A giant cell enhancer achieves cell-type specificity through activation via TCP and repression by Dof transcription factors

Lilan Hong[1,2,5], Clint S. Ko[2,3,5], 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: Adrienne Roeder (ahr75@cornell.edu)

Running title: Giant cell specific enhancer dissection

Abstract

Proper pattern formation relies on the tight coordination of cell fate specification and cell cycle regulation in growing tissues. How this can be organized at enhancers that activate gene expression necessary for differentiation is not well understood. One such example is the patterning of the *Arabidopsis thaliana* sepal epidermis where giant cell fate specification is associated with the endoreduplication cell cycle. Previously, we identified an enhancer region capable of driving giant cell-specific expression. 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 achieved primarily through the combination of two elements: an activator and a repressor. TCP transcription factors are involved in activation of non-specific expression throughout the epidermis with higher expression in endoreduplicated giant cells than small cells. Dof transcription factors act via the second element to repress activity of the enhancer

31 and limit expression to giant cells. Thus, we find that cell-type specific expression emerges from
32 the combined activities of two broadly acting enhancer elements.
33
34
35 **Keywords**: Arabidopsis, flower development, sepal, giant cell, enhancer, endoreduplication, Dof,
36 TCP
37
38 **Introduction**
39 Plant and animal organs are composed of different types of cells that perform specialized
40 roles. Cell-type specification is driven by the establishment of distinct patterns of gene expression,
41 as has been confirmed recently by single-cell RNA-seq (Lee & Schiefelbein 2002; Brady et al.
42 2007; Denyer et al. 2019; Ryu et al. 2019; Jaitin et al. 2014). Cell-type specific expression patterns
43 are generally orchestrated through transcription factors binding to enhancers, short non-coding
44 DNA sequences (Long et al. 2016). Enhancers typically contain clusters of transcription factor
45 binding sites, acting as platforms to integrate information (Buecker & Wysocka 2012). The spatial
46 and temporal specificity of transcriptional regulation through enhancers has been extensively
47 studied in several model organisms, including mouse, sea urchins, chicks, and fruit flies (Davidson
48 2010; Spitz & Furlong 2012). Recently, many enhancers have been identified in *Arabidopsis*
49 *thaliana* (henceforth Arabidopsis) and other plants through whole genome methods, including
50 hypersensitivity to DNase I (Zhu et al. 2015; Yan et al. 2019; Oka et al. 2017) and ATAC-seq
51 (Ricci et al. 2020; Tannenbaum et al. 2018). However, relatively little is known about how
52 individual enhancers function mechanistically in plants to generate cell-type specific expression
53 patterns.

54 To elucidate the regulatory logic through which an individual enhancer drives cell-type
55 specific expression, we used the Arabidopsis sepal giant cell model system. Sepals, the outermost
56 floral organ, protect the inner androecium and gynoecium during floral development. On the
57 abaxial epidermis of Arabidopsis sepals, epidermal pavement cells form patterns of two cell types:
58 mitotically dividing small cells and elongated, endoreduplicating giant cells, which scatter between
59 the small cells (Figure 1B) (Roeder et al. 2010). Endoreduplication is a specialized cell cycle
60 wherein cells continue to grow and replicate their DNA but fail to undergo mitosis, generating
61 enlarged, polyploid cells (Traas et al. 1998). Thus, giant cells are a good model system for

62 understanding the intersection of cell-type specification with cell division and growth. Giant cell
63 fate specification and differentiation is promoted by the epidermal specification pathway,
64 including HD-ZIP Class IV transcription factors *Arabidopsis thaliana* MERISTEM LAYER 1
65 (ATML1) and HOMEODOMAIN GLABROUS 11 (HDG11), as well as intercellular signaling
66 proteins DEFECTIVE KERNEL 1 (DEK1) and *Arabidopsis* CRINKLY 4 (ACR4) (Roeder et al.
67 2012). The endomembrane trafficking protein SEC42A restricts the formation of giant cells by
68 modulating this pathway (Qu et al. 2014). Downstream of the epidermal specification pathway,
69 the SIAMESE family Cyclin Dependent Kinase Inhibitor (CKI) LOSS OF GIANT CELLS FROM
70 ORGANS (LGO, also known as SMR1) is required for endoreduplication of giant cells (Roeder
71 et al. 2010; Roeder et al. 2012; Schwarz & Roeder 2016; Meyer et al. 2017). Giant cells affect
72 sepal curvature (Roeder et al. 2012) and are hypothesized to play a role in defense against
73 pathogens based on transcriptomic analysis (Schwarz & Roeder 2016). Stochasticity in gene
74 expression appears to be important for initiating the scattered pattern of giant cells between smaller
75 cells. Fluctuation of ATML1 expression to a high level during the G2 phase of the cell cycle
76 specifies giant cell fate (Meyer et al. 2017).

77 Our previous work described an enhancer region capable of driving giant cell-specific
78 expression in sepals (Roeder et al. 2012). The region was identified based on an enhancer trap line
79 (YJ158) in which the T-DNA was inserted 4.7 kb upstream of AT5G17700 (Figure 1A) (Eshed et
80 al. 2004). AT5g17700 encodes a MATE efflux family protein. The giant cell-specific enhancer
81 (hereafter referred to as the giant cell enhancer) was functionally defined as the 1024 bp region
82 immediately upstream of the enhancer trap T-DNA insertion; this region was sufficient for giant-
83 cell-specific expression in either orientation (Figure 1A) (Roeder et al. 2012). Expression of a
84 reporter driven by the enhancer is regulated by ATML1 and the other members of the giant cell
85 specification pathway (Roeder et al. 2012). In the leaf, the giant cell enhancer activates reporter
86 expression both in giant epidermal pavement cells and leaf margin cells (Roeder et al. 2012; Eshed
87 et al. 2004). The full 4.7 kb upstream region of AT5g17700 that includes the giant cell enhancer
88 drives reporter expression in giant cells of young sepals as well as in other cell types (e.g. cells in
89 petals and the style), suggesting that the giant cell enhancer is part of a larger regulatory region for
90 At5g17700 (Roeder et al. 2012).

91 Here we dissected the 1kb giant cell enhancer into smaller regions and revealed that the

92 enhancer is comprised of an element promoting broad epidermal expression and a region that limits
93 this expression to highly endoreduplicated giant cells in the sepal. A genome-wide yeast one-
94 hybrid screen for transcription factors binding to the enhancer fragments identified TCP and Dof
95 transcription factors, among others. We show that TCP transcription factors are involved in
96 activating a broad epidermal pattern of expression. The Dof transcription factors broadly repress
97 expression, leading to giant cell-specific expression.

98

99 **Results**

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

102 To identify regulatory modules, the 1024 bp giant cell enhancer was dissected into four
103 regions: 1 to 208 bp (Region 1), 209 to 449 bp (Region 2), 450 to 760 bp (Region 3), and 761 to
104 1024 bp (Region 4) (Figure 1A, and S1). Fragments containing these regions or combinations
105 thereof were cloned into a plasmid containing a minimal 35S promoter (mini35S) driving a
106 nuclear-localized yellow fluorescent reporter (3XVenus-N7) (Roeder, et al. 2012). Patterns of
107 Venus reporter expression were characterized in T1 transgenic lines from different constructs to
108 assess the capabilities of the different enhancer fragments to activate transcription. The full 1024
109 bp enhancer (Region 1-2-3-4, contained in *pAR111*) activated expression specifically in giant cells
110 in the sepal (Figure 1B), as previously shown (Roeder et al. 2012). Region 1 alone (in *pAR280*),
111 Region 4 alone (in *pAR256*), or Regions 3 and 4 together (Region 3-4, in *pAR258*) did not activate
112 reporter expression (Figure 1A and Figure S2). Region 2 alone (in *pAR254*), Region 2 with Region
113 3 (Region 2-3, in *pAR261*), and Region 2 with Regions 3 and 4 (Region 2-3-4, in *pAR260*) all
114 drove ubiquitous expression in the epidermis, not specific to giant cells (Figure 1A, D-F). Although
115 the reporter was expressed in all epidermal pavement cells for this broad epidermal expression
116 pattern, the expression level was about 2 fold higher in giant cells than small cells, as quantified
117 by mean fluorescence intensity of the reporter in the nucleus (Figure 1G). Regions 1, 2, and 3
118 together (Region 1-2-3, in *pAR262*) were sufficient for driving reporter expression specifically in
119 endoreduplicated giant cells (Figure 1C).

120 Regions 1 and 2 together (Region 1-2, in *pAR257*) drove a varied expression pattern in
121 independent T1 transgenic lines, among which the majority showed giant cell specificity. To

122 further characterize the expression specificity of Region 1-2, we classified the expression pattern
123 of these T1 transgenic plants into three categories: (1) giant cell specific (type I) (Figure 2A); (2)
124 enhanced signal in giant cells with additional signal in the small cells at the sepal tip (type II, giant
125 cell enhanced) (Figure 2B); (3) broad epidermal signal (type III) (Figure 2C). We quantified the
126 percentage of *pAR257* (Region 1-2) independent T1 transgenic plants showing these three
127 expression patterns (47.8% giant cell specific, 34.8% giant cell enhanced, and 17.4% epidermal),
128 and compared it with those of *pAR111* (Region 1-2-3-4), *pAR254* (Region 2) and *pAR261* (Region
129 2-3) plants. *pAR111* (Region 1-2-3-4) transgenic plants all showed giant cell specific signal. For
130 *pAR254* (Region 2) and *pAR261* (Region 2-3), the majority of the lines had type III, broad
131 epidermal signal, although occasionally, they showed type II, giant cell enhanced expression
132 pattern (Figure 2D). Together, these data suggest that Region 1-2 play an important role in
133 generating the giant cell specificity of enhancer activity. Region 2 is sufficient to produce broad
134 epidermal expression, while Region 1 is required for limiting expression to giant cells. In addition,
135 Region 3 and 4 play a small role in enhancing giant cell specificity when present in combination
136 with Region 1; therefore, we focused our further analysis on Regions 1 and 2.
137
138 **TCP transcription factors promote activity of the enhancer**
139 To identify transcription factors interacting with Regions 1 and 2 of the enhancer, we
140 performed a full-genome yeast one-hybrid screen with a library of 1956 transcription factors. We
141 analyzed Region 1, Region 2, and a 153 bp region overlapping the junction of these regions
142 separately (Figure S1); the overlap was intended to catch interactions that might be disrupted at
143 the boundary between Region 1 and Region 2. We identified 111 high-confidence (Table 1) and
144 43 low-confidence interactions (Table S1) in total for all three assays (Datasets 1-3). We paid
145 particular attention to transcription factors that specifically bound one region and not the other
146 because they might play roles in driving the distinct expression patterns conferred by these regions.
147 In the screen, several Class II CINCINNATA-like TEOSINTE BRANCHED1,
148 CYCLOIDEA, and PCF (TCP) transcription factors (TCP2, TCP3, TCP4, TCP10) interacted with
149 Region 2 and the overlap of the enhancer, but not Region 1. To test for regulation of the enhancer
150 by Class II TCPs, we crossed the full 1kb enhancer reporter (*pAR111*) to the *jaw-1D* mutant
151 (Palatnik et al. 2003). The *jaw-1D* line overexpresses miR319a, a microRNA that targets *TCP2*,
152 *TCP3*, *TCP4*, *TCP10*, and *TCP24*, thus leading to downregulation of these *TCP*s (Palatnik et al.

153 2003). In *pAR111* and *jaw-1D* F1 heterozygotes, sepals had fewer cells expressing the reporter
154 (Figure 3B and 3C), compared with sepals from *pAR111* and wild-type (Col-0) F1 heterozygotes
155 (Figure 3A and 3C). Knockdown of TCPs only affected the expression of the pAR111 reporter
156 and not the presence of giant cells in the sepal, as identified by morphology (Figure 3B, PI stain).
157 In *pAR111* X *jaw-1D* F1 hybrid sepals, the number of giant cells was similar to that in *pAR111* X
158 wild-type F1 sepals (18.6 ± 2.3, mean ± SD, n = 5 in *pAR111* X *jaw-1D* F1; 17 ± 4.1, mean ± SD,
159 n = 5 in *pAR111* X wild-type F1; non-significant difference by t test); however, in *jaw-1D* X
160 *pAR111* F1 sepals, most of these giant cells were not detectably expressing the *pAR111* reporter
161 (Figure 3B and 3C).
162 DNA affinity purification sequencing (DAP-seq) analysis (O'Malley et al. 2016) indicated
163 that TCP transcription factors bind to Region 2 (Figure S3) and throughout the intergenic region
164 upstream of AT5G17700. However, there are no obvious canonical TCP binding sites in Region
165 2. We selected three sites with TGGG, which is half of the consensus TCP binding sequence, as
166 putative TCP binding sites in Region 2 (Figure S4). When all three of the TGGG sites in the Region
167 2 reporter *pAR254* were mutated to TAAA (*pLH173*), the reporter lines expressed almost no Venus
168 signal in sepal cells (Figure 3D and S5A). However, with extremely strong laser power, dim signal
169 could be detected in some *pLH173* sepal cells (Figure S5B). This indicated that disruption of all
170 the TGGG sites significantly reduced transcriptional activity in Region 2 (Figure 3D). Thus, our
171 results suggest that binding of TCPs to Region 2 is critical for this region to activate expression.
172
173 **Lowering the expression level increased giant cell specificity**
174 While we were dissecting the enhancer into regions (Figures 1-2), we noticed that different
175 constructs required different laser intensity to generate the same fluorescence intensity of the
176 reporter in giant cells (as determined by giant cell signal saturation). We recorded the laser
177 intensity that different constructs required in order to reach the same giant cell fluorescence
178 intensity and found that *pAR257* (Region 1-2) needed much higher laser intensity than *pAR254*
179 (Region 2) alone to saturate the fluorescence in giant cells (Figure 3E), suggesting that Region 1
180 reduces transcriptional activity of Region 2.
181 Our analysis suggests that Region 1 reduces the activity of Region 2 and also limits reporter
182 expression mainly to giant cells. We hypothesized that these two functions of Region 1 are
183 connected such that globally reducing expression activity would be sufficient to shift a broad

184 epidermal expression pattern toward a giant cell-specific expression pattern. To test this
185 hypothesis, we lowered the transcriptional activity of Region 2 by crossing the Region 2 reporter
186 (*pAR254*) to the *jaw-1D* line. Plants heterozygous for both *pAR254* and *jaw-1D* (F1) had sepals
187 that displayed a giant cell-specific expression pattern (Figure 3G), instead of the broad epidermal
188 expression pattern in sepals heterozygous for *pAR254* alone (Figure 3F; *pAR254* X wild-type F1
189 heterozygote), which confirmed that reducing Region 2 transcriptional activity increased the
190 specificity of the expression pattern to giant cells. These data are consistent with the hypothesis
191 that Region 1 limits the transcriptional activity driven by Region 2 to giant cells in part by reducing
192 the overall expression level, thus reducing the relatively low expression in small cells (Figure 1G)
193 and allowing only the relatively high expression in giant cells to persist.
194
195 **Dof transcription factors repress expression of the enhancer by lowering the expression level**
196 **and limiting it to giant cells.**
197 We next searched for transcription factor binding sites in Region 1 responsible for
198 repressing expression, which give rise to giant cell specificity. Our yeast one-hybrid screen showed
199 that several DNA-binding One Zinc Finger (Dof) transcription factors bind to Region 1 and the
200 overlap but not Region 2. Based on the AthaMap website (http://www.athamap.de) transcription
201 factor binding site prediction, a 24 bp region 14 bp before the end of Region 1 contained three Dof
202 transcription factor binding sites (AAAG) (Sani et al. 2018; Noguero et al. 2013). We generated a
203 reporter construct containing Region 2 and extending 40 bp into the end of Region 1, which
204 contained all three potential Dof binding sites (Region 2+Dof in *pAR307*, Figure 4A and S4). As
205 a control, we also generated a construct containing Region 2 and 17 bp at the end of Region 1,
206 which excluded the Dof binding sites (*pAR308*, Figure 4A and S4). The expression pattern of
207 *pAR307* also varied among transgenic lines. Categorizing the expression pattern showed that
208 *pAR307* retained most of the giant cell specific expression that Region 1-2 *pAR257* has (Figure
209 4B). Removing the Dof binding sites region led to an obvious shift in the expression pattern, with
210 a majority of the *pAR308* plants showing epidermal expression (Figure 4B), indicating the
211 importance of the Dof binding sites in generating the giant cell-specific expression pattern. Site-
212 directed mutagenesis at these three putative Dof binding sites in the 40 bp region (Region2-Dof
213 mutant, *pLH166*) reproduced the epidermal expression pattern in *pAR308* (Figure 4B). There is
214 also a trihelix site in the 24 bp Dof binding site region, and mutation of the trihelix site did not

alter the expression pattern (data not shown). We investigated the transcriptional activity of 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 of reporter signal strength (detected with a lower power laser; Figure 4C), consistent with our previous observations (Figure 3E).

Based on DAP-seq analysis, several Dof transcription factors bind to the 24 bp Dof binding site in Region 1 (Figure S6). The ones showing the highest binding affinity include: *AtDOF2.2* (*AT2G28810)*, *HIGH CAMBIAL ACTIVITY2* (*HCA2 AT5G62940)*, and *AtDOF5.8* (*AT5G66940)*. We overexpressed each of the genes encoding these Dofs under the constitutive 35S promoter in the *pAR111* full length giant cell enhancer background to assess the impact of Dofs on enhancer activity. Overexpression of each of these Dofs strongly decreased *pAR111* expression in sepal giant cells (Figure 5D-L) compared with the sepal giant cells from the *pAR111* control (Figure 5A-C), indicating that Dof transcription factors repressed transcriptional activity of the giant cell enhancer. Overexpression of the Dofs also severely inhibited sepal development, resulting in shorter, narrow sepals in *35S::AtDOF2.2* and *35S::HCA2* (Figure 5D-E and G-H), and longer, narrow sepals in *35S::AtDOF5.8* (Figure 5J-K). Even in these morphologically altered sepals, giant cells were present, based on size and morphology. A couple 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, I, and L). Thus, our results demonstrate that binding of Dofs to Region 1 reduce overall reporter expression to confer giant cell-specific expression (Figure 4 and 5).

Discussion

In this study, we used giant cells as a model system to investigate how a cell type-specific expression pattern is achieved and reported the dissection of an enhancer region driving a giant-cell-specific expression pattern. This enhancer has a modular organization: Region 2 activated by TCP transcription factors drives expression in all epidermal cells with higher levels of expression in giant cells than small cells, while Region 1 is bound by Dof transcription factors to represses expression, thus limiting expression to giant cells (Figure 6). This repression contributes to cell type specificity due to the differences in expression level between giant cells and small cells driven by Region 2, which allows only the giant cell expression to remain after repression. Intriguingly,

246 neither the activation nor the repression appears to be cell-type specific, yet the combination results
247 in giant cell specificity.

248 Rough dissection (Figure 6C) indicated that transcription factors interacting with Region 2
249 are candidates for activating broad epidermal expression. The results from the large-scale yeast
250 one-hybrid assay suggest 37 transcription factors showed high-confidence interaction with this
251 region, including proteins from the TCP gene family (Table 1). The TCP family is divided into
252 two main classes, based on differences in the TCP domains and binding sites. In general, class II
253 TCP proteins are attributed to functions that inhibit plant cell division and growth, whereas class
254 I proteins promote proliferation. For example, class II protein CINCINNATA (CIN) controls leaf
255 surface curvature in *Antirrhinum* by making cells more sensitive to an arrest signal, inhibiting
256 proliferation (Nath et al. 2003; Crawford et al. 2004). Interestingly, class II protein TCP4 regulates
257 both density and branching of trichomes, another highly endoreduplicated cell type in Arabidopsis
258 (Vadde et al. 2017; Vadde et al. 2019). Class I TCP15 regulates expression of boundary-specific
259 genes (Uberti-Manassero et al. 2011) and TCP20 promotes cell division (C. Li et al. 2005).
260 Intriguingly, TCP15 also regulates endoreduplication in Arabidopsis (Zi-Yu et al. 2011).
261 Moreover, specific repression of TCP15 by fusing to an EAR repressor domain has been reported
262 to result in a loss of giant cells in the abaxial sepal epidermis (Uberti-Manassero et al. 2011). The
263 TCP proteins identified from the screen, which bound to Region 2, (TCP2, TCP3, and TCP4) are
264 all in the class II CINCINATA subfamily (Martín-Trillo & Cubas 2010). As proteins in this
265 subfamily act to repress cell proliferation, especially at the leaf margin, it is plausible, though
266 untested, that these factors serve to coordinate activation of expression at the enhancer with the
267 transition from proliferation to differentiation, since the giant-cell enhancer is only active after
268 cells have stopped dividing (Roeder et al. 2012).

269 Among the yeast one-hybrid candidates, another group of transcription factors that might
270 be involved in giant cell specific expression were the three HD-ZIP Class IVs: HDG11, ANL2,
271 and GL2. HD-ZIP Class IV transcription factors are typically restricted to the L1 and act to specify
272 epidermal cell type identity (Nakamura et al. 2006), making these strong candidates for activating
273 epidermal expression at the enhancer. HDG11 has been shown to have a small role in promoting
274 giant cell formation (Roeder et al. 2012). ATML1, a HD-ZIP Class IV not included in the yeast
275 one-hybrid assay, has been shown to be required for giant cell patterning as well as specification
276 of epidermal identity (Abe et al. 2003; Roeder et al. 2012; Meyer et al. 2017). We have previously

277 shown that overexpression of ATML1 is sufficient to create ectopic giant cells covering the sepal
278 and that these ectopic giant cells express the *pAR111* full length giant cell enhancer reporter
279 (Meyer et al. 2017), suggesting ATML1 may be involved in activating enhancer activity. Enhancer
280 activity is downregulated but not completely absent in *atml1* mutants, consistent with a role for
281 ATML1 in activation of enhancer activity (Roeder et al. 2012). The canonical binding site of
282 ATML1 and other HD-ZIP Class IV transcription factors, the L1 box (TAAATGCA) (Abe et al.
283 2001; Nakamura et al. 2006), does not appear in the enhancer sequence. However, we speculate
284 that HD-ZIP Class IV proteins may recognize another motif in the enhancer or interact with other
285 transcription factors. Non-L1-box regions have been shown to specify epidermal identity; the
286 promoter of ATML1 activates epidermis-specific expression even when its L1 box is deleted
287 (Takada & Jurgens 2007).

288 It is currently unclear what gives rise to higher expression in giant cells than small cells in
289 the broad epidermal expression pattern generated by Region 2. Higher expression in giant cells
290 could be caused by a combination of factors. One is that giant cells are endoreduplicated, generally
291 reaching a ploidy of 16C; thus, higher DNA content and associated increases in cell size could
292 enhance expression (Roeder et al. 2010; Robinson et al. 2018). In trichomes, endoreduplication
293 has been shown to be important for the maintenance of cell identity (Bramsiepe et al. 2010).
294 However, as we have previously described, this giant cell enhancer is not strictly a marker of
295 ploidy and additional factors contributing to increased expression in giant cells are likely (Roeder
296 et al. 2012). HD-ZIP Class IV transcription factor ATML1 reaches higher levels of expression in
297 giant cells than small cells during their specification, which could initiate this pattern of higher
298 expression (Meyer et al. 2017). Finally, it is possible that TCPs or other factors binding to Region
299 2 generate this difference in expression level. Further elucidating the cause of this difference in
300 expression will be an important future goal in understanding the relationship between
301 endoreduplication and cell identity in giant cells.

302 Giant cell specificity arises through the repressive activity of Region 1 (Figure 6C). Both
303 yeast one hybrid and DAP-seq results suggest that Dof transcription factors bind to Region 1 and
304 overexpression analyses and enhancer dissections suggest Dofs repress transcription to create giant
305 cell specificity. Region 1 alone does not activate any expression, suggesting it acts solely as a
306 repressor and detecting its output requires the activators bound to Region 2. Dof transcription
307 factors bind AAAG sites in DNA through a single N-terminal C_2C_2-zinc-finger like motif

308 (Noguero et al. 2013). Dof proteins interact more tightly with the DNA when two proximate
309 binding sites are present than they do with only one binding site (Sani et al. 2018), which is
310 consistent with the three nearby putative Dof sites in Region 1. There are 36 Dof family members
311 in *Arabidopsis* which tend to function redundantly (Yanagisawa 2002; Noguero et al. 2013; Gupta
312 et al. 2015; Yanagisawa 2015). Overexpression of all three *Dof* genes we tested was sufficient to
313 repress expression of the giant cell enhancer. Dof overexpression also altered sepal development,
314 making sepals narrow, and in most cases shorter, suggesting that these *Dof* genes are important for
315 sepal morphogenesis, not just regulation of this enhancer. Dofs have been shown to regulate many
316 developmental processes and environmental responses including seed germination, light response,
317 vascular cambium development, and cell division (Noguero et al. 2013). Dofs can act as either
318 activators or repressors in different contexts (Noguero et al. 2013). The Dof transcription factor
319 OPB1 promotes cell cycle entry by directly activating CYCLIN D3;3 and AtDOF2.3 which is
320 associated with replication (Skirycz et al. 2008). In seeds, DOF3.2 binds to the DELLA co-
321 repressor RGA-like 2 (RGL2) and DOF AFFECTING GERMINATION 1 (DAG1) binds to the
322 DELLA corepressor GA INSENSITIVE (GAI) to repress seed germination in the absence of
323 gibberellin (GA), linking Dofs with GA signaling (Ravindran et al. 2017; Boccaccini et al. 2014;
324 Ruta et al. 2020). This observation raises the possibility that the repressive activity conferred by
325 the Dofs could be mediated by co-repressors. Of the Dofs we tested, the functions of DOF2.2
326 (AT2g28810) remains unknown (Yanagisawa 2015). HCA2/DOF5.6 (AT5G62940) promotes the
327 development of the vascular cambium and consequently the radial growth of the root (Guo et al.
328 2009; Miyashima et al. 2019). DOF5.8 (AT5G66940) regulates ANAC069 in response to abiotic
329 stress (He et al. 2020). In the future, it will be interesting to determine which Dofs are required for
330 the repressive activity of Region 1 in wild type sepals, since our overexpression analysis have
331 shown that all of three Dofs we tested, despite coming from different clades, were sufficient to
332 repress activity (Yanagisawa 2002; Yanagisawa 2015). Given the binding of multiple Dofs,
333 including the three we tested, to this enhancer region in the DAP-seq database, we expect multiple
334 Dof proteins to act redundantly in this role.

335 Enhancers tend to be conserved across closely related species because selection preserves
336 their function whereas nonfunctional intergenic sequences are not under selection and diverge
337 faster. This conservation has been used to identify enhancer sequences in non-coding DNA
338 (Pennacchio et al. 2006). The giant cell enhancer includes a previously identified conserved

339 noncoding sequence (Haudry et al. 2013). Alignment of Region 1-2 enhancer sequences from six
340 species across the Brassicaceae family revealed that two of the three putative Dof binding sites we
341 identified are conserved, and all three of the putative TCP binding sites are conserved in
342 *Arabidopsis lyrata*, *Cardamine hirsuta*, *Brassica rapa*, *Sisimbrium irio*, and *Eutrema salsugineum*
343 (Figure S7). In the future, it will be interesting to determine whether these enhancers produce
344 similar expression patterns in these other species.

345 Previous studies in animals have often shown spatial restriction of enhancer activity is
346 generated by broad activation combined with localized repression (Spitz & Furlong 2012). For
347 example, during early stages of sea urchin development, the *endo16* gene is precisely expressed in
348 the endoderm through broad activation in the embryo by the enhancer Region A combined with
349 localized repression in the neighboring ectoderm by Regions E and F and localized repression in
350 the neighboring skeletogenic mesenchyme by Region DC (Davidson 2010). Likewise, in
351 *Drosophila* the *even-skipped* (*eve*) stripe two enhancer is activated broadly by Bicoid (Bcd) and
352 Hunchback (Hb), and narrowed to a precise stripe through repression by Giant (Gt) on the anterior
353 and Krüppel (Kr) on the posterior (Stanojevic et al. 1991). Genetic analysis suggests a similar
354 regulatory logic is found in Arabidopsis during the establishment of the narrow band of
355 *SHATTERPROOF* (*SHP*) expression precisely at the valve margin through repression by
356 FRUITFULL (FUL) in valves and REPLUMLESS (RPL) in the replum (Ferrándiz et al. 2000;
357 Roeder et al. 2003). Our findings in the giant cell enhancer are distinct in that both the activator
358 and repressor do not appear to be spatially specific, and yet the combination gives rise to cell type
359 specificity. Broad activation of the enhancer in all epidermal cells through a combination of TCP
360 and other transcription factors likely including HD-ZIP Class IV transcription factors generates a
361 pattern with high expression in giant cells and low expression in small cells. The Dof transcription
362 factors repress expression by lowering levels until the enhancer no longer activates expression in
363 small cells, and only expression in giant cells remains detectable (Figure 6). Thus, we suggest that
364 cell-type specificity can emerge from the combination of two modules neither of which is itself
365 cell-type specific.
366

367 **Methods**

368

369 **Plant Growth**

370 *Arabidopsis thaliana* Col-0 plants were grown on Lambert Mix LM-111 soil in Percival growth
371 chambers at 22 °C with 24 hours fluorescent light (~100 µmol m^{-2} s^{-1}) conditions. Seeds were
372 planted on soil and cold stratified for about two days before placing in the growth chambers.
373
374 **Giant cell enhancer dissection**
375 Cloning of the full length 1024 bp enhancer construct pAR111 and verification that it was
376 sufficient to drive giant cell-specific expression in either orientation was previously described
377 (Roeder et al. 2012). This 1024 bp region was divided into four arbitrary regions with Region 1
378 falling closest to the downstream gene AT5G17700 and farthest from the original enhancer trap
379 insertion YJ158 (Figure 1). Enhancer fragments were tested for their activity in the orientation
380 they would have relative to downstream gene AT5G17700 such that Region 1 is closest to the
381 reporter gene.
382 We first cloned a gateway destination reporter vector (pSL12) so that we could rapidly
383 clone and assay the reporter expression patterns driven by fragments of the giant cell enhancer.
384 pSL12 contains a gateway site upstream of the -60 minimal promoter form 35S and the 3XVenus-
385 N7 super-bright yellow-fluorescent nuclear-localized reporter in the pMLBart binary vector
386 backbone. pSL12 confers Basta resistance in plants and spectinomycin resistance in bacteria. To
387 create pSL12, the gateway site in which the *Not*I restriction site had been removed was amplified
388 with oAR505 and oXQ6 and cloned into pGEM-T Easy (Promega) to create pSL10. The gateway
389 site was cut from pSL10 with *Xho*I and *Kpn*I and cloned into -60 3XVenus-N7-BJ36 (Roeder et
390 al. 2012) to create pSL11. pSL11 was cut with *Not*I and the gateway -60 3XVenus-N7 fragment
391 was cloned into pMLBart to create pSL12.
392 Fragments of the enhancer were amplified through PCR with Pfu Ultra II (Agilent) or
393 Phusion (NEB) from the *Arabidopsis thaliana* genome BAC clone MAV3 or Col-0 genomic DNA
394 using the primers specified in Table S2 for each fragment. The primers sequences are listed in
395 Table S3. The PCR products were cloned into pENTR D TOPO (ThermoFisher) according to the
396 manufacturer's instructions to create the entry clones listed in Table S2. LR reactions
397 (ThermoFisher) between the entry clones and pSL12 generated the final constructs listed in Table
398 S2. Putative Dof and TCP sites were mutated through changes in the primer sequences. All
399 constructs were verified by sequencing the inserts.

400 These constructs were transformed into Columbia (Col-0) plants through *Agrobacterium*-
401 mediated floral dipping and were selected for Basta Resistance. For most constructs the expression
402 patterns in approximately 20 T1 transgenic plants were analyzed. Note the original pAR111
403 transgenic plants described in (Roeder et al. 2012) were in the Landsberg *erecta* (L*er*) accession,
404 but for this project pAR111 was transformed into Col-0. Transformation of the empty vector
405 pSL12 into plants generated no expression (Figure S2D). As typical with the independent insertion
406 of the T-DNA into the genome (Schubert et al. 2004), we saw differences in expression level and
407 sometimes pattern between different T1 lines as characterized in the results. However, within a
408 single plant the expression pattern was similar across sepals. Likewise the expression pattern was
409 consistent in subsequent T2 and T3 generations for those lines examined.
410
411 **Confocal Microscopy**
412 To minimize the morphological or gene expression variability caused by different plant vitality,
413 the 12th to 25th flowers on the main stem were used for observation (Hong et al. 2016). Mature
414 sepals (stage 14 according to (Smyth et al. 1990)) were dissected from the flowers with tweezers
415 and needles and stained for 15 minutes in propidium iodide (PI; 0.15 mg/ml in water), mounted
416 on slides in 0.01% triton X-100 under a cover slip, and imaged with a Zeiss 710 confocal laser
417 scanning microscope. A 514 nm excitation laser was used to excite both the fluorophores. The
418 3xVenus-N7 enhancer reporter signal was collected between 519 to 566 nm, while the PI signal
419 was collected between 603-650 nm. Images were taken using 10X (Plan-APOCROMAT NA =
420 0.45 air) or 20X (Plan-APOCROMAT NA = 1.0 water dipping) objective lenses. Different T1
421 transgenic lines have different levels of expression likely due to effects of the position of insertion
422 of the T-DNA. For the experiments in Figures 1, 2, and 4 we varied the laser power to achieve
423 saturation of the reporter in a few of the brightest giant cell nuclei (the other settings including
424 gain were kept the same). This allowed us to compare reporter expression between cells while
425 keeping the expression level in giant cells relatively constant (ranges of laser power used are
426 reported in Figure 3E). In contrast, Figure S5 shows the same sepal imaged with different laser
427 powers (from 3% and 40% as indicated in the figure), to detect very low levels of expression. For
428 most images the maximum intensity projection is shown. For Figure 5C,F,I, and L the image was
429 volume rendered in MorphoGraphX and subsequently Auto Contrast was used in photoshop to
430 visualize the epidermal cell layer without the underlying layers.

431

432 **Quantitative Image Processing**

433 FIJI (https://fiji.sc) with the Costanza plugin (http://home.thep.lu.se/~henrik/Costanza/) was used
434 to quantify the fluorescence intensity and size of nuclei expressing pAR254. In the images used,
435 the fluorescence intensity of the reporter in the nuclei was not saturated. The .lsm stack image
436 from the microscope was opened in FIJI. The color of Channel 1 (with the reporter expression)
437 was converted to greys using the Channels tool. The channels were split using Split Channels. The
438 maximum intensity projection of Channel 1 (reporter expression) was created with Z project. The
439 maximum intensity projection image was analyzed with the Costanza plugin with the following
440 settings. In the general menu, "Mark intensity plateau with single maximum" was checked as were
441 "Mark cell centers. Marker pixel radius: 3", Display basins of attractions (BOA)", and "Display
442 basins of attractions according to measured intensity." In the pre-processor queue, "Background
443 extraction" with an intensity threshold of 20 was executed first followed by "Mean filter" with
444 radius 0.1 and number of times 10. In the post-processor, "BOA remover" with a size threshold of
445 10 and an intensity threshold of 10 was executed first followed by a "BOA merger" with a radius
446 of 15. Use ImageJ stack calibration was used for scaling. The results were analyzed in Microsoft
447 Excel. Giant cells have enlarged nuclei, so a threshold of 100 μm^2 was used to separate giant cell
448 from small cell nuclei.

449

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

455

456 **Statistics**

457 Two-tailed students T-tests were used for pairwise comparisons.

458

459 **High Throughput Yeast-One-Hybrid Screen**

460 Yeast-one-hybrid screens were conducted with a nearly genome-wide transcription factor
461 collection arrayed in 384 well plates (Pruneda-Paz et al. 2014) according to the protocol in (Z. Li

462 et al. 2019) using a luciferase reporter (Bonaldi et al. 2017). Bait plasmids were created through
463 LR reactions of entry clones containing Region 1, Region 2 or a region overlapping the junction
464 between Regions 1 and 2 into the pY1-gLUC59_GW yeast reporter plasmid (Table S4). The
465 resulting constructs were integrated into the URA3 locus on the chromosome of the yeast strain
466 YM4271.
467
468 **Genetic interaction with *jaw-1D***
469 The *jaw-1D* mutant (CS6948) was ordered from the Arabidopsis Biological Resource Center
470 (ABRC). The homozygous *jaw-1D* mutant was crossed to *pAR111* plants and to *pAR254* plants.
471 The expression patterns were analyzed in the F1 generation which were heterozygous for the
472 reporter and the *jaw-1D* dominant mutant. The expression patterns were compared with the F1
473 generation of the cross of each reporter with Col-0 wild type as a control.
474
475 **Dof overexpression**
476 Entry clones containing Dof family member cDNAs were ordered from the ABRC: TOPO-U09-
477 G04 (AT2G28810), TOPO-U20-E12 (AT5G62940), and TOPO-U13-H01 (AT5G66940).
478 Through LR reactions these were recombined into the binary 35S destination vector pK7WG2 to
479 make pLH160 (35S::AT2G28810 CDS), pLH162 (35S::AT5G62940 CDS), and pLH163
480 (35S::AT5G66940 CDS). These constructs were transformed into pAR111 (in Col-0) plants
481 through agrobacterium-mediated floral dipping and were selected for kanamycin resistance.
482
483 **Flower pictures**
484 Images of flowers were taken on a Zeiss Stemi 2000-C stereomicroscope with a Cannon Powershot
485 A640 digital camera.
486
487 **DAP-seq database analysis**
488 The C_2C_2-Dof and TCP transcription factor families DAP-seq data was analyzed from the Plant
489 Cistrome Database (http://neomorph.salk.edu/dev/pages/shhuang/dap_web/pages/index.php
490 (O'Malley et al. 2016)) using the Genome Browser to examine the giant cell enhancer region on
491 chromosome 5 coordinates 5837111-5837559 for Region 1-2.
492

493 **Alignment of the giant cell enhancer from Brassicaceae species**

494 A conserved noncoding sequence (Chr5:5,837,330-5,837,417) has been identified among
495 Brassicaceae species (Haudry et al. 2013) overlapping with Region 2 of the giant cell enhancer
496 (http://mustang.biol.mcgill.ca:8885/cgi-bin/hgGateway). To examine sequence conservation of
497 Regions 1 and 2 of the enhancer and the putative Dof and TCP binding sites we aligned the
498 corresponding sequences from *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Cardamine hirsuta*
499 (CoGe id 36106 Cardamine hirsute v1 unmasked Chromosome 6 18239653-18239154), *Brassica*
500 *rapa* (COGE id 24668 Brasica DB Chr unmasked v1.5 chromosome A10 11517566-11517500 and
501 A02 3259372-3259603), *Eutrema parvulum* (formerly *Thellungiella parvula* 12384 UIUC
502 unmasked v2 chomosome 6-6 5884609-5883951), *Eutrema salsugineum* (Id 19492 JGI unmasked
503 6007159-6007397), *Sisimbrium irio* (Id 20245 VEGI unmasked vVEGI 2.5 Chromosome
504 scaffold_57 2159183-2159617). All sequences were checked that they fell in syntenic regions
505 upstream of a homolog of AT5g17700. The sequences were aligned with Clustal Omega
506 (https://www.ebi.ac.uk/Tools/msa/clustalo/) and the alignments were formatted and displayed
507 with Boxshade (https://embnet.vital-it.ch/software/BOX_form.html). Alignments were adjusted
508 slightly by hand around gaps and the ends of sequences were trimmed or extended to give the best
509 alignment (Figure S7).

510

511 **Accession Numbers**

512 Giant cell enhancer associated MATE efflux family protein, AT5g17700.

513 *jaw-1D* seed (CS6948) and associated gene miR319a (*AT4G23713*)

514 Dof family members analyzed: AT2G28810, AT5G62940, AT5G66940

515

516 **Distribution of Materials and Data**

517 Requests for materials and data should be addressed to Adrienne H. K. Roeder
518 (ahr75@cornell.edu).

519

520 **Author Contributions**

521 Conception and design of experiments: L.H., C.S.K., A.H.K.R.

522 Giant cell enhancer dissection experiments and analysis: L.H., C.S.K., A.H.K.R.

523 High throughput yeast-one-hybrid screen: S.E.K., J.L.P.-P.

524 Writing of the manuscript: L.H., A.H.K.R.

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

526

527 **Acknowledgements**

528 We thank Joseph Cammarata, Frances Clark, Kate Harline, and Shuyao Kong for critical reading
529 and comments on the manuscript. We thank Samuel Leiboff and Dana Robinson for technical
530 assistance in cloning constructs for this project.
531 This work was supported by NSF IOS Plant, Fungal, and Microbial Developmental Mechanisms
532 grants IOS-1256733 (AHKR) and IOS-1553030 (AKHR), NSF IOS 1755452 (JLPP), NSF MCB
533 1158254 (JLPP) and NIH R01GM056006 (JLPP). No conflicts of interest declared.

534

535 **References**

536 Abe, M. et al., 2003. Regulation of shoot epidermal cell differentiation by a pair of
537 homeodomain proteins in *Arabidopsis*. *Development*, 130(4), pp.635–643.

538 Abe, M., Takahashi, T. & Komeda, Y., 2001. Identification of a *cis*-regulatory element for L1
539 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein.
540 *The Plant Journal*, 26(5), pp.487–494.

541 Boccaccini, A. et al., 2014. The DOF Protein DAG1 and the DELLA Protein GAI Cooperate in
542 Negatively Regulating the AtGA3ox1 Gene. *Molecular plant*, 7(9), pp.1486–1489.

543 Bonaldi, K. et al., 2017. Novel cell surface luciferase reporter for high-throughput yeast one-
544 hybrid screens. *Nucleic acids research*, 45(18), p.e157.

545 Brady, S.M. et al., 2007. A high-resolution root spatiotemporal map reveals dominant expression
546 patterns. *Science*, 318(5851), pp.801–806.

547 Bramsiepe, J. et al., 2010. Endoreplication controls cell fate maintenance. *PLoS genetics*, 6(6),
548 p.e1000996.

549 Buecker, C. & Wysocka, J., 2012. Enhancers as information integration hubs in development:
550 lessons from genomics. *Trends in Genetics*, 28(6), pp.276–284.

551 Crawford, B.C.W. et al., 2004. CINCINNATA controls both cell differentiation and growth in
552 petal lobes and leaves of Antirrhinum. *Plant Physiology*, 135(1), pp.244–253.

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

555 Denyer, T. et al., 2019. Spatiotemporal Developmental Trajectories in the Arabidopsis Root
556 Revealed Using High-Throughput Single-Cell RNA Sequencing. *Developmental cell*, 48(6),
557 pp.840–852.e5.

558 Eshed, Y. et al., 2004. Asymmetric leaf development and blade expansion in Arabidopsis are
559 mediated by KANADI and YABBY activities. *Development*, 131(12), pp.2997–3006.

560 Ferrándiz, C., Liljegren, S.J. & Yanofsky, M.F., 2000. Negative regulation of the
561 SHATTERPROOF genes by FRUITFULL during Arabidopsis fruit development. *Science*,
562 289(5478), pp.436–438.

563 Guo, Y. et al., 2009. Dof5.6/HCA2, a Dof Transcription Factor Gene, Regulates Interfascicular
564 Cambium Formation and Vascular Tissue Development in Arabidopsis. *The Plant cell*,
565 21(11), pp.3518–3534.

566 Gupta, S. et al., 2015. Insights into structural and functional diversity of Dof (DNA binding with
567 one finger) transcription factor. *Planta*, 241(3), pp.549–562.

568 Haudry, A. et al., 2013. An atlas of over 90,000 conserved noncoding sequences provides insight
569 into crucifer regulatory regions. *Nature Genetics*, pp.1–10.

570 He, L. et al., 2020. ATDOF5.8 protein is the upstream regulator of ANAC069 and is responsive
571 to abiotic stress. *Biochimie*, 110, pp.17–24.

572 Hong, L. et al., 2016. Variable cell growth yields reproducible organ development through
573 spatiotemporal averaging. *Developmental cell*, 38(1), pp.15–32.

574 Jaitin, D.A. et al., 2014. Massively parallel single-cell RNA-seq for marker-free decomposition
575 of tissues into cell types. *Science*, 343(6172), pp.776–779.

576 Lee, M.M. & Schiefelbein, J., 2002. Cell pattern in the Arabidopsis root epidermis determined
577 by lateral inhibition with feedback. *The Plant cell*, 14(3), pp.611–618.

578 Li, C. et al., 2005. Arabidopsis TCP20 links regulation of growth and cell division control
579 pathways. *Proceedings of the National Academy of Sciences of the United States of America*,
580 102(36), pp.12978–12983.

581 Li, Z. et al., 2019. High-Throughput Yeast One-Hybrid Screens Using a Cell Surface gLUC
582 Reporter. *Current Protocols in Plant Biology*, 4(1), p.e20086.

583 Long, H.K., Prescott, S.L. & Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional
584 Enhancers in Development and Evolution. *Cell*, 167(5), pp.1170–1187.

585 Martín-Trillo, M. & Cubas, P., 2010. TCP genes: a family snapshot ten years later. *Trends in*
586 *Plant Science*, 15(1), pp.31–39.

587 Meyer, H.M. et al., 2017. Fluctuations of the transcription factor ATML1 generate the pattern of
588 giant cells in the *Arabidopsis* sepal. *eLife*, 6, p.e19131.

589 Miyashima, S. et al., 2019. Mobile PEAR transcription factors integrate positional cues to prime
590 cambial growth. *Nature*, pp.1–23.

591 Nakamura, M. et al., 2006. Characterization of the class IV homeodomain-Leucine Zipper gene
592 family in Arabidopsis. *Plant Physiology*, 141(4), pp.1363–1375.

593 Nath, U. et al., 2003. Genetic control of surface curvature. *Science*, 299(5611), pp.1404–1407.

594 Noguero, M. et al., 2013. The role of the DNA-binding One Zinc Finger (DOF) transcription
595 factor family in plants. *Plant Science*, 209, pp.32–45.

596 Oka, R. et al., 2017. Genome-wide mapping of transcriptional enhancer candidates using DNA
597 and chromatin features in maize. pp.1–24.

598 O'Malley, R.C. et al., 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA
599 Landscape. *Cell*, 165(5), pp.1280–1292.

600 Palatnik, J.F. et al., 2003. Control of leaf morphogenesis by microRNAs. *Nature*, 425(6955),
601 pp.257–263.

602 Pennacchio, L.A. et al., 2006. In vivo enhancer analysis of human conserved non-coding
603 sequences. *Nature*, 444(7118), pp.499–502.

604 Pruneda-Paz, J.L. et al., 2014. A Genome-Scale Resource for the Functional Characterization of
605 Arabidopsis Transcription Factors. *Cell Reports*, 8(2), pp.622–632.

606 Qu, X., Chatty, P.R. & Roeder, A.H.K., 2014. Endomembrane trafficking protein SEC24A
607 regulates cell size patterning in Arabidopsis. *Plant Physiology*, 166(4), pp.1877–1890.
608 Available at: http://www.plantphysiol.org/cgi/doi/10.1104/pp.114.246033.

609 Ravindran, P. et al., 2017. A Novel RGL2-DOF6 Complex Contributes to Primary Seed
610 Dormancy in Arabidopsis thaliana by Regulating a GATA Transcription Factor. *Molecular*
611 *plant*, 10(10), pp.1307–1320.

612 Ricci, W.A. et al., 2020. Widespread long-range cis-regulatory elements in the maize genome.
613 *Nature Plants*, pp.1–18.

614 Robinson, D.O. et al., 2018. Ploidy and Size at Multiple Scales in the Arabidopsis Sepal. *The*
615 *Plant Cell*, 30, pp.2308–2329.

616 Roeder, A.H.K. et al., 2012. Cell cycle regulates cell type in the *Arabidopsis* sepal. *Development*,
617 139, pp.4416–4427.

618 Roeder, A.H.K. et al., 2010. Variability in the control of cell division underlies sepal epidermal
619 patterning in *Arabidopsis thaliana*. *PLoS biology*, 8(5), p.e1000367.

620 Roeder, A.H.K., Ferrándiz, C. & Yanofsky, M.F., 2003. The role of the REPLUMLESS
621 homeodomain protein in patterning the *Arabidopsis* fruit. *Current biology*, 13(18), pp.1630–
622 1635.

623 Ruta, V. et al., 2020. The DOF Transcription Factors in Seed and Seedling Development. *Plants*,
624 9(2), pp.218–15.

625 Ryu, K.H. et al., 2019. Single-Cell RNA Sequencing Resolves Molecular Relationships Among
626 Individual Plant Cells. *Plant Physiology*, 179(4), pp.1444–1456.

627 Sani, H.M. et al., 2018. Expression, purification and DNA-binding properties of zinc finger
628 domains of DOF proteins from Arabidopsis thaliana. *Tabriz University of Medical Sciences*,
629 8(3), pp.167–176.

630 Schubert, D. et al., 2004. Silencing in Arabidopsis T-DNA Transformants: The Predominant
631 Role of a Gene-Specific RNA Sensing Mechanism versus Position Effects. *The Plant cell*,
632 16(10), pp.2561–2572.

633 Schwarz, E.M. & Roeder, A.H.K., 2016. Transcriptomic Effects of the Cell Cycle Regulator
634 LGO in Arabidopsis Sepals. *Frontiers in plant science*, 7(e0129), pp.635–22.

635 Skirycz, A. et al., 2008. The DOF transcription factor OBP1 is involved in cell cycle regulation
636 in Arabidopsis thaliana. *The Plant Journal*, 56(5), pp.779–792.

637 Smyth, D.R., Bowman, J.L. & Meyerowitz, E.M., 1990. Early flower development in
638 *Arabidopsis*. *The Plant cell*, 2(8), pp.755–767.

639 Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to
640 developmental control. *Nature reviews Genetics*, pp.1–14.

641 Stanojevic, D., Small, S. & Levine, M., 1991. Regulation of a segmentation stripe by
642 overlapping activators and repressors in the Drosophila embryo. *Science*, 254(5036),
643 pp.1385–1387.

644 Takada, S. & Jurgens, G., 2007. Transcriptional regulation of epidermal cell fate in the
645 Arabidopsis embryo. *Development*, 134(6), pp.1141–1150.

646 Tannenbaum, M. et al., 2018. Regulatory chromatin landscape in Arabidopsis thaliana roots
647 uncovered by coupling INTACT and ATAC-seq. *Plant Methods*, pp.1–12.

648 Traas, J. et al., 1998. Endoreduplication and development: rule without dividing? *Current*
649 *opinion in plant biology*, 1(6), pp.498–503.

650 Uberti-Manassero, N.G. et al., 2011. The class I protein AtTCP15 modulates plant development
651 through a pathway that overlaps with the one affected by CIN-like TCP proteins. *Journal of*
652 *experimental botany*, 63(2), pp.809–823.

653 Vadde, B.V.L. et al., 2019. The TCP4 Transcription Factor Directly Activates
654 TRICHOMELESS1and 2and Suppresses Trichome Initiation. *Plant Physiology*, 181(4),
655 pp.1587–1599.

656 Vadde, B.V.L., Challa, K.R. & Nath, U., 2017. The TCP4 transcription factor regulates trichome
657 cell differentiation by directly activating GLABROUS INFLORESCENCE STEMSin
658 Arabidopsis thaliana. *The Plant Journal*, 93(2), pp.259–269.

659 Yan, W. et al., 2019. Dynamic control of enhancer activity drives stage-specific gene expression
660 during flower morphogenesis. *Nature communications*, pp.1–16.

661 Yanagisawa, S., 2015. *Structure, Function, and Evolution of the Dof Transcription Factor*
662 *Family*, Elsevier Inc.

663 Yanagisawa, S., 2002. The Dof family of plant transcription factors. *Trends in Plant Science*,
664 7(12), pp.555–560.

665 Zhu, B. et al., 2015. Genome-Wide Prediction and Validation of Intergenic Enhancers in
666 Arabidopsis Using Open Chromatin Signatures. *The Plant Cell*, 27(9), pp.2415–2426.

667 Zi-Yu, L., Bin, L. & Ai-Wu, D., 2011. The Arabidopsis Transcription Factor AtTCP15 Regulates
668 Endoreduplication by Modulating Expression of Key Cell-cycle Genes. *Molecular plant*,
669 5(1), pp.270–280.

670

671 **Supplemental Information**

672 Dataset S1: Yeast-one-hybrid results for Region 1.

673 Dataset S2: Yeast-one-hybrid results for Region 2.

674 Dataset S3 : Yeast-one-hybrid results for the overlap between Regions 1 and 2.

675

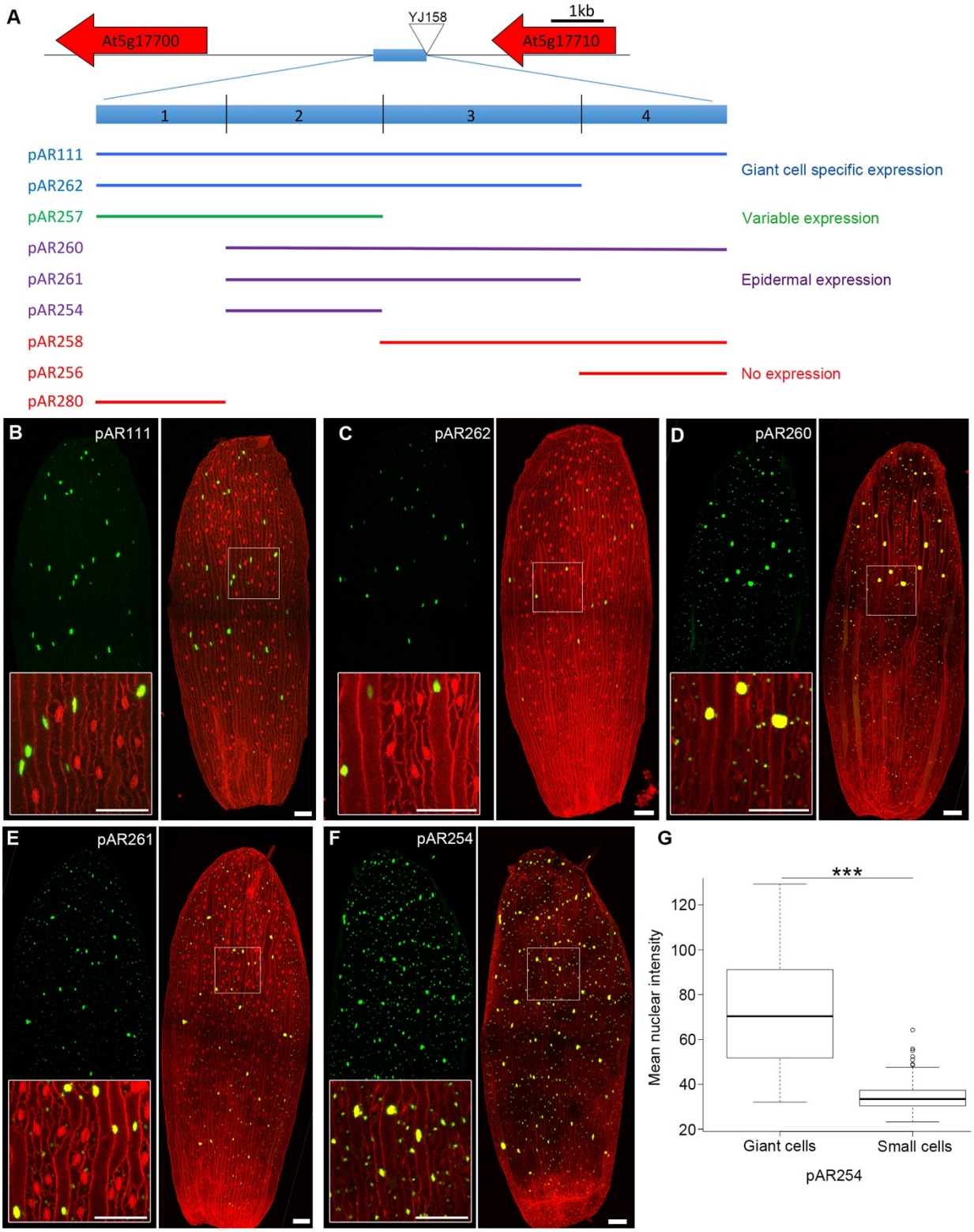


Figure 1. Dissection of 1kb giant cell enhancer.
(A) The giant cell enhancer is a 1 kb sequence 3.2 kb upstream of *AT5G17700*, a gene encoding a MATE efflux family protein. The 1 kb enhancer was originally identified because it flanked the YJ158 enhancer

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 expression of the nuclear localized 3XVenus-N7 reporter under 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, green for expression patterns that vary between giant cell-specific and broad epidermal independent T1 lines (further classified in Figure 2), and red for no expression.

(B-F) Confocal images of stage 14 sepals from different reporter lines. Images on the left, 3XVenus (green) marking the nuclei of cells expressing the reporter; Images on the right, 3XVenus (green) signals merged with propidium iodide (PI, red) stained cell wall signals. Insets overlapping the base of the green channel images, show magnified views of the cells outlined with the box. Expression of the 3XVenus is restricted to giant cells in reporter lines *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 red oblongs. Scale bars: 100 µM.

(G) Mean fluorescent intensity of the 3XVenus reporter in giant cell nuclei versus small cell nuclei of pAR254 sepals, which is expressed in a broad epidermal pattern. Note that the expression (intensity) is about 2-fold higher in giant cell nuclei. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend past 1.5 of the interquartile range. ***$p < 0.001$, significant difference by two tailed t-test, $p = 4.29$ E-13; n=210 small cells and 48 giant cells.

See also Figures S1 and S2.

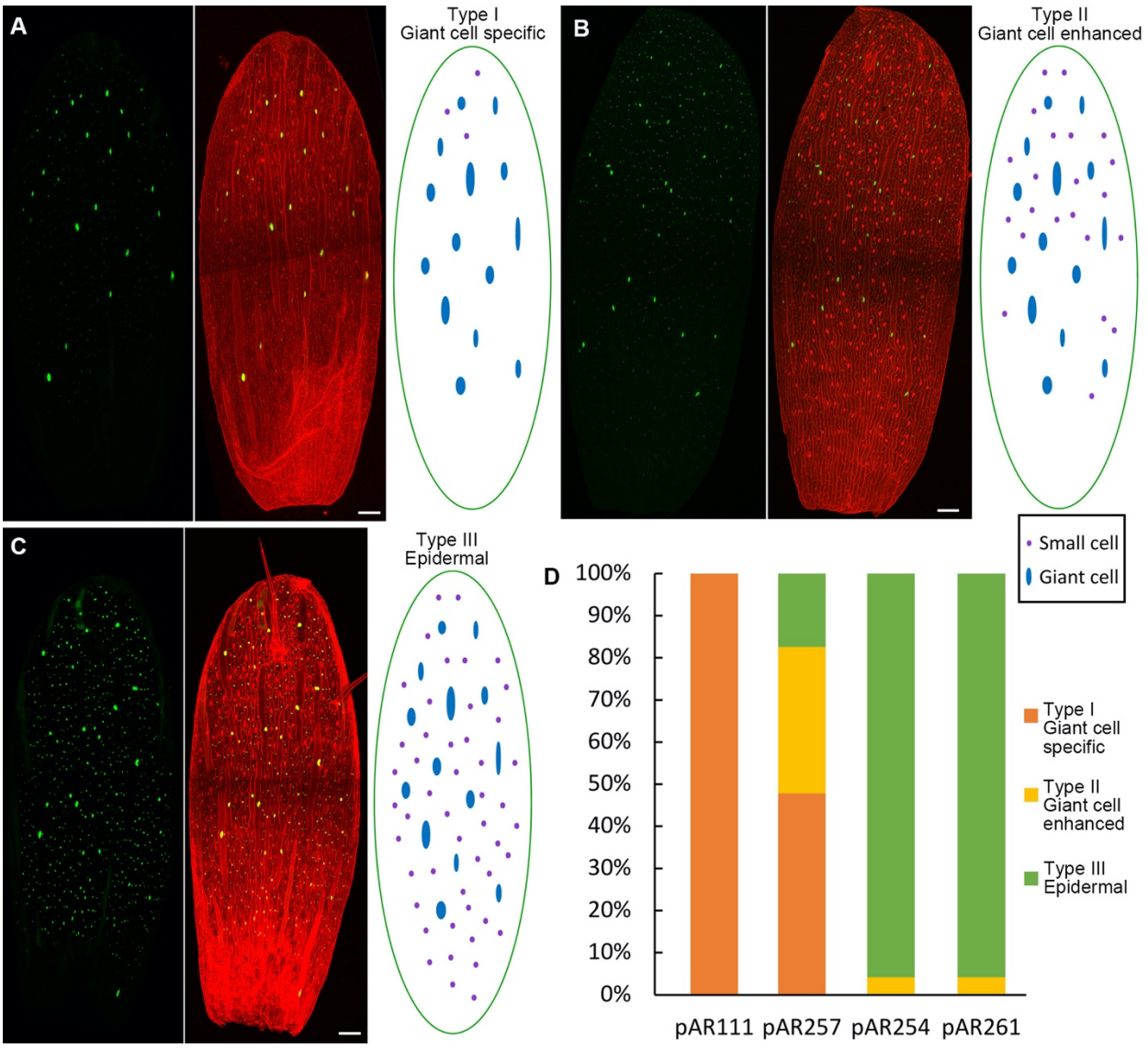


Figure 2. The reporter lines show three types of expression pattern.

(A-C) The three types of expression pattern detected from reporter lines with *pAR257* (Region 1-2). Images on the left, 3XVenus (green) marking the nuclei of cells expressing the reporter; images in the middle, 3XVenus (green) signals merged with PI (red) stained cell wall signals; images on the right, schematic summarizing the pattern. In type I (giant cell specific) expression pattern, 3XVenus is expressed mainly in giant cells (A). In type II (giant cell enhanced) expression pattern, 3XVenus is expressed both in giant cells and in small cells located in the tip part (B). In type III (epidermal) expression pattern, 3XVenus is expressed broadly throughout the epidermis (C). Scale bars: 100 µM.

(D) Percentages of the three types of reporter expression patterns observed across different reporter lines. *pAR111* (Region 1-2-3-4) sepals all show the type I pattern. *pAR254* (Region 2) and *pAR261* (Region 2-3) have predominantly type III expression pattern.

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

Gene family	Region 1	Overlap	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 Table S1

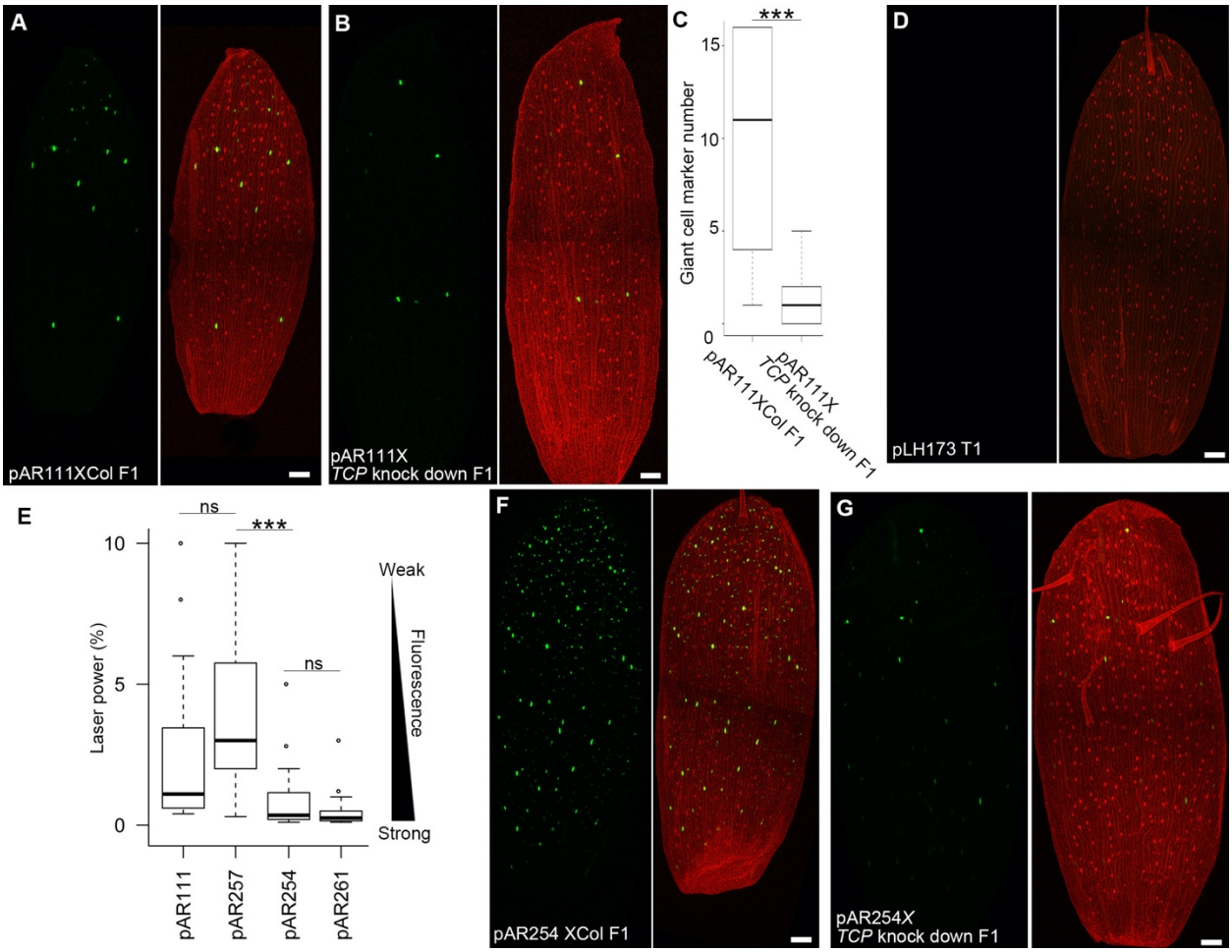


Figure 3. TCP transcription factors affect expression activity.

(A) Confocal images of a sepal from *pAR111* and Col-0 (Col) hybrid plants.

(B) Confocal images of a sepal from *pAR111* and *TCP* knock-down (*jaw-1D*) hybrid plants.

(C) The number of nuclei expressing 3XVenus in *pAR111* X Col hybrid sepals and *pAR111*X *TCP* knock-down (*jaw-1D*) hybrid sepals. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend past 1.5 of the interquartile range. ***$p < 0.001$, significant difference by two tailed t-test, p=9.86 E-08; n=20.

(D) Confocal images of a sepal from a *pLH173* (all TGGG putative TCP sites mutated in giant cell enhancer fragment 2) T1 plant. Hardly any nuclei expressing 3XVenus can be detected.

(E) The laser intensity required to saturate the 3XVenus signal in giant cell for each reporter line. *pAR257* needs much higher laser intensity, indicating it has much lower expression level of 3XVenus. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend past 1.5 of the interquartile range. ns, non-significant difference (t test); ***$p < 0.001$, significant difference (t test). n = 17 for pAR111, n = 22 for pAR257, n = 24 for pAR254, and n = 24 for pAR261.

(F) Confocal images of a sepal from a hemizygous *pAR254* plant (*pAR254* X Col F1).

(G) Confocal images of a sepal from *pAR254* X *TCP* knock-down (*jaw-1D*) F1 hybrid plants.

Sepals were imaged at the same setting, with 3% laser power. Scale bars: 100 µm. See also Figures S3, S4, and S5.

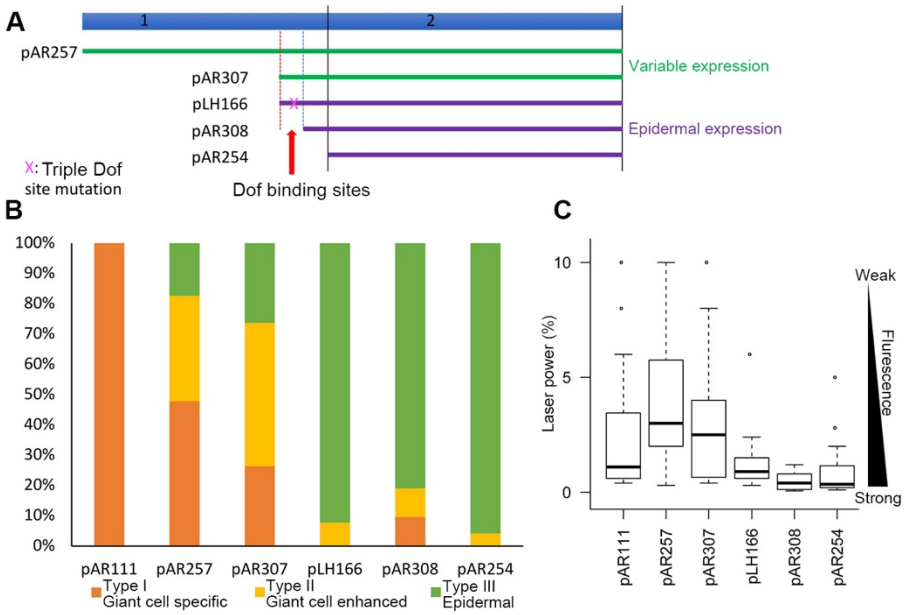


Figure 4. Dof binding sites in Region 1 confer giant cell enhanced activity.

(A) Fine-scale dissecting of Region 1 in the enhancer. Schematics of reporter lines containing different fragments or mutation of Region 1 were shown.

(B) Percentages of the three expression pattern types in different reporter lines. Data for *pAR111*, *pAR254* and *pAR257* were duplicated from Figure 2D for comparison.

(C) The laser intensity required to saturate the 3XVenus signal in giant cell for each reporter line. The higher laser intensity signifies weaker fluorescence of the reporter. The boxes extend from the lower to upper quartile values of the data, with a line at the median. The whiskers extend past 1.5 of the interquartile range. See also Figure S4, S6, and S7.

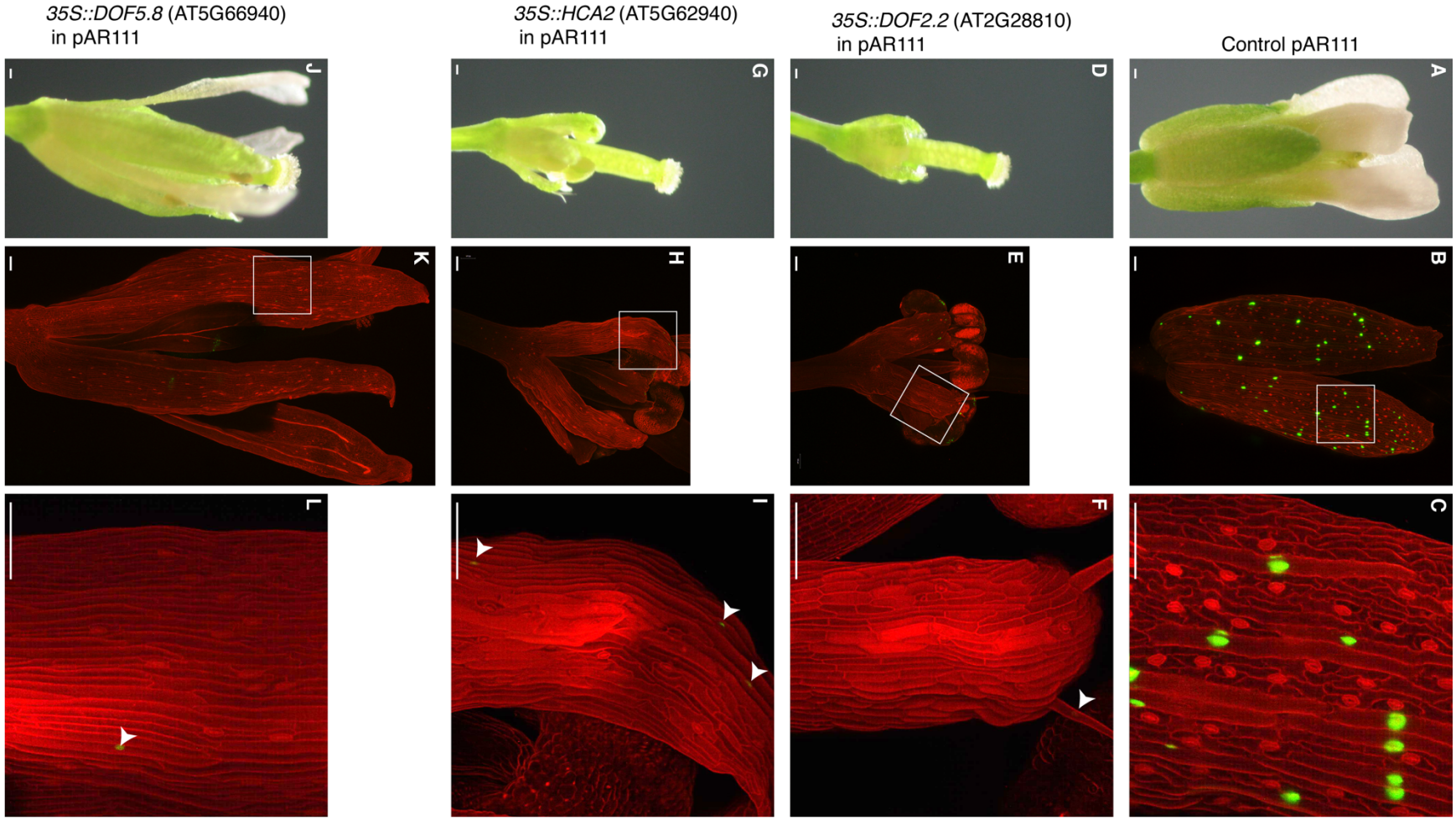

35S::DOF5.8 (AT5G66940)
in pAR111
35S::HCA2 (AT5G62940)
in pAR111
35S::DOF2.2 (AT2G28810)
in pAR111
Control pAR111
A
B
C
D
E
F
G
H
I
J
K
L

Figure 5. Dof transcription factor overexpression suppresses giant cell enhancer activity.
pAR111 is the full length giant cell enhancer which drives 3XVenus-N7 expression in giant cells (nuclear localized green signal). Cell walls were stained with PI (red). Arrowheads indicate nuclei expressing weak signal.

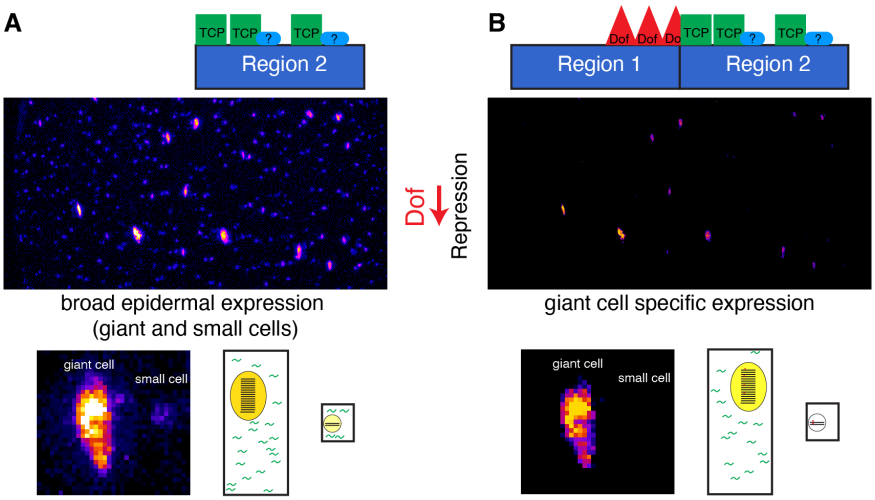
(A-C) Flowers from a control *pAR111* plant. (A) A dissecting microscope image. (B) A confocal image showing bright *pAR111* full length giant cell enhancer reporter signal in giant cell nuclei. (C) Magnification of the white box in B.

(D-F) Flowers from a *pAR111* plant expressing *p35S::DOF2.2* (*AT2G28810*). (D) A dissecting microscope image showing stunted sepals, petals, and stamens. (E) A 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 *pAR111* plant expressing *p35S::HCA2* (*AT5G62940*). (G) A dissecting microscope image showing stunted sepals, petals, and stamens. (H) A 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 *pAR111* plant expressing *p35S::DOF5.8* (*AT5G66940*). (J) A dissecting microscope image showing the sepals are narrow and highly elongated. (K) A 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. C, F, I, and L were rendered in MorphoGraphX adjusted in Photoshop with auto contrast. See also Figure S6.



C

- Dofs
- TCPs
- others
- giant cell specific expression
- intermediate expression
- epidermal expression

Diagrams	Constructs	Expression pattern	Expression level
	35S::Dof in pAR111	Largely repressed	low
	pAR111		high
	pAR262		low
	pAR257		low
	pAR307		low
	pLH166		high
	pAR254		high
	pAR254 TCP knock down		low
	pAR260		
	pAR261		high
	pLH173	Essentially no expression	very low
	pAR111 TCP knock down	Reduced numbers giant cells	low
	pAR258	no expression	no
	pAR280	no expression	no

Figure 6. Summary: Giant cell specificity arises from the repression of a broad epidermal expression pattern

In the giant cell enhancer, cell type specificity of expression arises primarily from the combination of two elements: Region 1, an element driving broad epidermal expression, with higher expression levels in giant cells than small cells, and Region 2, an element repressing expression, and making the expression pattern giant cell specific.

(A) Region 2 of the giant cell drives broad epidermal expression. Nevertheless this expression is stronger in giant cells (large, highly endoreduplicated nuclei) than small cells (smaller nuclei) as highlighted by the fire LUT (using FIJI) applied in A. Our evidence suggests that TCP transcription factors (represented as green rectangles) bind to and activate expression via Region 2 of the enhancer. Our yeast-one hybrid results suggest that other factors (represented generally as blue oblongs) including HD-ZIP Class IV transcription factors bind to Region 2. We speculate these may be important in either activating epidermal expression or activating stronger expression in giant cells than small cells. One factor that may contribute to the increased expression in giant cells is the increased ploidy of these cells, providing many more copies of the enhancer (as represented in the giant cell with 16C DNA content) and lots of transcripts (represented in green) to drive strong enhancer expression (strong yellow in the nucleus) in comparison to the small diploid cell with few transcripts and low expression in the nucleus.

(B) When combined with Region 1, Region 2 becomes much more giant cell specific as detailed in the results. Our evidence suggests that Dof transcription factors (red ovals) bind to this element and repress expression.

(C) Table summarizing the results of each of our dissection constructs together with site specific mutations and genetic manipulation of TCPs and Dofs.

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

Figure S1. The whole sequence of the giant cell enhancer, related to Figure 1.
Green: Region 1, 1-206 bp; magenta: Region 2, 207-450 bp; blue: Region 3, 451-760 bp; purple: Region 4, 761-1024 bp. The underlined region is the overlapping region used in the yeast one hybrid assay.

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

Gene family	Region 1	Overlap	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

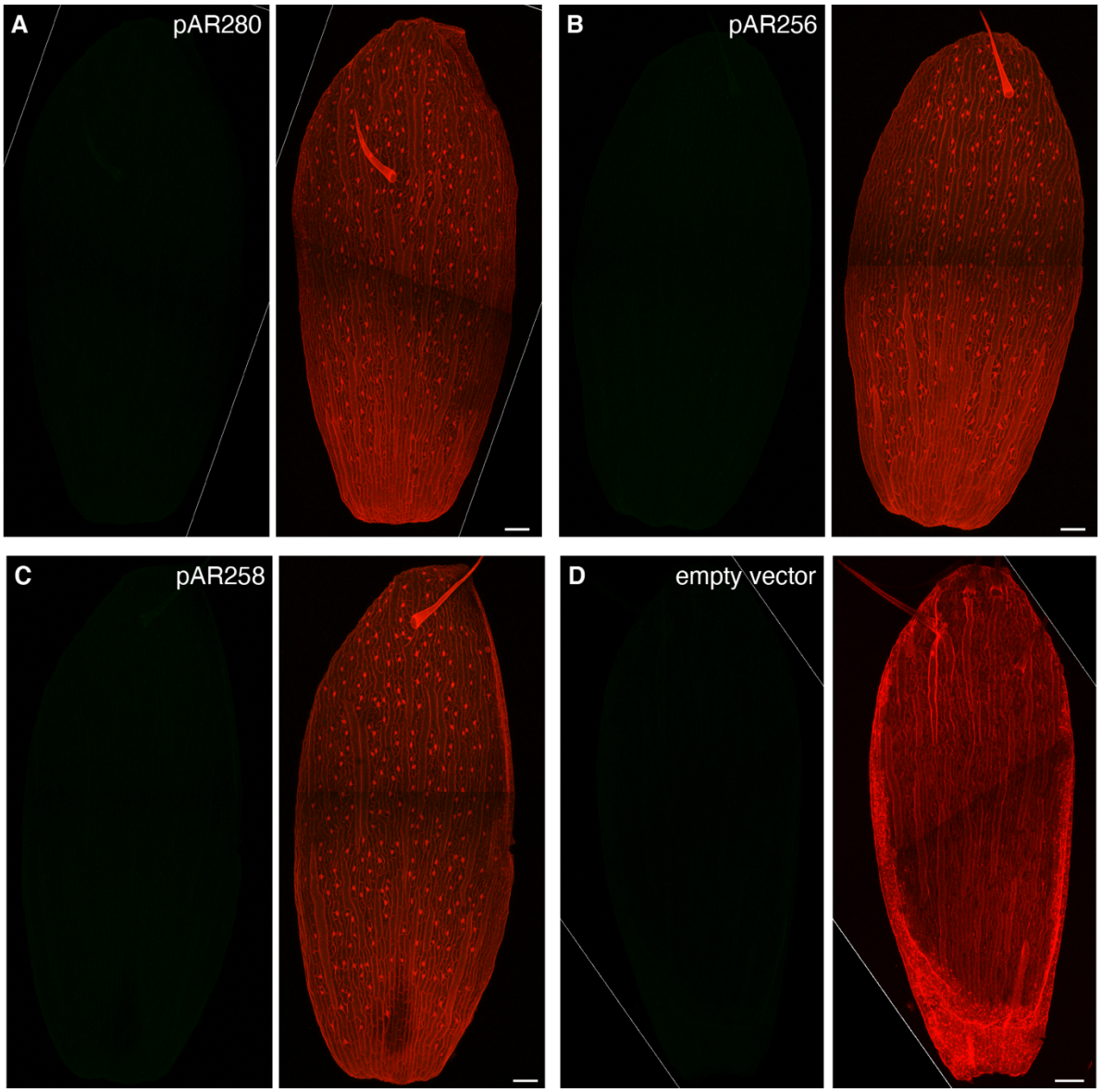


Figure S2. Region 1 alone, Region 4 alone and Regions 3 and 4 together are not sufficient to drive any reporter expression, related to Figure 1.

Confocal images of stage 14 sepals from different reporter lines. Images on the left show 3XVenus (green) signal marking the nuclei of cells expressing the reporter. Images on the right show 3XVenus (green) signals merged with PI (red) stained cell walls. Images were taken with 3% laser power except where noted. Increasing laser power to 40% did not reveal any expression.

(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 represent 100 µm.

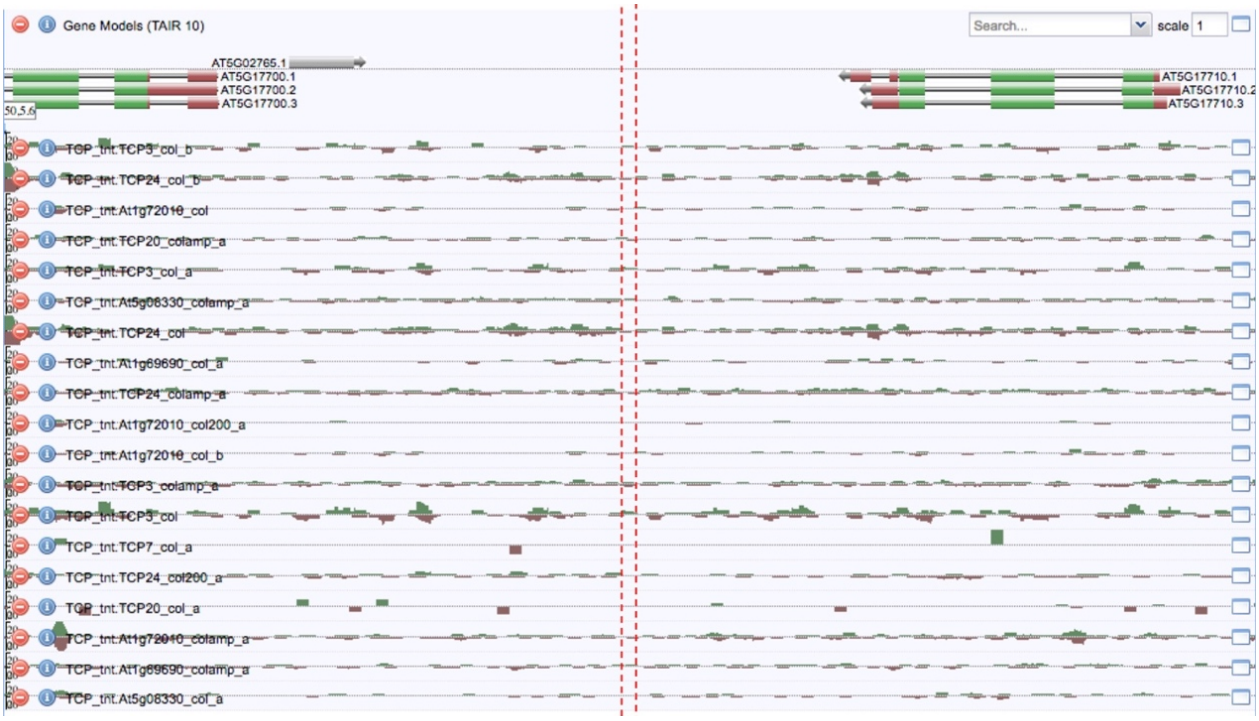


Fig. S3. TCP transcription factors bind to the giant cell enhancer and nearby region, related to Figure 3.

Binding sites of TCP transcription factors in the giant cell enhancer and nearby region. The red dash lines flank the three TCP binding site region in Region 2. Data are from DNA affinity purification sequencing (DAP-seq) analysis (http://neomorph.salk.edu/dev/pages/shhuang/dap_web/pages/aj_browsers.php). We observe moderate binding throughout the intergenic region including the region of the enhancer. We speculate that this moderate binding might reflect spatial restriction of binding either to sepals or epidermal cell layers, that would not be accurately captured by DAP-seq.

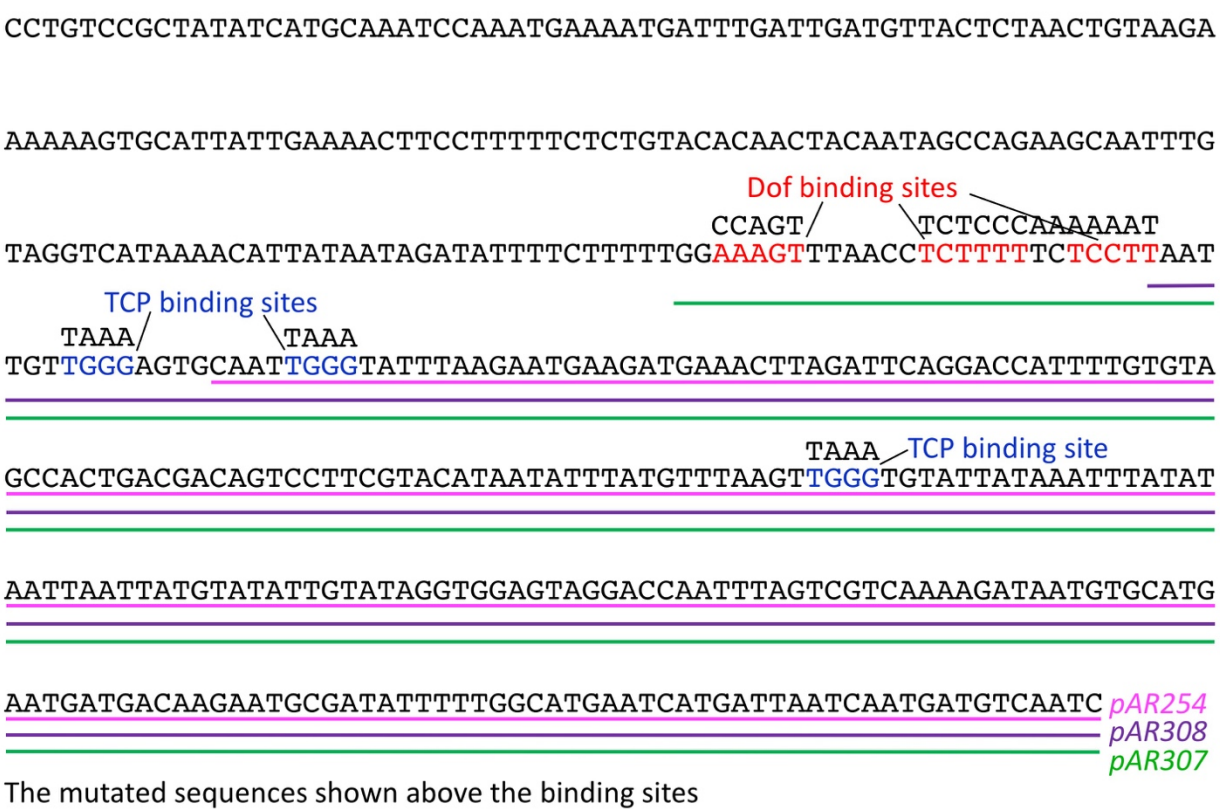


Fig. S4. Sequences and locations of the putative Dof binding sites, putative TCP binding sites and different enhancer fragments, related to Figures 3 and 4.

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 sites are highlighted with red letters. The putative TCP binding sites are highlighted in blue. The mutated sequences for the putative Dof and TCP binding sites are shown above the original sequences.

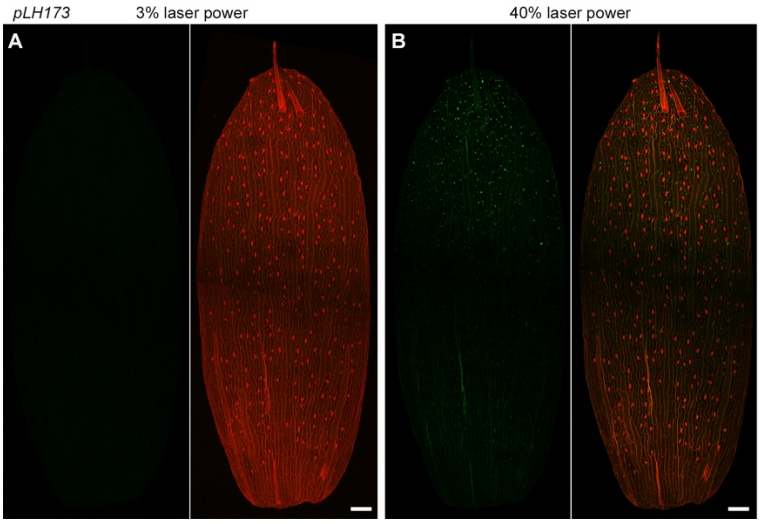


Figure S5. *pLH173* sepals expressed weak 3XVenus in some nuclei, related to Figure 3

Confocal images of a *pLH173* sepal imaged with 3% (A) and 40% (B) laser power. With 3% laser power, hardly any 3XVenus could be observed on *pLH173* sepal epidermis. With laser power increased to 40%, some nuclei at the sepal tip expressed weak 3XVenus. Images on the left, 3XVenus (green) marking the nuclei of cells expressing the enhancer reporter; Images on the right, 3XVenus (green) signals merged with PI (red) stained cell wall signals. A was duplicated from Fig.3D for comparison.

Figure S6 Binding sites of Dof transcription factors in the giant cell enhancer region and nearby, related to Figure 6.

Results by DAP-seq. Dof transcription factors framed in red showed up in the yeast one-hybrid (Y1H) assay in this study; framed in blue were overexpressed in *pAR111* Arabidopsis. The red dash line marks the Dof binding site region.

```
A_thaliana        1 ---------------------------------------------CCTGTCCG-CTATATCA
A_lyrata          1 ---------------------------------------------CCTGTCCG-CTATATCA
C_hirsuta         1 ---------------------------------------------CCTGTCCACTGTTATCA
B_rapa_ch_A10     1 ---------------------------------------------GATGTCCGCTATTATCA
S_irio            1 ---------------------------------------------ACAGACTATTGGTAGTC
E_salsugineum     1 ---------------------------------------------CGATCCCGCTCTTATCA
S_parvula         1 GATTTCATCTAAAAACAGATGAAAGATAACAAAAAAAGATATATATATATATTTTTT
B_rapa_ch_A02     1 -----------AATCCAGGATGCCCGCTGTCGAAACCG-----------------------

A_thaliana       17 TGCAAATCCAA-------------ATG---AAAATGATTTGATTG-----------------
A_lyrata         17 TTAAAATCCAA-------------ATG---AAAATGATTTGATTGATTGATTTTTTTTTT
C_hirsuta        18 TTGAATTCCAA-------------AAT---AA-------------------------------
B_rapa_ch_A10    18 TTGAAACCAAA-------------ATAATAAAAATGATTTGATTGATGGTATTCTAACTTC
S_irio           18 TGAAAATAGAAATAT--AGTAAAATAATAGAAACATCGAGGATGTCCGCTATT-----TA
E_salsugineum    18 CTGAAAGGGT--------------GAAATAAAAATGATTTGA-------------------
S_parvula        61 TTAAAAAATTAATATACAACAAATTGTTGATAGTTTCAAAATAGAAAGACGTGGA--TAT
B_rapa_ch_A02    28 ----------------------AGTAATAAAAATGATTTGATTGATGGTATTCTAAAAAG

A_thaliana       46 ---------------------------------------------ATGTTACTCTAACTG
A_lyrata         62 TTTTTTTTTTGATTG-----------------------------ATGTTACTCTAACTG
C_hirsuta        34 ---AATGATTTGATTG-----------------------------ACGGTACTCTAACTG
B_rapa_ch_A10    66 CAGGAAAAAGTGCATTAT------------------------TGAAAGTTTCCTTTAGTT
S_irio           71 TAAAAAAAAATGATTTGA------------------------TTGATGGTACTCTAACTG
E_salsugineum    46 -----------TTGATGAT------------------------TTGAAGGTATTCTAACTG
S_parvula       119 CAAAATAATAATAATA-A------------------------TAGAAAATCTTCTAGT--
B_rapa_ch_A02    66 TATATTAGCCAAAATCAATTTTAGGTCATGAACATAATCATATATAGGGTATTTTCTCTC

A_thaliana       61 TA--AG--AAAAAAGTGCATTATTGAA---A----ACTTCCTTTTTC-------------
A_lyrata         93 TATGAG--AAAAAAGTGCATTATTGAC---A----GGTTCCTTTTTTT------------
C_hirsuta        62 CA--AG--AAAAAAGTGCATTATTTAAAGGG----TTCTCCTTTTTCT------------
B_rapa_ch_A10   102 TTATAC--TGGAAAGTATATTAGCCACACAACTCACAAGCAATATGTT------------
S_irio          107 CAAGTAAGGAAAAAGTGCATTATTGCAAG------CTTCCTTTTATTA------------
E_salsugineum    72 CA------AGGAAAGTGCATTATTAAAAGT-----TTCCTTTTTCTTA------------
S_parvula       152 --------CGGCAAGCGAACTAATCCAGGATGTCCGCTATTTTATCACTGAAACCGAAAT
B_rapa_ch_A02   126 TAA-----TTTTCAATTTATGGTTCCAAAAATTTAGATCCCTATGCATC-----------

A_thaliana       97 -------TCTGTACACAACTACAATAGCCAGAAGCAATT---TGTAGGTCATAAAACA-T
A_lyrata        132 -------GCACACAAAAAGTATATTAGCCAGAAGCAATT---TGTAGGTCACAAAACA-T
C_hirsuta       102 -------TGTAGAAGAAAGTATATTAGCCAGAAGCAATT---TGTAGGTCATAGAACA-T
B_rapa_ch_A10   148 -------GGTCATGAACA--TTATCATAGAGAAAGCCTT---GTTACAAAAAAAGCA-T
S_irio          149 -------TATACTGGAAAATATAATAGCCAGAAACAATA---TGTTGGTTATGAACAT--
E_salsugineum   109 -------TATACTGGAAAGTATATTAGCCAGAAGCAATT---TGTAGGTCATGAACAC-T
S_parvula       204 AATAAAAATGATTTGATTGATGGTATACCAAAAGCAATT---CCTGGGTCATGAACAT--
B_rapa_ch_A02   170 -------CCCACCGCAATTAATACTAACCAAAATATCTACTGTGTATATATATATATATG

                                                   pAR307  Dof              Dof   Dof
A_thaliana      146 TATAATAGATATTTTCTTTT-----------TG-----GAAAGTTTAACCTCTTTTTCTC
A_lyrata        181 TATCATAGTCATTTTTCTGA---------AATT-----TTAATTTATGTCTCTTTTTCTC
C_hirsuta       151 TATCATAGTTTTTTTTTGGTTCT----GAAATT-----AAAATTTATGTCTCTTTTTCT
B_rapa_ch_A10   195 CATCAAAGATTTTTTTTCTC--T--------GA-----AAATTAATATGTTTCTTTTTCT
S_irio          197 TATCATAGAGCTTTCTTTTT-TT----CTCTGA-----AAGTTTAATTATTTTTTTCT
E_salsugineum   158 ATCATAGAGGTTTTTATTTTTCTTTTTGTTCTGAAAT-TTTGGTTTAAATTCTTCTCCTT
S_parvula       259 GATCATAGAGTTTTTTTTTTCTT---TGCTCTGAAAT-TTTAATTTATGCTTCTTTTTTTT
B_rapa_ch_A02   223 TATCTACATAATTTTATTAC-TATTTAGACCGCACCCATGGTCTCCTATGTTTTACTTT

                    pAR308   TCP      pAR254TCP
A_thaliana      190 CTTAATTGTTGGGAGTGCAATTGGGTA----TTTAAGAATGAAGATGAAACTTAGATTCA
A_lyrata        227 CTTAATTGCTGGGAGTGCAATTGGGTA----TTTAAGAATGAAGATGAAGCTTTAGG---
C_hirsuta       202 CCTTAATTGTTGTGGTGCAATTGGGTA----TTTAAGAATGAATATGAAACGTAGGTTAT
B_rapa_ch_A10   240 GCTTAACTGTTGTGGTGCAATTGGGTA----TTTATGAATGAATATGAAACTTAGGTTCA
S_irio          247 TTTTAAATGTTCTGGTGCAATTGAGTA----TTTAAGAATGAATATGAACGTAGGTTC-A
E_salsugineum   217 AATTAATTATTCTGGTCCAGTTGGGTATTTAATTAAGAACGAATATGAAACGTAGGTTCA
S_parvula       315 GTAGCTCTTAATTATTGTGGTGCAATAGGTATTTAAGAATGA---ATATGAATAGGTTCA
B_rapa_ch_A02   282 TTTTCTCTTATATGAAGCAATTGGGTA----TTAAAGTATGAATATGAAATGTA-TGTCA
```

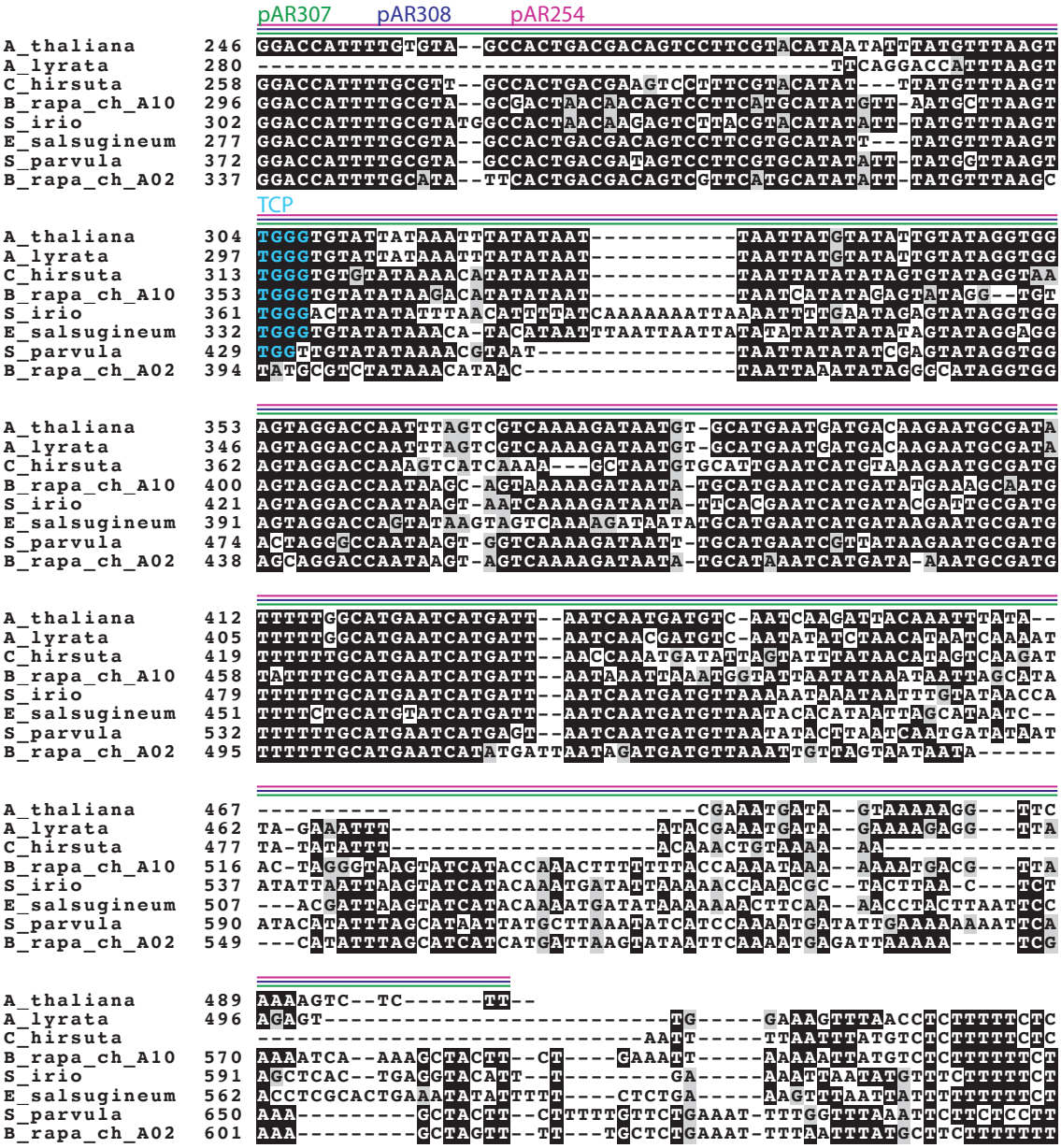


Figure S7 Conservation of Regions 1 and 2 of the giant cell enhancer across Brassicaceae species, related to Figure 4.

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*. Although *Brassica rapa* is a diploid, it underwent genome duplication during evolution, so there are two regions on chromosomes A10 and A02 which 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 sites predicted in *Arabidopsis thaliana* and their conserved counterparts are highlighted with red letters. The putative TCP binding sites and their conserved counterparts are highlighted with aqua letters.

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
pLH173	Region 2+ with 3 putative TCP sites mutated (oAR630-oAR631 used to mutate third TCP site) 190 – 449 bp	oAR497	oAR612	pDR9
pAR307	Region2 with Dof region 167-449 bp	oAR497	oAR556	pAR299
pLH166	Region 2 with 3 putative Dof sites mutated 167-449 bp	oAR497	oLH412	pLH164
pAR308	Region 2+ (without Dof region) 190 – 449 bp	oAR497	oAR609	pAR301

Associated with Methods

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	ggaaagtttaacctcttttctcc	Reverse primer for overlap region
oAR557	CACCagtggctacacaaaatggtcc	Forward primer for overlap region
oAR609	cttaattgttgggagtgcaattgggtatttaagaatgaagatgaaac	Reverse primer for Region 2+ without Dof
oAR612	cttaattgttAAAagtgcaattAAAtatttaagaatgaagatgaaac	Primer for mutating two putative TCP sites (compare with oAR609)
oAR630	cgtacataatatttatgtttaagttAAAtgtattataaatt	Primer for mutating third putative TCP site
oAR631	aatttataatacaTTTaacttaaacataaatattatgtacg	Primer for mutating third putative TCP site
oLH412	ggCCagtttaacctctCCCAAAAAAtaattgttgggagtg	Primer for mutating 3 putative Dof sites
oXQ6	CGGTACCatcaaccactttgtacaagaaagct	Reverse primer for Gateway sequence with *Kpn*I

Capital letters in primers indicate bases that do not match the template sequence, whereas lower case letters indicate bases that do match the template sequence. Associated with Methods

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	Overlap of junction 207-499 bp	oAR557	oAR556	pAR279

Associated with Methods