

ARTICLE

Received 18 May 2015 | Accepted 2 Mar 2016 | Published 21 Apr 2016

DOI: 10.1038/ncomms11222

OPEN

# A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor

Camille Sayou<sup>1,†</sup>, Max H. Nanao<sup>2</sup>, Marc Jamin<sup>3</sup>, David Posé<sup>4,5</sup>, Emmanuel Thévenon<sup>1</sup>, Laura Grégoire<sup>1</sup>, Gabrielle Tichtinsky<sup>1</sup>, Grégoire Denay<sup>1</sup>, Felix Ott<sup>4,†</sup>, Marta Peirats Llobet<sup>1,†</sup>, Markus Schmid<sup>4,6</sup>, Renaud Dumas<sup>1</sup> & François Parcy<sup>1,7</sup>

Deciphering the mechanisms directing transcription factors (TFs) to specific genome regions is essential to understand and predict transcriptional regulation. TFs recognize short DNA motifs primarily through their DNA-binding domain. Some TFs also possess an oligomerization domain suspected to potentiate DNA binding but for which the genome-wide influence remains poorly understood. Here we focus on the LEAFY transcription factor, a master regulator of flower development in angiosperms. We have determined the crystal structure of its conserved amino-terminal domain, revealing an unanticipated Sterile Alpha Motif oligomerization domain. We show that this domain is essential to LEAFY floral function. Moreover, combined biochemical and genome-wide assays suggest that oligomerization is required for LEAFY to access regions with low-affinity binding sites or closed chromatin. This finding shows that domains that do not directly contact DNA can nevertheless have a profound impact on the DNA binding landscape of a TF.

<sup>1</sup>Laboratoire de Physiologie Cellulaire et Végétale, Université Grenoble Alpes, CNRS UMR5168, CEA/DRF/BIG, INRA UMR 1417, 17, avenue des Martyrs, 38054 Grenoble, France. <sup>2</sup>European Molecular Biology Laboratory, 71, avenue des Martyrs, 38042 Grenoble, France. <sup>3</sup>Institut de Biologie Structurale CEA/DRF, CNRS, Université Grenoble Alpes, 71, avenue des Martyrs, 38044 Grenoble, France. <sup>4</sup>Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany. <sup>5</sup>Instituto de Hortofruticultura Subtropical y Mediterránea, Universidad de Málaga-Consejo Superior de Investigaciones Científicas, Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain. <sup>6</sup>Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, SE-901 87 Umeå, Sweden. <sup>7</sup>Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia, Canada V5Z 4H4. † Present addresses: Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Max Born Crescent, Edinburgh EH9 3BF, UK (C.S.); CeGaT GmbH, 72076 Tübingen, Germany (F.O.); Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia, ES-46022 Valencia, Spain (M.P.L.). Correspondence and requests for materials should be addressed to M.N. (email: mnanao@embl.fr) or to R.D. (email: renaud.dumas@cea.fr) or to F.P. (email: francois.parcy@cea.fr).

**T**ranscription factors (TFs) play key roles during development and differentiation<sup>1</sup>. By binding to short stretches of DNA called *cis*-elements or TF binding sites (TFBS) and thereby regulating gene expression, they are able to decode the regulatory information present in genomes<sup>2</sup>. Based on structural studies and genome-wide analyses, the rules governing TF DNA interactions are becoming clearer and DNA-binding models with increased prediction power are available<sup>3,4</sup>. TF DNA-binding specificity is primarily determined by their DNA-binding domains (DBDs) through contact with bases and DNA shape readout<sup>4</sup>. DBD are often dimeric or sometimes tetrameric, thereby increasing the TFBS length and the TF specificity<sup>3</sup>. In addition to their DBD, some TF also possess an oligomerization domain<sup>5</sup>. Such a feature is found in several types of factors such as members of the E26 transformation-specific or E-twenty-six (ETS) family<sup>6</sup>, auxin response factors<sup>7</sup>, some zinc-finger TFs<sup>8</sup> or the PRH/Hex<sup>9</sup> TF. How such domains modify the repertoire of TF target regions at the genome-wide scale is largely unknown.

Here we focus on LEAFY (LFY), a plant-specific TF essential for flower development<sup>10</sup>. LFY is one of the few master regulators of flower development, as it integrates environmental and endogenous signals to orchestrate the whole floral network<sup>11–14</sup>. It first triggers the early emergence of a bulge of stem cells on the flanks of the shoot apex<sup>15,16</sup> and subsequently specifies their floral identity by directly inducing floral homeotic genes such as *APETALA1* (*AP1*), *APETALA3* (*AP3*) or *AGAMOUS* (*AG*) and by repressing the shoot meristem identity gene *TERMINAL FLOWER1* (*TFL1*)<sup>10,11</sup>. Comprehensive lists of genes bound and possibly regulated by LFY have been established by combining genome-wide chromatin immunoprecipitation (ChIP) experiments and transcriptomic approaches<sup>17,18</sup>.

The function of LFY as a master regulator of floral development is most obvious when expressed from a constitutive promoter: it is sufficient to trigger early flowering and ectopic flower production<sup>19</sup>. Together with co-regulators such as the homeodomain TF *WUSCHEL* or the F-Box protein *UNUSUAL FLORAL ORGANS*, it can even induce flowers from root or leaf tissue, respectively<sup>20,21</sup>. This acquisition of floral fate is remarkable, as several LFY targets including *AG* and *AP3* are known to be under the repression of two Polycomb repressive complexes (PRC1 and PRC2) (refs 22,23), which ectopic LFY expression is apparently sufficient to override.

Structure–function studies have demonstrated that LFY binds semi-palindromic 19-bp DNA elements through its highly conserved C-terminal DBD, a unique helix–turn–helix fold that by itself dimerizes on DNA (Fig. 1a)<sup>17,18,24,25</sup>. A biophysical model describing LFY DNA-binding specificity *in vitro* was built and accurately predicts LFY-binding sites (LFYbs) in the *Arabidopsis thaliana* genome<sup>17,26</sup>. In addition to its well-characterized DBD, LFY possesses a second conserved domain at its amino terminus (LFY-N) (Fig. 1a). This domain has been proposed to mediate LFY dimerization by forming a leucine zipper<sup>27</sup>. Here we report the crystallographic structure of LFY-N and elucidate its function using a combination of biochemical and genomic approaches. Our results demonstrate that LFY-N is a Sterile Alpha Motif (SAM) domain that mediates LFY oligomerization. It allows LFY to bind to regions lacking high-affinity LFYbs and confers on LFY the ability to access closed chromatin regions.

## Results

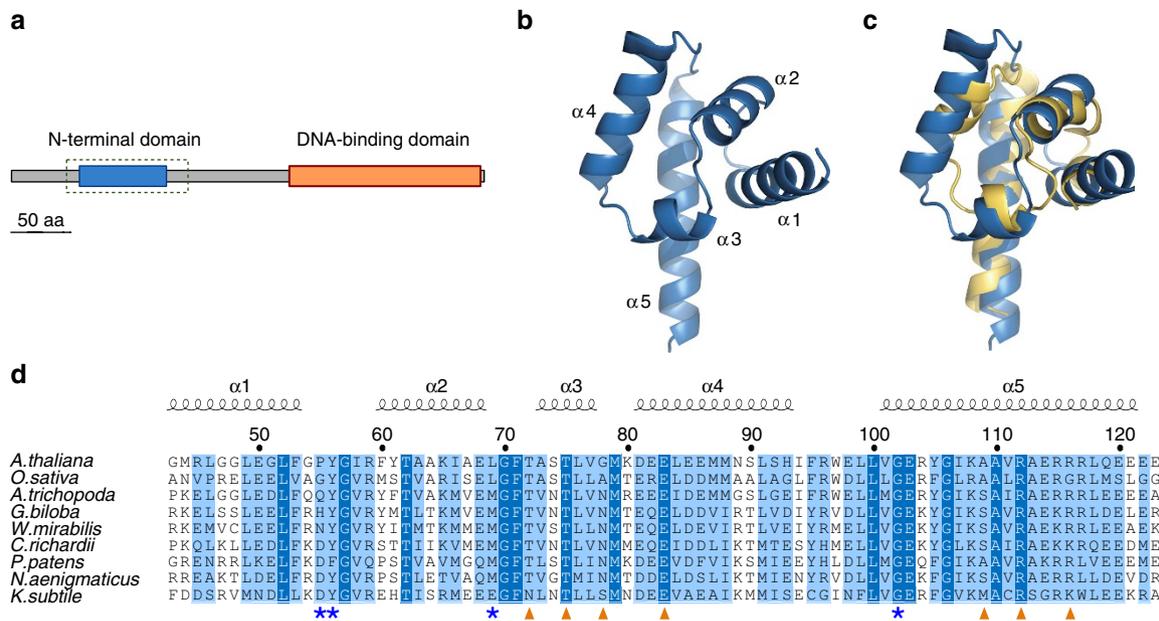
**LFY N-terminal domain is a SAM.** We used a structural approach to unravel the function of the N-terminal domain of LFY. After heterologously overexpressing LFY-N domains from different plant species, we were able to determine the crystal

structure for GbLFY-N from the gymnosperm *Ginkgo biloba* at 2.3 Å resolution (Table 1 and Supplementary Fig. 1). This structure showed that the GbLFY-N monomer is made up of five  $\alpha$ -helices separated by four loops (Fig. 1b). Comparison against the Protein Data Bank revealed a strong structural similarity with the SAM, also called Pointed domain<sup>28</sup> (Fig. 1c), a resemblance that could not be predicted from LFY-N sequences. SAM domains have not been characterized in plants, but they are common protein motifs in other eukaryotes. They interact with proteins, DNA, RNA and lipids, and are involved in many cellular functions including transcriptional regulation and signal transduction<sup>6,28</sup>. Some SAM domains are able to self-associate as oligomers through the so-called mid-loop (ML) and end-helix (EH) surfaces<sup>29,30</sup>. GbLFY-SAM appears to belong to this class: in the crystal, it forms a head-to-tail polymer (Fig. 2a) with monomers contacting their neighbours through two polar surfaces (Fig. 2b). The ML surface comprises negative and polar residues (Fig. 2c), including T72, T75 and E83, which interact with the positively charged EH surface through hydrogen bonds and salt bridges, in particular via R112 and R116 (residue numbering refers to LFY from *A. thaliana*—AtLFY—sequence; Fig. 1d).

**LFY-SAM is an oligomerization domain.** To validate the interactions observed in the crystal in solution, we analysed the oligomerization state of the wild-type (WT) GbLFY-SAM protein and of mutants at the ML surface (T75E substitution, GbLFY-SAM<sub>TE</sub>), EH surface (R112E substitution, GbLFY-SAM<sub>RE</sub>) or at both surfaces (GbLFY-SAM<sub>TERE</sub>) (Fig. 2d). Size-exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS) showed that the WT GbLFY-SAM domain formed oligomers of variable size in solution, depending on protein concentration, containing up to eight monomers (Fig. 2d, Supplementary Fig. 2 and Supplementary Table 1). In comparison, all mutations affecting the interaction interface completely abolished LFY oligomerization and GbLFY-SAM<sub>TE</sub>, GbLFY-SAM<sub>RE</sub> and GbLFY-SAM<sub>TERE</sub> proteins were found to be exclusively monomeric (Fig. 2d and Supplementary Table 1). However, when mixed at an equimolar ratio, the GbLFY-SAM<sub>TE</sub> and GbLFY-SAM<sub>RE</sub> single-face mutants retained the ability to interact, forming a GbLFY-SAM<sub>[RE+TE]</sub> dimer (Fig. 2d and Supplementary Table 1), thus showing that the different mutations did not alter the overall fold of the SAM domain, and that single face mutant proteins still displayed one functional interaction surface. These results further showed that the head-to-tail arrangement observed in the crystal structure reflects the state of LFY-SAM in solution, which forms an oligomer of limited size in the conditions of this study. The point mutations we designed allowed us to control the oligomerization state, creating either monomeric or dimeric variants of LFY-SAM instead of higher-order oligomers.

Comparing LFY sequences from multiple species throughout the plant kingdom revealed that the key interaction residues of the SAM domain are well conserved (Fig. 1d and Supplementary Fig. 2d), suggesting that the capacity to oligomerize is a general characteristic of LFY-SAM in all species from algae to angiosperms. This hypothesis is supported by the behaviour of the SAM domain of *A. thaliana* LFY (AtLFY-SAM) whose molecular mass in solution increased with its concentration, similar to what was observed for GbLFY-SAM (Supplementary Fig. 2).

**LFY-SAM<sub>TERE</sub> mutants have highly reduced function in planta.** Several mutations that compromise LFY function in rice and *A. thaliana* map to the SAM domain<sup>27,31</sup>. However, according to



**Figure 1 | LFY N-terminus is a SAM domain.** (a) Domain representation of GbLFY protein with the N-terminal domain in blue and DBD in orange. Dashes indicate the region used to obtain the crystal structure. Scale bar, 50 amino acids (aa). (b) Crystal structure of GbLFY-N monomer.  $\alpha$ ,  $\alpha$ -helices. (c) Superimposition of GbLFY-N (blue) and the SAM domain of p63 $\alpha$  (yellow, PDB 2Y9U)<sup>66</sup>. The two structures are very similar and can be superimposed with a root mean square deviation of 1.9 Å for 62 C $\alpha$ . (d) Sequence alignment of LFY-N domain. The residue numbering refers to *A. thaliana* LFY sequence; the secondary structure from GbLFY-SAM crystal structure is indicated above the alignment; orange triangles and blue stars indicate residues involved in the interaction between SAM monomers through the lateral and main chain, respectively.

**Table 1 | Crystallographic data collection and refinement statistics.**

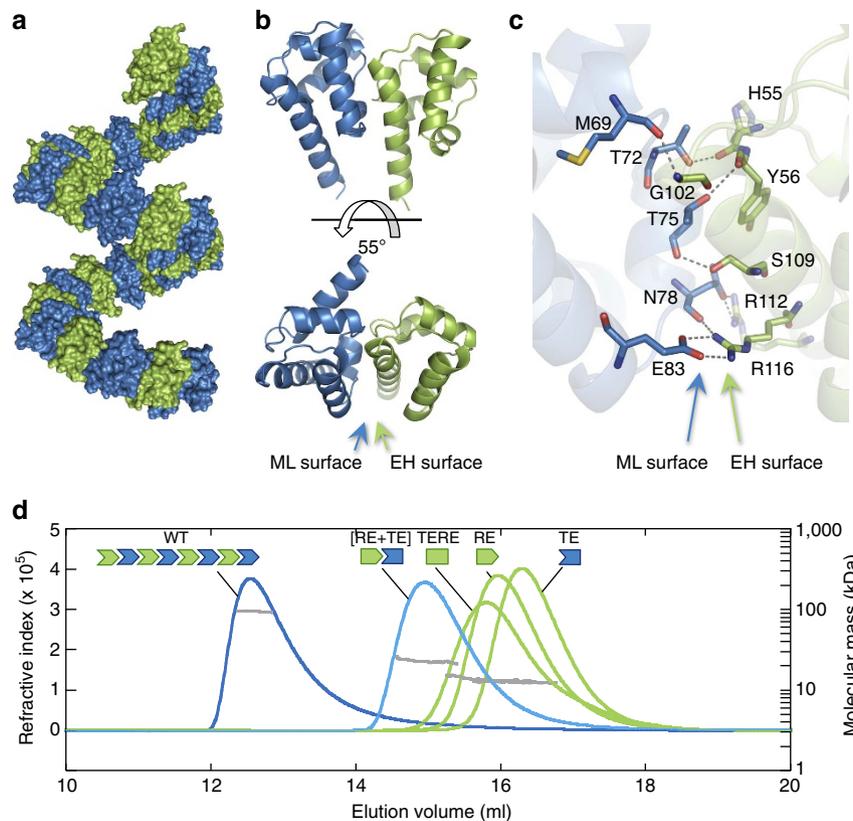
	GbLFY-N (Native)	GbLFY-N (Se-Met)
<i>Data collection</i>		
Wavelength (Å)	0.872	0.872
Resolution (Å) range	52.34–2.25 (2.33–2.25)	100–2.61 (2.77–2.61)
Space group	<i>P</i> 6 <sub>5</sub>	<i>P</i> 6 <sub>5</sub>
Cell dimensions (Å, °)	<i>a</i> = <i>b</i> = 81.1, <i>c</i> = 78.5	<i>a</i> = <i>b</i> = 80.6, <i>c</i> = 78.5
Number of total reflections	57,705	83,685
Number of unique reflections	14,004	8,879
Multiplicity	4.1(3.9)	9.4 (9.2)
Completeness (%)	99.3 (99.0)	99.6 (97.5)
Mean <i>I</i> / $\sigma$ ( <i>I</i> )	15.67 (2.89)	15.28 (2.54)
<i>R</i> -sym	0.069 (0.514)	0.127 (0.915)
<i>Refinement statistics</i>		
Wilson B-factor	44.1	35.85
<i>R</i> -factor	0.1859 (0.2564)	
<i>R</i> -free	0.2298 (0.2767)	
Number of atoms	1,462	
Macromolecule	1,335	
Ligands	25	
Water	102	
Protein residues	163	
Root mean square (bonds)	0.016	
Root mean square (angles)	1.78	
Ramachandran favoured (%)	99	
Ramachandran outliers (%)	0	
Clashscore	13.40	
Average B-factor	46.80	
Macromolecule	46.00	
Solvent	51.80	

Statistics for the highest-resolution shell are shown in parentheses.

our structural data, these positions correspond to buried hydrophobic residues that are unlikely to directly affect oligomerization but are more likely to be important for proper domain folding. Only the very weak *lfy-22* allele in *A. thaliana*<sup>32</sup> carries a mutation (G70D) that localizes to the SAM ML surface and possibly slightly weakens oligomerization. Therefore, to specifically assess the importance of oligomerization for LFY function, we tested whether the TERE mutation affects AtLFY ability to control flower development in the genetically amenable *A. thaliana* plant.

We first tested whether AtLFY<sub>TERE</sub> could complement the *lfy-12*-null mutant, in which flowers are replaced by shoots or sterile shoot/flower intermediate structures lacking petals and stamens (Fig. 3a)<sup>33</sup>. When expressed under the control of the LFY endogenous promoter (*pLFY*), AtLFY<sub>TERE</sub> complemented the *lfy-12* mutant much less efficiently than AtLFY (Fig. 3a,b and Supplementary Table 2, Mann–Whitney rank test, *P* = 0.01006). Whereas most of *pLFY:LFY lfy-12* primary transformants produced fertile flowers, the majority of *pLFY:LFY<sub>TERE</sub> lfy-12* primary transformants produced only a few fertile flowers at the base of the inflorescence. These results indicated that LFY activity was reduced, although not entirely abolished, when its oligomerization was compromised.

To further assess the importance of the SAM domain for LFY master regulator function, we expressed AtLFY from the *p35S* constitutive promoter and monitored its effect at an early developmental stage, when other pathways contributing to flower development have only a minimal impact. When expressed constitutively, AtLFY is known to induce early flowering, and to trigger the precocious termination of the shoots and the formation of ectopic flowers from the axils of rosette leaves<sup>19</sup>. About half of the 61 *p35S:LFY* T1 plants we generated showed this characteristic phenotype. In contrast, ectopic flowers were never observed in 51 *p35S:LFY<sub>TERE</sub>* T1 transformants (Fig. 3c,d, Fisher’s exact test, *P*-value =  $1.5 \times 10^{-10}$ ) despite the fact that



**Figure 2 | LFY-SAM is an oligomerization domain.** (a) Helical polymer of GbLFY-SAM in the crystal structure. Twenty-eight monomers are shown, alternatively coloured in blue and green. (b) Side and top views (upper and bottom panels, respectively) of the GbLFY-SAM dimer in the asymmetric unit of the crystal. ML and EH interaction surfaces are indicated. (c) Detailed view of GbLFY-SAM oligomerization interface. Residues mediating contacts between the two monomers are shown as sticks; residues on the ML and EH surfaces are in blue and green, respectively. (d) SEC-MALLS analysis of GbLFY-SAM (WT), GbLFY-SAM<sub>TE</sub> (TE, T75E substitution at the ML surface) and GbLFY-SAM<sub>RE</sub> (RE, R112E substitution at the EH surface), GbLFY-SAM<sub>TERE</sub> (TERE, T75E, R112E) double mutant and GbLFY-SAM<sub>[RE+TE]</sub> ([RE+TE], an equimolar mixture of GbLFY-SAM<sub>TE</sub> and GbLFY-SAM<sub>RE</sub>). 50  $\mu$ l at  $\sim$ 4.5 mg ml<sup>-1</sup> of proteins were used. Elution profiles were monitored by excess refractive index (left ordinate axis). The grey line under each elution peak shows the molecular mass distribution (right ordinate axis). Measured molecular masses (reported in Supplementary Table 1) show that GbLFY-SAM is oligomeric; GbLFY-SAM<sub>TE</sub>, GbLFY-SAM<sub>RE</sub> and GbLFY-SAM<sub>TERE</sub> are monomeric; and GbLFY-SAM<sub>[RE+TE]</sub> is dimeric.

AtLFY and AtLFY<sub>TERE</sub> proteins were expressed at similar levels (Supplementary Fig. 3). Taken together, these results provide evidence that LFY-SAM domain oligomerization is required for the function of LFY as a master switch in flower initiation.

#### SAM<sub>TERE</sub> mutations do not alter single LFYbs binding *in vitro*.

Next, we aimed to pinpoint how a functional SAM domain contributes to LFY function at the molecular level. As a TF, LFY specifically binds to *cis*-elements present in the regulatory regions of its target genes. Hence, we tested whether LFY mutants with altered oligomerization status differed in DNA-binding ability. For other SAM proteins, such analyses were hampered by the difficulty in producing recombinant soluble proteins due to the aggregation and precipitation triggered by the SAM domain<sup>34</sup>. By optimizing the purification protocol, we successfully produced a near full-length GbLFY protein (GbLFY $\Delta$ ) lacking only the 53 non-conserved N-terminal amino acids preceding the SAM domain.

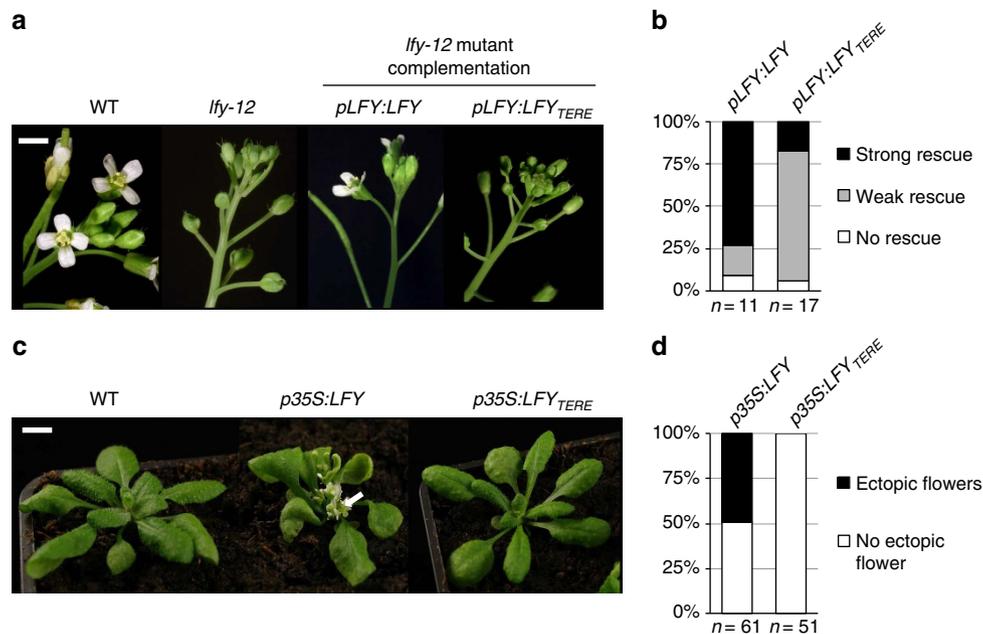
As GbLFY and AtLFY have comparable DNA-binding specificities<sup>24</sup>, we tested GbLFY $\Delta$  binding to a DNA fragment carrying a single high-affinity AtLFYbs (*API* probe) from the promoter of the *A. thaliana API* gene<sup>35</sup>. Electrophoretic mobility shift assay (EMSA) and SEC-MALLS experiments showed that GbLFY $\Delta$  and GbLFY $\Delta$ <sub>TERE</sub> bind *API* DNA mostly as dimers (Fig. 4c and Supplementary Table 1) and with the same apparent

affinity (Fig. 4a). Thus, the SAM domain is not essential for the formation of the LFY dimer/DNA complex. This is consistent with previous studies showing that LFY-DBD by itself dimerizes on DNA<sup>25</sup>.

We noticed, however, a slight migration shift for GbLFY $\Delta$ <sub>TE</sub>, GbLFY $\Delta$ <sub>RE</sub> and GbLFY $\Delta$ <sub>TERE</sub> complexes when compared with GbLFY $\Delta$  in EMSA (Fig. 4a,b). This shift disappeared when GbLFY $\Delta$ <sub>RE</sub> and GbLFY $\Delta$ <sub>TE</sub> (hereafter called GbLFY $\Delta$ <sub>[RE+TE]</sub>) were mixed at equimolar concentration (Fig. 4b). This difference in migration behaviour might be due to a slightly different conformation of the complexes when the SAM domain is monomeric. A comparable result was obtained for the AtLFY protein (Supplementary Fig. 4a). Taken together, these findings indicate that LFY-SAM oligomerization is not required to trigger dimer formation on DNA but it appears to only slightly influence the migration of the LFY/DNA complex.

#### The SAM domain prevents LFY DNA binding as monomer. In

addition to the main dimeric protein/DNA complex, GbLFY $\Delta$ <sub>TE</sub>, GbLFY $\Delta$ <sub>RE</sub> and GbLFY $\Delta$ <sub>TERE</sub> protein variants showed a weak monomeric complex with the *API* probe in EMSA (Fig. 4b). This was particularly obvious when using a mutated probe, *API*<sub>m</sub>, which contained mutations on each side of the LFYbs pseudo-palindrome and was not bound by WT GbLFY $\Delta$  (Fig. 4a bottom panel). GbLFY $\Delta$ <sub>TERE</sub> was also able to bind to DNA probes that



**Figure 3 | Mutations to the LFY-SAM oligomerization domain impair LFY floral function in *A. thaliana*.** (a) Pictures of WT, *lfy-12* mutant and *lfy-12* expressing pLFY:LFY or pLFY:LFY<sub>TERE</sub> inflorescences from 40-day-old plants grown under long day conditions (scale bar, 2 mm). (b) Distribution of pLFY:LFY and pLFY:LFY<sub>TERE</sub> primary transformants in phenotypic complementation classes (see Supplementary Table 2 for details). (c) Pictures of WT and transgenic plants expressing p35S:LFY or p35S:LFY<sub>TERE</sub> grown for 30 days under long-day conditions (scale bar, 1 cm). The white arrow indicates ectopic rosette flowers. (d) Proportion of the p35S:LFY and p35S:LFY<sub>TERE</sub> primary transformants showing ectopic rosette flowers.

carried mutations in either half of the palindromic binding site (*AP1m1* and *AP1m2* probes) as a monomer, whereas GbLFYΔ showed very little binding to such probes (Supplementary Fig. 4b). These experiments suggest that the LFY-SAM domain might prevent monomeric LFY binding on the numerous half sites present in a genome. To further investigate this possibility, we performed competition EMSA with increasing amounts of nonspecific unlabelled DNA competitor (fish genomic DNA). We found that GbLFYΔ<sub>TE</sub>, GbLFYΔ<sub>RE</sub> and GbLFYΔ<sub>TERE</sub> dissociated from *API* at lower competitor concentrations than GbLFYΔ and GbLFYΔ<sub>[RE+TE]</sub> (Fig. 4d and Supplementary Fig. 4c). Thus, the SAM domain appears to favour LFY dimeric binding at the expense of a less specific monomeric interaction with DNA.

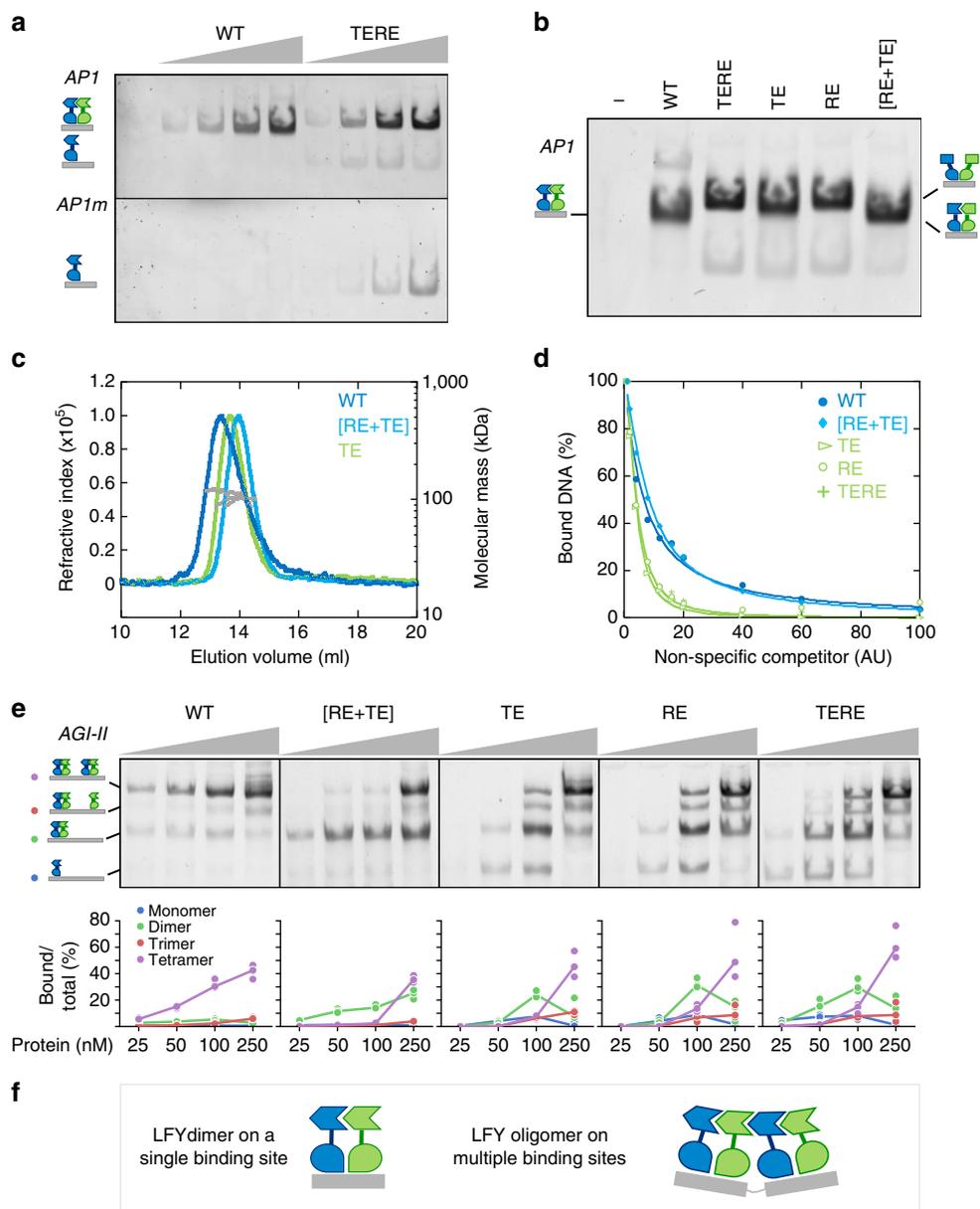
#### LFY-SAM<sub>TERE</sub> mutants show impaired cooperative DNA binding.

Because of their ability to oligomerize, SAM domains have been proposed to allow cooperative DNA binding on regions with multiple TFBS<sup>36</sup>. This property has been inferred from SAM protein variants whose oligomerization potential was limited to dimerization<sup>34</sup> but never established for a WT SAM protein. To test whether LFY-SAM could affect binding on DNA fragments with multiple binding sites *in vitro*, we used probes carrying two nearby LFYbs, either synthetic or from the *AGI-II*<sup>17</sup>. With these probes, the GbLFYΔ protein formed one main complex corresponding to two bound dimers (Fig. 4e and Supplementary Figs 4d and 5). At low concentrations, GbLFYΔ<sub>[RE+TE]</sub>, which mimics a WT dimer unable to further oligomerize (Fig. 2d), was not as efficient as GbLFYΔ at forming a tetrameric complex and the tetramer/dimer ratio was lower for GbLFYΔ<sub>[RE+TE]</sub> than for GbLFYΔ (Fig. 4e and Supplementary Fig. 5). As expected, all monomeric mutants (GbLFYΔ<sub>TE</sub>, GbLFYΔ<sub>RE</sub> and GbLFYΔ<sub>TERE</sub>) were also impaired in tetramer formation at low concentrations (Fig. 4e and Supplementary Fig. 5). In these assays, the AtLFY protein behaved similarly as GbLFYΔ (Supplementary Fig. 4a), indicating that mediating cooperative binding to multiple LFYbs is an intrinsic property of

the LFY-SAM domain and is important for the formation of tetrameric complexes (Fig. 4f).

#### A functional SAM domain is required for DNA binding *in vivo*.

To test the importance of a functional SAM domain for LFY DNA binding *in vivo*, we performed biological duplicate ChIP sequencing (ChIP-Seq) experiments with p35S:LFY, p35S:LFY<sub>TERE</sub> and non-transgenic *A. thaliana* (Col-0) 2-week-old seedlings. At this stage, the expression of the endogenous LFY is minimal as compared with that of the transgenes, enabling us to use the WT seedlings as a negative control. We chose to use ectopic expression in seedlings so that we could compare LFY and LFY<sub>TERE</sub> DNA-binding activity in the same tissue, a condition that would not be fulfilled by comparing *lfy* inflorescences expressing LFY or LFY<sub>TERE</sub>. We are aware that, when ectopically expressed, LFY might contact some genomic regions that are not genuine targets. However, it is known that the LFY-bound regions in seedlings and inflorescences significantly overlap and the LFY-binding motifs in the two tissues are extremely similar<sup>17,18</sup>. We therefore compared the intensity of LFY binding in p35S:LFY and p35S:LFY<sub>TERE</sub> for high confidence bound regions identified in p35S:LFY. After signal normalization (Supplementary Fig. 6), we found that the TERE mutation drastically reduced the LFY binding *in vivo* (Fig. 5a,b and Supplementary Fig. 7a). For all subsequent analyses, we define the coverage fold reduction (CFR) as the ratio between the LFY and the LFY<sub>TERE</sub> ChIP-Seq coverages. The CFR values ranged from 1.3 to >150 (Fig. 5a). Ninety-five per cent of the regions have a CFR > 3 and 54% have a CFR > 10. We also performed the converse analysis by selecting the highest confidence bound regions in p35S:LFY<sub>TERE</sub> and comparing their binding with that observed in p35S:LFY. Except for a few regions for which the binding by LFY<sub>TERE</sub> was slightly stronger, most regions were bound better by LFY (Supplementary Fig. 7b). The binding reduction is unlikely to be due to a less efficient recognition of LFY<sub>TERE</sub> as compared with LFY by the antibody used in the ChIP, as it was raised against LFY-DBD



**Figure 4 | Role of the LFY-SAM oligomerization domain in *in vitro* DNA binding.** (a) EMSA with 10 nM *AP1* or *AP1m* DNA probe and 0, 50, 100, 250 and 500 nM GbLFY $\Delta$  (WT) or GbLFY $\Delta_{\text{TERE}}$  (TERE) protein. This experiment was performed four times with similar results. (b) EMSA with 10 nM *AP1* DNA and 500 nM GbLFY $\Delta$  (WT), GbLFY $\Delta_{[\text{RE}+\text{TE}]}$  [RE + TE], GbLFY $\Delta_{\text{TE}}$  (TE), GbLFY $\Delta_{\text{RE}}$  (RE) or GbLFY $\Delta_{\text{TERE}}$  (TERE) protein. This experiment was performed twice with similar results. (c) SEC-MALLS molecular mass analysis of GbLFY $\Delta$ , GbLFY $\Delta_{\text{TE}}$  and GbLFY $\Delta_{[\text{RE}+\text{TE}]}$  in complex with the *AP1* probe. Fifty microlitres containing 2.6 mg ml $^{-1}$  (60  $\mu$ M) protein and 15  $\mu$ M DNA were used. Elution profiles were monitored by excess refractive index (left ordinate axis). The line under each elution peak shows the molecular mass distribution (right ordinate axis). The measured molecular mass for each complex is reported in Supplementary Table 1 and is consistent with that of a protein dimer on DNA. The SEC experiment was performed twice and the MALLS measurement was performed once. (d) Percentage of DNA bound to LFY as a function of nonspecific unlabelled competitor DNA concentration. Quantifications were based on EMSA shown in Supplementary Fig. 4c. This experiment was performed four times with similar results. (e) EMSA with 10 nM *AGI-II* DNA probe and 25, 50, 100 and 250 nM GbLFY $\Delta$ , GbLFY $\Delta_{[\text{RE}+\text{TE}]}$ , GbLFY $\Delta_{\text{TE}}$ , GbLFY $\Delta_{\text{RE}}$  or GbLFY $\Delta_{\text{TERE}}$  protein. Quantifications show the percentage of total DNA bound by one (monomeric complex, blue), two (dimeric complex, green), three (trimeric complex, red) or four (tetrameric complex, violet) LFY proteins. EMSA have been performed in triplicate and the fit follows the median value. Formation of the tetrameric complex starts at lower concentrations for WT GbLFY as compared with the mutant GbLFY versions. (f) Schematic representation of LFY bound as dimer on a single binding site (left) and as tetramer on two binding sites (right) deduced from the EMSA and SEC-MALLS analysis. DNA is in grey and LFY in blue and green. For all EMSAs, only the protein-DNA complexes are shown.

alone and yielded a comparable signal in western blotting on both proteins (Supplementary Fig. 3). The differences in CFR were apparently neither due to alterations to the sequence specificity of AtLFY $_{\text{TERE}}$ , as the binding site motifs derived from AtLFY and AtLFY $_{\text{TERE}}$  ChIP-Seq were very similar (Supplementary Fig. 7c).

This result is consistent with previously published SELEX experiments demonstrating that LFY N-terminal domain deletion does not affect LFY DNA-binding specificity<sup>15,17,24</sup>. Taken together, these findings indicate that LFY SAM domain is crucial for its genome-wide DNA binding *in vivo*.

**LFY-SAM facilitates binding to sites of suboptimal affinity.** The general reduction in DNA binding of LFY<sub>TERE</sub> is probably partially explained by its capacity to bind as a monomer and scatter over the numerous monomeric sites of the genome. However, as illustrated by three examples (Fig. 5b), the reduction in binding by mutants that are unable to oligomerize varies greatly between genomic regions (Fig. 5a). Gene ontology analysis for genes neighbouring the top high- and low-CFR regions did not detect any specific enrichment. We tried to understand whether the nature and number of LFYbs present in a region might help explaining the binding reduction, as our biochemical analysis suggested. To identify LFYbs, we used a previously validated and highly predictive position weight matrix that computes a score between 0 (highest affinity site) and  $-56$  (lowest affinity) for each 19 bp sequence<sup>17,26</sup>. To avoid using an arbitrary threshold above which a DNA stretch qualifies as LFYbs, we devised a specific procedure to compute the threshold (see methods). The low threshold we obtained ( $-25$ ) indicates that LFY-bound regions are enriched in LFYbs of very weak affinity (Supplementary Fig. 8a). When applied to regions bound by LFY<sub>TERE</sub>, the same procedure yielded a higher value ( $-20$ ) (Supplementary Fig. 8b), suggesting that oligomerization promotes binding to low-affinity sites. Consistent with this hypothesis, the bound regions that lack high-affinity LFYbs tend to have higher CFR than regions with high-affinity LFYbs (Fig. 5c). We also analysed the possible influence of LFYbs density and found that regions with high LFYbs density have higher CFR (Fig. 5d) than regions with less LFYbs. Thus, a functional SAM domain helps LFY to bind to DNA regions with either sub-optimal LFYbs or clusters of LFYbs. To investigate whether these clusters display a specific structure, we characterized the distribution of distances between LFYbs in bound regions. We found that binding sites separated by 1 bp were specifically overrepresented (three- to fourfold) in regions with high CFR (Fig. 5e). Reciprocally, the presence of one or more 1-bp distant sites significantly increased the CFR value (Fig. 5f). These analyses suggest that SAM-mediated oligomerization promotes LFY binding to a specific LFYbs configuration. We confirmed this hypothesis biochemically, by showing in EMSA that a distance of 1 bp between LFYbs most favoured LFY tetrameric binding over other spacing (0, 2, 6 and 11 bp) (Supplementary Fig. 5). Similar results were obtained when we limited the same analyses to the subset of genomic regions that are both bound by LFY in seedlings and by endogenous LFY in inflorescences<sup>18</sup> (Supplementary Fig. 9). Except for the effect of binding site density that lost its statistical significance, the conclusions are overall very similar despite the smaller size data set. Altogether, these analyses provide evidence that oligomerization through the LFY SAM domain facilitates LFY binding to regions that lack high-affinity binding sites or display multiple and adjacent sites.

**LFY-SAM allows access to closed chromatin regions.** As other SAM domain proteins have been suggested to regulate chromatin status<sup>37,38</sup>, we also tested whether the CFR could depend on the accessibility of DNA regions. This genomic feature can be monitored using DNaseI hypersensitivity (DHS) coupled to high-throughput sequencing (DNaseI-Seq). Such data were publicly available for 2-week-old WT *A. thaliana* seedlings<sup>39</sup>, the same developmental stage we used for the ChIP-Seq experiments. We therefore calculated the percentage of opened regions (following the criteria defined in ref. 39) in deciles of LFY-bound regions, sorted according to their CFR. In the low CFR regions, only 10–15% of the LFY-bound regions were closed. Strikingly, this number reached 70% in regions with high CFR

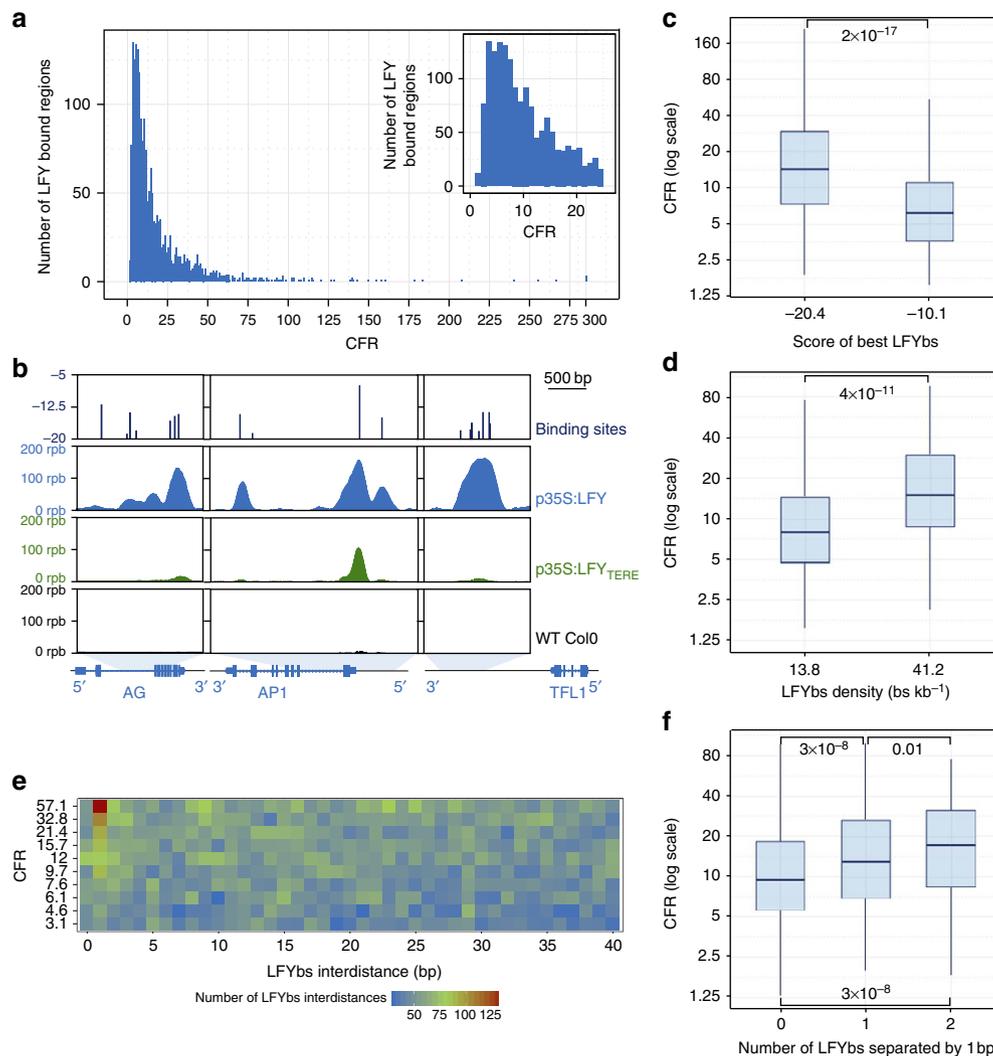
(Fig. 6a), suggesting that the chromatin state of the bound regions has a strong impact on the need for LFY oligomerization. To complement this analysis, we studied the correlation between quantitative DHS levels and the CFR. We found that the CFR values remarkably decreased when the accessibility (that is, the DHS level) increased (Fig. 6b) and poorly accessible regions have higher CFR values than accessible ones (Fig. 6c). An independent DNaseI-Seq data set from slightly younger seedlings (7-day-old) yielded similar results (Supplementary Fig. 7d)<sup>40</sup>. Again, these conclusions held true for the subset of genomic regions for which binding by LFY binding has also been validated in inflorescences (Supplementary Fig. 9g–i). Taken together, these results provide strong evidence that LFY binding to closed chromatin regions depends on the presence of a functional SAM domain.

## Discussion

The crystal structure of the conserved LFY N-terminus revealed a SAM oligomerization domain. This finding was unexpected not only because the LFY-N primary sequence has no homology with other SAM domains but also because LFY-N was initially proposed to be a putative leucine zipper dimerization domain<sup>27</sup>. The SAM domain appears to be highly conserved throughout LFY evolution (Fig. 1d and Supplementary Fig. 2d) and probably performs the same oligomerization function in all plant species. Both LFY from *G. biloba* (gymnosperm) and *A. thaliana* (angiosperm) oligomerize. Moreover, chimeric proteins bearing the AtLFY-DBD combined with the SAM from the moss *Physcomitrella patens* or the fern *Ceratopteris richardii* complemented the *A. thaliana lfy* mutant, indicating that the biochemical properties of the SAM domain are evolutionary conserved<sup>41</sup>.

In *planta* experiments revealed that altering the capacity of LFY to oligomerize compromised its floral function and drastically reduced its genome-wide DNA binding. The fact that TF oligomerization could contribute to DNA binding was expected based on the biochemical characterization we performed on LFY, as well as previous work on PRH-Hex<sup>9</sup> and on TFs from the ETS family<sup>34</sup>. Specifically, our study shows the extent to which genome-wide DNA binding can be altered when oligomerization is disrupted. The single TF for which this question was addressed so far is YAN, a *Drosophila* ETS TF<sup>38</sup>. The analysis of a non-oligomeric version of its SAM domain indicated that YAN oligomerization did not appear to be the primary determinant for its spreading over extended chromatin regions. Another study focused on PRC1, which is not a TF but a SAM domain member of the polycomb repressive complex in mammals. This work showed that mutation of its oligomerization capacity affected its DNA binding in only 12% of its target regions in mouse embryonic cells<sup>37</sup>. Our study thus offers evidence that TF oligomerization can have a profound impact on its genome-wide DNA-binding landscape.

We identified several reasons why the SAM oligomerization domain might influence LFY DNA binding. First, the SAM domain appears to limit LFY binding as a monomer and therefore might prevent LFY scattering over the genome due to nonspecific binding. How the SAM oligomerization domain reduces monomeric binding is unknown. However, it is tempting to speculate that it might impose constraints on the DBD, reducing its ability to interact with sub-optimal sites as a dimer. Second, combining *in vitro* and *in vivo* experiments, we showed that the SAM oligomerization domain facilitates cooperative higher-order LFY complex formation on regions with multiple or low-affinity LFYbs. Even for LFYbs that appear relatively isolated in the genome, it is conceivable that LFY oligomerization could contribute to cooperative binding through DNA looping.



**Figure 5 | Mutations altering LFY oligomerization affect genome-wide LFY DNA binding in planta.** **(a)** Histogram showing the distribution of CFR values for the LFY-bound regions in *p35S:Lfy*. The inset shows an enlargement for low CFR. **(b)** AtLFY and AtLFY<sub>TERE</sub> binding profiles on the *AG* (left), *API* (middle) and the *TFL1* (right) genomic regions. ChIP-Seq read coverage is shown for *p35S:Lfy* (blue), *p35S:Lfy<sub>TERE</sub>* (green) and Col-0 (black, background signal). The upper panel shows the LFYbs locations and scores. Blue boxes at the bottom indicate exons. rpb, read per base. **(c)** Boxplot showing the CFR of the LFY-bound regions with LFYbs of lowest and highest score (first and last deciles). The x axis shows the median LFYbs best score value. *P*-value for Mann-Whitney rank test is indicated on the graph. **(d)** Boxplot showing the CFR of the LFY-bound regions with lower and higher LFYbs density (first and last deciles of LFY-bound regions). The x axis shows the median density value (bs kb<sup>-1</sup>). *P*-value for Mann-Whitney rank test is indicated on the graph. **(e)** Heat map representing the distribution of distances between LFYbs. The LFY-bound regions in *p35S:Lfy* were pooled in deciles according to their CFR value (y axis shows the groups median CFR value). The x axis shows the distance (in bp) separating two 19-bp LFYbs. Only distances < 40 bp are shown. **(f)** Boxplot showing the CFR of regions depending on the number of pairs of LFYbs separated by 1 bp (zero, one and two pairs of LFYbs). *P*-values for Mann-Whitney rank test are indicated on the graph.

Cooperative TF DNA binding was previously shown to be important to trigger developmental switches<sup>42,43</sup> and the development of flowers is one of the best examples of a phase transition in plants. When coupled to clusters of low-affinity binding sites, TF oligomerization was shown to increase the sharpness of expression patterns<sup>44</sup>. Therefore, as levels of LFY progressively increase<sup>45</sup>, oligomerization may induce target genes with specific spatio-temporal patterns depending on their promoter *cis*-element topology (LFYbs affinity and density). For example, the early target *API* possesses only a few LFYbs, including one of the highest affinity sites found in the genome (score -7) and might thus be induced early during flower development, whereas genes such as *AP3*, *AG* or *TFL1* that contain a number of lower affinity sites (scores between -10 and -15) are regulated (induced or repressed) later during

flower development. SAM-mediated oligomerization could thus contribute to the timing in the expression of LFY target genes. Finally, although we consider it unlikely, we cannot totally exclude that the TERE mutations could have an indirect effect on LFY DNA binding *in vivo*, independent of oligomerization. These mutations could, for example, affect the interactions with yet unidentified regulators that contribute to the efficiency of LFY DNA binding.

By combining ChIP-Seq data with maps of DNA accessibility<sup>39,40</sup>, we have also provided evidence for a novel and unexpected role for SAM domain-mediated oligomerization: it enables LFY to access closed chromatin regions. Regulatory regions embedded in closed chromatin are usually poorly accessed by TFs<sup>4,46,47</sup>. However, a handful of factors, termed 'pioneer TF', are able to interact with packed chromatin to affect

transcriptional activity<sup>48,49</sup>. Several of those factors are potent developmental regulators involved in cell programming or reprogramming<sup>49</sup>. LFY is known to play a key role in stem cell emergence and floral fate determination<sup>10,15,16</sup>. Whether it qualifies as a pioneer TF will require testing its effect on the chromatin status of bound regions<sup>49</sup>. In recent times, the MADS

TF SEPALLATA3 and AP1 have also been proposed to play a pioneer role during floral organ development<sup>50–52</sup>. Interestingly, these TFs, as LFY<sup>50</sup>, interact with the several chromatin-remodelling factors<sup>51</sup> that, once recruited to closed chromatin regions, could contribute to their opening.

Determining the crystal structure of LFY-N not only revealed its identity as a SAM oligomerization domain but also provided the tools to uncover its function during floral development. We provide evidence that oligomerization can have a profound effect on a TF binding landscape by promoting cooperative binding of LFY to DNA, as was proposed for other oligomeric TFs, and, more unexpectedly, also gives LFY access to closed chromatin regions that are notably refractory to TF binding.

## Methods

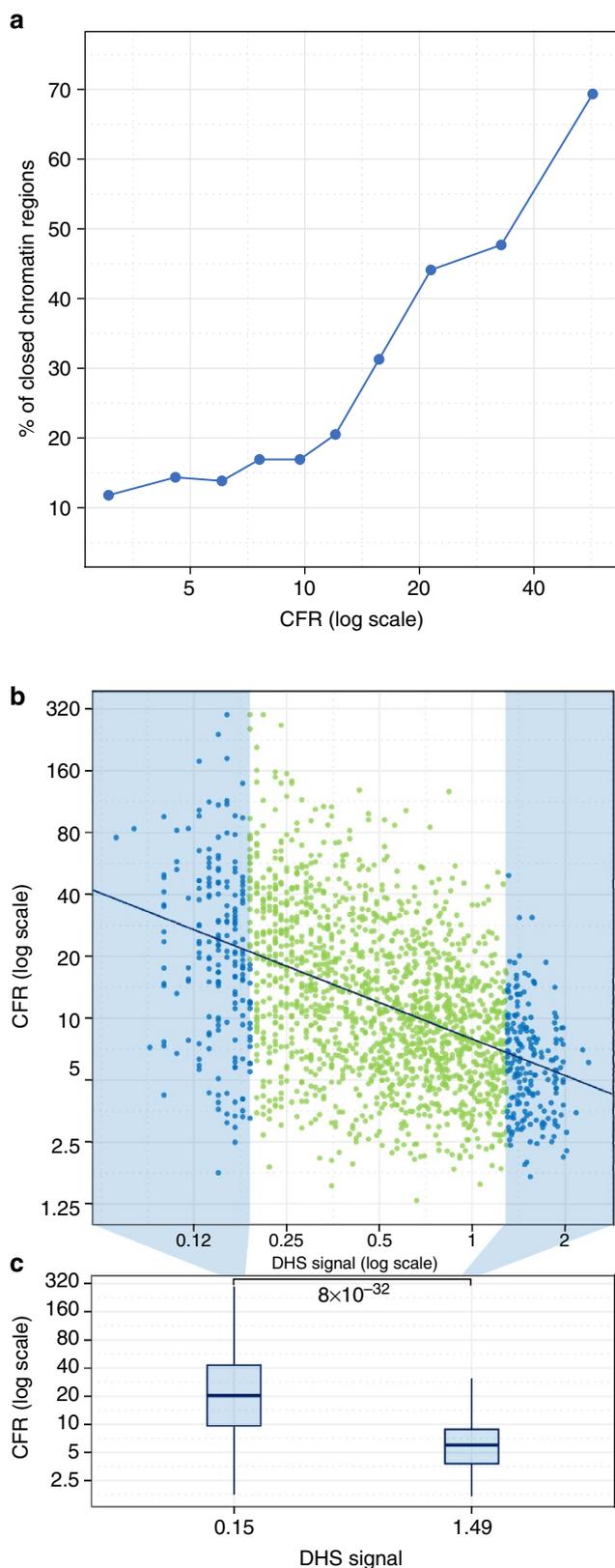
**Plant growth conditions and transformation.** WT, mutants and transgenic lines used are in the *A. thaliana* Columbia-0 background. *A. thaliana* seeds were sown on 0.5 × Murashige and Skoog (Duchefa biochemie) basal salt mixture medium. Plates were stratified 3 days at 4 °C, grown at 22 °C under long-day conditions and seedlings were transferred to soil. *Lfy-12* mutation genotyping was performed as described<sup>35</sup>. *A. tumefaciens* C58C1 pMP90 was used for stable transformation of *lfy-12* heterozygous plants by the floral dip method<sup>53</sup>.

**Plasmid constructions for protein expression in *Escherichia coli*.** Plasmids were constructed using primers listed in the Supplementary Table 3. GbLFYΔ (*pETH164*; residues 55–402 from *G. biloba* LFY complementary DNA) was amplified with primers oETH1067 and oETH1068, cloned into the PCR-Blunt vector (Invitrogen) to yield the pCA04 vector and transferred to the *pETM-11* expression vector using NcoI and XhoI restriction sites. GbLFY-SAM (*pETH195*; residues 54–159 from *G. biloba* LFY cDNA) and AtLFY-SAM (*pETH201*; residues 38–151 from *A. thaliana* LFY cDNA) were amplified with the primers oETH1126/oETH1128 and oETH1130/oETH1131, respectively. The PCR products were cloned into the pETM-11 expression vector<sup>54</sup> using NcoI and XhoI restriction sites, generating fusions with an N-terminal 6 × His tag, cleavable by the tobacco etch virus (TEV) protease. AtLFY (*pETH94*; residues 1–420 from *A. thaliana* LFY cDNA) was amplified with the primer oETH1031/oETH1032, cloned into the PCR-Blunt vector and transferred into the *pET-30a* vector (Novagen) (C-terminal 6 × His tag) using NcoI and XhoI restriction sites.

Mutagenesis was done by site-directed mutagenesis with primers listed in the Supplementary Table 4. GbLFYΔ<sub>TE</sub> (*pCA21*), GbLFYΔ<sub>RE</sub> (*pCA22*) and GbLFYΔ<sub>TERE</sub> (*pCA39*) were derived from *pETH164*; AtLFY<sub>TE</sub> (*pCA23*), AtLFY<sub>RE</sub> (*pCA24*) and AtLFY<sub>TERE</sub> (*pCA25*) were derived from *pETH94*; GbLFY-SAM<sub>TE</sub> (*pCA15*), GbLFY-SAM<sub>RE</sub> (*pCA17*) and GbLFY-SAM<sub>TERE</sub> (*pCA20*) were derived from *pETH195*. All plasmids were verified by appropriate digestions or sequencing.

**Binary vector constructions for *A. thaliana* transformation.** *pLFY:LFY<sub>TERE</sub>* (*pCA35*) and *p35S:LFY<sub>TERE</sub>* (*pCA29*) were obtained by inserting a PstI-SalI fragment containing the TERE mutations from *pCA25* in the *pETH29* and *pCA26* (ref. 15), respectively.

**Protein expression and purification.** Proteins were expressed using *E. coli* Rosetta2 (DE3) strain (Novagen). Cells were grown in Luria-Bertani medium supplemented with Kanamycin (50 μg ml<sup>-1</sup>) and Chloramphenicol (34 μg ml<sup>-1</sup>) at 37 °C under agitation up to an optical density of 600 nm of 0.6. Betaine (2 mM) was added and cultures were shifted to 17 °C for 1 h before addition of 0.4 mM isopropyl β-D-1-thiogalactopyranoside. After overnight growth at 17 °C, cells were pelleted. Production of selenomethionine (SeMet) GbLFY-SAM for crystallography was carried out in *E. coli* B834 (DE3) (met<sup>-</sup>) strain (Novagen) transformed with the GbLFY-SAM plasmid. The growth of bacteria was initiated in 0.5 l of Luria-Bertani medium supplemented with Kanamycin (50 μg ml<sup>-1</sup>) and Chloramphenicol (34 μg ml<sup>-1</sup>) at 37 °C under agitation up to an optical density of



**Figure 6 | Effect of the chromatin state on the genome-wide LFY DNA binding in planta.** (a) Percentage of closed regions in deciles of *p35S:LFY*-bound regions sorted according to their CFR. (b) Comparison of the CFR of the LFY-bound regions in *p35S:LFY* and the chromatin accessibility in 15-day-old seedlings (DHS signal)<sup>39</sup>. Pearson's correlation test was performed between the CFR and the DHS signal:  $r^2 = 21.71\%$ ,  $P$ -value =  $7. \text{e} - 106$ . Data are shown in light blue and green, and fitted values are shown in dark blue. (c) Boxplot representing the CFR signal in the regions of highest and lowest DHS (first and last deciles).  $P$ -value for Mann-Whitney rank test on the two extreme groups is indicated on the graph.

600 nm of 0.5. Bacteria were harvested by centrifugation and washed with M9 minimal medium. The growth was continued at 37 °C until an optical density of 600 of 0.6 in 1 litre of SeMet buffer ( $2 \times$  M9 minimal medium +  $\text{MgSO}_4$  (2 mM),  $\text{FeSO}_4$  (25 mg  $\text{ml}^{-1}$ ), glucose (4 g  $\text{l}^{-1}$ ), vitamins (thiamine, pyridoxine, riboflavin and niacinamide at 1 mg  $\text{ml}^{-1}$ ), a mix of all amino acids, except methionine (40 mg  $\text{ml}^{-1}$ ), SeMet (40 mg  $\text{ml}^{-1}$ ) and antibiotics (34  $\mu\text{g ml}^{-1}$  of Chloramphenicol and 50  $\mu\text{g ml}^{-1}$  of Kanamycin). The expression of the SeMet protein and growth of the cells at 17 °C were further performed as described above for the GbLFY-SAM protein. All purification steps were performed at 4 °C. All proteins were solubilized and first purified by an affinity chromatography column. The buffer composition used for each protein is given in Supplementary Table 5. Pellets corresponding to 0.5 l culture containing the recombinant protein were sonicated in 50 ml buffer supplemented by one protease inhibitor cocktail tablet Complete EDTA-free (Roche) and centrifuged for 30 min at 20,000 g. The clear supernatant was transferred on a column containing 1 ml Ni-Sepharose High Performance resin (GE Healthcare), washed with buffer containing 20 and 40 mM imidazole, and eluted with buffer containing 300 mM imidazole. Eluted fractions were immediately diluted three times in buffer without imidazole and dialysed overnight.

For GbLFY-SAM, GbLFY-SAM<sub>TE</sub>, GbLFY-SAM<sub>RE</sub>, GbLFY-SAM<sub>TERE</sub> and AtLFY-SAM, after dialysis, the 6  $\times$  His tag was cleaved overnight by the TEV protease (5% w/w). The cleavage product was loaded on a Ni-Sepharose High Performance resin to remove the 6  $\times$  His tag, the non-cleaved protein and the 6  $\times$  His-tagged TEV protease. The protein was eluted with buffer A (Supplementary Table 5) plus 20 mM imidazole. Eluted fractions were applied to a Hi-load Superdex-200 16/60 Prep Grade Column (GE Healthcare) equilibrated with buffer A. After concentration using Amicon Ultra Centrifugal filters (Millipore), the protein concentration was assessed using a NanoDrop-2000 spectrophotometer (Thermo Fisher Scientific Inc.). For crystallography, the SeMet GbLFY-SAM was purified as the native protein.

After dialysis in buffer B, GbLFY $\Delta$ , GbLFY $\Delta$ <sub>TE</sub>, GbLFY $\Delta$ <sub>RE</sub> and GbLFY $\Delta$ <sub>TERE</sub> proteins were concentrated using Amicon Ultra Centrifugal filters and applied to a Hi-load Superdex-200 16/60 Prep Grade Column equilibrated with buffer C. Protein concentration was assessed using a NanoDrop-2000 spectrophotometer.

After dialysis in buffer D, AtLFY, AtLFY<sub>TE</sub>, AtLFY<sub>RE</sub> and AtLFY<sub>TERE</sub> proteins were concentrated using Amicon Ultra Centrifugal filters. Protein concentrations were assessed using the Bradford assay<sup>55</sup>. All proteins were flash frozen in liquid nitrogen and stored at  $-80$  °C.

**Crystal structure determination.** Initial crystallization conditions were identified using the high-throughput crystallization platform at EMBL Grenoble (embl.fr/htxlab). The optimum condition was obtained at 20 °C with the hanging drop vapour diffusion method by mixing 1  $\mu\text{l}$  GbLFY-N, native or Se-Met substituted, at 2.5 mg  $\text{ml}^{-1}$  with 1  $\mu\text{l}$  of reservoir solution containing 25 mM Tris-HCl pH 8.8 and 40 mM ammonium sulfate. Crystals were cryoprotected by plunging into liquid nitrogen, after incubation in well conditions to which 20% glycerol had been supplemented. The structure was solved by the single-wavelength anomalous dispersion method from Se-Met substituted protein). A highly redundant data set was collected on the microfocus beamline ID23-2 (ref. 56). The selenium substructure was determined by SHELXC/D<sup>57</sup> and the selenium positions were refined in SHARP<sup>58</sup>. Phases were calculated, solvent-flattened maps were obtained with PIRATE<sup>59</sup> and a partial model was automatically built with BUCCANEER<sup>60</sup>. The model was manually extended into the experimental electron density map and the resultant model was then used as a search model for molecular replacement into the higher-resolution native data (preserving and extending the reflections previously selected for the free set). Successive rounds of model improvement in COOT<sup>61</sup> and refinement in BUSTER<sup>62</sup> were then performed. The built region in the two monomers present in the asymmetric unit comprised residues 54–134 and 56–133, respectively.

**Western blotting.** *A. thaliana* leaves from 15-day-old seedlings grown under long-day conditions were harvested and one leaf from six to eight seedlings were pooled for each transgenic line. The leaves were flash frozen in liquid nitrogen, mechanically crushed, immediately suspended in 80  $\mu\text{l}$  denaturing buffer (50 mM Tris-HCl pH 6.8, 10% w/v glycerol, 1% w/v SDS, 0.0025% w/v Bromophenol blue and 0.4% w/v dithiothreitol) for 10 mg of fresh matter and denatured for 5 min at 95 °C. The samples were run on a SDS-PAGE gel and transferred on a nitrocellulose membrane (Immobilon P transfer membrane, Millipore) for blotting. The equivalent of 18.75 and 12.5 mg of fresh matter were used for anti-LFY and anti-KARI blots, respectively. Five nanograms of recombinant LFY protein (LFY $\Delta$ , produced in *E. coli*) were used as a positive control. Two anti-LFY antibodies were used: JA70 raised in rabbit against the recombinant LFY protein<sup>17</sup> (used at 1/6,000 dilution in Supplementary Fig. 3a) and 4028 (ref. 17) raised in rabbit against the LFY C-terminal amino acids 223–424 (BioGenes). Anti-KARI antibodies (used at 1/3,000) were described in ref. 63. Horseradish peroxidase-conjugated AffiniPure goat anti-rabbit IgG (used at 1/15,000, 111-035-144, Jackson ImmunoResearch) was used as secondary antibody for detection. For Supplementary Fig. 3b, antibodies J70 and 4028 were used at 1/25,000 followed by the secondary antibody at 1/50,000 dilution. Revelation was done using the Pierce ECL2 Western Blotting Substrate kit (Thermo Scientific) and scanned on a Typhoon 9400 scanner

(Molecular Dynamics). Uncropped western blottings and protein ladders are presented as Supplementary Fig. 10.

**Protein sequence and structure analysis.** Three-dimensional structure images were done using PyMOL (www.pymol.org); sequence alignment was done using Multalin (http://multalin.toulouse.inra.fr/multalin/) and visualized using ESPrit (http://espriti.ibcp.fr/ESPrIt/ESPrIt/). Root mean squared deviations were computed using the Dali server (http://ekhidna.biocenter.helsinki.fi/dali\_server/). Surface conservation was computed using the ConSurf server (http://consurf.tau.ac.il/; (ref. 64)).

**Electrophoretic mobility shift assay.** Oligonucleotides used for EMSA are listed in Supplementary Table 6. For *API*, *APIm1*, *APIm2*, *AGI-II* and *S-AGI*, complementary single-stranded oligonucleotides were annealed in annealing buffer (10 mM Tris pH 7.5, 150 mM NaCl and 1 mM EDTA). The resulting double-stranded DNA with a protruding G was fluorescently labelled by end filling: 4 pmol of double-stranded DNA was incubated with 1 unit of Klenow fragment polymerase (Ozyme) and 8 pmol Cy3-dCTP or Cy5-dCTP (GE Healthcare) in Klenow buffer during 2 h at 37 °C, followed by 10 min enzyme inactivation at 65 °C. Binding reactions were performed in 20  $\mu\text{l}$  binding buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% glycerol, 0.25 mM EDTA, 2 mM  $\text{MgCl}_2$ , 0.01% Tween-20 and 3 mM TCEP) with 10 nM labelled probe,  $1 \times$  (28 ng  $\text{ml}^{-1}$ ) fish sperm DNA (Roche) as nonspecific competitor and 25–500 nM proteins.

Competition assays were performed in duplicates and  $1-100 \times$  fish sperm DNA (Roche) was used in the binding reaction. Signal quantification was performed in using ImageLab v2.0.1 (Bio-Rad Laboratories). Signal of each protein-DNA complex was quantified relatively to total DNA signal. For Fig. 4e, each binding reaction was performed in triplicate. Uncropped gels are presented as Supplementary Fig. 10.

**SEC and SEC-MALLS.** The molecular mass of GbLFY-SAM and AtLFY-SAM was estimated at 4 °C using a Superdex-200 10/300GL column (GE Healthcare), equilibrated with buffer A and calibrated with low- and high-molecular-weight protein standards (gel filtration calibration kit; GE Healthcare). Accurate molecular mass determination using SEC-MALLS was carried out with a Superdex-200 10/300GL column (GE Healthcare). GbLFY-SAM, WT and mutants were analysed in buffer A. Protein-DNA complexes containing GbLFY $\Delta$ , WT or mutants and *API* DNA were analysed in 20 mM Tris-HCl pH 8, 150 mM NaCl, 0.25 mM EDTA, 2 mM  $\text{MgCl}_2$  and 1 mM TCEP. Separations were performed at 20 °C with a flow rate of 0.5 ml  $\text{min}^{-1}$ . Elutions were monitored by using a DAWN-EOS detector with a laser emitting at 690 nm for online MALLS measurement (Wyatt Technology Corp., Santa Barbara, CA) and with a RI2000 detector for online refractive index measurements (Schambeck SFD). Molecular mass calculations were performed using the ASTRA software using a refractive index increment ( $dn/dc$ ) of 0.185 ml  $\text{g}^{-1}$ .

**Chromatin immunoprecipitation sequencing.** Fifteen-day-old seedling from lines CA26 #15 (*p35S:LFY*), CA29 #17 (*p35S:LFY<sub>TERE</sub>*) and Col-0 were grown on 0.5  $\times$  Murashige and Skoog medium under long-day conditions. The experiment was performed twice to produce biological replicates. Seedlings ( $\sim 1$  g) were fixed with 1% formaldehyde in MC buffer (10 mM sodium phosphate pH 7.0, 50 mM NaCl and 0.1 M sucrose) for 1 h under vacuum. Fixation was stopped with 0.125 M glycine, followed by three washes with MC buffer. The tissue was ground in liquid nitrogen, the powder was suspended in 15 ml M1 buffer (10 mM sodium phosphate pH 7.0, 0.1 M NaCl, 1 M 2-methyl 2,4-pentanediol, 10 mM  $\beta$ -mercaptoethanol, Complete Protease Inhibitor Cocktail (Roche Diagnostics GmbH, Mannheim, Germany)). The slurry was filtrated three times through 55- $\mu\text{m}$  mesh (Miracloth, Calbiochem) and centrifuged at 1,000 g for 10 min at 4 °C. Filtration and centrifugation were repeated twice. Subsequent steps were at 4 °C, unless indicated otherwise. The pellet was washed five times with 1 ml M2 buffer (M1 buffer with 10 mM  $\text{MgCl}_2$  and 0.5% Triton X-100), once with 1 ml M3 buffer (M1 without 2-methyl 2,4-pentanediol) and resuspended in 1 ml Sonic buffer (10 mM sodium phosphate pH 7.0, 0.1 M NaCl, 0.5% Sarkosyl, 10 mM EDTA, Complete Protease Inhibitor Cocktail and 1 mM PEFA BLOC SC (Roche Diagnostics)), and sonicated with a Focused-ultrasonicator S2 (Covaris) (Duty cycle: 20%, intensity: 5, cycles per burst: 200, cycle time: 2 min). After sonication, the suspension was centrifuged (microcentrifuge, top speed) for 5 min and the supernatant was mixed with one volume of IP buffer (50 mM Hepes pH 7.5, 150 mM KCl, 5 mM  $\text{MgCl}_2$ , 10  $\mu\text{M}$   $\text{ZnSO}_4$ , 1% Triton X-100 and 0.05% SDS). The solubilized chromatin was incubated overnight with 2.5  $\mu\text{l}$  anti-LFY serum antibody 4028. The immunoprotein-chromatin complexes were captured by incubation with protein A-agarose beads (Santa Cruz Biotechnology) for 1 h on a rotating wheel. The protein A-agarose beads were washed five times with 1 ml IP buffer for 10 min at room temperature. Beads were vortexed for 30 s with 100  $\mu\text{l}$  cold glycine elution buffer (0.1 M glycine, 0.5 M NaCl and 0.05% Tween-20 pH 2.8) and pelleted in a microfuge (room temperature, 1 min, top speed). The supernatant was mixed with 50  $\mu\text{l}$  of 1 M Tris pH 9, to neutralize the eluant. Elution and neutralization were repeated twice. The eluted sample was spun at top speed for 2 min in a microcentrifuge at room temperature. The eluate was treated with 1  $\mu\text{l}$  RNase A/T1

mix 10 mg ml<sup>-1</sup> (Fermentas) and 1.5 µl Proteinase K, recombinant, PCR grade (Roche). After overnight incubation, a second aliquot of Proteinase K was added and incubated at 65 °C for 6 h. DNA was purified using Minelute columns (Qiagen). Libraries for high-throughput sequencing were prepared using standard Illumina protocols. Deep sequencing was performed on an Illumina GAIIX instrument following the manufacturer's instructions.

**Bioinformatics analyses.** Bioinformatics analyses were performed using the SHORE suite (version 0.9.3) (ref. 65) and scripts in python (version 2.7.3) using BioPython (version 1.60), pybedtools (version 0.6.4) and Matplotlib (version 1.2.0) modules. All statistical analyses were managed using R (version 3.0.2) and figures were generated using the ggplot2 module (version 1.0.0). All scripts are available on request.

Read mapping and duplicates removal were performed using the SHORE pipeline (preprocess, import, mapflowcell subprocesses) using -H 1,1 -M 0,4 -X 130 -B 1 as parameters. The number of unique reads mapped is indicated in the Supplementary Table 7. Peakfinding was performed with the shore peak subprogramme using -B 1 as parameter. We selected as bound regions all peaks with a BH-FDR-*q* < 10<sup>-7</sup> in all four comparisons of the two replicates against the two controls. A python script was used to build bound peak bed files from SUMMARY.txt shore peak output. We obtained 1,954 peaks for *p35S:LFY* and 176 for *p35S:LFY<sub>TERE</sub>*. Corresponding bed files with peak rank and CFR values have been uploaded on the GEO database.

For each genomic position, the read number corresponds to the number of reads extended by 130 bp covering this position. For each bound region (ChIP-Seq peak), the coverage is defined as the area above the read number curve.

The normalization procedure was performed in two steps based on peak coverage comparisons. In *p35S:LFY* and *p35S:LFY<sub>TERE</sub>* we used peaks detected by SHORE, and in the Col control we used a python script as rudimentary peak finder using 20 as a cutoff value for signal (Supplementary Fig. 6). First, we performed intra genotype normalization between ChIP-Seq replicates from the same genotype (Col, *p35S:LFY* or *p35S:LFY<sub>TERE</sub>*). For each genotype, we used the most significant peaks and plotted the coverage values of one replicate against the other. The regression coefficient *m* as shown in Supplementary Fig. 6 was used to normalize one replicate against the other.

Replicates were then fused by calculating a normalized read count at each position

$$\text{intraNormReadCount} = \frac{1}{2} * \left( \text{readCount}_{\text{sample1}} + \frac{\text{readCount}_{\text{sample2}}}{m} \right) \quad (1)$$

Second, inter genotype normalization was performed based on the background peaks detected in the Col control sample. For each background peak, coverages were compared between normalized *p35S:LFY* or *p35S:LFY<sub>TERE</sub>* and the Col control sample (Supplementary Fig. 6b), and the regression coefficient *m'* was computed and used for final normalization:

$$\text{NormReadCount} = \text{intraNormReadCount} / m' \quad (2)$$

for *p35S:LFY* and *p35S:LFY<sub>TERE</sub>* ChIP-Seq. All coefficients used are listed in Supplementary Table 8.

To calculate read coverages (based on ChIP-Seq or DnaseI-Seq signals), we selected the genomic information included in '.wig' or '.bedgraph' files at bound regions using the bedtools intersect subprogram (-wa -wb option) followed by a python script that calculates the coverage of each of those genomic regions. The CFR due to the TERE mutation was computed with R using equation (3) on normalized wig files for each bound regions.

$$\text{CFR} = \text{coverage}(p35S : LFY) / \text{coverage}(p35S : LFY_{TERE}) \quad (3)$$

When computing CFR for the 1,954 LFY-bound regions (Fig. 5a), the very low coverage in *p35S:LFY<sub>TERE</sub>* for two regions was leading to very high or infinite CFR values. Those two CFR values were arbitrarily set up to 300.

The chromatin information about chromatin state (open versus closed) for each bound region was retrieved with a python script from *dhleaf.txt* and *dhflower.txt* files from ref. 39. A bound region was considered open when its centred half was entirely open.

To find the LFYBs in ChIP-Seq peaks, we used a previously optimized position weight matrix<sup>17</sup> and python script computing a score between -56 and 0 to each 19-pb sequence. Zero corresponds to the site of highest affinity and -56 corresponds to the lowest affinity. To determine what score threshold should be used for binding-site identification, we compared bound regions with a negative set of unbound regions. The negative set was built with a python script that takes a bed file with bound regions as input and randomly selects in the *Arabidopsis* genome regions of same size, similar GC content and with various annotation type (CDS, intron, upstream, downstream or intergenic). The LFYBs scores for each bound and unbound regions were calculated. We then used an R script to compute the density of LFYBs in sliding score windows (going from [0; -1] to [-55; -56]). As shown in Supplementary Fig. 8a, the density of LFYBs is significantly higher in the positive set as compared with the negative one for scores better than -25, as judged using a Mann-Whitney rank test on each window of scores (*P*-value < 0.05). The corresponding binding site density might appear surprisingly high but we believe that our score threshold calculation is more valid than

commonly used arbitrary cutoff score values. Moreover, our analyses were robust to threshold change within a reasonable range [-20, -26].

For the heat map of interdistances (Fig. 5e), we used a python script to divide regions in deciles according to their CFR, calculate all distances between LFYBs and plot their number within the decile. To perform the analysis presented in Supplementary Fig. 9, we kept only the 436 regions found as bound in *p35S:LFY* seedlings that intersected with regions bound by LFY in inflorescences according to ref. 18 and performed the whole pipeline of analysis. R was used for all figures and statistical analyses, except that Supplementary Fig. 6 was generated by the Matplotlib module of python.

## References

- Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
- Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
- Weirauch, M. T. & Hughes, T. R. in *A Handbook of Transcription Factors* **52**, 25–73 (2011).
- Slattery, M. *et al.* Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014).
- Bienz, M. Signalosome assembly by domains undergoing dynamic head-to-tail polymerization. *Trends Biochem. Sci.* **39**, 487–495 (2014).
- Hollenhorst, P. C., McIntosh, L. P. & Graves, B. J. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu. Rev. Biochem.* **80**, 437–471 (2011).
- Nanao, M. H. *et al.* Structural basis for oligomerization of auxin transcriptional regulators. *Nat. Commun.* **5**, 3617 (2014).
- Katsani, K. R., Hajibagheri, M. A. N. & Verrijzer, C. P. Co-operative DNA binding by GAGA transcription factor requires the conserved BTB / POZ domain and reorganizes promoter topology. *EMBO J.* **18**, 698–708 (1999).
- Williams, H., Jayaraman, P.-S. & Gaston, K. DNA wrapping and distortion by an oligomeric homeodomain protein. *J. Mol. Biol.* **383**, 10–23 (2008).
- Moyroud, E., Tichtinsky, G. & Parcy, F. The LEAFY floral regulators in angiosperms: conserved proteins with diverse roles. *J. Plant Biol.* **52**, 177–185 (2009).
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R. & Parcy, F. LEAFY blossoms. *Trends Plant Sci.* **15**, 346–352 (2010).
- Irish, V. F. The flowering of *Arabidopsis* flower development. *Plant J.* **61**, 1014–1028 (2010).
- Chandler, J. W. Floral meristem initiation and emergence in plants. *Cell. Mol. Life Sci.* **69**, 3807–3818 (2012).
- Posé, D., Yant, L. & Schmid, M. The end of innocence: flowering networks explode in complexity. *Curr. Opin. Plant Biol.* **15**, 45–50 (2012).
- Chahtane, H. *et al.* A variant of LEAFY reveals its capacity to stimulate meristem development by inducing RAX1. *Plant J.* **74**, 678–689 (2013).
- Yamaguchi, N. *et al.* A molecular framework for auxin-mediated initiation of flower primordia. *Dev. Cell* **24**, 271–282 (2013).
- Moyroud, E. *et al.* Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* LEAFY transcription factor. *Plant Cell* **23**, 1293–1306 (2011).
- Winter, C. M. *et al.* LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev. Cell* **20**, 430–443 (2011).
- Weigel, D. & Nilsson, O. A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377**, 495–500 (1995).
- Gallois, J.-L., Nora, F. R., Mizukami, Y. & Sablowski, R. WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem. *Genes Dev.* **18**, 375–380 (2004).
- Risseuw, E. *et al.* An activated form of UFO alters leaf development and produces ectopic floral and inflorescence meristems. *PLoS ONE* **8**, e83807 (2013).
- Kaufmann, K., Pajoro, A. & Angenent, G. C. Regulation of transcription in plants: mechanisms controlling developmental switches. *Nat. Rev. Genet.* **11**, 830–842 (2010).
- Engelhorn, J., Blanvillain, R. & Carles, C. C. Gene activation and cell fate control in plants: a chromatin perspective. *Cell. Mol. Life Sci.* **71**, 3119–3137 (2014).
- Sayou, C. *et al.* A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* **343**, 645–648 (2014).
- Hamès, C. *et al.* Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.* **27**, 2628–2637 (2008).
- Minguet, E. G., Segard, S., Charavay, C. & Parcy, F. MORPHEUS, a webtool for transcription factor binding analysis using position weight matrices with dependency. *PLoS ONE* **10**, e0135586 (2015).
- Siriwardana, N. S. & Lamb, R. S. A conserved domain in the N-terminus is important for LEAFY dimerization and function in *Arabidopsis thaliana*. *Plant J.* **71**, 736–749 (2012).
- Qiao, F. & Bowie, J. U. The many faces of SAM. *Sci. STKE* **2005**, re7 (2005).

29. Meruelo, A. D. & Bowie, J. U. Identifying polymer-forming SAM domains. *Proteins* **74**, 1–5 (2009).
30. Kim, C. a *et al.* Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J.* **20**, 4173–4182 (2001).
31. Ikeda-Kawakatsu, K., Maekawa, M., Izawa, T., Itoh, J.-I. & Nagato, Y. ABERRANT PANICLE ORGANIZATION 2/RFL, the rice ortholog of *Arabidopsis* LEAFY, suppresses the transition from inflorescence meristem to floral meristem through interaction with APO1. *Plant J.* **69**, 168–180 (2012).
32. Levin, J. Z. & Meyerowitz, E. M. UFO: an *Arabidopsis* gene involved in both floral meristem and floral organ development. *Plant Cell* **7**, 529–548 (1995).
33. Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F. & Meyerowitz, E. M. LEAFY controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843–859 (1992).
34. Green, S. M., Coyne, H. J., McIntosh, L. P. & Graves, B. J. DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J. Biol. Chem.* **285**, 18496–18504 (2010).
35. Benlloch, R. *et al.* Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of APETALA1. *Plant J.* **67**, 1094–1102 (2011).
36. Zhang, J. *et al.* Sterile alpha motif domain-mediated self-association plays an essential role in modulating the activity of the *Drosophila* ETS family transcriptional repressor Yan. *Mol. Cell. Biol.* **30**, 1158–1170 (2010).
37. Isono, K. *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26**, 565–577 (2013).
38. Webber, J. L. *et al.* The relationship between long-range chromatin occupancy and polymerization of the *Drosophila* ETS family transcriptional repressor Yan. *Genetics* **193**, 633–649 (2013).
39. Zhang, W., Zhang, T., Wu, Y. & Jiang, J. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* **24**, 2719–2731 (2012).
40. Sullivan, A. M. *et al.* Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep.* **8**, 2015–2030 (2014).
41. Maizel, A. *et al.* The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* **308**, 260–263 (2005).
42. Cherry, J. L. & Adler, F. R. How to make a biological switch. *J. Theor. Biol.* **203**, 117–133 (2000).
43. Lebrecht, D. *et al.* Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*. *Proc. Natl Acad. Sci. USA* **102**, 13176–13181 (2005).
44. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**, 535–540 (2008).
45. Blázquez, M. A., Soowal, L. N., Lee, I. & Weigel, D. LEAFY expression and flower initiation in *Arabidopsis*. *Development* **124**, 3835–3844 (1997).
46. Sherwood, R. I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
47. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* **15**, 69–81 (2014).
48. Guertin, M. J. & Lis, J. T. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr. Opin. Genet. Dev.* **23**, 116–123 (2013).
49. Iwafuchi-doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Cell* **158**, 2679–2692 (2014).
50. Wu, M.-F. *et al.* SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. *Proc. Natl Acad. Sci. USA* **109**, 3576–3581 (2012).
51. Smaczniak, C. *et al.* Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl Acad. Sci.* **109**, 1560–1565 (2012).
52. Pajoro, A. *et al.* Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol.* **15**, R41 (2014).
53. Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W. & Chua, N.-H. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* **1**, 641–646 (2006).
54. Dümmler, A., Lawrence, A.-M. & de Marco, A. Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb. Cell Fact.* **4**, 34 (2005).
55. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
56. Flot, D. *et al.* The ID23-2 structural biology microfocus beamline at the ESRF. *J. Synchrotron Radiat.* **17**, 107–118 (2010).
57. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
58. Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2023–2030 (2003).
59. Cowtan, K. Generic representation and evaluation of properties as a function of position in reciprocal space. *J. Appl. Crystallogr.* **35**, 655–663 (2002).
60. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
61. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
62. Bricogne, G. *et al.* BUSTER. Available at <http://www.globalphasing.com> (2011).
63. Dumas, R., Joyard, J. & Douce, R. Purification and characterization of acetohydroxyacid reductoisomerase from spinach chloroplasts. *Biochem. J.* **262**, 971–976 (1989).
64. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533 (2010).
65. Ossowski, S. *et al.* Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18**, 2024–2033 (2008).
66. Sathyamurthy, A., Freund, S. M. V., Johnson, C. M., Allen, M. D. & Bycroft, M. Structural basis of p63 $\alpha$  SAM domain mutants involved in AEC syndrome. *FEBS J.* **278**, 2680–2688 (2011).

## Acknowledgements

We thank E. Moyroud for the gift of GbLFY cDNA; A. Mathelier, A. Fejes, R. Worsley-Hunt, W.W. Wasserman, C. Petosa, B. Ycart and P. Laufs for helpful discussions; and C. Zubieta, J. Zubieta, J. Engelhorn, C. Carles and J. Govin for critical reading of the manuscript. This work was supported by a DFG-funded ERA-PG grant (BloomNet; SCHM1560/7-1) to M.S., the ANR (Blanc – SVSE2 – 2011- Charmful) and CEA NanoProt3D programme to F.P. and R.D., the iRTeLIS CEA PhD programme to C.S. and the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement numbers 227764 (P-CUBE) and 283570 (BioStruct-X) to R.D. This work used the platforms of the Grenoble Instruct Center (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

## Author contributions

C.S., R.D., M.S. and F.P. designed the experiments. C.S., M.P.L., E.T., G.D., M.J. and R.D. performed the biochemical experiments. M.H.N., C.S. and R.D. performed the crystallographic experiments. C.S. and G.T. performed plant experiments. D.P. and M.S. performed the ChIP experiment. G.D., L.G., F.O. and F.P. performed the bioinformatics analyses. All authors interpreted the results. F.P. wrote the paper with the help of C.S., R.D., M.H.N. and M.S., and input from all authors.

## Additional information

**Accessions codes:** The GbLFY-N structure has been deposited in the RCSB with code 4UDE. The ChIP-Seq data have been submitted to the NCBI GEO database with accession number GSE64245.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Sayou, C. *et al.* A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nat. Commun.* **7**:11222 doi: 10.1038/ncomms11222 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>