The lncRNA *MARS* modulates the epigenetic reprogramming of the marneral cluster in response to ABA

Thomas Roulé[1,2], Aurelie Christ[1,2], Nosheen Hussain[4], Ying Huang[1,2], Caroline Hartmann[1,2], Moussa Benhamed[1,2], Jose Gutierrez-Marcos[4], Federico Ariel[3], Martin Crespi [1,2]* and Thomas Blein[1,2]*

[1] Université Paris-Saclay, CNRS, INRAE, Université Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91190, Gif sur Yvette, France

[2] Université de Paris, Institute of Plant Sciences Paris-Saclay (IPS2), 91190, Gif sur Yvette, France

[3] Instituto de Agrobiotecnología del Litoral, CONICET, FBCB, Universidad Nacional del Litoral, Colectora Ruta Nacional 168 km 0, 3000 Santa Fe, Argentina

[4] School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK

*Correspondence to: MC (martin.crespi@cnrs.fr) and TB (thomas.blein@cnrs.fr)

SHORT SUMMARY

The Arabidopsis *MARneral Silencing* (*MARS*) long non-coding RNA participates in the marneral metabolic genes cluster co-expression. In response to ABA, *MARS* promotes the formation of a chromatin loop bringing together the *MARNERAL SYNTHASE 1* locus and a distal ABA-responsive enhancer. LncRNAs might constitute an additional regulatory layer driving the regulation of metabolic gene clusters.

ABSTRACT

Clustered organization of biosynthetic non-homologous genes is emerging as a characteristic feature of plant genomes. The co-regulation of clustered genes seems to largely depend on epigenetic reprogramming and three-dimensional chromatin conformation. Here we identified the long noncoding RNA (lncRNA) *MARneral Silencing* (*MARS*), localized inside the Arabidopsis marneral cluster, which controls the local epigenetic activation of its surrounding region in response to ABA. *MARS* modulates the POLYCOMB REPRESSIVE COMPLEX 1

(PRC1) component LIKE-HETEROCHROMATIN PROTEIN 1 (LHP1) binding throughout the cluster in a dose-dependent manner, determining H3K27me3 deposition and chromatin condensation. In response to ABA, *MARS* decoys LHP1 away from the cluster and promotes the formation of a chromatin loop bringing together the *MARNERAL SYNTHASE 1* (*MRN1*) locus and a distal ABA-responsive enhancer. The enrichment of co-regulated lncRNAs in clustered metabolic genes in Arabidopsis suggests that the acquisition of novel noncoding transcriptional units may constitute an additional regulatory layer driving the evolution of biosynthetic pathways.

KEYWORDS: lncRNA, enhancer, cluster, chromatin conformation, LHP1, ABA, seed germination, epigenetics, marneral

INTRODUCTION

In eukaryotes, functionally related genes are usually scattered across the genome. However, a growing number of operon-like clustered organization of non-homologous genes participating in common metabolic pathways point at an emerging feature of animal, fungi and plant genomes (Nützmann et al., 2016).

In plants, synthesis of numerous secondary metabolic compounds is important for the dynamic interaction with their environment, affecting their life and survival (Go et al., 2012). Terpenoids are bioactive molecules of diverse chemical structure (Yasumoto et al., 2016). In *Arabidopsis thaliana*, the biosynthesis and further elaboration of four triterpene scaffolds, namely thalianol (Field and Osbourn, 2008), tirucalla-7,24-dien-3b-ol (Boutanaev et al., 2015), arabidiol (Castillo et al., 2013) and marneral (Field et al., 2011), is governed by enzymes encoded by genes organized in clusters (Nützmann et al., 2016). The thalianol and marneral related genes are located in the smallest metabolic clusters identified in plants to date, each being less than 40kb in size (Nützmann et al., 2016). Both compounds are derived from 2,3-oxidosqualene and the corresponding gene clusters contain the oxidosqualene cyclases (OSCs), thalianol synthase (THAS) and marneral synthase (MRN1), respectively. The marneral cluster includes two additional protein-coding genes, *CYP705A12* and *MRO*, participating in marneral oxidation (Field et al., 2011).

Growing evidence indicates that the co-regulation of clustered genes relies on epigenetic mechanisms. It has been shown that the deposition of the histone variant H2A.Z

positively correlates with transcriptionally active clusters. Accordingly, nucleosome stability precluding gene expression is dependent on ARP6, a component of the SWR1 chromatin remodeling complex required for the deposition of H2A.Z into nucleosomes (Nützmann and Osbourn, 2015). Additionally, it was shown that the thalianol and marneral clusters exhibit increased expression in the Polycomb mutant *curly leaf* (*clf*) with compromised H3K27me3 deposition, and reduced expression in the trithorax-group protein mutant *pickle* (*pkl*), a positive regulator that counteracts H3K27me3 silencing (Yu et al., 2016). Strikingly, it has been recently shown that biosynthetic gene clusters are embedded in local hot spots of three-dimensional (3D) contacts that segregate cluster regions from the surrounding chromosome environment in a tissue-dependent manner. Notably, H3K27me3 appeared as a central feature of the 3D domains at silenced clusters (Nützmann et al., 2020).

Long noncoding RNAs (lncRNAs) have emerged as important regulators of eukaryotic gene expression at different levels (Rinn and Chang, 2020). In plants, several lncRNAs have been shown to interact with the Polycomb Repressive Complex 1 and 2 components LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) and CLF, respectively, which are related to H3K27me3 distribution (Berry et al., 2017; Lucero et al., 2020). Furthermore, it has been proposed that lncRNAs can modulate the transcriptional activity of neighboring genes by shaping local 3D chromatin conformation (Ariel et al., 2014; Kim and Sung, 2017; Gagliardi et al., 2019). Here we show that the marneral cluster in Arabidopsis includes three noncoding transcriptional units. Among them, the lncRNA *MARS* influences the expression of marneral cluster genes in response to ABA through modification of the epigenetic landscape. *MARS* deregulation affects H3K27me3 distribution, LHP1 deposition and chromatin condensation throughout the cluster. Furthermore, an ABA responsive chromatin loop dynamically regulates *MRN1* transcriptional activation by bringing together the *MRN1* proximal promoter and an enhancer element enriched in ABA-related transcription factors (TF) binding sites. *MARS*-mediated control of the marneral cluster affects seed germination in response to ABA. The general co-regulation of genes located within lncRNA-containing clusters in Arabidopsis points to noncoding transcription as an important feature in coordinated transcriptional activity of clustered loci.

RESULTS

The marneral gene cluster contains three noncoding transcriptional units

The small marneral cluster includes three genes: marneral synthase (*MRN1*), *CYP705A12* and *MRO/CYP71A16* that are two P450 cytochrome-encoding genes (**Figure 1A**), all participating in the biosynthesis and metabolism of the triterpene marneral (Field et al., 2011).

The advent of novel sequencing technologies has allowed the identification of an increasing number of lncRNAs throughout the *Arabidopsis* genome. According to the latest annotation (Araport 11 (Cheng et al., 2017)), three additional transcriptional units are located within the marneral cluster, between the *MRO* and the *MRN1* loci. The *AT5G00580* gene and the pair of antisense genes *AT5G06325* and *AT5G06335* are located upstream of the *MRN1* gene at 6kbp and 3kbp, respectively (**Figure 1A**). The 1,941bp-long *AT5G00580* locus generates four transcript isoforms ranging from 636 nt to 1,877 nt in length (**Figure 1B**). In contrast, each of the antisense genes *AT5G06325* and *AT5G06335* are transcribed into only one RNA molecule of 509 nt and 367 nt, respectively (**Figure 1A**). Additionally, two nascent transcripts can be detected within the marneral cluster (Ivanov et al., 2021). One within the *AT5G00580* locus and another one located in 3' and reverse strand of *MRO* with 1,505 and 564 nt in length, respectively (**Figure 1A and 1B**).

Focusing our analyses on the stable non-nascent transcripts, we assessed the coding potential of the Araport11-based transcripts and found that all of them can be classified as lncRNAs when using two coding prediction tools, CPC (Kong et al., 2007) and CPC2 (Kang et al., 2017) because of their low coding potential and their length (over 200 nt), similarly to previously characterized lncRNAs (*COLDAIR* (Heo and Sung, 2011); *APOLO* (Ariel et al., 2014); and *ASCO* (Bardou et al., 2014)) (**Figure 1C**).

According to available transcriptomic datasets (Araport11), *AT5G00580* transcriptional accumulation positively correlates with that of marneral genes, whereas *AT5G06325* and *AT5G06335* RNAs do not (**Figure S1**). Notably, our analysis of the transcriptional dynamics of the noncoding gene *AT5G00580* and the marneral cluster protein-coding genes revealed a correlated expression in response to phosphate and nitrate starvation, heat stress, as well as to exogenous auxin and ABA (**Figure 1D**). Interestingly, the *AT5G00580* lncRNA exhibited the strongest transcriptional induction in response to heat stress and exogenous ABA, in comparison with *MRN1* and the two *CYP* genes (**Figure 1D**). Notably, using isoform-specific probes, we showed that the *AT5G00580* isoforms 1, 2 and 3

are similarly induced in response to exogenous ABA, whereas we could not detect the isoform 4 (**Figure S2**).

Altogether, our observations uncovered that the marneral cluster includes three noncoding transcriptional units, one of which is actively transcribed and co-regulated with its neighboring protein-coding genes.

The lncRNA *MARS* shapes the transcriptional response of the marneral gene cluster to ABA

It has been shown that lncRNAs can regulate the expression of their neighboring genes through epigenetic mechanisms (Jarroux et al., 2017). Thus, we wondered if the lncRNAs derived from the *AT5G00580* locus may regulate the transcriptional activity of the protein-coding genes included in the marneral cluster. To this end, we modified the lncRNA expression without affecting the cluster DNA region by using an RNAi construct targeting *AT5G00580* transcripts, and isolated three independent lines. The RNAi lines impaired the transcriptional accumulation of *AT5G00580* without affecting significantly the rest of the cluster (**Figure 2A**). Notably, RNAi-mediated silencing did not trigger DNA methylation over the *AT5G00580* locus (**Figure S3**). Strikingly, the response of the three protein-coding genes of the marneral cluster to exogenous ABA was significantly deregulated in the RNAi lines (**Figure 2B and S4B**) compared to mock treatment (**Figure S5**), in contrast to the expression of two *AT5G00580*-unrelated ABA marker genes taken as a control of ABA treatment (**Figure S4C, D**). Remarkably, the *AT5G00580*-mediated deregulation was limited to the marneral cluster genes and did not affect the two closest neighboring genes upstream and downstream from the cluster (**Figure S6A**). Therefore, we named *the AT5G00580*-derived noncoding transcript *MARneral Silencing* (*MARS*) lncRNA. Transcriptional levels of *MRN1* and the two *CYP* genes increased earlier in RNAi-*MARS* seedlings (15 min) than in the wild-type (Col-0, 30 min) (**Figure 2B bottom panel**). In addition, the transcriptional accumulation of these genes later reached two-fold higher levels in the RNAi-*MARS* lines compared to Col-0 (**Figure 2B top panel and S4B**). The same behavior was observed using a higher concentration of ABA (**Figure S7**). Notably, none of the marneral cluster genes exhibit any transcriptional oscillation during the day (Covington et al., 2008; Hsu and Harmer, 2012; Romanowski et al., 2020), indicating that the transcriptional modulation of *MARS* and the marneral gene cluster linked to an ABA-mediated pathway.

To further support our observations, we isolated a transgene insertional mutant (*SALK_133089*) located 325 bp upstream the transcription start site (TSS) of *MARS* gene, according to public TSS/CAGE-seq datasets (Nielsen et al., 2019; Thieffry et al., 2020), that we named *mrs1-1* (**Figure S8**). We found that *mrs1-1* partially impairs the transcriptional accumulation of *MARS* and protein-coding genes in the marneral cluster, mainly *MRO* (**Figure 2A**). In agreement with the RNAi-*MARS* lines, compared to wild-type plants, *MRN1* and *CYP705A12* genes responded earlier to ABA and reached higher levels in *mrs1-1* (**Figure 2B**). In contrast, *MRO*, whose promoter region may be locally affected by the T-DNA insertion, was no longer responsive to ABA. In addition, we characterized a transgene insertional mutant in *MRO gene* (*mro1-2;* (Field et al., 2011)), which did not influence the expression of the other marneral cluster genes (**Figure S9A, B**) or known ABA-responsive genes (**Figure S9C, D**), suggesting that the transcriptional misregulation observed in *mrs1-1* mutant could be caused by the down-regulation of this lncRNA. In agreement, the deregulation of *MRN1* in over-expression and in knock-out *mrn1* mutant did not affect the expression of the *CYP* genes of the marneral cluster, nor ABA-responsive genes (**Figure S10**), but resulted in a decrease in *MARS* transcripts abundance under control condition (**Figure S10A, B**).

To further demonstrate that the *MARS* locus participates in the regulation of marneral cluster genes, we deleted the entire *MARS* locus by CRISPR/Cas9-mediated genome editing (**Figure S11A, B**). As observed in the *mrs1-1* mutant (**Figure 2A**), the *MARS* deletion impaired the transcriptional accumulation of the marneral protein-coding genes in control condition (**Figure S11C**), suggesting that the integrity of marneral cluster locus is necessary for normal expression of marneral clustered genes. As in the other lines deregulated for *MARS* (RNAi and *mrs1-1*), the marneral cluster genes were significantly more responsive to the exogenous ABA in the CRISPR-*MARS* line compared to wild-type plants (**Figure S11D, E**), in contrast to typical ABA marker genes (**Figure S11F, G**) and the expression of genes neighboring to the marneral cluster were not affected by the *MARS* deletion (**Figure S6B**).

Collectively, our results indicate that the noncoding transcriptional activity of *MARS* represses the dynamic expression of the marneral cluster genes, mainly *MRN1*, in response to ABA.

MARS affects seed germination and root growth under osmotic stress

The phytohormone ABA has been implicated in the perception and transduction of environmental signals participating in a wide range of growth and developmental events such as seed development, germination and root growth response to environmental stimuli (Vishwakarma et al., 2017).

Considering that the marneral cluster exhibited a strong *MARS*-dependent response to ABA, we wondered what was the physiological impact of *MARS* deregulation during seed germination. We assessed seed germination in Col-0 and *MARS* deregulated lines with or without exogenous ABA. Notably, *MARS* silencing resulted in a delayed germination compared to wild type seeds, both in response to ABA and in control conditions as revealed by an increase in T50 (time for 50% of germination; **Figure S12A-F**). Accordingly, *35S:MRN1* and *mro1-2* seeds also exhibit a delayed germination phenotype regardless of the treatment with ABA, whereas the germination speed rate of *mrn1* was only impaired in response to ABA (**Figure S12A-D**). The physiological behavior of the cluster-related mutants suggests that the misregulation of marneral genes in the *MARS* down-regulated lines could be linked to an increased sensitivity to ABA during germination (**Figure 2B**).

Considering the influence of the marneral cluster in the regulation of seed germination, we decided to assess root growth response to ABA and ABA-related environmental stimuli such as osmotic stress and salt. Except for *mrs1-1* which showed a significantly reduced root growth, all the other knock-down and mutant lines tested were not affected in root growth under normal growth conditions. However, all the genotypes presented a tendency to reduce root growth in response to ABA or hyperosmotic salt stress (**Figure S12G and S12H**). Notably, in response to osmotic stress mediated by mannitol, *MARS* knock-down lines and the lines with modified *MRN1* expression exhibited a weaker impact on lateral root development with increased lateral root density and length (**Figure S12G;** Lateral root length and Lateral root density). The similar behavior between *MRN1* deregulated lines and RNAi-*MARS* lines suggest that the decreased sensitivity to osmotic stress observed in the RNAi-*MARS* lines could be linked to *MRN1* misregulation.

To further understand the physiological phenotypes linked to *MARS* deregulation, we analyzed by RNA-seq the complete transcriptome of RNAi-*MARS* line 1 in response to ABA, compared to wild type plants. Even though none of the marneral cluster genes was identified as significantly deregulated due to a high heterogeneity between the replicates, their response to ABA is enhanced upon *MARS* downregulation (**Table S2**), in agreement with the RT-qPCR

experiments (**Figure 2B**). Globally, exogenous ABA triggers a strong transcriptional reprogramming in the wild type. Interestingly, in the RNAi line we found even more differentially expressed genes in response to ABA (3 238 versus 2 849 in Col-0), both in ABA induced or repressed genes (**Figure 3A and Table S2**).

However, under control conditions, only 27 genes were differentially expressed between the two genotypes, none of them implicated in seed germination, ABA or stress-related mechanisms (**Table S2**). To further investigate the impact of the *MARS*-deregulation on the plant ABA response we conducted a hierarchical clustering of all the differentially expressed genes (**Figure 3B and Table S2**). Two large clusters (cluster 1 and 2), grouping the great majority of the differentially expressed genes, exhibit a similar behavior in the two genotypes. Two smaller clusters (cluster 3 and 4) present opposite expression profiles between the genotypes (**Figure 3B, S13A and Table S2**). Inside these last clusters, we could identify a known regulator of the carbon/nitrogen balance and cell oxidation status, such as *QUA-QUINE STARCH* (*QQS*) (Li et al., 2015), *GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE 2* (*GGT2*) (Ohkama-Ohtsu et al., 2007), *ABC1-LIKE KINASE 3* (*ABC1K3*) (Martinis et al., 2013), *NITRITE REDUCTASE 1* (*NIR1*) (Shi et al., 2015) and *PHOSPHOENOLPYRUVATE CARBOXYLASE 2* (*PPC2*) (Duermeyer et al., 2018) (**Figure S13B**), in agreement with the seed germination and osmotic-mediated root growth defects observed in the *MARS*-related lines. In short, the downregulation of *MARS* mainly increases the reprogramming already induced by ABA (cluster 1 and 2), as for the marneral gene cluster, and affects only a few genes in an opposite way (cluster 3 and 4).

Collectively, our results indicate that *MARS* can modulate various ABA-related physiological responses at the transcriptional and physiological levels possibly through the regulation of *MRN1* expression.

MARS controls the epigenetic status of the marneral locus

It has been shown that gene clusters in plants are tightly regulated by epigenetic modifications, including the repressive mark H3K27me3 (Yu et al., 2016). ChIP-Seq datasets (Veluchamy et al., 2016) reveals that the marneral cluster region is highly enriched in H3K27me3 in shoots and overlaps with the deposition of LHP1 (**Figure S14)**. ATAC-Seq data (Sijacic et al., 2018) also revealed that the marneral cluster exhibits a high chromatin condensation in shoots (**Figure S14**). These data suggest that the marneral cluster is in an

epigenetically silent state in aerial organs, thus correlating with its low expression level in leaves (Yu et al., 2016). Even though marneral genes are preferentially expressed in roots compared to shoots (Field et al 2011; **Figure S15A**), their transcriptional induction by ABA is comparable in these organs and similarly impaired upon *MARS* deregulation (**Figure S15B, C**).

We wondered if the transcriptional activation of the marneral cluster in response to exogenous ABA was associated with a dynamic epigenetic reprogramming. We first assessed H3K27me3 deposition across the marneral cluster, including the gene body of *MRN1*, *MARS* and the two *CYP* loci (**Figure 4A and S16**). Interestingly, exogenous ABA triggered a strong reduction of H3K27me3 deposition throughout the marneral cluster (**Figure 4A and S16**). Markedly, H3K27me3 basal levels were also significantly lower in RNAi-*MARS* seedlings. Remarkably, H3K27me3 deposition was even lower across the body of all genes of the cluster in response to ABA in the RNAi-*MARS* lines when compared with Col-0, in agreement with the stronger induction by ABA of this subset of genes upon *MARS* silencing (**Figure 4A, S16 and Figure 2B**). Furthermore, we assessed the deposition of LHP1 on different regions of the marneral cluster and found that LHP1 was enriched at the *MRN1* promoter and more weakly across *MARS* gene body and the intergenic region between *MRO* and *MARS* (**Figure 4B and S17**). Remarkably, LHP1 recognition was strongly impaired in response to ABA as well as in RNAi-*MARS* seedlings (**Figure 4B and S17**). Therefore, our results indicate that ABA triggers an epigenetic reprogramming of the marneral cluster, likely in a process involving the lncRNA *MARS*.

MARS is directly recognized by LHP1 and modulates local chromatin condensation

It has been shown that the deposition of the repressive mark H3K27me3 and the concomitant recognition of the plant PRC1 component LHP1 are correlated with high chromatin condensation (Yang et al., 2017). Therefore, we determined the chromatin condensation of the whole marneral cluster by Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE). In contrast to Col-0 showing a highly condensed chromatin, RNAi-*MARS* seedlings exhibit a lower chromatin condensation in control conditions, including the *MARS* locus (**Figure S18).** Notably, the global chromatin status of the cluster was less condensed in RNAi-*MARS* seedlings in response to ABA (**Figure 5A and S18**), in agreement with a decrease of both H3K27me3 deposition and LHP1 binding (**Figure 4A, B, S16 and S17**) and the concomitant transcriptional activation of the clustered genes (**Figure 2B**).

Consistently, *lhp1* mutant seedlings also showed a global chromatin decondensation in control conditions, comparable to Col-0 in response to ABA. Notably, chromatin decondensation triggered by ABA was completely impaired in *lhp1* (**Figure 5B and S19**), supporting the role of LHP1 in the dynamic epigenetic silencing of the marneral cluster. Concomitantly, the increased chromatin decondensation of *lhp1* mutant seedlings correlates with increased abundance of marneral genes transcripts (**Figure S20A**), as observed in the RNAi-*MARS* seedlings (**Figure 2B**) with decondensed chromatin (**Figure 5A and S18**).

It has been shown that LHP1 can recognize RNAs *in vitro* (Berry et al., 2017) and the lncRNA *APOLO in vivo* (Ariel et al., 2014). Moreover, it has been proposed that *APOLO* over-accumulation can decoy LHP1 away from target chromatin (Ariel et al., 2020). Therefore, we wondered whether *MARS* lncRNA was able to interact with the chromatin-related protein LHP1 participating in the modulation of the local epigenetic environment. Thus, we first determined that two of the four *MARS* isoforms (isoform 1 and 2) were enriched in the nucleus, as the previously characterized lncRNAs *APOLO* and *ASCO* that interact respectively with nuclear epigenetic and splicing machineries, and the nuclear structural ncRNA *U6*, involved in the spliceosome (**Figure S20B**). Then, we investigated by RNA immunoprecipitation (RIP) whether LHP1 can interact with *MARS* RNA molecules *in vivo*. We found that LHP1 can interact with the two nuclear-enriched *MARS* RNA isoforms (isoform 1 and 2) but not with the third, non-nuclear enriched one (isoform 3), the *MRN1* RNA or a randomly selected housekeeping gene RNA (*PP2A*) used as negative controls (**Figure 5C**).

LHP1 is part of the PRC1 complex (Yang et al., 2017) suggesting that its interaction with *MARS* may be mediated by other components of the complex. Hence, to investigate it, we conducted an *in vitro* RNA-protein interaction assay. We incubated plant purified LHP1 protein from non-crosslinked *ProLHP1:LHP1:GFP* (Nakahigashi et al., 2005) seedlings with a mix of *in vitro* transcribed RNAs, including *GFP*, *APOLO*, *ASCO* and the *MARS* isoforms 1 and 2, the two *MARS* isoforms showing interaction with LHP1 *in vivo* (**Figure 5C**). The *MARS* RNA isoforms 1 and 2 were able to directly interact with the purified LHP1 protein, as the *APOLO* RNA used as a positive control, whereas the *ASCO* and *GFP* RNAs did not (**Figure 5D**). Thus, the ability of the *MARS* isoforms 1 and 2 to interact with LHP1 *in vivo* and *in vitro* suggests they may be both involved in the control of LHP1 binding to the chromatin.

Altogether, we showed that the *MARS* lincRNA physically interacts with the LHP1 protein, likely affecting LHP1 binding to the marneral cluster participating in the modulation of the dynamic chromatin organization of the locus.

***MARS* expression modulates an LHP1-dependent chromatin loop bringing together the *MRN1* locus and an ABA enhancer element**

It has been reported that the spatial conformation of cluster-associated domains differs between transcriptionally active and silenced clusters. In Arabidopsis, segregating 3D contacts are distinguished among organs, in agreement with the corresponding transcriptional activity of clustered genes (Nützmann et al., 2020). Therefore, we explored whether *MARS* could participate in the dynamic regulation of the local 3D chromatin conformation modulating the transcription of the marneral cluster.

According to available HiC datasets (Liu et al., 2016; Veluchamy et al., 2016) there is a significant 3D DNA-DNA contact linking the intergenic region between *MRO* and *MARS* and the *MRN1* locus (Chromatin Loop 1 (CL1)), together with another 3D contact linking the *MARS* locus and the intergenic region between *MARS* and *MRN1* (CL2) (**Figure S22A**). Notably, the longest chromatin loop (CL1; **Figure 6A**) was the only one detected in the two independent datasets (**Figure S22A**) (Liu et al., 2016; Veluchamy et al., 2016).

We used Chromatin Conformation Capture (3C) to monitor the formation of this chromatin loop and found that it increased drastically after 30 min exposure of seedlings to exogenous ABA and that this chromatin loop remained for at least 4 hours after the treatment (**Figure 6B**). These data indicate that the formation of this chromatin loop positively correlates with the transcriptional accumulation of the marneral cluster genes in response to ABA (**Figure 2B**). To further support the formation of the CL1 chromatin loop, we tested several primers within the region taking the forward and reverse primers used for CL1 quantification as anchors (1F/1R). Coherently, we were able to detect CL1 using different sets of primers (1F/8R and 2F/1R). In addition, we detected a new chromatin loop linking the same intergenic region with the *MARS* locus (CL3; **Figure S22 B, C**).

The *MARS* locus is encompassed in the ABA-dependent chromatin loop (**Figure 6A**). In order to determine the role of *MARS* in the modulation of local 3D chromatin conformation, we assessed the formation of the chromatin loop in RNAi-*MARS* lines.

Notably, RNAi-*MARS* seedlings exhibit enhanced chromatin loop formation, which remained unchanged in response to exogenous ABA (**Figure 6B**). Interestingly, LHP1 has been implicated in shaping local 3D conformation of target regions (Veluchamy et al., 2016), suggesting that the LHP1-*MARS* module may dynamically switch the epigenetic status of the marneral cluster from a condensed-linear to a decondensed-3D structured chromatin conformation. Supporting this hypothesis, *lhp1* mutant seedlings exhibited enhanced chromatin loop formation compared to Col-0 (**Figure 6C**). Overall, our results suggest that the formation of a chromatin loop within the marneral cluster is regulated by LHP1 through the interaction with *MARS* lncRNA transcripts.

To better understand the role of the *MARS*-dependent chromatin loop in response to ABA we looked for ABA-related *cis* regulatory sequences throughout the marneral cluster. We analyzed the distribution of binding sites for 21 ABA-related transcription factors (TFs) determined experimentally (Song et al., 2016). Interestingly, we found a high enrichment for ABA TF binding sites at the *MARS* locus, as well as in the intergenic region between the *MRO* and *MARS* loci, in particular at regions surrounding the contact point brought into close spatial proximity with the *MRN1* locus by the ABA-dependent 3D chromatin loop (**Figure 6A and S23**). We thus assessed the capacity of these genomic regions to enhance the transcriptional activity of *MRN1*. To this end, we generated transcriptional reporter lines combining the candidate distant enhancer elements to a minimal 35S promoter and *ß-glucuronidase* (*GUS*) gene (Yan et al., 2019). We also included as controls two genomic regions nearby the putative enhancers, one between *CYP705A12* and *MRO* and the other at the 3' end of *AT5G42620* locus (**Figure 6A**). Among the two putative distal enhancers tested, one was able to activate *GUS* expression in both heterologous and homologous system (Intergenic region 2, **Figure 6D and Figure S24),** coinciding with the region showing a high enrichment of ABA-related TF binding sites close to the chromatin loop anchor point (**Figure 6A**). Our results indicate that an ABA-driven chromatin loop brings into close spatial proximity the *MRN1* locus and a transcriptional activation region likely acting as an ABA enhancer element and that this chromatin reorganization process depends on the LHP1-*MARS* module.

Globally, our results indicate that low and high levels of *MARS* RNA lead to similar behavior for the marneral cluster. Indeed, upon *MARS* downregulation or ABA-mediated *MARS* upregulation, we observe the same chromatin conformation change dependent on LHP1 binding (**Figure 7A)**. Since *MARS* RNA interacts with LHP1 *in vivo* and *in vitro*

(**Figure 5C and 5D**), we hypothesize that basal levels of *MARS* RNA might drive LHP1 binding to the marneral cluster loci (**Figure 7A upper**) while at higher levels it might titrate it away (**Figure 7A right**), hinting at a potential *MARS*-LHP1 stoichiometric effect, regulating the binding of LHP1 to the chromatin (see discussion).

Hence, the physical interaction of the nuclear-enriched lncRNA *MARS* to LHP1 may modulate its binding to proximal chromatin, thus directly influencing the 3D chromatin conformation state of the marneral cluster in response to ABA.

Long noncoding RNAs as emerging regulators of gene clusters

Physically linked genes organized in clusters are generally coregulated (Nützmann et al., 2016). Considering that the lncRNA *MARS* is implicated in the regulation of the marneral cluster, we wondered whether the presence of noncoding transcriptional units may constitute a relevant feature of gene cluster organization. Therefore, we looked for the presence of lncRNAs in other gene clusters using two different datasets, one with co-expressed neighboring gene clusters (Yu et al., 2016) and one with metabolic gene clusters (PlantiSMASH (Kautsar et al., 2017)). Among the 197 clusters of co-expressed neighboring genes, 86 (44%) contained at least one lncRNA embedded within the cluster which is significantly less than for random regions of similar sizes (**Figure S25A**). However, among the 45 metabolic clusters, 33 (73%) include lncRNAs inside the cluster, which is more than for random genomic regions of similar sizes (**Figure 7C**). We then asked whether the presence of lncRNA inside a cluster could influence the observed expression correlation. Strikingly, the presence of at least one lncRNA in metabolic clusters significantly increases the coding gene expression correlation as opposed to comparable random genomic sequences (**Figure 7D**), whereas it is not the case for generally co-expressed clusters (**Figure S25B**). In agreement, lncRNAs present inside metabolic clusters tend to be more correlated with the other genes of the cluster compared to random clusters containing lncRNAs (**Figure 7E**). In contrast, in the case of generally co-expressed clusters, lncRNAs are significantly less correlated (**Figure S25C**). Altogether, our analyses suggest that lncRNAs embedded into clustered genes might be a conserved feature of metabolic clusters where the lncRNA likely participates in the clustered genes coexpression, as shown here for *MARS*.

DISCUSSION

The cell nucleus is a dynamic arrangement of DNA, RNAs and proteins (Cavalli and Misteli, 2013). Genome topology has emerged as an important feature in the complex network of mechanisms regulating gene activity and genome connectivity, leading to regionalized chromosomal spatial distribution and the clustering of diverse genomic regions with similar expression patterns (Rodriguez-Granados et al., 2016).

In the last few years, noncoding transcription has been implicated in shaping 3D nuclear organization (Quinodoz and Guttman, 2014). Notably, RNase-A micro-injection into the nucleus revealed that long nuclear-retained RNAs maintained euchromatin in a biologically active decondensed state, whereas heterochromatin domains exhibited an RNA-independent structure (Caudron-Herger et al., 2011; Caudron-Herger and Rippe, 2012). More recently, HiC analyses were performed in mammalian cells exposed or not to RNase, before and after crosslinking, or upon transcriptional inhibition (Barutcu et al., 2019). As a result, it was observed that topologically associated domains (TAD) boundaries remained mostly unaffected by RNase treatment, whereas compartmental interactions suffered a subtle disruption. In contrast, transcriptional inhibition led to weaker TAD boundaries, hinting at different roles of steady-state RNA vs. active transcription in nuclear organization (Barutcu et al., 2019).

In plants, several lncRNAs have been implicated in local chromatin conformation dynamics affecting the transcriptional activity of neighboring genes (Gagliardi and Manavella, 2020; Lucero et al., 2020). Notably, the lncRNA *COLDWRAP* participates in the formation of an intragenic chromatin loop blocking the transcription of the flowering time regulator *FLOWERING LOCUS C* (*FLC* (Kim and Sung, 2017)) in response to cold, in a process involving the recruitment of PRC2 by direct interaction with the component CLF. The lncRNA *APOLO* also controls the transcriptional activity of its neighboring gene *PINOID* (*PID*) by dynamically modulating the formation of an intergenic chromatin loop encompassing the divergent promoter of *PID* and *APOLO* (Ariel et al., 2014), in a process involving the PRC1 component LHP1. More recently, it was proposed that high levels of *APOLO* can decoy LHP1 away from multiple loci in *trans*, modulating the 3D conformation of distal target genes (Ariel et al., 2020). In rice, the expression of the leucine-rich repeat receptor kinase clustered genes *RLKs* is modulated by the locally-encoded lncRNA *LRK ANTISENSE INTERGENIC RNA* (*LAIR*). It was proposed that *LAIR* may directly recruit OsMOF (MALES ABSENT ON THE FIRST) and OsWDR5 (WD REPEAT DOMAIN 5), involved in H4K16 acetylation and chromatin remodeling (Wang et al., 2018). Conversely,

the Arabidopsis lncRNA *SVALKA*, embedded into the three cold-related *CBF* clustered genes (Medina et al., 2011), directly influences the activity of only one gene of the cluster (Kindgren et al., 2018) by transcriptional read-through and Pol II collision, indicating that functional lncRNA can present localized or cluster-spanning effects for the regulation of clustered genes. Here, we showed that the lncRNA *MARS* contributes to the co-regulation of a set of physically linked genes in *cis* in Arabidopsis. We demonstrated that the relative abundance of *in vitro*-transcribed *MARS* fine-tunes LHP1 binding to the cluster region in a stoichiometry-dependent manner, thus explaining how *MARS* levels affect H3K27me3 deposition and chromatin condensation. It has been shown in yeast that histone depletion boosts chromatin flexibility and facilitates chromatin loop formation on the kilobase pair scale (Diesinger et al., 2010). In agreement thereof, we uncovered here the dynamic role of the LHP1-*MARS* module affecting nucleosome distribution across the marneral cluster in response to ABA, thus promoting the formation of an intra-cluster chromatin loop.

It has been recently observed that biosynthetic gene clusters are embedded in local three-dimensionally organized hot spots that segregate the region from the surrounding chromosome environment (Nützmann et al., 2020). Here, we showed that active noncoding transcriptional units within the cluster may contribute to 3D conformation dynamics switching from silent to active states. Our results indicated that a *MARS*-dependent chromatin loop may bring the *MRN1* locus and a distal ABA-responsive element into close spatial proximity, likely acting as an enhancer. Notably, *MARS*-dependent LHP1 and H3K27me3 removal in Col-0, RNAi-*MARS* and the *lhp1* mutant correlated with chromatin decondensation, loop formation and increased marneral genes transcriptional activity in response to ABA. According to this model, we hypothesize that chromatin loop conformation depends on LHP1 binding and is modulated by *MARS* in a dual manner. In control condition, the basal *MARS* RNA levels might promote the binding of LHP1 to the chromatin of the marneral region (**Figure 4B and S17**), maintaining a condensed chromatin configuration of the cluster (**Figure 5A and S18**), in agreement with the low abundance of marneral protein-coding genes transcripts (**Figure 2B and 7A**). On one hand, the decrease of *MARS* RNA levels in the RNAi lines might reduce LHP1 binding correlating with a chromatin decondensation and enhancer-*MRN1* loop formation (CL1; **Figure 4B, S17, 5A, S18 and 6B**), priming the *MRN1* gene for expression (**Figure 7A left**). On the other hand, the ABA-dependent over-accumulation of *MARS* transcripts might titrate LHP1 from the chromatin region (**Figure 2B, 4B and S17**), thus promoting chromatin decondensation and facilitating the CL1 chromatin loop formation

(**Figure 5A, S18 and 6B)**, resulting in increased marneral genes expression (**Figure 2B and 7A right**). Hence, *MARS* RNA levels may directly affect the binding of LHP1 to the marneral cluster in a dual way, as already demonstrated for the *APOLO* lncRNA controlling the WRKY42 binding to the *RHD6* promoter (Moison et al 2021).

In mammals, growing evidence supports the role of lncRNAs in chromatin conformation determination (Gil and Ulitsky, 2020) and enhancer activity (e.g. *PVT1* (Cho et al., 2018) and *CCAT1-L* (Xiang et al., 2014)). Here, we showed that the nuclear-enriched lncRNA *MARS* brings together the *MRN1* proximal promoter and a putative enhancer element enriched in ABA-responsive TF binding sites. Interestingly, it has been shown that human lncRNAs can modulate the binding of TFs to their target chromatin (*DHFR* (Martianov et al., 2007)) and *PANDA* (Hung et al., 2011), whereas TFs have been implicated in chromatin loop formation in plants (Rodriguez-Granados et al., 2016). Furthermore, it was shown that in addition to the TF NF-YA, the lncRNA *PANDA* interacts with the scaffold-attachment-factor A (SAFA) as well as with PRC1 and PRC2 to modulate cell senescence (Puvvula et al., 2014). Therefore, further research will be needed to determine what ABA-responsive TFs are in control of the marneral cluster and to elucidate how they participate in chromatin loop formation along the area, in relation with the PRC1-interacting lncRNA *MARS*.

Plants are a tremendous source of diverse chemicals which are important for their life and survival (Yu et al., 2016). Marneral biosynthesis has been linked to root and leaf development, flowering time and embryogenesis (Go et al., 2012). Here we found that the Arabidopsis marneral cluster is activated by the phytohormone ABA, in a lncRNA-dependent epigenetic reprogramming. *MARS* deregulation affects the cluster response to ABA, impacting seed germination and root sensitivity to osmotic stress. Furthermore, whole transcriptome analysis revealed that the downregulation of *MARS* mainly increases the effect of the ABA treatment (**Figure 3A**), and drastically affects only a small number of genes independently of this hormonal stimulus (**Figure 3B and S13**). Interestingly, lncRNAs had already been associated with seed germination and environmental stress. For example, the overexpression of the cotton *lncRNA973* resulted in an increased seed germination rate and salt-tolerance in *Arabidopsis* (Zhang et al., 2019). Concomitantly, the decrease in *lncRNA973* transcript abundance in cotton was associated with hypersensitivity to salt stress. In addition, the *Arabidopsis* lncRNA *DRIR* regulates plant response to drought and salt stress by altering the expression of stress-responsive genes (Qin et al., 2017).

It was proposed that the marneral cluster was founded by the duplication of ancestral genes, independent events of gene rearrangement and the recruitment of additional genes (Field et al., 2011). The exploration of the noncoding transcriptome in Arabidopsis recently served to identify ecotype-specific lncRNA-mediated responses to the environment (Blein et al., 2020). It was suggested that the noncoding genome may participate in multiple mechanisms involved in ecotype adaptation. Collectively, our results indicate that the acquisition of novel noncoding transcriptional units within biosynthetic gene clusters may constitute an additional regulatory layer behind their natural variation in plant evolution.

METHODS

Lines selection and generation

All plants used in this study are in Columbia-0 background. RNAi-*MARS* were obtained using the pFRN binary vector (Ariel et al., 2012) bearing 250bp of the first exon of *MARS* gene (see primers in **Table S1**), previously sub-cloned into the pENTR/D-TOPO vector. Genomic deletion line for *MARS* was produced using two sgRNAs (**Table S1**) and CRISPR/Cas9 (Durr et al., 2018). Arabidopsis plants were transformed using *Agrobacterium tumefaciens* Agl-0 (Clough and Bent, 1998). The T-DNA inserted line *SALK_133089* was ordered to NASC (N633089). Homozygous mutants were identified by PCR (see primers in **Table S1**).

Seeds of *mrn1* (Go et al., 2012), *35S:MRN1* (Field et al., 2011), and *mro1-2* (*MRO*, (Field et al., 2011)) mutants were kindly provided by Dr. Ben Field (BIAM, CEA Cadarache, France) and Pr. Suh (Chonnam National University, Department of Bioenergy Science and Technology, Korea), respectively.

Growth conditions and phenotypic analyses

Seeds were sown in plates vertically placed in a growing chamber in long day conditions (16 h in light 150uE; 8 h in dark; 21°C) for all the experiments. Plants were grown on solid half-strength MS medium (MS/2) (Duchefa, M0222) supplemented with 0.7% sucrose, or without sucrose for the germination assay. For nitrate starvation assay, KNO_3 and $Ca(NO_3)_2$ were replaced from the MS/2 composition by a corresponding amount of KCl and $CaCl_2$ respectively, 2.25 mM NH_4HCO_3 was added for nitrate-containing medium. For the phosphate starvation assay, growth medium contained 0.15 mM $MgSO_4$, 2.1 mM NH_4NO_3, 1.9

mM KNO_3, 0.34 mM $CaCl_2$, 0.5 µM KI, 10 µM $FeCl_2$, 10 µM H_3BO_3, 10 µM $MnSO_4$, 3 µM $ZnSO_4$, 0.1 µM $CuSO_4$, 0.1 µM $CoCl_2$, 0.1 µM Na_2MoO_4, 0.5 $g.L^{-1}$ sucrose supplemented with 500µM NaH_2PO_4 for Pi containing medium versus 10µM for Pi free medium. All media were supplemented with 0.8g/L agar (Sigma-Aldrich, A1296 #BCBL6182V) and buffered at pH 5.6 with 3.4mM 2-(N-morpholino) ethane sulfonic acid.

For the molecular experiments (gene expression and epigenetic status analyses) seedlings were sprayed with water, 10µM to 100µM ABA and 10µM 1-Naphthaleneacetic acid (NAA), respectively. For heat stress, plates were transferred to a growth chamber at 37°C under the same lighting conditions. For nitrate and phosphate starvation assays, seedlings were transferred at day 12 after sowing (DAS) from respectively nitrate and phosphate containing medium to nitrate and phosphate free medium.

For root phenotyping, seeds were sown in control media and transferred at day 6 in control medium or root-growth-inhibiting medium containing 2µM ABA, 200mM mannitol or 100mM NaCl, respectively. After 3 days of growth, the root length was measured using RootNav software (Pound et al., 2013) from images taken with a flat scanner.

For seed germination assay, 0.5µM ABA was supplemented to the medium for ABA condition to slow down germination without completely restricting it. Germination rate was evaluated twice a day. Seeds were considered germinated when the seed coat was perforated by elongating radicles. Germination half time (T50) was estimated using time-to-event methods with the drc R package (v3.0 (Ritz et al., 2015)). Three-parameter log-logistic models were fitted on the seed germination data to each combination of genotype and ABA concentration:

$$G(t) = d1 + \exp \{b[\log(t) - \log(e)]\}$$

Where *G* is the fraction of germinated seeds at time *t*, *d* the germinable fraction, *e* the median germination time for the germinable fraction and *b* the slope around the inflexion point. Germination half time corresponds to the *e* parameter of these models.

For all the molecular experiments (qPCR, FAIRE, ChIP-qPCR, 3C), samples were taken from 12 DAS starting two hours after light illumination, at different time-points.

RT-qPCR

Total RNA was extracted from whole seedlings using TRI Reagent (Sigma-Aldrich) and treated with DNase (Fermentas) as indicated by the manufacturers. Reverse transcription was performed using 1µg total RNA and the Transcriptase inverse Maxima H minus Reverse (Thermo Scientific). qPCR was performed on a Light Cycler 480 with SYBR Green master I (Roche) in standard protocol (40 cycles, 60°C annealing). Primers used in this study are listed in **Table S1**. Data were analyzed using the ΔΔCt method with *PROTEIN PHOSPHATASE 2A SUBUNIT A3* (*AT1G13320*) for gene normalization (Czechowski et al., 2005) and time 0 for time-course experiments.

Library construction and sequencing

Three biological replicates of 12 DAS whole seedlings grown in control condition or sprayed with 10µM ABA were collected, four hours after the ABA treatment. RNA samples were extracted using TRI Reagent (Sigma-Aldrich) and treated with DNase (Fermentas) as indicated by the manufacturers. Libraries were processed using Illumina Truseq Stranded mRNA library preparation kit following the manufacturer's instructions, starting with one microgram of total RNA. 75-nt single-end reads were sequenced on a NextSeq 500 Sequencing System (Illumina). Sequence files generated in this study have been deposited in the NCBI GEO database under the accession GSE192382.

Differential expression analysis and clustering

Adapter and poor-quality sequences were trimmed using Trimmomatic and ribosomal sequences were removed using sortMeRNA (Kopylova et al., 2012). Cleaned mRNA reads were aligned on the TAIR10 genome (Lamesch et al., 2012) using STAR (version 2.7.2a; (Dobin et al., 2013)) with the following arguments: --runThreadN 2 --sjdbGTFfile Araport11.gtf --readFilesCommand zcat --alignIntronMin 20 --alignIntronMax 3000 --outSAMtype BAM SortedByCoordinate --outReadsUnmapped Fastx --outBAMsortingBinsN 100. FeatureCounts from the subread package (v1.6.5, (Liao et al., 2014)) were used for read counting using strand specific mode (-s 2 -O -M --fraction). Differential gene expression analysis was performed with DEseq2 (v1.16.1; (Love et al., 2014)) using a linear model and as factors the genotype and the treatment (two levels each). Low counts were discarded using DESeq2 independent filtering with default parameters and raw p-values were adjusted with Bonferroni method. Differentially expressed genes were defined as having an adjusted p-value lower than 0.01.

To generate the heatmap, the DEseq2 normalized count of all the differentially expressed genes were processed into a maximum parsimony phylogenetic tree with four factors: the two genotypes (Col-0 and RNAi-*MARS*) and the two conditions (control and ABA-treated) with the help of the pheatmap (v1.0.12; (Kolde, 2019)) R packages. The optimal number of clusters were defined through the Elbow method with pkgs (v1.8.0; (Zhang et al., 2018)). Clustering profiles were determined from the clustered heatmap.

Chromatin Immunoprecipitation (ChIP)

ChIP was performed using anti-IgG (Millipore, Cat#12-370), anti-H3K27me3 (Millipore, Cat#07-449) and anti-LHP1 (Covalab, Pab0923-P), as previously described (Ariel et al., 2014), starting from two grams of seedlings crosslinked in 1% (v/v) formaldehyde. Chromatin was sonicated in a water bath Bioruptor Plus (Diagenode; 30 cycles of 30s ON and 30s OFF pulses at high intensity). ChIP was performed using an SX-8G IP-Star Compact Automated System (Diagenode). Antibody-coated Protein A Dynabeads (Invitrogen) were incubated 12 hours at 4 °C with the samples. Immunoprecipitated chromatin was reverse cross-linked with 20µg of Proteinase K (Thermo, EO0491) overnight at 65°C. Finally, DNA was recovered using Phenol:Chloroform:Isoamilic Acid (25:24:1, Sigma) followed by ethanol precipitation. For input samples, ten percent of sonicated chromatin were collected for each sample before the immunoprecipitation and were reverse cross-linked and extracted as the immunoprecipitated samples. Results are expressed as an enrichment, corresponding to the H3K27me3 or LHP1 percent of input divided by the background provided by the IgG percent of input.

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)

FAIRE was performed as described by (Simon et al., 2012). Briefly, cross-linked chromatin was isolated as for the ChIP and sonicated for 10 cycles in a water bath Bioruptor Plus (Diagenode; cycle of 30s ON and 30s OFF at high intensity). Ten percent of the sonicated chromatin was collected and diluted in 10 mM Tris-HCl pH 8 to a final volume of 500 µl.

Decondensed chromatin was separated from the condensed one, by a first extraction in Phenol:Chloroform:Isoamilic Acid (25:24:1, Sigma) followed by ethanol precipitation. After a reverse-crosslink overnight at 65 °C with 20 µg of proteinase K (Thermo, EO0491), this decondensed chromatin fraction was recovered a second time using

Phenol:Chloroform:Isoamilic Acid (25:24:1, Sigma) extraction, followed by ethanol precipitation. For input samples, diluted sonicated chromatin was directly reverse-crosslinked and total DNA, disregarding its condensed state, was recovered using Phenol:Chloroform:Isoamilic Acid (25:24:1, Sigma) extraction followed by ethanol precipitation. Decondensed vs total DNA quantification was performed by qPCR using the same set of primers as for ChIP. Results were expressed as the percent of input.

Immunoprecipitation of methylated DNA (meDIP)

MeDIP was performed as described by (Nagymihály et al., 2017). For genomic DNA purification, 100mg of non-cross-linked seedlings were incubated 30min at 65°C in 600µL of cetyltrimethylammonium bromide (CTAB) buffer (2% CTAB, 1.4M NaCl, 100mM Tris pH8, 20mM EDTA and 0.2% B-mercaptoethanol). A Chloroform:Isoamyl Alcohol (24:1) wash was performed prior to precipitation with isopropanol. After RNAse A treatment, 1µg of pure DNA was sonicated in a water bath Bioruptor Plus (Diagenode; 10 cycles of 30s ON and 30s OFF pulses at low intensity). The IP of the methylated DNA was performed overnight at 4°C using Protein A Dynabeads coated with anti-5mC (Diagenode, C15200081) or anti-IgG (Diagenode, C15400001). Immunoprecipitated DNA was recovered using Phenol:Chloroform:Isoamyl Alcohol (25:24:1, Sigma) followed by ethanol precipitation and quantified by qPCR. For input samples, non-immunoprecipitated sonicated chromatin was processed in parallel. Results are expressed as an enrichment, corresponding to the 5mC percent of input divided by the background provided by the IgG percent of input.

Nuclear purification

Non-cross-linked seedlings were used to assess the sub-cellular localization of RNAs. To obtain the nuclear fraction, chromatin was purified as for ChIP and resuspended, after the sucrose gradient, into 1mL of TRI Reagent (Sigma-Aldrich). For the total fraction, 200 µL of cell suspension from the first step of the ChIP protocol, were treated with 800 µL of TRI Reagent to follow with the RNA extraction. RNA samples were treated with DNase, and RT was performed using random hexamers prior to qPCR analysis. Data were analyzed using the ΔΔCt method using *PROTEIN PHOSPHATASE 2A SUBUNIT A3* (*AT1G13320*) for gene normalization (Czechowski et al., 2005) and the total fraction to assess nuclear enrichment.

RNA immunoprecipitation (RIP)

For *in vivo* RIP, *lhp1* mutant complemented with the *ProLHP1:LHP1:GFP* (Nakahigashi et al., 2005) was treated for 4 hours with ABA. After crosslinking and chromatin extraction as for ChIP. Chromatin was sonicated in a water bath Bioruptor Plus (Diagenode; 5 cycles of 30 s ON and 30 s OFF pulses at high intensity) and ten percent of sonicated chromatin was kept at -20 °C as the input.

The IP of the sonicated chromatin was separated into two fractions and incubated overnight at 4°C with pre-coated anti-GFP beads (Chromotek, gtma-20) or non-coated control beads (Chromotek, bmab-20), both supplemented with 1µL of RNAse inhibitor (RNaseOUT, Invitrogen, 10777019). Beads were washed twice with each of the cold RIP Washing buffer 1 (150mM NaCl, 1% Triton X100, 0.1% SDS, 0.5% Na deoxycholate, 50mM Tris HCl pH8), 2 (1% Na deoxycholate, 1% NP-40, 1mM EDTA, 10mM Tris-HCl pH8) and 3 (20mM Tris HCl pH8) for 5min at 4°C under soft agitation. Proteins were degraded with 20 µg of proteinase K (Thermo, EO0491) for 2 hours at 55 °C followed by 15 min at 95 °C. 1µL of RNAse inhibitor (RNaseOUT, Invitrogen, 10777019) were added prior reverse crosslinking to protect RNAs. RNAs were recovered treating supernatants with TRI Reagent, and followed by DNase treatment and reverse transcription using random hexamers to be finally quantified by qPCR. Results are expressed as an enrichment, corresponding to the GFP percent of input divided by the background provided by the control percent of input.

In vitro RNA-protein interaction

LHP1 protein was isolated from the *ProLHP1:LHP1:GFP* (Nakahigashi et al., 2005) line through immunoprecipitation and washes. Briefly, purified non-crosslinked chromatin was prepared as for ChIP from five grams of *ProLHP1:LHP1:GFP* (Nakahigashi et al., 2005). LHP1's IP was realized overnight at 4 °C using pre-coated anti-GFP beads (Chromotek, gtma-20) together with 100U/mL of TURBO DNAse (Invitrogen, AM1907) and 100µg/mL of RNAseA (Sigma-Aldrich) enabling removal of all LHP1 associated nucleic acids. Beads were washed twice with each of the cold ChIP Dilution buffer (1.1% Triton X100, 1.2mM EDTA, 16.7mM Tris-HCl pH8 and 167mM NaCl), Low Salt buffer (150mM NaCl, 0.1% SDS, 1% Triton X100, 2mM EDTA and 20mM Tris-HCl ph8), High Salt buffer (500mM NaCl, 0.1% SDS, 1% Triton X100, 2mM EDTA and 20mM Tris-HCl ph8) and LiCl buffer (0.25M LiCl, 1% Na Deoxycholate, 1% NP-40, 1mM EDTA and 10mM Tris-HCl pH8).

Detection and enrichment of LHP1 protein were checked using Western Blot and silver nitrate staining. For both, proteins were extracted from beads with 2X SDS-loading Buffer for 10 min at 75°C and loaded on two SDS PAGE gels. Controls consist of anti-GFP beads (Chromotek, gtma-20) treated as the samples. Samples were transferred from the first gel onto a nitrocellulose membrane and blotted with GFP antibodies (abcam, ab6789). The second gel was silver nitrate stained as (Mortz et al., 2001).

The washed beads were resuspended into 1mL of Binding Buffer (10mM HEPES pH7, 50mM KCl, 10% glycerol, 1mM EDTA, 1mM DTT, 0.5% Triton X-100) supplemented with 1µL of RNase inhibitor (RNaseOUT, Invitrogen, 10777019) and incubated for 1 hour at 4 °C with a mix of *in vitro* transcribed RNAs (HiScribe T7 High Yield RNA Synthesis Kit, NEB), among which 1µg of *GFP* RNA, 100ng of *ASCO* RNA, 100ng of *APOLO* RNA, 100ng of each *MARS* isoforms 1 and 2. Notably, the *GFP* RNA was added in excess as compared to the other plant RNAs, strengthening the relevance of the RNA-protein interactions detected. The *ASCO*, *APOLO* and *MARS* isoform 2 RNAs were obtained from transcription of a PCR product amplified from wild-type genomic DNA. The *MARS* isoform 1 was first cloned into the pJET1.2 plasmid (Thermo, K1231) and amplified from this plasmid (**Table S1**). Finally, the *GFP* RNA was amplified from the pB7FWG2 plasmid (**Table S1**). Random RNA/GFP antibody interaction was determined by incubating pre-coated anti-GFP beads (Chromotek, gtma-20) with the same RNA mix. Washes and RNA recovery from the beads were then performed as for *in vivo* RIP, described above. Results are expressed as an enrichment, corresponding to the immunoprecipitated GFP percent of input divided by the background provided by the non-immunoprecipitated GFP percent of input.

Chromosome conformation capture (3C)

3C was performed as previously described (Louwers et al., 2009). Briefly, chromatin was extracted from two grams of cross-linked seedlings as for ChIP. Overnight digestion at 37 °C was performed using 400U of Hind III enzyme (NEB). Digested DNA was ligated for 5 h at 16 °C with 100 U of T4 DNA ligase (NEB). DNA was recovered after reverse crosslinking and Proteinase K treatment (Invitrogen) by Phenol:Chloroform:Isoamyl Acid (25:24:1; Sigma) extraction and ethanol precipitation. Interaction frequency was determined by qPCR using probes upstream and downstream indicated HindIII restriction sites. Results are expressed as the relative enrichment of an HindIII uncut genomic region to normalize the

amount of DNA across samples (background) (Kim and Sung, 2017; Ariel et al., 2020; Zhao et al., 2021).

Transcriptional activation assay in tobacco leaves

The *GUS* reporter system for validation of putative enhancer element activity was adapted from (Yan et al., 2019). Different promoter regions were cloned in the pGGA000 vector using GreenGate system (Lampropoulos et al., 2013) and fused to a minimal 35S promoter element from CAMV (synthesized by Eurofins Genomics). The sub-unit B3 of the 35S promoter element from CAMV (Moreno-risueno et al., 2010) was synthesized and used as a positive control. All primers used for cloning are indicated in **Table S1**.

Heterologous *A. tumefaciens*-mediated transient transformation was performed on 5-weeks-old tobacco plants using a needleless syringe. Together with enhancer constructs, a vector containing mCherry driven by a 35S promoter was co-transfected to determine transformation efficiency by fluorescence observation under epifluorescent microscope. Three days after infiltration, two leaf discs were collected near the infiltration site, one, to determine the transfection efficiency and the second one for GUS staining, as previously described (Jefferson et al., 1987). In addition, a single leaf was also infiltrated with the four different constructs to compare by GUS staining their relative activity.

Homologous *A. tumefaciens*-mediated transient transformation was performed on 4-weeks-old Arabidopsis plants using a needleless syringe as described in (Zhang et al., 2020). For this assay, complete 35S promoter driving GUS construct was used as positive control. All samples were incubated overnight in the dark at 37 °C before observation.

Identification of lncRNA loci in Arabidopsis gene clusters

The genes of co-expressed clusters (197 clusters) were retrieved from (Yu et al., 2016). The boundaries of the gene clusters were extracted using Araport11 annotations. The boundaries of the metabolic clusters were extracted from the plantiSMASH predicted clusters (45 clusters) on Arabidopsis (Kautsar et al., 2017). As control for each cluster, 1000 random genomic regions of the same size were computed. These regions were then regrouped by type of cluster, resulting in 197 000 random genomic regions as control for the coexpressed genes clusters and 45 000 random genomic regions as control for the metabolic clusters. Using Araport11 GFF, for each cluster (co-expressed, metabolic or random), the content in coding

genes and lncRNAs (genes with a locus type annotated as “long_noncoding_rna”, “novel_transcribed_region” or “other_rna”) present within the boundaries of the cluster were retrieved. Based on these annotations, the number of clusters harboring lncRNA was extracted for each cluster and compared to corresponding random genomic regions.

Gene expression correlation analyses

To compute the correlation of expression in different Arabidopsis organs we used the 113 RNA-seq datasets previously considered for the Araport11 annotations (Cheng et al., 2017). These datasets were generated from untreated or mock-treated wild-type Col-0 plants. After removal of the adaptors with Trim Galore using default parameters, reads were mapped on TAIR10 with STAR v2.7.2a (Dobin et al., 2013) and the parameters ‘--alignIntronMin 20 --alignIntronMax 3000’. Gene expression was then quantified with featureCounts v2.0.0 (Liao et al., 2014) with the parameters “-B -C -p -s 0” using the GFF of Araport11. Raw counts were then normalized by median of ratios using the DESeq2 R package (Love et al., 2014).

For the expression correlation analysis inside the marneral cluster, the transcript levels of the genes included in the cluster and 25kb around it (four genes upstream and two downstream) were considered. Pearson's correlations for each pair of genes were computed after log 2 transformation of the normalized counts. The correlation value and associated p-value were plotted with the corrplot R package (Wei et al., 2017).

Inside each co-expressed and metabolic cluster of genes and their respective random set, and after removing poorly expressed genes (detected in less than 20 libraries), Pearson’s correlation was computed between every possible pair of lncRNA and coding gene as well as for the genes inside the marneral cluster. The co-expressed clusters of Yu et al., 2016 are not continuous: some non-coregulated genes might be present inside the genomic regions of the cluster. Therefore, we compute the correlation analysis comprising all the genes included in the cluster boundaries (complete cluster) or with only the one annotated as co-expressed (co-expressed cluster). The maximum correlation value was kept as an indication of lncRNAs correlation with the genes of the cluster. As an indication of gene coexpression, we also compute the median Pearson’s correlation between all the coding genes of the clusters.

Quantification and statistical analyses

For all the experiments, at least three independent biological samples were considered. For RT-qPCR, each sample was prepared from a pool of 5 to 10 individual seedlings. For biochemistry assays (ChIP, FAIRE, nuclear purification, RIP and 3C) two to five grams of seedlings were prepared for each independent biological sample. For validation of enhancer function, the four leaf discs were taken from four or three independent tobacco or Arabidopsis plants, respectively. An additional replicate was performed on three independent tobacco plants upon the agroinfiltration of all the different constructs on the same leaf. The tests used for statistical analyses are indicated in the respective figure legends. Statistical tests and associated plots have been generated using R software (v3.6.3(R Core, 2004)) with the help of the tidyverse package (Wickham et al., 2019).

FUNDING

This work was supported by BBSRC grant (BB/L016966/1) to J.G-M and Saclay Plant Sciences-SPS (ANR-17-EUR-0007) and CNRS (Laboratoire International Associé NOCOSYM) to MC and FA. TR was awarded a PhD scholarship from the French “Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation” (MESRI).

AUTHOR CONTRIBUTIONS

T.R., F.A., T.B. and M.C. conceived and designed the experiments. T.R. performed the experiments. Y.H. participates in the ChIP experiments. N.H. and A.C. generated the CRISPR line. T.R. and T.B. analyzed the data. All authors discussed the results and edited the manuscript.

ACKNOWLEDGEMENT

We thank Jeremie Bazin for the helpful discussion about results interpretation and design of the experiments. We thank Camille Fonouni-Farde for her wise suggestions on the artistic improvement of our model. We thank Olivier Martin for critical reading of the manuscript.

AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Martin Crespi (martin.crespi@cnrs.fr).

Plant lines generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

REFERENCES

Ariel, F., Brault-Hernandez, M., Laffont, C., Huault, E., Brault, M., Plet, J., Moison, M., Blanchet, S., Ichanté, J. L., Chabaud, M., et al. (2012). Two direct targets of cytokinin signaling regulate symbiotic nodulation in medicago truncatula. *Plant Cell* **24**:3838–3852.

Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., and Crespi, M. (2014). Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* **55**:383–396.

Ariel, F., Lucero, L., Christ, A., Mammarella, M. F., Jegu, T., Veluchamy, A., Mariappan, K., Latrasse, D., Blein, T., Liu, C., et al. (2020). R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. *Mol. Cell* **77**:1–11.

Bardou, F., Ariel, F., Simpson, C. G., Romero-Barrios, N., Laporte, P., Balzergue, S., Brown, J. W. S., and Crespi, M. (2014). Long Noncoding RNA Modulates Alternative Splicing Regulators in Arabidopsis. *Dev. Cell* **30**:166–176.

Barutcu, A. R., Blencowe, B. J., and Rinn, J. L. (2019). Differential contribution of steady-state RNA and active transcription in chromatin organization . *EMBO Rep.* **20**:1–13.

Berry, S., Rosa, S., Howard, M., Bühler, M., and Dean, C. (2017). Disruption of an RNA-binding hinge region abolishes LHP1-mediated epigenetic repression. *Genes Dev.* **31**:2115–2120.

Blein, T., Balzergue, C., Roulé, T., Gabriel, M., Scalisi, L., François, T., Sorin, C., Christ, A., Godon, C., Delannoy, E., et al. (2020). Landscape of the non-coding

transcriptome response of two Arabidopsis ecotypes to phosphate starvation. *Plant Physiol.* **183**:pp.00446.2020.

Boutanaev, A. M., Moses, T., Zi, J., Nelson, D. R., Mugford, S. T., Peters, R. J., and Osbourn, A. (2015). Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**:E81–E88.

Castillo, D. A., Kolesnikova, M. D., and Matsuda, S. P. T. (2013). An effective strategy for exploring unknown metabolic pathways by genome mining. *J. Am. Chem. Soc.* **135**:5885–5894.

Caudron-Herger, M., and Rippe, K. (2012). Nuclear architecture by RNA. *Curr. Opin. Genet. Dev.* **22**:179–187.

Caudron-Herger, M., Müller-Ott, K., Mallm, J. P., Marth, C., Schmidt, U., Fejes-Tóth, K., and Rippe, K. (2011). Coding RNAs with a non-coding function: Maintenance of open chromatin structure. *Nucleus* **2**.

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**:290–299.

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **89**:789–804.

Cho, S. W., Xu, J., Sun, R., Mumbach, M. R., Carter, A. C., Chen, Y. G., Yost, K. E., Kim, J., He, J., Nevins, S. A., et al. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173**:1398-1412.e22.

Clough, S. J., and Bent, A. F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* **16**:735–743.

Covington, M. F., Maloof, J. N., Straume, M., Kay, S. A., and Harmer, S. L. (2008). Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol.* **9**.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W. (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. *Plant Physiol.* **139**:5–17.

Diesinger, P. M., Kunkel, S., Langowski, J., and Heermann, D. W. (2010). Histone depletion facilitates chromatin loops on the kilobasepair scale. *Biophys. J.* **99**:2995–3001.

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.

Duermeyer, L., Khodapanahi, E., Yan, D., Krapp, A., Rothstein, S. J., and Nambara, E. (2018). Regulation of seed dormancy and germination by nitrate. *Seed Sci. Res.* **28**:150–157.

Durr, J., Papareddy, R., Nakajima, K., and Gutierrez-Marcos, J. (2018). Highly efficient heritable targeted deletions of gene clusters and non-coding regulatory regions in Arabidopsis using CRISPR/Cas9. *Sci. Rep.* **8**:1–11.

Field, B., and Osbourn, A. E. (2008). Clusters in Different Plants. *Science (80-.).* **194**:543–547.

Field, B., Fiston-Lavier, A.-S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A. E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci.* **108**:16116–16121.

Gagliardi, D., and Manavella, P. A. (2020). Short-range regulatory chromatin loops in plants. *New Phytol.* Advance Access published 2020, doi:10.1111/nph.16632.

Gagliardi, D., Cambiagno, D. A., Arce, A. L., Tomassi, A. H., Giacomelli, J. I., Ariel, F. D., and Manavella, P. A. (2019). Dynamic regulation of chromatin topology and transcription by inverted repeat-derived small RNAs in sunflower. *Proc. Natl. Acad. Sci. U. S. A.* **116**:17578–17583.

Gil, N., and Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* **21**:102–117.

Go, Y. S., Lee, S. B., Kim, H. J., Kim, J., Park, H. Y., Kim, J. K., Shibata, K., Yokota, T., Ohyama, K., Muranaka, T., et al. (2012). Identification of marneral synthase, which is critical for growth and development in Arabidopsis. *Plant J.* **72**:791–804.

Heo, J. B., and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science (80-.).* **331**:76–79.

Hsu, P. Y., and Harmer, S. L. (2012). Circadian Phase Has Profound Effects on Differential Expression Analysis. *PLoS One* **7**:18–21.

Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., Horlings, H. M., Shah, N., Umbricht, C., Wang, P., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**:621–629.

Ivanov, M., Sandelin, A., and Marquardt, S. (2021). TrancriptomeReconstructoR: data-driven annotation of complex transcriptomes. *BMC Bioinformatics* **22**:1–15.

Jarroux, J., Morillon, A., and Pinskaya, M. (2017). Long Non Coding RNA Biology. *Adv. Exp. Med. Biol.* **1008**:1–46.

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987). GUS fusions: ‚B-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**:3901–3907.

Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., and Gao, G. (2017). CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**:W12–W16.

Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A., and Medema, M. H. (2017). PlantiSMASH: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* **45**:W55–W63.

Kim, D.-H., and Sung, S. (2017). Vernalization-triggered intragenic chromatin-loop formation by long noncoding RNAs. *Dev. Cell* **176**:100–106.

Kindgren, P., Ard, R., Ivanov, M., and Marquardt, S. (2018). Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation. *Nat. Commun.* **9**.

Kolde, R. (2019). Package ‘pheatmap’: pretty heat map. *Advance Access published* 2019.

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., and Gao, G. (2007). CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**:345–349.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**:3211–3217.

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., et al. (2012). The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **40**:1202–1210.

Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegele, I., Lohmann, J. U., and Forner, J. (2013). GreenGate - A novel, versatile, and efficient cloning system for plant transgenesis. *PLoS One* **8**.

Li, L., Zheng, W., Zhu, Y., Ye, H., Tang, B., Arendseea, Z. W., Jones, D., Li, R., Ortiz, D., Zhao, X., et al. (2015). QQS orphan gene regulates carbon and nitrogen partitioning across species via NF-YC interactions. *Proc. Natl. Acad. Sci. U. S. A.* **112**:14734–14739.

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**:923–930.

Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M., and Weigel, D. (2016). Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. *Genome Res.* **26**:1057–1068.

Louwers, M., Splinter, E., van Driel, R., de Laat, W., and Stam, M. (2009). Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). *Nat. Protoc.* **4**:1216–1229.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**:1–21.

Lucero, L., Fonouni-Farde, C., Crespi, M., and Ariel, F. (2020). Long noncoding RNAs shape transcription in plants. *Transcription* **00**:1–12.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**:666–670.

Martinis, J., Glauser, G., Valimareanu, S., and Kessler, F. (2013). A chloroplast ABC1-like kinase regulates vitamin E metabolism in arabidopsis. *Plant Physiol.* **162**:652.

Medina, J., Catalá, R., and Salinas, J. (2011). The CBFs: Three arabidopsis transcription factors to cold acclimate. *Plant Sci.* **180**:3–11.

Moreno-risueno, M. A., Norman, J. M. Van, Moreno, A., Zhang, J., Ahnert, S. E., and Benfey, P. N. (2010). NIH Public Access **329**:1306–1311.

Nagymihály, M., Veluchamy, A., Györgypál, Z., Ariel, F., Jégu, T., Benhamed, M., Szücs, A., Kereszt, A., Mergaert, P., and Kondorosi, É. (2017). Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **114**:4543–4548.

Nakahigashi, K., Jasencakova, Z., Schubert, I., and Goto, K. (2005). The Arabidopsis HETEROCHROMATIN PROTEIN1 homolog (TERMINAL FLOWER2) silences genes within the euchromatic region but not genes positioned in heterochromatin. *Plant Cell Physiol.* **46**:1747–1756.

Nielsen, M., Ard, R., Leng, X., Ivanov, M., Kindgren, P., Pelechano, V., and Marquardt, S. (2019). Transcription-driven Chromatin Repression of Intragenic Promoters. *Plos Gen.* 1-33

Nützmann, H. W., and Osbourn, A. (2015). Regulation of metabolic gene clusters in Arabidopsis thaliana. *New Phytol.* **205**:503–510.

Nützmann, H. W., Huang, A., and Osbourn, A. (2016). Plant metabolic clusters – from genetics to genomics. *New Phytol.* **211**:771–789.

Nützmann, H., Doerr, D., Ramírez-colmenero, A., Sotelo-fonseca, J. E., Fernandez-valverde, S. L., and Osbourn, A. (2020). Active and repressed biosynthetic gene clusters have spatially distinct chromosome states Advance Access published 2020, doi:10.1073/pnas.1920474117.

Ohkama-Ohtsu, N., Radwan, S., Peterson, A., Zhao, P., Badr, A. F., Xiang, C., and Oliver, D. J. (2007). Characterization of the extracellular γ-glutamyl transpeptidases, GGT1 and GGT2, in Arabidopsis. *Plant J.* **49**:865–877.

Pound, M. P., French, A. P., Atkinson, J. A., Wells, D. M., Bennett, M. J., and Pridmore, T. (2013). RootNav: Navigating Images of Complex Root Architectures. *Plant Physiol.* **162**:1802–1814.

Puvvula, P. K., Desetty, R. D., Pineau, P., Marchio, A., Moon, A., Dejean, A., and Bischof, O. (2014). Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nat. Commun.* **5**.

Qin, T., Zhao, H., Cui, P., Albesher, N., and Xionga, L. (2017). A nucleus-localized long non-coding rna enhances drought and salt stress tolerance. *Plant Physiol.* **175**:1321–1336.

Quinodoz, S., and Guttman, M. (2014). Long non-coding RNAs: An emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* **24**:651–663.

Rinn, J. L., and Chang, H. Y. (2020). Long Noncoding RNAs: Molecular Modalities to Organismal Functions. *Annu. Rev. Biochem.* **89**:283–308.

Ritz, C., Baty, F., Streibig, J. C., and Gerhard, D. (2015). Dose-response analysis using R. *PLoS One* **10**:1–13.

Rodriguez-Granados, N. Y., Ramirez-Prado, J. S., Veluchamy, A., Latrasse, D., Raynaud, C., Crespi, M., Ariel, F., and Benhamed, M. (2016). Put your 3D glasses on: Plant chromatin is on show. *J. Exp. Bot.* **67**:3205–3221.

Romanowski, A., Schlaen, R. G., Perez-Santangelo, S., Mancini, E., and Yanovsky, M. J. (2020). Global transcriptome analysis reveals circadian control of splicing events in Arabidopsis thaliana. *Plant J.* **103**:889–902.

Shi, J., Yi, K., Liu, Y., Xie, L., Zhou, Z., Chen, Y., Hu, Z., Zheng, T., Liu, R., Chen, Y., et al. (2015). Phosphoenolpyruvate carboxylase in arabidopsis leaves plays a crucial role in carbon and nitrogen metabolism. *Plant Physiol.* **167**:671–681.

Sijacic, P., Bajic, M., McKinney, E. C., Meagher, R. B., and Deal, R. B. (2018). Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. *Plant J.* **94**:215–231.

Simon, J. M., Giresi, P. G., Davis, I. J., and Lieb, J. D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat. Protoc.* **7**:256–267.

Song, L., Huang, S. S. C., Wise, A., Castanoz, R., Nery, J. R., Chen, H., Watanabe, M., Thomas, J., Bar-Joseph, Z., and Ecker, J. R. (2016). A transcription factor hierarchy defines an environmental stress response network. *Science (80-.).* **354**:598.

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.

Veluchamy, A., Jégu, T., Ariel, F., Latrasse, D., Mariappan, K. G., Kim, S. K., Crespi, M., Hirt, H., Bergounioux, C., Raynaud, C., et al. (2016). LHP1 Regulates H3K27me3 Spreading and Shapes the Three-Dimensional Conformation of the Arabidopsis Genome. *PLoS One* **11**:1–25.

Vishwakarma, K., Upadhyay, N., Kumar, N., Yadav, G., Singh, J., Mishra, R. K., Kumar, V., Verma, R., Upadhyay, R. G., Pandey, M., et al. (2017). Abscisic acid signaling and abiotic stress tolerance in plants: A review on current knowledge and future prospects. *Front. Plant Sci.* **8**:1–12.

Wang, Y., Luo, X., Sun, F., Hu, J., Zha, X., Su, W., and Yang, J. (2018). Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat. Commun.* **9**:1–9.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., and Zemla, J. (2017). Visualization of a Correlation Matrix. *Statistician* **56**:316–324.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* **4**:1686.

Xiang, J. F., Yin, Q. F., Chen, T., Zhang, Y., Zhang, X. O., Wu, Z., Zhang, S., Wang, H. Bin, Ge, J., Lu, X., et al. (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **24**:513–531.

Yan, W., Chen, D., Schumacher, J., Durantini, D., Engelhorn, J., Chen, M., Carles, C. C., and Kaufmann, K. (2019). Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. *Nat. Commun.* **10**:1–16.

Yang, X., Tong, A., Yan, B., and Wang, X. (2017). Governing the silencing state of chromatin: The roles of polycomb repressive complex 1 in arabidopsis. *Plant Cell Physiol.* **58**:198–206.

Yasumoto, S., Fukushima, E. O., Seki, H., and Muranaka, T. (2016). Novel triterpene oxidizing activity of Arabidopsis thaliana CYP716A subfamily enzymes. *FEBS Lett.* **590**:533–540.

Yu, N., Nützmann, H. W., Macdonald, J. T., Moore, B., Field, B., Berriri, S., Trick, M., Rosser, S. J., Kumar, S. V., Freemont, P. S., et al. (2016). Delineation of

metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res.* **44**:2255–2265.

Zhang, A. X., Bjorklund, N., Alvare, G., Ryzdak, T., Sparling, R., Fristensky, B., Alvare, M. G., and Zhang, X. (2018). Package ‘ ELBOW ’ *Advance Access published* 2018.

Zhang, X., Dong, J., Deng, F., Wang, W., Cheng, Y., Song, L., Hu, M., Shen, J., Xu, Q., and Shen, F. (2019). The long non-coding RNA lncRNA973 is involved in cotton response to salt stress. *BMC Plant Biol.* **19**:459.

Zhang, Y., Chen, M., Siemiatkowska, B., Toleco, M. R., Jing, Y., Strotmann, V., Zhang, J., Stahl, Y., and Fernie, A. R. (2020). A Highly Efficient Agrobacterium-Mediated Method for Transient Gene Expression and Functional Studies in Multiple Plant Species. *Plant Commun.* **1**:100028.

Zhao, B., Xi, Y., Kim, J., and Sung, S. (2021). Chromatin architectural proteins regulate flowering time by precluding gene looping. *Sci. Adv.* **7**:1–11.

TABLE AND FIGURES LEGENDS

Figure 1. *AT5G00580* is a lncRNA transcribed from the marneral cluster locus and its expression correlates with its neighboring genes

(A) Schematic illustration of the marneral cluster. Genes are indicated with a plain rectangle and white arrows indicate the sense of transcription. The square indicates the region displayed in (B).

(B) Schematic illustration of transcripts found within the *AT5G00580* gene locus. First line corresponds to the *AT5G00580* genomic region whereas the other lines present the various Araport11-based isoforms and nascent transcript. For each isoform, exons are indicated with rectangles and introns with solid lines.

(C) Coding potential of the transcripts located in the marneral cluster genomic region. Scores were determined using CPC1 (left) and CPC2 (right) algorithms (Kong et al., 2007; Kang et al., 2017). For each, the threshold between coding and noncoding genes is displayed with a horizontal solid black line. Coding genes are situated above the threshold, whereas non-

coding genes are situated under. *COLDAIR*, *APOLO* and *ASCO* are used as positive controls for non-coding transcripts.

(D) Dynamic transcriptional levels of genes of the marneral cluster in response to phosphate and nitrate starvation, heat stress, and exogenous ABA and auxin treatment or stimuli. Gene expression data are shown as the mean ± standard error (n = 3) of the log2 fold change compared to time 0h.

Figure 2. *MARS* transcriptional activity modulates the response to ABA of the marneral cluster

(A) Transcript abundance of the marneral cluster genes in control conditions in RNAi lines targeting *AT5G00580*/*MARS* and *mrs1-1* mutant (*SALK_133089)*. Transcriptional abundance is shown as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype. Results were analyzed by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test. For each gene, different letters indicate statistical differences between genotypes ($p \leq 0.05$).

(B) Transcript levels of the genes of the marneral cluster in response to ABA treatment in RNAi lines targeting *AT5G00580*/*MARS* and *mrs1-1* mutant (*SALK_133089*). Gene expression data are shown as the mean ± standard error (n = 3) of the log2 fold change compared to time 0h.

Figure 3. *MARS* downregulation increases ABA effect on the transcriptome

(A) Number of differentially expressed (DE) genes under ABA treatment in Col-0 and RNAi-*MARS* line 1.

(B) Hierarchical clustering based on the z-score (scaled and centered expression) of all DE genes according to genotype or response to ABA. Colors correspond to the z-score in each condition.

Figure 4. *MARS* modulates the epigenetic landscape of *MRN1* locus

(A) and (B) H3K27me3 deposition (A) and LHP1 binding (B) over the *MRN1* locus in Col-0 and RNAi-*MARS* 1 and 2 seedlings under control conditions and in response to ABA. Higher values indicate more H3K27me3 or LHP1 deposition, respectively. Values under the dotted

line are considered as not enriched. Results are shown as the mean ± standard error ($n \geq 4$) of the H3K27me3/IgG or LHP1/IgG ratio resulting from the fitting of a linear mixed model taking into account genotype and replicates. Results were analyzed by two-way analysis of variance with (ANOVA) including genotype and time as additive factors. For each genotype and condition, different letters indicate statistical differences according to Tukey's post-hoc test ($p \leq 0.05$).

Figure 5. *MARS* influences chromatin condensation of *MRN1* gene through its interaction with LHP1 protein

(A) and (B) Evolution of the chromatin condensation state of the *MRN1* locus in Col-0, RNAi-*MARS* lines and *lhp1* mutant subjected to ABA treatment determined by Formaldehyde Assisted Isolation of Regulatory Element (FAIRE) qPCR. Results are shown as the mean ± standard error ($n = 3$) of the percentage of input (signal measured on total chromatin fraction, before isolation of a decondensed region). Lower value indicates more condensed chromatin. Results were analyzed by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test ($p \leq 0.05$) or Student t-test between the two genotypes for (A) and (B), respectively.

(C) *In vivo* LHP1-*MARS* interaction was assessed by RNA immunoprecipitation (RIP) using *LHP1-GFP* seedlings. Negative controls include a housekeeping gene (*PP2A*) and *MRN1* mRNA. *MRN1* and *PP2A* transcript levels in nuclei samples are comparable to *MARS*. The interaction between *APOLO* and LHP1 was taken as a positive control (Ariel et al., 2014). Results are shown as the mean ± standard error ($n = 4$) of the GFP/IgG ratio.

(D) *In vitro* LHP1-*MARS* interaction was assessed by RNA immunoprecipitation (RIP) using purified LHP1 proteins from *LHP1-GFP* seedlings incubated with various *in vitro* transcribed RNA. Negative controls include the *GFP* and *ASCO* RNAs. The interaction between *APOLO* and LHP1 was taken as a positive control (Ariel et al., 2014). Results are shown as the mean ± standard error ($n = 4$) of the GFP/control ratio.

In (C) and (D) numbers are FDR corrected p-values of the difference between the different corresponding genes determined by Student t-test.

Figure 6. An LHP1-dependent chromatin loop approaches the *MRN1* locus with a putative enhancer element in response to ABA

(A) Schematic illustration of the Chromatin Loop 1 (CL1) linking the *MRN1* locus with the intergenic region between *MRO* and *MARS*. Forward (F) and Reverse (R) oligonucleotides used for 3C-qPCR (in B–C) are indicated with arrows. The orange and blue tracks show the number of ChIP peaks for different ABA-related transcription factor (HB5, HB6, HB7, GBF2, GBF3, MYB3, MYB44, NF-YC2, NF-YB2, ANAC102, ANAC032, ABF1, ABF3, ABF4, RD26, ZAT6, FBH3, DREB2A, AT5G04760, HAT22 and HSFA6A) found on the marneral cluster in ABA and mock conditions, respectively (Song et al., 2016). Individual raw ChIP peaks are provided in Figure S23. Green and red rectangles indicate the putative enhancer region and the negative controls, respectively, tested for the GUS-based reporter system in (D).

(B) and (C) Relative chromatin loop formation in response to ABA in Col-0, RNAi-*MARS* and *lhp1* mutant seedlings. Results are shown as the mean ± standard error ($n \geq 3$) from 3C-qPCR using primer F and R shown on (A), compared to time 0h. Letters indicate a statistical group determined by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test ($p \leq 0.05$).

(D) Constructs used for the GUS-based reporter system are illustrated on the left. Corresponding transformed tobacco leaf discs are on the right ($n = 4$). First line represents the positive control in which the 35S sub-unit controls *GUS* expression. The second and third lines show two independent negative controls in which the *GUS* gene is driven by a genomic region that does not contain ABA-related binding sites indicated in (A). In the remaining lines, the transcriptional activity is assessed for the two intergenic regions indicated in (A).

Figure 7. LHP1-dependent *MARS*-mediated regulation of a chromatin loop spanning the marneral clustered genes

(A) The lncRNA *MARS* regulates the expression of the marneral cluster genes through epigenetic reprogramming and chromatin conformation. In control conditions (upper panel) the chromatin of the marneral cluster is enriched in H3K27me3 and LHP1, which results in a condensed chromatin conformation (illustrated by shorter gene length and increased DNA strand thickness). Upon *MARS* repression (bottom left panel), LHP1 recruitment to the cluster is impaired. The decrease of LHP1 deposition diminishes H3K27me3 distribution, relaxes the chromatin and as a consequence allows the formation of the chromatin loop CL1 that brings into close proximity the enhancer element and the *MRN1* proximal promoter. Under this

chromatin state, the clustered genes become highly responsive to the ABA treatment. In response to ABA (bottom right panel) *MARS* over-accumulated transcripts titrate LHP1 from the cluster, thus leading to a similar chromatin state: decrease in H3K27me3 mark, chromatin decondensation and increase in CL1 chromatin loop conformation, leading to an ABA-mediated transcriptional activation.

(B) Number and proportion of lncRNAs containing metabolic clusters compared to control random genomic regions. The number shown above indicates the p-value of the difference between the frequency of lncRNA found within these regions determined by a binomial exact test.

(C) Influence of the presence of lncRNA on the median level of coding gene expression correlation inside metabolic clusters or in control random genomic regions. The numbers shown above indicate the FDR corrected p-values of the differences for the genomic regions with and without lncRNAs determined by Student t-test. The marneral metabolic cluster is indicated with a red dot.

(D) Maximum level of correlation between a lncRNA and any coding gene of the cluster in metabolic clusters or in control random genomic regions. The number shown above indicates the p-value of the difference between metabolic clusters and control random genomic regions determined by Student t-test. The marneral metabolic cluster is indicated with a red dot.

LEGENDS TO SUPPLEMENTAL FIGURES

Figure S1. *MARS* is coregulated with the genes of the marneral cluster

Pearson correlation analysis derived from transcriptomics data from Araport11. Correlations between two genes are indicated with scores ranging from -1 to +1 where -1 corresponds to a negative correlation and +1 a positive correlation. A color scale indicates the Pearson correlation score. Each correlation was tested for significant differences (* for $p \leq 0.05$, ** for $p \leq 0.01$, *** for $p \leq 0.001$).

Figure S2. Detected *MARS* isoforms are similarly induced in response to ABA.

Transcript abundance of the various *MARS* isoforms in response to ABA treatment in Col-0. Gene expression data are shown as the mean ± standard error (n = 4) of the log2 fold change compared to time 0h. The isoform 4 has not been detected.

Figure S3. The RNAi-transgene does not affect DNA methylation at the *AT5G00580/MARS* locus.

DNA methylation of the *AT5G00580/MARS* locus in Col-0 and RNAi-*MARS* seedlings assessed by MeDIP-qPCR under control condition. Higher values indicate 5mC enrichment. *APOLO* region has been taken as a positive control of 5mC enrichment. Data are expressed as the mean ± standard error (n = 2) of the 5mC/IgG ratio. Letters indicate a statistical group determined by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test. For each genotype, letters indicate statistical difference between genomic position ($p \leq 0.05$).

Figure S4. *MARS* modulates the response of marneral genes to ABA without altering the plant sensitivity to an exogenous treatment

(A) Transcript abundance of *MARS* in response to ABA treatment in RNAi-*MARS* lines and *mrs1-1*.

(B) Average genotype effect on transcript levels of each marneral cluster gene in RNAi-*MARS* lines and *mrs1-1* compared to Col-0 in response to 10µM ABA. Data are presented in Fig 2B.

(C) Transcript abundance of two ABA marker genes, *RAB18* and *RD29B*, in response to 10µM ABA in RNAi-*MARS* lines and *mrs1-1*.

(D) Average genotype effect on the transcript levels of two ABA marker genes, *RAB18* and *RD29B*, in RNAi-*MARS* lines and *mrs1-1* compared to Col-0 in response to 10µM ABA. Data are presented in Fig S4C.

(C) Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype at time 0h. (B) and (D) The mean genotype effect was estimated with two-way analysis of variance (ANOVA) including genotype and time as additive factors. For each effect, numbers indicate the p-value of the difference between the RNAi lines and Col-0 by Tukey's post-hoc test.

Figure S5. Marneral cluster genes do not exhibit a circadian rhythm behavior

(A) Transcript abundance of the marneral cluster genes in response to water treatment in RNAi-*MARS* lines. Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype at time 0h.

(B) Average genotype effect on the transcript levels of the marneral cluster genes in independent RNAi-*MARS* lines compared to Col-0 along water treatment according to two-way analysis of variance (ANOVA) including genotype and time as additive factors. Data are presented in Fig S5A. For each effect, numbers indicate the p-value of the difference between the RNAi lines and Col-0 by Tukey's post-hoc test.

(C) Transcript abundance of two ABA marker genes, *RAB18* and *RD29B*, in response to water treatment in RNAi-*MARS* lines. Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype at time 0h.

(D) Average genotype effect on the transcript levels of two ABA marker genes, *RAB18* and *RD29B*, in the independent RNAi-*MARS* lines compared to Col-0 in response to water treatment according to two-way analysis of variance (ANOVA) including genotype and time as additive factors. Data are presented in Fig S5C. For each effect, numbers indicate the p-value of the difference between the RNAi lines and Col-0 by Tukey's post-hoc test.

Figure S6. *MARS* action is restricted to the marneral cluster genes.

(A) and (B) Transcript abundance of the marneral cluster's upstream and downstream genes in response to 10 µM ABA in (A) RNAi-*MARS* lines and *mrs1-1* and in (B) CRISPR-*MARS*. Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype (A) or the control genotype (WT for the *MARS* locus on the segregating population) (B) at time 0h.

Figure S7. *MARS* modulates the response of marneral genes to high concentrations of ABA

(A) Transcript abundance of marneral cluster genes in response to 100 µM ABA in RNAi-*MARS* lines. Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype at time 0h.

(B) Average genotype effect on the transcript levels of marneral cluster genes in RNAi-*MARS* lines compared to Col-0 in response to 100µM ABA according to two-way analysis of variance (ANOVA) including genotype and time as additive factors. Data are presented in Fig S7A. For each effect, numbers indicate the p-value of the difference between the RNAi lines and Col-0 by Tukey's post-hoc test.

(C) Transcript abundance of two ABA marker genes in response to 100µM ABA in RNAi-*MARS* lines. Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype at time 0h.

(D) Average genotype effect on the transcript levels of two ABA marker genes in the different RNAi lines targeting AT5G00580/*MARS* compared to Col-0 in response to 100µM ABA according to two-way analysis of variance (ANOVA) including genotype and time as additive factors. Data are presented in Fig S7C. For each effect, numbers indicate the p-value of the difference between the RNAi lines and Col-0 by Tukey's post-hoc test.

Figure S8. The *SALK_133089/mrs1-1* insertional mutant is located within the *MARS* promoter

Genome browser view of TSS-seq (Nielsen et al 2019) and CAGE-seq (Thieffry et al 2020). The T-DNA from the *SALK_133089/mrs1-1* is located at 325bp upstream from the TSS of *MARS*.

Figure S9. Deregulation of *MRO* does not modulate marneral cluster genes expression nor plant sensitivity to ABA

(A) Transcript abundance of marneral cluster genes in response to 10µM ABA in *mro1-2 (MRO* knock-out*)* mutant.

(B) Average genotype effect on the transcript levels of the marneral cluster genes in the *mro1-2 (MRO* knock-out*)* mutant compared to Col-0 in response to 10µM ABA. Data are presented in Fig S9A.

(C) Transcript abundance of two ABA marker genes in response to 10µM ABA in *the mro1-2 (MRO* knock-out*)* mutant.

(D) Average genotype effect on the transcript levels of two ABA marker genes in the *mro1-2 (MRO* knock-out*)* mutant compared to Col-0 in response to 10µM ABA. Data are presented in Fig S9C.

(A) and (C) Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype at time 0h. (B) and (D). The mean genotype effect was estimated with two-way analysis of variance (ANOVA) including genotype and time as additive factors. For each effect, numbers indicate the p-value of the difference between the *MRO* mutant and Col-0 by Tukey's post-hoc test.

Figure S10. Deregulation of *MRN1* does not modulate marneral cluster genes expression nor plant sensitivity to ABA

(A) Transcript abundance of the marneral cluster genes in response to 10µM ABA in *35S:MRN1* and *mrn1* mutants.

(B) Average genotype effect on the transcript levels of the marneral cluster genes in *35S:MRN1* and *mrn1* mutants compared to Col-0 in response to 10µM ABA. Data are presented in Fig S10A.

(C) Transcript abundance of two ABA marker genes in response to 10µM ABA in *35S:MRN1* and *mrn1* mutants.

(D) Average genotype effect on the transcript levels of two ABA marker genes in *35S:MRN1* and *mrn1* mutants compared to Col-0 in response to 10µM ABA Data are presented in Fig S10C.

(A) and (C) Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype at time 0h. (B) and (D). The mean genotype effect was estimated with two-way analysis of variance (ANOVA) including genotype and time as additive factors. For each effect, numbers indicate the p-value of the difference between the *MRO* mutant and Col-0 by Tukey's post-hoc test.

Figure S11. Complete *MARS* locus deletion disturbs marneral cluster genes expression

(A) Schematic representation of the marneral locus showing the CRISPR/Cas9-mediated genomic deletion of *MARS*. Forward (F) and Reverse (R) oligonucleotides used for genotyping (in B) are indicated with arrows.

(B) DNA amplification of the genomic region encompassing the *MARS* gene. The position of the primers used for the amplification are displayed on A. MWM stands for Molecular Weight Marker (GeneRuler 1 kb Plus DNA Ladder, Thermo Scientific).

(C) Transcript abundance of the marneral cluster genes in the CRISPR-*MARS* line. Gene expression data are expressed as the mean ± standard error (n = 4) of the log2 fold change compared to the control genotype (WT for the *MARS* locus on the segregating population). Letters indicate a statistical group determined by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test. For each genotype, letters indicate statistical difference between T50 ($p \leq 0.05$).

(D) Transcript abundance of the marneral cluster genes in response to 10µM ABA in the CRISPR-*MARS* line.

(E) Average genotype effect on the transcript levels of the marneral cluster genes in the CRISPR-*MARS* line compared to the control genotype (WT for the *MARS* locus on the segregating population) in response to 10µM ABA. Data are presented in Fig S11D.

(F) Transcript abundance of two ABA marker genes in response to 10µM ABA in the CRISPR-*MARS* line.

(G) Average genotype effect on the transcript levels of the genes of two ABA marker genes in the CRISPR-*MARS* line compared to the control genotype (WT for the *MARS* locus on the segregating population) in response to 10µM ABA. Data are presented in Fig S11D.

(D) and (F) Gene expression data are expressed as the mean ± standard error (n = 4) of the log2 fold change compared to time 0h. (E) and (G) The mean genotype effect was estimated with two-way analysis of variance (ANOVA) including genotype and time as additive factors. For each effect, numbers indicate the p-value of the difference between the CRISPR-*MARS* line and control by Tukey's post-hoc test.

Figure S12. *MARS* modulates seed germination and mannitol-dependent root growth through the regulation of *MRN1* expression

(A) to (F) Percentage of germinated seeds in a (A)(C)(E) control or (B)(D)(F) 0.5µM ABA supplemented medium. Results are expressed as the mean ±standard error (n = 7) from a batch of 50 seeds collected from plants grown together. Time for 50% germination (T50) is indicated on the right.

(G) Mean primary root length, lateral root length and lateral root density according to the genotype and the condition of 11-day-old seedlings.

Results were analyzed by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test. For each genotype, different letters indicate statistical differences ($p \leq 0.05$) between T50 (A-F) or root architecture parameter (G).

(H) Representative pictures of roots of each genotype and conditions from (G). Scale bar = 1 cm.

Figure S13. *MARS* downregulation impaired genes implicated in carbon/nitrogen equilibrium and cell oxidation status

(A) Profile of expression (z-score) in each identified cluster from (B). Each grey line corresponds to a gene. Red lines represent the mean of all the genes of the cluster.

(B) Transcript abundance of selected genes from cluster 3 and 4 implicated in the Carbon/Nitrogen equilibrium or cell oxidation status. Gene expression data are expressed as the mean ± standard error (n = 3) of the normalized counts. Numbers are Bonferroni corrected p-values of the difference between the two genotypes determined by Wald test. Genes with low counts were removed through independent filtering independently for each conducted test.

Figure S14. Epigenetic landscape of the marneral cluster and surrounding genomic region

First track represents DNA accessibility determined by ATAC-seq (Sijacic et al., 2018). ATAC-peaks are indicated in blue and correspond to relaxed chromatin. Second and third tracks show H3K27me3 and LHP1 ChIP-Seq, respectively (Veluchamy et al., 2016). The

three experiments shown here have been performed using Arabidopsis shoot. Gene annotation is shown at the top.

Figure S15. *MARS* deregulation disturbs the ABA responsiveness of marneral cluster genes in shoots and roots

(A) Transcript abundance of the marneral cluster genes in shoots and roots of Col-0 plants.

(B) and (C) Transcript abundance of the marneral cluster genes in shoots and roots of (B) RNAi-*MARS* and *mrs1-1* and (C) CRISPR-*MARS* line in response to 10µM ABA treatment.

(A-C) Gene expression data are expressed as the mean ± standard error (n = 4) of the log2 fold change compared to shoots (A) and to Col-0 or the control genotype (WT for the *MARS* locus on the segregating population) at time 0h (B or C, respectively). For each gene, each letter indicates statistical difference between time ($p \leq 0.05$).

Figure S16. *MARS* influences H3K27me3 deposition in the marneral cluster region

H3K27me3 deposition in the intergenic region, *CYPs* and *MARS* loci was assessed by ChIP-qPCR in Col-0 and RNAi-*MARS* seedlings in control conditions and after 4 hours of ABA treatment. Higher values indicate H3K27me3 enrichment. Values under the dotted line are considered as not enriched. Data are expressed as the mean ± standard error ($n \geq 4$) of the H3K27me3/Igg ratio estimated through the fitting of a linear mixed model taking into account genotype and replicates. Results were analyzed by two-way analysis of variance with (ANOVA) including genotype and time as additive factors. For each genotype and condition, different letters indicate statistical differences according to Tukey's post-hoc test ($p \leq 0.05$).

Figure S17. *MARS* modulates LHP1 binding across the marneral cluster region

LHP1 binding to the intergenic region, *CYPs* and *MARS* loci was assessed by ChIP-qPCR in Col-0 and RNAi-*MARS* seedlings in control conditions and in response to ABA treatment. Higher values indicate LHP1 enrichment. Values under the dotted line are considered as not enriched. Data are expressed as the mean ± standard error ($n \geq 4$) of the LHP1/Igg ratio. Results were analyzed by two-way analysis of variance with (ANOVA) including genotype and time as additive factors. For each genotype and condition, different letters indicate statistical differences according to Tukey's post-hoc test ($p \leq 0.05$).

Figure S18. *MARS* modulates chromatin condensation of the marneral cluster genomic region

Chromatin condensation in the intergenic region, *CYPs* and *MARS* loci was assessed by Formaldehyde Assisted Isolation of Regulatory Element (FAIRE)-qPCR in Col-0 and RNAi-*MARS* seedlings in control conditions and in response to ABA treatment. Lower value indicates more condensed chromatin. Results are expressed as the mean ± standard error (n = 3) of the percentage of input (signal measured before isolation of decondensed region of chromatin; free of nucleosomes). Results were analyzed by two-way analysis of variance with (ANOVA) including genotype and time as additive factors. For each genotype and condition, different letters indicate statistical differences according to Tukey's post-hoc test ($p \leq 0.05$).

Figure S19. LHP1 is involved in chromatin condensation modulation of the marneral cluster region

Chromatin condensation in the intergenic region, *CYPs* and *MARS* loci was assessed by Formaldehyde Assisted Isolation of Regulatory Element (FAIRE)-qPCR in Col-0 and *lhp1* mutant seedlings in control conditions and in response to ABA treatment. Lower value indicates more condensed chromatin. Results are expressed as the mean ± standard error (n = 3) of the percentage of input (signal measured before isolation of decondensed region of chromatin; free of nucleosomes). Results were analyzed by two-way analysis of variance with (ANOVA) including genotype and time as additive factors. For each time, numbers represent the p-value of the difference to Col-0 according to Tukey's post-hoc test.

Figure S20. The *MARS* isoform1 and 2 are enriched in the nucleus and may participate in the LHP1-mediated regulation of marneral cluster genes expression

(A) Transcript levels of the marneral cluster genes in response to ABA treatment in *lhp1* mutant. Results are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to time-point 0h. Results were analyzed by two-way analysis of variance with (ANOVA) including genotype and time as additive factors. For each time, numbers represent the p-value of the difference to Col-0 according to Tukey's post-hoc test.

(B) Nuclear enrichment of the lncRNA *MARS* compared to other nuclear-enriched lncRNAs is determined as the ratio of transcript abundance in the nuclear fraction compared to total cellular RNA fraction. Higher value indicates nuclear enrichment. *APOLO*, *ASCO* and *U6*

RNA have been used as positive controls whereas *PP2A* (*AT1G13320;* housekeeping gene) has been used as negative control. Results are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the total cell fraction. Numbers are FDR corrected p-values of the difference between the corresponding RNA determined by Student t-test.

Figure S21. Isolation of the plant LHP1 protein

(A) Western blot image of the 76kDa GFP-LHP1 protein (26.9 kDa and 48.6 kDa for GFP and LHP1, respectively), shown in the IP well and revealed with anti GFP antibody. The control corresponds to anti-GFP beads (Chromotek, gtma-20) solely. Raw and enhanced contrasted images are shown in the left and right panel, respectively.

(B) Silver staining image of the 76kDa GFP-LHP1 protein (26.9 kDa and 48.6 kDa for GFP and LHP1, respectively). The control corresponds to anti-GFP beads (Chromotek, gtma-20) solely. Raw and enhanced contrasted images are shown in the left and right panel, respectively.

Figure S22. Chromatin loop formation between *MRO, MARS* and the *MRN1* 5' locus

(A) Chromatin loops found in the marneral cluster in two independent HiC datasets (Liu et al 2016; Veluchamy et al 2016) displayed using the Spring model webservice from Kadlof et al 2019. The Chromatin Loop 1 (CL1) has been found in the two independent HiC datasets whereas the CL2 has only been found in Veluchamy et al 2016.

(B) Relative chromatin loop formation in Col-0 seedlings treated with ABA. Data are shown as the mean ± standard error (n = 3) from 3C-qPCR using primer 1F and 1R as anchor and several primers across the region. HindIII restriction sites are indicated in grey lines.

(C) Chromatin loops detected in (B) over the background level are displayed using the Spring model webservice from (Kadlof et al 2019).

Figure S23. ABA-related TFs binding over the intergenic *MRO-MARS* region

ChIP peaks of different ABA-related TFs (HB6, HB7, GBF2, GBF3, MYB3, MYB44, NF-YC2, NF-YB2, ANAC102, ANAC032, ABF1, ABF3, ABF4, RD26, ZAT6, FBH3, DREB2A, AT5G04760, HAT22 and HSFA6A) found on the marneral cluster (Sato et al 2016) in mock and ABA treated plants.

Figure S24. Transcription activation by the *MRO-MARS* intergenic region

(A) Representative GUS staining, over three independent stained tobacco leaves, after agroinfiltration of the described constructs. Stars (*) indicate agroinfiltration points and letters correspond to agroinfiltrated construct.

(B) Pictures of GUS stained arabidopsis leaves after agroinfiltration of the same constructs described in (A).

Figure S25. Regulation of metabolic clusters in plants by lncRNA

(A) Number and proportion of lncRNAs containing coexpressed gene clusters (Yu et al 2016) compared to control random genomic regions. The number shown above indicates the p-value of the difference between the frequency of lncRNA found within these regions determined by a binomial exact test.

(B) Influence of the presence of lncRNAs on the median level of coding gene expression correlation inside co-expressed gene clusters or in control random genomic regions. The numbers shown above indicate the FDR corrected p-value of the differences for the genomic regions with and without lncRNAs determined by Student t-test.

(C) Maximum level of correlation between a lncRNA and any coding gene of the cluster in co-expressed gene clusters or in control random genomic regions. The numbers shown above indicate the FDR corrected p-values of the difference between genomic regions determined by Student t-test.

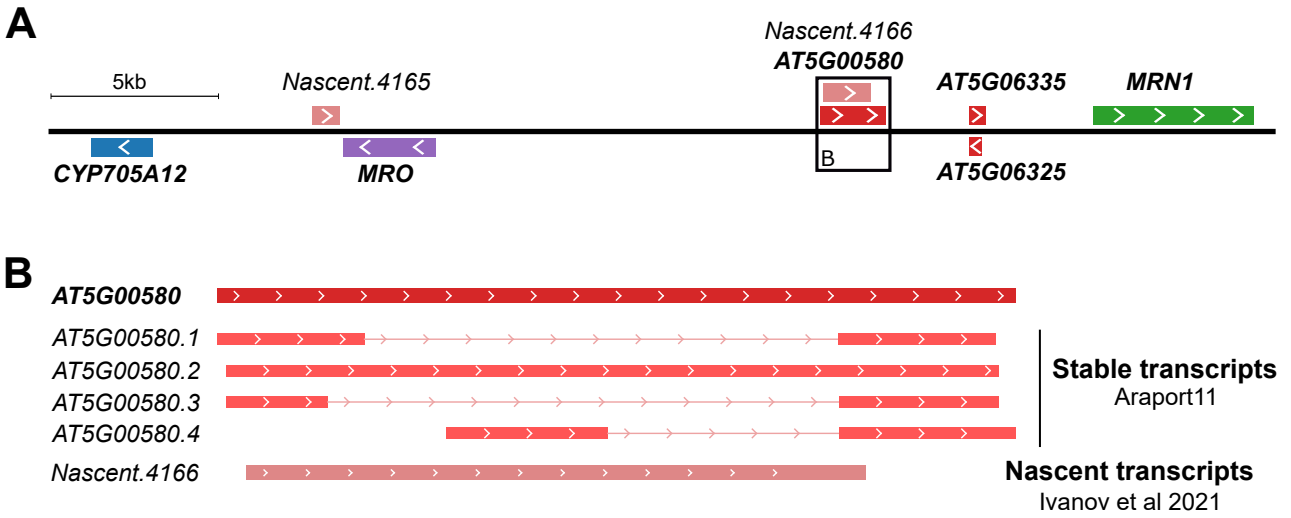
The co-expressed clusters of Yu et al. 2016 are not continuous: some non-coregulated genes might be present inside the genomic regions of the cluster. Therefore, we compute the correlation analysis comprising all the genes included in the cluster boundaries (complete cluster) or with only the one annotated as co-expressed (co-expressed cluster).

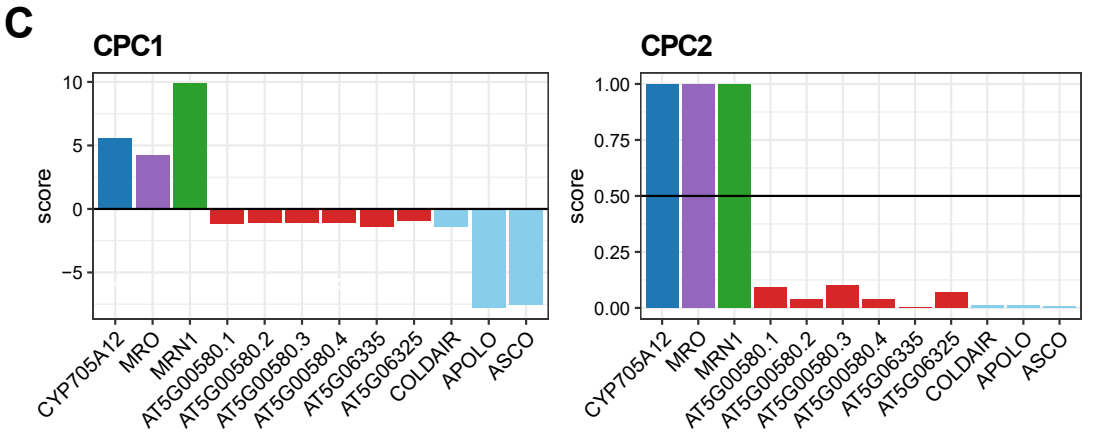
Table S1

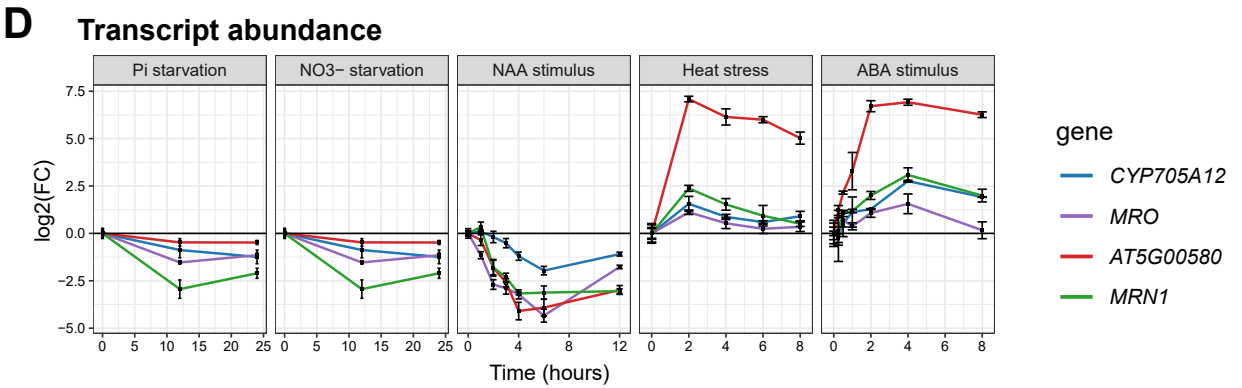
Sequence of primers used in this study.

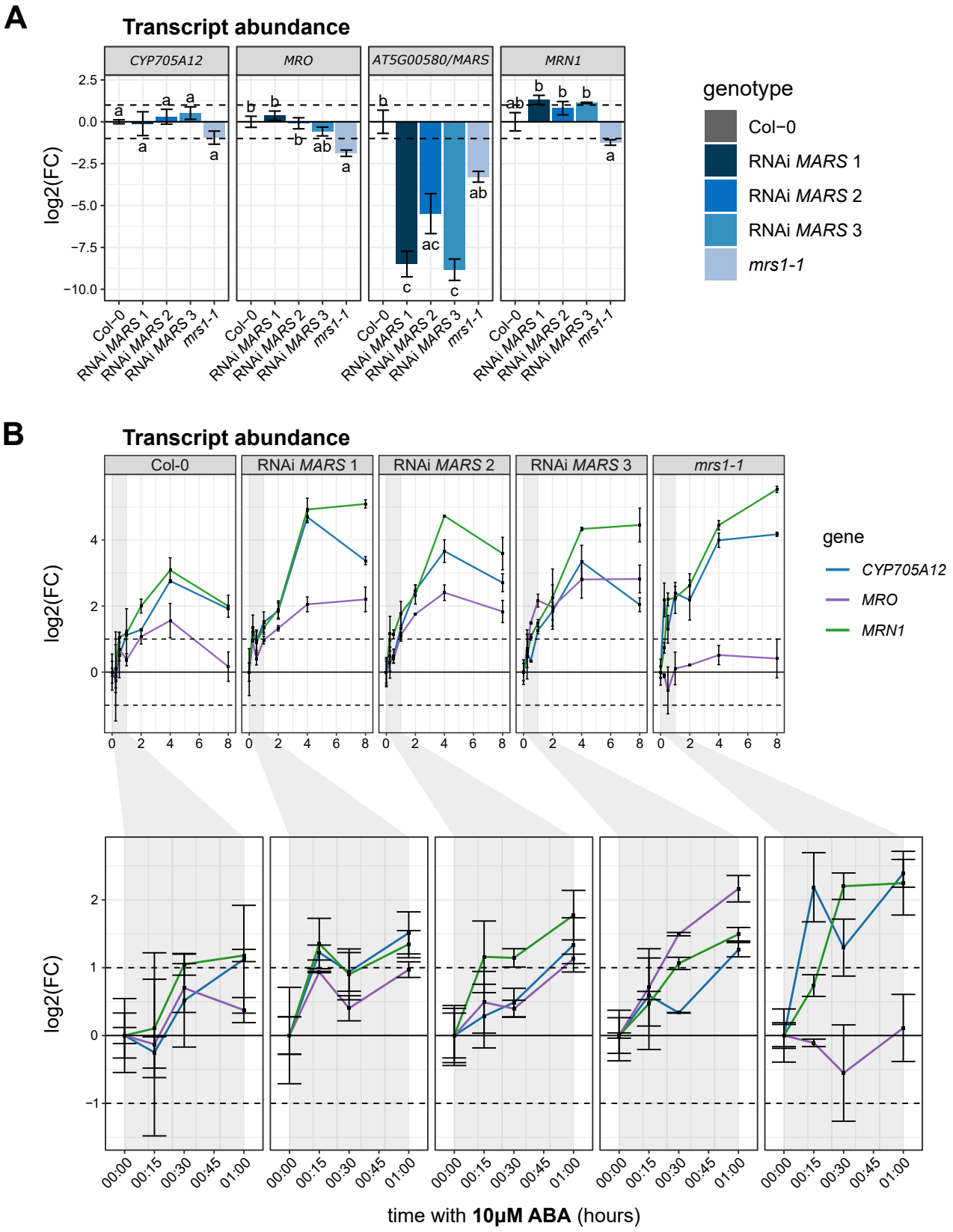
Table S2

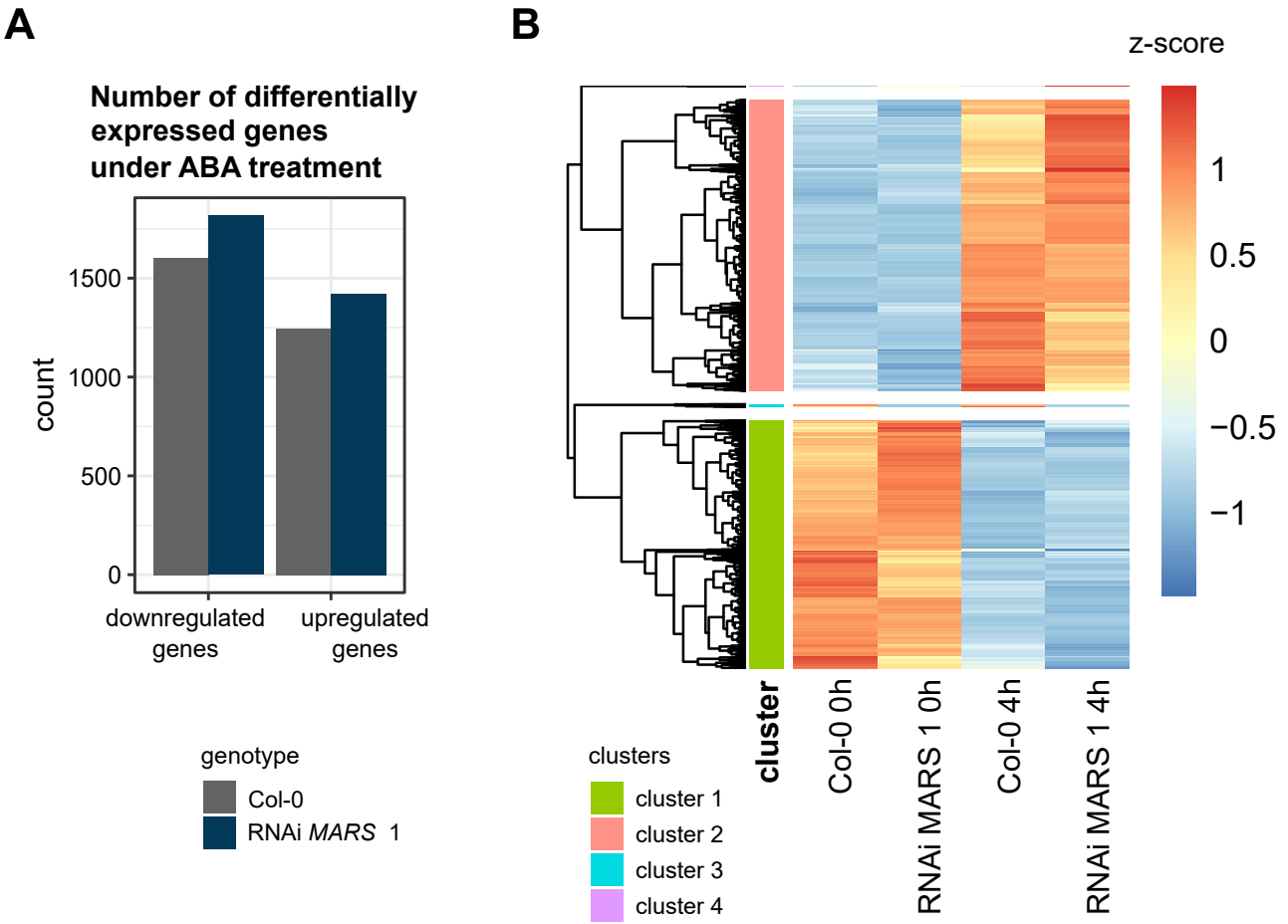
Differential gene expression analysis of RNAi-*MARS* line 1 in response to ABA.

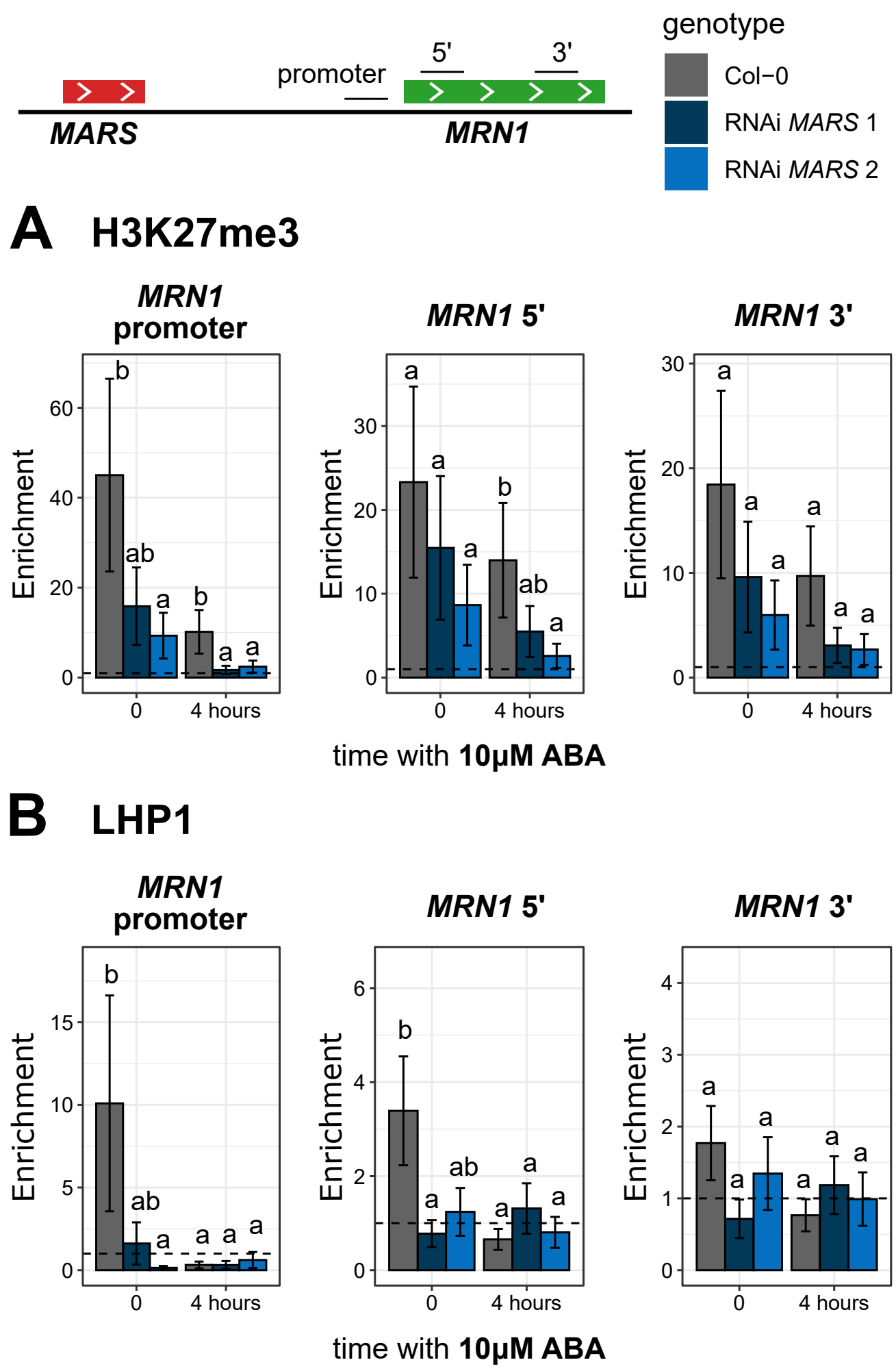

A
5kb
Nascent.4165
Nascent.4166
AT5G00580
AT5G06335
MRN1
CYP705A12
MRO
B
AT5G06325
B
AT5G00580
AT5G00580.1
AT5G00580.2
AT5G00580.3
AT5G00580.4
Nascent.4166
Stable transcripts
Araport11
Nascent transcripts
Ivanov et al 2021

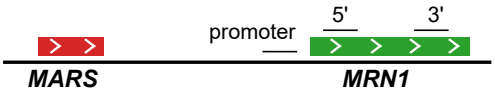

C
CPC1
CPC2
score
10
5
0
−5
1.00
0.75
0.50
0.25
0.00
CYP705A12
MRO
MRN1
AT5G00580.1
AT5G00580.2
AT5G00580.3
AT5G00580.4
AT5G06335
AT5G06325
COLDAIR
APOLO
ASCO


D
Transcript abundance
Pi starvation
NO3− starvation
NAA stimulus
Heat stress
ABA stimulus
log2(FC)
7.5
5.0
2.5
0.0
−2.5
−5.0
Time (hours)
gene
CYP705A12
MRO
AT5G00580
MRN1


A
Transcript abundance
CYP705A12
MRO
AT5G00580/MARS
MRN1
log2(FC)
2.5
0.0
−2.5
−5.0
−7.5
−10.0
Col−0
RNAi MARS 1
RNAi MARS 2
RNAi MARS 3
mrs1-1
genotype
Col−0
RNAi MARS 1
RNAi MARS 2
RNAi MARS 3
mrs1-1
B
Transcript abundance
Col-0
RNAi MARS 1
RNAi MARS 2
RNAi MARS 3
mrs1-1
log2(FC)
gene
CYP705A12
MRO
MRN1
00:00
00:15
00:30
00:45
01:00
time with 10µM ABA (hours)

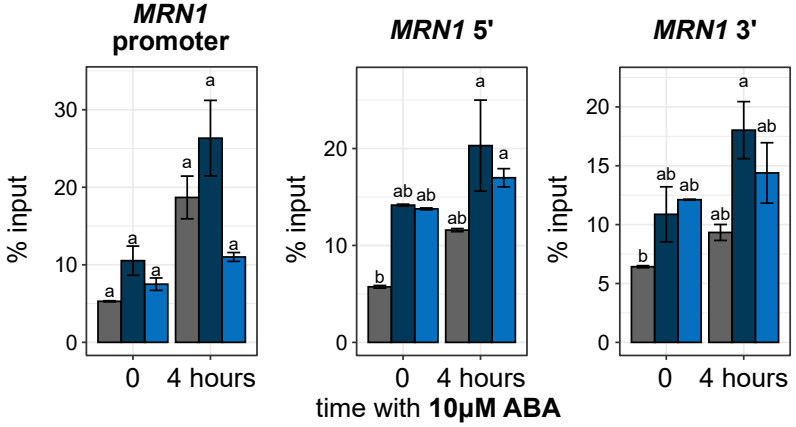

A
Number of differentially expressed genes under ABA treatment
count
1500
1000
500
0
downregulated genes
upregulated genes
genotype
Col-0
RNAi MARS 1
B
z-score
1
0.5
0
−0.5
−1
cluster
Col-0 0h
RNAi MARS 1 0h
Col-0 4h
RNAi MARS 1 4h
clusters
cluster 1
cluster 2
cluster 3
cluster 4


genotype
Col−0
RNAi MARS 1
RNAi MARS 2
promoter
5'
3'
MARS
MRN1
A
H3K27me3
MRN1 promoter
MRN1 5'
MRN1 3'
Enrichment
0
4 hours
time with 10µM ABA
B
LHP1
MRN1 promoter
MRN1 5'
MRN1 3'
Enrichment
0
4 hours
time with 10µM ABA

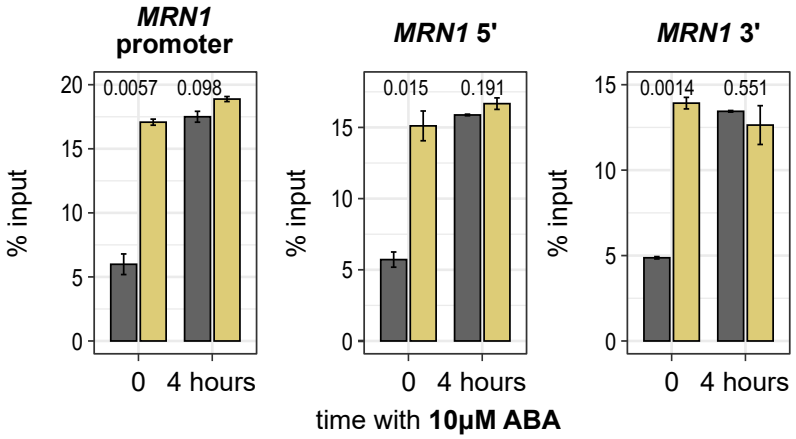


genotype
- Col−0
- RNAi *MARS* 1
- RNAi *MARS* 2
- *lhp1*

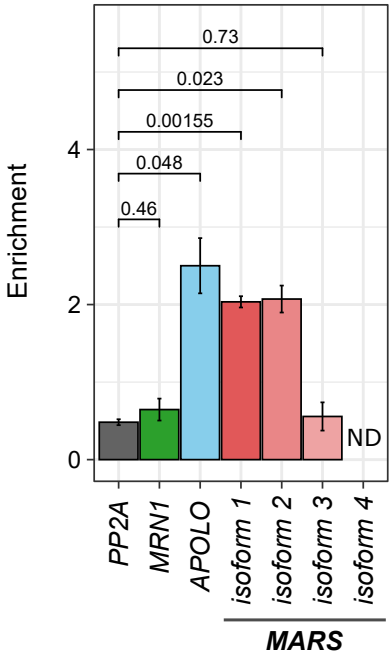
A **FAIRE**



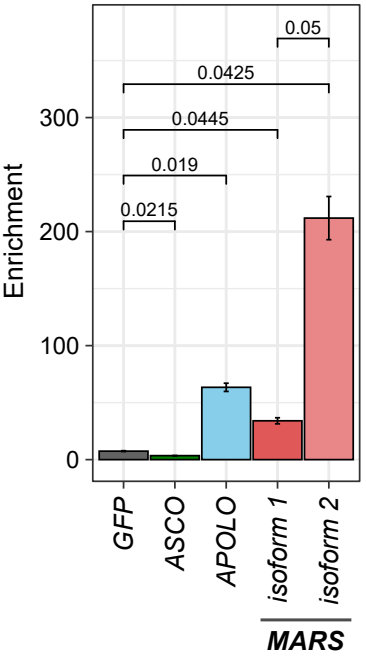
B **FAIRE**

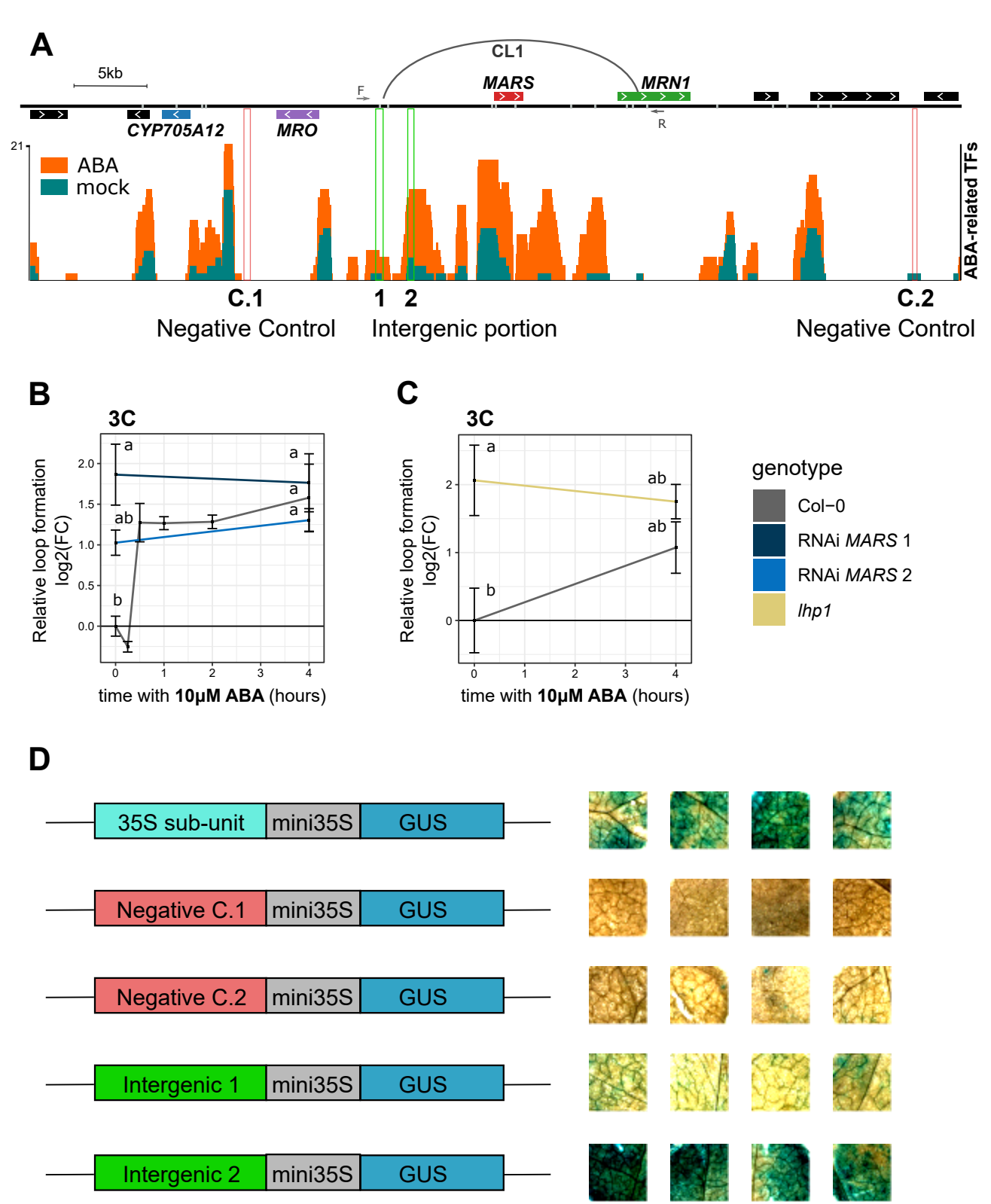


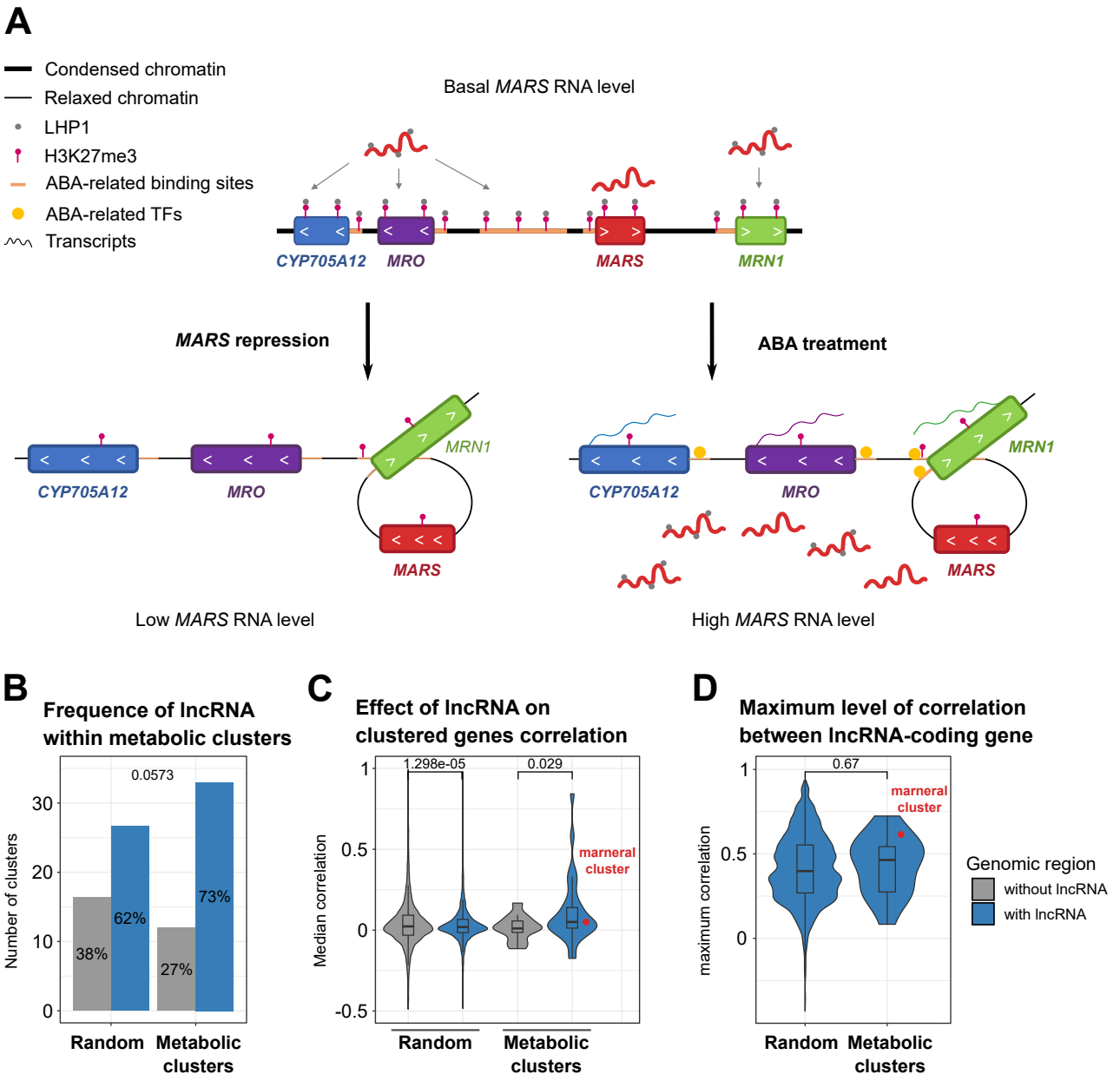
C **LHP1 RIP *in vivo***



D **LHP1 RIP *in vitro***




A
5kb
CL1
MARS
MRN1
F
R
CYP705A12
MRO
21
ABA
mock
ABA-related TFs
C.1
1
2
C.2
Negative Control
Intergenic portion
Negative Control
B
3C
Relative loop formation
log2(FC)
time with 10µM ABA (hours)
C
3C
Relative loop formation
log2(FC)
time with 10µM ABA (hours)
genotype
Col−0
RNAi MARS 1
RNAi MARS 2
lhp1
D
35S sub-unit
mini35S
GUS
Negative C.1
mini35S
GUS
Negative C.2
mini35S
GUS
Intergenic 1
mini35S
GUS
Intergenic 2
mini35S
GUS


A
Condensed chromatin
Relaxed chromatin
LHP1
H3K27me3
ABA-related binding sites
ABA-related TFs
Transcripts
Basal MARS RNA level
CYP705A12
MRO
MARS
MRN1
MARS repression
ABA treatment
Low MARS RNA level
High MARS RNA level
B
Frequence of lncRNA within metabolic clusters
Number of clusters
0.0573
38%
62%
27%
73%
Random
Metabolic clusters
C
Effect of lncRNA on clustered genes correlation
Median correlation
1.298e-05
0.029
marneral cluster
Random
Metabolic clusters
D
Maximum level of correlation between lncRNA-coding gene
maximum correlation
0.67
marneral cluster
Random
Metabolic clusters
Genomic region
without lncRNA
with lncRNA