The CPSF phosphatase module links transcription termination to chromatin silencing

Eduardo Mateo-Bonmatí[1,3], Xiaofeng Fang[1,4], Robert Maple[1], Marc Fiedler[2], Lori A Passmore[2], and Caroline Dean[1]*

[1]John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom;

[2]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom;

[3]Present address: Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM) – Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)/CSIC, 28223 Pozuelo de Alarcón, Madrid, Spain;

[4]Present address: Centre for Plant Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China.

*Correspondence to: Caroline Dean (caroline.dean@jic.ac.uk).

Keywords: RNA processing, chromatin, transcription, *FLC*, *COOLAIR*, noncoding RNA, CPSF

ORCID (E.M.-B.): 0000-0002-2364-5173 @emateobonmati
ORCID (X.F.): 0000-0003-1161-9186 @XiaofengFangLab
ORCIDC (R.M.): 0000-0002-9175-5024 @Bobby_Mapes
ORCID (M.F.): 0000-0003-4269-7873
ORCID (L.P.): 0000-0003-1815-3710 @lapassmore
ORCID (C.D.): 0000-0002-6555-3525 @CarolineDeanLab

Word count (total): 9214

ABSTRACT

RNA-mediated chromatin silencing plays important roles in development and environmental response in many organisms. However, it is still unclear how the RNA, and/or the transcriptional processes producing them, trigger chromatin silencing. Through the study of developmental timing in plants we have found co-transcriptional RNA processing modulates a chromatin regulatory mechanism controlling quantitative expression of Arabidopsis *FLOWERING LOCUS C* (*FLC*). Here, we show that the phosphatase module of the mRNA 3' processing CPSF complex is central to this mechanism. The phosphatase module components APRF1 (yeast Swd2/human WDR82), TOPP4 (Glc7/PP1) and LD (which we show is functionally equivalent to Ref2/PNUTS) all function genetically downstream of the CPSF-mediated cleavage and polyadenylation and are necessary for chromatin modification that results in *FLC* transcriptional silencing. LD has been previously shown to stoichiometrically co-associate *in vivo* with the histone demethylase FLD. We show both proteins result in removal of H3K4 monomethylation in the central section of the *FLC* locus. This links transcription termination to delivery of chromatin modifications that then influence subsequent transcription. Thus, physical association of the CPSF phosphatase module with chromatin modifiers generates a transcription-

coupled silencing mechanism. The conservation of the factors involved suggest this mechanism may be generally relevant to transcription-mediated chromatin silencing.

INTRODUCTION

How chromatin and transcription interact to regulate gene expression is a major area of interest. Extensive feedback mechanisms make this interaction difficult to dissect. Chromatin affects co-transcriptional processes that influence RNA Polymerase II (RNA Pol II) elongation rate, which can lead to alternative splicing and early termination/processivity changes i.e., the likelihood of transcription reaching the end of the gene. In turn, these co-transcriptional processes feed-back to influence the chromatin state (Berry *et al.* 2015, Muniz *et al.* 2021, Zumer *et al.* 2021). One locus where this complexity has been studied in detail is Arabidopsis *FLC*. *FLC* encodes a floral repressor and quantitative variation in *FLC* transcription has been an important determinant in adaptation of Arabidopsis accessions to a wide range of climates. For example high *FLC* expression necessitates an overwintering reproductive strategy through winter-induced epigenetic silencing of *FLC* (Menon *et al.* 2021). In contrast, low *FLC* expression leads to a rapid-cycling reproductive strategy, enabling multiple generations a year in some climates.

There has been extensive analysis of the components regulating *FLC* transcription. Forward genetic screens identified repressors that include RNA binding proteins, RNA 3' processing factors and chromatin modifiers (Wu *et al.* 2020). These factors have been found to promote proximal termination of both the antisense transcript *COOLAIR*, but also the sense *FLC* transcript in the embryo (Schon *et al.* 2021). Suppressor genetics and chromatin analysis showed that these co-transcriptional activities function through the chromatin modifiers FLD (a H3K4 demethylase)-LD (TFIIS domain protein)-SDG26 (a SET domain protein) (Liu *et al.* 2007, Fang *et al.* 2020) to suppress *FLC*. Proximal termination induces histone H3K4 demethylation across the locus, delivered by FLD, LD and SDG26. These three factors tightly associate with each other *in vivo* based on proteomic interactions, physically associate with Cleavage and Polyadenylation Specificity Factor (CPSF; CPF in yeast) factors after cross-linking (Fang *et al.* 2020), and together reduce H3K4me1 and H3K36me3 accumulation at *FLC*. This antagonizes transcription enabling a switch to PRC2-mediated H3K27me3 accumulation, so reducing transcriptional initiation and elongation rates (Wu et al 2016). The *FLC* silencing mechanism thus involves co-transcriptional processes that mediate chromatin modifications, which in turn feedback to reinforce the co-transcriptional processing. The factors involved are all evolutionarily conserved, raising the likelihood that this mechanism will be broadly relevant for gene regulation in many organisms; for example, there are parallels with mammalian aging processes (Gyenis *et al.* 2023).

A key question that remains is how proximal termination delivers a changed chromatin environment. Here, we used proteomic immunoprecipitation experiments to identify additional partners of the FLD-LD-SDG26 complex. We find robust interaction with Arabidopsis APRF1 and TOPP4. These proteins are homologous to different members of the CPF phosphatase module (Swd2 and Glc7) that associates with RNA Pol II and mediates transcription termination and RNA polyadenylation in yeast (Carminati *et al.* 2022). We use molecular and genetic tools to show that these components function in the same co-transcriptional pathway with the FLD-LD-SDG26 complex to alter *FLC* chromatin environment, and thus transcriptional output. Upon close examination of all the phosphatase module components we also show that the LD component of the complex is structurally related to Ref2 in yeast and PNUTS in mammals; they all contain intrinsically disordered regions and a TFIIS domain. Ref2/PNUTS act in yeast/mammalian complexes as the regulatory subunit of the phosphatase module that influences phosphorylation of the RNA Pol II and its partners via the Glc7/PP1 phosphatase. This activity promotes transcription termination through phosphorylation-dependent effects on RNA Pol II elongation/processivity (Rodríguez-Molina *et al.* 2023). We show that LD-APRF1-TOPP4 have similar functional consequences promoting proximal termination of the antisense *COOLAIR* transcripts. Thus, our proteomic and genetic dissection identifies many factors of this floral repression pathway as conserved members of the 3' CPSF phosphatase module. This clarifies the overall mechanism of how RNA processing is linked to chromatin silencing at Arabidopsis *FLC*. The CPSF phosphatase module directly links RNA processing with histone demethylase activity to alter the local chromatin environment as the RNA Pol II transcribes downstream of the polyadenylation site. This chromatin environment reinforces proximal termination choice so providing the molecular feedback necessary to stably maintain a low transcription state. Our accompanying paper describes a mathematical model of this co-transcriptional regulatory mechanism revealing how co-transcriptional 3' processing sets the level of productive transcription via H3K4 monomethylation, thus dictating the degree of antagonism to Polycomb silencing, which then determines the rate at which Polycomb repression is established (Menon *et al.* 2023). Overall, our studies shed light on the tight interconnections/feedbacks between co-transcriptional RNA processing, chromatin environment and transcriptional mode. These have proven difficult to dissect at the molecular level but are likely to be at the basis of most epigenetic switching mechanisms.

RESULTS

The FLD complex robustly associates with the homolog of yeast Swd2.

We previously reported that the histone demethylase homolog FLD (FLOWERING LOCUS D) associates with LD (LUMINIDEPENDENS) and SDG26 (SET DOMAIN GROUP 26) *in vivo*; each tagged version of these three proteins stoichiometrically enrich the other two partners in co-immunoprecipitation experiments (Fang *et al.* 2020). Each of these proteins also co-

immunoprecipitated with ANTHESIS PROMOTING FACTOR1 (APRF1; Figure 1A). This was also found in another analysis of proteins that co-immunoprecipitate with the FLD complex [Figure 1B; Qi et al. (2022)].

APRF1 is a WD40-repeat protein encoded by At5g14530. We obtained a T-DNA insertion line [WiscDsLox_489; *aprf1-9*; Kapolas et al. (2016)] and analysed the flowering time and *FLC* expression. Both *FLC* spliced and unspliced transcripts were significantly upregulated in the mutant line, and accordingly, *aprf1-9* plants were late flowering (Figure 1C-E). Given that the insertion is predicted to interrupt the 9th of 10 exons, the plants might retain some APRF1 function. To address this, we designed a CRISPR-Cas9 transgene to generate a mutant null for *APRF1*. Using a sgRNA targeting the second exon and screening for edited plants, we found transgene-free T_2 plants carrying a 5-nt deletion which creates an in frame early stop codon (Figure S1). This new allele was named *aprf1-10* and was as late flowering as *aprf1-9* (Figure 1D). We crossed *aprf1-9* to *aprf1-10* and analysed *FLC* expression levels of the single and heterozygous mutants. All of them had similarly upregulated *FLC*, confirming their allelism and the role of APRF1 on *FLC* repression (Figure 1E).

APRF1 is one of the two Arabidopsis orthologs of the *Saccharomyces cerevisiae* Swd2 [(Figure S2A)(Fiorucci *et al.* 2019)]. Swd2 has been found to play a role in two very different complexes; it is one of the WD40 components of the COMPlex of Proteins Associated with Set1 (COMPASS). This complex catalyses the co-transcriptional deposition of H3K4me3 and Swd2 has been shown to play a key role in the interaction between Set1 and the RNA Pol II Carboxy Terminal Domain (CTD) (Bae *et al.* 2020). Swd2 has also been found as a structural component of the phosphatase module of the Cleavage and Polyadenylation Factor (CPF; CPSF in higher eukaryotes), which signals transcriptional termination (Casanal *et al.* 2017). This apparently opposite role on transcription in yeast made us test if the Arabidopsis orthologs *APRF1* (also known as *Swd2-like A/S2LA*) and *S2LB* have functionally diverged. *FLC* spliced and unspliced levels were assayed in an insertional allele of *S2LB* (Figure S2B) and the double mutant *aprf1-9 s2lb*. Opposite to *aprf1-9*, *s2lb* showed a significant reduction in *FLC* expression, while the double mutant shows up-regulated *FLC* indistinguishable from the *aprf1-9* single mutant (Figure S2C). These contrasting phenotypes suggest that in Arabidopsis, the ancestral Swd2 may have been sub-functionalized, with S2LB working through the COMPASS complex and therefore activating *FLC*, and APRF1 working with the 3' processing machinery to silence the locus (Figure S2D).

FLD-mediated repression of *FLC* requires H3K4me1 removal (Fang *et al.* 2020, Inagaki *et al.* 2021). Considering the tight link between the FLD complex and APRF1, we performed chromatin immunoprecipitation coupled with quantitative PCR (ChIP-qPCR) to quantify H3K4me1 at *FLC*. We found a H3K4me1 increase across the locus in *aprf1-10* compared to Col-0 (Figure 1G). H3K4me1 is thought to be read by SDG8 (Liu and Huang 2018), a histone methyltransferase that deposits

H3K36me3, an active histone modification, which is mutually exclusive to the repressive PRC2-deposited H3K27me3 (Yang *et al.* 2014). Consistent with the *FLC* upregulation and the increased H3K4me1 levels in *aprf1-10*, ChIP-qPCR analyses also showed that H3K36me3 and H3K27me3 were upregulated and downregulated, respectively. Thus, APRF1 has a role in establishing a silent chromatin status at *FLC* (Figure 1H, I).

Working upstream of the FLD complex, FCA promotes proximal termination on both strands of *FLC* (Simpson *et al.* 2003, Schon *et al.* 2021). To obtain genetic evidence of the relationship between *APRF1* and *FLD, LD* or *FCA*, we crossed *aprf1-9* with mutants in subunits of the FLD complex such as *fld-4* or *ld-1*, and the core component of the pathway *fca-9*. *FLC* levels in the double mutants revealed an epistatic relationship between *FLD, LD* or *FCA* and *APRF1* (Figure 1J, K), further demonstrating the genetic connection between *APRF1* and the *FLC* silencing machinery. To study the effects of loss of *APRF1* in a high transcriptional environment, but without perturbing the FCA pathway, we introgressed a functional *FRIGIDA* (*FRI*) into *aprf1-9*. FRI functions as an *FLC* transcriptional activator, and like other anti-terminators (Gregersen *et al.* 2019), promotes distal polyadenylation of both *FLC* and *COOLAIR*, thus antagonizing the co-transcriptional repression mechanism (Schon *et al.* 2021). Levels of *FLC* in *aprf1-9 FRI* were the same as Col*FRI*, likely due to an overwriting effect of *FRI* compared to the relatively modest *FLC* upregulation of *aprf1-9* (Figure 1L, M). Finally, we crossed *aprf1-10* into the C2 background which harbours a $35S_{pro}$*:FCA* transgene in the presence of a functional *FRI*. *aprf1-10* significantly released the $35S_{pro}$*:FCA* transgene-mediated *FLC* silencing convincingly demonstrating a role for APRF1 downstream of FCA (Figure 1N, O).

APRF1, LD, and TOPP4 form a plant equivalent of a CPSF phosphatase sub-module.

To further characterize the role of APRF1 on *FLC* silencing we generated a translational fusion of APRF1 ($APRF1_{pro}$*:APRF1-VENUS*) and performed CrossLinked Nuclear ImmunoPrecipitation and Mass Spectrometry (CLNIP-MS) (Fang *et al.* 2020), detecting LD in the pulldown, thus confirming their reciprocal interaction (Figure 2A). We also detected a significant number of peptides for TYPE ONE SERINE/THREONINE PROTEIN PHOSPHATASE4 (TOPP4) (Figure 2A). These interactions were recently validated by independent experiments of differentially tagged lines [(Qi et al. 2022; Figure 2B)].

TOPP4 shares high sequence and structural homology with Glc7, another member of the CPF phosphatase module in yeast (Figure S3A). The similarity was even higher than that showed by APRF1 and Swd2 (Figure S4). Identifying homologues of two members of the CPF phosphatase module prompted us explore further parallels. $Swd2^{APRF1}$ and $Glc7^{TOPP4}$ are tightly associated with each other in yeast, and to Ref2 (Casanal *et al.* 2017). Ref2 is key for the interaction between CPF and

the RNA Pol II and behaves as a phosphatase regulatory subunit, potentially providing substrate specificity to the phosphatase Glc7 (Nedea *et al.* 2008, Carminati *et al.* 2022). Ref2 and its putative metazoan ortholog PNUTS, are largely disordered proteins apart from a TFIIS motif at the N terminal region of the protein which consists of a compact four-helix bundle. Intriguingly, despite its apparent key role in an otherwise highly conserved 3' processing machinery, Ref2 shows no obvious homology with an Arabidopsis protein. We performed BLAST searches to find potential orthologs in Arabidopsis for PNUTS, the counterpart of Ref2 in mammals. BLAST algorithms showed the best hit for PNUTS in Arabidopsis is LD (Figure S5). We then aligned the sequence of Ref2, PNUTS and LD. The degree of conservation is low (Figure S6A), although slightly improved when aligning only LD and PNUTS (Figure S7). However, all of them share similar features, a TFIIS domain in the N-terminal region of the protein and a highly disordered region (Figure 2D, S6B-D). PNUTS is known to provide substrate specificity to PP1 phosphatases, like Glc7 and TOPP4 (Figure S3), and to interact with WDR82, a close ortholog of Swd2/APRF1 (Figure S4). Further exploring these parallels, we carried out an *in-silico* prediction using AlphaFold2 of the interaction between APRF1-LD and their yeast and human counterparts (Figure 2E). Despite the low sequence homology, predictions support conservation of structural features and contact points between APRF1/Swd2/WDR82 and LD/Ref2/PNUTS supporting their functional equivalence (Figure 2F). Taken together, the robust immunoprecipitation of LD and TOPP4 by APRF1, the structural parallels between Ref2-PNUTS-LD and the similar roles in co-transcriptional processing lead us to propose that LD, APRF1, and TOPP4 may form a functional CPSF phosphatase module in plants.

FLD complex and RNA Pol II co-occupy *FLC* chromatin independently of FCA function.

Available genome-wide data shows FLD binding correlates with actively transcribed genes and elongating RNA Pol II CTD phosphorylated at Ser2 or Ser5 (Inagaki *et al.* 2021). RNA Pol II occupancy in the Arabidopsis genome often shows peaks near transcription termination sites (TTS), potentially linked to slow co-transcriptional termination events (Wu *et al.* 2016). We found this was the case for *FLC* in Col*FRI* vs Col-0 (Figure 3A) and has been shown to be the case in *fca-9* or *fld-4* (Wu *et al.* 2016). To address co-occupancy with FLD, we performed ChIP-qPCR using a transgenic FLAG-tagged FLD with and without the transcriptional activator FRI, and non-transgenic control plants. Agreeing with the reported data (Inagaki *et al.* 2021), FLAG-FLD showed high enrichment at the 3' region of *FLC* in a *FRI* genotype compared to Col-0 (*fri)*, with the latter close to the background signal (Figure 3B). To rule out the possibility that this enrichment was an effect of the transcriptional activator FRI and not the high transcription itself, we introgressed the FLAG-FLD transgene into the *fca-9* background. FLD enrichment at the 3´end of *FLC* in *fca-9 FLAG-FLD* was indistinguishable from *FRI FLAG-FLD* confirming FLD enrichment primarily associated to transcriptional activity (Figure 3B). *FLD* functions genetically downstream of *FCA* (Liu *et al.* 2007) so it was interesting that FLD-Pol II co-occupancy association was not affected by *fca-9*. Thus, even

when FLD is located at *FLC*, it cannot function properly without FCA-mediated 3' processing of the nascent transcript.

We introgressed a GFP-tagged LD (Fang *et al.* 2020) into Col*FRI* to create a line with high *FLC* transcription and performed ChIP-qPCR experiments. LD enrichment was only detected in the line where *FLC* transcription is high, peaking at the *FLC* TTS (Figure 3C). The shared enrichment of elongating Pol II, FLD- and LD is consistent with an *in vivo* association of the transcriptional machinery and the FLD complex as suggested by genome-wide data (Inagaki *et al.* 2021). Thus, we propose that FLD and LD associate with RNA Pol II, and therefore work co-transcriptionally, fitting with LD working as a CPSF equivalent to Ref2 and PNUTS.

Inefficient termination in *aprf1* mutants leads to transcriptional read-through.

Previous work on the FCA pathway found that most of the factors involved were directly affected on *FLC* antisense transcription (Wu *et al.* 2020). To pursue a potential role of LD-APRF1-TOPP4 as a CPSF phosphatase module we analysed the transcription and processing of *COOLAIR*. In contrast to *FLC* levels which did not show any significant difference (Figure 1J, K), total *COOLAIR* levels were upregulated in double mutants containing *aprf1-9* (Figure 4A, B). This was particularly striking for the *aprf1-9 FRI* combination. *COOLAIR* transcripts are generally polyadenylated at either a cluster of proximal sites (Class I) or a cluster of distal sites coincident with the *FLC* promoter (Class II) (Liu *et al.* 2010). A moderate increase in proximal *COOLAIR* (Figure 4C) and no change in distal *COOLAIR* were found in *aprf1-9 FRI* (Figure 4D), with the proximal / distal ratio increased in *aprf1-9* double mutant combinations compared to the single mutant (Figure 4E), unlike other mutations affecting the FCA pathway (Xu *et al.* 2021a). However, like *PEP1* in *Arabis alpina* (Castaings *et al.* 2014), we had also detected low abundance spliced *COOLAIR* transcripts polyadenylated at a medial site (Liu and Dean, unpublished), which we had termed *COOLAIR* Class III (Figure 4A). This isoform was significantly up-regulated specifically in *aprf1-9 FRI* (Figure 4F), with one specific spliced variant *COOLAIR* Class III.3 becoming the most abundant isoform, enriched in high transcription situations such as *fca-9* or Col*FRI* (Figure 4G, S8). We interpret the high levels of the *COOLAIR* class III isoform as reflecting inefficient polyadenylation/transcript termination at the proximal site. To analyse the polyadenylation site choice in an unbiased and strand-specific manner in these two genotypes, we carried out a Quant-seq analysis, which detects polyadenylation events, of Col*FRI* and *aprf1-9 FRI*. For the *FLC* (sense) transcripts there were a few reads potentially indicating transcriptional readthrough events (Figure S9). However, the differences for *COOLAIR* were more dramatic with reads compatible with the use of the medial poly A site, others pointing to distal *COOLAIR* readthrough, and even alternative *COOLAIR* transcriptional start sites, potential consequence of *FLC* sense readthrough (Figure 4H, S9).

To understand which were likely to be co-transcriptional changes we analysed chromatin-bound RNA (chRNA) of Col-0, *aprf1-9*, Col*FRI* and *aprf1-9 FRI*. Not unexpectedly, the *FLC* (sense) chRNA levels of *aprf1-9* were higher than Col-0, as *aprf1-9* compromises *FLC* silencing. PCR amplicons for sense *FLC* levels in Col*FRI* and *aprf1-9 FRI* gave similar values at the 5' and 3' ends of the locus, but there were differences in the centre, pointing to complex effects of APRF1 on *FLC* Pol II processivity (Figure 5A). However, the most striking differences were on the *COOLAIR* strand (antisense) (Figure 5A). RNA was detected in regions corresponding to the 5' half of *COOLAIR* in Col*FRI*, but in *aprf1-9* genotypes (*fri* and *FRI*) this signal extended to regions that cover both Class I and Class III *COOLAIR* transcripts. This likely represents antisense transcriptional readthrough, supporting a role for APRF1 as part of the Pol II termination machinery (Figure 5A). Overall, the loss of *APRF1* appears to affect the ability of CPSF to facilitate transcription termination leading to transcriptional readthrough.

LD and APRF1 affect RNA Pol II CTD Ser5 hyperphosphorylation

Different mechanistic models, not mutually exclusive, have been proposed for transcription termination (Mo *et al.* 2021, Rodríguez-Molina *et al.* 2023). Current thinking based on studies in different organisms is that CPSF-RNA Pol II recognition of polyadenylation site (PAS) triggers the nascent RNA cleavage. This generates a free 5' end on the cleaved RNA which is a substrate for 5'-3' ribonucleases such as XRNs that degrade the nascent transcript and eventually dislodge RNA Pol II from the chromatin (Cortazar *et al.* 2019, Rodríguez-Molina *et al.* 2023). Transcription of the PAS also triggers a conformational change in RNA Pol II, potentially driven by dephosphorylation of the CTD and or co-factors, which slows-down RNA Pol II. To ascertain if *aprf1* mutants affect RNA Pol II CTD modifications, we carried out ChIP-qPCR experiments using antibodies targeting the RNA Pol II CTD phosphorylated residues Ser2, Ser5 and Tyr1 comparing *aprf1-10* with Col-0. Not surprisingly, all the modified RNA Pol II were detected at higher levels in *aprf1-10*, since the locus is more actively transcribed (Figure 5B, S10A). We repeated our analysis comparing Col*FRI* and *aprf1-9 FRI*, given that both genotypes have similar *FLC* RNA levels. In this case, RNA Pol II Ser2P and Tyr1P levels were identical (Figure 5C, S10B), while there was a slight increase in RNA Pol II Ser5P towards the 5' end of the locus (Figure 5C). FRI forms a super-molecular complex recruiting transcriptional activators and repressing proximal termination so may mask APRF1 direct effects on RNA Pol II phosphorylation status (Li *et al.* 2018, Zhu *et al.* 2021). Thus, to investigate direct effects of APRF1-LD (a potential CPSF phosphatase module) on RNA Pol II CTD phosphorylation we compared the FRI mutant Col-0 [which is a *fri* null (Johanson *et al.* 2000)], *ld-1* and *aprf1-9 ld-1*. Both, *ld-1* and *aprf1-9 ld-1* showed similarly high *FLC* expression levels (Figure 1J, K), similar levels of total Pol II, Ser2P and Tyr1P (Figure 5D, S10C), but *aprf1-9 ld-1* showed higher RNA Pol II Ser5P (Figure 5D). This suggests a role for the Arabidopsis CPSF phosphatase module in removal of RNA Pol II Ser5P at *FLC*, however, we cannot rule out the possibility that the CTD Ser5 phosphorylation

effects are indirect, through phosphorylation of elongation factors such as SPT5 (Cortazar *et al.* 2019).

DISCUSSION

The study of developmental timing in plants has led to mechanistic dissection of chromatin silencing mechanisms at the gene encoding the floral repressor Arabidopsis FLC. Since quantitative variation of *FLC* expression affects the reproductive strategy of Arabidopsis any molecular variation can be subject to strong evolutionary selection. *FLC* is thus an excellent system to dissect RNA processing, chromatin regulation, and their interconnections and molecular feedbacks that generate low and high transcriptional states that underpin adaptively important variation in transcriptional output.

The work here identifies the importance of the CPSF phosphatase module, a regulator of RNA Pol II, in *FLC* regulation. We identify LD, a protein characterized as a flowering regulator nearly 20 years ago as part of the CPSF phosphatase module, functionally equivalent to yeast Ref2 and metazoan PNUTS proteins. Given its interaction partners, LD thus acts as a bridge between chromatin modifiers and the Pol II-machinery (Figure 6). We propose that actively transcribed *FLC* chromatin is poised with H3K4me1, which promotes RNA Pol II processivity, i.e. the likelihood of transcription reaching the end of the gene. Any pause in RNA Pol II functioning, for example coincident with the formation of the R-loop during *COOLAIR* transcription, would stimulate the 3' processing machinery (carried along in the RNA Pol II supercomplex) to polyadenylate the transcript at the proximal site. PAS recognition would be signalled by the CPSF phosphatase module to Pol II via dephosphorylation of either elongating factors or the CTD or both and trigger a conformational change on RNA Pol II (as proposed by Carminati *et al.* 2022) to activate FLD function. RNA Pol II downstream of the PAS would proceed slowly due to the conformational change, causing FLD to remove H3K4me1 co-transcriptionally until the RNA Pol II is terminated by 5'-3' XRNs. This creates a less processive chromatin environment for subsequent rounds of transcription. Each round of proximal polyadenylation-termination-H3K4me1 removal will reinforce the next round of transcription, creating an intrinsic feedback loop in the mechanism (Menon *et al.* 2023). Intriguingly, our chRNA data detected very clear readthrough effects on the antisense (*COOLAIR*) strand, but also in certain parts of the sense (*FLC*) transcription unit when *APRF1* was disrupted. Proximal termination of *FLC* sense transcription during early embryo development is associated with establishment of the silenced state (Schon *et al.* 2021). Proximally polyadenylated sense transcripts do not accumulate in seedling tissue, but it is possible that these are particularly sensitive to RNA degradation pathways. We thus envisage the same mechanism functioning on both *FLC* and *COOLAIR* transcription to both establish and then maintain the transcriptionally silenced state, with the conserved CPSF phosphatase factors central to that chromatin silencing mechanism.

Arabidopsis genome-wide data indicates that FLD shows a clear enrichment at sites of convergent transcription (Inagaki *et al.* 2021) and CPSF factors have been shown to resolve DNA transcription/replication conflicts in the Arabidopsis genome (Baxter *et al.* 2021). Thus, FCA and the CPSF may constitute a plant equivalent to the metazoan Integrator-PP2C (INTAC) (Hu *et al.* 2023, Wagner *et al.* 2023). Indeed, FCA and FPA, have been previously shown to play a role in defining the transcriptional boundaries, with transcriptional readthrough in hundreds of genes detected in the double mutant *fca fpa* (Sonmez *et al.* 2011).

The evolutionary variation in the co-transcriptional complexes provides regulatory flexibility. *APRF1* and *S2LB* constitute an unusually clear example of gene sub-functionalization. In yeast, Swd2 is a single copy gene whose product was found in two protein complexes with antagonistic functions: COMPASS (Bae *et al.* 2020) and the mRNA 3'-end processing machinery (Casanal *et al.* 2017). Likely due to this dual role, Swd2 is the only essential COMPASS component in yeast (Soares and Buratowski 2012). S2LB was found to interact *in vivo* with the methyltransferase SDG2 and the structural component of COMPASS WDR5, controlling H3K4me3 levels genome-wide at thousands of gene promoters (Fiorucci *et al.* 2019). In contrast, APRF1 interacts with LD as well as the other components of the FLD complex, and with TOPP4, functionally equivalent to the CPF phosphatase, Glc7 (Fang *et al.* 2020, Qi *et al.* 2022). The functional divergence suggested by the differential *in vivo* association was confirmed through analysis of mutants affecting *FLC* expression. Like other mutants in the Arabidopsis COMPASS machinery (Jiang *et al.* 2011), *s2lb* mutants reduce *FLC* expression due to their impaired ability to deposit H3K4me3 at the locus and promote transcriptional initiation.

The accompanying paper (Menon *et al.* 2023) describes a theoretical analysis producing a dynamic mathematical model of this co-transcriptional mechanism. The model quantifies the molecular feedbacks linking co-transcriptional processing, chromatin modification and transcriptional output in an interlinked three-way association that sets the timescale for the slow switch into the digital Polycomb silenced state. The model explains how the low transcriptional state influenced by proximal termination and H3K4 demethylation reduces the antagonism to Polycomb silencing and leads to a stable PRC2 epigenetic switch, with sufficient feedbacks to maintain the silenced state through DNA replication and cell division. We have generated an animation (weblink) to help explain these molecular feedbacks, and how this mechanism leads to the stable expression of *FLC* in one of two stable expression states.

This transcription-coupled silencing mechanism is likely to be generally relevant to heterochromatin establishment in genomes of many eukaryotic organisms. A clear genome-wide correlation has been found between human FLD and APRF1 homologs (LSD1 and WDR82) and Pol II (Kim *et al.* 2022, Estell *et al.* 2023), and an unbiased search for LSD1 regulators in human cells found a direct

relationship between the histone demethylase, and an RNA helicase involved in R-loop resolution and PRC2-mediated silencing (Pinter *et al.* 2021). Indeed, 3' processing has been extensively linked to chromatin silencing in *S. pombe* (Kowalik *et al.* 2015, Grewal 2023). Continued mechanistic dissection is therefore likely to elaborate concepts important in chromatin silencing generally.

MATERIALS AND METHODS

Plant material and growth conditions

All the plants were homozygous for the indicated genotype. The *aprf1-9* mutant seeds (WiscDsLox489-492K11, N858279) were obtained from NASC (Nottingham Arabidopsis Stock Centre). The *s2lb* mutant and the *aprf1-9 s2lb* double mutant, previously described in (Fiorucci *et al.* 2019), were kindly provided by Fredy Barneche. Transgenic *fld-4*; FLD_{pro}*:3xFLAG-FLD* were previously described in (Inagaki *et al.* 2021) and shared by Soichi Inagaki.

The *aprf1-10* mutant harbours a 5-nt deletion in *APRF1*. The *aprf1-10* deletion creates a new target for the MaeII (target ACGT) restriction enzyme. To generate the *aprf1-10* we employed the CRISPR/Cas9 plasmid pKI1.1R following the protocol described (Tsutsui and Higashiyama 2017). Briefly, pKI1.1R plasmid (Addgene #85808) was linearized by incubating 1.5 µg of the plasmid with AarI restriction enzyme for 16 h, and then dephosphorylated using the alkaline phosphatase rAPid (Roche). A target-specific gRNA was designed using CRIPR-P 2.0 (http://crispr.hzau.edu.cn/CRISPR2). Oligonucleotides harbouring the gRNA target (sgRNA_APRF1_F and sgRNA_APRF1_R; Table S1) were hybridised by slow cooling down from 95-25°C and then phosphorylated using the T4 Polynucleotide Kinase (NEB). The digested plasmid and the hybridised oligonucleotides were ligated using the T4 ligase (NEB) and then transformed in *Escherichia coli* HST08 competent cells (Takara). The sequence integrity of inserts carried by transformants were verified by Sanger sequencing. The plasmid was then transfer to *Agrobacterium tumefaciens* C58C1 strain by electroporation. T_1 plants carrying the construct were selected on MS media supplemented with 15 µg/ml of Hygromycin. Next generation plants were counter-selected to find transgene-free individuals carrying the homozygous mutation. The *aprf1-9* $APRF1_{pro}$*-APRF1-mVENUS* transgenic was generated by cloning the *APRF1pro:APRF1* genomic sequence without the stop codon upstream of the coding sequence of *mVENUS* at pCAMBIA1300. Vector was mobilized to Agrobacterium C58C1 and transform to *aprf1-9* by floral dipping. T_1 transformant plants were selected on Hygromycin.

Seeds were surface sterilized in 40 % v/v commercial bleach for 10 min and rinsed 4 times with sterile distilled water. Seeds were then sown on standard half-strength Murashige and Skoog (MS) medium (0.22% MS, 8% plant agar) media plates and kept at 4°C in darkness for 3 days before being transferred to long day photoperiod conditions (16 h of light, 8 h dark). All RNA and protein experiments were done using 14-days old seedlings unless otherwise specified.

Gene expression analyses

Seedlings were harvested, and RNA was extracted with the hot phenol method as previously described (Zhu *et al.* 2021). TURBO DNase (Ambion) was used to remove genomic DNA contamination before reverse transcription. cDNA was synthesized using SuperScript IV (Invitrogen) and gene-specific primers (Table S1). qPCR analyses were performed, and data was normalized to the indicated housekeeping gene or genes.

ChIP

2.5 gr of seedlings were crosslinked with 1% formaldehyde in 1X PBS for 12 min by vacuum infiltration, followed by addition of glycine (final concentration 125 mM) with another 7 min of vacuum infiltration. Tissue was then ground to fine powder with liquid nitrogen. Ground tissue was resuspended in 35 mL of Honda Buffer (20 mM Hepes, 0.44 M sucrose, 1.25 % Ficoll, 2.5% Dextran, 10 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT, 1x Roche protease inhibitor mixture), filtered through two layers of Miracloth, and centrifuged at 2500 xg for 15 min. Nuclei pellet was then washed once more with 1.6 mL of Honda Buffer.

For histone ChIP, nuclear pellets were resuspended in Nuclei Lysis Buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1 % SDS), and sonicated 4 x 5 min (30 sec ON/ 30 sec OFF) using a Diagenode Bioruptor on Medium setting. IP was performed by incubating 140 µl of sonicated chromatin diluted ten times with ChIP dilution buffer (16.7 mM Tris-HCl pH 8, 1.2 mM EDTA, 1.1 % Triton X-100, 167 mM NaCl) with 15 µl of Protein A-coated Dynabeads (Invitrogen) previously incubated for 2 h with either 2.5 µg of anti-H3 (ab176842), anti-H3K27me3 (ab192985), anti-H3K36me3 (ab9050), or H3K4me1 (ab8895) and incubated overnight at 4°C on a rotator wheel.

For FLAG-FLD and GFP-LD ChIP, nuclei were obtained as described above, but for FLAG-FLD the crosslinking buffer was supplemented with 1.5 mM of EGS [ethylene glycol bis(succinimidyl succinate); ThermoFisher]. Nuclear pellets were suspended in RIPA buffer (50 mM Tris-HCl, 150 mM NaCl, 1% Nonidet P-40, 0.5% NaDeoxycholate, 0.1% SDS, 1x Roche protease inhibitor mixture) and sonicated 5 times x 5 min (30 s ON/ 30 s OFF) with the Bioruptor in high setting. Undiluted chromatin was incubated overnight at 4°C with either 1.5 mg Dynabeads M-270 Epoxy preincubated with 1.5 µl anti-FLAG (Anti-FLAG® M2 / F1804, Merck) for FLAG-FLD or 15 µl Protein-A coated Dynabeads with 2.5 µl anti-GFP (ab290, Abcam), for GFP-LD.

Beads were then washed twice with Low Salt Wash Buffer (150 mM NaCl, 0.1 % SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 1x Roche protease inhibitor mixture), twice with High Salt Wash Buffer (500 mM NaCl, 0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris-HCl 8, 1x Roche protease inhibitor mixture), and twice with TE wash buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 1x Roche protease inhibitor mixture).

For RNA Pol II ChIP, nuclei were obtained as described for histone ChIP, complementing the Honda Buffer with 1x of PhosSTOP (Merck). Nuclear pellet was suspended in 1 mL of TAP buffer (100 mM

NaCl, 20 mM Tris-HCl pH 8, 2.5 mM EDTA, 10 % Glycerol, 1 % Triton, 1x of PhosSTOP, 1x Roche protease inhibitor mixture) and given 20 strokes with the Dounce Homogenizer. The resulting solution was sonicated 4 x 10 min (15 s ON/ 45 s OFF) with the Bioruptor in low setting. 250 µl of undiluted chromatin was incubated overnight at 4°C with either preincubated Protein-G coated Dynabeads with 2 µl anti-Total Pol II (C15200004, Diagenode) or Dynabeads M-270 Epoxy preincubated with anti-Tyr1P (MABE350, Merck), anti-Ser2P (C15200005-50, Diagenode), or anti-Ser5P (C15200007-50, Diagenode). Then beads were washed twice for 15 min with Low Salt and High Salt buffers like for histones but including 1x of PhosSTOP. Then beads were washed for 15 min with the LiCl buffer (250 mM LiCl, 0.5 % NP40, 2.5 mM EDTA, 0.05 % NaDeoxycholate, 20 mM Tris-HCl pH 8, 1x of PhosSTOP, 1x Roche protease inhibitor mixture), and the TE buffer (same as for histones plus 1x PhosSTOP). In all cases, after IP, DNA was then eluted and reverse-crosslinked by incubating the beads at 95°C for 10 min in presence of 100 µl of 10 % Chelex resin (BioRad), treated with Proteinase K (Roche) for 1 h at 45°C, and incubated again at 95°C for 10 min to inactivate the Proteinase K. Finally, DNA was purified using the ChIP DNA Clean & Concentrator kit (Zymo Research).

CLINIP-MS

CLNIP-MS was performed as previously described in (Fang *et al.* 2019). Briefly, 12-days-old *aprf1-9 APRF1$_{pro}$:APRF1-VENUS* seedlings were crosslinked with 1% formaldehyde. Tissue was then ground to a fine powder and nuclei isolated in lysis buffer (20 mM Tris HCl pH 7.5, 20 mM KCl, 2 mM EDTA pH 8, 2.5 mM $MgCl_2$, 25 % glycerol, 250 mM sucrose, 1x Roche protease inhibitor mixture), filtered through miracloth, washed three times with NRBT (20 mM Tris HCl pH 7.5, 2.5 mM MgCl2, 25 % glycerol, 0.2 % Triton X-100, 1x Roche protease inhibitor mixture), and resuspended in RIPA (50 mM Tris-HCl, 150 mM NaCl, 1 % Nonidet P-40, 0.5 % NaDeoxycholate, 0.1 % SDS, 1x Roche protease inhibitor mixture) before sonication. Nuclear extract was incubated with GFP-Trap Magnetic Agarose beads (ChromoTek, gtma-10) for 2 h and washed sequentially with Low Salt, High Salt, and TE buffers as for ChIP. The immunoprecipitates were then boiled at 95°C for 15 min to reverse crosslinking. The protein samples were gel-purified and subjected to mass spectrometry analysis by nanoLC-MS/MS on a Orbitrap Fusion Tribrid Mass Spectrometer coupled to an UltiMate 3000 RSLCnano LC system Thermo Scientific, Hemel Hampstead, UK. Data were searched using Mascot server (Matrixscience, London, UK) and analysed using MaxQuant software.

Preparation of Chromatin-bound RNA

Chromatin-bound RNA was isolated as previously described (Wu *et al.* 2016). Nuclei from 2-2.5 gr of non-crosslinked seedlings were obtained with Honda Buffer, supplemented with 20 U/mL RNase inhibitor RNase Out (Invitrogen), 1 mM PMSF, and 50 ng/µL of yeast tRNA. Nuclear pellet was rinsed with 500 µL of resuspension buffer (20 mM Tris pH 8, 75 mM NaCl, 0.5 mM EDTA, 1 mM

DTT, 0.125 mM PMSF, 50 % glycerol, 1x Roche complete, 20 U/mL RNase Out) and centrifuged at 4000 xg at 4°C for 3 min. Nuclei pellet was weighed and resuspended in an equal volume of Resuspension Buffer. The suspension was then washed with two volumes of Urea Wash Buffer (20 mM Tris pH8, 300 mM NaCl, 7.5 mM MgCl2, 0.25 mM EDTA, 1 mM DTT, 1 M Urea, 1 % NP-40, 1x Roche complete, 20 U/mL RNase Out), pipetting up and down 30 times, and spun at 8000 xg for 1 min at 4°C. Supernatant was kept as the Free RNA samples. Nuclei pellet was resuspended again with 1 volume of Resuspension buffer and washed with 1 volume of Urea Wash Buffer, pipetting 30 times up and down, and spun for 1 min at 8000 xg and 4°C. Supernatant was combined with the previous one to constitute the Free RNA sample. Finally, nuclear pellet was dissolved in 1 mL of TRIzol (Invitrogen), adding 0.2 mL of chloroform and shaking vigorously by hand for 15 s. Then the suspension was incubated at RT for 2 min and centrifuged at 12 000 xg for 15 min at 4°C. The aqueous phase was taken and mixed with an equivalent volume of Phenol:Chloroform:Isoamyl alcohol (25:24:1, Sigma), shaken for 10 min at room temperature, and centrifuge at RT and 12 000 xg for 10 min. The aqueous phase was then transferred to another tube and followed two precipitations, first with isopropanol, sodium acetate, and RNA GlycoBlue and another one with LiCl. Finally, RNA was dissolved and treated with DNase Turbo (Ambion) and used as template for reverse transcription with gene-specific primers.

Quant-seq

For Quant-seq experiments, total RNA was isolated as for qPCR and further cleaned up using the Qiagen RNeasy miniprep kit (74106). Library preparation, sequencing, and data analysis were carried out by Lexogen GmbH (Austria). Sequencing-ready libraries were generated from 100 ng of input RNA using a QuantSeq 3' mRNA-Seq Library Prep Kit REV for Illumina (015UG009V0271) following standard procedures. RNA integrity, and Indexed libraries quality were assessed on a Fragment Analyzer device (Agilent Technologies) using a DNF-471 RNA Kit and HS- DNA assay, respectively. Libraries were quantified using a Qubit dsDNA HS assay (Thermo Fisher). A sequencing-ready pool of indexed libraries were sequenced on an Illumina NextSeq 2000 with a 100-cycle cartridge using the Custom Sequencing Primer (CSP). FastQC version v0.11.7 was used to verify the read quality and cutadapt version 1.18 (Martin 2011) for read adapter trimming. Clean reads were mapped to the latest version of the Arabidopsis genome (TAIR10) with a splice-aware aligner STAR version 2.6.1a (Dobin *et al.* 2013). Raw reads have been deposited on Short Read Archive (SRA) under the reference PRJNA978558.

AlphaFold2 protein interaction prediction

AlphaFold models were predicted using the Colab notebook running a slightly simplified version of AlphaFold v2.3.2. (Jumper *et al.* 2021). For the Arabidopsis LD–APRF1 complex LD aa566-661 and APRF1 aa1-330 were used. For the human PNUTS–WDR82 complex PNUTS aa380-530 and

WDR82 aa1-313, and for the yeast Ref2–Swd2 complex Ref2 aa406-533 and Swd2 aa1-329 were used. The number of recycles were set to 3 and models included a final relaxation stage. PAE plots indicating the quality of the prediction and LDDT plots are shown in Figure S11. Final models were imported into and colour figures were prepared with PyMOL (v2.1, Schrödinger).

Bioinformatic analyses

To design the sgRNA to edit *APRF1* via CRISPR/Cas9 we used the CRISPR-P 2.0 (http://crispr.hzau.edu.cn/cgi-bin/CRISPR2/CRISPR), selecting the canonical NGG PAM motif, and the Arabidopsis TAIR10 as a target genome. Protein alignments were performed with MEGA X (Kumar *et al.* 2018) using the MUSCLE (Edgar 2004) algorithm with default parameters, and shaded with the "Colour Align Conservation" tool from the Sequence Manipulation Suite (Stothard 2000). To find putative PNUTS homologs in Arabidopsis, BLASTP (Altschul *et al.* 1997), PSI-BLAST, and DELTA-BLAST (Boratyn *et al.* 2012) searches were performed at NCBI site using an alignment score threshold of 80 for DELTA-BLAST, and default parameters for the rest. Individual protein structure predictions were retrieved from the Alphafold website (https://alphafold.ebi.ac.uk/).

ACKNOWLEDGEMENTS

For genetic materials, we are indebted to Fredy Barneche for providing seeds of *s2lb* and *aprf1-9 s2lb* (Fiorucci *et al.* 2019) and Soichi Inagaki for sharing their *3xFLAG-FLD* [in both (*fld-4; FRI*) and (*fld-4*; *fri*) backgrounds] transgenic lines (Inagaki *et al.* 2021). We are indebted to Yusheng Zhao for initiating the introgression of the GFP-LD line into Col*FRI* and to Minglei Yang for writing the code used to analyse Quant-seq data. Authors would also thank Shuqin Chen, Aida Sánchez, and Tina Zhang for their excellent technical assistance. This work was funded by the European Research Council Advanced Grant (EPISWITCH, 833254), Welcome Trust (210654/Z/18/Z), BBSRC Institute Strategic Programmes (BB/J004588/1 and BB/P013511/1), and the Royal Society Professorship (RP\R1\180002) to Caroline Dean; and the Medical Research Council (MRC) as part of United Kingdom Research and Innovation (MRC file reference number MC_U105192715) and the Wellcome Trust (225217/Z/22/Z) to Lori A Passmore. Eduardo Mateo-Bonmatí would like to thank grant RYC2021-030895-I, funded by MCIN/AEI and by European Union NextGenerationEU/PRTR.

FIGURE LEGENDS

Figure 1. APRF1, a homolog of the yeast Swd2 is a robust interactor of the FLD complex, and genetically functions in the FCA pathway. Tables in (A) and (B) summarize LC-MS/MS proteomic experiments using tagged versions of the FLD complex components FLD, LD, SDG26. Experiments from (A) were performed on formaldehyde-treated seedlings (Fang et al. 2020), and from (B) were carried out using non-cross-linked tissue (Qi et al. 2022). Different columns represent different replicates. (C) Architecture of the *APRF1* gene and illustration of the nature and location of the mutations studied in this work. Boxes indicate exons and lines introns. White boxes represent

untranslated regions. The triangle represents a T-DNA insertion (*aprf1-9*); the red arrow points to the location of the CRISPR/Cas9-derived deletion (*aprf1-10*). (D) Boxplot representing the leaf number at bolting of the wild-type Col-0 and both *aprf1* mutants. Each dot represents the score for one plant. (E) Relative values of *FLC* spliced (left) and unspliced (right) in the wild-type Col-0, both *aprf1* mutants and F1 *aprf1-9/aprf1-10* hybrid plants. Values were normalized to the housekeeping *UBC* gene and to Col-0. (F) Schematic diagram showing *FLC* gene structure following the guidelines described for (C). +1 indicates the transcriptional start site (TSS). (G-I) ChIP analysis of H3K4me1 (G), H3K36me3 (H), and H3K27me3 (I) levels at *FLC* in Col-0 and *aprf1-10*. Numbers in x axis represent the distance in kilobases to the *FLC* TSS and numbers in the y axis correspond to relative enrichment of the corresponding histone mark. Each dot represents an amplicon. Values were normalized to H3 and to *ACT* (for H3K4me1 and H3K36me3) or *STM* (for H3K27me3). (J-O) Relative values of *FLC* spliced (J, L, N), and unspliced (K, M, O) in various genetic backgrounds. Values were normalized to the housekeeping gene *UBC* (J-O) and Col-0 (J, K) or *PP2A* (L-O). Asterisks indicate statistically significant differences to Col-0 (D, E), and to C2 ($35S_{pro}$*:FCA; FRI*) (K), in a two-way Student's t test (** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$). The n.s. stands for not statistically different. Scale bars indicate 500 bp (C, F). Experiments were performed in 2-week-old seedlings grown in long days conditions with $n \geq 3$ (E, G-O).

Figure 2. LD-APRF1-TOPP4 form a plant equivalent of the CPSF phosphatase module. (A) Summary of the CLNIP results using *aprf1-9* $APRF1_{pro}$*:APRF1-VENUS* transgenic line. (B) Summary of the different proteomic experiments obtained in Qi et al (2022) using an APRF1-FLAG transgenic line in different backgrounds. (C) Components of the CP(S)F phosphatase domain in yeast, humans and Arabidopsis. (D) Schematic representation of protein size, annotated domains and disorder probability score for LD, PNUTS (*Homo sapiens*) and Ref2 (*Saccharomyces cerevisiae*). Individual amino acid score for potential disorder were obtained with the online Protein DisOrder prediction System (PrDOS) and plotted using Graphpad. All the proteins harbour a TFIIS-I domain (IPR003617), followed by highly disordered region. Additionally, LD carries an upstream homeodomain (IPR001356), PNUTS a Zinc-finger domain in the C-terminal region, and it was reported that the C-terminal region (348-523) of Ref2 is key to interact with Glc7 (Nedea et al., 2008). (E) AlphaFold2 predictions of complexes between APRF1-LD, WDR82-PNUTS, and Swd2-Ref2. WD40 orthologs are shown in grey surface representation, whereas TFIIS orthologs are shown in cartoon; the N-terminal amino acid of the predicted TFIIS proteins are denoted by a black sphere. (F) Overlay of the three predictions shown in (E).

Figure 3. ChIP-qPCR co-occupancy profiles indicate that FLD and LD work co-transcriptionally to control *FLC* transcription. (A) Elongating RNAPol II (Ser2P) ChIP profiles over *FLC* in a high (Col*FRI*) and low (*fri* or Col-0) transcriptional background. (B) ChIP binding profile over *FLC* of the *fld-4* FLD_{pro}*:3xFLAG-FLD* transgenic plants in different genetic backgrounds. (C) ChIP binding profile over *FLC* of the *ld-1* LD_{pro}*:GFP-LD* transgenic plants in different genetic backgrounds. *FLC* gene structure following the indications for Figure 1F. All the experiments were done with 2-week-old seedlings. Dots and error bars represent average ± s.e.m. of three replicates. Results in (A) are normalized as in Mikulski *et al*. 2022.

Figure 4. Mutations in *APRF1* trigger *COOLAIR* upregulation and an increase in a new medially polyadenylated *COOLAIR* isoform. (A) *FLC* architecture following the representation of Figure 1J with indication of different *COOLAIR* isoforms in grey. For simplicity, Classes I and II (proximal and distal) are each represented by one isoform. The new three *COOLAIR* class III isoforms are highlighted in pale blue. Triangles represent primers pairs (not drawn to scale) used to measure the relative abundance by qRT-PCR. Dashed lines indicate primers spanning two *COOLAIR* exons. Green triangles indicate the primer used for *COOLAIR* class III retro-transcription. Red vertical bars represent polyadenylation sites. Scale bar indicates 500 bp. (B-D) Relative expression analyses of (B) total *COOLAIR*, (C) proximal *COOLAIR*, and (D) distal *COOLAIR* in different genotypes. (E) *COOLAIR* proximal-to-distal ratio in different genotypes. (F, G) Relative expression analyses of *COOLAIR* class (F) III-1 and III-2, and (G) III-3. Each dot represents a biological replicate analysed in triplicate. Expression values were normalized to the *UBC* housekeeping gene and to Col-0.

Asterisks indicate statistically significant differences to the indicated genotypes in a Student's t test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$). N.s. stands for not statistically different. (c, f). All the experiments were performed in 2-week-old seedlings grown in long days conditions. (H) Quant-seq experiment results from the *COOLAIR* strand carried out on Col*FRI* and *aprf1-9 FRI* seedlings. *FLC* locus is represented as in panel A. The blue spikes indicate reads supporting a polyadenylation site. Vertical pale orange bars highlight polyA sites found in *aprf1-9 FRI* absent in the wild type Col*FRI*.

Figure 5. *COOLAIR* transcriptional readthrough correlates with changed phosphorylation of the RNA Pol II carboxy terminal domain (CTD). (A) Chromatin-bound RNA levels of *FLC* and *COOLAIR* in *aprf1-9* mutants with and without functional *FRI*. Data was normalized to *UBC* and *PP2A* and is shown as fold-change to wildtype Col-0. Dots correspond to amplicons of the *FLC* (upper chart) and *COOLAIR* (bottom chart) transcripts, represented as average ± s.e.m. of three biological replicates. (B, C) RNA Pol II ChIP of Col-0 vs *aprf1-10* (B), and Col*FRI* vs *aprf1-9 FRI* (C) using antibodies raised against different phosphoisoforms (Ser2P and Ser5P) of the CTD. (D) RNA Pol II ChIP of Col-0, *ld-1*, and *ld-1 aprf1-9* using antibodies recognizing total CTD (left), CTD Ser2P (centre), and CTD Ser5P (right). Pol II ChIP data was normalized as in Fig 3A.

Figure 6. Proposed model for transcription-mediated chromatin silencing. (A) Open *FLC* chromatin represented by white nucleosomes and marked with H3K4me1 (green circles) is actively transcribed by RNA Pol II machinery (solid maroon), which carries a non-active FLD complex (blue circle) as well as the 3'-end processing machinery (CPSF), including the phosphatase module formed by LD-APRF1-TOPP4, and elongation factors (green circle). Pol II CTD and elongation factors harbour some posttranslational modifications including phosphorylation (yellow circles). Nascent *COOLAIR* forms an R-loop in the 3' end of the locus (Sun *et al.* 2013, Xu *et al.* 2021b). (B) Formation of the R-loop stimulates the 3'-end processing machinery to terminate transcription at the proximal PAS. This termination is also signalled by the phosphatase module to Pol II via dephosphorylation of either elongating factors or the CTD or both (dashed line). PAS recognition triggers a conformational change on RNA Pol II, illustrated by a solid to pale maroon colour change, which also activates FLD (now solid blue circle). (C) After *COOLAIR* is released, RNA Pol II continues transcribing an uncapped transcript that is the substrate of 5'-3' exonucleases (XRNs). During this non-productive transcription, the FLD complex co-transcriptionally removes H3K4me1 marks from nucleosomes, creating a less processive chromatin environment for subsequent rounds of transcription.

REFERENCES

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs Nucleic Acids Res *25*, 3389-3402.

Bae, H.J., Dubarry, M., Jeon, J., Soares, L.M., Dargemont, C., Kim, J., Geli, V., and Buratowski, S. (2020). The Set1 N-terminal domain and Swd2 interact with RNA polymerase II CTD to recruit COMPASS. Nat Commun *11*, 2181.

Berry, S., Hartley, M., Olsson, T.S., Dean, C., and Howard, M. (2015). Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance. eLife *4*, e07205.

Boratyn, G.M., Schäffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J., and Madden, T.L. (2012). Domain enhanced lookup time accelerated BLAST. Biology Direct *7*, 12.

Carminati, M., Manav, M.C., Bellini, D., and Passmore, L.A. (2022). A direct interaction between CPF and Pol II links RNA 3□-end processing to transcription. https://doi.org/10.1101/2022.07.28.501803

Casanal, A., Kumar, A., Hill, C.H., Easter, A.D., Emsley, P., Degliesposti, G., Gordiyenko, Y., Santhanam, B., Wolf, J., Wiederhold, K., et al. (2017). Architecture of eukaryotic mRNA 3'-end processing machinery. Science *358*, 1056-1059.

Castaings, L., Bergonzi, S., Albani, M.C., Kemi, U., Savolainen, O., and Coupland, G. (2014). Evolutionary conservation of cold-induced antisense RNAs of *FLOWERING LOCUS C* in *Arabidopsis thaliana* perennial relatives. Nat Commun *5*, 4457.

Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D.L. (2019). Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a "Sitting Duck Torpedo" mechanism. Mol Cell *76*, 896-908.

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.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Estell, C., Davidson, L., Eaton, J.D., Kimura, H., Gold, V.A.M., and West, S. (2023). A restrictor complex of ZC3H4, WDR82, and ARS2 integrates with PNUTS to control unproductive transcription. Mol Cell.

Fang, X., Wang, L., Ishikawa, R., Li, Y., Fiedler, M., Liu, F., Calder, G., Rowan, B., Weigel, D., Li, P., and Dean, C. (2019). Arabidopsis FLL2 promotes liquid-liquid phase separation of polyadenylation complexes. Nature *569*, 265-269.

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.

Fiorucci, A.S., Bourbousse, C., Concia, L., Rougee, M., Deton-Cabanillas, A.F., Zabulon, G., Layat, E., Latrasse, D., Kim, S.K., Chaumont, N., et al. (2019). Arabidopsis S2Lb links AtCOMPASS-like and SDG2 activity in H3K4me3 independently from histone H2B monoubiquitination. Genome Biol *20*, 100.

Gregersen, L.H., Mitter, R., Ugalde, A.P., Nojima, T., Proudfoot, N.J., Agami, R., Stewart, A., and Svejstrup, J.Q. (2019). SCAF4 and SCAF8, mRNA anti-terminator proteins. Cell *177*, 1797-1813.

Grewal, S.I.S. (2023). The molecular basis of heterochromatin assembly and epigenetic inheritance. Mol Cell *83*, 1767-1785.

Gyenis, A., Chang, J., Demmers, J., Bruens, S.T., Barnhoorn, S., Brandt, R.M.C., Baar, M.P., Raseta, M., Derks, K.W.J., Hoeijmakers, J.H.J., and Pothof, J. (2023). Genome-wide RNA polymerase stalling shapes the transcriptome during aging. Nat Genet *55*, 268-279.

Hu, S., Peng, L., Song, A., Ji, Y.X., Cheng, J., Wang, M., and Chen, F.X. (2023). INTAC endonuclease and phosphatase modules differentially regulate transcription by RNA polymerase II. Mol Cell *83*, 1588-1604.

Inagaki, S., Takahashi, M., Takashima, K., Oya, S., and Kakutani, T. (2021). Chromatin-based mechanisms to coordinate convergent overlapping transcription. Nat Plants *7*, 295-302.

Jiang, D., Kong, N.C., Gu, X., Li, Z., and He, Y. (2011). Arabidopsis COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. PLoS Genet *7*, e1001330.

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.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.

Kapolas, G., Beris, D., Katsareli, E., Livanos, P., Zografidis, A., Roussis, A., Milioni, D., and Haralampidis, K. (2016). *APRF1* promotes flowering under long days in *Arabidopsis thaliana*. Plant Sci *253*, 141-153.

Kim, H.J., Li, P., Kim, T., Oldfield, A.J., Zheng, X., and Yang, P. (2022). Integrative analysis reveals histone demethylase LSD1 promotes RNA polymerase II pausing. iScience *25*, 105049.

Kowalik, K.M., Shimada, Y., Flury, V., Stadler, M.B., Batki, J., and Bühler, M. (2015). The Paf1 complex represses small-RNA-mediated epigenetic gene silencing. Nature *520*, 248-252.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution *35*, 1547-1549.

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.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3' processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. Science *327*, 94-97.

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.

Liu, Y., and Huang, Y. (2018). Uncovering the mechanistic basis for specific recognition of monomethylated H3K4 by the CW domain of *Arabidopsis* histone methyltransferase SDG8. J Biol Chem *293*, 6470-6481.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10-12.

Menon, G., Mateo-Bonmatí, E., Reeck, S., Maple, R., Wu, Z., Ietswaart, R., Dean, C., and Howard, M. (2023). Alternative polyadenylation controls a transcriptional state that determines the rate of Polycomb silencing *Co-submitted with this manuscript*.

Menon, G., Schulten, A., Dean, C., and Howard, M. (2021). Digital paradigm for Polycomb epigenetic switching and memory. Curr Opin Plant Biol *61*, 102012.

Mikulski, P., Wolff, P., Lu, T., Nielsen, M., Echevarria, E.F., Zhu, D., Qüesta, J.I., Saalbach, G., Martins, C., and Dean, C. (2022). VAL1 acts as an assembly platform co-ordinating co-transcriptional repression and chromatin regulation at Arabidopsis *FLC*. Nat Commun *13*, 5542.

Mo, W., Liu, B., Zhang, H., Jin, X., Lu, D., Yu, Y., Liu, Y., Jia, J., Long, Y., Deng, X., et al. (2021). Landscape of transcription termination in Arabidopsis revealed by single-molecule nascent RNA sequencing. Genome Biol *22*, 322.

Muniz, L., Nicolas, E., and Trouche, D. (2021). RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. EMBO J *40*, e105740.

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. Mol Cell *29*, 577-587.

Pinter, S., Knodel, F., Choudalakis, M., Schnee, P., Kroll, C., Fuchs, M., Broehm, A., Weirich, S., Roth, M., Eisler, S.A., et al. (2021). A functional LSD1 coregulator screen reveals a novel transcriptional regulatory cascade connecting R-loop homeostasis with epigenetic regulation. Nucleic Acids Res *49*, 4350-4370.

Qi, P.L., Zhou, H.R., Zhao, Q.Q., Feng, C., Ning, Y.Q., Su, Y.N., Cai, X.W., Yuan, D.Y., Zhang, Z.C., Su, X.M., et al. (2022). Characterization of an autonomous pathway complex that promotes flowering in *Arabidopsis*. Nucleic Acids Res *50*, 7380-7395.

Rodríguez-Molina, J.B., West, S., and Passmore, L.A. (2023). Knowing when to stop: Transcription termination on protein-coding genes by eukaryotic RNAPII. Mol Cell *83*, 404-415.

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.

Simpson, G.G., Dijkwel, P.P., Quesada, V., Henderson, I., and Dean, C. (2003). FY is an RNA 3' end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. Cell *113*, 777-787.

Soares, L.M., and Buratowski, S. (2012). Yeast Swd2 is essential because of antagonism between Set1 histone methyltransferase complex and APT (associated with Pta1) termination factor. J Biol Chem *287*, 15219-15231.

Sonmez, C., Bäurle, I., Magusin, A., Dreos, R., Laubinger, S., Weigel, D., and Dean, C. (2011). RNA 3' processing functions of *Arabidopsis* FCA and FPA limit intergenic transcription. Proc Natl Acad Sci USA *108*, 8508-8513.

Stothard, P. (2000). The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques *28*, 1102-1104.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis *FLC* locus. Science *340*, 619-621.

Tsutsui, H., and Higashiyama, T. (2017). pKAMA-ITACHI Vectors for Highly Efficient CRISPR/Cas9-Mediated Gene Knockout in *Arabidopsis thaliana*. Plant Cell Physiol *58*, 46-56.

Wagner, E.J., Tong, L., and Adelman, K. (2023). Integrator is a global promoter-proximal termination complex. Mol Cell *83*, 416-427.

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.

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.

Xu, C., Fang, X., Lu, T., and Dean, C. (2021a). Antagonistic cotranscriptional regulation through ARGONAUTE1 and the THO/TREX complex orchestrates *FLC* transcriptional output. Proc Natl Acad Sci USA *118*.

Xu, C., Wu, Z., Duan, H.C., Fang, X., Jia, G., and Dean, C. (2021b). R-loop resolution promotes co-transcriptional chromatin silencing. Nat Commun *12*, 1790.

Yang, H., Howard, M., and Dean, C. (2014). Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at *Arabidopsis FLC*. Curr Biol *24*, 1793-1797.

Zhu, P., Lister, C., and Dean, C. (2021). Cold-induced *Arabidopsis* FRIGIDA nuclear condensates for *FLC* repression. Nature *599*, 657-661.

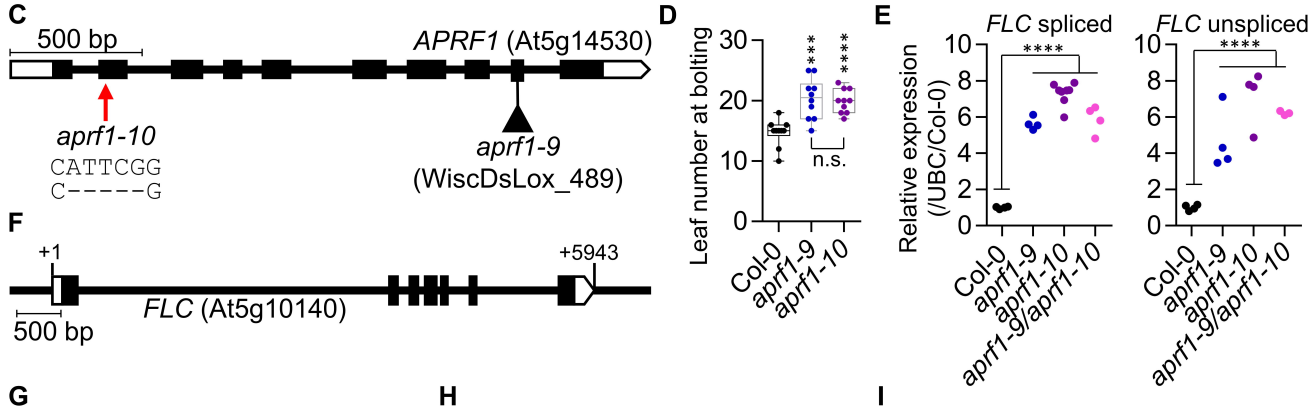
Zumer, K., Maier, K.C., Farnung, L., Jaeger, M.G., Rus, P., Winter, G., and Cramer, P. (2021). Two distinct mechanisms of RNA polymerase II elongation stimulation *in vivo*. Mol Cell *81*, 3096-3109.

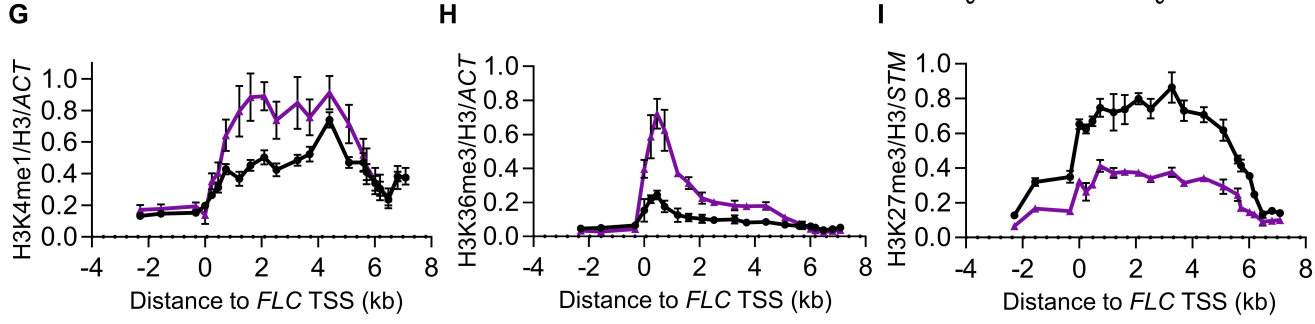
A

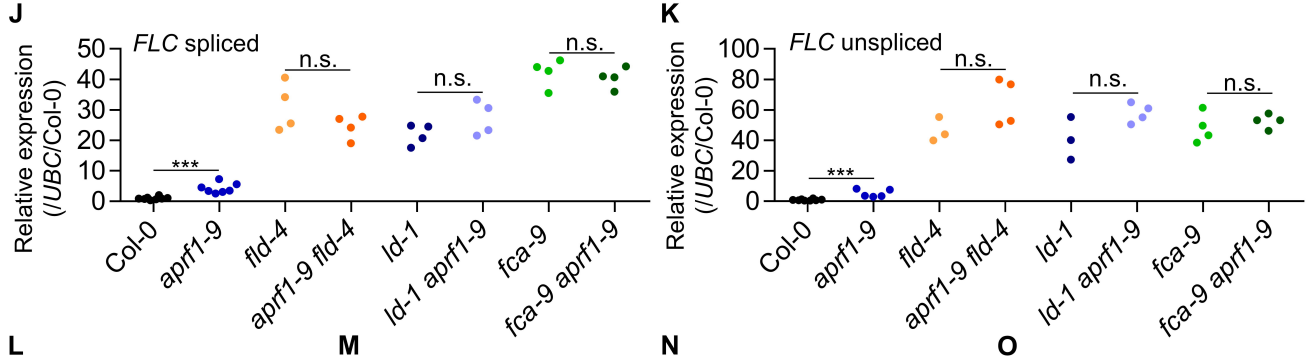
	Number of unique peptides							
	SDG26-GFP			GFP-LD			FLAG-FLD	
	sdg26-3			*ld-1*			*fld-4*	
FLD	29	68	79	0	61	35	225	206
LD	34	80	91	9	134	83	97	73
SDG26	58	108	115	1	36	21	44	35
APRF1	11	25	30	0	38	22	41	40

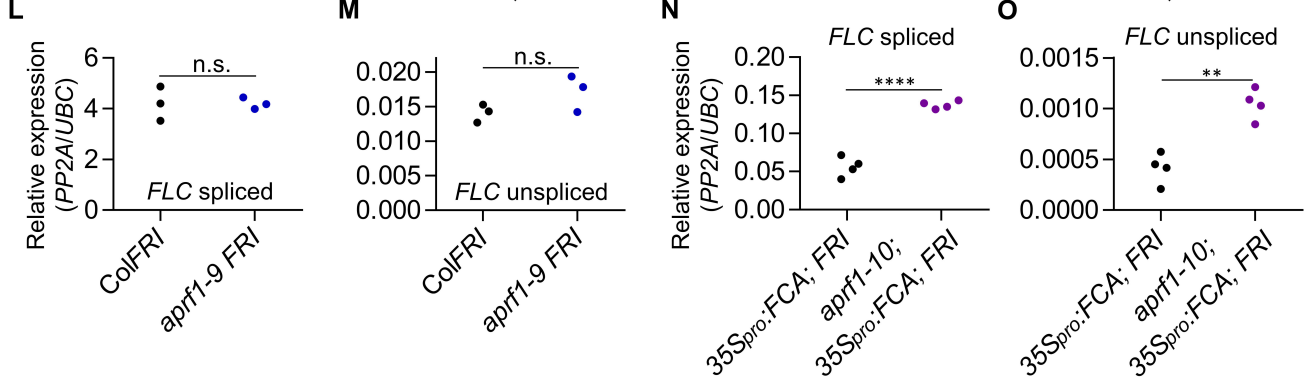
B

	Number of unique peptides		
	FLD-FLAG	SDG26-FLAG	LD-FLAG
	fld	*sdg26*	*ld*
FLD	23	20	21
LD	8	21	42
SDG26	8	42	21
APRF1	4	6	15









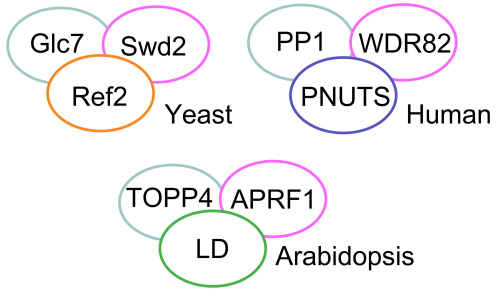
A

	Number of peptides APRF1-VENUS *aprf1-9*
FLD	0
LD	24
SDG26	0
APRF1	86
TOPP4	18

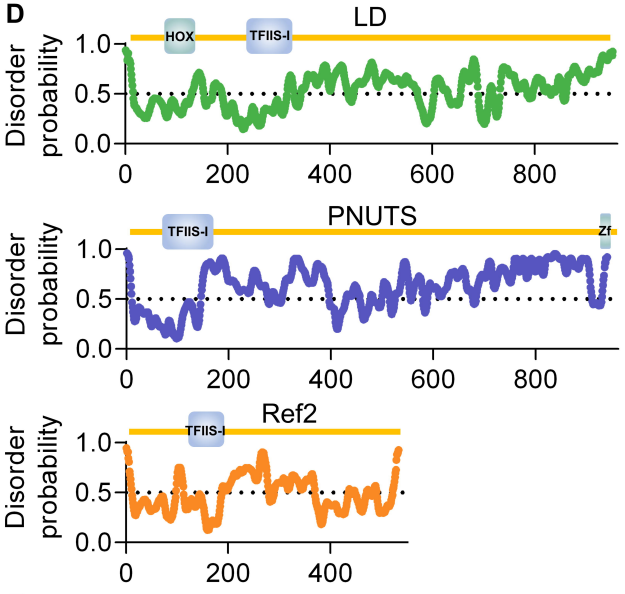
B

	Number of peptides APRF1-FLAG *APRF1*	*fld*	*ld*
FLD	15	0	0
LD	27	6	0
SDG26	8	4	0
APRF1	16	28	26
TOPP4	14	11	14

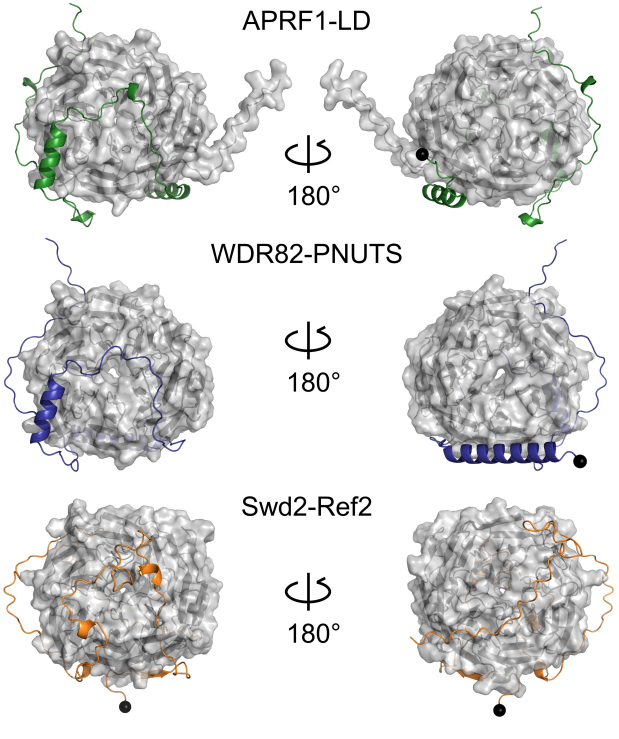
C



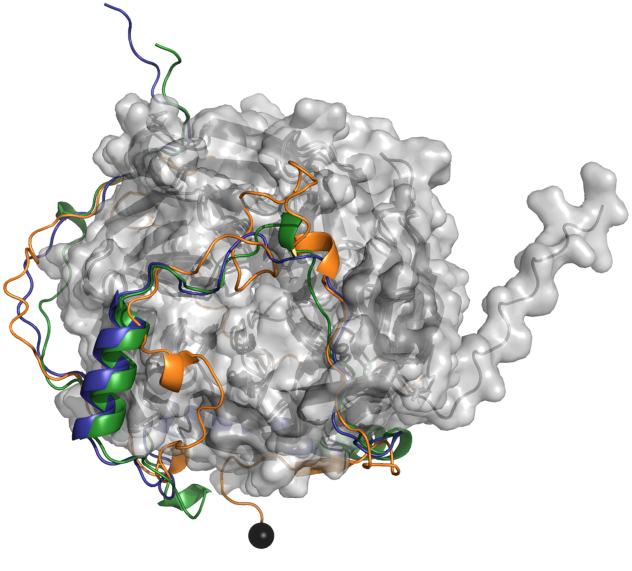
D

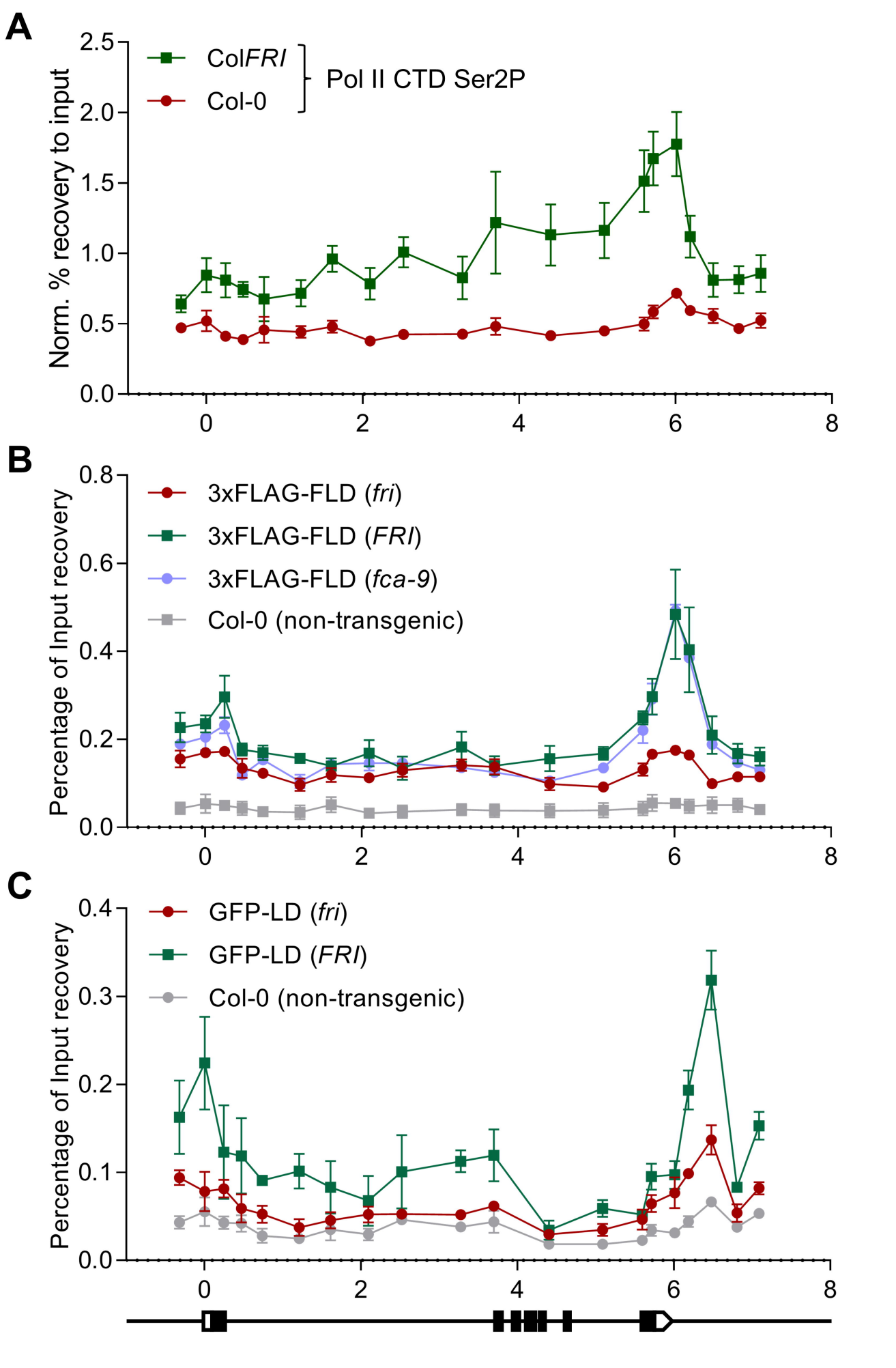


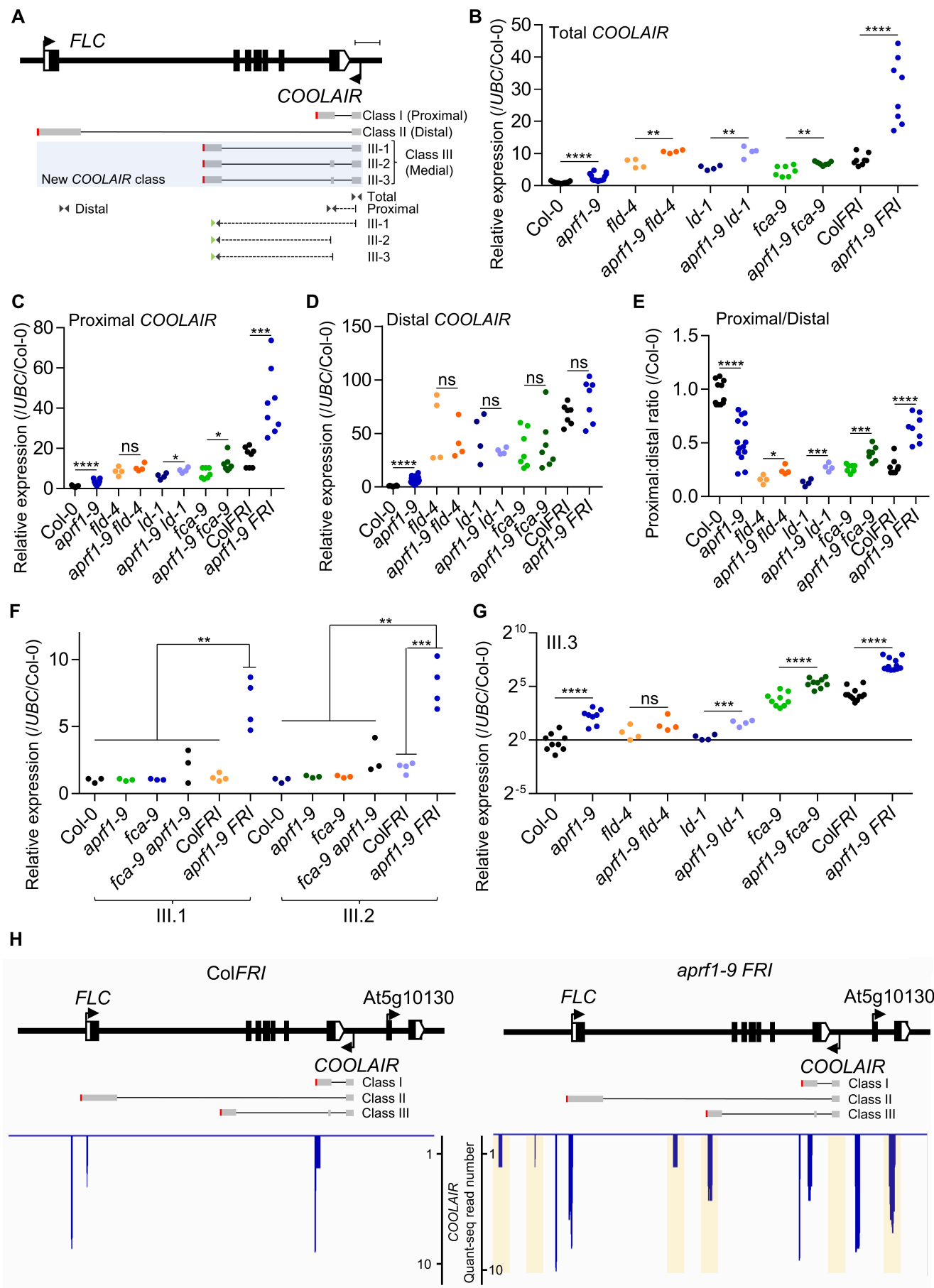
E

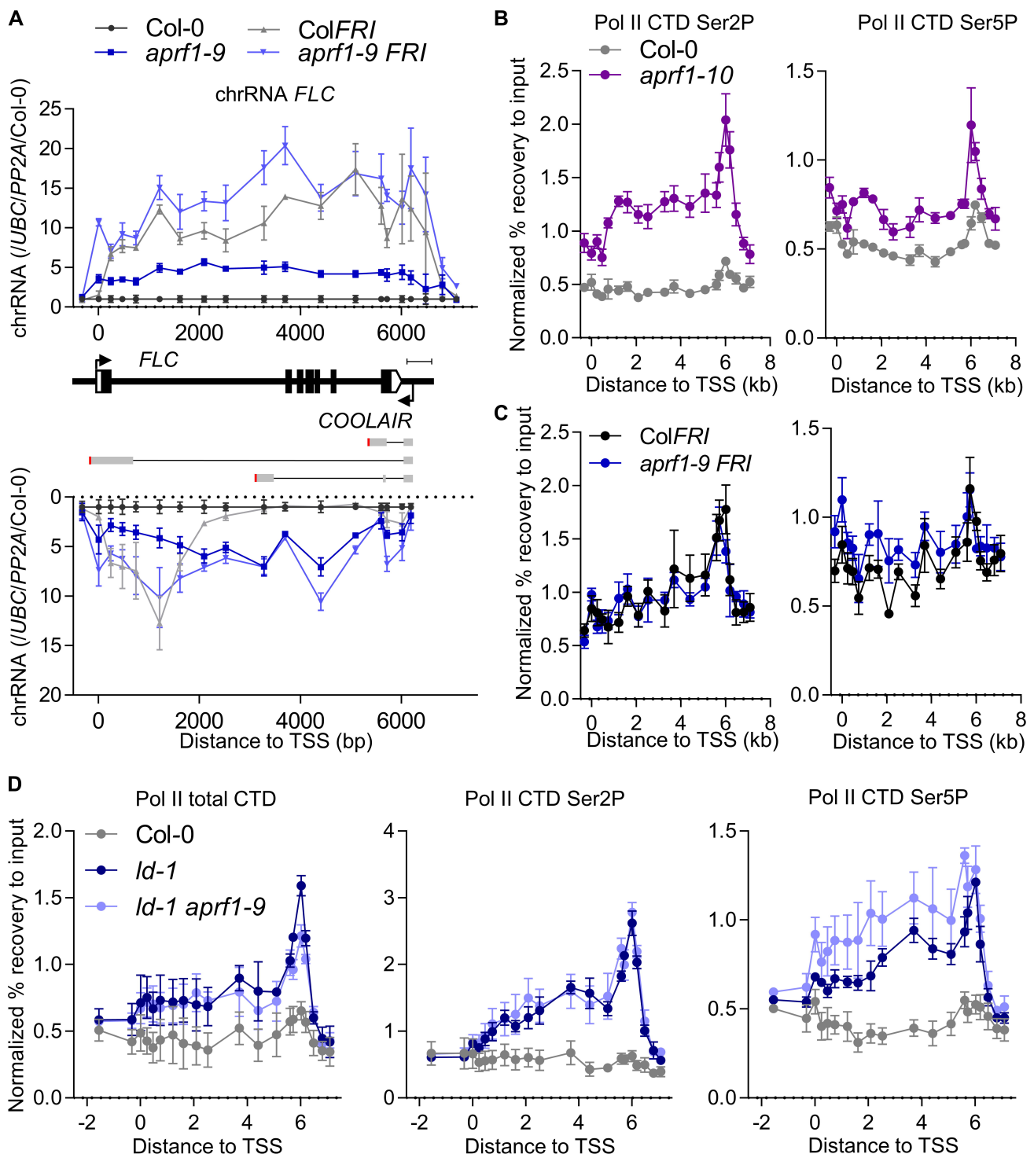


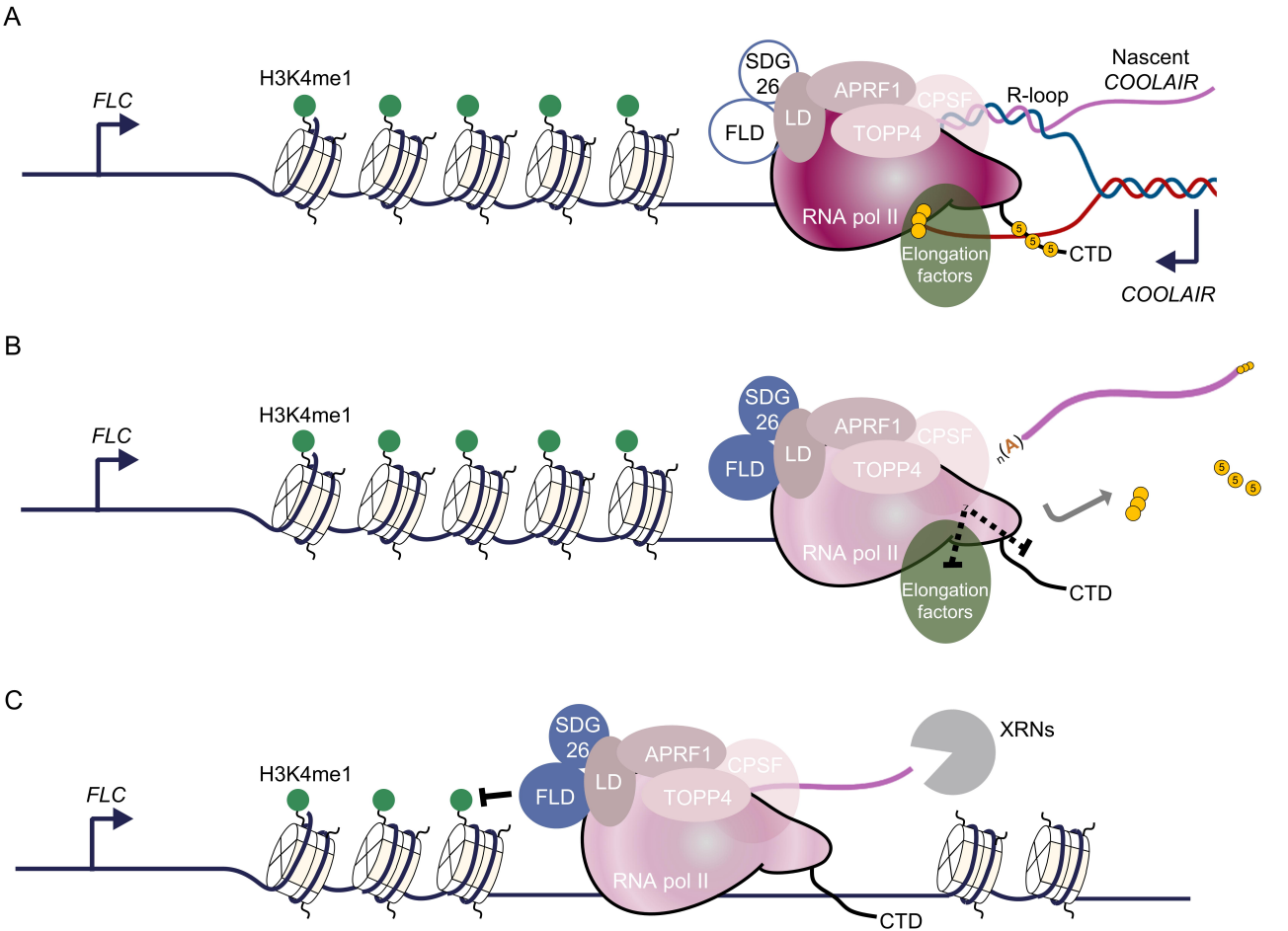
F




A
Norm. % recovery to input
2.5
2.0
1.5
1.0
0.5
0.0
ColFRI
Col-0
Pol II CTD Ser2P
0
2
4
6
8
B
Percentage of Input recovery
0.8
0.6
0.4
0.2
0.0
3xFLAG-FLD (fri)
3xFLAG-FLD (FRI)
3xFLAG-FLD (fca-9)
Col-0 (non-transgenic)
0
2
4
6
8
C
Percentage of Input recovery
0.4
0.3
0.2
0.1
0.0
GFP-LD (fri)
GFP-LD (FRI)
Col-0 (non-transgenic)
0
2
4
6
8


A
FLC
COOLAIR
Class I (Proximal)
Class II (Distal)
III-1
III-2
III-3
Class III (Medial)
New COOLAIR class
Distal
Total
Proximal
III-1
III-2
III-3
B
Total COOLAIR
Relative expression (/UBC/Col-0)
Col-0
aprf1-9
fld-4
aprf1-9 fld-4
ld-1
aprf1-9 ld-1
fca-9
aprf1-9 fca-9
ColFRI
aprf1-9 FRI
C
Proximal COOLAIR
D
Distal COOLAIR
E
Proximal/Distal
Proximal:distal ratio (/Col-0)
F
III.1
III.2
G
III.3
H
ColFRI
aprf1-9 FRI
At5g10130
COOLAIR
Class I
Class II
Class III
COOLAIR Quant-seq read number


A
Col-0
aprf1-9
ColFRI
aprf1-9 FRI
chrRNA FLC
chrRNA (/UBC/PP2A/Col-0)
FLC
COOLAIR
Distance to TSS (bp)
B
Pol II CTD Ser2P
Pol II CTD Ser5P
Col-0
aprf1-10
Normalized % recovery to input
Distance to TSS (kb)
C
ColFRI
aprf1-9 FRI
D
Pol II total CTD
Pol II CTD Ser2P
Pol II CTD Ser5P
Col-0
ld-1
ld-1 aprf1-9
Distance to TSS


A
FLC
H3K4me1
SDG 26
FLD
LD
APRF1
CPSF
TOPP4
R-loop
Nascent COOLAIR
RNA pol II
Elongation factors
CTD
COOLAIR
B
FLC
H3K4me1
SDG 26
FLD
LD
APRF1
CPSF
TOPP4
n(A)
RNA pol II
Elongation factors
CTD
C
FLC
H3K4me1
SDG 26
FLD
LD
APRF1
CPSF
TOPP4
XRNs
RNA pol II
CTD