**Title:**

# Chromatin architectural proteins regulate flowering time by precluding gene looping

Bo Zhao[1], Yanpeng Xi[1], Junghyun Kim[1], Sibum Sung[1*]

[1] Department of Molecular Biosciences, The University of Texas at Austin, TX 78712, USA

* Corresponding Author: Sibum Sung, sbsung@austin.utexas.edu

**Short Title:**

**Regulation of gene loop formation**

**Abstract**

Chromatin structure is critical for proper gene expression as well as many other cellular processes. In *Arabidopsis thaliana*, the major floral repressor *FLC* adopts a self-loop chromatin structure via bridging of its flanking regions. This local gene loop is necessary for active *FLC* expression. However, the molecular mechanism underlying the formation of this class of gene loops is not known. Here we report the characterization of a group of linker histone-like proteins, named the GH1-HMGA family in *Arabidopsis*, which act as chromatin architecture modulators. We demonstrate that these family members redundantly promote the floral transition through the repression of *FLC*. A genome-wide study revealed that this family preferentially binds to the 5' and 3' ends of gene bodies. The loss of this binding increases *FLC* expression by stabilizing the *FLC* 5' to 3' gene looping. Our study provides mechanistic insights into how a family of evolutionarily conserved proteins regulates the formation of local gene loops.

## Introduction

Eukaryotic DNA is spatially and functionally organized with its associated proteins in the form of chromatin. Nucleosomes are the fundamental subunit of chromatin. Each nucleosome is a complex of ~146 base pairs (bp) of DNA wrapped around a histone octamer. Nucleosomes play an essential role in the formation of higher-order chromatin structures and orchestrate transcriptional regulation (*1-3*). Research into the role of nucleosome structures, histone modifications, and nucleosome-binding proteins is beginning to reveal sophisticated mechanisms by which the fate of gene expression is determined in response to developmental and environmental stimuli in eukaryotes (*3, 4*). In addition, there are growing evidences that chromosome structure plays a vital role in controlling gene expression, although there is limited understanding of the nuclear proteins that contribute to structural interactions among nucleosomes (*5-7*).

Nucleosomes are connected through a segment of linker DNA, which often associates with other proteins like linker histone proteins (H1 or H5). Linker histones are the most divergent class of histones, but they all contain an evolutionarily conserved N-terminal globular domain (GH1 domain), which binds to the nucleosome dyad and interacts with the linker DNA (*8, 9*). In addition to the GH1 domain, linker histones generally contain a positively charged C-terminal domain that can interact with DNA. It has been known that linker histones can function as architectural proteins that induce chromatin conformation changes through cooperative binding of both the GH1 domain and the C-terminal domain to its target (*10, 11*).

The high mobility group (HMG) proteins are another set of chromatin architectural proteins. The HMG proteins were originally isolated via biochemical purification of chromatin proteins, and they are the most abundant non-histone proteins (*12, 13*). They bind to DNA and nucleosomes and generally act as architectural elements that modulate multiple DNA-dependent processes, including replication and transcription (*14, 15*). Higher eukaryotes contain three classical families of HMG proteins based on their DNA-binding domains: HMGA, HMGB, and HMGN (*16, 17*). The

HMGB family contains HMG-boxes, and the HMGN family contains nucleosome-binding domains (*16, 17*). The HMGA subfamily was grouped together because these proteins preferentially bind to the minor groove of AT-rich regions of DNA via several AT-hook motifs (*16, 17*). The AT-hook motif is a conserved DNA-binding motif commonly found in eukaryotes (*18*). HMGA proteins affect local chromatin structure in several ways, including bending, straightening, unwinding, and looping of substrate DNA (*19*), and they have been implicated in numerous DNA-based cellular processes.

In the flowering plant *Arabidopsis thaliana*, GH1 domain-containing proteins have been systematically annotated (*20*). Interestingly, a subgroup of plant GH1 domain-containing proteins possesses a C-terminal domain that has similarity to the mammalian HMGA proteins and thus was designated as the GH1-HMGA clade. Furthermore, similar arrangements of the GH1 domain and AT-hook motifs also exist in animals as well as in yeast, nematode, and insect species. However, plant GH1-HMGA proteins are restricted to angiosperms, implying that they are newly evolved in the plant lineage. Therefore, convergent evolution may have resulted in this group of H1 variants in diverse organisms. The biological function of plant GH1-HMGA proteins is not known, but they play fundamental roles in chromatin structure in other organisms (*20*).

In *Arabidopsis*, *FLOWERING LOCUS C* (*FLC*) has been an excellent model system to identify chromatin regulators, both protein and non-coding RNA components, and to unravel mechanistic details of epigenetic regulation (*21-23*). In addition, *FLC* chromatin contains loops that influence transcriptional activity (*22, 24, 25*). Although the presence of topologically associated domains (TADs) in *Arabidopsis* is not as clear as in mammals, plants also utilize a three-dimensional spatial organization of the genome, including gene loops, as means of architecturally regulating gene expression (*26-30*). Indeed, the formation of gene loops is prominent in *Arabidopsis* (*27*) and perhaps constitutes the basic unit of higher-order nucleosome structures that affect transcriptional activity. However, there is only a limited understanding of the mechanism underlying the formation of gene loops.

Here, we report that members of the *Arabidopsis* GH1-HMGA family are chromatin architectural factors and redundantly promote floral transition through the repression of *FLC* expression. We demonstrated that GH1-HMGA proteins directly repress *FLC* by preventing the formation of the 5' to 3' gene loop, which facilitates *FLC* transcriptional activation.

## Results

### Characterization of the roles of GH1-HMGA family genes in flowering in *Arabidopsis*

The mammalian HMGA family of proteins plays roles in various biological processes through influencing chromatin structure and transcription (*19*). Our phylogenetic analysis identified members of the GH1-HMGA clade in *Arabidopsis* as the closest homologs to mammalian HMGA proteins (Fig. 1A). Unlike the canonical mammalian HMGA, three members of *Arabidopsis* GH1-HMGA, including GH1-HMGA1 (HON4), GH1-HMGA2 (HON5), and GH1-HMGA3 contain a conserved GH1 domain at their N-terminus in addition to four to six AT-hook motifs (*20*). A distant member, which we arbitrarily called GH1-HMGA4, has an N-terminal GH1 domain but contains C-terminus without recognizable AT-hook, was also grouped with the GH1-HMGA cluster (*20*) (Fig. 1A and fig. S1, A and B). Both the GH1 domain and the AT-hook motifs can bind to nucleosomes, indicating that *Arabidopsis* GH1-HMGA proteins may function as architectural factors that influence chromatin structure (*19*).

Considering that the biological functions of plant GH1-HMGA proteins are mostly unknown, we isolated and analyzed corresponding T-DNA insertional loss-of-function mutants (fig. S2, A and B). Interestingly, we found *hon4*, *hon5,* and *gh1-hmga3* single mutant plants show a slight but reproducible late-flowering phenotype (Fig. 1, B and C, and fig. S2C). Subsequent genetic analyses showed that *hon4hon5* (*hon45*) double mutant plants display a more pronounced late-flowering phenotype, and the late-flowering was further enhanced by stepwise introgressions of *gh1-hmga3* and *gh1-hmga4* mutations; thus, the members of the GH1-HMGA

gene family redundantly control the floral transition in *Arabidopsis* (Fig. 1, B and C, and fig. S2C). A delay in the flowering of mutants was also observed in short days, showing that the late flowering is not due to a compromised photoperiodic response in the mutants (Fig. 1D).

**Transcriptome analysis of *hon45* mutants**

To explore the roles of the GH1-HMGA gene family in plant development, we performed RNA-Seq analysis using *hon45* mutants. The transcriptome analysis identified 525 differentially expressed genes (DEGs) in *hon45* mutants (Fig. 1E and Data Set S1). Gene ontology (GO) analysis showed no significant enrichment of GO terms for down-regulated genes. On the other hand, several significant GO terms were identified for up-regulated genes, including biological pathways involved in photosynthesis, disease, and responses to environmental stimuli including light and temperature, implying that GH1-HMGA family of proteins may function to repress these classes of genes (Fig. 1F). Next, we sought genes implicated in floral transition that may contribute to the late-flowering phenotype of the *hon45* mutant. Notably, transcripts of the major floral repressor *FLC*, increase significantly in the *hon45* mutant, and this observation was further confirmed by quantitative RT-PCR assay (Fig. 1G and Data Set S1). More importantly, *FLC* mRNA levels in higher-order of mutants display a positive correlation with their flowering times (Fig. 1, C and G), supporting that the members of GH1-HMGA clade redundantly repress *FLC* to promote flowering.

**Functional characterization of HON4 and HON5**

Our results indicated that members of the GH1-HMGA gene family positively promote floral transition through the repression of *FLC*. Accordingly, the *hon45* mutant plants were transformed with *HON4* or *HON5* genomic sequences fused to the Myc-epitope for molecular complementation. We found that the *hon45* late-flowering phenotype could be complemented by either *gHON4-Myc* or *gHON5-Myc* (Fig. 2A, and fig. S3A). A representative complemented line for each transgene was selected and further verified by detecting the Myc-fused proteins (Fig. 2A,

and fig. S3B). Correlated with the flowering phenotype, the elevated level of *FLC* mRNA expression in *hon45* mutants was restored in the complementation lines to a comparable level to that in the wild type (Fig. 2C). In addition, the expression level of *FT*, a floral integrator downstream of FLC, is also recovered in the complementation lines (Fig. 2D). By analyzing transgenic plants in which β-glucuronidase (GUS) gene is fused in frame with *HON4* or *HON5* genomic copies, we found *HON4* and *HON5* has similar expression pattern (Fig. 2E and fig.S3E); both HON4 and HON5 are expressed in the shoot apical meristem, supporting their redundant role in promoting floral meristem formation (Fig. 2E). We also detected strong GUS-staining in root tissues as well as in the vasculature of cotyledons and expanded leaves (Fig. 2E and fig.S3E). Similar expression patterns of HON4 and HON5 were observed from GFP-tagged transgene lines, with a strong signal at the shoot apex and the GFP signal is also observed throughout cotyledons (fig. S3F). Consistent with their potential role as chromatin architectural proteins, HON4-GFP and HON5-GFP exclusively localize in the nucleus of plant cells (Fig. 2F).

**GH1-HMGA gene family act through *FLC* to regulate flowering**

Our gene expression analysis suggests that *FLC* may be a target of GH1-HMGA family of proteins in regulating flowering time (Fig. 1G, and Fig. 2C). We addressed their genetic relationship by introducing the null *flc-3* mutation (*31*) into *hon45* and *gh1-hmga quadruple* (*honq*) mutants. Genetic assays revealed that the GH1-HMGA gene family promotes flowering mainly through *FLC*, as the *flc-3* mutation could mostly reverse the late-flowering of both *hon45* and *honq* mutants back to that of wild type (Fig. 3, A and B, and fig. S4A). FLC represses flowering by suppressing the transcription of floral integrator genes, including *FT* (*32*). In agreement with the flowering trait, the dramatic reduction of *FT* mRNA in *honq* mutant was also restored to the level similar to the wild-type by *flc-3* mutation (Fig. 3C), demonstrating that the members of the GH1-HMGA gene family modulate *FT* expression through FLC.

To explore the tissues in which HON4 and HON5 regulate the *FLC* transcription, we crossed the transgenic line carrying the GUS-fused with the entire *FLC* genomic region (*FLC-GUS*) (*33*) into *hon45* mutant plant (fig. S4B). Consistent with previous reports (*33, 34*), we detected FLC-GUS signals throughout vascular tissues of young seedlings (fig. S4C). Compared to the wild-type Col-0 background, the overall patterns of FLC-GUS staining are not altered in *hon45* mutants; however, much higher level of FLC-GUS signal is detected in *hon45* mutants (fig. S4C). Moreover, the *FLC* spatial expression patterns overlap with *HON4* and *HON5* expression domains (Fig. 2E and fig. S4C), implying that the GH1-HMGA family directly represses *FLC* transcription in the tissues where *FLC* is actively expressed.

Considering that the GH1-HMGA family proteins are novel regulators of *FLC* transcription, we tested their genetic relationship with several other *FLC* regulators. Mutations in *HON4* and *HON5* show additive effects on the late-flowering phenotypes of autonomous pathway mutants, *fca-9*, *fve-4,* and *fld-3* (Fig. 3D), implying that the GH1-HMGA family functions independently of FCA, FVE, and FLD in the repression of *FLC*. We also tested the *hon45* mutant for its vernalization response by the introgression of *hon45* mutants into the winter-annual *FRI*-Col genetic background (*35*). Additive effects were observed in terms of both flowering time and *FLC* expression with vernalization treatment, indicating that FRI and the GH1-HMGA family act in parallel to regulate *FLC* (Fig. 3E). Interestingly, *hon45* mutants show a slow reduction of *FLC* and later-flowering upon vernalization, although gradual repression of *FLC* by cold is still observed (Fig. 3E).

Tri-methylation at Histone H3 Lys 27 (H3K27me3) and tri-methylation at Histone H3 Lys 36 (H3K36me3) are two epigenetic markers that antagonize each other to fine-tune *FLC* expression (*36*). Although clear de-repression of *FLC* in *hon45* and higher-order of mutants was observed, no apparent difference in either H2K27me3 or H3K36me3 at *FLC* is observed in *hon45*

mutants (fig. S4, D and E). Our above results collectively suggested that the GH1-HMGA family proteins may regulate *FLC* through a previously unknown mechanism.

**GH1-HMGA family members bind to *FLC* chromatin**

Next, we utilized the complemented lines harboring the Myc-epitope to test whether HON4 and HON5 directly associate with *FLC* chromatin by chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR). We detected significant enrichment of both HON4-Myc and HON5-Myc to the same region of ~600 bp upstream of *FLC* transcription start site, corresponding to the canonical promoter of *FLC* (Fig. 4A), demonstrating that members of the GH1-HMGA family directly regulate *FLC* expression through physical association with *FLC* chromatin.

To better understand the molecular function of the GH1-HMGA family, we identified HON5 targets at the genome-wide level by employing ChIP-seq (Supplementary Table. S2). More than 21,000 HON5 binding peaks were determined (Data Set S2), including the one at the *FLC* promoter that is identical to the region detected by ChIP-qPCR (Fig. 4, A and C). Interestingly, HON5 shows distinct binding patterns that peak at 5' and 3' flanking regions of protein-coding genes (Fig. 4C). The majority (~76%) of the HON5 binding sites are clustered within 3 kb upstream of transcription start site (TSS) or 1 kb downstream of the transcription end site (TES), with relatively higher occupancy towards the 5' end of genes (Fig. 4, B and D). However, few binding signals were observed across the gene body region. Therefore, genome-wide distribution patterns of HON5 indicate that it generally functions at the flanking regions of protein-coding genes (Fig. 4B).

Motif analysis of HON5 binding sites identified DNA motifs primarily composed of adenines and thymines, known as the binding motifs of AT-rich interaction domain (ARID)-containing family proteins (*37*) (Fig. 4E). Six such AT-rich motifs are clustered within 150 bp of HON5-enriched regions at the *FLC* promoter (Fig. 4, A and C, and fig. S5A), supporting their importance in mediating the binding of HON5 to *FLC* chromatin. Therefore, plant GH1-HMGA family proteins

show similar functional property to mammalian HMGAs in terms of DNA binding preference toward AT-rich regions of DNA (*15, 19, 38*). Given that GC- and AT-rich chromatin may differ in conformation and modification, we tested whether HON5 enrichment is associated with certain histone modifications. However, we did not find any genome-wide correlation among tested histone modifications, including H3K4me3, H3K27me3, and H3K36me3 (fig. S6). This is consistent with our finding that no obvious change in these modifications was observed at *FLC* in *hon45* mutants (fig. S4 D and E). In addition, there is no correlation between HON5 occupancy and the level of gene expression or gene size (fig. S6), implying that the GH1-HMGA family proteins may function in a previously unknown manner.

**The GH1-HMGA family proteins preclude *FLC* gene looping**

The distinct patterns of HON5 occupancy revealed by ChIP-Seq analysis prompted us to check whether it also binds to the 3' end of *FLC* (Fig. 4B). Indeed, a sharp HON5 binding signal was observed at the 3' region of the *FLC* locus (Fig. 4C and fig. S5B), which corresponds to the promoter of antisense long non-coding RNA, *COOLAIR* (*39, 40*). Given that *COOLAIR* is known to be involved in the downregulation of *FLC* transcription, we examined whether HON5 and its related members regulate *FLC* by altering *COOLAIR* transcription. However, there is no significant change in the level of both distal and proximal *COOLAIR* transcripts in *hon45*-FRI compare to the wild-type FRI-containing line (fig. S7, A and B).

Two competing chromatin loops have been identified at the *FLC* locus (*22, 24*). A gene loop between the 5' and 3' of *FLC* flanking regions is known to be necessary for the active *FLC* transcription (*22, 24, 25, 28*). However, the regulatory factors involved in the formation of the *FLC* gene loop are not known. Because HON5 binds to the same regions where the 5' to 3' *FLC* gene loop forms, we investigated whether the GH1-HMGA gene family plays a role in the formation of the *FLC* gene loop. By using chromosome conformation capture (3C) followed by quantitative PCR (*22, 25, 41*), we found that the frequency of the 5' to 3' gene looping at *FLC* dramatically

increased (> 4-fold) in the *honq* mutant compared to the wild-type Col-0 (Fig. 5A). Moreover, the enhanced *FLC* gene looping in the *honq* mutant is restored to near the wild-type level in the complemented line (Fig. 5A, and fig. S8, A and B). Similarly, we observed that the *FLC* gene looping significantly increased in *honq*-FRI compared to the wild-type *FRI*-Col (Fig. 5B). The frequency of *FLC* gene looping is more robust in *honq*-Col compared to *FRI*-Col (Fig. 5C), despite that the level of *FLC* transcription in *FRI*-Col is 4 times higher than that in *honq* mutants in Col-0 background (fig. S8B). Therefore, the GH1-HMGA family proteins contribute to the repression of *FLC* by preventing the 5' to 3' gene looping at *FLC*, independent of the FRI complex.

It has been proposed that the *FLC* 5' to 3' looping may create a favorable condition for transcription by facilitating the recycling of the RNA polymerase II (RNA Pol II) at *FLC* (*24, 25*). This prompted us to examine the level of transcription-initiation form of RNA Pol II, Ser 5-phosphorylated Pol II (Ser5-P Pol II) (*42*) at *FLC* locus. Consistent with the change in the level of *FLC* gene looping and transcription, we found that the level of Ser5-P Pol II at *FLC* increases in *hon45* mutant compared to the wild-type, and this accumulation is restored in the complementation line (Fig. 5D). Moreover, we detected a relatively higher level of Ser5-P Pol II at the region corresponding to the HON4 and HON5 binding sites (Fig. 4A and Fig. 5D). Therefore, our results demonstrated that the binding of GH1-HMGA family proteins to *FLC* flanking regions disrupts the formation of gene loop and thus alters local chromatin structures necessary for effective *FLC* transcription (Fig. 5E).

**Discussion**

Here, we characterized the GH1-HMGA gene family for their roles in floral transition in *Arabidopsis*. We showed that the late-flowering observed in higher-order of mutants is due to the elevated level of the floral repressor *FLC*. By a classical definition (*43, 44*), the GH1-HMGA gene family belongs to the autonomous-pathway genes which regulate *FLC*. *hon45* and higher-order

mutants among the members of the family still retain the photoperiod response, and their late-flowering phenotypes are suppressed by *flc* mutation, which is a classical definition of autonomous-pathway, flowering-time mutants (*43, 44*) (Fig. 1, C and D). Moreover, additive effects of GH1-HMGA family mutations were observed in all tested mutant backgrounds (Fig. 3, D to F), suggesting that this group of proteins regulate *FLC* through a novel molecular mechanism.

Our analysis revealed the *Arabidopsis* GH1-HMGA family members are the closest homologs to mammalian HMGA proteins (Fig. 1a). However, the *Arabidopsis* GH1-HMGA family proteins are unique in that they contain the GH1 domain, which is the signature motif of H1 linker histone proteins (*20*). A recent study showed that mammalian HMGA proteins display widespread bindings with only a preference to AT-rich regions (*45*). Our ChIP-seq analysis also showed that HON5 has pervasive genome-wide occupancy with over 21,000 peaks (Data Set S2), and GH1-HMGA proteins preferentially bind to AT-rich regions as well (Fig. 4E, and fig. 5, A and B). 5,611 genes have at least one nearby HON5 binding signal, and 108 genes with the HON5 peak are differentially expressed in *hon45* mutants (Fig. 1E, and Data Set S3). Relatively minor changes in transcriptome were also reported in mouse embryonic stem cells (*45*), suggesting that only a limited number of loci are sensitive to the loss of this class of chromatin architectural proteins. A previous study reported that *hon4* mutants exhibited multiple growth defects, including short roots, small and sharp leaves, short inflorescences, and total sterility (*46*). However, we did not observe any developmental abnormality in single mutants or in any higher-order of mutants (Fig. 1b), except for the late-flowering due to the de-repression of *FLC*.

Although the GH1-HMGA family of proteins share some similarities with known HMGA and H1-linker proteins, genome-wide occupancy patterns of the GH1-HMGA family of proteins are unique. Their occupancies peak at both 5’ and 3’ end of protein-coding genes (Fig. 4B), and the depletion of the GH1-HMGA family of proteins resulted in the enhanced formation of a gene loop at *FLC* locus. In *Arabidopsis*, gene loops have been systematically identified and, the packing of its genome is predicted to adopt units of gene bodies (*27, 28*). In a previous study, 1,792 genes

were found to contain self-loops between the 5' and 3' portion of their transcribed region (*27, 28*). It should also be noted the formation of a gene loop could be inducible in response to stimuli and also be expected to be tissue-specific (*22, 24, 27-29, 47*). Therefore, the number of genes with self-looping is likely to be underestimated. Besides the *FLC* locus, whether GH1-HMGA family proteins control gene loop formation at other loci remains to be determined.

One of *FLC* transcriptional activators, the FRI complex, has been shown to be necessary for *FLC* 5' to 3' gene loop formation (*24*). Our data revealed that the FRI complex is not required for GH1-HMGA family proteins to govern gene looping at *FLC* (Fig. 5, A to C). Although all examined H3K4me3, H3K27me3, and H3K36me3 histone modifications unlikely contribute to the regulatory role of GH1-HMGA on *FLC* gene looping (fig. S4, D and E and fig. S6), a recent study showed that the RNA Pol II complex plays an active role in gene loop formation (*48*). Our ChIP-qPCR data show that the enrichment of transcription-initiation form of RNA Pol II at *FLC* promoter occurs in a HON4 and HON5 dependent manner (Fig. 5D). The binding of GH1-HMGA family proteins to *FLC* promoter and region downstream of the terminator appears to create chromatin structures that adversely affect the recycling of RNA Pol II, and thus prevents 5' to 3' gene looping (Fig. 5E).

Although the presence of gene loop has been reported in many species (*6, 22, 27, 49*), the regulators that affect the formation of gene loop is not well understood and may be divergent among species (*49, 50*). Our work identified the GH1-HMGA family proteins as regulators of the formation of gene loop at *FLC*. Further characterization of this group of proteins will shed light on the molecular mechanisms underlying gene loop formation and their function in various biological processes.

## Materials and Methods

### Plant materials and growth conditions

The *hon4* (SALK_071403), *hon5* (SALK_116292), *gh1-hgma3* (SALK_078336), *gh1-hmga4* (CS824818) mutants in Columbia (Col-0) background were obtained from Arabidopsis Biological Resource Center (ABRC). Mutants were cross with the *FRI*-Col to generate lines in *FRI* background. Primers for T-DNA insertion genotyping are listed in Supplementary Table S1. Sterilized seeds were sown on agar plates, stratified at 4°C for 3 days, then moved to the growth chamber with long-day condition (16 hours light, 8 hours dark) at 22°C for 7 days. After that, plants were transplanted to soil and transferred to either long-day or short-day (8 hours light, 16 hours dark) growth chambers for further assay. Flowering time was measured by counting the total number of leaves (rosette and cauline leaves) at the bolting stage. For the vernalization treatment, seeds were germinated on agar plates for 10 days and vernalized at 4°C under short-day condition. After the vernalization treatment, plants were transplanted to soil and transferred to growth chambers (22°C) under long-day condition for flowering time test or harvested for RNA isolation.

### Transgenic plants

Genomic sequences of *HON4* and *HON5* were amplified by PCR and cloned into pENTR, then transferred into pGWB16, pGWB203, and pGWB604 binary vectors using Gateway System (Invitrogen). Sequences were confirmed by Sanger sequencing and used for complementation of the mutant lines. All binary vectors were transformed into *Agrobacterium tumefaciens* GV3101 strain. Plants of *hon4hon5* double mutant were transformed with a flower dip method. Homozygous transgenic plants harboring single T-DNA insertion were selected on antibiotic plates. Primers used for gene cloning were listed in Supplementary Table S1.

### Phylogenetic tree analysis

Protein sequences were obtained from The Arabidopsis Information Resource (TAIR) and UniProt databases. MEGA 7 software was used to construct phylogenetic trees with maximum-likelihood estimation and 1000 bootstrap. The tree was rooted using human and mouse HMGA proteins as an outgroup for both GH1-HMGA and histone H1 in Arabidopsis.

**RNA expression analysis**

Total RNA was extracted from whole seedlings 10 days after germination unless otherwise specified using TRIzol (Invitrogen). Extracted RNA was treated with DNase I (Promega) for 30 minutes at 37°C to remove genomic DNA. Purified RNA was quantified on NanoDrop (Thermo Scientific) and 1 μg RNA was used for first-strand cDNA synthesis using oligo (dT) primers. Synthesized cDNA products were diluted three-fold with water and then used for real-time qRT-PCR analyses with Maxima SYBR green master mix (Thermo Scientific) on a ViiA 7 real-time system (Life Technologies). Relative gene expressions were determined by normalizing to the levels of *PP2A*. Primer sequences for qRT-PCR are listed in Supplementary Table S1.

**Transcriptomic analysis**

Whole seedlings grew on half-strength MS medium under short-day condition were collected at zeitgeber time (ZT) 6. Total RNAs were extracted using TRIzol (Invitrogen) and treated with DNase I (Promega) to eliminate traces of genomic DNA. Sequencing libraries were prepared with 500 ng total RNA following NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7420). Libraries were assessed on a bioanalyzer (Agilent High Sensitivity DNA Assay) and sequenced on Illumina NextSeq 500 platform. RNA-seq clean reads were aligned to TAIR10 genome release using HISAT2 with default parameters. Gene expression was quantified as counts per million reads mapped (CPM). Differentially expressed genes were determined with edgeR over two biological replicates. Genes with more than 1.5-fold change relative to Col-0 and FDR < 0.05 were

considered as DEGs. Gene ontology term enrichment was performed over the sets of DEGs with the online tools (http://geneontology.org).

**Chromatin immunoprecipitation (ChIP)**

About two grams of 10-days-old seedlings were harvested and cross-linked in 1% formaldehyde solution under a vacuum for 25 minutes. Cross-linking was stopped by adding 0.125 M glycine and vacuumed for 5 minutes. Cross-linked seedlings were rinsed in 10 mM HEPES buffer three times and dry with paper towels. Samples were ground into fine powder in liquid nitrogen. ChIP assays were performed following the Abcam ChIP protocol (https://www.abcam.com/protocols) with minor adjustments. Immunoprecipitations were performed by using c-Myc antibody (9E10, Santa Cruz Biotechnology) combined with protein G magnetic beads (Thermo Scientific). Input DNA and immunoprecipitated DNA samples were purified by PCR purification kits from Qiagen. Eluted DNA samples were used for either ChIP-qPCR or sequencing.

**ChIP-sequencing analysis**

ChIP assays were conducted by using both Col-0 and *HON5-Myc* transgenic plants. One immunoprecipitated DNA sample from Col-0 (Col-0-IP) and immunoprecipitated DNA from two replicates of *HON5-Myc* (HON5-IP) as well as pooled input DNA (HON5-Input) were selected for sequencing. ChIP-seq libraries were prepared using the NEBNext ChIP-Seq library prep kit and sequenced on an Illumina Nextseq 500 platform. Sequencing reads were mapped to the *Arabidopsis* reference genome (TAIR10) with Bowtie2. Mapped reads were normalized using Deeptools and visualized using IGV. As both Col-0-IP and HON5-Input track show background signals, we selected Col-0-IP as the control for the following assays. To check the enrichment of HON5 relative to gene start and end positions, we calculated the scores of HON5 per gene using deepTools. Each gene was defined as the interval from TSS to TES plus 3 kb upstream and 1kb downstream. In total, 33602 such regions annotated from TAIR10 were analyzed. All the regions

were then scaled and stacked, and the average score was plotted to show the relative enrichment of HON5 over genes. Genome-wide HON5 peak distribution was analyzed by categorizing Arabidopsis genome into non-overlapping elements including TSS, TES, 5 'UTR, 3' UTR, Exon, Intron, Intergenic, and TSS & TES. The TSS &TES is where two genes are closely located, and thus the TSS of one gene overlaped with the TES of another gene. The percentage of HON5 peaks that fell into each category was calculated and showed in the pie chart. Motif analysis was carried out by extracting the +/- 300bp sequences surrounding HON5 peak summits and submitting these regions to MEME-ChIP motif discovery module against DAP motifs (*51*). The estimated statistical significance (E-value) and sequence logo was generated for each motif.

To check the correlation between HON5 enrichment and histone modifications of its neighboring gene (*52*), we extracted the pairs of HON5 peak and its closest gene using bedtools. In total, 21164 HON5-gene unique pairs were obtained (median distance 223 bp; mean distance 526 bp). In cases where the downstream and upstream genes showed the same distance to the HON5 peak, we assigned two pairs to include both genes. The level of HON5 was calculated by averaging the coverage within each peak. The levels of H3K27me3, H3K36me3, and log-transformed transcription were calculated by averaging the coverage within each gene. Unlike those modifications that spread across the gene body, H3K4me3 is largely concentrated in 5' end regardless of gene length, therefore we calculated H3K4me3 levels by extracting the max coverage value within each gene. Pairwise correlation analysis was carried out using R stats package. The corresponding Pearson correlation coefficients were calculated and displayed together with the scatter plot and linear trendline for each pair.

**Chromosome Conformation Capture (3C)**

3C assays were conducted as previously described with minor modifications (*41*). Nuclei were isolated from 1% formaldehyde cross-linked 10-day-old Arabidopsis seedlings and treated with

0.3% SDS at 65°C for 40 minutes followed by 30 min at 37°C. SDS was sequestered with 1% Triton X-100 for 60 min at 37°C. Chromatin was digested overnight by 400U DpnII restriction enzyme (NEB) at 37°C. Restriction enzymes were inactivated by the addition of 1.6% SDS and incubate at 65°C for 20 minutes, and then, 2% Triton X-100 was added to sequester SDS. Ligations were performed for 5 hours at 16°C using 200 U of T4 DNA ligase (Invitrogen) followed by 2 hours at room temperature. Reverse cross-linking was performed at 65°C for 6 hours. After Proteinase K (NEB) treatment, ligated DNA was purified by phenol/chloroform/isoamyl-alcohol (25:24:1) extraction and ethanol precipitation. Quantitative PCR was performed to calculate the relative interaction frequencies between the two regions. An *FLC* region without *Dpn*II digestion was amplified as a loading control to normalize the DNA concentrations of different samples. The primer efficiencies were corrected using a control template that contains equal amounts of all possible ligation products from a *Dpn*I digested plasmid harboring 11 kilobases of assayed *FLC* genomic region. Primers used for 3C-qPCR are listed in Supplementary Table S1.

**Histochemical β-glucuronidase staining**

Plant materials were submerged in 0.5 mg/mL X-Gluc solution (0.1 M Monosodium phosphate, pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 0.5 mM potassium ferrocyanide, 0.5 mM potassium ferricyanide), vacuumed for five minutes and kept at 37°C. Subsequent materials were decolorized in 70% ethanol and imaged with a stereo microscope.

**Statistical analysis**

Two-tailed Student's *t*-test and one-way ANOVA were conducted using Excel.

**Accession Numbers**

Arabidopsis Genome Initiative gene identifiers are as follows: *FLC* (AT5G10140), *FT* (AT1G65480), *HON4* (AT3G18035), *HON5* (AT1G48620), *GH1-HMGA3* (AT1G14900), *GH1-*

*HMGA4* (AT5G08780), *PP2A* (AT1G13320), *ACT2* (AT3G18780), *ACT7* (AT5G09810), *H1.1* (AT1G06760), *H1.2* (AT2G30620), *H1.3* (AT2G18050), *GH1-Myb1* (AT1G49950), *GH1-Myb2* (AT5G67580), *GH1-Myb3* (AT3G49850), *GH1-Myb4* (AT1G17520), *GH1-Myb5* (AT1G72740), *GH1-Myb6* (AT1G54230), *GH1-Myb7* (AT1G54240), *GH1-Myb8* (AT1G54260).

**Acknowledgments**

The authors acknowledge the Texas Advanced Computing Center (TACC; http://www.tacc.utexas.edu) at The University of Texas at Austin for providing high performance computing resources that have contributed to the research results reported within this paper. We wish to thank Richard Amasino for comments on the manuscript.

**Funding**: This work was supported by NIH R01GM100108 and NSF IOS 1656764 to S. S.

**Author contributions:**

B. Z., Y. P., J. K., and S. S. conceived of and implemented the method, performed the experiments and data analysis. B. Z. and S. S. drafted the manuscript. S. S. advised on the design and implementation and interpretation of results and edited the manuscript. All authors read and approved the final manuscript.

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

**Data and Materials Availability**

The data supporting the findings of this study are available within the paper and its supplementary information files. A reporting summary for this article is available as a supplementary information file. ChIP-seq and RNA-seq data, including raw reads and FPKM expression tables, were

deposited in the NCBI Gene Expression Omnibus (GEO) database under accession GSE163850. Specific materials generated during this study are available upon request.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available

## Supplementary Information

**Data Set S1**: List of differentially expressed genes in *hon45* mutant

**Data Set S2**: List of HON5 ChIP-seq peaks

**Data Set S3**: List of genes with HON5 binding signal

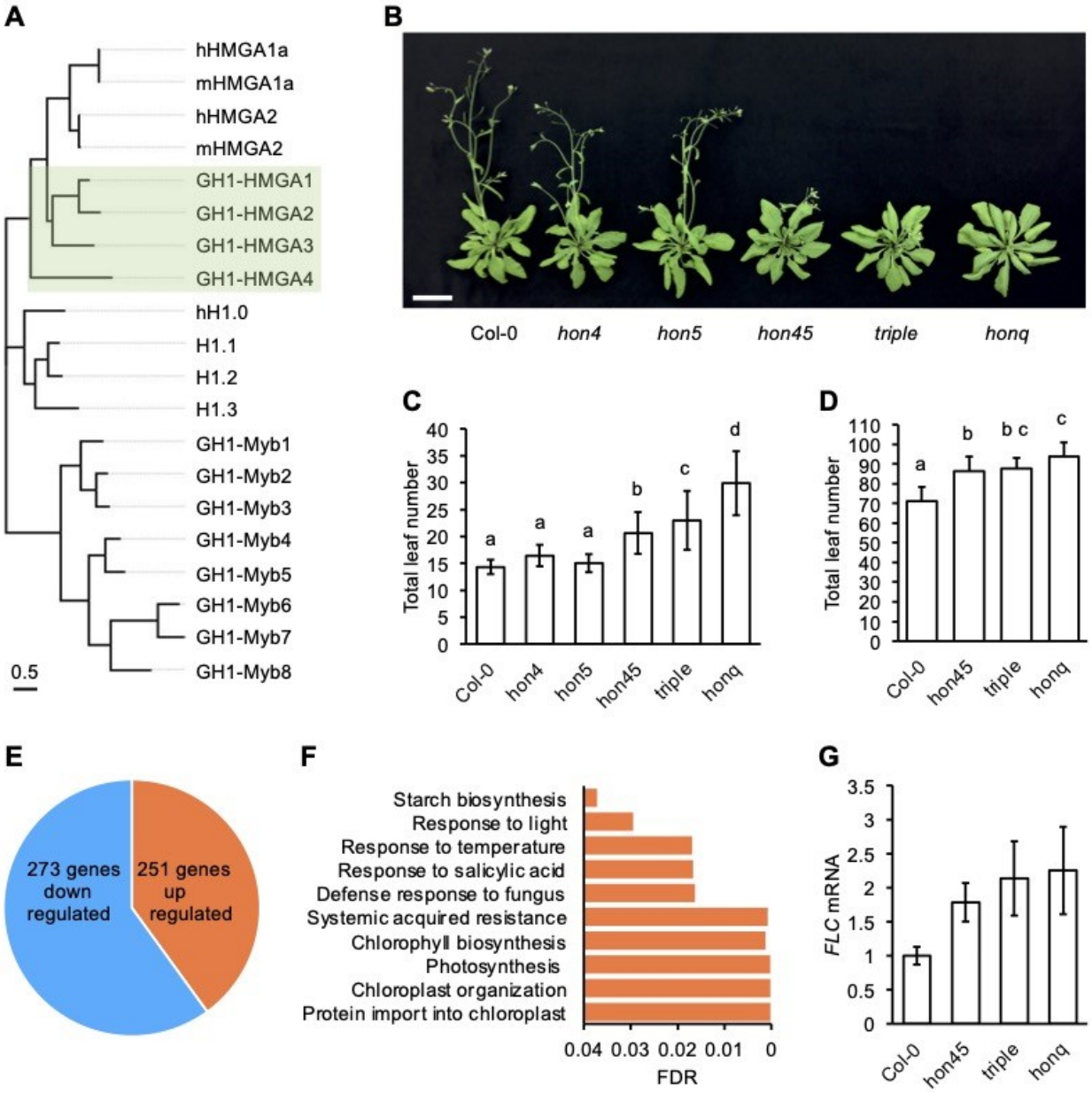



**Fig. 1. Characterization of the HMGA family of proteins in *Arabidopsis*.** (**A**) Phylogenetic tree of *Arabidopsis* GH1 domain-containing proteins. Human and mouse HMGA variants were used as outgroups. (**B**) Morphology of representative five-week-old plants grown under long-day (LD) at 22 °C. Scale bar, 5 cm. (**C**) Flowering times of plants measured under LD. Error bars: ± s.d. ($n \geq 30$); Significantly distinct groups were determined by one-way ANOVA followed by Tukey HSD test for multiple comparisons (letters indicate statistically distinct groups; $P < 0.05$). (**D**) Flowering

time of plants grown in short-day (SD) condition at 22 °C. Error bars: ± s.d. ($n \geq 15$); Significantly distinct groups were determined by one-way ANOVA followed by Tukey HSD test for multiple comparisons (letters indicate statistically distinct groups; $P < 0.05$). (**E**) Differentially expressed genes (DEG) identified by RNA-seq of two biological replicates. Genes with more than 1.5-fold change were defined as differentially expressed genes, FDR<0.05. (**F**) Enriched GO biological pathways of the up-regulated differentially expressed genes in *hon45* mutant compare to Col-0. (**G**) qRT-PCR quantification of *FLC* mRNA level in 10-day-old seedlings grown under LD condition at 22 °C. Error bars: ± s.d. ($n$=3).

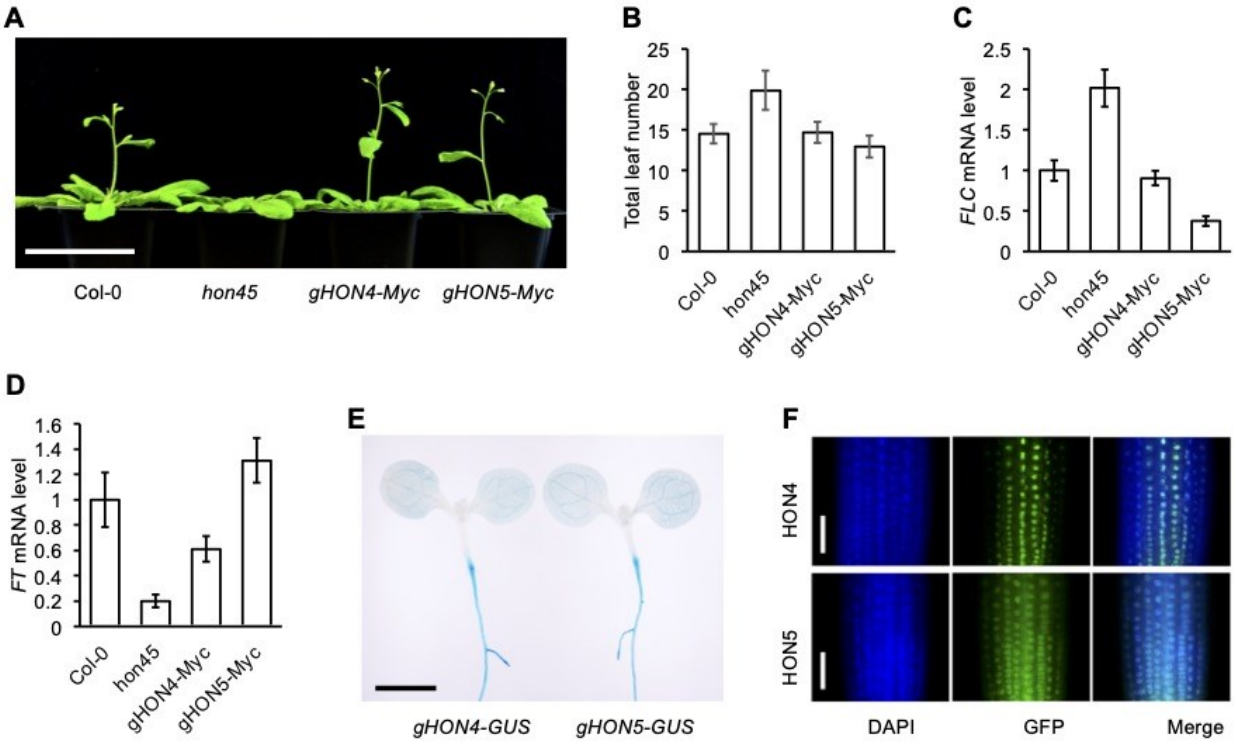



**Fig. 2. Functional characterization of HON4 and HON5.** (**A**) Representative 35-day-old plants grown under LD at 22 °C showing the molecular complementation by Myc-tagged transgenes. Scale bar, 5 cm. (**B**) Flowering times of representative complementation lines grown under LD at 22 °C. Error bars: ± s.d. ($n \geq 30$). (**C**) Relative expression levels of *FLC* mRNA in 10-day-old seedlings grown under LD at 22 °C. Bars indicate s.d. of three biological replicates. (**D**) The relative expression level of *FT* mRNA in 10-day-old seedlings grown under LD at 22 °C condition. Error bars: ± s.d. ($n = 3$). (**E**) GUS staining of 7-day-old seedlings that carry a transgene to express HON4-GUS and HON5-GUS fusion proteins in *hon45* mutant background. Scale bar, 2 mm. (**F**) HON4-GFP and HON5-GFP fusion proteins localize in the nucleus of root cells. Same subcellular localizations were observed from other tissues. Scale bar, 200 μm.

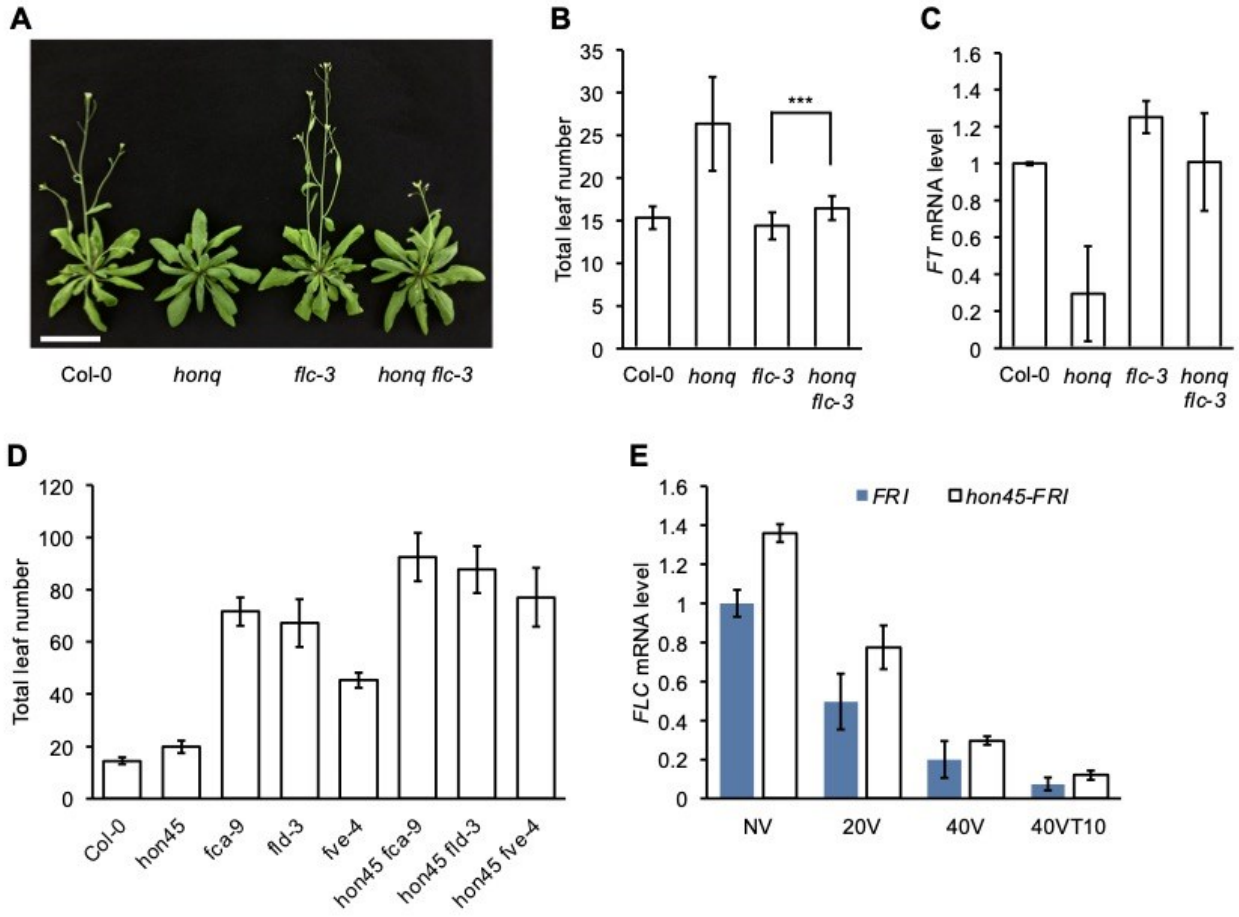
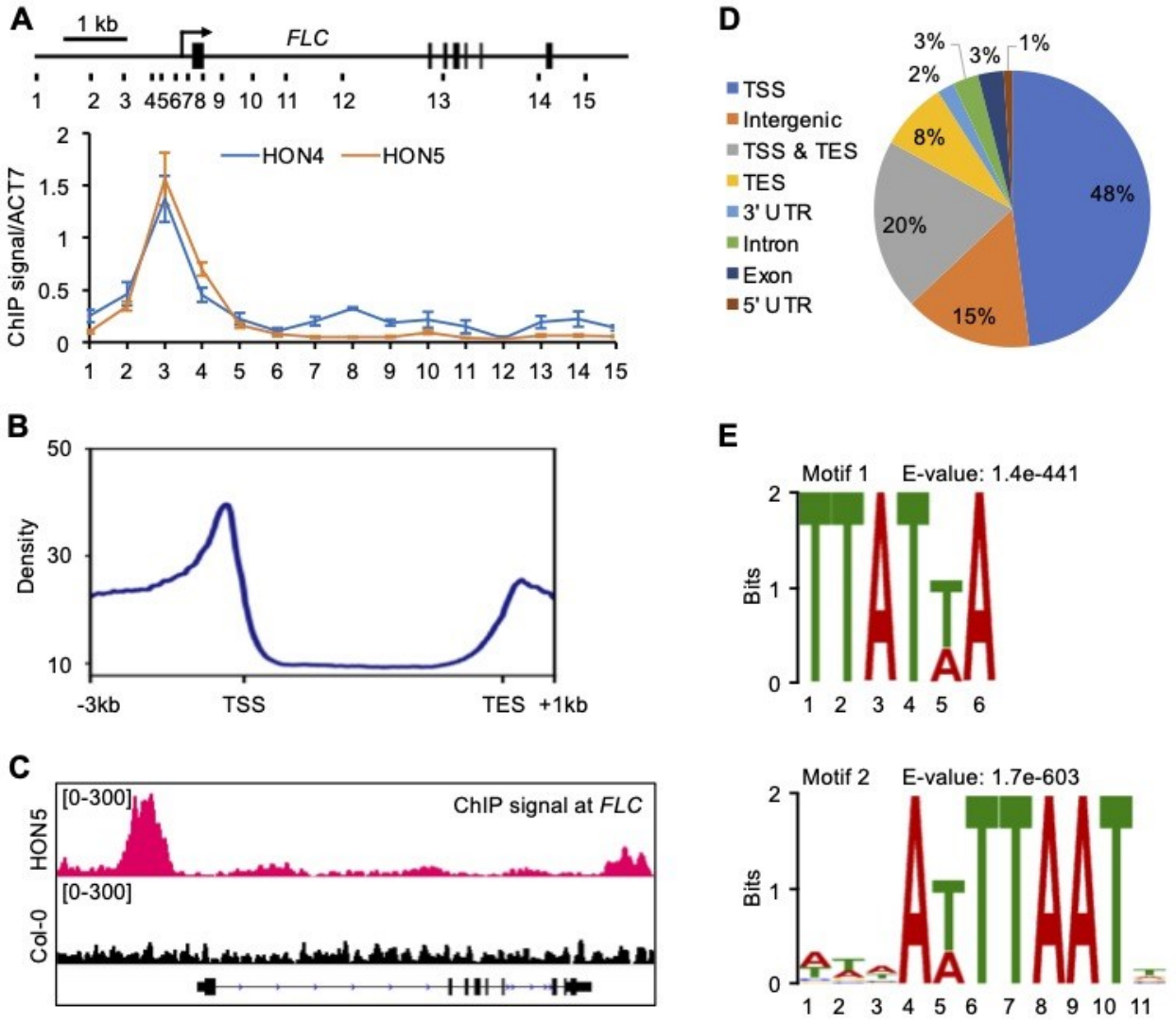



**Fig. 3. *Arabidopsis* GH1-HMGA family genetically acts through *FLC* to regulate flowering.** (**A**) Introduction of *flc-3* mutation rescues the late-flowering phenotype of *honq*. Scale bar, 5 cm. (**B**) Total leaf number of plants grown under LD at 22 °C. Error bars: ± s.d. ($n \geq 30$); two-tailed Student's *t*-test, ***P < 0.001. (**C**) Expression changes of *FT* show that *FLC* is required for the GH1-HMGA family proteins to promote the floral transition. (**D**) Genetic analysis of *hon45* mutant with autonomous pathway mutants. Error bars: ± s.d. ($n \geq 20$). (**E**) Changes in *FLC* expression during vernalization treatment.




488

489 **Fig. 4. Genome-wide study of HON5 occupancy.** ((**A**) The upper part is a diagram of *FLC* gene
490 structure with numbers marking the positions of PCR amplicons used for ChIP-qPCR. The lower
491 panel is the ChIP-qPCR results performed across the *FLC* locus. Error bars: ± s.d. (*n*=3). Primers
492 used for qPCR are listed in Supplementary Table S1. (**B**) The genome-wide average profile of
493 HON5-Myc ChIP-seq signals. (**C**) IGV browser track of HON5 binding at *FLC* locus. Track in red
494 color shows HON5-Myc ChIP signal; Track in black color shows background from Col-0
495 immunoprecipitated with anti-Myc antibody. (**D**) Distribution of annotated HON5-Myc ChIP-seq
496 peaks within defined genomic regions. (**E**) The two most significant DNA motifs associated with

HON5 binding are AT-rich sequences. Those two motifs are enriched in the HON5 binding region at *FLC* and the positions of corresponding motifs were highlighted in fig. S5.

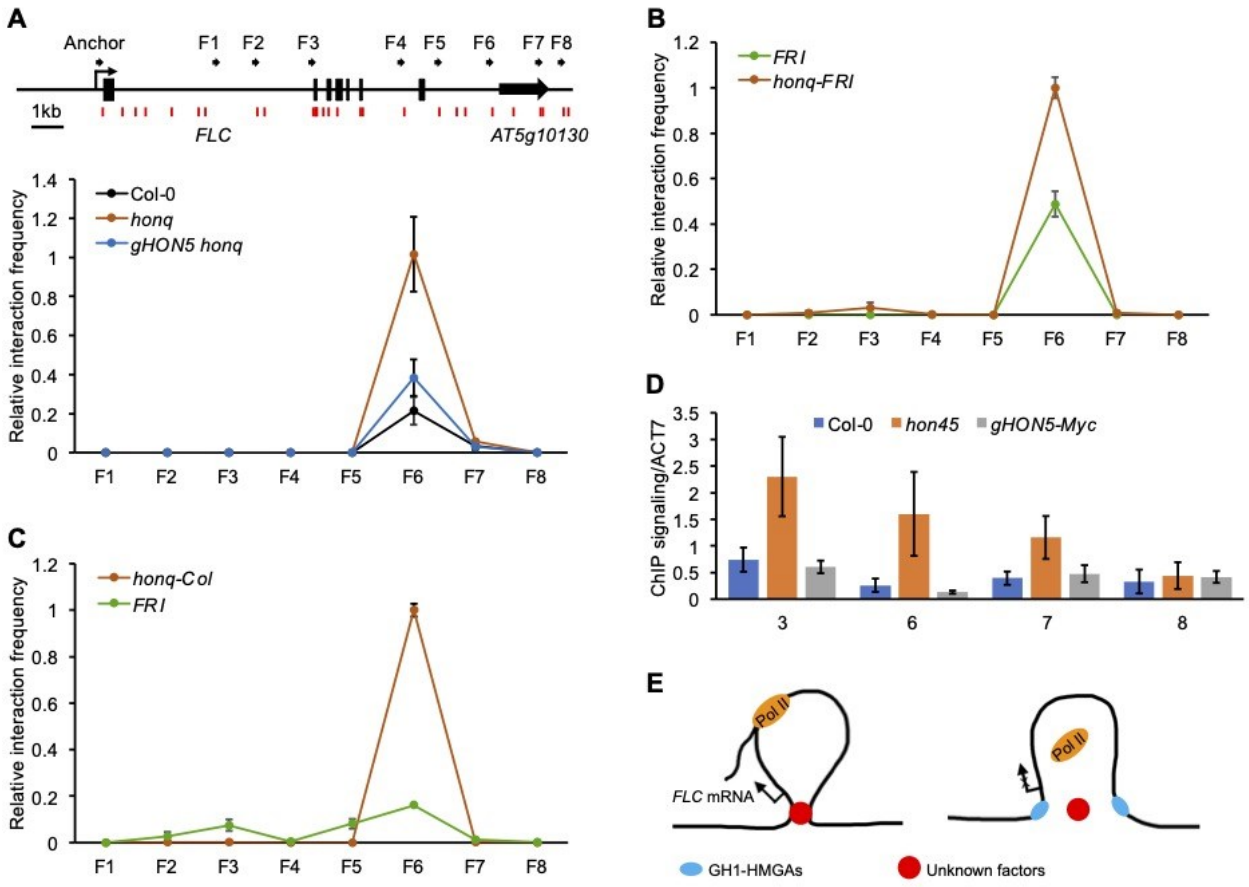



**Fig. 5. GH1-HMGA family of proteins preclude *FLC* gene loop formation.** (**A**) The upper panel shows the relative locations at *FLC* in 3C-qPCR experiments. *Dpn*II restriction sites are indicated with vertical red lines. The lower panel is the quantitative relative interacting frequency of *FLC* 5' and 3' regions determined by 3C-qPCR. Error bars: ± s.d. (*n* =2X2 ; biological replicates X technical replicates). (**B, C**) Quantitative relative interacting frequency of *FLC* 5' and 3' regions determined by 3C-qPCR. Error bars: ± s.d. (*n* =2X2 ; biological replicates X technical replicates). Primers used for 3C are listed in Supplementary Table S1. (**D**) Detection of transcriptional initiation form of RNA Pol II levels at the *FLC* promoter region. Error bars: ± s.d. (*n* =2X2 ; biological replicates X technical replicates). (**E**) A model to depict the role of the GH1-HMGA family proteins in precluding the *FLC* gene looping. The left part shows that *FLC* self-looping is stabilized by unknown factors, which promote *FLC* transcription. The right part shows that members of the GH1-HMGA family, including HON5, bind to the *FLC* promoter and region

512 downstream of the terminator to prevent gene loop formation by antagonizing with the unknown
513 factors and RNA Pol II, which in turn suppress *FLC* transcription.

## References


1. K. Zhou, G. Gaullier, K. Luger, Nucleosome structure and dynamics are coming of age. *Nat Struct Mol Biol* **26**, 3-13 (2019).
2. G. Li, D. Reinberg, Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* **21**, 175-186 (2011).
3. E. H. Finn, T. Misteli, Molecular basis and biological function of variability in spatial genome organization. *Science* **365**, (2019).
4. M. Ueda, M. Seki, Histone modifications form epigenetic regulatory networks to regulate abiotic stress response. *Plant Physiol* **182**, 15-26 (2020).
5. B. Bintu, L. J. Mateo, J. H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, X. W. Zhuang, Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, 419-+ (2018).
6. S. S. P. Rao, S. C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K. R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, E. L. Aiden, Cohesin loss eliminates all loop domains. *Cell* **171**, 305-320.e324 (2017).
7. M. H. Kagey, J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, D. J. Taatjes, J. Dekker, R. A. Young, Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435 (2010).
8. D. V. Fyodorov, B. R. Zhou, A. I. Skoultchi, Y. W. Bai, Emerging roles of linker histones in regulating chromatin structure and function. *Nature Reviews Molecular Cell Biology* **19**, 192-206 (2018).
9. D. S. Singer, M. F. Singer, Studies on interaction of H-1 histone with superhelical DNA - characterization of recognition and binding regions of H-1 Histone. *Nucleic Acids Research* **3**, 2531-2547 (1976).
10. T. J. Stasevich, F. Mueller, D. T. Brown, J. G. McNally, Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis. *EMBO J* **29**, 1225-1234 (2010).
11. S. J. McBryant, V. H. Adams, J. C. Hansen, Chromatin architectural proteins. *Chromosome Res* **14**, 39-51 (2006).
12. G. H. Goodwin, E. W. Johns, Isolation and characterization of 2 calf-thymus chromatin non-histone proteins with high contents of acidic and basic amino-acids. *European Journal of Biochemistry* **40**, 215-219 (1973).
13. T. Lund, J. Holtlund, M. Fredriksen, S. G. Laland, On the presence of 2 new high mobility group-like proteins in Hela S3 cells. *Febs Letters* **152**, 163-167 (1983).
14. R. Reeves, Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* **277**, 63-81 (2001).
15. R. Reeves, Nuclear functions of the HMG proteins. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1799**, 3-14 (2010).
16. M. Bustin, Revised nomenclature for high mobility group (HMG) chromosomal proteins. *Trends in Biochemical Sciences* **26**, 152-153 (2001).

17. F. Catez, R. Hock, Binding and interplay of HMG proteins on chromatin: Lessons from live cell imaging. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1799**, 15-27 (2010).
18. L. Aravind, D. Landsman, AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res* **26**, 4413-4421 (1998).
19. N. Ozturk, I. Singh, A. Mehta, T. Braun, G. Barreto, HMGA proteins as modulators of chromatin structure during transcriptional activation. *Front Cell Dev Biol* **2**, 5 (2014).
20. M. Kotlinski, L. Knizewski, A. Muszewska, K. Rutowicz, M. Lirski, A. Schmidt, C. Baroux, K. Ginalski, A. Jerzmanowski, Phylogeny-based systematization of *Arabidopsis* proteins with histone H1 globular domain. *Plant Physiol* **174**, 27-34 (2017).
21. C. Whittaker, C. Dean, The FLC locus: a platform for discoveries in epigenetics and adaptation. *Annu Rev Cell Dev Biol* **33**, 555-575 (2017).
22. D. H. Kim, S. Sung, Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. *Dev Cell* **40**, 302-312 e304 (2017).
23. D. H. Kim, S. Sung, Coordination of the vernalization response through a VIN3 and FLC gene family regulatory network in *Arabidopsis*. *Plant Cell* **25**, 454-469 (2013).
24. Z. Li, D. Jiang, Y. He, FRIGIDA establishes a local chromosomal environment for FLOWERING LOCUS C mRNA production. *Nat Plants* **4**, 836-846 (2018).
25. P. Crevillen, C. Sonmez, Z. Wu, C. Dean, A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J* **32**, 140-148 (2013).
26. C. Liu, Y. J. Cheng, J. W. Wang, D. Weigel, Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*. *Nat Plants* **3**, 742-748 (2017).
27. C. Liu, C. Wang, G. Wang, C. Becker, M. Zaidem, D. Weigel, Genome-wide analysis of chromatin packing in *Arabidopsis thaliana* at single-gene resolution. *Genome Res* **26**, 1057-1068 (2016).
28. C. Wang, C. Liu, D. Roqueiro, D. Grimm, R. Schwab, C. Becker, C. Lanz, D. Weigel, Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res* **25**, 246-256 (2015).
29. F. Ariel, T. Jegu, D. Latrasse, N. Romero-Barrios, A. Christ, M. Benhamed, M. Crespi, Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol Cell* **55**, 383-396 (2014).
30. L. Wang, C. M. Zhou, Y. X. Mai, L. Z. Li, J. Gao, G. D. Shang, H. Lian, L. Han, T. Q. Zhang, H. B. Tang, H. Ren, F. X. Wang, L. Y. Wu, X. L. Liu, C. S. Wang, E. W. Chen, X. N. Zhang, C. Liu, J. W. Wang, A spatiotemporally regulated transcriptional complex underlies heteroblastic development of leaf hairs in *Arabidopsis thaliana*. *Embo j* **38**, (2019).
31. S. D. Michaels, R. M. Amasino, FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949-956 (1999).
32. I. Searle, Y. He, F. Turck, C. Vincent, F. Fornara, S. Krober, R. A. Amasino, G. Coupland, The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes Dev* **20**, 898-912 (2006).

599 33. S. Y. Kim, Y. He, Y. Jacob, Y. S. Noh, S. Michaels, R. Amasino, Establishment of the
600 vernalization-responsive, winter-annual habit in Arabidopsis requires a putative histone
601 H3 methyl transferase. *Plant Cell* **17**, 3301-3310 (2005).
602 34. S. D. Michaels, E. Himelblau, S. Y. Kim, F. M. Schomburg, R. M. Amasino, Integration of
603 flowering signals in winter-annual Arabidopsis. *Plant Physiol* **137**, 149-156 (2005).
604 35. I. Lee, R. M. Amasino, Effect of Vernalization, Photoperiod, and Light Quality on the
605 Flowering Phenotype of Arabidopsis Plants Containing the FRIGIDA Gene. *Plant Physiol*
606 **108**, 157-162 (1995).
607 36. H. Yang, M. Howard, C. Dean, Physical coupling of activation and derepression activities
608 to maintain an active transcriptional state at FLC. *Proc Natl Acad Sci U S A* **113**, 9369-9374
609 (2016).
610 37. A. Patsialou, D. Wilsker, E. Moran, DNA-binding properties of ARID family proteins. *Nucleic*
611 *Acids Res* **33**, 66-80 (2005).
612 38. M. Xu, P. Sharma, S. Pan, S. Malik, R. G. Roeder, E. Martinez, Core promoter-selective
613 function of HMGA1 and Mediator in Initiator-dependent transcription. *Genes Dev* **25**,
614 2513-2524 (2011).
615 39. T. Csorba, J. I. Questa, Q. Sun, C. Dean, Antisense COOLAIR mediates the coordinated
616 switching of chromatin states at FLC during vernalization. *Proc Natl Acad Sci U S A* **111**,
617 16160-16165 (2014).
618 40. S. Swiezewski, F. Liu, A. Magusin, C. Dean, Cold-induced silencing by long antisense
619 transcripts of an Arabidopsis Polycomb target. *Nature* **462**, 799-802 (2009).
620 41. M. Louwers, E. Splinter, R. van Driel, W. de Laat, M. Stam, Studying physical chromatin
621 interactions in plants using Chromosome Conformation Capture (3C). *Nat Protoc* **4**, 1216-
622 1229 (2009).
623 42. H. P. Phatnani, A. L. Greenleaf, Phosphorylation and functions of the RNA polymerase II
624 CTD. *Genes Dev* **20**, 2922-2936 (2006).
625 43. S. D. Michaels, R. M. Amasino, Loss of FLOWERING LOCUS C activity eliminates the late-
626 flowering phenotype of FRIGIDA and autonomous pathway mutations but not
627 responsiveness to vernalization. *Plant Cell* **13**, 935-941 (2001).
628 44. M. Koornneef, C. Alonso-Blanco, H. Blankestijn-de Vries, C. J. Hanhart, A. J. Peeters,
629 Genetic interactions among late-flowering mutants of Arabidopsis. *Genetics* **148**, 885-892
630 (1998).
631 45. D. F. Colombo, L. Burger, T. Baubec, D. Schübeler, Binding of high mobility group A
632 proteins to the mammalian genome occurs as a function of AT-content. *PLoS Genet* **13**,
633 e1007102 (2017).
634 46. C. Charbonnel, O. Rymarenko, O. Da Ines, F. Benyahya, C. I. White, F. Butter, S. Amiard,
635 The Linker Histone GH1-HMGA1 Is Involved in Telomere Stability and DNA Damage Repair.
636 *Plant Physiol* **177**, 311-327 (2018).
637 47. A. Veluchamy, T. Jegu, F. Ariel, D. Latrasse, K. G. Mariappan, S. K. Kim, M. Crespi, H. Hirt,
638 C. Bergounioux, C. Raynaud, M. Benhamed, LHP1 Regulates H3K27me3 Spreading and
639 Shapes the Three-Dimensional Conformation of the Arabidopsis Genome. *PLoS One* **11**,
640 e0158936 (2016).

48. P. Allepuz-Fuster, M. J. O'Brien, N. Gonzalez-Polo, B. Pereira, Z. Dhoondia, A. Ansari, O. Calvo, RNA polymerase II plays an active role in the formation of gene loops through the Rpb4 subunit. *Nucleic Acids Res* **47**, 8975-8987 (2019).
49. K. P. Eagen, E. L. Aiden, R. D. Kornberg, Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc Natl Acad Sci U S A* **114**, 8764-8769 (2017).
50. Q. Wang, Q. Sun, D. M. Czajkowsky, Z. Shao, Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells. *Nat Commun* **9**, 188 (2018).
51. R. C. O'Malley, S. C. Huang, L. Song, M. G. Lewsey, A. Bartlett, J. R. Nery, M. Galli, A. Gallavotti, J. R. Ecker, Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **166**, 1598 (2016).
52. Z. Du, H. Zheng, Y. K. Kawamura, K. Zhang, J. Gassler, S. Powell, Q. Xu, Z. Lin, K. Xu, Q. Zhou, E. A. Ozonov, N. Veron, B. Huang, L. Li, G. Yu, L. Liu, W. K. Au Yeung, P. Wang, L. Chang, Q. Wang, A. He, Y. Sun, J. Na, Q. Sun, H. Sasaki, K. Tachibana, A. Peters, W. Xie, Polycomb Group Proteins Regulate Chromatin Architecture in Mouse Oocytes and Early Embryos. *Mol Cell* **77**, 825-839 e827 (2020).