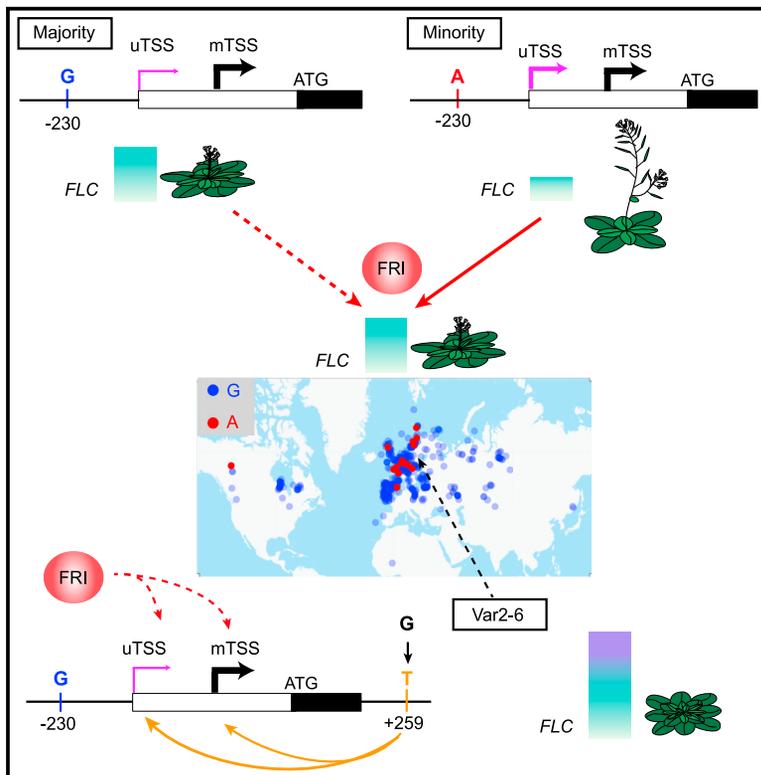


Current Biology

Causal role of a promoter polymorphism in natural variation of the *Arabidopsis* floral repressor gene *FLC*

Graphical abstract



Authors

Pan Zhu, Michael Schon, Julia Questa, Michael Nodine, Caroline Dean

Correspondence

caroline.dean@jic.ac.uk

In brief

Zhu et al. show a promoter SNP is causative for alternative transcription start site choice at the *FLC* gene in natural *Arabidopsis* accessions. Combined with other SNPs and allelic variation of *trans* regulators, it generates quantitative variation in *FLC* expression, thus contributing to the flowering time variation underpinning plant adaptation.

Highlights

- An upstream TSS is alternatively used at *FLC* in natural *Arabidopsis* accessions
- A promoter polymorphism (SNP–230) is causative for changed *FLC* upstream TSS usage
- SNP–230 and FRIGIDA redundantly regulate *FLC* transcription and flowering time
- Combination of different SNPs extends *FLC* expression and flowering time variation



Article

Causal role of a promoter polymorphism in natural variation of the *Arabidopsis* floral repressor gene *FLC*

Pan Zhu,¹ Michael Schon,^{2,3} Julia Questa,^{1,4} Michael Nodine,^{2,3} and Caroline Dean^{1,5,*}¹John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK²Laboratory of Molecular Biology, Wageningen University, Wageningen 6708 PB, the Netherlands³Gregor Mendel Institute, Vienna Biocenter, Vienna 1030, Austria⁴Present address: Center for Research in Agricultural Genomics, Barcelona 08193, Spain⁵Lead contact*Correspondence: caroline.dean@jic.ac.uk<https://doi.org/10.1016/j.cub.2023.08.079>

SUMMARY

Noncoding polymorphism frequently associates with phenotypic variation, but causation and mechanism are rarely established. Noncoding single-nucleotide polymorphisms (SNPs) characterize the major haplotypes of the *Arabidopsis thaliana* floral repressor gene *FLOWERING LOCUS C (FLC)*. This noncoding polymorphism generates a range of *FLC* expression levels, determining the requirement for and the response to winter cold. The major adaptive determinant of these *FLC* haplotypes was shown to be the autumnal levels of *FLC* expression. Here, we investigate how noncoding SNPs influence *FLC* transcriptional output. We identify an upstream transcription start site (uTSS) cluster at *FLC*, whose usage is increased by an A variant at the promoter SNP–230. This variant is present in relatively few *Arabidopsis* accessions, with the majority containing G at this site. We demonstrate a causal role for the A variant at –230 in reduced *FLC* transcriptional output. The G variant upregulates *FLC* expression redundantly with the major transcriptional activator FRIGIDA (FRI). We demonstrate an additive interaction of SNP–230 with an intronic SNP+259, which also differentially influences uTSS usage. Combinatorial interactions between noncoding SNPs and transcriptional activators thus generate quantitative variation in *FLC* transcription that has facilitated the adaptation of *Arabidopsis* accessions to distinct climates.

INTRODUCTION

Arabidopsis thaliana accessions have a worldwide distribution associated with extensive variation in flowering time.¹ Two major determinants of this natural variation are the floral repressor locus *FLOWERING LOCUS C (FLC)* and its transcriptional activator FRIGIDA (FRI).^{2–4} Variation in *FLC* expression determines reproductive strategy, and low expression leads to a rapid-cycling (summer annual) habit, so flowering occurs without the need for cold exposure. By contrast, high expression results in a winter annual habit, where overwintering is required before flowering.⁵

FRI confers a winter annual habit by establishing an active transcriptional environment at *FLC*, delaying flowering.^{6,7} FRI interacts with many co-transcriptional regulators, such as RNA polymerase II (RNA Pol II) and Mediator subunits, Paf1 complex components, and histone modification factors.^{8,9} FRI function also requires both subunits of the nuclear cap-binding complex (CBC), CBP80 and CBP20, which binds to the 5' cap of mRNAs.^{10,11} Extensive nucleotide sequence variation is found at *FRI* among *Arabidopsis* accessions, and most rapid cyclers, such as the common lab strain Columbia (Col-0), carry nonfunctional alleles of *FRI* with premature stop codons or large deletions.^{3,12}

In winter annual accessions, the FRI-upregulated *FLC* expression is antagonized by cold in a process called vernalization.^{9,13}

This starts in autumn when temperatures widely fluctuate over hourly and daily timescales.^{14–16} Short-term cold (>3 h to 2 weeks) exposure promotes FRI nuclear condensate formation sequestering the protein away from the *FLC* promoter.⁹ This facilitates downregulation of *FLC* transcription, which occurs in parallel with the epigenetic switch to a silenced state. The epigenetic switching is established and maintained by Polycomb repressive complex 2 (PRC2)^{17–19} and requires VERNALIZATION INSENSITIVE3 (VIN3) whose expression is induced by cold.²⁰ VIN3 is a PRC2 accessory protein that facilitates PRC2 nucleation at *FLC*, delivering H3K27me3 for long-term epigenetic memory.^{18,20,21} *FLC* transcription is also silenced in the warm through a mechanism involving the RNA binding protein FLOWERING CONTROL LOCUS A (FCA) and alternative polyadenylation of *COOLAIR*, the antisense transcripts at *FLC*. Proximal polyadenylation is linked to histone demethylation and PRC2-induced chromatin changes over the body of *FLC*.^{22–24}

Natural variation in *FLC* transcription influences *FLC* expression both as the plants germinate and during the early response to cold temperatures (collectively called autumnal *FLC* expression). This variation is determined by noncoding *cis*



polymorphism at the *FLC* locus that results in different cold requirements in natural accessions.^{5,25,26} For example, a noncoding single nucleotide polymorphism (SNP) within *FLC* intron 1 (SNP+259) in the *FLC* haplotype contained in the Swedish accession Var2-6 leads to increased *FLC* transcription and a requirement for 8 weeks of vernalization, through alteration of *COOLAIR* splicing.²⁷ Given that autumn temperatures fluctuate widely and show extensive year-on-year differences, variable autumnal *FLC* expression enables plants to adapt to different climates influencing when epigenetic silencing initiates. How most of the SNPs influence *FLC* autumnal expression, however, has been largely unexplored.

Here, we investigate this natural variation and discover an upstream transcription start site (uTSS) cluster at the *FLC* locus important for quantitative regulation of *FLC* expression in natural *Arabidopsis* accessions. Alternative transcriptional initiation is widely found in eukaryotes.^{28–30} Generally, there is a cluster of TSSs rather than one specific site,^{28,29,31} and large-scale shifts in TSS usage are commonplace during development and response to environment,^{32–35} altering 5′ untranslated regions (UTRs), thereby influencing transcript stability and translation.^{34–36} In addition, an upstream TSS frequently inhibits the use of a downstream TSS.^{37,38}

TSS position can be influenced by DNA motifs, such as the well-known TATA box and other transcription factor binding sites.³⁹ Clear TATA-box motifs only occur in ~10% of mammalian promoters³¹ and ~30% of *Arabidopsis* genes^{29,40} and transcription factor binding, enhancer-promoter connections, or epigenetic status have been shown to be modulated through SNPs in promoter elements.^{41–44} In plants, such promoter elements are the Y Patch, GA, and CA elements, which are different from those in yeast and mammals.^{45,46} Our natural variation analysis shows that a G variant at –230 (230 base pairs [bp] upstream of ATG) in the *FLC* promoter of the majority of natural accessions (>88%) results in higher *FLC* expression, possibly through the formation of a binding site for plant-specific NAM, ATAF1/2, and CUC2 (NAC) transcription factors.^{47–49} By contrast, an A variant at this site promotes uTSS usage, reduces *FLC* expression, and accelerates flowering, possibly by forming a CA-rich motif.^{44,46} Our genetic evidence further demonstrates a strong interaction between this SNP and *FRI*, as well as an additive effect with an intronic noncoding SNP at +259. Overall, this work provides important insights into the synergistic transcriptional control by *cis* SNPs and *trans* regulators that have produced variation in *FLC* transcription levels, defining adaptively important *FLC* haplotypes.

RESULTS

Multiple TSS sites are used at *FLC* throughout development

Given autumnal *FLC* expression is an important variable in vernalization response in *Arabidopsis* accessions, we asked whether *FLC* transcriptional initiation was affected in different haplotypes. The *FLC* TSS was mapped in the Col-0 (*fri*), Col *FRI* (*FRI* from SF2 accession introgressed into Col-0), and Var2-6 near isogenic line (NIL), in which *FLC* from the Swedish accession Var2-6 had been introgressed into Col *FRI*.²⁵ Although the summer annual Col-0 does not need vernalization to flower,

Col *FRI* flowers late but has a rapid vernalization (RV) response—acceleration of flowering is saturated by 6 weeks of cold exposure. The Var2-6 NIL has a slow vernalization (SV) response—8 weeks of cold exposure is required to fully accelerate flowering.^{25,26} 5′ rapid amplification of cDNA ends (5′ RACE) analysis using seedlings grown in warm and cold conditions showed most *FLC* TSSs are distributed between –60 and –120 from the ATG, consistent with the annotation in The *Arabidopsis* Information Resource (TAIR) and recently published datasets^{38,50} (Figures 1A and 1B). Interestingly, an uTSS cluster at *FLC*, 100 bp upstream of the major TSS (mTSS) was mapped in both Col *FRI* (in the warm and cold) and Var2-6 NIL (in the cold) (Figures 1A and S1). The failure to detect it in Col-0 by 5′ RACE likely reflects the lower expression, as it is occasionally detected in cap analysis gene expression sequencing (CAGE-seq) and TSS sequencing (TSS-seq)^{38,50} (Figures 1A and 1B).

We next examined the transcripts produced from uTSS by 3′ RACE and found that the majority produced full-length mature spliced *FLC* mRNAs with no alternative splicing sites detected (Figures S2A–S2D). However, some less abundant proximally polyadenylated transcripts terminating within *FLC* intron1 were found in every genotype (Figures S2A–S2C). A cluster of short non-polyadenylated transcripts with termination sites in the *FLC* 5′ UTR and exon1 were also detected with increased abundance in cold-treated samples (Figures S2A and S2B). This is likely to correspond to the previously described *COLDWRAP* noncoding RNA transcribed from even further upstream at a TSS located ~–315 bp from the *FLC* ATG.⁵¹ Given that *COLDWRAP* is not co-amplified with either unspliced or spliced *FLC* mRNA and exhibits an opposite expression pattern to *FLC* during vernalization, it is thought to be independent of *FLC* transcription.⁵¹ Therefore, we conclude that the mapped uTSS mainly produce *FLC* mRNA with an extended 5′ UTR that has no potential to contain any upstream open reading frames (uORFs) (Figure S2E).^{35,52}

Prematurely terminated transcripts that initiate at a TSS upstream of the protein-coding gene can play important roles in gene expression.^{53,54} Our previous work has shown that *FLC* can be proximally polyadenylated within the first intron in a process promoted by *FCA* and antagonized by *FRI* and that this sets *FLC* expression during embryo development.⁵⁵ Given the tight correspondence between the site of transcription initiation and 3′ end selection,³⁹ we wanted to know if this proximal polyadenylated event was associated with the uTSS. *FLC* uTSS was detected at all eight stages of embryo development analyzed, in each contributing around 20% of the total TSS, in contrast to the decreasing usage of the proximal polyadenylation site (Figures S3A–S3C). Consistently, compared with Col-0, neither Col *FRI*, with less proximally polyadenylated *FLC*, nor 35S:: *FCA*, with more proximally polyadenylated *FLC*, changed uTSS usage (Figures 1C–1E). In addition, uTSS produces the same proportion of distal *FLC* as uTSS + mTSS (Figures S3D and S3E). Collectively, these data suggest that *FLC* uTSS is used during both embryo and seedling stages, and its usage does not associate with the changed poly(A) site during embryo development.

We calculated relative uTSS usage from the ratio of the relative expression level of uTSS *FLC* to uTSS + mTSS *FLC* (hereafter called total *FLC*) by reverse transcriptase qPCR (Figures 2A

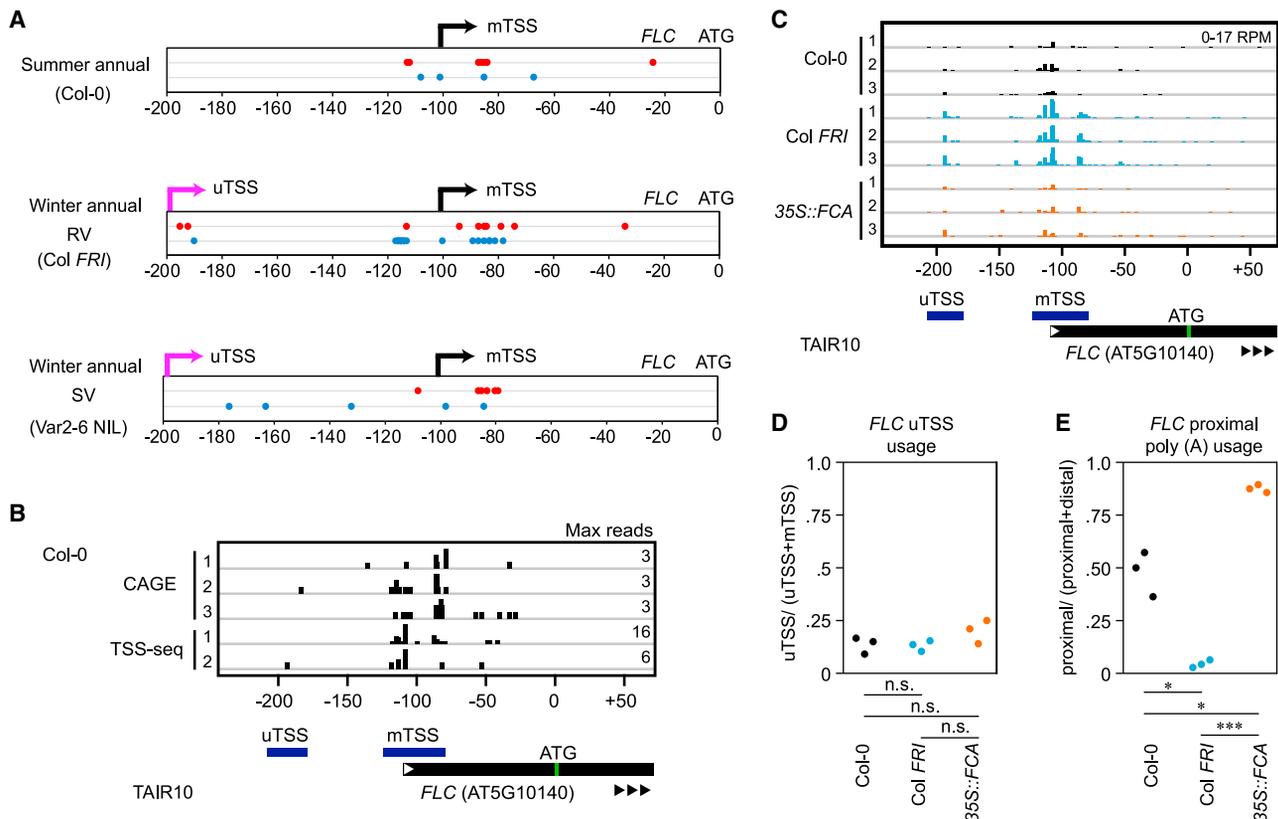


Figure 1. Characterization of upstream transcription start sites (uTSSs) at *FLC*

(A) *FLC* TSS mapped by 5' RACE in total RNA from seedlings, *TUB6* used as control as shown in Figure S1. See also Table S4. Red dots indicate TSS in non-vernalized (NV) plants, and blue dots for TSS in plants given 2 weeks cold (2WV). Black arrows are mTSS, and pink arrows are uTSS. (B) Genome browser screenshot of CAGE-seq, $n = 3^{50}$ and TSS-seq $n = 2^{38}$ showing TSS at *FLC* in Col-0 seedlings. (C) 5' end-labeled RNA sequencing read abundance at *FLC* ($n = 3$) of Col-0, Col *FRI* and 35S::*FCA* early heart embryos. (B and C) 5' end of *FLC* gene is shown below. Black triangles show transcriptional direction, blue bars are regions for TSS usage quantification. (A–C) The x axis indicates the distance in base pairs from ATG. (D) Ratio of transcripts from uTSS: transcripts from (uTSS + mTSS) in the same samples as in (C). (E) Proximal poly(A): (proximal + distal) ratio in the same sample as in (C). One-way ANOVA, p values were adjusted with Dunnett's multiple comparisons tests. n.s., not significant; * $p < 0.05$; *** $p < 0.001$. See also Figures S2 and S3.

and S3D). There is no change in uTSS usage between proximal + distal and distal *FLC* transcripts in Col-0 and Col *FRI* (Figure S3F), supporting the conclusion that *FRI* promotes *FLC* transcription from both uTSS and mTSS without changing uTSS usage (Figures 1C and 1D). Expression of *FLC* from uTSS was significantly reduced in both *cbp20 fri* and *cbp20 FRI* (Figures 2B and 2C), the mutant of the small subunit of the nuclear CBC,¹⁰ reinforcing the role of uTSS as a *bona fide* TSS. The usage of uTSS was not changed in either *cbp20 fri* or *cbp20 FRI* (Figures 2D–2F), further confirming that *FRI* via *CBP20* promotes *FLC* transcription in Col *FRI* using both uTSS and mTSS.

A naturally occurring A/G variation at –230 changes *FLC* uTSS usage

The *FLC* uTSS was not detected by 5' RACE in warm-grown Var2-6 NIL seedlings (Figure 1A). A comparative analysis showed uTSS *FLC* expression is almost 50% lower in Var2-6 NIL compared with Col *FRI*, despite a nearly 3-fold higher total *FLC* RNA level (Figures 2G–2I). Var2-6 is one of the five major *FLC* haplotypes defined by noncoding sequence variation that

confer variable vernalization response^{25,26,56,57} (Table S1). We also compared uTSS usage in NILs of the other haplotype groups including Uil2-5 NIL, Lov-1 NIL2, and Edi-0 NIL (Table S1). They all had low relative uTSS usage and high total *FLC* levels (except for Edi-0 NIL) in comparison to Col *FRI* (Figures 2G–2I), revealing a link between reduced uTSS usage and higher *FLC* expression in most accessions.

A 75-bp sequence between –272 and –197 in the uTSS promoter had previously been defined as essential for non-vernalized (NV) *FLC* expression.⁵⁸ We found a SNP located within the uTSS core promoter (230 bp upstream of *FLC* ATG, hereafter named SNP–230), and A in Col-0 but G in most other haplotypes (Figure 3A). Further investigation revealed that the A variant was present in only four of the 47 *A. thaliana* accessions that represent the major *FLC* haplotypes (Table S1), with the majority having a G variant at SNP–230. Of the 1,135 sequenced accessions, 1.6% (18 accessions) have the A variant, while 88.2% have G at –230, with the remaining small portion 10.2% carrying various deletions of that sequence (Figure 3B). The geographical distribution of the accessions carrying the A variant (Figure 3B)

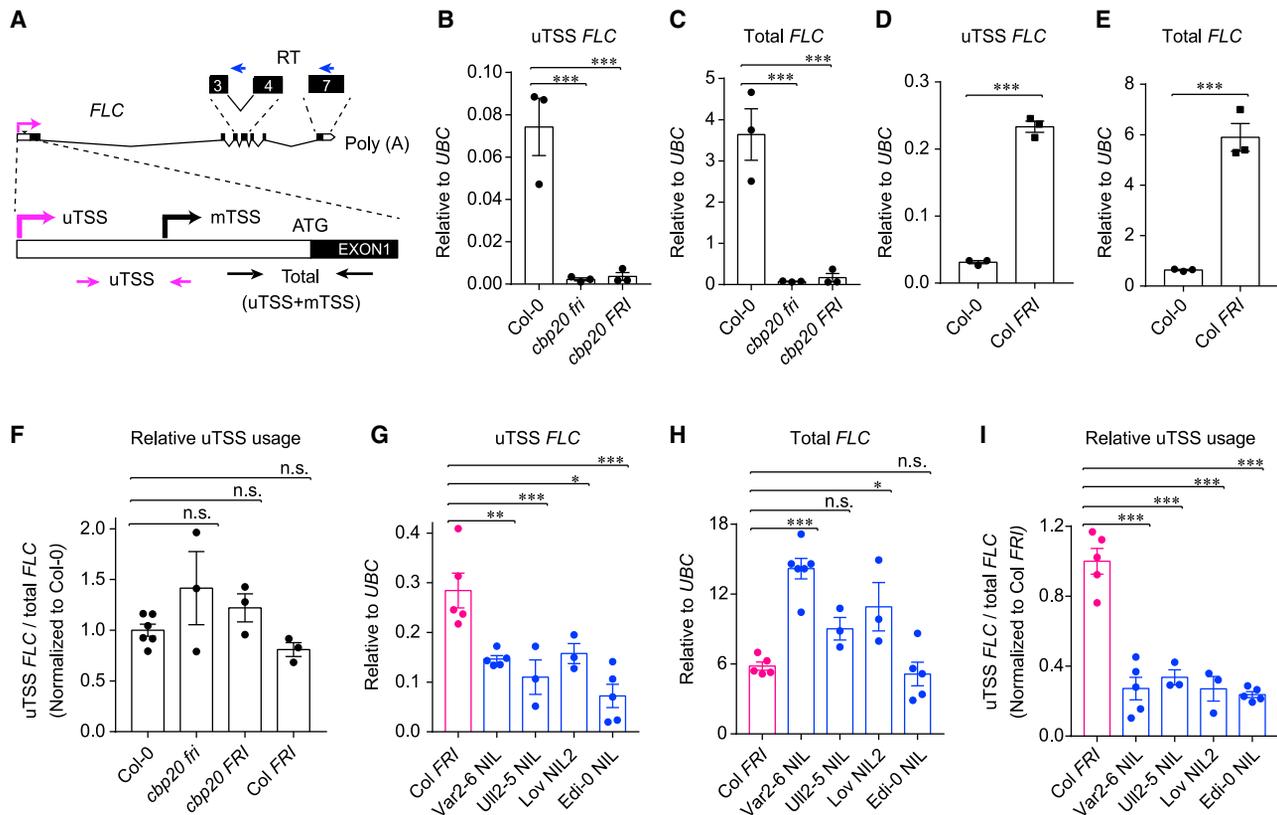


Figure 2. uTSS usage varies in *FLC* haplotypes

(A) Schematic illustration of full-length *FLC* transcript from uTSS (expanded below for the 5' UTR-exon1 region). Black boxes represent exons, and white boxes represent UTR. Primers for RT are shown by blue arrows. Primer pairs for qPCR are indicated with head-to-head arrows. See also Figure S3. (B–E and G–H) The relative expression level of *FLC* transcripts from uTSS (B, D, and G) and uTSS + mTSS (total *FLC*) (C, E, and H) in the indicated genotypes. (F and I) Relative uTSS usage was measured by the ratio of uTSS *FLC* to total *FLC*. Data are represented as mean \pm SEM. (B, C, and F–I) One-way ANOVA, *p* values were adjusted with Dunnett's multiple comparisons tests. (D and E) Unpaired t test (two-tailed) with Welch's correction. n.s., not significant; * *p* < 0.05; ** *p* < 0.01; *** *p* < 0.001.

does not suggest any obvious association of the functional consequences of this A variant with a specific climatic variable. Moreover, most species within the Brassicaceae family contain the G variant at this site⁵⁹ (Figure 3C), supporting a G variant as the ancestral state.

The accessions carrying the A variant at –230 had been classified as rapid vernalizing with lower *FLC* expression and early flowering (Table S1; Figure S4). To investigate whether SNP–230 variation is causal for TSS selection and changed *FLC* expression, we generated transgenic plants carrying *Col FLC* with a single point A to G mutation at SNP–230 (A-230G). Since the A variant occurred in both accessions with an active *FRI* allele and those with inactive *FRI* (Figures 3B and S4B), the influence of A-230G on uTSS usage was tested in both *fri* and *FRI* backgrounds (Figures 3D–3F and S5A–S5D). The G substitution significantly lowered uTSS usage in both backgrounds but increased *FLC* expression levels, particularly in *fri* background (Figures 3D–3F and S5A–S5D). This resulted in A-230G *fri* flowering much later than Col-0, whereas A-230G *FRI* flowered at a similar time to *Col FLC FRI* (Figures 3G–3I). Overall, these data support the view that SNP–230 is indeed the SNP responsible for the reduced uTSS usage in the NILs.

SNP–230 falls within the distal exon of the *COOLAIR* transcript close to its distal poly(A) site (Figure S5A), so we also asked whether SNP–230 influences *COOLAIR* processing. Except for the alternative splicing event in Var2-6 NIL caused by SNP+259,²⁷ no alternatively spliced isoforms were detected in Edi-0 NIL, Lov NIL2, UII2-5 NIL, or the A-230G transgenic lines, either in *fri* or *FRI* backgrounds (Figures S5E–S5G). However, the G variant in a *fri* background increased *COOLAIR* transcript levels and promoted distal polyadenylation, like *FRI* and mutants such as *fca-9*^{22–24} (Figures S5H–S5J). Distal *COOLAIR* RNA forms elaborate secondary structures *in vivo* that regulate *FLC* expression.⁶⁰ SNP–230 is located within the H13 of the central structural domain (Figure S5K), which is highly stable in both warm and cold conditions.⁶⁰ The A to G substitution changes U·G pairing into a stronger C·G pairing, reinforcing the RNA helix formation (Figure S5K). Therefore, the changed transcriptional output of *FLC* in *fri* by SNP–230 is unlikely to be associated with a changed *COOLAIR* structure.

The reduction in uTSS usage by the G variant is independent of *FRI* (Figure 3F) suggesting SNP–230 works genetically upstream of *FRI* to upregulate *FLC*. This genetic relationship was also detected in early heart-stage embryos (Figures 3J and 3K). At this developmental stage, *FRI* antagonizes FCA-induced proximal

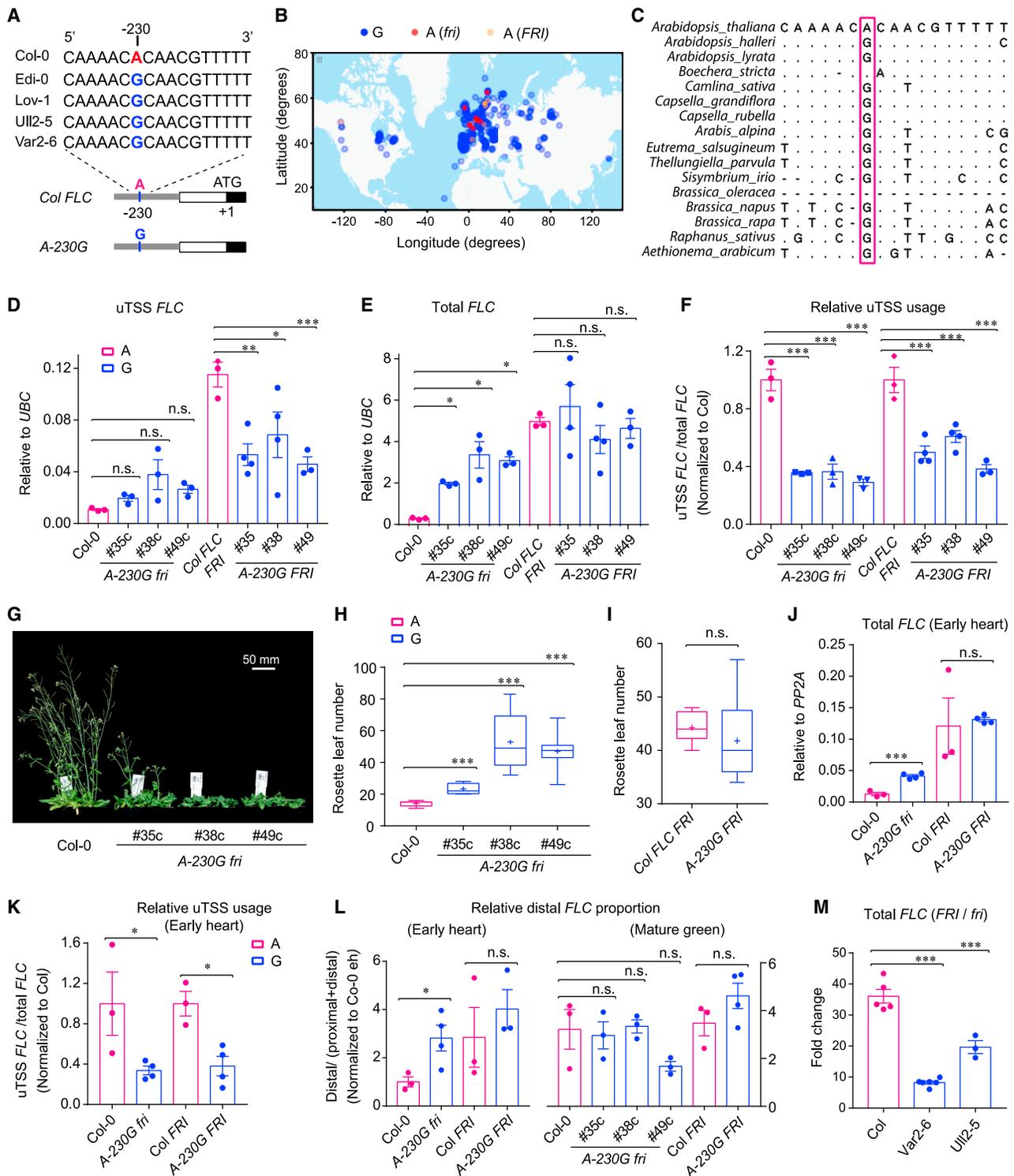


Figure 3. SNP-230 is causative for changes of FLC uTSS usage and transcriptional output

(A) Sequence alignment of uTSS promoter region in the 5 representative *A. thaliana* accessions from the major FLC haplotypes in Table S1. –230 (distance from ATG, bp) indicates the causal SNP. Transgenes Col FLC and A-230G carry the FLC genomic fragment either from Col-0 or with a single A to G substitution at –230. See also Table S1.

(B) Geographic distribution of the *Arabidopsis thaliana* accessions carrying A or G variant from the 1,135 accessions in 1001 Genome database (<https://1001genomes.org/>). The remaining 116 accessions with various lengths of deletions at this promoter region are not shown. See also Figure S4.

(C) Sequence alignment of the same genomic fragment in (A) within Brassicaceae (data from PlantRegMap⁵⁹). The sequence of *Arabidopsis thaliana* is Col-0 accession. SNP-230 is outlined.

(legend continued on next page)

polyadenylation of the *FLC* transcript to maximally increase *FLC* expression and delay flowering⁵⁵ (Figures 1E, S3B, and S3E). The G variant also has an FRI-like function in promoting the proportion of distal *FLC* isoforms at the early heart embryonic stage, linking to alternative TSS usage⁶¹ (Figures 3L and S5L). When the embryos reach the mature green stage the majority of *FLC* isoforms are distally polyadenylated (Figures S3B and S3E), and at this stage of development the influence of the G variant and FRI on polyadenylation site selection is reduced (Figures 3L and S3E). This could explain the lack of additivity of A-230G and FRI on total *FLC* expression in both seedlings and early heart-stage embryos (Figures 3E and 3J). This epistatic interaction could prevent excessive upregulation of *FLC* expression in natural accessions carrying a G variant at -230 and a functional *FRI* allele. Consistent with this hypothesis, compared with *Col* carrying the A allele at SNP-230, the upregulation of total *FLC* by FRI is compromised in both Var2-6 and Ull2-5 containing the G variant (Figure 3M). By contrast, accessions carrying the A variant in the absence of FRI show an even earlier flowering phenotype (Figure S4). The multiple combinations of *FRI* allelic variation and A-230G variation thus have the potential to extend the flowering time range of *Arabidopsis thaliana* accessions.

SNP-230 and SNP+259 synergistically upregulate *FLC* in Var2-6

Most of the *FLC* haplotypes, when combined with an active *FRI* allele, showed higher total *FLC* levels than *Col FRI* (Figure 2H). However, the transgene A-230G *FRI* did not (Figures 3E and 3J), so we speculated that other SNPs may function in an FRI-independent manner to cause an additive effect on *FLC* transcription. To find such SNPs, we examined SNPs in Var2-6 *FLC*, which of all the NILs shows the highest total *FLC* expression (Figure 2H). We had previously determined that SNP+259 located within *FLC* intron 1 contributes to the high *FLC* expression in Var2-6 (Figure 4A). A G/T variant influences the choice of splice acceptor site in the large *COOLAIR* intron, with a G to T substitution causative for higher transcription of *FLC*.²⁷ In addition to Var2-6 NIL, we selected Ull2-5 NIL, which contains the same G variant as Var2-6 at -230 but the *Col-0* G variant at +259 (Figure 4A), enabling us to disentangle the contribution of the different SNPs. Ull2-5 NIL shows much lower total *FLC* than Var2-6 NIL but still higher than *Col FRI* (Figure 2H). Both Ull2-5 NIL and Var2-6 NIL were crossed with *Col fri*, and both Var2-6 NIL *fri* and Ull2-5 NIL *fri* expressed higher levels of *FLC* with lower uTSS usage than *Col-0*, like that in the *FRI* background (Figures 4B-4D). This is consistent with the A to G substitution at -230 causative for the FRI-independent uTSS usage reduction and total *FLC* upregulation.

Whether SNP +259 affects *FLC* expression levels in an FRI-dependent manner was examined by comparing the *Col FLC* allele carrying the G to T substitution at SNP +259 (*Col-G259T*) and the reciprocal mutant, i.e., a Var2-6 *FLC* allele carrying a T-G substitution (*Var-T259G*), in both *fri* and *FRI* backgrounds (Figure 4A). Consistent with our hypothesis, *Col-G259T* gave rise to higher total *FLC* independently of *FRI*, whereas *Var-T259G FRI* had an unchanged total *FLC* level compared with *Col FLC FRI* (Figures 4E and 4F). Next, the influence of SNP+259 on uTSS usage was analyzed. *Col-G259T* slightly increased uTSS usage, indicating a different regulatory mechanism from SNP-230 (Figures 3F and 4G), and *Var-T259G* showed lower uTSS usage (Figure 4G), resembling A-230G *FRI* (Figure 3F). These data further support that the highest *FLC* expression in Var2-6 NIL (Figure 2H) is mainly caused by the G to T substitution at +259. However, the reduced uTSS usage in Var2-6 NIL demonstrates the SNP-230 mediated TSS selection is genetically upstream of SNP+259. In summary, in the Var2-6 NIL, the G variant caused reduced uTSS usage and FRI redundantly activate *FLC* transcription, while SNP +259 upregulates *FLC* independently of FRI. Their combination thus synergistically affects flowering time.

SNP-230 and SNP+259 regulate *FLC* uTSS usage independently of vernalization

Higher *FLC* levels determine the requirement for overwintering and vernalization, the cold-induced epigenetic silencing of *FLC*. We therefore asked whether and how uTSS usage was influenced during vernalization. *FLC* expressed from uTSS and mTSS progressively decreased with cold exposure in all the genotypes tested (Figures 5A, 5B, 5D, 5E, 5G, and 5H). Interestingly, relative uTSS usage increased in all genotypes after cold treatment (Figures 5C, 5F, and 5I). The difference in *FLC* expressed from uTSS between *Col FLC* and A-230G was smaller after 5 weeks of vernalization (5WV) compared with NV and 2WV (Figures 5A and 5D). Thus, the major effect caused by SNP-230 mediated TSS selection seems to predominantly influence *FLC* expression in warm conditions or after short cold exposure. Consistently, the upregulation of *FLC* by SNP+259 is also independent of vernalization, with G to T substitution on *Col FLC* giving higher *FLC* expression without reducing uTSS usage (Figures 5D-5F). The reciprocal mutant (*Var-T259G*) behaved the same as A-230G in all the conditions (Figures 5A-5F).

In addition, the expression level of *FLC* from uTSS in both *Col FRI* and Var2-6 NIL was reduced after cold (Figure 5G) but the relative uTSS usage in the Var2-6 NIL remained lower than *Col FRI* (Figure 5I). The genetic interaction between SNP-230 and

(D-F) Relative expression level of uTSS *FLC* (D), total *FLC* (E), and uTSS usage (F) in 10-day-old seedlings. Numbers indicate independent transgenic lines. The comparison of A-230G *fri* to *Col FLC fri* is shown in Figure S5.

(G-I) Flowering phenotype of A-230G transgenic plants. Center lines show the median, box edges delineate the 25th and 75th percentiles, bars extend to minimum and maximum values, and "+" indicates the mean value. n = 8 plants per genotype in (G) and (H), and n = 13 for A-230G *FRI* in (I), including #35 (n = 4), #38 (n = 6), and #49 (n = 3).

(J and K) Expression level of total *FLC* (J) and relative uTSS usage (K) in embryos at the early heart stage. For A-230G transgenic plants, at least one replicate from each of the three independent lines in (D)-(F) was pooled.

(L) The relative proportion of distal *FLC* isoforms (distal *FLC*/proximal *FLC*+ distal *FLC*) in early heart and mature green embryos by RT-qPCR. See also Figure S5.

(M) The fold change of total *FLC* expression level calculated by the ratio of total *FLC* in *FRI* to that in *fri* within different *FLC* haplotypes. For example, the fold change in *Col* = total *FLC* in *Col FRI*/total *FLC* in *Col-0 (fri)*. For all qPCR, data are represented as mean ± SEM.

(D-F, H, and K-M) One-way ANOVA with p values adjusted in Dunnett's multiple comparisons tests. (I and J) Unpaired t test (two-tailed) with Welch's correction. n.s., not significant; * p < 0.05; ** p < 0.01; *** p < 0.001.

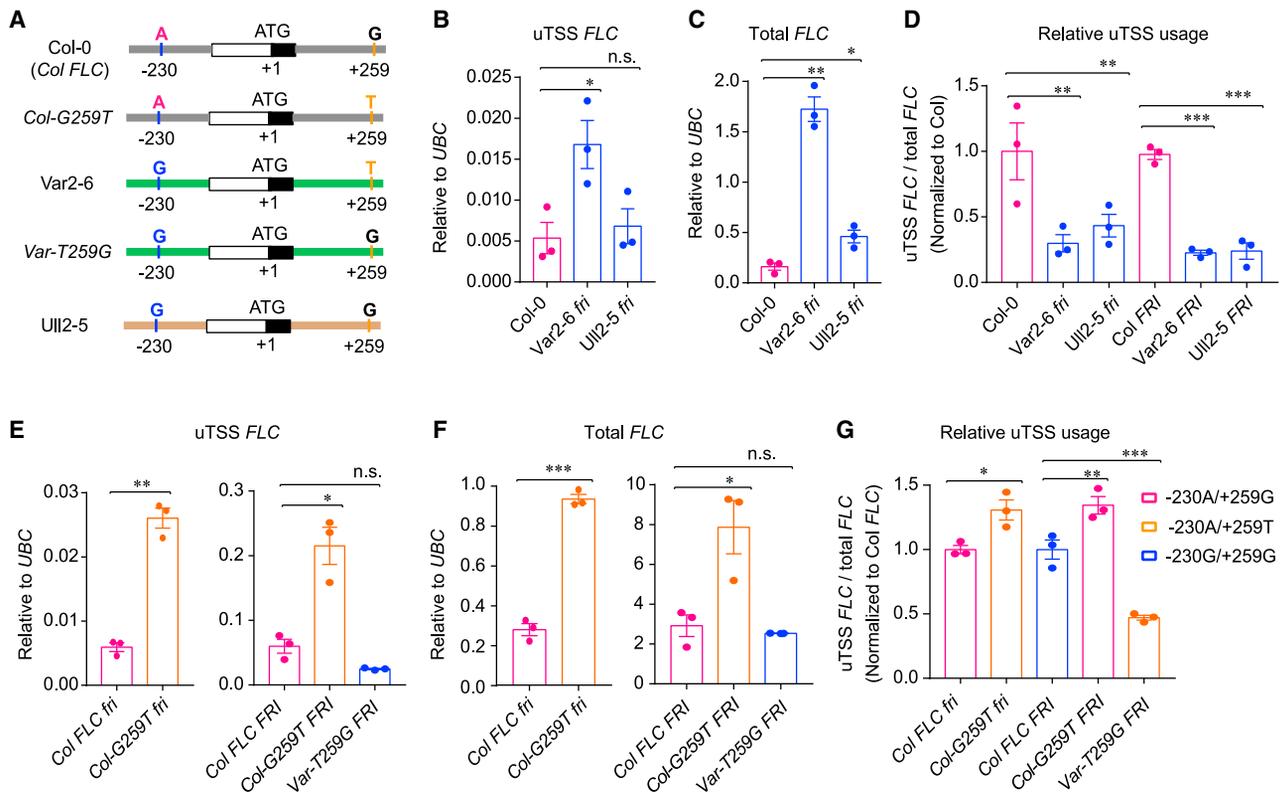


Figure 4. SNP-230 and SNP+259 cooperate to upregulate *FLC* in *Var2-6*

(A) Schematic illustration of SNP-230 and SNP+259 in *Col-0*, *Var2-6*, and *Ull2-5* *FLC* haplotypes shown by colored bars. *Col-G259T* and *Var-T259G* are the reciprocal SNP+259 exchange made in *Col* and *Var 2-6* *FLC* constructs.

(B–G) The relative expression level of uTSS *FLC* (B and E), total *FLC* (C and F), and uTSS usage (D and G). For all qPCR, data are represented as mean \pm SEM. (B and C) Unpaired t test (two-tailed) with Welch's correction. (D–G) One-way ANOVA with p values adjusted in Dunnett's multiple comparisons tests. n.s., not significant, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

SNP+259 would therefore not appear to change during vernalization. Of note, uTSS usage was unaltered by *vin3-4* in either *Col FRI* or *Var2-6 NIL*, although *FLC* expression was universally increased due to the defective cold-induced epigenetic silencing of *FLC* (Figures 5G–5I). Thus, the two SNPs mediate uTSS usage regulation independently of VIN3. Interestingly, *FLC* levels were further increased in vernalized *Var2-6 NIL vin3-4* compared with *Var2-6 NIL* or *vin3-4* and to a milder extent in *Ull2-5 NIL vin3-4* and *Edi-0 NIL vin3-4* (Figures 5G–5I and S7). This demonstrates an additive effect between *cis* SNPs and the VIN3-PRC2 silencing pathway, further confirming that SNP+259 and SNP-230 regulate *FLC* expression independently of VIN3. These data reinforce the conclusion that most SNP-induced variation at *FLC* affects autumnal *FLC* expression (VIN3-independent pathway) rather than cold-induced epigenetic silencing.²⁵

DISCUSSION

Arabidopsis thaliana flowering time has emerged as an important system to explore how noncoding polymorphism mechanistically alters gene regulation that underpins adaptive variation.^{25–27} Here, through analysis of *FLC* haplotypes, we show how a promoter SNP (SNP-230) modulates *FLC* transcription by influencing TSS selection, thus contributing to flowering time

variation (Figure 6). The accessions carrying A or G variants at this SNP position represent the majority (~90%) of the worldwide natural accessions. The variants at SNP-230 interact differentially with the transcriptional activator FRI, and these effects combine additively with variant polymorphisms at SNP+259. This detailed analysis reveals how combinations of *cis* SNPs and *trans* activators cooperate to quantitatively regulate transcriptional output.⁶² The -230 and +259 SNP effects are independent of the VIN3-mediated epigenetic silencing pathway that delivers the cold-induced silencing at *FLC*.^{15,18,20} Thus, the most variable part of the vernalization mechanism in *Arabidopsis thaliana* accessions lies in the autumnal *FLC* expression.²⁵ Due to the widely and rapidly fluctuating environmental conditions, as well as varied lengths of autumn at different latitudes, the interplay between SNP-230 and SNP+259 plus the highly variable and temperature-sensitive FRI function would provide important plasticity for the plant's response to the environment.⁹

Quantitative regulation of *FLC* via co-transcriptional mechanisms involves many activators (e.g., FRI) and repressors such as FCA, which is a component of the autonomous floral pathway.^{22–24,63} FRI and FCA antagonistically regulate *FLC* proximal polyadenylation and set *FLC* expression during the early stages of embryo development (Figures 1C–1E and S3). A connection between transcription termination and transcription

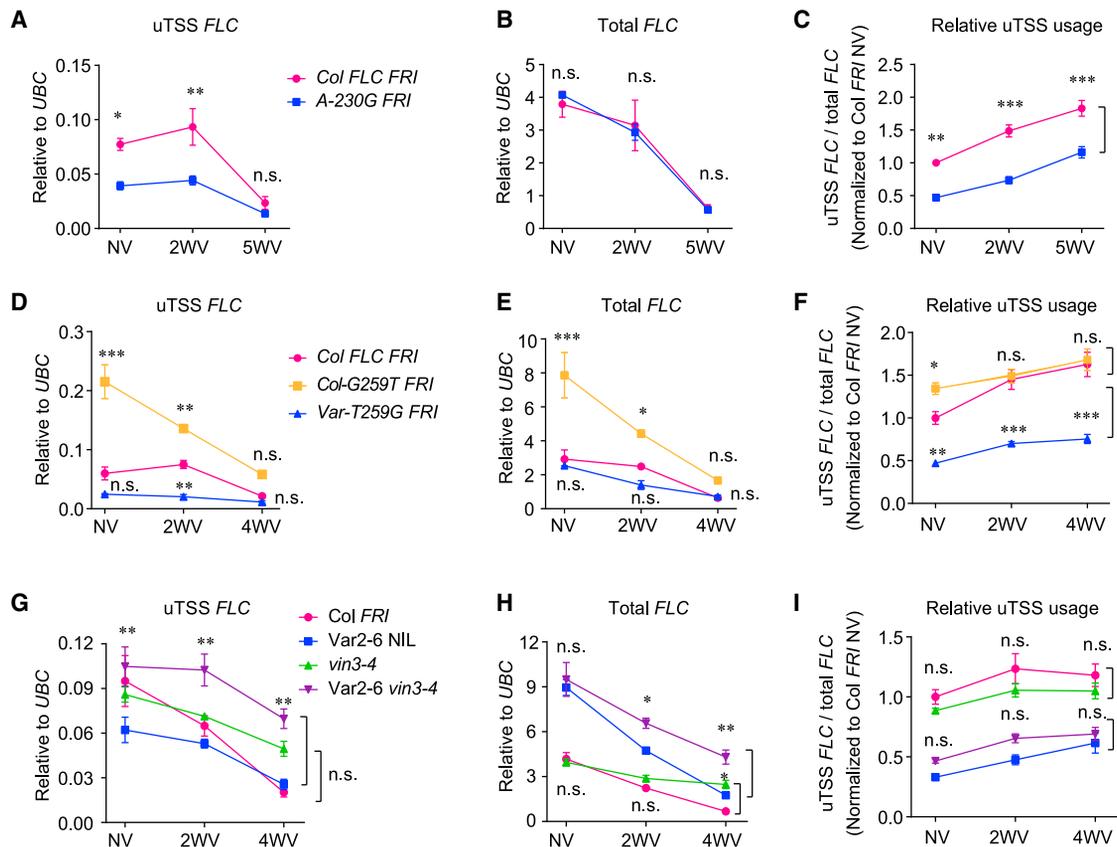


Figure 5. SNP-230 and SNP+259 regulate *FLC* uTSS usage independently of vernalization

(A, D, and G) Relative expression level of uTSS *FLC* with different lengths of vernalization.

(B, E, and H) Relative expression level of *FLC* transcripts from uTSS + mTSS (total *FLC*).

(C, F, and I) Relative usage of uTSS measured by the ratio of uTSS *FLC* to total *FLC*. Plants were given no cold treatment (NV, non-vernalized) or 2, 4, or 5 weeks of cold treatment (2WV, 4WV, or 5WV). Data are represented as mean \pm SEM. Two-way ANOVA with p values adjusted in Sidak's multiple comparisons tests. In (A)–(F), all analyses were performed by comparing to *Col FLC FRI*, and in (G)–(I), *Var2-6 NIL* was compared with *Col FRI* and *Var2-6 vin3-4* to *vin3-4*. n.s., not significant, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. See also Figure S6.

initiation sites has been shown to influence transcriptional output^{53,61} and indeed, we did find an association between the G variant reduced uTSS usage and an increased proportion of distal *FLC* transcripts in early heart-stage embryos (Figures 3J–3L). This would explain the redundancy of the G variant and *FRI* in upregulating *FLC*. However, *FLC* uTSS usage is not changed through different embryo developmental stages in *Col-0* or in the presence of *FRI* or overexpression of *FCA* (Figures 1C–1E and S3), suggesting the highly regulated alternative polyadenylation of sense *FLC* transcript during embryo development is not through alternative uTSS usage.

There are several possibilities as to how the A to G substitution at SNP-230 reduces uTSS usage, resulting in higher *FLC* transcription.⁶⁴ The most straightforward explanation is that it affects a promoter *cis* element. For example, SNP-230 lies within an ACGCAA element, which has been shown to function as a core binding site for NAC transcription factors.^{47–49} The G variant may also affect the functioning of a CA-rich promoter motif (AAAACACAAC),^{44,45,50,65} or influence co-transcriptional regulation of *FLC* regulation through disruption of a gene loop (as it is close to the *COLDWRAP* TSS) that shapes the transcription initiation process.^{7,42,51,66}

In summary, we identify how a noncoding promoter SNP is causative for *FLC* transcriptional variation in different *Arabidopsis* accessions. We also demonstrate that the combinatorial effects of different noncoding SNPs and an important transcriptional activator provide quantitative variation in *FLC* transcriptional output, which is a major determinant underpinning adaptive changes in *Arabidopsis thaliana* accessions (Figure 6). An understanding of how natural variation mechanistically endows transcriptional plasticity is likely to become even more important in the context of climate change.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability

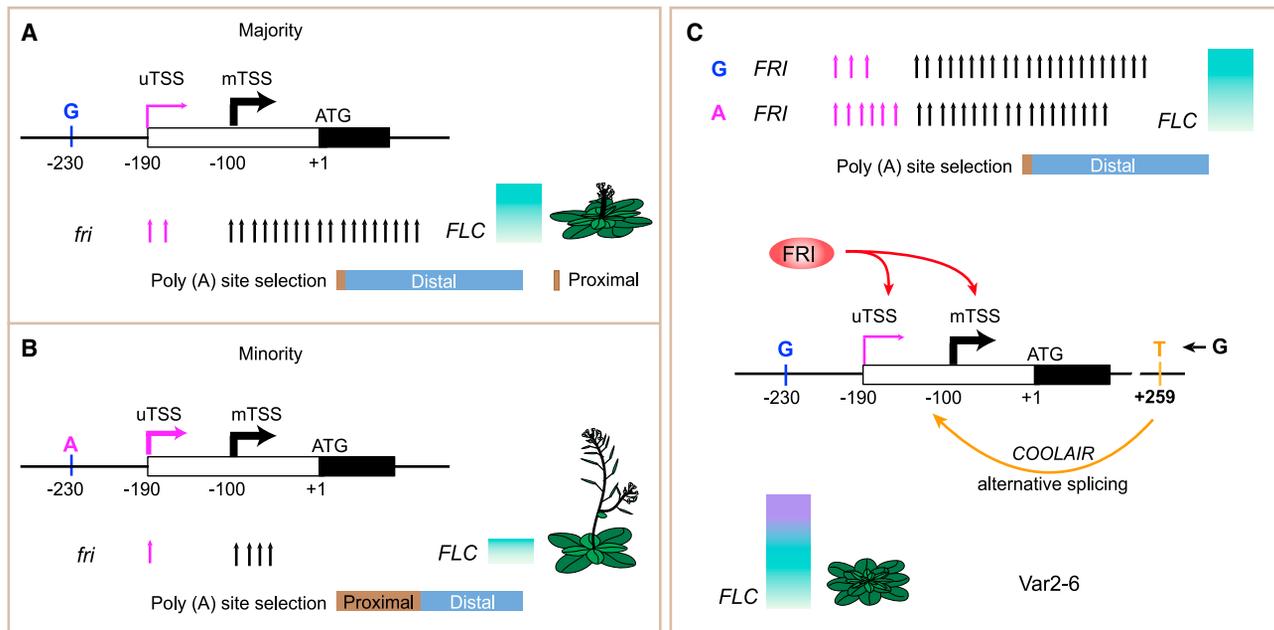


Figure 6. Combinatorial interaction between noncoding SNPs and FRI variation in *FLC* expression in natural *Arabidopsis* accessions

G allele at SNP-230 in most *Arabidopsis* accessions is causative for the reduced uTSS usage that is associated with higher *FLC* expression and late flowering (A). The naturally occurring G to A substitution in a minority of accessions results in decreased *FLC* expression and early flowering with higher uTSS usage (B). Compared with A, the G variant-mediated uTSS reduction is associated with a preference for distal polyadenylated *FLC* in the early heart-stage embryos when *FLC* transcriptional state is set (A and B), redundantly with FRI (C, top). SNP+259 additionally upregulates *FLC* expression and further delays flowering in Var2-6 (C, bottom). In all panels, the numbers of the pink and black upward arrows indicate the relative levels of *FLC* transcripts produced from uTSS and mTSS, respectively. Gradient boxes show the total *FLC* expression levels, and graphs show the corresponding flowering phenotypes.

- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
- **METHOD DETAILS**
 - Plasmid construction and generation of transgenic lines
 - 5' RACE
 - 3' RACE
 - 5' end RNA-seq analysis of *FLC* polyadenylation sites and uTSS usage
 - RNA extraction and RT-qPCR
 - Accession numbers
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2023.08.079>.

ACKNOWLEDGMENTS

We thank J. Hepworth for the NIL *vin3-4* genotypes, S. Marquardt for discussion, and the laboratories of C.D. and M. Howard for their critical input and reading of the manuscript. This work was supported by the UK Biotechnology and Biological Sciences Research Council Institute Strategic Program GEN (BB/P013511/1), a Royal Society Professorship to C.D., and a European Research Council Advanced Investigator grant (EPISWITCH-833254).

AUTHOR CONTRIBUTIONS

P.Z. and C.D. conceived the project and wrote the manuscript. P.Z. performed most of the experiments and the data analysis. M.S. analyzed the RNA-seq

data in *Arabidopsis* embryos with the supervision of M.N., and J.Q. generated the A-230G *FRI* transgenic lines. C.D. obtained funding and supervised the work. All authors contributed to the data interpretation and editing of the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 1, 2023

Revised: July 6, 2023

Accepted: August 25, 2023

Published: September 19, 2023

REFERENCES

1. Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631.
2. Shindo, C., Aranzana, M.J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., and Dean, C. (2005). Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiol.* 138, 1163–1173.
3. Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C. (2000). Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* 290, 344–347.

4. Strange, A., Li, P., Lister, C., Anderson, J., Warthmann, N., Shindo, C., Irwin, J., Nordborg, M., and Dean, C. (2011). Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis* accessions. *PLoS One* 6, e19949.
5. Shindo, C., Lister, C., Crevillen, P., Nordborg, M., and Dean, C. (2006). Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalization response. *Genes Dev.* 20, 3079–3083.
6. Jiang, D.H., Gu, X.F., and He, Y.H. (2009). Establishment of the winter-annual growth habit via *FRIGIDA*-mediated histone methylation at *FLOWERING LOCUS C* in *Arabidopsis*. *Plant Cell* 21, 1733–1746.
7. Li, Z., Jiang, D., and He, Y. (2018). *FRIGIDA* establishes a local chromosomal environment for *FLOWERING LOCUS C* mRNA production. *Nat. Plants* 4, 836–846.
8. Choi, K., Kim, J., Hwang, H.J., Kim, S., Park, C., Kim, S.Y., and Lee, I. (2011). The *FRIGIDA* complex activates transcription of *FLC*, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. *Plant Cell* 23, 289–303.
9. Zhu, P., Lister, C., and Dean, C. (2021). Cold-induced *Arabidopsis* *FRIGIDA* nuclear condensates for *FLC* repression. *Nature* 599, 657–661.
10. Geraldo, N., Bäurle, I., Kidou, S., Hu, X., and Dean, C. (2009). *FRIGIDA* delays flowering in *Arabidopsis* via a cotranscriptional mechanism involving direct interaction with the nuclear cap-binding complex. *Plant Physiol.* 150, 1611–1618.
11. Bezerra, I.C., Michaels, S.D., Schomburg, F.M., and Amasino, R.M. (2004). Lesions in the mRNA cap-binding gene *ABA HYPERSENSITIVE 1* suppress *FRIGIDA*-mediated delayed flowering in *Arabidopsis*. *Plant J.* 40, 112–119.
12. Zhang, L., and Jiménez-Gómez, J.M. (2020). Functional analysis of *FRIGIDA* using naturally occurring variation in *Arabidopsis thaliana*. *Plant J.* 103, 154–165.
13. Song, J., Angel, A., Howard, M., and Dean, C. (2012). Vernalization - a cold-induced epigenetic switch. *J. Cell Sci.* 125, 3723–3731.
14. Antoniou-Kourouniotti, R.L., Hepworth, J., Heckmann, A., Duncan, S., Qüesta, J., Rosa, S., Säll, T., Holm, S., Dean, C., and Howard, M. (2018). Temperature sensing is distributed throughout the regulatory network that controls *FLC* epigenetic silencing in vernalization. *Cell Syst.* 7, 643–655.e9.
15. Hepworth, J., Antoniou-Kourouniotti, R.L., Bloomer, R.H., Selga, C., Berggren, K., Cox, D., Collier Harris, B.R.C., Irwin, J.A., Holm, S., Säll, T., et al. (2018). Absence of warmth permits epigenetic memory of winter in *Arabidopsis*. *Nat. Commun.* 9, 639.
16. O'Neill, C.M., Lu, X., Calderwood, A., Tudor, E.H., Robinson, P., Wells, R., Morris, R., and Penfield, S. (2019). Vernalization and floral transition in autumn drive winter annual life history in oilseed rape. *Curr. Biol.* 29, 4300–4306.e2.
17. Buzas, D.M., Robertson, M., Finnegan, E.J., and Helliwell, C.A. (2011). Transcription-dependence of histone H3 lysine 27 trimethylation at the *Arabidopsis* polycomb target gene *FLC*. *Plant J.* 65, 872–881.
18. Yang, H., Berry, S., Olsson, T.S.G., Hartley, M., Howard, M., and Dean, C. (2017). Distinct phases of Polycomb silencing to hold epigenetic memory of cold in *Arabidopsis*. *Science* 357, 1142–1145.
19. Costa, S., and Dean, C. (2019). Storing memories: the distinct phases of Polycomb-mediated silencing of *Arabidopsis FLC*. *Biochem. Soc. Trans.* 47, 1187–1196.
20. Zhao, Y.S., Antoniou-Kourouniotti, R.L., Calder, G., Dean, C., and Howard, M. (2020). Temperature-dependent growth contributes to long-term cold sensing. *Nature* 583, 825–829.
21. Fiedler, M., Franco-Echevarría, E., Schulten, A., Nielsen, M., Rutherford, T.J., Yeates, A., Ahsan, B., Dean, C., and Bienz, M. (2022). Head-to-tail polymerization by VEL proteins underpins cold-induced Polycomb silencing in flowering control. *Cell Rep.* 41, 111607.
22. Xu, C., Wu, Z., Duan, H.C., Fang, X., Jia, G., and Dean, C. (2021). R-loop resolution promotes co-transcriptional chromatin silencing. *Nat. Commun.* 12, 1790.
23. Fang, X., Wu, Z., Raitskin, O., Webb, K., Voigt, P., Lu, T., Howard, M., and Dean, C. (2020). The 3' processing of antisense RNAs physically links to chromatin-based transcriptional control. *Proc. Natl. Acad. Sci. USA* 117, 15316–15321.
24. Wu, Z., Fang, X., Zhu, D., and Dean, C. (2020). Autonomous pathway: *FLOWERING LOCUS C* repression through an antisense-mediated chromatin-silencing mechanism. *Plant Physiol.* 182, 27–37.
25. Hepworth, J., Antoniou-Kourouniotti, R.L., Berggren, K., Selga, C., Tudor, E.H., Yates, B., Cox, D., Collier Harris, B.R.C., Irwin, J.A., Howard, M., et al. (2020). Natural variation in autumn expression is the major adaptive determinant distinguishing *Arabidopsis FLC* haplotypes. *eLife* 9, e57671.
26. Li, P.J., Filiault, D., Box, M.S., Kerdaffrec, E., van Oosterhout, C., Wilczek, A.M., Schmitt, J., McMullan, M., Bergelson, J., Nordborg, M., et al. (2014). Multiple *FLC* haplotypes defined by independent *cis*-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes Dev.* 28, 1635–1640.
27. Li, P.J., Tao, Z., and Dean, C. (2015). Phenotypic evolution through variation in splicing of the noncoding RNA *COOLAIR*. *Genes Dev.* 29, 696–701.
28. Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y.A., Ohler, U., and Zhu, J. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods* 7, 521–527.
29. Morton, T., Petricka, J., Corcoran, D.L., Li, S., Winter, C.M., Carda, A., Benfey, P.N., Ohler, U., and Megraw, M. (2014). Paired-end analysis of transcription start sites in *Arabidopsis* reveals plant-specific promoter signatures. *Plant Cell* 26, 2746–2760.
30. Lu, Z., and Lin, Z. (2019). Pervasive and dynamic transcription initiation in *Saccharomyces cerevisiae*. *Genome Res.* 29, 1198–1210.
31. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38, 626–635.
32. Steinthorsdottir, V., Stefansson, H., Ghosh, S., Birgisdottir, B., Bjornsdottir, S., Fasquel, A.C., Olafsson, O., Stefansson, K., and Gulcher, J.R. (2004). Multiple novel transcription initiation sites for *NRG1*. *Gene* 342, 97–105.
33. Demircioğlu, D., Cukuroglu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N.A., Kahles, A., Lehmann, K.V., Stegle, O., Brazma, A., et al. (2019). A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. *Cell* 178, 1465–1477.e17.
34. Ushijima, T., Hanada, K., Gotoh, E., Yamori, W., Kodama, Y., Tanaka, H., Kusano, M., Fukushima, A., Tokizawa, M., Yamamoto, Y.Y., et al. (2017). Light controls protein localization through phytochrome-mediated alternative promoter selection. *Cell* 171, 1316–1325.e12.
35. Thieffry, A., López-Márquez, D., Bornholdt, J., Malekroudi, M.G., Bressendorff, S., Barghetti, A., Sandelin, A., and Brodersen, P. (2022). PAMP-triggered genetic reprogramming involves widespread alternative transcription initiation and an immediate transcription factor wave. *Plant Cell* 34, 2615–2637.
36. Kurihara, Y., Makita, Y., Kawashima, M., Fujita, T., Iwasaki, S., and Matsui, M. (2018). Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 115, 7831–7836.
37. Chia, M.H., Tresenrider, A., Chen, J.X., Spedale, G., Jorgensen, V., Ünal, E., and van Werven, F.J. (2017). Transcription of a 5' extended mRNA isoform directs dynamic chromatin changes and interference of a downstream promoter. *eLife* 6, e27420.
38. Nielsen, M., Ard, R., Leng, X.Y., Ivanov, M., Kindgren, P., Pelechano, V., and Marquardt, S. (2019). Transcription-driven chromatin repression of intragenic transcription start sites. *PLoS Genet.* 15, e1007969.
39. Chou, S.P., Alexander, A.K., Rice, E.J., Choate, L.A., and Danko, C.G. (2022). Genetic dissection of the RNA polymerase II transcription cycle. *eLife* 11, e78458.

40. Hetzel, J., Duttke, S.H., Benner, C., and Chory, J. (2016). Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci. USA* *113*, 12316–12321.
41. Lou, S.L., Guo, X., Liu, L., Song, Y., Zhang, L., Jiang, Y.Z., Zhang, L.S., Sun, P.C., Liu, B., Tong, S.F., et al. (2022). Allelic shift in *cis*-elements of the transcription factor *RAP2.12* underlies adaptation associated with humidity in *Arabidopsis thaliana*. *Sci. Adv.* *8*, eabn8281.
42. Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.M., Gibson, T.J., and Zaugg, J.B. (2017). CTCF-mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. *Cell Syst.* *5*, 628–637.e6.
43. Le, N.T., Harukawa, Y., Miura, S., Boer, D., Kawabe, A., and Saze, H. (2020). Epigenetic regulation of spurious transcription initiation in *Arabidopsis*. *Nat. Commun.* *11*, 3224.
44. Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J.D., Pai, A., Song, C., Zhuang, M., Klasfeld, S., He, C., et al. (2017). *Cis* and *trans* determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. *Nat. Genet.* *49*, 1546–1552.
45. Yamamoto, Y.Y., Ichida, H., Abe, T., Suzuki, Y., Sugano, S., and Obokata, J. (2007). Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. *Nucleic Acids Res.* *35*, 6219–6226.
46. Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K., and Obokata, J. (2009). Heterogeneity of *Arabidopsis* core promoters revealed by high-density TSS analysis. *Plant J.* *60*, 350–362.
47. Zhang, Z., Dong, J., Ji, C., Wu, Y., and Messing, J. (2019). NAC-type transcription factors regulate accumulation of starch and protein in maize seeds. *Proc. Natl. Acad. Sci. USA* *116*, 11223–11228.
48. Bi, Y., Wang, H., Yuan, X., Yan, Y., Li, D., and Song, F. (2023). The NAC transcription factor ONAC083 negatively regulates rice immunity against Magnaporthe oryzae by directly activating transcription of the RING-H2 gene OsRFPH2-6. *J. Integr. Plant Biol.* *65*, 854–875.
49. Wang, X., Liu, Y., Hao, C., Li, T., Majeed, U., Liu, H., Li, H., Hou, J., and Zhang, X. (2023). Wheat NAC-A18 regulates grain starch and storage proteins synthesis and affects grain weight. *Theor. Appl. Genet.* *136*, 123.
50. Thieffry, A., Vigh, M.L., Bornholdt, J., Ivanov, M., Brodersen, P., and Sandelin, A. (2020). Characterization of *Arabidopsis thaliana* promoter bidirectionality and antisense RNAs by inactivation of nuclear RNA decay pathways. *Plant Cell* *32*, 1845–1867.
51. Kim, D.H., and Sung, S. (2017). Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. *Dev. Cell* *40*, 302–312.e4.
52. Pedersen, A.G., and Nielsen, H. (1997). Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* *5*, 226–233.
53. Kamieniarz-Gdula, K., and Proudfoot, N.J. (2019). Transcriptional control by premature termination: a forgotten mechanism. *Trends Genet.* *35*, 553–564.
54. Colin, J., Libri, D., and Porrua, O. (2011). Cryptic transcription and early termination in the control of gene expression. *Genet. Res. Int.* *2011*, 653494.
55. Schon, M., Baxter, C., Xu, C., Enugutti, B., Nodine, M.D., and Dean, C. (2021). Antagonistic activities of cotranscriptional regulators within an early developmental window set *FLC* expression level. *Proc. Natl. Acad. Sci. USA* *118*. e2102753118.
56. Duncan, S., Holm, S., Questa, J., Irwin, J., Grant, A., and Dean, C. (2015). Seasonal shift in timing of vernalization as an adaptation to extreme winter. *eLife* *4*, e06620.
57. Qüesta, J.I., Antoniou-Kourouniotti, R.L., Rosa, S., Li, P.J., Duncan, S., Whittaker, C., Howard, M., and Dean, C. (2020). Noncoding SNPs influence a distinct phase of Polycomb silencing to destabilize long-term epigenetic memory at *Arabidopsis FLC*. *Genes Dev.* *34*, 446–461.
58. Sheldon, C.C., Conn, A.B., Dennis, E.S., and Peacock, W.J. (2002). Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. *Plant Cell* *14*, 2527–2537.
59. Tian, F., Yang, D.C., Meng, Y.Q., Jin, J., and Gao, G. (2020). PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* *48*, D1104–D1113.
60. Yang, M., Zhu, P., Cheema, J., Bloomer, R., Mikulski, P., Liu, Q., Zhang, Y., Dean, C., and Ding, Y. (2022). *In vivo* single-molecule analysis reveals *COOLAIR* RNA structural diversity. *Nature* *609*, 394–399.
61. Alfonso-Gonzalez, C., Legnini, I., Holec, S., Arrigoni, L., Ozbulut, H.C., Mateos, F., Koppstein, D., Rybak-Wolf, A., Bönisch, U., Rajewsky, N., et al. (2023). Sites of transcription initiation drive mRNA isoform selection. *Cell* *186*, 2438–2455.e22.
62. Signor, S.A., and Nuzhdin, S.V. (2018). The evolution of gene expression *in cis* and *trans*. *Trends Genet.* *34*, 532–544.
63. Wu, Z., Ietswaart, R., Liu, F., Yang, H., Howard, M., and Dean, C. (2016). Quantitative regulation of *FLC* via coordinated transcriptional initiation and elongation. *Proc. Natl. Acad. Sci. USA* *113*, 218–223.
64. Xiao, R., Chen, J.Y., Liang, Z., Luo, D., Chen, G., Lu, Z.J., Chen, Y., Zhou, B., Li, H., Du, X., et al. (2019). Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. *Cell* *178*, 107–121.e18.
65. Tokizawa, M., Kusunoki, K., Koyama, H., Kurotani, A., Sakurai, T., Suzuki, Y., Sakamoto, T., Kurata, T., and Yamamoto, Y.Y. (2017). Identification of *Arabidopsis* genic and non-genic promoters by paired-end sequencing of TSS tags. *Plant J.* *90*, 587–605.
66. Crevillén, P., Sonmez, C., Wu, Z., and Dean, C. (2013). A gene loop containing the floral repressor *FLC* is disrupted in the early phase of vernalization. *EMBO J.* *32*, 140–148.
67. Zhao, Y., Zhu, P., Hepworth, J., Bloomer, R., Antoniou-Kourouniotti, R.L., Doughty, J., Heckmann, A., Xu, C., Yang, H., and Dean, C. (2021). Natural temperature fluctuations promote *COOLAIR* regulation of *FLC*. *Genes Dev.* *35*, 888–898.
68. Liu, F., Quesada, V., Crevillén, P., Bäurle, I., Swiezewski, S., and Dean, C. (2007). The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate *FLC*. *Mol. Cell* *28*, 398–407.
69. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26.
70. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.
71. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* *34*, 525–527.
72. Schon, M.A., Kellner, M.J., Plotnikova, A., Hofmann, F., and Nodine, M.D. (2018). NanoPARE: parallel analysis of RNA 5' ends from low-input RNA. *Genome Res.* *28*, 1931–1942.
73. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* *17*, 10–12.
74. Csorba, T., Questa, J.I., Sun, Q., and Dean, C. (2014). Antisense *COOLAIR* mediates the coordinated switching of chromatin states at *FLC* during vernalization. *Proc. Natl. Acad. Sci. USA* *111*, 16160–16165.
75. Plotnikova, A., Kellner, M.J., Schon, M.A., Mosiolek, M., and Nodine, M.D. (2019). MicroRNA dynamics and functions during *Arabidopsis* embryogenesis. *Plant Cell* *31*, 2929–2946.
76. Box, M.S., Coustham, V., Dean, C., and Mylne, J.S. (2011). Protocol: a simple phenol-based method for 96-well extraction of high quality RNA from *Arabidopsis*. *Plant Methods* *7*, 7.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
NEB 5-alpha Competent <i>E. coli</i> (High Efficiency)	New England BioLabs	Cat #: C2987
<i>Agrobacterium tumefaciens</i> pGV2260	N/A	N/A
Chemicals, peptides, and recombinant proteins		
FirstChoice™ RLM-RACE Kit	ThermoFisher Scientific	Cat #: AM1700M
Cap-Clip™ Acid Pyrophosphatase	Cambridge Bioscience Ltd	Cat #: C-CC15011H
T4 RNA Ligase 1 (ssRNA Ligase)	New England BioLabs	Cat #: M0204S
RNaseOUT RNase Inhibitor	ThermoFisher Scientific	Cat #: 10777019
TURBO DNA-free Kit	ThermoFisher Scientific	Cat #: AM1907
Phenol solution saturated with 0.1 M CIT	Merk Life Science UK Ltd	Cat #: P4682
Superscript IV Reverse Transcriptase	ThermoFisher Scientific	Cat #: 18090050
GoTaq G2 Green Master Mix	Promega UK Ltd	Cat #: M7823
T4 DNA LIGASE	New England Biolabs Ltd	Cat #: M0202S
pGEM-T Easy Vector System	Promega UK Ltd	Cat #: A1360
Lightcycler 480 Sybr Green I Master	Roche Diagnostics Ltd	Cat #: 04887352001
Agarose Molecular Biology Grade	Melford Biolaboratories Ltd	Cat #: A20080
Gateway LR Clonase II Enzyme Mix	ThermoFisher Scientific	Cat #: 11791-020
Experimental models: Organisms/strains		
<i>Arabidopsis</i> : Col-0	Hepworth et al. ²⁵ and Zhao et al. ⁶⁷	N/A
<i>Arabidopsis</i> : Col <i>FRI</i> (<i>FRI-SF2</i> in Col)	Hepworth et al. ²⁵ and Zhao et al. ⁶⁷	N/A
<i>Arabidopsis</i> : <i>vin3-4</i>	Hepworth et al. ²⁵ and Zhao et al. ⁶⁷	N/A
<i>Arabidopsis</i> : Var2-6 NIL	Hepworth et al. ²⁵ and Zhao et al. ⁶⁷	N/A
<i>Arabidopsis</i> : UII2-5 NIL	Hepworth et al. ²⁵ and Zhao et al. ⁶⁷	N/A
<i>Arabidopsis</i> : Lov NIL2	Hepworth et al. ²⁵ and Zhao et al. ⁶⁷	N/A
<i>Arabidopsis</i> : Edi-0 NIL	Hepworth et al. ²⁵ and Zhao et al. ⁶⁷	N/A
<i>Arabidopsis</i> : Var2-6 <i>vin3-4</i>	This paper	N/A
<i>Arabidopsis</i> : UII2-5 <i>vin3-4</i>	This paper	N/A
<i>Arabidopsis</i> : Edi-0 <i>vin3-4</i>	This paper	N/A
<i>Arabidopsis</i> : 35S:: <i>FC</i> A	Schon et al. ⁵⁵	N/A
<i>Arabidopsis</i> : <i>cbp20</i>	Geraldo et al. ¹⁰	N/A
<i>Arabidopsis</i> : <i>cbp20 FRI</i>	This paper	N/A
<i>Arabidopsis</i> : Col-G259T	Li et al. ²⁷	N/A
<i>Arabidopsis</i> : Var-T259G	Li et al. ²⁷	N/A
<i>Arabidopsis</i> : Col <i>FLC FRI</i>	Csorba et al. ⁷⁴	N/A
<i>Arabidopsis</i> : Col <i>FLC fri</i>	Li et al. ²⁷	N/A
<i>Arabidopsis</i> : Col <i>FLC A-230G FRI</i>	This paper	N/A
<i>Arabidopsis</i> : Col <i>FLC A-230G fri</i>	This paper	N/A
Oligonucleotides		
3' RACE adaptor (5'-UUUAAACCGCATC CTTCTCTCTACCTACCATTGACCTGTT-3')	Liu et al. ⁶⁸	N/A
RT primers for distal <i>FLC</i> isoforms (<i>FLC</i> -RT (Intron 3): 5'-CTTTGTAATCAAAGGTGGAG AGC-3'; <i>FLC</i> -RT (Exon 7): 5'-TTGTCCAG CAGGTGACATC-3')	This paper	N/A
RT primers for proximal <i>FLC</i> + distal <i>FLC</i> isoforms (<i>FLC</i> - RT (Intron 1) 5': -AGGGG GAACAAATGAAAACC-3'; <i>FLC</i> - RT (Exon1): 5' -GAGAAGCTGTAGAGCTTGCC G - 3')	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
Plasmid: <i>pSLJ-Col FLC A-230G(A-230G)</i>	This paper	N/A
Software and algorithms		
Microsoft Excel	Microsoft	Microsoft® Official Site
GraphPad Prism 10	GraphPad	https://www.graphpad.com/updates/prism-1000-release-notes
Integrative Genomics Viewer	Robinson et al. ⁶⁹	https://software.broadinstitute.org/software/igv/
STAR v2.7.3a	Dobin et al. ⁷⁰	https://github.com/alexdobin/STAR
Kallisto v0.46.0	Bray et al. ⁷¹	https://pachterlab.github.io/kallisto/
NanoPARE Analysis Tools v1.1	Schon et al. ⁷²	https://github.com/Gregor-Mendel-Institute/nanoPARE
Cutadapt v1.9.1	Martin ⁷³	https://github.com/marcelm/cutadapt

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Caroline Dean (caroline.dean@jic.ac.uk).

Materials availability

The plasmids and transgenic plants generated in this study are available from the [lead contact](#) upon request.

Data and code availability

- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the [method details](#). All qPCR data reported in this paper will be shared by the [lead contact](#) upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Arabidopsis thaliana plants were used in this study. All mutants and transgenic lines were in Columbia (Col) background. The Col *FRI* (*FRI-SF2* in Col), *vin3-4*, *Var2-6 NIL*, *Ull2-5 NIL*, *Lov NIL2* and *Edi-0 NIL* were described previously.^{25,67} The introgressed lines NILs were then crossed with *vin3-4* to generate *Var2-6 vin3-4*, *Ull2-5 vin3-4* and *Edi-0 vin3-4*. *35S:FCA*⁵⁵ and *cbp20*¹⁰ were the same as previously reported. *Col-G259T* and *Var-T259G* were progenies from the parental transgenic plants generated in Li et al.²⁷ and *Col FLC FRI* was described in Csorba et al.⁷⁴ *Col-G259T*, genomic fragment of Col-0 *FLC* with G at +259 substituted by T; and *Var-T259G*, genomic fragment of *Var2-6 FLC* with T at +259 replaced by G (as shown in [Figure 4A](#)).

Arabidopsis thaliana plants were grown at the same conditions as previously described.⁹ Seeds were surface-sterilized and sown on Murashige and Skoog (MS) agar plates without glucose or sucrose and stratified at 4 °C for 3 days. For non-vernalization (NV), seedlings were grown in warm conditions (16 h light, 8 h darkness with constant 20 °C) for 10 days. For cold treatment, seedlings were pre-grown in warm conditions (16 h light, 8 h darkness with constant 20 °C) for 10 days, and then transferred to cold (8 h light, 16 h darkness with constant 5 °C) for long-term vernalization, such as two weeks and four weeks (2WV and 4WV). For embryonic qPCR analysis, mother plants were vernalized for four weeks before flowering and ovules at 5 days after pollination (DAP) (early heart stage) or 11 DAP (mature green stage) were harvested for RNA analysis. For flowering time analysis, plants were pre-grown on plates in warm conditions for 10 days before being transferred from petri dish to soil and were kept in warm conditions for those with *fri*. Four weeks vernalization was applied to those with *FRI* after the 10-day pre-growth. Rosette leaves were counted until bolting (when stems reached 2 cm in height).

METHOD DETAILS

Plasmid construction and generation of transgenic lines

Col FLC carrying *A-230G* substitution was generated by mutagenizing the *pENTRY-FLC prom::FLC* using Megaprimer method of site-directed mutagenesis. *pENTRY-FLC prom::FLC* carrying ~12 kb genomic DNA that comprises ~6 kb of the genomic sequence

of the *FLC* (Col) gene with its flanking natural 5' (~3.5 kb) and 3' (~2.6 kb) sequences. *pENTRY Col FLC A-230G* was subsequently subcloned into a pSLJ-DEST vector (a modified version of the pSLJ series) using Gateway Cloning technology (Thermo Fisher Scientific). The construct was transformed into *flc-2 FRI* plants by floral dipping to generate *Col A-230G FRI*. 15 out of 52 independent transgenic lines had a 3:1 segregation for Basta resistance and were further grown for T3 homozygous lines and used for RNA expression analysis. All of them complemented the early flowering phenotype of *flc-2 FRI* and showed reduced uTSS usage. Three independent lines showing the median or average *FLC* expression level²⁷ were used for further analysis and were crossed with *flc-2 fri* to generate *Col A-230G fri* (numbers with suffix c), of which the F3 and F4 homozygous lines were used for uTSS usage analysis. The primers used in the genotyping were listed in [Table S2](#).

5' RACE

5' RACE was performed as previously described.⁶⁷ Sequences of oligos for 5' RACE are listed in [Table S3](#). At least 48 colonies from each RNA sample were examined by PCR and at least two colonies from each different sized PCR product were sent to Sanger sequencing. The clone numbers sequenced for each TSS site were listed in [Table S4](#).

3' RACE

3' RACE was performed as previously described.⁶⁸ Briefly, 5 µg of total RNA was ligated to 0.5 µg 3' RACE adaptor, a DNA oligonucleotide with the first three nucleotides as UUU and the last nucleotide as 3' inverted deoxythymidine using T4 RNA ligase. Ligated RNA was converted into cDNA using adaptor primer RT by reverse transcriptase ([Figure S2A](#)). Gene-specific primers and adaptor primers were used for nested PCR ([Figure S2A](#)). The PCR products were cloned and sequenced. Sequences of the primers in reverse transcription and nested PCR are listed in [Table S3](#).

5' end RNA-seq analysis of *FLC* polyadenylation sites and uTSS usage

FLC RNA 5' and 3' ends during embryogenesis were analyzed in previously published Smart-seq2 and nanoPARE RNA-seq datasets, downloaded from the Gene Expression Omnibus (GEO) under the series accessions GSE132066 (Col-0 time series⁷⁵) and GSE166728 (early heart stage mutants⁵⁵). Quantification was performed as previously described for end-labeled Smart-seq2⁵⁵ and nanoPARE,⁷⁵ respectively.

RNA extraction and RT-qPCR

Unless otherwise stated, seedlings were sampled and RNA was extracted with the hot phenol method followed by lithium chloride precipitation.⁷⁶ Genomic DNA contamination was removed by TURBO DNA-free kit (Invitrogen, AM1907) following the manufacturer's guidelines. cDNA was synthesized with SuperScript IV reverse transcriptase (Invitrogen, 18090050) using gene-specific primers. For analysing the ratio of distal *FLC*: (proximal *FLC* + distal *FLC*) in embryos, two RT-qPCR were conducted separately. For detection of distal *FLC* isoforms the primers FLC-RT (Intron 3) and FLC-RT (Exon 7) are used for RT reaction; for proximal *FLC* + distal *FLC* the primers are used for RT reaction. These primers are listed in [key resources table](#). All the primers in other analyses are listed in [Table S5](#). Standard reference genes *UBC* and *PP2A* for gene expression were used for normalization.

Accession numbers

The genes mentioned in this article have been given the following accession numbers by the Arabidopsis Information Resource (TAIR) database: *FLC*, AT5G10140; *FRI*, AT4G00650; *COOLAIR*, AT5G01675; *VIN3*, AT5G57380; *FCA*, AT4G16280; *CBP20*, AT5G44200; *UBC*, AT5G25760; *PP2A*, AT1G13320. The accession numbers for the CAGE-seq⁵⁰ and TSS-seq³⁸ datasets are Gene Expression Omnibus (GEO): GSE136356 and GEO: GSE113677.

QUANTIFICATION AND STATISTICAL ANALYSIS

All the statistical analyses in this study were performed by GraphPad Prism 10. All the statistical details of experiments, including the statistical tests used, the exact values of sample number, and precision measures, can be found in the figure legends. All comparisons and *P* values were indicated in the figures.