataset as a text file (see *Data availability* section), with the list of underlying Blast evidences. Enrichment analysis for test genes was performed using all genes expressed in the WT petal bulk RNA-Seq as background genes. A hypergeometric test (R version 4.2.2[106]) was applied to assess the significance of enrichment of each subset[132,133]. GO terms enriched (p-value < 0.01 with a hypergeometric test) and with at least 2 genes in the pathway were analyzed with REVIGO[134] (Tiny mode) to reduce redundancy. Enriched GO terms for biological processes were sorted by their log10(p-value) and the 10 best terms were displayed in barplots generated with ggplot2[135] in R.

Data and code availability

All data and code are publicly available as of the date of publication. Single-cell RNA-seq raw fastq files, h5 matrices and rds images after integration have been deposited at GEO (GSE290697). ChIP-Seq fastq and BED files have been deposited at GEO (GSE308255). Scripts for bulk RNA-Seq

analysis, GO database update and GO term enrichment, and scRNA-Seq analysis are available at gitbio.ens-lyon.fr/rdp/phdef_flower_layer. ChIP-Seq BAM and BAI files, ChIP-Seq and scRNA-Seq FastQC html reports, and functional annotation of the *Petunia axillaris* proteome are available at https://entrepot.recherche.data.gouv.fr/dataverse/PhDEF_Flower_layer.

REFERENCES

1. Bowman, J. L. & Moyroud, E. Reflections on the ABC model of flower development. *Plant Cell* **36**, 1334–1357 (2024).
2. Peraldi, R. & Kmita, M. 40 years of the homeobox: mechanisms of Hox spatial-temporal collinearity in vertebrates. *Dev. Camb. Engl.* **151**, dev202508 (2024).
3. Hubert, K. A. & Wellik, D. M. Hox genes in development and beyond. *Development* **150**, dev192476 (2023).
4. Buffry, A. D. & McGregor, A. P. Micromanagement of Drosophila Post-Embryonic Development by Hox Genes. *J. Dev. Biol.* **10**, 13 (2022).
5. Dornelas, M. C., Patreze, C. M., Angenent, G. C. & Immink, R. G. H. MADS: the missing link between identity and growth? *Trends Plant Sci.* **16**, 89–97 (2011).
6. Satina, S., Blakeslee, A. F. & Avery, A. G. Demonstration of the Three Germ Layers in the Shoot Apex of Datura by Means of Induced Polyploidy in Periclinal Chimeras. *Am. J. Bot.* **27**, 895–905 (1940).
7. Stewart, R. N. & Burk, L. G. Independence of Tissues Derived from Apical Layers in Ontogeny of the Tobacco Leaf and Ovary. *Am. J. Bot.* **57**, 1010–1016 (1970).
8. Meyerowitz, E. M. Genetic control of cell division patterns in developing plants. *Cell* **88**, 299–308 (1997).
9. Scheres, B. Plant cell identity. The role of position and lineage. *Plant Physiol.* **125**, 112–114 (2001).
10. Schrick, K., Ahmad, B. & Nguyen, H. V. HD-Zip IV transcription factors: Drivers of epidermal cell fate integrate metabolic signals. *Curr. Opin. Plant Biol.* **75**, 102417 (2023).
11. Robinson, D. O. & Roeder, A. H. K. Themes and variations in cell type patterning in the plant epidermis. *Curr. Opin. Genet. Dev.* **32**, 55–65 (2015).
12. Coen, E. S. & Meyerowitz, E. M. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31–37 (1991).
13. Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. & Sommer, H. Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus. *Science* **250**, 931–936 (1990).
14. Theissen, G. & Saedler, H. Plant biology. Floral quartets. *Nature* **409**, 469–471 (2001).

15. Theißen, G., Melzer, R. & Rümpler, F. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* **143**, 3259–3271 (2016).
16. Käppel, S., Rümpler, F. & Theißen, G. Cracking the Floral Quartet Code: How Do Multimers of MIKCC-Type MADS-Domain Transcription Factors Recognize Their Target Genes? *Int. J. Mol. Sci.* **24**, 8253 (2023).
17. Melzer, R. & Theissen, G. Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. *Nucleic Acids Res.* **37**, 2723–2736 (2009).
18. Causier, B., Schwarz-Sommer, Z. & Davies, B. Floral organ identity: 20 years of ABCs. *Semin. Cell Dev. Biol.* **21**, 73–79 (2010).
19. Litt, A. & Kramer, E. M. The ABC model and the diversification of floral organ identity. *Semin. Cell Dev. Biol.* **21**, 129–137 (2010).
20. Morel, P. *et al.* Divergence of the Floral A-Function between an Asterid and a Rosid Species. *Plant Cell* **29**, 1605–1621 (2017).
21. Galliot, C., Stuurman, J. & Kuhlemeier, C. The genetic dissection of floral pollination syndromes. *Curr. Opin. Plant Biol.* **9**, 78–82 (2006).
22. Rodrigues, D. M. *et al.* Do we truly understand pollination syndromes in Petunia as much as we suppose? *AoB PLANTS* **10**, ply057 (2018).
23. Vandenbussche, M., Zethof, J., Royaert, S., Weterings, K. & Gerats, T. The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in Petunia hybrida flower development. *Plant Cell* **16**, 741–754 (2004).
24. Rijpkema, A. S. *et al.* Analysis of the Petunia TM6 MADS Box Gene Reveals Functional Divergence within the DEF/AP3 Lineage. *Plant Cell* **18**, 1819–1832 (2006).
25. Tröbner, W. *et al.* GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis. *EMBO J.* **11**, 4693–4704 (1992).
26. Schwarz-Sommer, Z. *et al.* Characterization of the Antirrhinum floral homeotic MADS-box gene deficiens: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* **11**, 251–263 (1992).

27. McGonigle, B., Bouhidel, K. & Irish, V. F. Nuclear localization of the Arabidopsis APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. *Genes Dev.* **10**, 1812–1821 (1996).
28. Winter, K.-U. *et al.* Evolution of Class B Floral Homeotic Proteins: Obligate Heterodimerization Originated from Homodimerization. *Mol. Biol. Evol.* **19**, 587–596 (2002).
29. Lenser, T., Theissen, G. & Dittrich, P. Developmental robustness by obligate interaction of class B floral homeotic genes and proteins. *PLoS Comput. Biol.* **5**, e1000264 (2009).
30. Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E. & Yanofsky, M. F. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405**, 200–203 (2000).
31. Pelaz, S., Tapia-López, R., Alvarez-Buylla, E. R. & Yanofsky, M. F. Conversion of leaves into petals in Arabidopsis. *Curr. Biol. CB* **11**, 182–184 (2001).
32. Immink, R. G. H. *et al.* Analysis of the petunia MADS-box transcription factor family. *Mol. Genet. Genomics MGG* **268**, 598–606 (2003).
33. Ferrario, S., Immink, R. G. H., Shchennikova, A., Busscher-Lange, J. & Angenent, G. C. The MADS box gene FBP2 is required for SEPALLATA function in petunia. *Plant Cell* **15**, 914–925 (2003).
34. Morel, P. *et al.* Divergent Functional Diversification Patterns in the SEP/AGL6/AP1 MADS-Box Transcription Factor Superclade. *Plant Cell* **31**, 3033–3056 (2019).
35. Wuest, S. E. *et al.* Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13452–13457 (2012).
36. Manchado-Rojo, M., Delgado-Benarroch, L., Roca, M. J., Weiss, J. & Egea-Cortines, M. Quantitative levels of Deficiens and Globosa during late petal development show a complex transcriptional network topology of B function. *Plant J. Cell Mol. Biol.* **72**, 294–307 (2012).
37. Bednarczyk, D. *et al.* The homeotic gene PhDEF regulates production of volatiles in petunia flowers by activating EOBI and EOBII. *Plant Cell* **37**, koaf027 (2025).
38. Chopy, M. *et al.* Cell layer-specific expression of the homeotic MADS-box transcription factor PhDEF contributes to modular petal morphogenesis in petunia. *Plant Cell* **36**, 324–345 (2024).
39. Satina, S. & Blakeslee, A. F. Periclinal Chimeras in Datura Stramonium in Relation to Development of Leaf and Flower. *Am. J. Bot.* **28**, 862–871 (1941).

40. Bombarely, A. *et al.* Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. *Nat. Plants* **2**, 16074 (2016).
41. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
42. Liu, S., Thennavan, A., Garay, J. P., Marron, J. S. & Perou, C. M. MultiK: an automated tool to determine optimal cluster numbers in single-cell RNA sequencing data. *Genome Biol.* **22**, 232 (2021).
43. Heijmans, K. *et al.* Redefining C and D in the petunia ABC. *Plant Cell* **24**, 2305–2317 (2012).
44. Kim, J.-Y. *et al.* Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* **33**, 511–530 (2021).
45. Hegebarth, D. & Jetter, R. Cuticular Waxes of Arabidopsis thaliana Shoots: Cell-Type-Specific Composition and Biosynthesis. *Plants* **6**, 27 (2017).
46. Nakamura, M. *et al.* Characterization of the class IV homeodomain-Leucine Zipper gene family in Arabidopsis. *Plant Physiol.* **141**, 1363–1375 (2006).
47. Boachon, B. *et al.* Natural fumigation as a mechanism for volatile transport between flower organs. *Nat. Chem. Biol.* **15**, 583–588 (2019).
48. Denyer, T. & Timmermans, M. C. P. Crafting a blueprint for single-cell RNA sequencing. *Trends Plant Sci.* **27**, 92–103 (2022).
49. Li, Z. & Ahammed, G. J. Plant stress response and adaptation via anthocyanins: A review. *Plant Stress* **10**, 100230 (2023).
50. Kang, M., Choi, Y., Kim, H. & Kim, S.-G. Single-cell RNA-sequencing of Nicotiana attenuata corolla cells reveals the biosynthetic pathway of a floral scent. *New Phytol.* **234**, 527–544 (2022).
51. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
52. Reale, L. *et al.* Patterns of cell division and expansion in developing petals of Petunia hybrida. *Sex. Plant Reprod.* **15**, 123–132 (2002).
53. Chen, D., Yan, W., Fu, L.-Y. & Kaufmann, K. Architecture of gene regulatory networks controlling flower development in Arabidopsis thaliana. *Nat. Commun.* **9**, 4534 (2018).

54. Kaufmann, K. *et al.* Orchestration of floral initiation by APETALA1. *Science* **328**, 85–89 (2010).

55. Kaufmann, K. *et al.* Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biol.* **7**, e1000090 (2009).

56. Bemer, M., van Dijk, A. D. J., Immink, R. G. H. & Angenent, G. C. Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. *Trends Plant Sci.* **22**, 66–80 (2017).

57. Wanamaker, S. A. *et al.* CrY2H-seq: a massively-multiplexed assay for deep coverage interactome mapping. *Nat. Methods* **14**, 819–825 (2017).

58. Zhao, Y. *et al.* TCP and MADS-Box Transcription Factor Networks Regulate Heteromorphic Flower Type Identity in Gerbera hybrida. *Plant Physiol.* **184**, 1455–1468 (2020).

59. Goslin, K., Finocchio, A. & Wellmer, F. Floral Homeotic Factors: A Question of Specificity. *Plants Basel Switz.* **12**, 1128 (2023).

60. Zhang, S. *et al.* Genome-Wide Identification, Characterization and Expression Analysis of TCP Transcription Factors in Petunia. *Int. J. Mol. Sci.* **21**, 6594 (2020).

61. Koes, R., Verweij, W. & Quattrocchio, F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* **10**, 236–242 (2005).

62. Spelt, C., Quattrocchio, F., Mol, J. N. M. & Koes, R. anthocyanin1 of Petunia Encodes a Basic Helix-Loop-Helix Protein That Directly Activates Transcription of Structural Anthocyanin Genes. *Plant Cell* **12**, 1619–1631 (2000).

63. Bainbridge, K. *et al.* Auxin influx carriers stabilize phyllotactic patterning. *Genes Dev.* **22**, 810–823 (2008).

64. Swarup, R. & Péret, B. AUX/LAX family of auxin influx carriers—an overview. *Front. Plant Sci.* **3**, 225 (2012).

65. Kubo, H., Peeters, A. J., Aarts, M. G., Pereira, A. & Koornneef, M. ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in Arabidopsis. *Plant Cell* **11**, 1217–1226 (1999).

66. Chopy, M. *et al.* A single MYB transcription factor with multiple functions during flower development. *New Phytol.* **239**, 2007–2025 (2023).

67. Spitzer-Rimon, B. *et al.* EOBII, a Gene Encoding a Flower-Specific Regulator of Phenylpropanoid Volatiles' Biosynthesis in Petunia[C][W]. *Plant Cell* **22**, 1961–1976 (2010).
68. Spitzer-Rimon, B. *et al.* The R2R3-MYB-like regulatory factor EOBI, acting downstream of EOBII, regulates scent production by activating ODO1 and structural scent-related genes in petunia. *Plant Cell* **24**, 5089–5105 (2012).
69. Colquhoun, T. A. *et al.* PhMYB4 fine-tunes the floral volatile signature of Petunia×hybrida through PhC4H. *J. Exp. Bot.* **62**, 1133–1143 (2011).
70. Dixon, R. A. *et al.* The phenylpropanoid pathway and plant defence-a genomics perspective. *Mol. Plant Pathol.* **3**, 371–390 (2002).
71. Lynch, J. H., Pichersky, E. & Dudareva, N. Floral Scent Metabolic Pathways and Their Regulation. in *Biology of Plant Volatiles* (CRC Press, 2020).
72. Mendes, M. A. *et al.* MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in Arabidopsis. *Plant Cell* **25**, 2560–2572 (2013).
73. Bouhidel, K. & Irish, V. F. Cellular Interactions Mediated by the HomeoticPISTILLATAGene Determine Cell Fate in theArabidopsisFlower. *Dev. Biol.* **174**, 22–31 (1996).
74. Carpenter, R. & Coen, E. S. Floral homeotic mutations produced by transposon-mutagenesis in Antirrhinum majus. *Genes Dev.* **4**, 1483–1493 (1990).
75. Carpenter, R. & Coen, E. S. Transposon induced chimeras show that floricaula, a meristem identity gene, acts non-autonomously between cell layers. *Dev. Camb. Engl.* **121**, 19–26 (1995).
76. Jenik, P. D. & Irish, V. F. The Arabidopsis floral homeotic gene APETALA3 differentially regulates intercellular signaling required for petal and stamen development. *Development* **128**, 13–23 (2001).
77. Sessions, A., Yanofsky, M. F. & Weigel, D. Cell-cell signaling and movement by the floral transcription factors LEAFY and APETALA1. *Science* **289**, 779–782 (2000).
78. Filippis, I., Lopez-Cobollo, R., Abbott, J., Butcher, S. & Bishop, G. J. Using a periclinal chimera to unravel layer-specific gene expression in plants. *Plant J. Cell Mol. Biol.* **75**, 1039–1049 (2013).
79. Cavallini-Speisser, Q., Morel, P. & Monniaux, M. Petal Cellular Identities. *Front. Plant Sci.* **12**, 745507 (2021).

80. Riglet, L., Gatti, S. & Moyroud, E. Sculpting the surface: Structural patterning of plant epidermis. *iScience* **24**, 103346 (2021).

81. Zuch, D. T. *et al.* Cell biology of the leaf epidermis: Fate specification, morphogenesis, and coordination. *Plant Cell* **34**, 209–227 (2021).

82. Borsuk, A. M., Roddy, A. B., Théroux-Rancourt, G. & Brodersen, C. R. Structural organization of the spongy mesophyll. *New Phytol.* **234**, 946–960 (2022).

83. Zhang, L. & Ambrose, C. Beauty is more than epidermis deep: How cell division and expansion sculpt the leaf spongy mesophyll. *Curr. Opin. Plant Biol.* **79**, 102542 (2024).

84. Lu, P., Porat, R., Nadeau, J. A. & O'Neill, S. D. Identification of a meristem L1 layer-specific gene in Arabidopsis that is expressed during embryonic pattern formation and defines a new class of homeobox genes. *Plant Cell* **8**, 2155–2168 (1996).

85. Abe, M., Katsumata, H., Komeda, Y. & Takahashi, T. Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. *Dev. Camb. Engl.* **130**, 635–643 (2003).

86. Bhatia, S. *et al.* A cellular expression map of epidermal and subepidermal cell layer-enriched transcription factor genes integrated with the regulatory network in Arabidopsis shoot apical meristem. *Plant Direct* **5**, e00306 (2021).

87. Schellmann, S. & Hülskamp, M. Epidermal differentiation: trichomes in Arabidopsis as a model system. *Int. J. Dev. Biol.* **49**, 579–584 (2005).

88. Fich, E. A., Segerson, N. A. & Rose, J. K. C. The Plant Polyester Cutin: Biosynthesis, Structure, and Biological Roles. *Annu. Rev. Plant Biol.* **67**, 207–233 (2016).

89. Yeo, M. T. S. & Moyroud, E. Under the rainbow: Novel insights on the mechanisms driving the development and evolution of petal pigmentation. *Curr. Opin. Plant Biol.* **86**, 102743 (2025).

90. Yadav, R. K., Tavakkoli, M., Xie, M., Girke, T. & Reddy, G. V. A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. *Dev. Camb. Engl.* **141**, 2735–2744 (2014).

91. Smaczniak, C., Muiño, J. M., Chen, D., Angenent, G. C. & Kaufmann, K. Differences in DNA Binding Specificity of Floral Homeotic Protein Complexes Predict Organ-Specific Target Genes. *Plant Cell* **29**, 1822–1835 (2017).

92. Swift, J. *et al.* Exaptation of ancestral cell-identity networks enables C4 photosynthesis. *Nature* **636**, 143–150 (2024).

93. Liu, X. *et al.* AGAMOUS terminates floral stem cell maintenance in Arabidopsis by directly repressing WUSCHEL through recruitment of Polycomb Group proteins. *Plant Cell* **23**, 3654–3670 (2011).

94. Smaczniak, C. *et al.* Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1560–1565 (2012).

95. Urbanus, S. L. *et al.* In planta localisation patterns of MADS domain proteins during floral development in Arabidopsis thaliana. *BMC Plant Biol* **9**, 5 (2009).

96. Urbanus, S. L. *et al.* Intercellular transport of epidermis-expressed MADS domain transcription factors and their effect on plant morphology and floral transition. *Plant J. Cell Mol. Biol.* **63**, 60–72 (2010).

97. Fujita, H. *et al.* An Arabidopsis MADS-box protein, AGL24, is specifically bound to and phosphorylated by meristematic receptor-like kinase (MRLK). *Plant Cell Physiol.* **44**, 735–742 (2003).

98. Quemener, E. & Corvellec, M. SIDUS—the solution for extreme deduplication of an operating system. *Linux J* **2013**, 3:3 (2013).

99. Bombarely, A. *et al.* Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. *Nat. Plants* **2**, 16074 (2016).

100. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

101. Yu, J., Tu, L., Subburaj, S., Bae, S. & Lee, G.-J. Simultaneous targeting of duplicated genes in Petunia protoplasts for flower color modification via CRISPR-Cas9 ribonucleoproteins. *Plant Cell Rep.* **40**, 1037–1045 (2021).

102. Fernandez-Da Silva, R. & Menéndez-Yuffá, A. Viability in protoplasts and cell suspensions of Coffea arabica cv. Catimor. *Electron. J. Biotechnol.* **9**, 0–0 (2006).

103. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. (2010).

104. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

105. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
106. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2024).
107. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329-337.e4 (2019).
108. Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinforma. Oxf. Engl.* **31**, 545–554 (2015).
109. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://doi.org/10.48550/arXiv.1802.03426 (2020).
110. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169 (2015).
111. Bouyer, D. *et al.* Genome-wide identification of RETINOBLASTOMA RELATED 1 binding sites in Arabidopsis reveals novel DNA damage regulators. *PLOS Genet.* **14**, e1007797 (2018).
112. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
113. Hannon, G. J. FASTX-Toolkit. http://hannonlab.cshl.edu/fastx_toolkit/ (2010).
114. Leemans, M. *et al.* A Mixture of Chemicals Found in Human Amniotic Fluid Disrupts Brain Gene Expression and Behavior in Xenopus laevis. *Int. J. Mol. Sci.* **24**, 2588 (2023).
115. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
116. Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence analysis. Preprint at https://doi.org/10.12688/f1000research.6924.1 (2015).
117. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
118. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, (2011).

119. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* **26**, 841–842 (2010).
120. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
121. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **52**, D174–D182 (2024).
122. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
123. Robinson, J. T., Thorvaldsdottir, H., Turner, D. & Mesirov, J. P. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* **39**, btac830 (2023).
124. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016).
125. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **9**, ISCB Comm J-304 (2020).
126. TransDecoder/TransDecoder, https://github.com/TransDecoder/TransDecoder. TransDecoder (2025).
127. UniProt: the Universal Protein Knowledgebase in 2023 | Nucleic Acids Research | Oxford Academic. https://academic-oup-com.inee.bib.cnrs.fr/nar/article/51/D1/D523/6835362.
128. Gish, W. AB-BLAST, https://blast.advbiocomp.com.
129. GOA database: Gene Ontology annotation updates for 2015 | Nucleic Acids Research | Oxford Academic. https://academic-oup-com.inee.bib.cnrs.fr/nar/article/43/D1/D1057/2437623.
130. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
131. QuickGO: a web-based tool for Gene Ontology searching | Bioinformatics | Oxford Academic. https://academic-oup-com.inee.bib.cnrs.fr/bioinformatics/article/25/22/3045/180464.
132. Pavlidis, P., Qin, J., Arango, V., Mann, J. J. & Sibille, E. Using the Gene Ontology for Microarray Data Mining: A Comparison of Methods and Application to Age Effects in Human Prefrontal Cortex. *Neurochem. Res.* **29**, 1213–1222 (2004).

133. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).

134. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms | PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021800.

135. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis, Https://Ggplot2.Tidyverse.Org*. (2016).

811