RESEARCH ARTICLE

Enhancer activation via TCP and HD-ZIP and repression by Dof transcription factors mediate giant cell-specific expression

Lilan Hong[1,2,5*], Byron Rusnak[2,5], Clint S. Ko[2,3,5], Shouling Xu[1], Xi He[1], Dengying Qiu[1], S. Earl Kang[4], Jose L. Pruneda-Paz[4], Adrienne H. K. Roeder[2*]

1 Institute of Nuclear Agricultural Sciences, Key Laboratory of Nuclear Agricultural Sciences of Ministry of Agriculture and Zhejiang Province, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China.

2 Weill Institute for Cell and Molecular Biology and School of Integrative Plant Science, Section of Plant Biology, Cornell University, Ithaca, NY 14853 USA.

3 Present address: Laboratory of Morphogenesis, The Rockefeller University, New York, NY 10065 USA.

4 Section of Cell and Developmental Biology, Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA

5 Co-first authors

* Correspondence: Lilan Hong (lilanhong@zju.edu.cn); Adrienne Roeder (ahr75@cornell.edu)

Running title: Giant cell-specific enhancer dissection

One-sentence summary: Dissection of an enhancer that drives giant cell-specific expression reveals that activation via TCPs and HD-ZIPs and repression via Dof transcription factors combine to mediate cell type specificity.

The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell/pages/) are: Adrienne H. K. Roeder (ahr75@cornell.edu) and Lilan Hong (lilanhong@zju.edu.cn).

Abstract

Proper cell-type identity relies on highly coordinated regulation of gene expression. Regulatory elements such as enhancers can produce cell type-specific expression patterns, but the mechanisms underlying specificity are not well understood. We previously identified an enhancer region capable of driving specific expression in giant cells, which are large, highly endoreduplicated cells in the *Arabidopsis thaliana* sepal epidermis. In this study, we use the giant cell enhancer as a model to understand the regulatory logic that promotes cell type-specific expression. Our dissection of the enhancer revealed that giant cell specificity is mediated primarily through the combination of two activators and one repressor. HD-ZIP and TCP transcription factors are involved in the activation of expression throughout the epidermis. High expression of HD-ZIP transcription factor genes in giant cells promoted higher expression driven by the enhancer in giant cells. Dof transcription factors repressed the activity of the enhancer such that only giant cells mainteained enhancer activity. Thus, our data are consistent with a conceptual model whereby cell type-specific expression emerges from the combined activities of three transcription factor families activating and repressing expression in epidermal cells.

In A Nutshell

Background: Specialized cell types carry out specific functions within plants. Some genes are expressed uniquely in one cell type and not expressed in other cells; these genes may be important for the function of that cell type. These cell type-specific expression patterns are created by transcription factors that bind to regions of the DNA called enhancers to regulate when and where a gene is expressed. Yet the regulatory logic behind how transcription factors that are more broadly expressed combine to create a cell type-specific expression pattern is not well understood. For example, giant cells are highly enlarged cells in the Arabidopsis sepal that control sepal curvature and promote defense from pathogens and insects. We previously identified a 1000-bp region of DNA that is sufficient to turn on the expression of a reporter gene specifically in giant cells.

Question: We asked how this 1000-bp DNA region activates gene expression specifically in giant cells.

Findings: Using Arabidopsis, we found that the combined activity of transcription factors from three families are important for generating the giant cell-specific pattern. TCP-type transcription factors promote expression from the enhancer in all epidermal cells. The transcription factor ATML1 modulates the expression level driven by the enhancer. As ATML1 protein is highly accumulated in giant cells, it can bind to the enhancer and drive high expression in giant cells. Dof transcription factors repress expression by binding to the enhancer, resulting in high expression in giant cells only. We created a conceptual model for how transcription factors whose encoding genes are expressed broadly in many cell types can bind to the same enhancer, combining their activities to produce cell type-specific expression patterns.

Next steps: A future challenge will be to use this information as a basis for engineering new synthetic enhancers with cell type-specific expression.

Keywords: Arabidopsis, flower development, sepal, giant cell, enhancer, endoreduplication, Dof, TCP, HD-ZIP

Introduction

Plant and animal organs are composed of different types of cells that perform specialized roles. Specialized cell types have distinct patterns of gene expression (Lee and Schiefelbein, 2002; Brady et al., 2007; Denyer et al., 2019; Ryu et al., 2019; Jaitin et al., 2014). One way cell-type specific expression patterns are established is through the activity of transcription factors binding to enhancers, short non-coding DNA sequences (Long et al., 2016). Enhancers typically contain clusters of transcription factor binding sites, acting as platforms to integrate varied information (Buecker and Wysocka, 2012). The spatial and temporal specificity of transcriptional regulation through enhancers has been extensively studied in several model organisms, including mouse (*Mus*

musculus), sea urchins (*Strongylocentrotus purpuratus*), chicks (*Gallus gallus domesticus*), and fruit flies (*Drosophila melanogaster*) (Davidson, 2010; Spitz and Furlong, 2012). Recently, many enhancers have been identified in Arabidopsis (*Arabidopsis thaliana*) and other plants through whole-genome chromatin profiling methods, including hypersensitivity to DNase I (Zhu et al., 2015; Yan et al., 2019; Oka et al., 2017) and Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) (Ricci et al., 2020; Tannenbaum et al., 2018). However, relatively little is known about how individual enhancers function mechanistically in plants to generate cell type-specific expression patterns.

To elucidate the regulatory logic through which an individual enhancer drives cell-type specific expression, we use the Arabidopsis sepal giant cell model system. Giant cells affect sepal curvature (Roeder et al., 2012) and are hypothesized to play a role in defense against insect predators and pathogens based on transcriptomic analysis (Schwarz and Roeder, 2016). Giant cells form on the abaxial epidermis of Arabidopsis sepals and are scattered between smaller cells (Figure 1B) (Figure 1B) (Roeder et al., 2010). Giant cells become enlarged through endoreduplication, a specialized cell cycle wherein cells continue to grow and replicate their DNA but fail to undergo mitosis, generating large, polyploid cells (Traas et al., 1998; Roeder et al., 2010; Schwarz and Roeder, 2016). Giant cell fate specification and differentiation are promoted by the epidermal specification pathway, consisting of the Homeodomain leucine zipper (HD-ZIP) Class IV transcription factors ATML1 (*ARABIDOPSIS THALIANA* MERISTEM LAYER 1) and HOMEODOMAIN GLABROUS 11 (HDG11), as well as the cell-to-cell signaling proteins DEFECTIVE KERNEL 1 (DEK1) and Arabidopsis CRINKLY 4 (ACR4) (Roeder et al., 2012; Qu et al., 2014;; Meyer et al., 2017). Stochasticity in gene expression appears to be important for initiating the scattered pattern of giant cells between the smaller cells of the Arabidopsis sepal. Fluctuation of ATML1 protein concentration to a high level during the G2 phase of the cell cycle specifies giant cell fate (Meyer et al., 2017). One outstanding question is how giant cell-specific gene expression patterns are established.

Our previous work described an enhancer region capable of driving giant cell-specific expression in sepals (Roeder et al., 2012). The region was identified based on an enhancer trap line (YJ158) in which the T-DNA was inserted 4.7 kb upstream of At5g17700 (Figure 1A) (Eshed et al., 2004). At5g17700 encodes a MATE (Multidrug And Toxic Compound Extrusion) efflux

family protein. The giant cell-specific enhancer (hereafter referred to as the giant cell enhancer) was functionally defined as the 1024-bp region immediately upstream of the insertion site of the enhancer trap T-DNA; this region was sufficient for giant cell-specific expression in either orientation (Figure 1A) (Roeder et al., 2012). The expression of a reporter driven by the enhancer was regulated by ATML1 and the other members of the giant cell specification pathway (Roeder et al., 2012). In the leaf, the giant cell enhancer activated reporter expression in both giant epidermal pavement cells and leaf margin cells (Roeder et al., 2012; Eshed et al., 2004). The full 4.7-kb upstream region of At5g17700 that includes the giant cell enhancer drove reporter expression in giant cells of young sepals as well as in other cell types (e.g. cells in petals and the style), suggesting that the giant cell enhancer is part of a larger regulatory region for At5g17700 (Roeder et al., 2012).

Here we dissected the ~1-kb giant cell enhancer into smaller regions and determined that the enhancer was comprised of a region promoting broad epidermal expression and a region that limited this expression to giant cells in the sepal. A yeast one-hybrid screen for transcription factors binding to the enhancer fragments identified transcription factors from the TCP (TEOSINTE BRANCHED1, CYCLOIDEA, PROLIFERATING CELL FACTOR), Dof (DNA-binding one zinc finger), and HD-ZIP families, among others. We show that TCP transcription factors activate a broad epidermal pattern of expression. HD-ZIP transcription factors are also essential for activating expression and increased expression of the ATML1 gene encoding an HD-ZIP transcription factor is sufficient to increase enhancer activity in giant cells. The Dof transcription factors broadly repressed expression, leading to giant cell-specific expression.

Results

Dissection of the giant cell enhancer delineates regions that drive broad epidermal expression and limit expression to giant cells

To identify regulatory modules, we dissected the 1024-bp giant cell enhancer into four regions: 1–208 bp (Region 1), 209–449 bp (Region 2), 450–760 bp (Region 3), and 761–1,024 bp (Region 4) (Figure 1, and Supplemental Figure S1). We then cloned fragments containing these regions or combinations thereof into a plasmid containing a minimal cauliflower mosaic virus

(CaMV) 35S promoter (mini35S) driving the expression of a cassette encoding a nucleus-localized yellow fluorescent reporter (3×Venus-N7) (Roeder, et al. 2012). We characterized patterns of Venus reporter fluorescence in T1 transgenic plants from different constructs to assess the capabilities of the different enhancer fragments to activate transcription. The full 1024-bp enhancer (Region 1-2-3-4, contained in the plasmid pAR111) activated expression specifically in giant cells in the sepal (Figure 1B), as previously shown (Roeder et al., 2012). Region 1 alone (in pAR280), Region 4 alone (in pAR256), or Regions 3 and 4 together (Region 3-4, in pAR258) did not activate reporter expression (Figure 1A and Supplemental Figure S2). Region 2 alone (in pAR254), Regions 2 and 3 together (Region 2-3, in pAR261), and Region 2 with Regions 3 and 4 (Region 2-3-4, in pAR260) all drove ubiquitous expression in the epidermis, with no specificity for giant cells (Figure 1A, D-F). Although the reporter was expressed in all epidermal pavement cells for this broad epidermal expression pattern, the expression level was about twice as high in giant cells compared to small cells, as quantified by mean fluorescence intensity of Venus in the nucleus (Figure 1G). Regions 1, 2, and 3 together (Region 1-2-3, in pAR262) were sufficient for driving reporter expression specifically in endoreduplicated giant cells (Figure 1C).

Regions 1 and 2 together (Region 1-2, in pAR257) drove a varied expression pattern in independent T1 transgenic plants, most of which showed partial giant cell specificity, termed intermediate (Figure 2). To further characterize the expression specificity of Region 1-2, we quantified the ratio of small cells to giant cells expressing *Venus* in independent T1 transgenic plants harboring pAR257 (Region 1-2), compared to pAR111 (Region 1-2-3-4), pAR254 (Region 2), pAR260 (Region 2-3-4), pAR261 (Region 2-3), and pAR262 (Region 1-2-3) (Figure 2D). Plants with a lower small:giant cell ratio signifies higher giant cell specificity (few small cells expressing *Venus*); by contrast, a higher small:giant cell ratio signifies broader epidermal expression (many small cells expressing *Venus*). We observed that transgenic plants carrying pAR111 (Region 1-2-3-4) or pAR262 (Region 1-2-3) all have low small:giant cell ratios (giant cell-specific signal; Figure 2D). For pAR254 (Region 2), pAR260 (Region 2-3-4), and pAR261 (Region 2-3), most plants had high small:giant cell ratios (broad epidermal signal), although the range of ratio values was wider than that from giant cell-specific plants. pAR257 (Region 1-2) showed a wide range of expression patterns, as determined by the small:giant cell ratios, but tended to be intermediate in value between the ratios of giant cell-specific and broad epidermal patterns (Figure 2D). Together, these data suggest that Region 1-2 plays an important role in generating

the giant cell specificity of enhancer activity. Region 2 is sufficient to produce a broad epidermal expression, while Region 1 appears to limit expression to giant cells. In addition, Regions 3 and 4 play a small role in enhancing giant cell specificity when present in combination with Region 1, but are not sufficient to increase giant cell specificity without Region 1 (Figure 2D). Therefore, we focused on Regions 1 and 2.

TCP transcription factors promote activity of the enhancer

To identify transcription factors interacting with Regions 1 and 2 of the enhancer, we performed a yeast one-hybrid screen with an arrayed library of 1,956 transcription factors from Arabidopsis (Pruneda-Paz et al. 2014). Through this plate-based technique, we expressed each of the of 1,956 transcription factor genes individually in a yeast strain carrying a construct containing a fragment of the enhancer upstream of the firefly luciferase (*LUC*) reporter gene. We quantified luciferase reporter activity as a readout for transcription factor binding in yeast. We tested three constructs containing Region 1 (1–208 bp), Region 2 (209–450 bp), and a 100-bp junction region (167–266 bp) overlapping the edges of these regions (Supplemental Figure S1); the 100-bp junction region was intended to catch interactions that might be disrupted at the boundary between Region 1 and Region 2. We identified 111 high-confidence (Table 1) and 43 low-confidence interactions (Supplemental Table S1) across all three assays (Supplemental Datasets S1-S3). We paid particular attention to transcription factors that specifically bound to a single region, as they might be driving the distinct expression patterns conferred by each region.

In the screen, several Class II CINCINNATA-like TCP transcription factors (TCP2, TCP3, TCP4, TCP10) interacted with Region 2 and the 100-bp junction between Regions 1 and 2, but not with Region 1. To test for regulation of the enhancer by Class II TCPs, we separately crossed the full 1-kb enhancer reporter line pAR111 and the Region 2 enhancer reporter pAR254 to the *jaw-1D* mutant (Figure 3; Palatnik et al., 2003). The *jaw-1D* line overexpresses *MIR319a*, whose derived microRNA targets *TCP2*, *TCP3*, *TCP4*, *TCP10*, and *TCP24* transcripts, thus leading to downregulation of these *TCP*s (Palatnik et al., 2003). In pAR111 × *jaw-1D* F1 plants, sepals had fewer giant cells expressing the reporter (Figure 3B and 3G), compared to sepals from pAR111 crossed to wild type (Col-0) (pAR111 × Col-0 F1) (Figure 3A and 3G). Knockdown of *TCP*s only affected the expression of the pAR111 reporter and not the presence of giant cells in sepals, as identified by morphology (Figure 3B, PI stain). In pAR111 × *jaw-1D* F1 sepals, the number of giant

cells was similar to that in pAR111 × Col-0 F1 sepals (18.6 ± 2.3, mean ± SD, n = 5 in pAR111 × *jaw-1D* F1; 17 ± 4.1, mean ± SD, n = 5 in pAR111 × Col-0 F1; not significant by t test); however, most of the giant cells in *jaw-1D* × pAR111 F1 sepals showed no detectable expression of the pAR111 reporter (Figure 3B and 3G). Likewise, plants heterozygous for both pAR254 and *jaw-1D* (pAR254 × *jaw-1D* F1) had far fewer giant cells and small cells with Venus fluorescence compared to pAR254 × Col-0 F1 plants (Figure 3C 3D and 3G and3H).

To assess the functional importance of binding by TCP transcription factors to the giant cell enhancer, we looked for evidence of TCP transcription factor binding to the enhancer DNA. To identify potential TCP binding motifs in the enhancer sequence, we used the motif-based sequence analysis tool Find Individual Motif Occurrences (FIMO) (Grant et al. 2011), which revealed two putative GGACC motifs in Region 2 (Supplemental Figure S3). Mutating both GGACC motifs in the Region 2 reporter (named pBR67) or the full 1-kb enhancer reporter (named pBR63) led to a nearly complete loss of Venus signal (Figure 3E-3H). To verify TCP interaction with the enhancer in vitro, we performed electrophoretic mobility shift assays (EMSAs) with the DNA binding domain of TCP4 on these putative binding motifs. Recombinant purified TCP4 DNA binding domain bound to both GGACC motifs (Figure 3I). Thus, our results suggest that binding of TCPs to Region 2 is critical for this region to activate expression.

Dof transcription factors repress expression from the enhancer by lowering the expression level and limiting it to giant cells.

We next searched for transcription factor binding motifs in Region 1 responsible for repressing transcription, thereby potentially giving rise to giant cell specificity. Our yeast one-hybrid screen showed that several Dof transcription factors bind to Region 1 and the 100-bp junction construct, but not to Region 2 in yeast. Based on transcription factor binding motif predictions at the AthaMap website (http://www.athamap.de), a 24-bp region located 14 bp before the 3' end of Region 1 contained three Dof transcription factor binding motifs (AAAG) (Sani et al., 2018; Noguero et al., 2013). We generated a reporter construct containing Region 2 and extending 40 bp into the end of Region 1, which contained all three potential Dof binding motifs (Region 2+Dof in pAR307, Figure 4 and Supplemental Figure S3). As a control, we also generated a construct containing Region 2 and 17 bp at the 3' end of Region 1, which excluded the Dof binding motifs (pAR308, Figure 4A and Supplemental Figure S4). Quantifying the expression

patterns of these constructs with small:giant cell ratios showed that pAR307 (Region 2+Dof) retains most of the giant cell-specific expression previously observed with Region 1-2 with the reporter pAR257 (Figure 4B). Removing the Dof binding motifs region (pAR308) led to a significant increase in the small:giant cell ratio, indicating a shift to an epidermal expression pattern (Figure 4B). These results indicate that the region containing the three Dof binding motifs is important in generating the giant cell-specific expression pattern. Site-directed mutagenesis of these three putative Dof binding motifs in the 40-bp region (Region 2-Dof mutant, pLH166) recapitulated the epidermal expression pattern seen in pAR308 (Figure 4B), indicating that the Dof binding motifs confer specificity. We also detected a trihelix binding motif in the 24-bp Dof binding motif region, but mutation of this putative trihelix transcription factor binding motif did not alter the expression pattern of the reporter (Supplemental Figure S4).

While we were dissecting the enhancer into regions (Figures 1, 2), we noticed that different constructs required different laser power output to generate the same fluorescence intensity of the reporter in giant cells (as determined by giant cell signal saturation). We recorded the laser intensity necessary to reach the same giant cell fluorescence intensity. pAR257 (Regions 1-2) needed a much higher laser intensity than pAR254 (Region 2 only) to saturate fluorescence in giant cells (Figure 4C), suggesting that Region 1 reduces the ability of Region 2 to drive reporter expression. We also investigated the expression levels driven by the enhancer fragments in pAR307, pAR308, and pLH166 by quantifying the laser intensity required for fluorescence to reach saturation in giant cells. The shift of giant cell-specific expression to epidermal expression was correlated with an increase in reporter signal strength (as indicated by a lower laser power output; Figure 4C). These results are consistent with a model in which this 40-bp Dof motif-containing region of the enhancer confers transcriptional repression.

Based on DAP-seq analysis, several Dof transcription factors bind to the 24-bp Dof binding motif of Region 1 in native genomic DNA (Supplemental Figure S5). Those Dofs showing the highest binding affinity included: *AtDOF2.2* (At2g28810*)*, *HIGH CAMBIAL ACTIVITY2* (*HCA2*, At5g62940*)*, and *AtDOF5.8* (At5g66940) (Yanagisawa, 2002; Guo et al., 2009). We confirmed that the DNA binding domain of DOF2.2 binds to the putative Dof binding motifs in Region 1 in vitro by EMSA (Figure 4D). We also overexpressed each *Dof* gene under the constitutive 35S promoter in a background harboring the pAR111 full length giant cell enhancer reporter to assess the influence of Dofs on enhancer activity (Figure 5). Overexpression of each *Dof* gene strongly

decreased pAR111 expression in sepal giant cells, as measured by Venus fluorescence intensity (Figure 5D-5L) compared to the sepal giant cells from the pAR111 control alone (Figure 5A-5C), indicating that Dof transcription factors repress the transcriptional activity derived from the giant cell enhancer. Overexpression of each *Dof* gene also severely inhibited sepal development, resulting in shorter, narrow sepals in *35S:AtDOF2.2* and *35S:HCA2* plants (Figure 5D-5E and 5G-5H), and longer, narrow sepals in *35S:AtDOF5.8* plants (Figure 5J-5K). Even in these morphologically altered sepals, giant cells formed, based on size and morphology. A few of these giant cells expressed low levels of the reporter, indicating that repression of pAR111 expression was not due to the absence of giant cells, but instead to repression of expression (arrowheads in Figure 5F, 5I, and 5L). Thus, our results suggest that binding of Dofs to Region 1 reduces overall reporter expression to confer giant cell-specific expression (Figures 4 and 5).

Modulating the expression of HD-ZIP transcription factor genes modulates expression driven by the enhancer in giant cells

We noticed that giant cells showed higher Venus fluorescence intensity than small cells in our Region 2 reporter lines (Figure 1G). This higher expression in giant cells may contribute to the cell type specificity of the entire enhancer. Hypothetically, if Region 2 drove 2x expression in giant cells and 1x expression in small cells, and Region 1 subtracted 1x expression from both giant cells and small cells, then 1x expression remains in giant cell (2x - 1x = 1x) and no (1x – 1x = 0x) expression remains in small cells (Figure 6N). This model generates cell type specificity. Based on this hypothetical model, we asked how Region 2 of the enhancer might drive higher expression in giant cells relative to small cells. Our yeast one-hybrid analysis showed that members from the HD-ZIP transcription factor family bind to Region 2 and the 100-bp junction between Region 1 and 2, but not to Region 1 (Table 1). HD-ZIP Class IV transcription factors are typically restricted to the L1 layer (the cell layer that makes up the epidermis) and act to specify epidermal cell type identity (Nakamura et al., 2006). Furthermore, HDG11 was shown to have a small role in promoting giant cell formation (Roeder et al., 2012). ATML1, an HD-ZIP Class IV transcription factor not included in the yeast one-hybrid assay, is required for both giant cell patterning and specification of epidermal identity (Abe et al., 2003; Roeder et al., 2012; Meyer et al., 2017). Thus, we focused on ATML1 as a representative for the role of HD-ZIP Class IV transcription factors such as HDG11 on the enhancer. We previously showed that overexpression of *ATML1* was

sufficient to create ectopic giant cells covering the sepal and that these ectopic giant cells expressed the pAR111 full-length giant cell enhancer reporter (Meyer et al., 2017), suggesting that ATML1 may be involved in activating the enhancer. Enhancer activity is downregulated but not completely absent in *atml1* mutants, consistent with a role for ATML1 in enhancer activation (Roeder et al., 2012). The canonical binding motif of ATML1 and other HD-ZIP Class IV transcription factors, the L1 box (TAAATGCA) (Abe et al., 2001; Nakamura et al., 2006), is not present in the enhancer sequence. However, FIMO analysis detected a putative HD-ZIP binding motif (ATAATTAATTA) in Region 2. Mutating this motif led to a drastic reduction in Venus fluorescence in both the Region 2 reporter (pBR69) and the Region 1-2-3-4 reporter (pBR65; Figure 6A-6D), indicating that this putative HD-ZIP binding motif is important for reporter expression across the entire epidermis. Interestingly, we noticed that at very high laser power outputs, we were able to observe very weak Venus fluorescence (Supplemental Figure S6), leading us to hypothesize that HD-ZIP transcription factors can modulate the expression levels driven by the enhancer. We performed an EMSA, which confirmed that the DNA binding domain of ATML1 can bind to the putative HD-ZIP motif on Region 2 in vitro (Figure 6M).

Although *ATML1* is expressed in all epidermal cells, we previously showed that ATML1 protein was more highly concentrated in giant cells than in small cells (Meyer et al. 2017). Based on this observation, we hypothesized that differential *ATML1* expression levels might contribute to higher reporter expression in giant cells of Region 2 reporters. To test whether changing *ATML1* expression levels caused corresponding changes in the intensity of reporter signals, we generated heterozygous plants harboring the reporter constructs and either *ATML1* overexpression or loss-of-function lines to modulate *ATML1* expression levels. Accordingly, we separately crossed the Region 2 reporter pAR254 and the Region 1-2-3-4 reporter pAR111 to the *ATML1* overexpression line *PDF1pro:Flag-ATML1*, to the *ATML1* knockout mutant *atml1-4*, and to wild-type Col-0 as a control and analyzed Venus signal intensity in giant cell nuclei of the resulting F1 plants. Both reporters showed a significant increase in signal intensity when crossed to the *ATML1* overexpression lines compared to the control (Figure 6H and 6L), suggesting that increasing *ATML1* expression levels can increase reporter expression in giant cells. We also tested the converse. The *atml1-4* mutation is semi-dominant (Roeder et al., 2012; Meyer et al. 2017), making *atml1-4*/*ATML1* heterozygous plants a sensitized background: we determined that reporter intensity decreases in the giant cell nuclei of *atml1-4/ATML1* sepals (Figure 6H and 6L). Our

results are consistent with our hypothesis that higher levels of ATML1 in giant cells binding to Region 2 drives higher *Venus* expression in giant cells compared to small cells. Further, this result is consistent with our model that Dofs bind to Region 1, repressing expression and lowering the overall expression levels such that only giant cell expression remains, thus achieving cell type specificity by the full enhancer (Figure 6N).

Discussion

In this study, we used giant cells as a model system to investigate how a cell type-specific expression pattern is achieved. We present the dissection of an enhancer region driving a giant-cell-specific expression pattern (Supplemental Figure S10). Our results support a modular conceptual model of enhancer organization (Figure 6N). In this model, Region 2 is activated by TCP and HD-ZIP transcription factors and drives expression in all epidermal cells. We previously showed that the HD-ZIP transcription factor gene *ATML1* is more highly expressed in giant cells than small cells. Here we showed that ATML1 activity can modulate the intensity of the reporter expression driven by the enhancer in giant cells. Thus, our model posits that higher concentrations of HD-ZIP proteins in giant cells relative to small cells drive higher levels of expression in giant cells. In the model, Region 1 is bound by Dof transcription factors to repress expression, thus limiting expression to giant cells (Figure 6N). This repression contributes to cell type specificity due to the differences in expression level between giant cells and small cells, which allows only the giant cell expression to remain after repression. We note that the model does not require either the activation or the repression to be completely cell type-specific, yet the combination results in giant cell specificity. In the real biological enhancer, it remains to be determined whether additional factors contribute to specificity.

Our rough enhancer dissection indicated that transcription factors interacting with Region 2 were candidates for activating broad epidermal expression. The results from the large-scale yeast one-hybrid assay suggested that 37 transcription factors showed high-confidence interaction with this region, including proteins from the TCP and HD-ZIP class IV families (Table 1). Although we focused on TCPs, Dofs, and HD-ZIPs in this analysis, it is possible that the other transcription factors have interesting roles in regulating expression driven by the enhancer. The TCPs identified from the screen (TCP2, TCP3, and TCP4) are all in the class II CINCINATA subfamily (Martín-Trillo and Cubas, 2010). Class II protein CINCINNATA (CIN) controls leaf surface curvature in

snapdragon (*Antirrhinum majus*) by making cells more sensitive to an arrest signal, inhibiting proliferation (Nath et al., 2003; Crawford et al., 2004). As proteins in this subfamily act to repress cell proliferation, it is plausible, although untested, that these factors serve to coordinate activation of expression at the enhancer with the transition from proliferation to differentiation, since the giant-cell enhancer is only active after cells have stopped dividing (Roeder et al., 2012). Interestingly, the class II protein TCP4 regulates both the density and branching of trichomes, another highly endoreduplicated cell type in Arabidopsis (Vadde et al., 2017; 2019). Class I TCP15 also regulates endoreduplication in Arabidopsis (Zi-Yu et al., 2011). Moreover, specific repression of TCP15 by fusing to an EAR repressor domain was reported to result in a loss of giant cells on the abaxial sepal epidermis (Uberti-Manassero et al., 2011).

In our conceptual model, the higher concentrations of HD-ZIP Class IV transcription factors in giant cells relative to small cells promoted higher reporter expression in giant cells driven by Region 2 of the enhancer (Figure 6N). In particular, overexpressing the HD-ZIP Class IV transcription factor gene *ATML1* increased Venus intensity driven by the enhancer in giant cells (Figure 6H and 6L). There was also a decrease in Venus intensity when crossed into the *atml1-4* knockout background. During sepal development, ATML1 levels fluctuate stochastically, leading to giant cell specification when ATML1 levels surpass a threshold in the G2 phase of the cell cycle. The resulting giant cells have higher total ATML1 levels than small cells (Meyer et al., 2017). Higher ATML1 protein accumulation may be responsible for the increased giant cell expression driven by enhancer region 2 in these cells once they develop. ATML1 also forms a weak positive feedback loop with itself (Abe et al., 2001), which may help maintain higher expression levels of *ATML1* and, subsequently, the expression driven by the enhancer in giant cells. Interestingly, the canonical L1 box that ATML1 binds to is not present in the giant cell enhancer. Previous work showed the L1 box as not necessary for epidermal expression pattern or the positive feedback loop seen with ATML1 (Takada et al., 2013; Takada and Jürgens 2007). Our study revealed that ATML1 can bind to the HD-ZIP motif present in Region 2 of the enhancer, although it is not a canonical L1 box. It is possible that other HD-ZIP transcription factors not studied in this work are active at this site, such as ANTHOCYANINLESS (ANL2) or HDG1, which appeared in the FIMO search that identified the HD-ZIP motif on the giant cell enhancer.

Giant cell specificity arises through the repressive activity of Dof transcription factors (Figure 6N). Yeast one hybrid, EMSA, and DAP-seq results suggested that Dof transcription

factors bound to Region 1 and overexpression analyses and enhancer dissection suggested that Dofs repressed transcription to create giant cell specificity. Region 1 alone did not activate any expression, suggesting that it acts solely as a repressor and detecting its output requires the activators bound to Region 2. Dof transcription factors bind to AAAG motifs in DNA through a single N-terminal C_2C_2-zinc-finger like domain (Noguero et al., 2013). Dof proteins interact more tightly with DNA when two proximal binding motifs are present than they do with only one binding motif (Sani et al., 2018), which is consistent with the three nearby putative Dof motifs identified in Region 1. We did not identify the specific Dof or Dofs that are responsible for repression when binding to the giant cell enhancer in wild type sepals because there are 36 Dof family members in Arabidopsis and they tend to function redundantly (Yanagisawa, 2002; Noguero et al., 2013; Gupta et al., 2015; Yanagisawa, 2016). Overexpression of each of the three *Dof* genes we tested was sufficient to repress expression driven by the giant cell enhancer. *Dof* gene overexpression also altered sepal development, making sepals narrow, and in most cases shorter, suggesting that these *Dof* genes are important for sepal morphogenesis, not just regulation of this enhancer. Although our data suggest that Dofs act as repressors in the context of the giant cell enhancer, Dofs can act as either activators or repressors in different contexts (Noguero et al., 2013). For example, the Dof transcription factor OBF BINDING PROTEIN 1 (OBP1) promotes cell cycle entry by directly activating CYCLIN D3;3, which is interesting because small cells divide mitotically (Skirycz et al., 2008). The repressive activity conferred by the Dofs may be mediated by co-repressors. For example, in seeds, DOF3.2 binds to the DELLA co-repressor REPRESSOR OF GA-like 2 (RGL2) and DOF AFFECTING GERMINATION 1 (DAG1) binds to the DELLA corepressor GA INSENSITIVE (GAI) to repress seed germination in the absence of gibberellins (Ravindran et al., 2017; Boccaccini et al., 2014; Ruta et al., 2020). Of the Dofs we tested, the functions of DOF2.2 (At2g28810) remain unknown (Yanagisawa, 2016). HCA2/DOF5.6 (At5g62940) promotes the development of the vascular cambium and consequently the radial growth of the root (Guo et al., 2009; Miyashima et al., 2019). DOF5.8 (At5g66940) regulates ANAC069 in response to abiotic stress (He et al., 2020). In the future, it will be interesting to determine which Dofs are required for the repressive activity of Region 1 in wild-type sepals, since our overexpression analysis showed that all three Dofs we tested were sufficient to repress reporter expression, despite coming from different clades (Yanagisawa, 2002; 2016). Given that multiple Dofs, including the three we tested, bind to this Region 1 in native

genomic DNA in the DAP-seq database, we expect multiple Dof proteins to act redundantly in this role.

The giant cell enhancer includes a previously identified conserved noncoding sequence (Haudry et al., 2013). Enhancers tend to be conserved across closely related species because selection preserves their function, whereas nonfunctional intergenic sequences are not under selection and diverge faster. This conservation has been used to identify enhancer sequences in non-coding DNA (Pennacchio et al., 2006). Alignment of Region 1-2 enhancer sequences from seven species across the Brassicaceae family revealed that both TCP binding motifs are conserved in lyrate rockcress (*Arabidopsis lyrata*), hairy bittercress (*Cardamine hirsuta*), cabbage (*Brassica rapa*), London rocket (*Sisimbrium irio*), saltwater cress (*Eutrema salsugineum*), and salt cress (*Schrenkiella parvula*) (Supplemental Figure S7). The HD-ZIP motif is also conserved in all but *S. irio*, and two of the three Dof motifs were conserved in most species (Supplemental Figure S7). In the future, it will be interesting to determine whether these enhancers produce similar expression patterns in these other species.

Although Regions 1 and 2 had the most notable effects on the expression driven by the enhancer and were the focus of this study, Regions 3 and 4 had minor and often statistically nonsignificant effects on the expression pattern as well. Removing Regions 3-4 from the enhancer caused an apparent increase in the small:giant cell ratio and therefore a slight loss of specificity, although this effect was not statistically significant (pAR257 Region 1-2; Figure 2D). Nevertheless, Region 1-2 had a very broad range of expression patterns, with some T1 plants close to producing a giant cell-specific expression pattern and other T1 plants close to an epidermal expression pattern (Figure 2A-C), suggesting that Regions 3-4 may function to stabilize the enhancer's expression pattern. FIMO analysis identified potential Dof binding motifs in Region 3 and, if the output threshold was raised to 0.001, FIMO also identified potential TCP motifs in Region 4. These additional transcription factor binding motifs may therefore be responsible for stabilizing the expression pattern.

Previous studies in animals have often shown that spatial restriction of enhancer activity is generated by broad activation combined with localized repression (Spitz and Furlong, 2012; Davidson, 2010). For example, in Drosophila the *even-skipped* (*eve*) stripe two enhancer is activated broadly by Bicoid and Hunchback and narrowed to a precise stripe through repression by Giant on the anterior and Krüppel on the posterior sides (Stanojevic et al., 1991). Similarly,

broad activation of the giant cell enhancer occurs in all epidermal cells through the activity of Class II TCP and HD-ZIP class IV transcription factors. Our evidence suggests that HD-ZIP Class IV transcription factors, while not constrained to a single cell type, contribute to generating a pattern with higher expression in giant cells and lower expression in small cells. The Dof transcription factors repress expression by lowering levels until the enhancer no longer activates expression in small cells, and only expression in giant cells remains detectable (Figure 6N). Thus, our model suggests that cell type specificity can emerge from the combination of three modules, none of which is itself completely cell-type specific.

Methods

Plant Growth

Arabidopsis (*Arabidopsis thaliana*) Col-0 plants were grown on Lambert Mix LM-111 soil in Percival growth chambers at 22°C with constant illumination provided by Philips 800 Series 32 Watt fluorescent bulbs (f32t8/tl841) (~100 μmol m^{-2} s^{-1}). Seeds were sown on soil and stratified for about two days at 4°C in darkness before release in the growth chambers.

Giant cell enhancer dissection

Cloning of the full-length 1,024-bp enhancer construct pAR111 and verification that it was sufficient to drive giant cell-specific expression in either orientation was previously described (Roeder et al., 2012). This 1,024-bp region was divided into four arbitrary regions with Region 1 falling closest to the downstream gene At5g17700 and farthest from the original enhancer trap insertion YJ158 (Figure 1). Enhancer fragments were tested for their activity in the orientation they would have relative to downstream gene At5g17700 such that Region 1 is closest to the reporter gene.

We first cloned a Gateway destination reporter vector (pSL12) so that we could rapidly clone and assay the reporter expression patterns driven by fragments of the giant cell enhancer. pSL12 contains a Gateway cassette (attR1 Cm^R ccdb attR2) upstream of the –60 minimal promoter from the CaMV 35S promoter and the sequence encoding a 3×Venus-N7 super-bright yellow-fluorescent nucleus-localized reporter in the pMLBart binary vector backbone. pSL12 confers Basta resistance in plants and spectinomycin resistance in bacteria. To create pSL12, the Gateway cassette in which the *Not*I restriction site had been removed was amplified with primers oAR505

and oXQ6 and cloned into pGEM-T Easy (Promega) to generate pSL10. The Gateway cassette was cut from pSL10 with *Xho*I and *Kpn*I and cloned into –60 3×Venus-N7-BJ36 (Roeder et al., 2012) to create pSL11. pSL11 was cut with *Not*I and the Gateway –60 3×Venus-N7 fragment was cloned into pMLBart to create pSL12.

Each enhancer fragment was amplified by PCR with Pfu Ultra II (Agilent) or Phusion (NEB) using the BAC clone MAV3 or Col-0 genomic DNA as template using the primers specified in Supplemental Table S2. The primers sequences are listed in Supplemental Table S3. The PCR products were cloned into pENTR D TOPO (ThermoFisher) according to the manufacturer's instructions to create the entry clones listed in Supplemental Table S2. LR reactions (ThermoFisher) between the entry clones and pSL12 generated the final constructs listed in Supplemental Table S2 and Supplemental Figure S10. Putative Dof and TCP motifs were mutated through changes in the primer sequences. All constructs were verified by sequencing.

These constructs were transformed into Columbia (Col-0) plants through Agrobacterium (*Agrobacterium tumefaciens*)-mediated floral dipping (Clough and Bent, 1998) and were selected for Basta Resistance. For most constructs, the expression patterns from approximately 20 T1 transgenic plants were analyzed. Note that the original pAR111 transgenic plants described in (Roeder et al., 2012) were in the Landsberg *erecta* (L*er*) accession, but for this project pAR111 was transformed into Col-0. Transformation of the empty vector pSL12 into plants generated no expression (Supplemental Figure S2D). As typical with random insertion of the T-DNA into the genome (Schubert et al., 2004), we saw differences in expression level and sometimes pattern between different T1 plants, as characterized in the results. However, within a single plant the expression pattern was similar across sepals. Likewise, the expression pattern was consistent in subsequent T2 and T3 generations for those lines examined.

Confocal microscopy

To minimize morphological or expression variability caused by different plant age, the 12th to 25th flowers on the main stem were used for observation (Hong et al., 2016). Mature sepals (stage 14 according to (Smyth et al., 1990)) were dissected from flowers with tweezers and needles and stained for 15 min in propidium iodide (PI; 0.15 mg/mL in water), mounted on slides in 0.01% (v/v) Triton X-100 under a cover slip, and imaged with a Zeiss 710 confocal laser scanning microscope. A 514-nm excitation laser was used to excite both fluorophores (Venus and PI). The

3×Venus-N7 enhancer reporter signal was collected between 519 nm and 566 nm, while the PI signal was collected between 603 nm and 650 nm. Tiled images were taken using 10X (Plan-APOCROMAT NA = 0.45 air) or 20X (Plan-APOCROMAT NA = 1.0 water dipping) objective lenses. For the experiments in Figure 4C, the laser power output was modulated to achieve saturation of the reporter in a few of the brightest giant cell nuclei (the other settings including gain were kept the same). This setup allowed us to compare reporter expression between cells while keeping the expression level in giant cells relatively constant. By contrast, Supplemental Figure S6 shows the same two sepals imaged with different laser power outputs (3% and 40% as indicated in the figure), to detect very low levels of expression. For signal intensity experiments with *ATML1* overexpression and *atml1-4* mutants, the laser power output was set constant for all lines to keep the brightest giant cell nuclei just under saturation in order to detect differences in brightness between lines (Figure 6E-L). For most images, the maximum intensity projection is shown. For Figure 5C,F,I, and L the image was volume-rendered in MorphoGraphX and subsequently Auto Contrast was used in Adobe Photoshop to visualize the epidermal cell layer without the underlying layers.

Quantitative image processing

FIJI (https://fiji.sc) with the Costanza plugin (http://home.thep.lu.se/~henrik/Costanza/) was used to quantify the fluorescence intensity and size of nuclei expressing pAR254 (Figure 1G). In the images used, the fluorescence intensity of the reporter in the nuclei was not saturated. The .lsm stack image from the microscope was opened in FIJI. The color of Channel 1 (with the reporter expression) was converted to gray scale using the Channels tool. The channels were split using Split Channels. The maximum intensity projection of Channel 1 (reporter expression) was created with Z project. The maximum intensity projection image was analyzed with the Costanza plugin with the following settings. In the general menu, “Mark intensity plateau with single maximum” was checked as “Mark cell centers. Marker pixel radius: 3”, Display basins of attractions (BOA)”, and “Display basins of attractions according to measured intensity.” In the pre-processor queue, “Background extraction” with an intensity threshold of 20 was executed first, followed by “Mean filter” with radius 0.1 and number of times 10. In the post-processor, “BOA remover” with a size threshold of 10 and an intensity threshold of 10 was executed first, followed by a “BOA merger” with a radius of 15. ImageJ stack calibration was used for scaling. The results were analyzed in

Microsoft Excel. Giant cells have enlarged nuclei, so a threshold of 100 μm^2 was used to separate giant cell from small cell nuclei.

To count the number of nuclei in Figure 3C, we kept the image settings including the laser power and gain the same when imaging the sepals for both pAR111 × Col-0 F1 and pAR111 × *jaw-1D* F1. Images were processed with maximal intensity projection using the Zeiss Zen software and nuclei that showed saturation of the reporter signals were counted as giant cell nuclei.

Small cell nuclei and giant cell nuclei were counted using FIJI (https://fiji.sc) version 1.53c for calculating small:giant cell ratios. The ImageJ macro script Weka_Macro_Script_v2.ijm (https://github.com/RoederLab/Giant_Cell_Enhancer_Project.git) was used to load the images into the FIJI plugin "Trainable Weka Segmentation" v3.2.34, to classify nuclei as either small or giant cell nuclei using the classifier file (classifier_GCE_nuclei_T6.model) and training data (GCE_training_segmentation_T6.arff). The macro script then segmented the nuclei from the classified images and counted them. Results were copied and pasted into a text file, which was then reformatted into a CSV file either manually or with the python script GCE_segmentation_results_parsing.py. The ratio of small cells to giant cells was calculated for each sepal in the CSV file, which was then loaded into R for statistical analysis and plotting with the R script Plotting_script_improved.R.

Images with significant background noise that caused the Trainable WEKA segmentation program to call nuclei outside of the sepal were processed with GIMP. The lasso tool was used to trace the sepal margin, the selection was inverted to select the area outside the sepal, and the fill tool was used to turn the background black. The edited images were then re-classified, segmented, and counted as described above.

The procedure for the signal intensity calculations in *ATML1* overexpression, *atml1-4* mutant, and Col-0 F1 plants (Figure 6H and 6L) was similar to that for counting nuclei, using the ImageJ macro script (Weka_cell_intensity_quantification.ijm) to segment giant cell nuclei using the classifier (Classifier_GCE_intentsity_ATML1_crosses_v3.model) and training data (Data_GCE_intentsity_ATML1_crosses_v3.arff). Using the same imageJ macro script, the intensity of all giant cells in a sepal was calculated in arbitrary units based on pixel intensity ranging from 0 to 255. Results were copied and pasted into a text file and reformatted into a CSV file either manually or with the python script GCE_intensity_segmentation_results_parsing.py,

then analyzed in R with the script GCE_intensity_graphs.R. All scripts and data files can be found in our GitHub repository (https://github.com/RoederLab/Giant_Cell_Enhancer_Project.git).

Statistics

Two-tailed students t-tests with unequal variances were used for pairwise comparisons. One-way Analysis of Variance (ANOVA) followed by Tukey's Honest Significant Difference (Tukey HSD) was used for comparisons between three or more samples. The distribution for giant cell intensity (Figure 6H and 6L) was highly right skewed, so to meet the normality assumption for ANOVA these data were inverse-transformed before calculating statistics. ANOVA was performed in R (version 4.1.0) using the aov() function and Tukey's HSD was performed in R using the HSD.test() function from the package 'agricolae' version 1.3-5. T-tests were performed in R or in Excel.

High-throughput yeast one-hybrid screen

Yeast one-hybrid screens were conducted with a nearly genome-wide transcription factor collection arrayed in 384-well plates (Pruneda-Paz et al., 2014) according to the protocol in (Li et al., 2019) using a luciferase reporter (Bonaldi et al., 2017). Bait plasmids were created through LR reactions of entry clones containing Region 1, Region 2 or the 100-bp junction region overlapping the edges of Regions 1 and 2 into the pY1-gLUC59_GW yeast reporter plasmid (Table S4). The resulting constructs were integrated into the *URA3* locus on the chromosome of yeast strain YM4271.

Genetic crosses

The *jaw-1D* mutant (CS6948) was ordered from the Arabidopsis Biological Resource Center (ABRC). The homozygous *jaw-1D* mutant was crossed to pAR111 plants and to pAR254 plants. The expression patterns were analyzed in the F1 plants heterozygous for both the reporter and the *jaw-1D* mutant. The expression patterns were compared to the F1 generation of the cross between each reporter and Col-0 as a control. pAR111 plants and pAR254 plants were also crossed to the *ATML1* knockout line *atml1-4*, the *ATML1* overexpression line *PDF1pro:Flag-ATML1*, and to Col-0 as a control. Venus intensity was analyzed in the F1 plants derived from each cross.

Dof overexpression

Entry clones containing individual coding sequences (cDNAs) from Dof family genes were ordered from ABRC: TOPO-U09-G04 (DF2.2 At2g28810), TOPO-U20-E12 (HCA2 At5g62940), and TOPO-U13-H01 (DF5.8 At5g66940). Through LR reactions, these were recombined into the binary destination vector pK7WG2 to make pLH160 *35S:DOF2.2* (At2g28810), pLH162 *35S:HCA2* (At5g62940), and pLH163 *35S:DOF5.8* (At5g66940). These constructs were transformed into Col-0 plants harboring the pAR111 reporter through Agrobacterium-mediated floral dipping and were selected based on kanamycin resistance.

Flower pictures

Images of flowers were taken on a Zeiss Stemi 2000-C stereomicroscope with a Cannon Powershot A640 digital camera.

DAP-seq database analysis

DAP-seq data from the C_2C_2-Dof transcription factor families were analyzed from the Plant Cistrome Database (http://neomorph.salk.edu/dev/pages/shhuang/dap_web/pages/index.php (O'Malley et al., 2016)) using the Genome Browser to examine the giant cell enhancer region on chromosome 5 coordinates 5,837,111 bp to 5,838,135 bp.

Identifying putative transcription factor binding motifs

We obtained binding motifs for TCP, DOF, and HD-ZIP members from CIS-BP build 2.00 (Catalog of Inferred Sequence Binding Preferences, http://cisbp.ccbr.utoronto.ca/) (Weirauch et al. 2014) by searching "*Arabidopsis thaliana*" as the model organism and searching by domain type for TCP, Dof, and Homeodomain. All other fields were left blank or in their default state. We downloaded the motif data for all available TCPs, all available Dofs, and the following Homeodomains: ANL2, HDG1, HDG7, HDG11, ATML1, and PDF2. We then converted the CIS-BP files to basic MEME format using a homemade python script, Cisbp2MEME.py. We then used these MEME format files and the full giant cell enhancer sequence with a command line version of Find Individual Motif Occurrences (FIMO), a part of the MEME suite version 5.3.3, to search for potential transcription factor binding motifs using default parameters for all searches except for TCP motifs, where we changed the output threshold to 0.001.

Alignment of the giant cell enhancer from Brassicaceae species

A conserved noncoding sequence (chromosome 5: 5,837,330 bp to 5,837,417 bp) was previously identified among Brassicaceae species (Haudry et al., 2013) and overlapped with Region 2 of the giant cell enhancer (http://mustang.biol.mcgill.ca:8885/cgi-bin/hgGateway). To examine sequence conservation of Regions 1 and 2 of the enhancer and the putative Dof and TCP binding motifs, we aligned the corresponding sequences from *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Cardamine hirsuta* (CoGe id 36106 Cardamine hirsuta v1 unmasked Chromosome 6 18,239,653-18,239,154), *Brassica rapa* (COGE id 24668 Brassica DB Chr unmasked v1.5 chromosome A10 11,517,566-11,517,500 and A02 3,259,372-3,259,603), *Eutrema parvulum* (formerly *Thellungiella parvula* 12384 UIUC unmasked v2 chomosome 6-6 5,884,609-5,883,951), *Eutrema salsugineum* (Id 19492 JGI unmasked 6,007,159-6,007,397), *Sisimbrium irio* (Id 20245 VEGI unmasked vVEGI 2.5 Chromosome scaffold_57 2,159,183-2,159,617). All sequences were checked for falling in syntenic regions upstream of a homolog of At5g17700. The sequences were aligned with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and the alignments were formatted and displayed with Boxshade (https://embnet.vital-it.ch/software/BOX_form.html). Alignments were adjusted slightly by hand around gaps and the ends of sequences were trimmed or extended to give the best alignment (Supplemental Figure S7).

Electrophoretic mobility shift assay (EMSA)

Truncated versions of TCP4 (amino acids [aa] 1 to 131), DOF2.2 (aa 90 to 148) and ATML1 (aa 1 to 195) containing the DNA binding domains fused to maltose-binding protein (MBP) were used in the EMSAs (Supplemental Figure S8). The MBP–ΔTCP4, MBP–ΔDOF2.2 and MBP–ΔATML1 fusion constructs were developed by cloning the corresponding parts of the coding sequences into the pMAL-c5X vector (New England Biolabs). The fusion proteins were produced in *Escherichia coli* strain Rosetta 2(DE3) and purified using an ÄKTA pure™ 25 protein purification system (Cytiva). Specific probes encompassing TCP4, DOF2.2 or ATML1 binding motifs (Supplemental Table S5) were labeled at the 3' end with biotin and annealed to form the double-stranded DNA probe before use. As negative controls, probes with mutated TCP4, DOF2.2 or ATML1 binding motifs were also prepared (Supplemental Figure S9). Unlabeled probes were used at 50- and 100-fold molar excess as competitors in the competing assays. EMSA was performed using a Chemiluminescent EMSA Kit (Beyotime Technology, Cat. GS009) following the manufacturer's

protocol. The labeled probes and their shifted protein complexes were detected on a Tanon 5200 chemiluminescent imaging system (Tanon Science & Technology Company).

Accession Numbers

Giant cell enhancer associated MATE efflux family protein, At5g17700; *MIR319a* (At4g23713); Dof family members (At2g28810, At5g62940, At5g66940); *ATML1* (At4g21750); *TCP4* (At3g15030).

Supplemental Data

Supplemental Figure S1. Supplemental Figure S1. Entire sequence of the giant cell enhancer (related to Figure 1).
Supplemental Figure S2. Region 1 alone, Region 4 alone and Regions 3 and 4 together are not sufficient to drive any reporter expression (related to Figure 1).
Supplemental Fig. S3. Sequences and locations of the putative Dof binding motifs, putative TCP binding motifs, putative HD-ZIP binding motifs, and different enhancer fragments (related to Figures 3 and 4).
Supplemental Figure S4. Mutating the putative trihelix binding motif does not alter expression pattern (Related to Figure 4).
Supplemental Figure S5. Binding of Dof transcription factors in the giant cell enhancer region and nearby (related to Figure 6).
Supplemental Figure S6. pBR65 and pBR69 sepals show weak 3×Venus fluorescence in some nuclei (related to Figure 6).
Supplemental Figure S7. Conservation of Regions 1 and 2 of the giant cell enhancer across Brassicaceae species (related to Figure 4).
Supplemental Figure S9. Probe sequences used in EMSAs (related to Figures 3, 4, and 6).
Supplemental Figure S10. Summary of constructs used in dissection and mutation experiments (related to Figures 1, 3, 4, 5, and 6).
Supplemental Table S1. Full-genome yeast one-hybrid screen results: low-confidence interactions (related to Table 1).
Supplemental Table S2: Giant cell enhancer dissection constructs
Supplemental Table S3. Primer sequences
Supplemental Table S4. Yeast one hybrid constructs
Supplemental Table S5. EMSA probe sequences

Supplemental Datasets

Supplemental Dataset S1. Yeast one-hybrid results for Region 1
Supplemental Dataset S2. Yeast one-hybrid results for Region 2
Supplemental Dataset S3. Yeast one-hybrid results for the 100-bp Junction Region overlapping the edges of Region 1 and Region 2
Supplemental Dataset S4. Statistical Tables.

Supplemental File S1.. FASTA format alignment of Regions 1 and 2 of the giant cell enhancer across Brassicaceae species associated with Supplemental Figure S7.

Author Contributions

Conception and design of experiments: L.H., B.Z.R., C.S.K., A.H.K.R.. Giant cell enhancer dissection experiments and analysis: L.H., B.Z.R., C.S.K., A.H.K.R.. Electrophoretic mobility shift assay: L.H., S.X., X.H., D.Q.. High throughput yeast one-hybrid screen: S.E.K., J.L.P.-P..

Writing of the manuscript: L.H., B.Z.R., A.H.K.R.. Revising and editing of the manuscript: L.H., B.Z.R., C.S.K., J.L.P.-P., A.H.K.R..

Acknowledgements

We thank Joseph Cammarata, Frances Clark, Kate Harline, Shuyao Kong, Xiaobo Zhao and Ming Zhou for critical reading and comments on the manuscript. We thank Samuel Leiboff and Dana Robinson for technical assistance in cloning constructs for this project. We thank Binqiang Wang for technical assistance in protein purification for this project. This work was supported by NSF IOS Plant, Fungal, and Microbial Developmental Mechanisms grants IOS-1256733 (AHKR) and IOS-1553030 (AKHR), NSF Graduate Research Fellowship DGE-2139899 (BR), National Natural Science Foundation of China grant no. 32070853 (LH), Hundred-Talent Program of Zhejiang University (LH), NSF IOS 1755452 (JLPP), NSF MCB 1158254 (JLPP) and NIH R01GM056006 (JLPP). No conflicts of interest declared.

Table 1. Full-genome yeast one-hybrid screen results: high-confidence candidates

Gene family	Region 1	100-bp Junction	Region 2
C_2C_2-Dof	DOF1.8 (At1g64620) DOF4.7 (At4g38000)		
TCP		TCP2 (At4g18390) TCP3 (At1g53230) TCP4 (At3g15030) TCP19 (At5g51910) TCP24 (At1g30210)	TCP2 (At4g18390) TCP3 (At1g53230) TCP4 (At3g15030) TCP10 (At2g31070
HD-ZIP		HDG11 (At1g73360)	ANL2 (At4g00730) GL2 (At1g79840) HDG11 (At1g73360) PHB (At2g34710)
AP2	CRF5 (At3g61630) DEWAX (At5g61590) DREB2 (At5g05410) ERF53 (At2g20880) ERF14 (At1g04370) ESE1 (At3g23220) RAP2.4 (At1g22190) At3g57600 At1g72570 At1g75490 At5g18450 At5g65130	AIL5 (At5g57390) DEWAX (At5g61590) DREB2 (At5g05410) EBP (At3g16770) ERF1 (At3g23240) ERF14 (At1g04370) ESE1 (At3g23220) ESE2 (At2g25820) RAP2.4 (At1g22190) RAP2.6 (At1g43160) At1g72570 At3g18960 At4g33280 At5g65130	AIL5 (At5g57390) ANT (At4g37750) At1g04370 At1g72570 At1g75490 At3g16770 At3g57600 At4g31060 At5g05410 At5g18450 At5g65130
ARR	ARR2 (At3g04280)	ARR1 (At3g16857) ARR3 (At1g59940) ARR22 (At3g04280)	ARR22 (At3g04280)
bHLH	AKS1 (At1g51140)	LHL2 (At2g31280) LHW (At2g27230)	AKS1 (At1g51140) PIF5 (At3g59060) PIL5 (At2g20180) At3g57800
others	BBX7 (At3g07650) BBX26 (At1g60250) E2FC (At1g47870) GEBP (At4g00270) HB7 (At2g46680) LBD18 (At2g45420) NAC078 (At5g04410)	ATCTH (At2g25900) BBX27 (At1g68190) EIL3 (At1g73730) EIN3 (At3g20770) EMB2746 (At5g63420) GEBP (At4g00270) GOA (At1g31140) HSFA9 (At5g54070) JLO (At4g00220) KAN2 (At1g32240) MBS1 (At3g02790) MYB88 (At2g02820) NAC042 (At2g43000) SAP (At5g35770) SRS4 (At2g18120) TCX2 (At4g14770) VOZ1 (At1g28520) WRKY7 (At4g24240) WRKY16 (At5g45050) At1g04500 At1g11950 At1g14580 At1g31040 At1g76590 At3g07500 At4g00390 At4g09450	ATX1 (At2g31650) bZIP19 (At4g35040) EIL3 (At1g73730) HB7 (At2g46680) HB30 (At5g15210) HB33 (At1g75240) WRKY47 (At4g01720) At1g01640 At1g55650 At3g04450 At4g00390 At4g08455

See also Supplemental Table S1

Figure Legends:

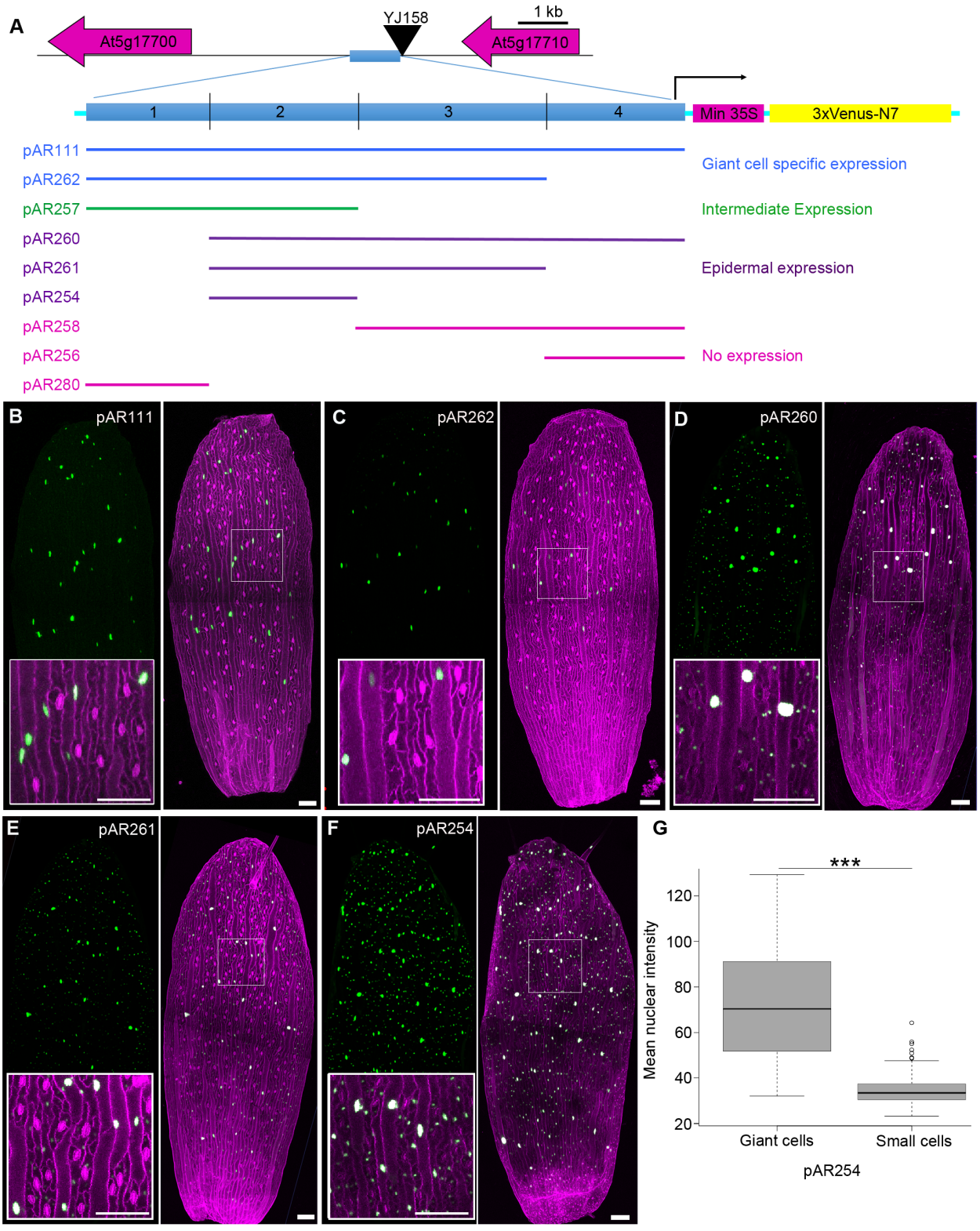


Figure 1. Dissection of the 1-kb giant cell enhancer reveals Region 1 and Region 2 drive major expression patterns.

(A) The giant cell enhancer is a ~1-kb sequence 3.2 kb upstream of At5g17700, encoding a MATE efflux family protein. The 1-kb enhancer was originally identified because it flanked the YJ158 enhancer trap T-DNA insertion (triangle), which produced the same giant cell-specific expression pattern. The enhancer was divided into four regions. Different constructs containing different regions of the enhancer for testing enhancer elements are diagrammed below. Reporter lines with these constructs were tested for their ability to produce fluorescence from nucleus-localized 3×Venus-N7 when expressed from a –60 minimal 35S promoter. Enhancer fragments in the different reporter lines are color-coded with their expression patterns: blue denotes giant cell-specific expression, purple for epidermal expression (giant cells + small cells), green for intermediate expression patterns (further classified in Figure 2), and magenta for no expression.

(B-F) Confocal images of stage-14 sepals from plants carrying the different reporters. Images on the left, 3×Venus (green) marking the nuclei of cells expressing the reporter; images on the right, 3×Venus (green) signals merged with propidium iodide (PI, magenta) staining the cell wall. (Note that overlap of green nuclear signal and magenta PI appears white, so white nuclei are expressing the reporter as well as green nuclei.) Insets in the green channel images show magnified views of the cells outlined with the white box. 3×Venus is restricted to giant cells in plants with the reporter pAR111 and pAR262 (B,C). The reporter has a broader expression in pAR260, pAR261 and pAR254 (D,E,F). Note that stomatal guard cells stain as magenta oblongs. Scale bars: 100 μm.

(G) Fluorescent intensity of the 3×Venus reporter in giant cell nuclei relative to the small cell nuclei of pAR254 sepals, which is expressed in a broad epidermal pattern. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend to 1.5 of the interquartile range. Points indicate outliers. ***$P < 0.001$, significant difference by two tailed t-test, $P = 4.29\text{x}10^{-13}$; n = 210 small cells and 48 giant cells (statistical details in Supplemental Dataset S4).

See also Supplemental Figures S1, S2, and S9.

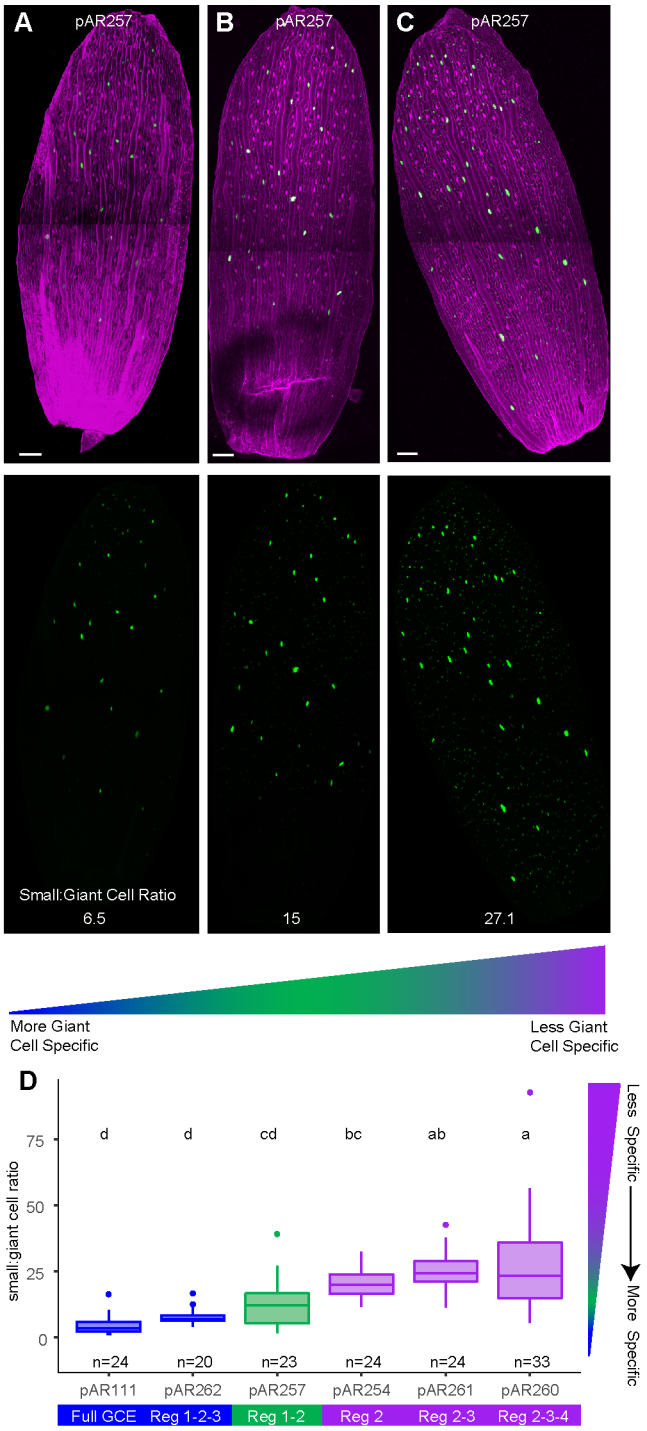


Figure 2. Region 1-2 of the enhancer drives an intermediate expression pattern

(A-C) Examples of the range of expression patterns driven by the pAR257 Region1-2 enhancer from independent T1 plants. Top, 3×Venus (green) marking the nuclei of cells expressing the reporter merged with PI (magenta) staining the cell wall. Bottom, 3×Venus (green) signals of the pAR257 reporter alone. The small cell to giant cell ratio is given at the bottom of the images, quantifying giant cell specificity of each of these sepals (see panel D). Scale bars, 100 μm.

(D) Expression patterns quantified by counting the number of small cells and giant cells expressing 3× Venus and calculating the ratio. Lower ratios (fewer small cells expressing 3×Venus) suggest greater giant cell specificity. Different lowercase letters indicate significant differences based on one-way analysis of variance (ANOVA) with Tukey's honest significant difference test ($P < 0.05$; statistical details in Supplemental Dataset S4). n is listed on the graph for each reporter and represents the number of independent T1 transgenic plants analyzed with one sepal quantified per plant. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend to 1.5 of the interquartile range. Points indicate outliers.

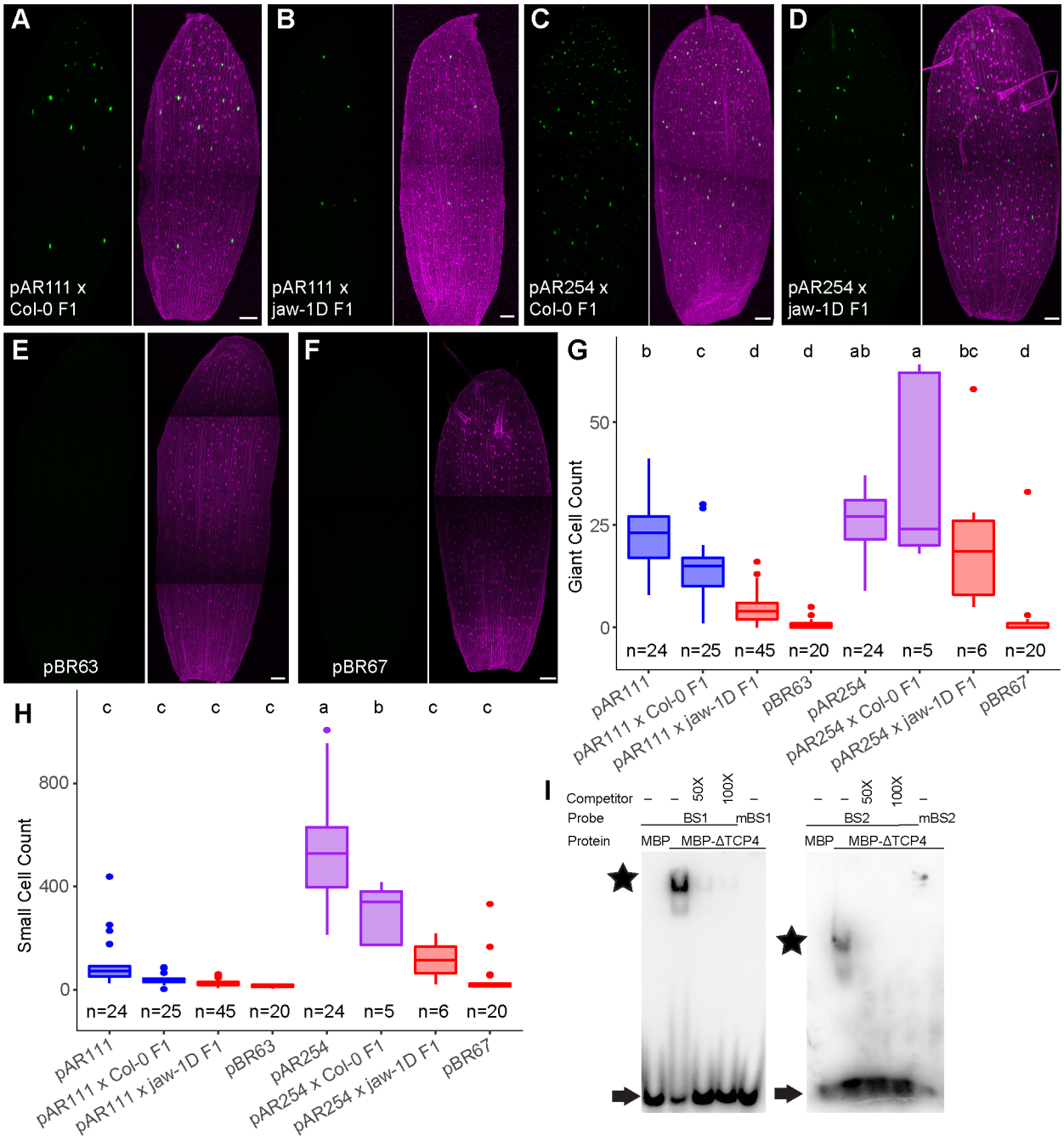


Figure 3. TCP transcription factors promote expression from the enhancer.

(A) Confocal images of a sepal from a pAR111 × Col-0 F1 plant (Full Giant Cell Enhancer Regions 1-2-3-4).

(B) Confocal images of a sepal from a pAR111 × *jaw-1D* F1 plant (Full Giant Cell Enhancer Regions 1-2-3-4), showing far fewer giant cells with 3×Venus fluorescence (green) compared to the sepal from the F1 plant in (A).

(C) Confocal images of a sepal from a pAR254 × Col-0 F1 plant (enhancer Region 2).

(D) Confocal images of a sepal from pAR254 × *jaw-1D* F1 plant (enhancer Region 2), showing far fewer giant cells with 3×Venus (green) compared to the sepal from the pAR254 F1 plant shown in (C).

(E) Confocal images of a sepal from a pBR63 (both GGACC putative TCP motifs mutated in the full Region 1-2-3-4 giant cell enhancer) T1 plant. Note that hardly any nuclei expressing 3×Venus can be detected.
(F) Confocal images of a sepal from a pBR67 (both GGACC putative TCP motifs mutated in the Region 2 giant cell enhancer reporter) T1 plant. Note that hardly any nuclei expressing 3×Venus can be detected.
(G) Number of giant cells expressing 3×Venus in the T1 plants shown in this figure, with pAR111 (Regions 1-2-3-4) and pAR254 (Region 2) homozygous plants for comparison.
(H) Number of small cells expressing 3×Venus in the T1 plants shown in this figure, with pAR111 (Regions 1-2-3-4) and pAR254 (Region 2) for comparison.
(I) Electrophoretic mobility shift assay performed using labeled TCP binding motif probes corresponding to the two putative TCP binding motifs (BS1 and BS2; see Supplemental Figure S10) in Region 2, together with recombinant MBP alone or the MBP–ΔTCP fusion (See also Supplemental Figure S9). Unlabeled probes were used as competitors in competition assays with increasing amounts of 50- and 100-fold molar excess. Assay using mutated probes (mBS1; see Supplemental Figure S9) with MBP–ΔTCP is also shown. The black stars (upper) and arrows (bottom) represent the probes binding to MBP–ΔTCP protein and free probes, respectively.
Sepals were imaged at the same setting, with 3% laser power output. Images on the left are 3×Venus (green) marking the nuclei of cells expressing the reporter; images on the right are 3×Venus (green) signals merged with propidium iodide (PI, magenta) staining of cell walls. Scale bars, 100 μm. Different lowercase letters in (G) and (H) indicate significant differences based on one-way ANOVA with Tukey's HSD test ($P < 0.05$; statistical details in Supplemental Dataset S4). n is listed in the graphs and is the number of sepals from that genotype analyzed. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend to 1.5 of the interquartile range. Points indicate outliers. See also Supplemental Figures S3, S7, and S8.

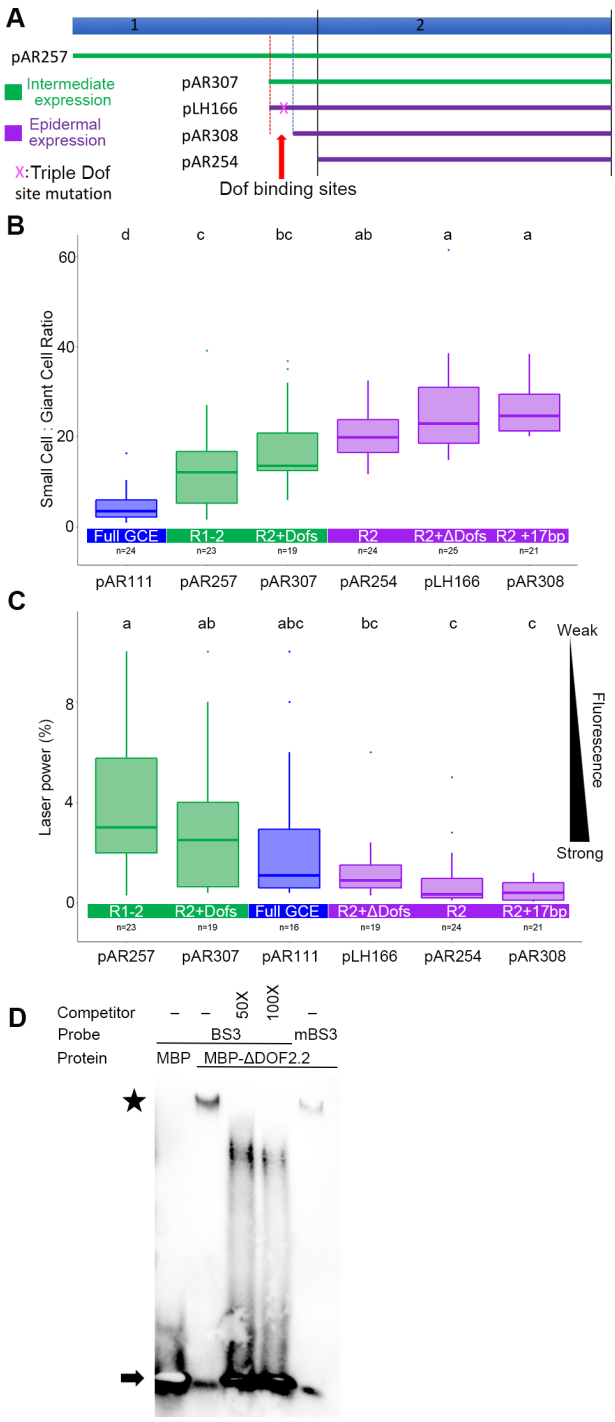


Figure 4. Dof binding motifs in Region 1 confer giant cell enhanced activity.
(A) Fine-scale dissection of Region 1 of the enhancer. Schematic diagram of reporter constructs containing different fragments or mutations of Region 1.
(B) Expression patterns quantified as the ratio of small cells to giant cells expressing 3×Venus. Lower ratios (fewer small cells expressing 3×Venus) suggest greater giant cell specificity. Data for pAR111, pAR254 and pAR257 are reproduced from Figure 2D for comparison.

(C) Laser intensity required to saturate the 3×Venus signal in giant cells for each reporter construct. A higher laser intensity signifies weaker fluorescence of the reporter.
(D) Electrophoretic mobility shift assay performed using a labeled Dof binding motif probe (BS3) in Region 2, together with MBP alone or the MBP–ΔDOF2.2 fusion (See also Supplemental Figures S9 and S10). Unlabeled probe DNA were used as competitors in competition assays with increasing amounts of 50- and 100-fold molar excess. Assay using mutated probes (mBS3) with MBP–ΔDOF2.2 is also shown. The black stars (upper) and arrows (bottom) represent the probes binding to MBP–ΔDOF2.2 protein and free probes, respectively.
Different lowercase letters in (B) and (C) indicate significant differences based on one-way ANOVA with Tukey's HSD test ($P < 0.05$; statistical details in Supplemental Dataset S4). n is listed on the graph for each reporter and represents the number of independent T1 transgenic plants analyzed with one sepal quantified per plant. Note that the data from Figure 2D is reproduced here for context. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend to 1.5 of the interquartile range. Points indicate outliers. See also Supplemental Figures S3, S4, S7, and S8.

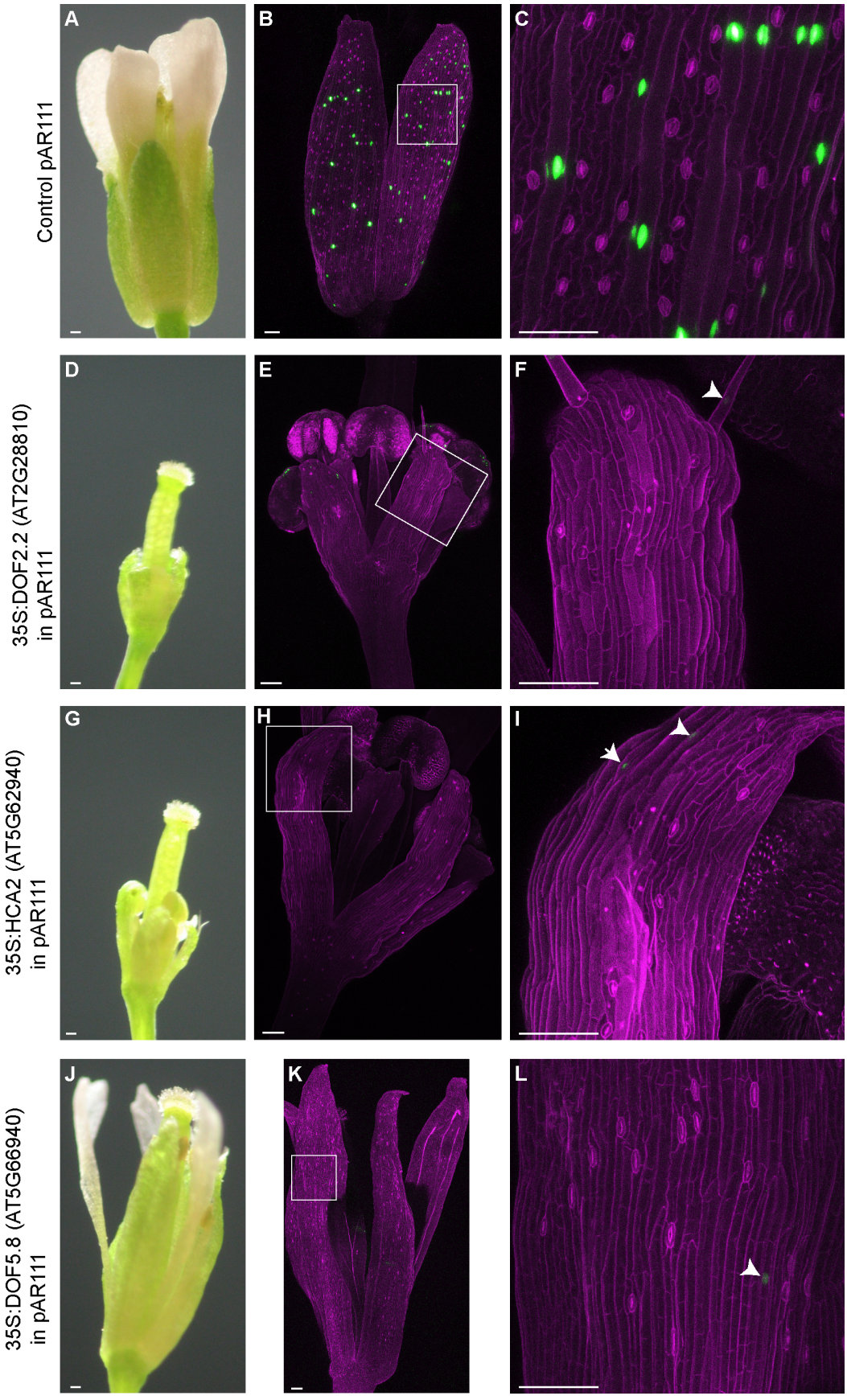

Control pAR111
35S:DOF2.2 (AT2G28810)
in pAR111
35S:HCA2 (AT5G62940)
in pAR111
35S:DOF5.8 (AT5G66940)
in pAR111
A
B
C
D
E
F
G
H
I
J
K
L

Figure 5. Dof transcription factor overexpression suppresses giant cell enhancer activity.
pAR111 is the full-length giant cell enhancer which drives 3×Venus-N7 expression in giant cells (nucleaus-localized green signal). Cell walls were stained with PI (magenta). Arrowheads indicate nuclei expressing weak signal.
(A-C) Flowers from a control pAR111 plant. (A) Dissecting microscope image. (B) Confocal image showing bright signal from the pAR111 full-length giant cell enhancer reporter in giant cell nuclei. (C) Magnification of the white box in B.
(D-F) Flowers from a pAR111 plant harboring *35S:DOF2.2* (At2g28810). (D) Dissecting microscope image showing stunted sepals, petals, and stamens. (E) Confocal microscope image showing strong repression of the pAR111 giant cell enhancer activity. (F) Magnification of the white box in E.
(G-I) Flowers from a pAR111 plant carrying *35S:HCA2* (At5g62940). (G) Dissecting microscope image showing stunted sepals, petals, and stamens. (H) Confocal microscope image showing strong repression of the pAR111 giant cell enhancer activity. (I) Magnification of the white box in H.
(J-L) Flowers from a pAR111 plant expressing *35S:DOF5.8* (At5g66940). (J) Dissecting microscope image showing the sepals are narrow and highly elongated. (K) Confocal microscope image showing strong repression of the pAR111 giant cell enhancer activity. (L) Magnification of the white box in K. Note that the arrowhead points to expression in a giant cell and that faint transverse walls that appear are in an underlying cell layer.
Scale bars, 100 μm. Confocal images were taken with the same settings at 3% laser power output. C, F, I, and L were rendered in MorphoGraphX adjusted in Photoshop with auto contrast.

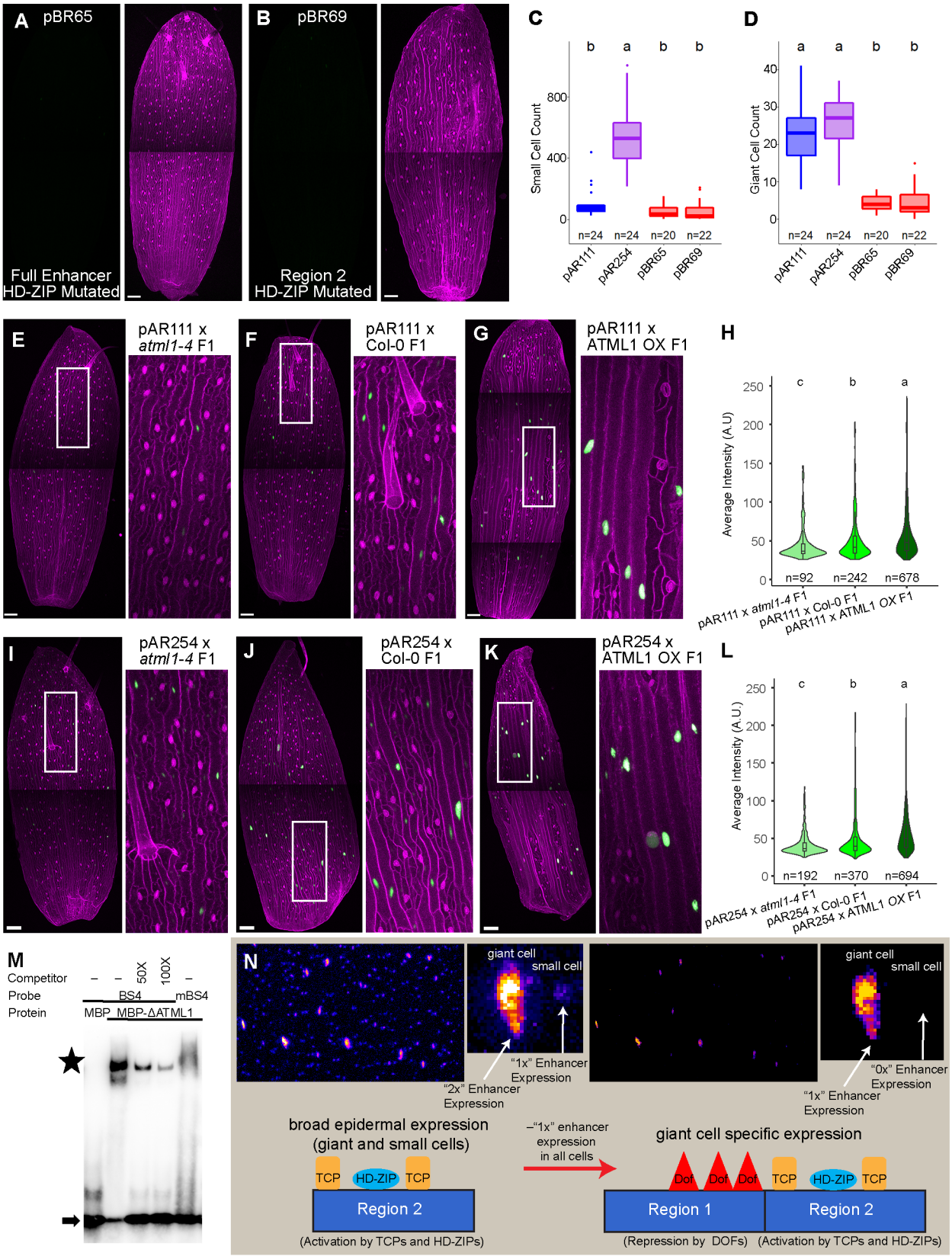

A pBR65
Full Enhancer HD-ZIP Mutated
B pBR69
Region 2 HD-ZIP Mutated
C
Small Cell Count
b a b b
n=24 n=24 n=20 n=22
pAR111 pAR254 pBR65 pBR69
D
Giant Cell Count
a a b b
n=24 n=24 n=20 n=22
pAR111 pAR254 pBR65 pBR69
E pAR111 x atml1-4 F1
F pAR111 x Col-0 F1
G pAR111 x ATML1 OX F1
H
Average Intensity (A.U)
c b a
n=92 n=242 n=678
pAR111 x atml1-4 F1
pAR111 x Col-0 F1
pAR111 x ATML1 OX F1
I pAR254 x atml1-4 F1
J pAR254 x Col-0 F1
K pAR254 x ATML1 OX F1
L
Average Intensity (A.U)
c b a
n=192 n=370 n=694
pAR254 x atml1-4 F1
pAR254 x Col-0 F1
pAR254 x ATML1 OX F1
M
Competitor – – 50X 100X –
Probe BS4 mBS4
Protein MBP MBP-ΔATML1
N
giant cell
small cell
"2x" Enhancer Expression
"1x" Enhancer Expression
"1x" Enhancer Expression
"0x" Enhancer Expression
broad epidermal expression (giant and small cells)
–"1x" enhancer expression in all cells
giant cell specific expression
TCP HD-ZIP TCP
Region 2
(Activation by TCPs and HD-ZIPs)
Dof Dof Dof TCP HD-ZIP TCP
Region 1
Region 2
(Repression by DOFs)
(Activation by TCPs and HD-ZIPs)

Figure 6. The HD-ZIP transcription factor ATML1 increases enhancer activity in giant cells
(A-B) Confocal images of sepals with the HD-ZIP binding motif mutated on the full Region 1-2-3-4 giant cell enhancer (pBR65) **(A)** or the Region 2 enhancer (pBR69) **(B)**. Mutations on both enhancer fragments nearly abolish 3×Venus expression.
(C-D) Boxplots showing the number of small cell nuclei **(C)** and giant cell nuclei **(D)** expressing 3× Venus from the HD-ZIP motif mutation reporters pBR65 and pBR69, with pAR111 (Region 1-2-3-4) and pAR254 (Region 2) shown for comparison. pAR111 and pAR254 counts obtained from the same dataset shown in figure 2D.
(E-G) Confocal images of F1 sepals from crosses between the full Region 1-2-3-4 giant cell enhancer reporter pAR111 and the *atml1-4* knockout line (**E)**, Col-0 **(F)**, and the *PDFpro:Flag-ATML1* overexpression line (*ATML1* OX) **(G)**. All plants are heterozygous for both the enhancer reporter and *ATML1* knockout or overexpression.
(H) Violin plot of 3×Venus signal intensity in nuclei of giant cells of lines shown in **(E-G)** measured in arbitrary units between 0 and 255.
(I-K) Confocal images of F1 sepals from crosses between the Region 2 giant cell enhancer reporter pAR254 and the *atml1-4* knockout line (**I)**, Col-0 **(J)**, and the *PDF1pro:Flag-ATML1* overexpression line **(K)**. All plants are heterozygous for both the enhancer reporter and the *ATML1* knockout or overexpression.
(L) Violin plot of 3×Venus signal intensity in nuclei of giant cells of lines shown in **(I-K)** measured in arbitrary units between 0 and 255.
(M) Electrophoretic mobility shift assay performed using a labeled HD-ZIP binding motif probe (BS4; see Supplemental Figure S10) in Region 2, together with MBP alone or the MBP–ΔATML1 fusion (See also Supplemental Figure S9). Unlabeled probe DNA were used as competitors in competition assays with increasing amounts of 50- and 100-fold molar excess. Assay using mutated probes (mBS4) with MBP–ΔATML1 is also shown. The black stars (upper) and arrows (bottom) represent the probes binding to MBP–ΔATML1 protein and free probes, respectively.
(N) Summary model. In the giant cell enhancer, cell type specificity of expression arises primarily from the combination of three elements across two regions. Region 2 of the giant cell enhancer, shown on the left, drives broad epidermal expression. Nevertheless, this expression is stronger (2x) in giant cells (large, highly endoreduplicated nuclei) than in small cells (1x, smaller nuclei) as highlighted by the fire LUT (using FIJI), see also quantification in Figure 1G. TCP transcription factors (represented as gold rectangles) bind to and activate expression via Region 2 of the enhancer. Additionally, HD-ZIP Class IV transcription factors (blue ovals) bind to Region 2, activating stronger expression in giant cells than small cells.
When Region 2 is combined with Region 1, the enhancer becomes much more giant cell-specific as detailed in the results. Our evidence suggests that Dof transcription factors (red triangles) bind to this Region 1 and broadly repress enhancer-driven expression (subtract 1x) across the sepal epidermis. This broad repression, which we represent here by decreasing the brightness of the image on the left until small cell nuclei are no longer visible (the image on the right), is capable of eradicating expression in the small cells (0x), but not in the giant cells (1x).
Note, the number of Dof, TCP, and HD-ZIP polygons is meant to convey the wild-type levels of transcription factors bound to their respective binding motifs, not an exact number of proteins.
Different lowercase letters in (C), (D), (H), and (L) indicate significant differences based on one-way ANOVA with Tukey's HSD test ($P < 0.05$; statistical details in Supplemental Dataset S4). In (C) and (D), n is listed on the graph for each reporter line and represents the number of independent T1 transgenic plants analyzed with one sepal quantified per plant. In (H) and (L), n is listed in the graph and represents the number of nuclei analyzed from about 20 sepals. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend to 1.5 of the interquartile range. Points indicate outliers. See also Supplemental Figures S3, S5, S7 and S8.

References

Abe, M., Katsumata, H., Komeda, Y., and Takahashi, T. (2003). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. Development 130: 635–643.

Abe, M., Takahashi, T., and Komeda, Y. (2001). Identification of a *cis*-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. Plant J 26: 487–494.

Boccaccini, A., Santopolo, S., Capauto, D., Lorrai, R., Minutello, E., Serino, G., Costantino, P., and Vittorioso, P. (2014). The DOF Protein DAG1 and the DELLA Protein GAI Cooperate in Negatively Regulating the AtGA3ox1 Gene. Mol Plant 7: 1486–1489.

Bonaldi, K., Li, Z., Kang, S.E., Breton, G., and Pruneda-Paz, J.L. (2017). Novel cell surface luciferase reporter for high-throughput yeast one-hybrid screens. Nucleic Acids Res 45: e157.

Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science 318: 801–806.

Bramsiepe, J., Wester, K., Weinl, C., Roodbarkelari, F., Kasili, R., Larkin, J.C., Hülskamp, M., and Schnittger, A. (2010). Endoreplication controls cell fate maintenance. PLoS Genet 6: e1000996.

Buecker, C. and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. Trends Genet 28: 276–284.

Clough SJ and Bent AF. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16:735-743.

Crawford, B.C.W., Nath, U., Carpenter, R., and Coen, E.S. (2004). CINCINNATA controls both cell differentiation and growth in petal lobes and leaves of Antirrhinum. Plant Phys 135: 244–253.

Davidson, E.H. (2010). Chapter 2 cis-Regulatory modules, and the structure/function basis of regulatory logic. In The Regulatory Genome (Elsevier), pp. 31–86.

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P. (2019). Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. Dev Cell 48: 840–852.

Eshed, Y., Izhaki, A., Baum, S.F., Floyd, S.K., and Bowman, J.L. (2004). Asymmetric leaf development and blade expansion in Arabidopsis are mediated by KANADI and YABBY activities. Development 131: 2997–3006.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). "FIMO: Scanning for occurrences of a given motif", Bioinformatics 27:1017-1018.

Guo, Y., Qin, G., Gu, H., and Qu, L.-J. (2009). Dof5.6/HCA2, a Dof Transcription Factor Gene, Regulates Interfascicular Cambium Formation and Vascular Tissue Development in Arabidopsis. Plant Cell 21: 3518–3534.

Gupta, S., Malviya, N., Kushwaha, H., Nasim, J., Bisht, N.C., Singh, V.K., and Yadav, D. (2015). Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. Planta 241: 549–562.

Haudry, A. et al. (2013). An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genet 45: 891–898.

He, L., Su, C., Wang, Y., and Wei, Z. (2020). ATDOF5.8 protein is the upstream regulator of ANAC069 and is responsive to abiotic stress. Biochimie 110: 17–24.

Hong, L. et al. (2016). Variable cell growth yields reproducible organ development through spatiotemporal averaging. Dev Cell 38: 15–32.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343: 776–779.

Lee, M.M. and Schiefelbein, J. (2002). Cell pattern in the Arabidopsis root epidermis determined by lateral inhibition with feedback. Plant Cell 14: 611–618.

Li, Z., Bonaldi, K., Kang, S.E., and Pruneda-Paz, J.L. (2019). High-Throughput Yeast One-Hybrid Screens Using a Cell Surface gLUC Reporter. Current Protocols in Plant Biology 4: e20086.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell 167: 1170–1187.

Martín-Trillo, M. and Cubas, P. (2010). TCP genes: a family snapshot ten years later. Trends Plant Sci 15: 31–39.

Meyer, H.M., Teles, J., Formosa-Jordan, P., Refahi, Y., San-Bento, R., Ingram, G., Jönsson, H., Locke, J.C.W., and Roeder, A.H.K. (2017). Fluctuations of the transcription factor ATML1 generate the pattern of giant cells in the *Arabidopsis* sepal. eLife 6: e19131.

Miyashima, S. et al. (2019). Mobile PEAR transcription factors integrate positional cues to prime cambial growth. Nature 565: 490–494.

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., and Takahashi, T. (2006). Characterization of the class IV homeodomain-Leucine Zipper gene family in Arabidopsis. Plant Phys 141: 1363–1375.

Nath, U., Crawford, B.C.W., Carpenter, R., and Coen, E.S. (2003). Genetic control of surface curvature. Science 299: 1404–1407.

Noguero, M., Atif, R.M., Ochatt, S., and Thompson, R.D. (2013). The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. Plant Sci 209: 32–45.

Oka, R., Zicola, J., Weber, B., Anderson, S.N., Hodgman, C., Gent, J.I., Wesselink, J.-J., Springer, N.M., Hoefsloot, H.C.J., Turck, F., and Stam, M. (2017). Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. Genome Biology 18: 137.

O'Malley, R.C., Huang, S.-S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165: 1280–1292.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature 425: 257–263.

Pennacchio, L.A. et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature 444: 499–502.

Pruneda-Paz, J.L., Breton, G., Nagel, D.H., Kang, S.E., Bonaldi, K., Doherty, C.J., Ravelo, S., Galli, M., Ecker, J.R., and Kay, S.A. (2014). A Genome-Scale Resource for the Functional Characterization of Arabidopsis Transcription Factors. Cell Reports 8: 622–632.

Qu, X., Chatty, P.R., and Roeder, A.H.K. (2014). Endomembrane trafficking protein SEC24A regulates cell size patterning in Arabidopsis. Plant Phys 166: 1877–1890.

Ravindran, P., Verma, V., Stamm, P., and Kumar, P.P. (2017). A Novel RGL2-DOF6 Complex Contributes to Primary Seed Dormancy in Arabidopsis thaliana by Regulating a GATA Transcription Factor. Mol Plant 10: 1307–1320.

Ricci, W.A. et al. (2020). Widespread long-range cis-regulatory elements in the maize genome. Nature Plants 5: 1237–1249.

Robinson, D.O., Coate, J.E., Singh, A., Hong, L., Bush, M.S., Doyle, J.J., and Roeder, A.H.K. (2018). Ploidy and Size at Multiple Scales in the Arabidopsis Sepal. Plant Cell 30: 2308–2329.

Roeder, A.H.K., Chickarmane, V., Cunha, A., Obara, B., Manjunath, B.S., and Meyerowitz, E.M. (2010). Variability in the control of cell division underlies sepal epidermal patterning in *Arabidopsis thaliana*. PLoS Biol 8: e1000367.

Roeder, A.H.K., Cunha, A., Ohno, C.K., and Meyerowitz, E.M. (2012). Cell cycle regulates cell type in the *Arabidopsis* sepal. Development 139: 4416–4427.

Ruta, V., Longo, C., Lepri, A., De Angelis, V., Occhigrossi, S., Costantino, P., and Vittorioso, P. (2020). The DOF Transcription Factors in Seed and Seedling Development. Plants 9: 218–15.

Ryu, K.H., Huang, L., Kang, H.M., and Schiefelbein, J. (2019). Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. Plant Phys 179: 1444–1456.

Sani, H.M., Hamzeh-Mivehroud, M., Silva, A.P., Walshe, J.L., Mohammadi, S.A., Rahbar-Shahrouziasl, M., Abbasi, M., Jamshidi, O., Low, J.K., Dastmalchi, S., and Mackay, J.P. (2018). Expression, purification and DNA-binding properties of zinc finger domains of DOF proteins from Arabidopsis thaliana. Bioimpacts Bi 8: 167–176.

Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., and Schmidt, R. (2004). Silencing in Arabidopsis T-DNA Transformants: The Predominant Role of a Gene-Specific RNA Sensing Mechanism versus Position Effects. Plant Cell 16: 2561–2572.

Schwarz, E.M. and Roeder, A.H.K. (2016). Transcriptomic Effects of the Cell Cycle Regulator LGO in Arabidopsis Sepals. Front Plant Sci 7: 635–22.

Skirycz, A., Radziejwoski, A., Busch, W., Hannah, M.A., Czeszejko, J., Kwaśniewski, M., Zanor, M.-I., Lohmann, J.U., De Veylder, L., Witt, I., and Mueller-Roeber, B. (2008). The DOF transcription factor OBP1 is involved in cell cycle regulation in Arabidopsis thaliana. Plant J 56: 779–792.

Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M. (1990). Early flower development in *Arabidopsis*. Plant Cell 2: 755–767.

Spitz, F. and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13: 613–626.

Stanojevic, D., Small, S., and Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science 254: 1385–1387.

Takada, S. and Jurgens, G. (2007). Transcriptional regulation of epidermal cell fate in the Arabidopsis embryo. Development 134: 1141–1150.

Takada, S.; Takada, N.; Yoshida, A. (2013). ATML1 promotes epidermal cell differentiation in *Arabidopsis* shoots. Development, 140: 1919–1923.

Tannenbaum, M., Sarusi-Portuguez, A., Krispil, R., Schwartz, M., Loza, O., Benichou, J.I.C., Mosquna, A., and Hakim, O. (2018). Regulatory chromatin landscape in Arabidopsis thaliana roots uncovered by coupling INTACT and ATAC-seq. Plant Methods 14: 113.

Traas, J., Hülskamp, M., Gendreau, E., and Höfte, H. (1998). Endoreduplication and development: rule without dividing? Curr Opin in Plant Biol 1: 498–503.

Uberti-Manassero, N.G., Lucero, L.E., Viola, I.L., Vegetti, A.C., and Gonzalez, D.H. (2011). The class I protein AtTCP15 modulates plant development through a pathway that overlaps with the one affected by CIN-like TCP proteins. J Exp Bot 63: 809–823.

Vadde, B.V.L., Challa, K.R., and Nath, U. (2017). The TCP4 transcription factor regulates trichome cell differentiation by directly activating GLABROUS INFLORESCENCE STEMSin Arabidopsis thaliana. Plant J 93: 259–269.

Vadde, B.V.L., Challa, K.R., Sunkara, P., Hegde, A.S., and Nath, U. (2019). The TCP4 Transcription Factor Directly Activates TRICHOMELESS1and 2and Suppresses Trichome Initiation. Plant Phys 181: 1587–1599.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J.C., Galli, M., Lewsey, M.G., Huang, E., Mukherjee, T., Chen, X., Reece-Hoyes, J.S., Govindarajan, S., Shaulsky, G., Walhout, A.J., Bouget, F.Y., Ratsch, G., Larrondo, L.F., Ecker, J.R., Hughes, T.R. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 158:1431-43.

Yan, W., Chen, D., Schumacher, J., Durantini, D., Engelhorn, J., Chen, M., Carles, C.C., and Kaufmann, K. (2019). Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. Nat Commun 10: 1705.

Yanagisawa, S. (2016). Structure, Function, and Evolution of the Dof Transcription Factor Family (Elsevier Inc.). Plant Transcription Factors: 183-197.

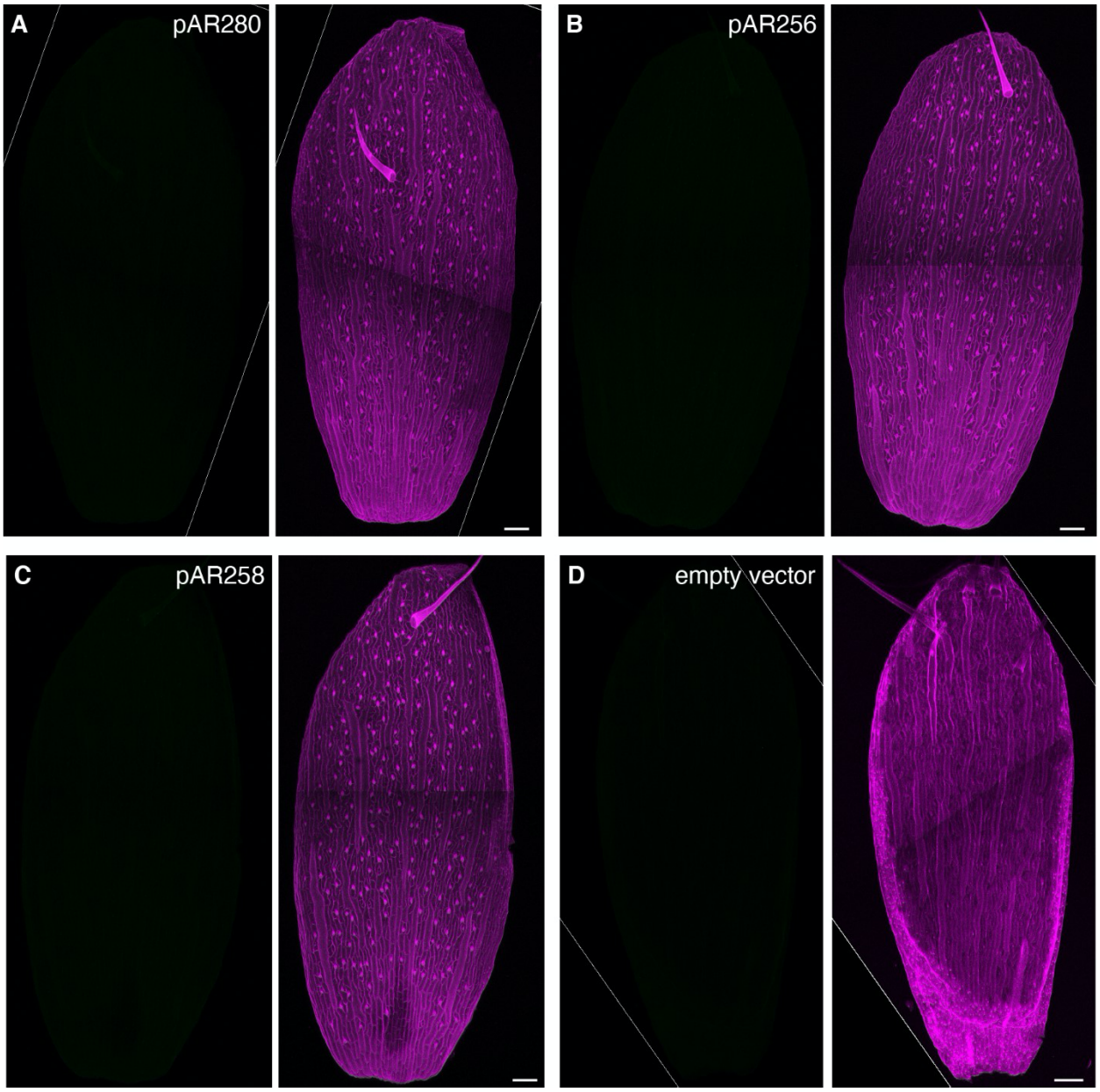
Yanagisawa, S. (2002). The Dof family of plant transcription factors. Trends Plant Sci 7: 555–560.

Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures. Plant Cell 27: 2415–2426.

Zi-Yu, L., Bin, L., and Ai-Wu, D. (2011). The Arabidopsis Transcription Factor AtTCP15 Regulates Endoreduplication by Modulating Expression of Key Cell-cycle Genes. Mol Plant 5: 270–280.

CCTGTCCGCTATATCATGCAAATCCAAATGAAAATGATTTGATTGATGTTACTCTAACTGTAAGA
AAAAAGTGCATTATTGAAAACTTCCTTTTTCTCTGTACACAACTACAATAGCCAGAAGCAATTTG
TAGGTCATAAAACATTATAATAGATATTTTCTTTTTGGAAAGTTTAACCTCTTTTTCTCCTTAAT
TGTTGGGAGTGCAATTGGGTATTTAAGAATGAAGATGAAACTTAGATTCAGGACCATTTTGTGTA
GCCACTGACGACAGTCCTTCGTACATAATATTTATGTTTAAGTTGGGTGTATTATAAATTTATAT
AATTAATTATGTATATTGTATAGGTGGAGTAGGACCAATTTAGTCGTCAAAAGATAATGTGCATG
AATGATGACAAGAATGCGATATTTTTGGCATGAATCATGATTAATCAATGATGTCAATCAAGATT
ACAAATTTATACGAAATGATAGTAAAAAGGTTCAAAAGTCTCTTAATTCTATTACACCAACAAAA
AAAAAGTCTCTTCATTCTACCTCACTAATATATGTTTGTTCGGTTTGGATTGTTTTTAAGAGGTC
CGAGCCGAGATCCGACTAATTTCTTAATGGATCTAAAACATGATACTAATTTTGACTAACTATGA
AAGAGAGAAATATCATCTACAAAATAAAAATTACTAAAAGTTTAAGAAGGTGTTCTAAATAGTTT
AGTACAATTTATTAAACTCTTGTGAACCTTTTCTAAATAAAATCTTGTAGTCTTGGTGTTTTCCT
AAATATTGTATTTTTAAATTAAAAACTTAATTTTGTCATATTTGAAGATATTTCTCTATACATAA
ATAAATATATCTTTCGAGCATACACGAACATTACATTTCTCTGGACAAATCAACATTAGGTTTAT
TCAATTGTCGATTTGACAAACGATGAAAGAAAACGAAACCCTACATATCTTTTAAGCATAAGTGA
CTCTGTGGTTCATGATCTCTATTTCTGGTTCAACGAACGCAAAAGGTAT

Supplemental Figure S1. Entire sequence of the giant cell enhancer (related to Figure 1).
Green: Region 1, 1–208 bp; magenta: Region 2, 209–450 bp; blue: Region 3, 451–760 bp; purple: Region 4, 761–1,024 bp. The underlined region is the 100-bp junction region (167– 266 bp) used in the yeast one-hybrid assay.



Supplemental Figure S2. Region 1 alone, Region 4 alone and Regions 3 and 4 together are not sufficient to drive any reporter expression (related to Figure 1).

Confocal images of stage-14 sepals from different reporter constructs. Images on the left show 3×Venus (green) signal. Images on the right show 3×Venus (green) signals merged with PI (magenta) staining cell walls. Images were taken with 3% laser power output except where noted.

Increasing laser power to 40% did not reveal any 3×Venus signal.

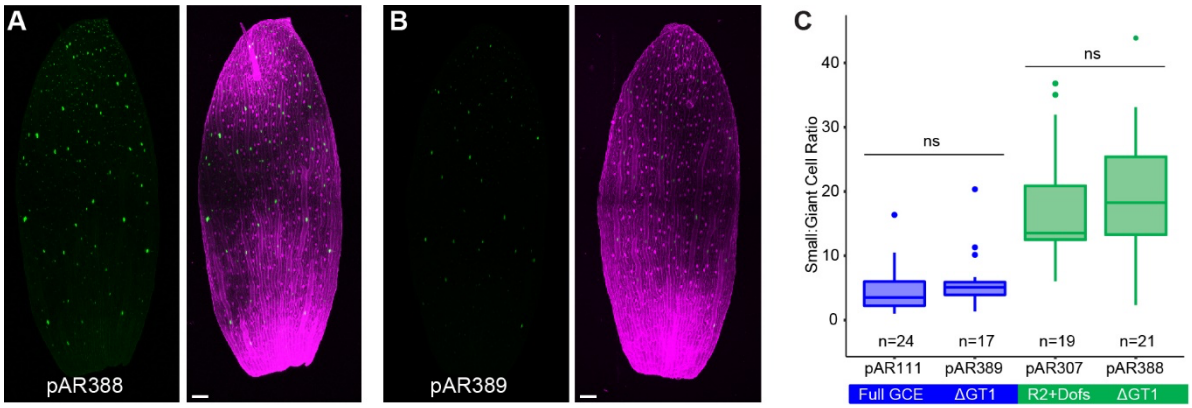
(A) pAR280 Region 1 alone does not drive any reporter expression.

(B) pAR256 Region 4 alone does not drive any reporter expression.

(C) pAR258 Region 3-4 do not drive any reporter expression.

(D) The empty vector (pSL12) does not drive any reporter expression. Imaged at 2.6% laser power. Scale bars, 100 µm.

Supplemental Fig. S3. Sequences and locations of the putative Dof binding motifs, putative TCP binding motifs, putative HD-ZIP binding motifs, and different enhancer fragments (related to Figures 3 and 4). The sequences included in pAR307 (Region 2 + Dof, green line), pAR308 (Region 2 plus, purple line), and pAR254 (Region 2 only, magenta line) are marked with lines below the sequence. The putative Dof binding motifs are highlighted with red letters. The putative TCP binding motifs are highlighted in gold. The putative HD-ZIP binding motif is highlighted in aqua. The mutated sequences for the putative Dof, TCP, and HD-ZIP binding motifs are shown above the original sequences.



Supplemental Figure S4. Mutating the putative trihelix binding motif does not alter expression pattern (Related to Figure 4).

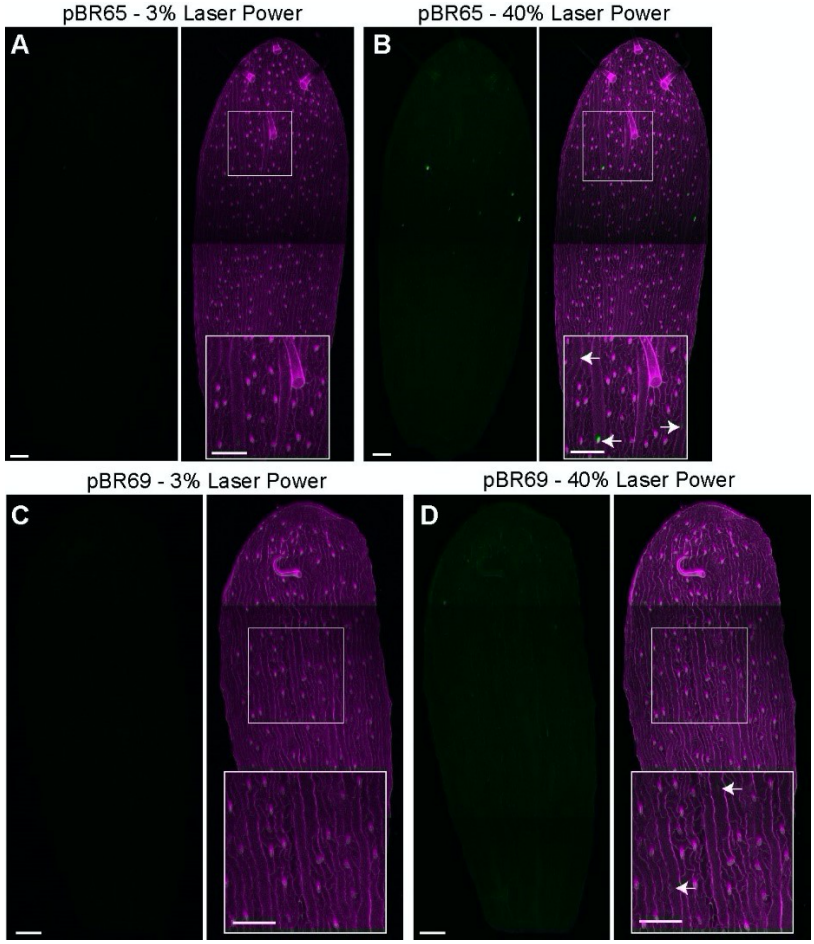
(A) Confocal image of pAR388, which has Region 2 + 40 bp of the end of Region 1 containing Dof motifs, but with the putative GT-1 trihelix motif mutated.

(B) Confocal image of pAR389, which has the full giant cell enhancer (Regions 1-2-3-4), but with the putative GT-1 trihelix motif mutated.

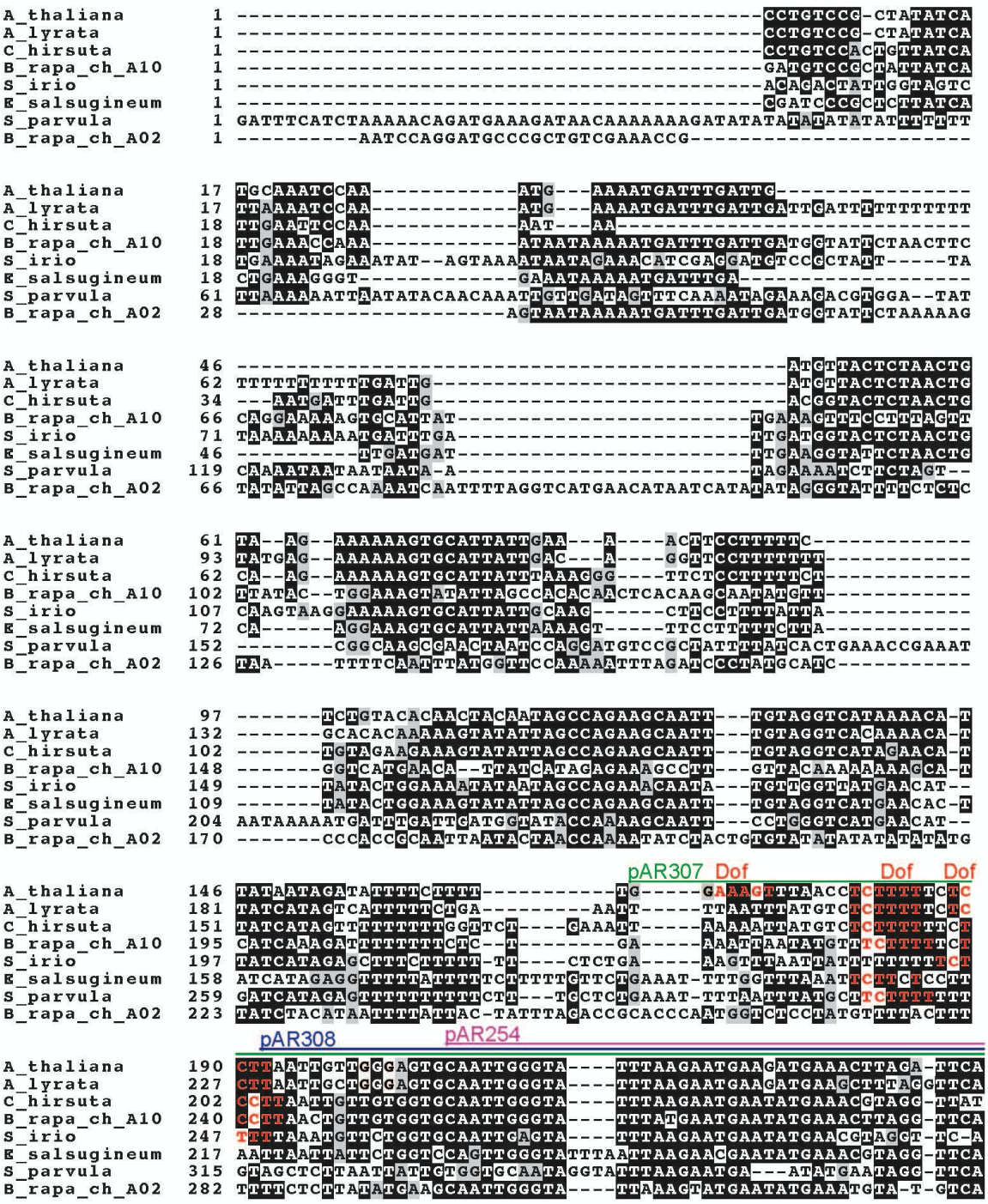
Images on the left are 3×Venus (green) marking the nuclei of cells expressing the enhancer reporter; Images on the right, 3×Venus (green) signals merged with PI (magenta) staining the cell walls. Scale bars, 100 μm.

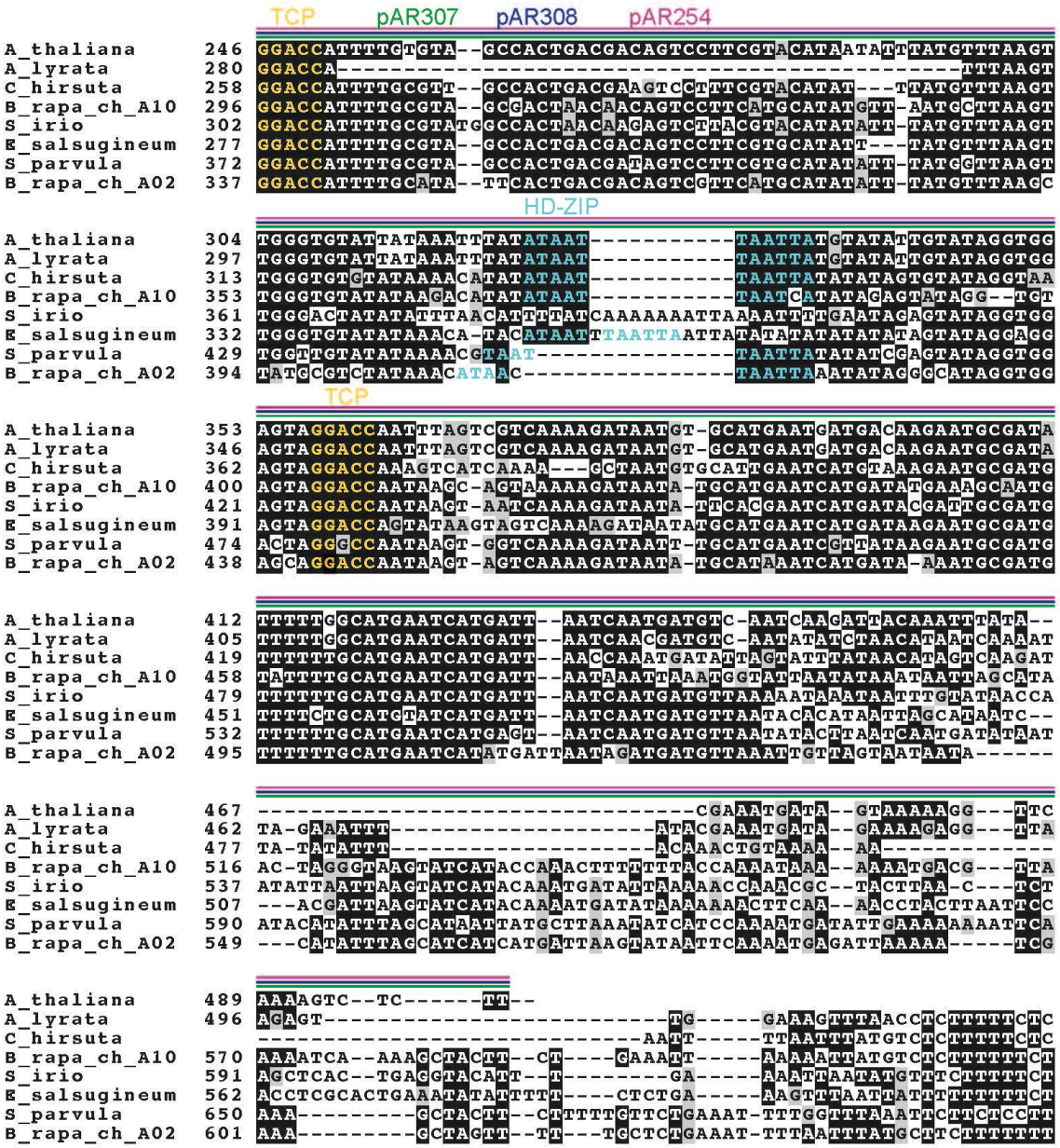
(C) Ratio of small cells to giant cells expressing 3×Venus for pAR388 and pAR389, with pAR111 (Region 1-2-3-4) and pAR307 (Region 2 + 40 bp of the end of Region 1) reproduced from Figure 2D and Figure 4B respectively for comparison. No significant differences were detected by two-tailed t-test with unequal variance, P = 0.33 (pAR111 and pAR389) and P = 0.58 (pAR307 and pAR388) (statistical details in Supplemental Dataset S4). n is listed on the graph for each reporter line and represents the number of independent T1 transgenic plants analyzed with one sepal quantified per plant. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend to 1.5 of the interquartile range. Points indicate outliers.

Supplemental Figure S5. Binding of Dof transcription factors in the giant cell enhancer region and nearby (related to Figure 6).
Results of DAP-seq. Dof transcription factors in red rectangles were identified in the yeast one-hybrid (Y1H) assay in this study; those in blue rectangles were overexpressed in pAR111 Arabidopsis. The red dashed lines mark the 1024-bp giant cell enhancer.



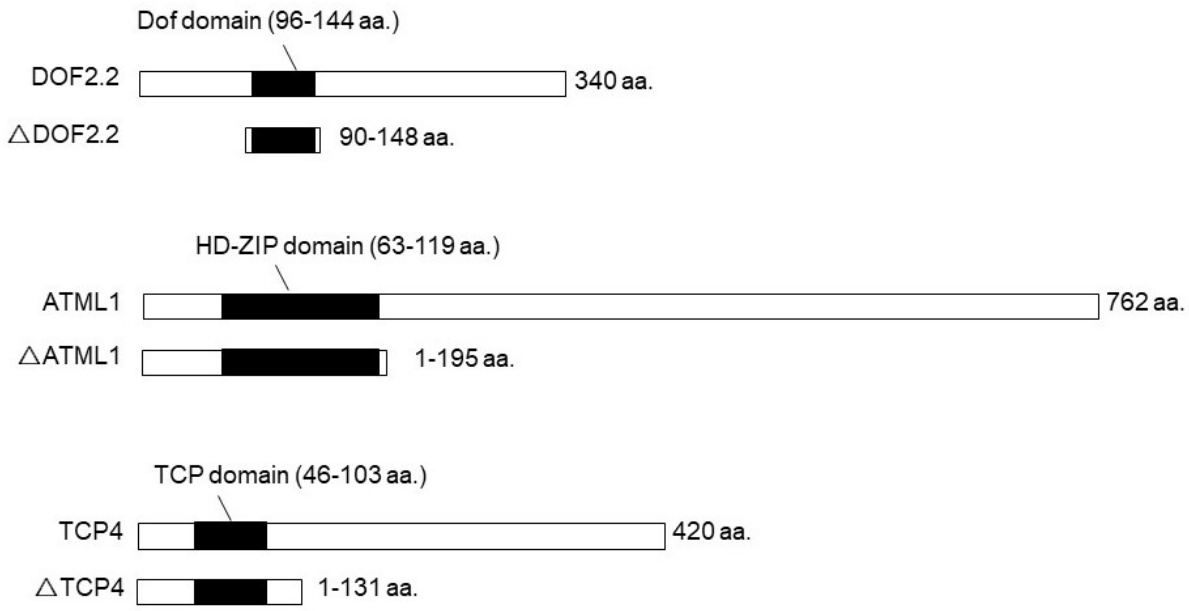
Supplemental Figure S6. pBR65 and pBR69 sepals show weak 3×Venus fluorescence in some nuclei (related to Figure 6). (A-B) Confocal images of a pBR65 (Region 1-2-3-4 enhancer with HD-ZIP motif knocked out) sepal imaged with 3% **(A)** or 40% **(B)** laser power output.
(C-D) Confocal images of a pBR69 (Region 2 enhancer with HD-ZIP motif knocked out) sepal imaged with 3% **(A)** or 40% **(B)** laser power output.
With 3% laser power output, hardly any 3×Venus can be observed on pBR65 or pBR69 sepal epidermis. With laser power output increased to 40%, some nuclei express weak 3×Venus (see arrows on enlarged portions), suggesting that mutating the HD-ZIP motif did not completely eradicate expression, but rather drastically decreased the intensity of enhancer expression.
Images on the left, 3×Venus (green) marking the nuclei of cells expressing the enhancer reporter; images on the right, 3×Venus (green) signals merged with PI (magenta) staining the cell walls. Scale bars, 100 μm.


pAR307
Dof
Dof
Dof
pAR308
pAR254



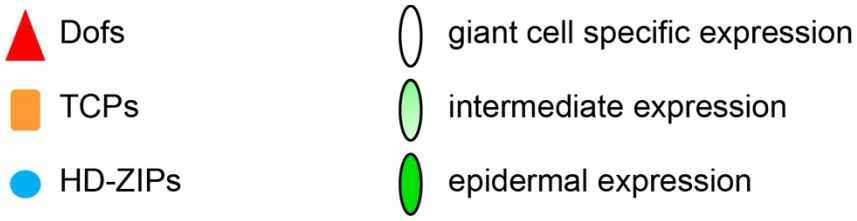
Supplemental Figure S7. Conservation of Regions 1 and 2 of the giant cell enhancer across Brassicaceae species (related to Figure 4).

Clustal omega sequence alignment of the giant cell enhancer Region 1-2 from *Arabidopsis thaliana* with *Arabidopsis lyrata*, *Cardamine hirsuta*, *Sisimbrium irio*, *Eutrema salsugineum*, *Schrenkiella parvula* (previously *Thellungiella parvula*) and *Brassica rapa* (FASTA file of alignment provided in Supplemental File S1). Although *Brassica rapa* is a diploid, it underwent genome duplication during evolution, so there are two regions on chromosomes A10 and A02 that both align to the giant cell enhancer. The sequences included in pAR307 (Region 2 + Dof, green line), pAR308 (Region 2 plus, purple line), and pAR254 (Region 2 only, magenta) are annotated with lines above the sequence. The putative Dof binding motifs predicted in *Arabidopsis thaliana* and their conserved counterparts are highlighted with red letters. The putative TCP binding motifs and their conserved counterparts are highlighted with gold letters. The putative HD-ZIP binding motifs and its conserved counterparts are highlighted with blue letters.



Supplemental Figure S8. Truncated DOF2.2, ATML1, and TCP4 transcription factors used for EMSAs (related to Figures 3, 4, and 6).

Supplemental Figure S9. Probe sequences used in EMSAs (related to Figures 3, 4, and 6).
Putative transcription factor binding motifs are highlighted in yellow. Mutated probe sequences are shown above the binding motifs in black text. Lowercase sequences in blue are part of Region 1 of the enhancer, while uppercase letters in blue are part of Region 2 of the enhancer.



Diagrams	Constructs	Expression pattern	Expression level
	35S:Dof pAR111	Largely repressed	low
	pAR111		high
	pAR262		low
	pAR257		low
	pAR307		low
	pLH166		high
	pAR254		high
	pAR254 *jaw-D*		low
	pAR260		high
	pAR261		high
	pBR67	Essentially no expression	none
	pAR111 *jaw-D*		low
	pAR258	no expression	none
	pAR280	no expression	none
	pBR63	Essentially no expression	none
	pBR69	Essentially no expression	very low
	pBR65	Essentially no expression	very low

Supplemental Figure S10. Summary of constructs used in dissection and mutation experiments (related to Figures 1, 3, 4, 5, and 6).

Table summarizing the results of each construct together with motif-specific mutations and genetic manipulation of TCPs and Dofs. The red triangles, orange rectangles, and blue ovals represent relative occupancy, not exact numbers of transcription factors bound to DNA.

Supplemental Table S1. Full-genome yeast one-hybrid screen results: low-confidence interactions (related to Table 1).

Gene family	Region 1	Junction	Region 2
C_2C_2-Dof	At2g28810	DOF4.5 (At4g21080)	
TCP			TCP13 (At3g02150) TCP22 (At1g72010)
AP2	RAP2.4 (At1g78080) At1g28160 At4g39780	At5g60142	CRF6 (At3g61630) DEWAX (At5g61590)
bHLH			AKS2 (At1g05805)
others	EIL3 (At1g73730) IAA26 (At3g16500) IAA31 (At3g17600) MYB27 (At3g53200) NAC038 (At2g24430) SPL2 (At5g43270) At3g10030	AGL14 (At4g11880) BZIP6 (At2g22850) POSF21 (At2g31370) SIG3 (AT3g53920) SUVR2 (At5g43990) WOX11 (At3g03660) At3g52250 At4g00390 At5g28040	AGL55 (At1g60920) BBX10 (At3g21880) BPC7 (At2g35550) DAZ2 (At4g35280) EMB2746 (At5g63420) GPL1 (At2g25650) HDG5 (At5g46880) NAC078 (At5g04410) NAC098 (At5g53950) TGA2 (At5g06950) UNE16 (At4g13640) At1g55760 At1g60240 At1g61730 At1g66420 At3g02400

Supplemental Table S2: Giant cell enhancer dissection constructs

Plant Construct	Region of enhancer	Forward primer	Reverse primer	Entry clone
pAR111	Full-length (Region 1-2-3-4) 1–1024 bp	oAR214	oAR215	NA (Roeder et al., 2012)
pAR262	Region 1-2-3 (1–760 bp)	oAR500	oAR215	pSL8
pAR257	Region 1-2 (1–449 bp)	oAR497	oAR215	pSL4
pAR260	Region 2-3-4 (207–1024 bp)	oAR214	oAR498	pSL7
pAR261	Region 2-3 (207–760 bp)	oAR500	oAR498	pSL8
pAR254	Region 2 (207–499 bp)	oAR497	oAR498	pSL1
pAR258	Region 3-4 (450–1024 bp)	oAR214	oAR501	pSL5
pAR256	Region 4 (761–1024 bp)	oAR214	oAR499	pSL3
pAR280	Region 1 (1–208 bp)	oAR532	oAR215	pAR278
pAR307	Region 2 with Dof region (167–449 bp)	oAR497	oAR556	pAR299
pLH166	Region 2 with 3 putative Dof motifs mutated (167–449 bp)	oAR497	oLH412	pLH164
pAR308	Region 2+ (without Dof region) (190–449 bp)	oAR497	oAR609	pAR301
pAR388	Region 2 with Dof region with Trihelix motif mutated	oAR497	oAR719	pAR384
pAR389	Region 1-2-3-4 with Trihelix motif mutated using oAR719 and oAR720	oAR714	oAR215	pAR385
pBR63	Region 1-2-3-4 with 2 putative TCP binding motif (GGACC) mutated using oBR210, oBR211, oBR212, and oBR213.	oAR214	oAR215	pBR62
pBR65	Region 1-2-3-4 with a putative HD-ZIP binding motif (ATAATTAATTA) mutated using oBR214 and oBR215	oAR497	oAR498	pBR64
pBR67	Region 2 with 2 putative TCP binding motif (GGACC) mutated using oBR210, oBR211, oBR212, and oBR213.	oAR214	oAR215	pBR66
pBR69	Region 2 with a putative HD-ZIP binding motif (ATAATTAATTA) mutated using oBR214 and oBR215	oAR497	oAR498	pBR68

Associated with Methods

Supplemental Table S3. Primer sequences

Primer	Sequence	Purpose
oAR214	CACCTCGAGatacctttgcgttcgttgaacca	Forward primer for Region 4
oAR215	GCTCGAGcctgtccgctatatcatgcaaatc	Reverse primer for Region 1
oAR497	CACCgattgacatcattgattaatcatg	Forward primer for Region 2
oAR498	caattgggtatttaagaatgaag	Reverse primer for Region 2
oAR499	tgtagtcttggtgttttcctaaat	Reverse primer for Region 4
oAR500	CACCagattttatttagaaaaggttcacaag	Forward primer for Region 3
oAR501	aagattacaaatttatacgaaatgatag	Reverse primer for Region 3
oAR505	CCTCGAGatcaacaagtttgtacaaaaaagct	Forward primer for Gateway sequence with *Xho*I
oAR532	tgcactcccaacaattaagg	Forward primer for Region 1
oAR556	ggaaagtttaacctctttttctcc	Reverse primer for junction region
oAR557	CACCagtggctacacaaaatggtcc	Forward primer for junction region
oAR609	cttaattgttgggagtgcaattgggtatttaagaatgaagatgaaac	Reverse primer for Region 2+ without Dof
oAR719	ggaaagtttaGGctctttttctcc	Forward Primer for mutating GT-1 Trihelix motif
oAR720	ggagaaaaagagCCtaaactttcc	Reverse Primer for mutating GT-1 Trihelix motif
oLH412	ggCCagtttaacctctCCCAAAAAAtaattgttgggagtg	Forward Primer for mutating 3 putative Dof motifs
oXQ6	CGGTACCatcaaccactttgtacaagaaagct	Reverse primer for Gateway sequence with *Kpn*I
oBR210	gtcagtggctacacaaaatTACTTtgaatctaagtttcatcttc	Reverse primer for mutating first GGACC motif
oBR211	gaagatgaaacttagattcaAAGTAattttgtgtagccactgac	Forward primer for mutating first GGACC motif
oBR212	cacattatcttttgacgactaaattTACTTtactccacctatac	Reverse primer for mutating second GGACC motif
oBR213	gtataggtggagtaAAGTAaatttagtcgtcaaaagataatgtg	Forward primer for mutating second GGACC motif
oBR214	ccacctatacaatatacaGGCGCCCGCGGataaatttataat acacccaac	Reverse primer for mutating HD-ZIP motif
oBR215	gttgggtgtattataaatttatCCGCGGGCGCCtgtatattgta taggtgg	Forward primer for mutating HD-ZIP motif

Uppercase letters in primers indicate bases that do not match the template sequence, whereas lowercase letters indicate bases that do match the template sequence. Associated with Methods

Supplemental Table S4. Yeast one hybrid constructs

Yeast Construct	Region of enhancer	Forward primer	Reverse primer	Entry clone
pAR276	Region 1 1-208 bp	oAR215	oAR532	pAR278
pAR275	Region 2 207-499 bp	oAR498	oAR497	pSL1
pAR277	100-bp Junction region 167-2666 bp	oAR557	oAR556	pAR279

Associated with Methods

Supplemental Table S5. EMSA probe sequences

Probe name	Primer sequences
TCP4 binding motif 1	Forward: 5'-TAGATTCAGGACCATTTTGT-3' Reverse: 5'-ACAAAATGGTCCTGAATCTA-3'
TCP4 binding motif 1 mutated	Forward: 5'-TAGATTCAAAGTAATTTTGT-3' Reverse: 5'-ACAAAATTACTTTGAATCTA-3'
TCP4 binding motif 2	Forward: 5'-GTGGAGTAGGACCAATTTAG-3' Reverse: 5'-CTAAATTGGTCCTACTCCAC-3'
TCP4 binding motif 2 mutated	Forward: 5'-GTGGAGTAAAGTAAATTTAG-3' Reverse: 5'-CTAAATTTACTTTACTCCAC-3'
DOF2.2 binding motif	Forward: 5'-TTTTTGGAAAGTTTAACCTCTTTTTCTCCTTA-3' Reverse: 5'-TAAGGAGAAAAAGAGGTTAAACTTTCCAAAAA-3'
DOF2.2 binding motif mutated	Forward: 5'-TTTTTGGCCAGTTTAACCTCTCCCAAAAAATA-3' Reverse: 5'-TATTTTTTGGGAGAGGTTAAACTGGCCAAAAA-3'
ATML1 binding motif	Forward: 5'-AAATTTATATAATTAATTATGTATATTG-3' Reverse: 5'-CAATATACATAATTAATTATATAAATTT-3'
ATML1 binding motif mutated	Forward: 5'-AAATTTATCCGCGGGCGCCTGTATATTG-3' Reverse: 5'-CAATATACAGGCGCCCGCGGATAAATTT-3'

Associated with Methods

Supplemental File S1. FASTA format alignment of Regions 1 and 2 of the giant cell enhancer across Brassicaceae species associated with Supplemental Figure S7**.**

Supplemental Datasets

Supplemental Dataset S1. Yeast one-hybrid results for Region 1

Supplemental Dataset S2. Yeast one-hybrid results for Region 2

Supplemental Dataset S3. Yeast one-hybrid results for the 100-bp Junction Region overlapping the edges of Region 1 and Region 2

Supplemental Dataset S4. Statistical Tables. ANOVA and T-TEST tables for Figures 1, 2, 3, 4, 6, and Supplemental Figure S5.

Structure of yeast one-hybrid dataset files

Sheets 1-12: contain Raw data for OD600 and luminescence for each of the 6x384-well plates containing the transcription factor (TF) collection. Each 384-well plate is generated by indexing 4x96-well plates.

Data sheet: Transfers the raw data in 384-well format to 96-well format (21 plates named U1-U21). Empty plasmid control is in well F4 of each 96-well plate. Negative control (no yeast) is in well F10 of each 96-well plate.

C.O values sheet: Establishes reference values for the data analysis.

Results sheet: uses the information in the data and C.O. values sheets to calculate the luciferase activity for each TF and determine if there is an interaction with the promoter bait

C.O values

In this sheet important reference values are calculated.

1-DO600 background value: calculated as the average value obtained for well F10 in plates U1 to U21.

2-Low and High confidence cut-off values for yeast growth are calculated based on the SD for the average OD600 in well F10. Low confidence is given to those OD600 that fall between 3SD and 6SD of this average value and high confidence to those above 6SD.
3-Luminescence background value: calculated as the average value obtained for well F10 in plates U1 to U21.
4-BASAL Luminescence value: calculated as the average value obtained for well F4 in plates U1 to U21. This well contains cells carrying and empty prey plasmid and the luminescence activity derived from this control is used to calculate the fold induction.
5-Low and High confidence cut-off values for basal luminescence are calculated based on the SD for the average Luminescence in well F04. Low confidence is given to those Luminescence values that fall between 3SD and 6SD of this average value and high confidence to those above 6SD. For this calculation the Average luminescence background is subtracted from each basal luminescence value.
6-BASAL luciferase activity: calculated based on the luminescence and OD600 for well F04 in plates U1 to U21.
7-Cut off values for basal luciferase activity are calculated based on the SD for the average luciferase activity in well F04. This value is used to normalize the luciferase activity across the screen.

Results

1-Columns F, G, H: each luminescence is corrected by subtracting the average luminescence background (cell G29 in the C.O values sheet).
2-Columns I, J, K: each OD600 is corrected by subtracting the average OD600 background (cell C29 in the C.O values sheet).
3-Column L: volume of cells in ml (in this screen all wells have the same cell volume so it has no influence on the final result)
4-Column M: time of reaction. This value is not used when the reporter is luciferase. In our screens it was set to "1" in order not to have an influence on the results.
5-Column N: Luciferase activity calculated as (Luminescence*1000)/(OD600*cell volume*time). Note that if the OD600 is below the average DO600 value (cell C29 in C.O. values sheet) the luciferase activity is not calculated and "FALSE" will be displayed in the corresponding cell).
6-Column O: Average luciferase activity for the plate to which this TF belongs.
7-Column P: Average luciferase activity for all plates.
8-Column Q: Fold-induction for the luciferase activity vs. the average luciferase activity for the plate to which the TF belongs (cell N/cell O)
8-Column R: Fold induction for the luciferase activity vs. the average luciferase activity for ALL plates (cell N/cell P)
9-Column S: Fold-induction for the luciferase activity vs. the average luciferase activity for the empty plasmid control (cell U29 in the C.O values sheet)
10-Column T: Fold-induction for the luciferase activity vs. the average luciferase activity for the empty plasmid control for each 384-well plate (plates U1 to U4, U5 to U8, U9 to U12, …)
11-column V: defines a conclusion for the results in each well. High confidence interactors are defined by OD600, Luminescence and Fold-induction values that fall above the limits defined by the user (Cells C34, M34 and U35 in the C.O values sheet). Low confidence interactors are defined by OD600 falling between the limits defined by cells C32 and C34 in the C.O values sheet, or luminescence readings falling between the limits defined by cells M32 and M34 in the C.O values sheet, or both. Note that regardless of the OD600 or luminescence value no interaction is considered positive unless the fold of induction is above the limit set (cell U35 in the C.O. values sheet).

Looking at High confidence interactions MAY minimize FALSE POSITIVES due to low growth alone.