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

Thomas Roulé[1,2], Federico Ariel[3], Caroline Hartmann[1,2], Martin Crespi [1,2]* and Thomas Blein[1,2]

[1] Institute of Plant Sciences Paris-Saclay, Centre Nationale de la Recherche, Institut National de la Recherche Agronomique, Université Evry, Université Paris-Saclay, 91405 Orsay, France

[2] Institute of Plant Sciences Paris-Saclay, Université de Paris, 91405 Orsay, France

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

*Correspondence to: MC (martin.crespi@universite-paris-saclay.fr)

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, and 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[1].

In plants, synthesis of numerous secondary metabolic compounds is important for the dynamic interaction with their environment, affecting their life and survival[2]. Terpenoids are bioactive molecules of diverse chemical structure[3]. In *Arabidopsis thaliana*, the biosynthesis of four triterpenes, namely thalianol[4], tirucalla-7,24-dien-3b-ol[5], arabidiol[6] and marneral[7], is governed by enzymes encoded by genes organized in clusters[8]. 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[8]. 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 *CYP71A16*, participating in marneral oxidation[7].

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[9]. 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[10]. 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[11].

Long noncoding RNAs (lncRNAs) have emerged as important regulators of eukaryotic gene expression at different levels[12,13]. 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[14,15]. Furthermore, it has been proposed that lncRNAs can modulate the transcriptional activity of neighboring genes by shaping local 3D chromatin conformation[16–18]. 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 includes three noncoding transcriptional units

The small marneral cluster includes three genes: marneral synthase (*MRN1*), *CYP705A12* and *CYP71A16* that *are* two P450 cytochrome-encoding genes (**Figure 1A**), all participating in the biosynthesis and metabolism of the triterpene marneral[7].

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[19]), three additional transcriptional units are located within the marneral cluster, between the *CYP71A16* and the *MRN1* loci. The *AT5G00580* and the pair of antisense genes *AT5G06325* and *AT5G06335* are located upstream of the *MRN1* gene at 6kpb 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 and 1C**). 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**). All these transcripts were classified as lncRNAs when using two coding prediction tools, CPC[20] and CPC2[21], because of their low coding

potential and their length (over 200 nt), similarly to previously characterized lncRNAs (*COLDAIR*[22]; *APOLO*[16]; and *ASCO*[23]) (**Figure 1D**).

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 behavior 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 1E**). Interestingly, the *AT5G00580* lncRNA exhibited the strongest induction in response to heat stress and exogenous ABA, in comparison with *MRN1* and the two *CYP* genes (**Figure 1E**). 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[15,24]. Thus, we wondered if the lncRNA 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 using an RNAi construct targeting the first exon of *AT5G00580*, and isolated two independent lines. The RNAi line 1 impaired the transcriptional accumulation of AT5G00580 without affecting the rest of the cluster. Interestingly, the RNAi line 2 exhibited a strong down-regulation of *AT5G00580* together with a significant basal repression of the two *CYP* genes (**Figure 2A**). Strikingly, the response of the three protein-coding genes of the marneral cluster to exogenous ABA was significantly deregulated in both independent RNAi lines. Transcriptional levels of *MRN1* and the two *CYP* genes increased earlier in RNAi 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 three-fold higher levels in the RNAi lines compared to Col-0 (**Figure 2B top panel**). To further support our observations, we isolated the insertional mutant *SALK_133089* located 200 bp upstream the transcription start site (TSS) of *AT5G00580* gene. Although this insertion did not affect *AT5G00580* basal levels (**Figure 2A**), it partially impaired its induction in response to ABA. In agreement with both RNAi lines, *MRN1* and *CYP705A12* genes were strongly induced by exogenous ABA (**Figure S2),** in contrast to *CYP71A16,* whose promoter region may be locally affected by the T-DNA insertion. Notably, *MRN1*, which encodes the marneral synthase, is the gene in the cluster most strongly affected by *AT5G00580* down-regulation. Collectively, our results indicate that the noncoding transcriptional activity derived from the *AT5G00580* locus represses the dynamic expression of the marneral cluster genes, mainly *MRN1*, in response to ABA. Therefore, we named *the AT5G00580*-derived noncoding transcript *MARneral Silencing* (*MARS*) lncRNA.

MARS affects seed germination likely through its impact on *MRN1* expression

The phytohormone ABA has been implicated in the perception and transduction of environmental signals participating also in a wide range of growth and developmental events such as seed maturation, dormancy and germination[25]. Considering that the marneral cluster

exhibited a strong *MARS*-dependent response to ABA, we wondered what the physiological impact of *MARS* deregulation was during seed germination. To this end, we assessed seed germination in Col-0 and *MARS* down-regulated lines (RNAi lines 1/2 and *SALK_133089*) with or without exogenous ABA. Notably, *MARS* silencing resulted in a delayed germination compared to Col-0 (**Figure S3A**), as revealed by an increase in T50 (time for 50% of germination) of nearly 5 hours (**Figure S3B**). Interestingly, in response to 0.5µM ABA, the germination of RNAi-*MARS* was further delayed than Col-0 with an increase of T50 of nearly 10 hours (**Figure S3C and S3D**). Accordingly, transgenic plants over-expressing *MRN1* also exhibited a delayed germination phenotype regardless of the treatment with ABA (**Figure S3**). The behavior of *35S:MRN1* seedlings suggests that *MRN1* up-regulation in RNAi-*MARS* lines could be linked to the increased sensitivity to ABA (**Figure 2B**). Altogether, our results indicate that *MARS* can modulate seed germination through the regulation of the expression of *MRN1*.

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[10]. According to publicly available ChIP-Seq datasets[26], the marneral cluster region is highly enriched in H3K27me3 deposition in shoots, extensively coinciding with LHP1 recognition (**Figure S4**). Consistently, an ATAC-Seq available dataset[27] revealed that the marneral cluster exhibits a high chromatin condensation in shoots (**Figure S4**). Thus, the marneral cluster is characterized by an epigenetically silent status in aerial organs coinciding with its low expression level[10].

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 3A and S5**). Interestingly, exogenous ABA triggered a strong reduction of H3K27me3 deposition throughout the marneral cluster (**Figure 3A and S5**). Strikingly, 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 3A, S5 and 2B**). Furthermore, we assessed LHP1 recognition of the marneral cluster. In agreement with previous observations (**Figure S4**[26]), LHP1 distribution was high in *MRN1* promoter and more weakly across *MARS* gene body and the intergenic region between *CYP71A16* and *MARS* (**Figure 3B and S6**). Remarkably, LHP1 recognition was strongly impaired in response to ABA as well as in RNAi-*MARS* seedlings (**Figure 3B and S6**). 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[28]. Therefore, we determined by Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) the chromatin condensation of the whole marneral cluster. In contrast to Col-0 showing a highly condensed chromatin, RNAi-*MARS* seedlings exhibit a lower chromatin condensation in control conditions. Notably, the global chromatin status of the cluster was even less condensed in RNAi-*MARS* seedlings in response to ABA (**Figure 4A and S7**), in agreement with a decrease of both H3K27me3 deposition and LHP1 binding (**Figure 3B and S6**) and the concomitant transcriptional activation of the clustered genes (**Figure 2B**). Consistently with our observations, *lhp1* mutant seedlings also showed a global

chromatin decondensation in control conditions, comparable to Col-0 in response to ABA. Notably, further chromatin decondensation triggered by ABA was completely impaired (**Figure 4B and S8**), supporting the role of LHP1 in the dynamic epigenetic silencing of the marneral cluster.

It has been shown that LHP1 can recognize RNAs *in vitro*[14] and the lncRNA *APOLO in vivo*[16]. Moreover, it has been proposed that *APOLO* over-accumulation can decoy LHP1 away from target chromatin[29]. 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 *MARS* was enriched in the nucleus, compared with total RNA, as the previously characterized lncRNAs *APOLO* and *ASCO* that interact respectively with nuclear epigenetic and splicing machineries, and the spliceosome structural ncRNA *U6* (**Figure S9A**). Then, we confirmed by RNA immunoprecipitation (RIP) that LHP1 can interact with *MARS in vivo,* in contrast to the *MRN1* mRNA or housekeeping RNAs taken as negative controls (**Figure 4C**).

LHP1 binding to the marneral cluster was impaired both in response to exogenous ABA (inducing *MARS*, **Figure 1E and 3B**) and in RNAi-*MARS* seedlings, hinting at a stoichiometry-dependent action of *MARS* on LHP1 recognition. Therefore, we used chromatin extracts from RNAi-*MARS* line 1 seedling, with very low *MARS* transcript levels (**Figure 2A**) to assess LHP1 recognition of the marneral cluster upon the addition of increasing concentrations of *in vitro*-transcribed *MARS* RNA. Strikingly, we found that low *MARS* RNA concentrations (between 0.01 and 0.1 µg of RNA; **Figure 4D and S9B**) successfully promoted LHP1 binding to the cluster, in contrast to higher concentrations (between 1 and 10 µg of RNA), supporting the relevance of *MARS*-LHP1 stoichiometry for LHP1-target recognition (**Figure 4D and S9B**). Altogether, our results suggest that the physical interaction of the nuclear-enriched lncRNA *MARS* to LHP1 modulates its binding to proximal chromatin likely participating in the modulation of the dynamic chromatin condensation of the marneral cluster.

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

It was recently observed 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[11]. 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[26,30] there is a significant interaction linking the intergenic region between *CYP71A16* and *MARS* and the *MRN1* locus (indicated as “Chromatin loop” in **Figure 5A**). By using Chromatin Conformation Capture (3C), we determined that the formation of this chromatin loop drastically increased over 30 min with exogenous ABA and remained formed for at least 4 hours (**Figure 5B**). Thus, the formation of this chromatin loop positively correlates with the transcriptional accumulation of the marneral cluster genes in response to ABA (**Figure 2B**).

The *MARS* locus is encompassed in the ABA-dependent chromatin loop (**Figure 5A**). 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 5B**). Interestingly, LHP1 has been implicated in shaping local 3D conformation of target regions[26], 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 5C**). Overall, our

results demonstrate that a chromatin loop within the marneral cluster is regulated by LHP1 and the interacting lncRNA *MARS*, encoded in the region encompassed by the loop.

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 extracted from ChIP-seq the binding distribution of 13 ABA-related transcription factors (TFs)[31]. Interestingly, a high enrichment in ABA TF binding sites was found at the *MARS* locus, as well as in the intergenic region between the *CYP71A16* and *MARS* loci, notably surrounding the contact point brought into close spatial proximity with the *MRN1* locus by the ABA-dependent 3D chromatin loop (**Figure 5A**). We thus assessed the activating capacity of this region, potentially acting as a distant enhancer element of the *MRN1* proximal promoter. To this end, we made use of a *GUS*-based reporter system (as described in [32]), fusing two regions of interest to a minimal *35S* promoter. Two additional DNA regions nearby the putative enhancers were used as negative controls: one between *CYP705A12* and *CYP71A16* and the other at the 3' end of *AT5G42620* locus (**Figure 5A** indicated in red). Among the two putative enhancers regions tested, one was able to activate *GUS* expression (Intergenic region 2, **Figure 5A and 5D**), coinciding with the region showing a high enrichment of ABA-related TF binding sites close to the chromatin loop anchor point (**Figure 5A**). Collectively, our results indicate that an ABA-driven chromatin loop brings into close spatial proximity the *MRN1* locus and a transcriptional activation site likely acting as an ABA enhancer element. Notably, this chromatin reorganization process depends on the LHP1-*MARS* module.

Long noncoding RNAs as emerging regulators of gene clusters

Physically linked genes organized in clusters are generally coregulated[8]. 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 collected two lists of Arabidopsis clustered genes, i.e. one of co-expressed neighboring genes[10] and one of predicted metabolic gene clusters (by PlantiSMASH[33]). According to the latest Arabidopsis genome annotation (Araport11), among the 390 clusters of co-expressed neighboring genes), 189 (48%) contain at least one embedded lncRNA inside the cluster. Most importantly, among the 45 metabolic clusters, 28 (62%) include lncRNAs (**Figure 6A**). Furthermore, among the clusters containing a lncRNA, a correlation analysis based on the maximum strength of co-expression between a lncRNA and any clustered gene revealed that the metabolic clusters exhibit a significantly higher correlation than co-expressed clusters (**Figure 6B**). Altogether, our analyses suggest that lncRNA-mediated local epigenetic remodeling may constitute an emerging feature of non-homologous genes metabolic clusters in plants.

DISCUSSION

The cell nucleus is a dynamic arrangement of DNA, RNAs and proteins[34,35]. 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[36].

In the last few years, noncoding transcription has been implicated in shaping 3D nuclear organization[37]. 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[38,39]. More recently, HiC analyses were performed in mammalian cells exposed or not to RNase, before and after crosslinking, or upon transcriptional inhibition[40]. 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[40].

In plants, several lncRNAs have been implicated in local chromatin conformation dynamics affecting the transcriptional activity of neighboring genes[15,41]. 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*[17]) 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*[16], 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 target genes[29]. 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[42]. 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* affects 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[43]. 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[11]. 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 the *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, chromatin loop conformation is related to LHP1 binding and is modulated by *MARS*. LHP1 recognition at basal *MARS* levels maintains a possibly linear conformation of the region, precluding the enhancer-*MRN1* locus interaction, whereas the positively activating chromatin loop is formed in the absence of LHP1. *MARS* transcriptional accumulation directly modulates LHP1 binding to the marneral cluster and high levels of *MARS* then titrate LHP1 away from the cluster (**Figure 6C**; in response to ABA). Interestingly, when *MARS* levels are too low compared to basal levels (as in the RNAi lines), recruitment of LHP1 to the cluster is also impaired (**Figure 6C;** *MARS* repression).

In mammals, growing evidence supports the role of lncRNAs in chromatin conformation determination[44] and enhancer activity (e.g. *PVT1*[45] and *CCAT1-L*[46]). 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*[47]) and *PANDA*[48], whereas TFs have been implicated in chromatin loop formation in plants[49–51]. 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[52]. 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[10]. Marneral biosynthesis has been linked to root and leaf development, flowering time and embryogenesis[2]. 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. Interestingly, noncoding transcription had already been associated with seed germination. The *DELAY OF GERMINATION 1* (*DOG1*) locus is a major actor regulating seed dormancy strength in Arabidopsis. Indeed, an antisense lncRNA (*asDOG1*) is able to repress *DOG1* transcription in mature plants. Notably, it was observed that *DOG1* suppression is released by shutting down antisense transcription, which is induced by ABA and drought[53].

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[7]. The exploration of the noncoding transcriptome in Arabidopsis recently served to identify ecotype-specific lncRNA-mediated responses to the environment[54]. 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

Lead contact and materials 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@ips2.universite-paris-saclay.fr).

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

Lines selection and generation

All plants used in this study are in Columbia-0 background. RNAi-*MARS* were obtained using the pFRN binary vector[55] bearing 250bp of the first exon of *MARS* gene (see primers in **Table S1**), previously sub-cloned into the pENTR/D-TOPO vector. Arabidopsis plants were transformed using *Agrobacterium tumefaciens* AgI-0[56]. The T-DNA inserted line *SALK_133089* was ordered to NASC (N633089).

Growth conditions

Seeds were sown in plates vertically disposed 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) supplemented with 0.7% sucrose, and without sucrose for the germination assay. For nitrate starvation assay, KNO_3 and $Ca(NO_3)_2$ were replaced from MS/2 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 500uM Pi for Pi containing medium versus 10uM 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 treatment with exogenous ABA or auxin, seedlings were sprayed with 10uM ABA and 10uM 1-Naphthaleneacetic acid (NAA), respectively. For heat stress, plates were transferred to a growth chamber at 37°C under the same lightning conditions. For nitrate and phosphate starvation assays, seedlings have been transferred at day 12 after sowing (DAS) from respectively nitrate and phosphate containing medium to nitrate and phosphate free medium. Finally, for seed germination assay, 0.5uM ABA was supplemented or not to the medium. Germination rate was evaluated twice a day. Seeds were considered germinated when the seed coats were perforated by elongating radicle. For all the experiments, samples were taken from 12 DAS starting two hours after light illumination, at different time-points, after cross-linking or not, depending on the experiment.

RT-qPCR

Total RNA was extracted from whole seedlings using TRI Reagent (Sigam-Aldrich) and treated with DNase (Fermentas) as indicated by the manufacturers. Reverse transcription was realized on 1ug total RNA using the Maxima Reverse Transcriptase (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 delta delta Ct method using *PROTEIN PHOSPHATASE 2A SUBUNIT A3* (*AT1G13320*) for gene normalization[57] and time 0 for time-course experiment.

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[16], starting from two grams of seedlings crosslinked in 1% (v/v) formaldehyde. Chromatin was sonicated in a water bath Bioruptor Plus (Diagenode; 60 cycles of 30s ON and 30s OFF pulses at high intensity). ChIP was performed in an SX-8G IP-Star Compact Automated System (Diagenode). Antibody coated into Protein A Dynabeads (Invitrogen) were incubated 12 hours at 4 °C with the samples. Recovering of immunoprecipitated DNA was realized using Phenol:Chloroform:Isoamilic Acid (25:24:1, Sigma) followed by ethanol precipitation and analyzed by qPCR. For input samples, non-immunoprecipitated sonicated chromatin was processed in parallel.

In-vitro transcribed *MARS* RNA was obtained from a PCR product amplified from wild-type genomic DNA using the T7 promoter on the 5' PCR primer (**TableS1**). PCR products were verified using agarose electrophoresis, and purified using NucleoSpin kit (Macherey-Nagel). 1µg of purified DNA was used for *in-vitro* transcription following the manufacturer instructions (HiScribe T7 High Yield RNA Synthesis Kit, NEB). Purified non-crosslinked chromatin obtained from five grams of *MARS* RNAi line 1 seedlings were resuspending in 1 mL of nuclei lysis buffer and split into five tubes. An increasing quantity of *MARS* RNA was added to each tube from 0 to 10 µg RNA and incubated under soft rotation during 3 h at 4 °C. Chromatin samples were then cross-linked using 1% (v/v) of formaldehyde for five minutes. Sonication and the following ChIP steps were performed as above.

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)

FAIRE was performed as described by [58]. After chromatin purification as for ChIP, only 50 µl from the 500 µl of purified chromatin were used (diluted to 500 µl with 10 mM Tris-HCl pH 8). For the qPCR, the same set of primers as for ChIP were used.

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 (Sigam-Aldrich). For the total fraction, 200 µL of cell suspension in cell resuspension solution, were collected and completed with 800 µL of TRI Reagent to follow with the RNA extraction. RNA's samples were treated by DNase, and RT was performed using random hexamers prior to qPCR analysis.

RNA immunoprecipitation (RIP)

For RIP, the *lhp1* mutant complemented with the *ProLHP1:LHP1:GFP*[59] were used after 4 hours of treatment with ABA. After crosslinking and chromatin extraction as for ChIP, ten percent of resuspended chromatin was conserved at -20 °C as the input. Chromatin was sonicated in a water bath Bioruptor Plus (Diagenode; 5 cycles of 30 s ON and 30 s OFF pulses at high intensity). Anti-LHP1 RIP was performed using the anti-GFP antibody (Abcam ab290), as previously described[16]. Results were expressed as the percentage of cDNA detected after IP taking the input value as 100%.

Chromosome conformation capture (3C)

3C was performed as described by [60]. Briefly, chromatin was extracted from two grams of cross-linked seedlings as for ChIP. Overnight digestion at 37 °C were performed using 400U of Hind III enzyme (NEB). Digested DNA was ligated during 5 h incubation 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 calculated by qPCR using a DNA region uncut by Hind III to normalize the amount of DNA.

Transcriptional activation assay in tobacco leaves

The *GUS* reporter system for validating the activity of the putative enhancer element was adapted from [32]. Different DNA fragments were cloned in the GreenGate system[61] fused to a minimal 35S promoter element from CAMV (synthesized by Eurofins Genomics). The sub-unit B3 from 35S promoter element from CAMV[62] was synthesized and used as a positive control. All primers used for cloning are indicated in **Table S1**.

A. tumefaciens-mediated transient transformation was performed on 5-week-old tobacco plants using a needle-less syringe. Together with enhancer constructs, another vector containing mCherry driven by 35S promoter was co-transfected to control the transformation efficiency. Two leaf discs were collected near the infiltration site. One, to determine the transfection efficiency by mCherry fluorescence observation under epifluorescent microscope. The second was used for GUS staining, as previously described[63]. Samples were incubated 4 h in the dark at 37 °C before observation.

Identification of lncRNA loci in Arabidopsis gene clusters

The genes of co-expressed clusters were retrieved from [10]. The boundaries of the gene clusters were extracted using Araport11 annotations. The boundaries of the metabolic clusters were extracted from the plantiSMASH predicted clusters on Arabidopsis[33]. Using Araport11 GFF, the 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.

Gene expression correlation analysis

To compute the correlation of expression in different organ of Arabidopsis we used the 113 RNA-seq datasets that were used for the Araport11 annotations (10.1111/tpj.13415). These datasets were generated from untreated or mock-treated wild-type Col-0 plants. After removing the adaptors with Trim Galore with default parameters, the reads were mapped on TAIR10 with STAR v2.7.2a (10.1093/bioinformatics/bts635) and the parameters '--alignIntronMin 20 --alignIntronMax 3000'. Gene expression was then quantified with featureCounts v2.0.0 (10.1093/bioinformatics/btt656) 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[64].

For the correlation of expression inside the marneral cluster, the expression of the genes of the clusters and 25kb around it (four genes upstream and two downstream) were collected for the correlation analysis. 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[65].

Inside each co-expressed and metabolic clusters of genes, Pearson's correlation was computed between every possible pairs of lncRNA and coding gene as for the genes inside the marneral cluster. The maximum correlation value was kept as an indication of lncRNAs correlation with the genes of the cluster.

Quantification and statistical analyses

For all the experiments, at least two 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 independent tobacco plants. The tests used for statistical analyses are indicated in the respective figure legends. Statistical test and associated plots have been generated using R software (v3.6.3[66]) with the help of the tidyverse package[67].

LEGENDS TO THE FIGURES

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 plain rectangles and white arrows indicate the sense of transcription. The square indicates the region displayed in (B).

(B) Schematic illustration of the different isoforms of *AT5G00580* transcripts. First line corresponds to *AT5G00580* genomic region whereas the other lines present the various isoforms. For each isoform, exons are indicated with rectangles and introns with solid lines. Black arrows indicate the primers used for the cDNA amplification in (C).

(C) cDNA amplification of *AT5G00580* isoforms. The position of the primers used for the amplification are displayed on B. MWM stands for Molecular Weight Marker (GeneRuler 1 kb Plus DNA Ladder, Thermo Scientific).

(D) Coding potential of the transcripts located in the marneral cluster genomic region. Scores were calculated using CPC1 (left) and CPC2 (right) programs[20,21]. For each algorithm, the threshold between coding and non-coding genes is displayed with a horizontal solid black line. Coding genes are situated above the threshold, whereas non coding genes are situated below. *COLDAIR*, *APOLO* and *ASCO* are used as positive controls of non-coding transcripts.

(E) Dynamic transcriptional levels of co-regulated genes of the marneral cluster in response to phosphate and nitrate starvation, heat stress, and exogenous ABA and auxin. Gene expression data are represented 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 levels of the marneral cluster genes in control conditions in RNAi lines targeting *AT5G00580*/*MARS* and *SALK_133089* line. Transcriptional levels are represented as the mean ± standard error (n = 3) of the log2 fold change compared to Col-0. Letters indicate statistic group determined by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test. For each gene, each letter indicates statistical difference 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*. Gene expression data are represented as the mean ± standard error (n = 3) of the log2 fold change compared to time 0h.

Figure 3. *MARS* modulates the epigenetic landscape of *MRN1* locus

(A) H3K27me3 deposition over the *MRN1* gene in Col-0 and RNAi-*MARS* seedlings under control conditions and in response to ABA. Higher values of ChIP-qPCR indicate more H3K27me3.

(B) LHP1 binding to the *MRN1* gene in Col-0 and RNAi-*MARS* seedlings in the same conditions as in (A). Higher values of ChIP-qPCR indicate more LHP1 deposition.In (A) and (B), values under the dotted line are considered as not enriched. Results are represented as the mean ± standard error (n = 2) of the H3K27me3/IgG or LHP1/IgG ratio. Numbers are p-value of the difference between the two genotypes determined by Student t-test.

Figure 4. *MARS* influences chromatin condensation of *MRN1* gene through its interaction with LHP1 protein

(A) Chromatin condensation in *MRN1* gene of Col-0 and RNAi-*MARS* seedlings in control conditions and in response to ABA, determined by Formaldehyde Assisted Isolation of Regulatory Element (FAIRE)-qPCR.

(B) Evolution of the chromatin condensation in *MRN1* gene of Col-0 and *lhp1* mutant subjected to ABA treatment determined by Formaldehyde Assisted Isolation of Regulatory Element (FAIRE) qPCR.

In (A) and (B), 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). Lower value indicates more condensed chromatin. Numbers are p-value of the difference between the two genotypes determined by Student t-test.

(C) LHP1-*MARS* interaction was assessed by RNA immunoprecipitation (RIP) using *LHP1-GFP* seedlings. Negative controls include a housekeeping gene (*PP2A*) and *MRN1* mRNA. The interaction between *APOLO* and LHP1 was taken as a positive control[16]. Results are expressed as the mean ± standard error (n = 4) of the percentage of input (signal measured before immunoprecipitation).

(D) LHP1 binding to the *MRN1* promoter region in chromatin from RNAi-*MARS* seedlings upon increasing amounts of *in-vitro* transcribed *MARS* RNA. After incubation (see Methods), the samples were crosslinked for LHP1 ChIP-qPCR. Higher values indicate more LHP1-DNA interaction. Results are expressed as the mean ± standard error (n = 2) of the LHP1/Igg ratio.

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

Figure 5. An LHP1-dependent chromatin loop brings together the *MRN1* locus and a putative enhancer element in response to ABA

(A) Schematic illustration of the loop linking the *MRN1* locus with the intergenic region between *CYP71A16* and *MARS*. Forward (F) and Reverse (R) oligonucleotides used for 3C-qPCR (in B–C) are indicated with arrows. The orange track shows the number of different ABA-related transcription factor binding sites (HB6, HB7, GBF2, GBF3, MYB3, MYB44, NF-YC2, NF-YB2, ANAC102, ANAC032, ABF1, ABF3, ABF4, RD26, ZAT6, FBH3, DREBA2A, AT5G04760, HAT22 and HSFA6A) found along the marneral cluster[31]. Green and red rectangles indicate the putative enhancer region and the negative controls, respectively, tested for the GUS-based reporter system in (D).

(B) Relative chromatin loop formation in response to ABA in Col-0 and RNAi-*MARS* seedlings. Results are expressed as the mean ± standard error (n = 2) from 3C-qPCR using primer F and R shown on (A), compared to time 0h.

(C) Relative chromatin loop formation in response to ABA treatment in Col-0 and *lhp1* mutant. Data are represented as the mean ± standard error (n = 3) from 3C-qPCR using primer F and R shown on (A), compared to time 0h.

(D) Constructs used for 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 35S sub-unit controls *GUS* expression. The second and third lines show two independent negative controls in which *GUS* gene is driven by a genomic region that does not contain ABA-related binding sites as indicated in (A). In the remaining lines, the transcriptional activity is assessed for the two intergenic regions indicated in (A).

Figure 6. Regulation of metabolic clusters in plants by lncRNA

(A) The proportion of metabolic clusters including lncRNA loci is higher than for other clusters. The co-expressed clusters were predicted in [10] and correspond to co-expressed neighboring genes. The metabolic clusters are co-expressed neighboring genes involved in the biosynthesis of a particular secondary metabolite predicted by plantiSMASH[33].

(B) Maximum level of correlation between a lncRNA and any clustered gene calculated in each cluster. The number shown above indicates the p-value of the difference between the two type of clusters determined by Student t-test.

(C) 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 and possibly linear chromatin conformation. In response to ABA (bottom left panel) *MARS* over-accumulated transcripts titrate LHP1 away from the cluster. The decrease of LHP1 deposition diminishes H3K27me3 distribution, relaxes the chromatin and as a consequence allows the formation of a chromatin loop that brings together the enhancer element and *MRN1* proximal promoter, leading to transcriptional activation. When *MARS* is repressed, LHP1 recruitment to the cluster is impaired, thus leading to a similar chromatin state: decrease in H3K27me3 mark, chromatin decondensation and increase in chromatin loop conformation. Under this chromatin state, the clustered genes become highly responsive to the ABA treatment.

LEGEND TO SUPPLEMENTAL FIGURES

Figure S1. *AT5G00580* 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. *MARS* deregulation enhances the response of *MRN1* and *CYP705A12* to exogenous ABA

Transcript levels of the marneral cluster genes in response to ABA in the insertional mutant line *SALK_133089*. Transcript levels are expressed as the mean ± standard error (n = 3) of the log2 fold change, compared to time 0h.

Figure S3. *MARS* modulates seed germination through the regulation of *MRN1* expression

(A) Percentage of germinated seeds in a control medium. Results are expressed as the mean ± standard error (n = 7) from a batch of ≈50 seeds.

(B) Time for 50% germination (T50) in control medium. Results are expressed as the mean ± standard error (n = 7) from a batch of ≈50 seeds. Letters indicate statistic 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 ($p \leq 0.05$).

(C) Percentage of germinated seeds in a medium containing 0.5uM ABA. Results are expressed as the mean ± standard error (n = 7) from batch of ≈50 seeds.

(D) Times needed for 50% germination (T50) in a medium containing 0.5uM ABA. Results are expressed as the mean ± standard error (n = 7) from a batch of ≈50 seeds. Letters indicate statistic 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 ($p \leq 0.05$).

Figure S4. The epigenetic landscape of the marneral cluster and surrounding genomic region

First track represents DNA accessibility determined by the ATAC-seq[27]. ATAC-peaks are indicated with a blue rectangle and correspond to relaxed chromatin. Second and third tracks

show H3K27me3 and LHP1 ChIP-Seq, respectively[26]. The three experiments shown here have been performed using Arabidopsis shoot. Gene annotation is shown at the top.

Figure S5. *MARS* influences H3K27me3 deposition in the marneral cluster region

H3K27me3 deposition in the intergenic region, *CYPs* and *MARS* loci was measured in Col-0 and RNAi-*MARS* seedlings was assessed by ChIP-qPCR in control conditions and in response to ABA. Higher values indicate H3K27me3 enrichment. Values under the dotted line are considered as not enriched. Data are represented as the mean ± standard error (n = 2) of the H3K27me3/Igg ratio. Numbers are p-value of the difference between the two genotypes determined by Student t-test.

Figure S6. *MARS* modulates LHP1 binding across the marneral cluster region

LHP1 binding to the intergenic region, *CYPs* and *MARS* loci was measured in Col-0 and RNAi-*MARS* seedlings was assessed by ChIP-qPCR in control conditions and in response to ABA. Higher values indicate LHP1 enrichment. Values under the dotted line are considered as not enriched. Data are represented as the mean ± standard error (n = 2) of the LHP1/Igg ratio. Numbers are p-value of the difference between the two genotypes determined by Student t-test.

Figure S7. *MARS* modulates chromatin condensation of the marneral cluster genomic region

Chromatin condensation in the intergenic region, *CYPs* and *MARS* loci was measured in Col-0 and RNAi-*MARS* seedlings was assessed by Formaldehyde Assisted Isolation of Regulatory Element (FAIRE)-qPCR in control conditions and in response to ABA. Lower value indicates more condensed chromatin. Results are expressed as the mean ± standard error (n = 2) of the percentage of input (signal measured before isolation of decondensed region of chromatin; free of nucleosomes). Numbers are p-value of the difference between the two genotypes determined by Student t-test.

Figure S8. LHP1 is involved in chromatin condensation modulation of the marneral cluster region

Chromatin condensation in the intergenic region, *CYPs* and *MARS* loci was measured in Col-0 and *lhp1* mutant seedlings was assessed by Formaldehyde Assisted Isolation of Regulatory Element (FAIRE)-qPCR in control conditions and in response to ABA. 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). Numbers are p-value of the difference between the two genotypes determined by Student t-test.

Figure S9. Nuclear-enriched *MARS* RNA modulates LHP1 binding to the marneral cluster and influences the subsequent ABA response of the marneral cluster genes

(A) Nuclear enrichment of the lncRNA *MARS* compared to other nuclear-enriched lncRNAs determined as the ratio of transcript abundance in the nuclear fraction compared to total cellular RNA. Higher value indicates nuclear enrichment. *APOLO*, *ASCO* and *U6* RNA have been used as positive controls whereas *RHIP1* (*AT4G2641;* 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 p-value of the difference between the corresponding RNA determined by Student t-test.

(B) LHP1 binding in RNAi-*MARS*-derived chromatin to different sites across the marneral cluster upon increasing amounts of *in-vitro* transcribed *MARS* RNA (ug), determined by ChIP-qPCR. Higher values indicate LHP1 enrichment. Results are expressed as the mean ± standard error (n = 2) of the LHP1/Igg ratio.

(C) Transcript levels of the marneral cluster genes in response to ABA treatment in *lhp1* mutant. Results are represented as the mean ± standard error (n = 3) of the log2 fold change compared to time-point 0h.

ACKNOWLEDGMENTS

IPS2 benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). We thank Jeremie Bazin and Aurélie Christ from IPS2 for the helpful discussion about results interpretation and design of the experiments. We thank Olivier Martin for critical reading of the manuscript.

AUTHORS' CONTRUBUTIONS

TR, FA, TB and MC conceived and designed the experiments. TR performed the experiments. TR and TB analyzed the data. All authors discussed the results and edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Nützmann, H.-W., Scazzocchio, C. & Osbourn, A. Metabolic Gene Clusters in Eukaryotes. *Annu. Rev. Genet.* **52**, 159–183 (2018).

2. Go, Y. S. *et al.* Identification of marneral synthase, which is critical for growth and development in Arabidopsis. *Plant J.* **72**, 791–804 (2012).

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

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

5. Boutanaev, A. M. *et al.* Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E81–E88 (2015).

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

7. Field, B. *et al.* Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci.* **108**, 16116–16121 (2011).

8. Nützmann, H. W., Huang, A. & Osbourn, A. Plant metabolic clusters – from genetics to

genomics. *New Phytol.* **211**, 771–789 (2016).

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

10. Yu, N. *et al.* Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res.* **44**, 2255–2265 (2016).

11. Nützmann, H. *et al.* Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. 1–10 (2020) doi:10.1073/pnas.1920474117.

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

13. Ariel, F., Romero-Barrios, N., Jégu, T., Benhamed, M. & Crespi, M. Battles and hijacks: Noncoding transcription in plants. *Trends Plant Sci.* **20**, 362–371 (2015).

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

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

16. Ariel, F. *et al.* Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* **55**, 383–396 (2014).

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

18. Gagliardi, D. *et al.* 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 (2019).

19. Cheng, C. Y. *et al.* Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **89**, 789–804 (2017).

20. Kong, L. *et al.* CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, 345–349 (2007).

21. Kang, Y. J. *et al.* CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).

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

23. Bardou, F. *et al.* Long Noncoding RNA Modulates Alternative Splicing Regulators in Arabidopsis. *Dev. Cell* **30**, 166–176 (2014).

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

25. Vishwakarma, K. *et al.* Abscisic acid signaling and abiotic stress tolerance in plants: A review on current knowledge and future prospects. *Front. Plant Sci.* **8**, 1–12 (2017).

26. Veluchamy, A. *et al.* LHP1 Regulates H3K27me3 Spreading and Shapes the Three-Dimensional Conformation of the Arabidopsis Genome. *PLoS One* **11**, 1–25 (2016).

27. Sijacic, P., Bajic, M., McKinney, E. C., Meagher, R. B. & Deal, R. B. Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific

transcription factor networks. *Plant J.* **94**, 215–231 (2018).

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

29. Ariel, F. *et al.* R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. *Mol. Cell* **77**, 1–11 (2020).

30. Liu, C. *et al.* Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. *Genome Res.* **26**, 1057–1068 (2016).

31. Song, L. *et al.* A transcription factor hierarchy defines an environmental stress response network. *Science (80-.).* **354**, 598 (2016).

32. Yan, W. *et al.* Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. *Nat. Commun.* **10**, 1–16 (2019).

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

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

35. Gibcus, J. H. & Dekker, J. The Hierarchy of the 3D Genome. *Mol. Cell* **49**, 773–782 (2013).

36. Rodriguez-Granados, N. Y. *et al.* Put your 3D glasses on: Plant chromatin is on show. *J. Exp. Bot.* **67**, 3205–3221 (2016).

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

38. Caudron-Herger, M. *et al.* Coding RNAs with a non-coding function: Maintenance of open chromatin structure. *Nucleus* **2**, (2011).

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

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

41. Gagliardi, D. & Manavella, P. A. Short-range regulatory chromatin loops in plants. *New Phytol.* 1–6 (2020) doi:10.1111/nph.16632.

42. Wang, Y. *et al.* Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat. Commun.* **9**, 1–9 (2018).

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

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

45. Cho, S. W. *et al.* Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173**, 1398-1412.e22 (2018).

46. Xiang, J. F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **24**, 513–531 (2014).

47. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the

human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).

48. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–629 (2011).

49. Liu, C. *et al.* A Conserved Genetic Pathway Determines Inflorescence Architecture in Arabidopsis and Rice. *Dev. Cell* **24**, 612–622 (2013).

50. Mendes, M. A. *et al.* MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in Arabidopsis. *Plant Cell* **25**, 2560–2572 (2013).

51. Cao, S. *et al.* A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in Arabidopsis. *Plant Cell* **26**, 1009–1017 (2014).

52. Puvvula, P. K. *et al.* Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nat. Commun.* **5**, (2014).

53. Yatusevich, R. *et al.* Antisense transcription represses Arabidopsis seed dormancy QTL DOG 1 to regulate drought tolerance . *EMBO Rep.* **18**, 2186–2196 (2017).

54. Blein, T. *et al.* Landscape of the non-coding transcriptome response of two Arabidopsis ecotypes to phosphate starvation. *Plant Physiol.* **183**, pp.00446.2020 (2020).

55. Ariel, F. *et al.* Two direct targets of cytokinin signaling regulate symbiotic nodulation in medicago truncatula. *Plant Cell* **24**, 3838–3852 (2012).

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

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

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

59. Nakahigashi, K., Jasencakova, Z., Schubert, I. & Goto, K. 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 (2005).

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

61. Lampropoulos, A. *et al.* GreenGate - A novel, versatile, and efficient cloning system for plant transgenesis. *PLoS One* **8**, (2013).

62. Moreno-risueno, M. A. *et al.* NIH Public Access. **329**, 1306–1311 (2010).

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

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

65. Taiyun Wei and Viliam Simko (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available online at https://github.com/taiyun/corrplot

66. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at https://www.R-project.org/.

67. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

Figure1

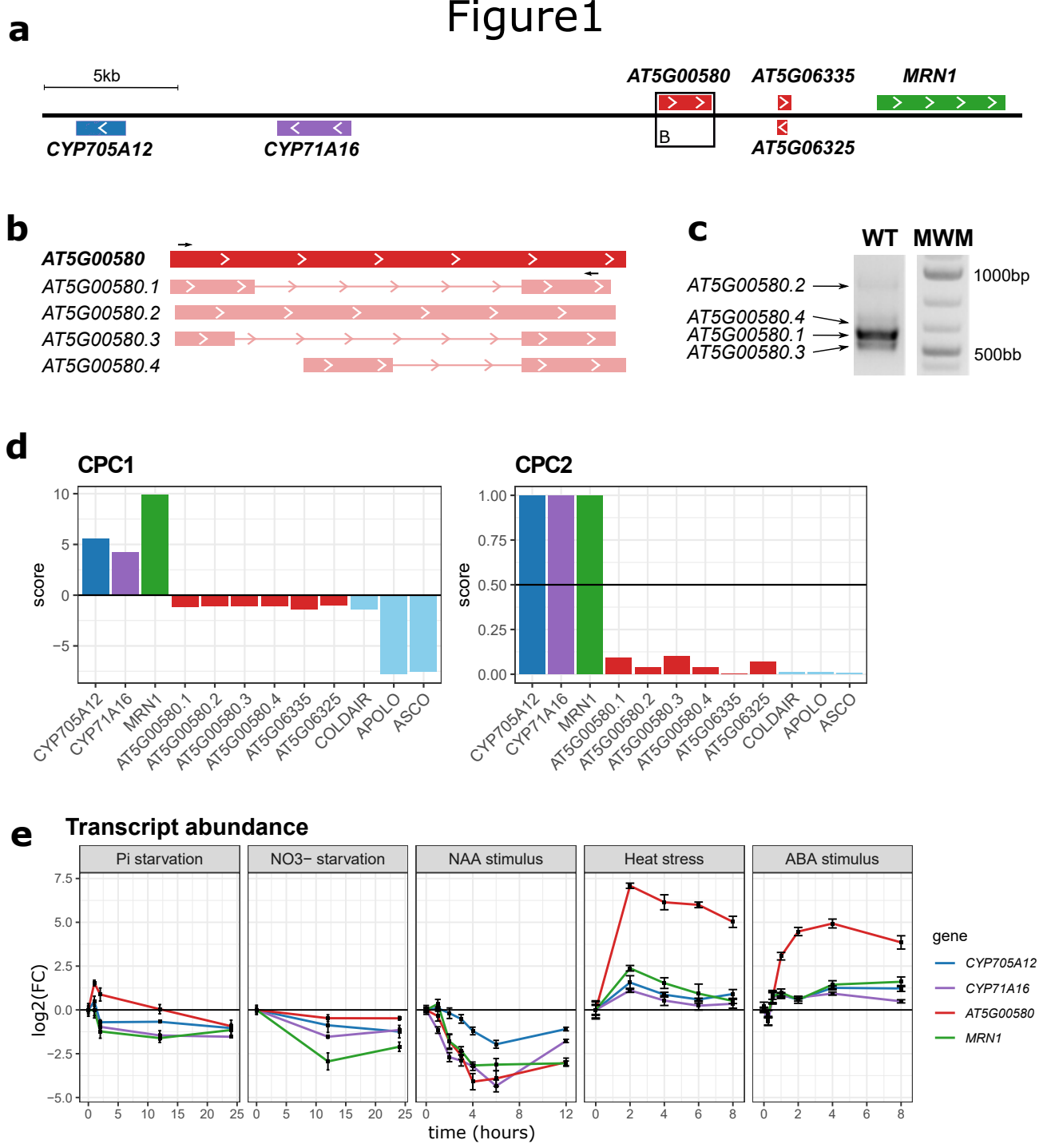
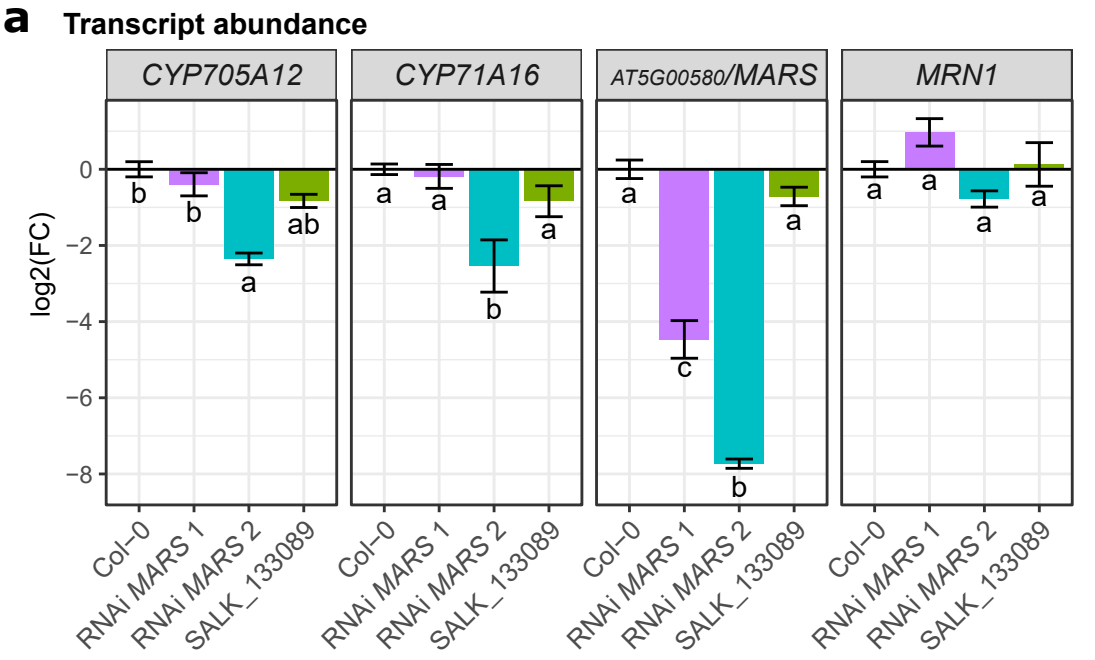

a
5kb
AT5G00580
AT5G06335
MRN1
CYP705A12
CYP71A16
B
AT5G06325
b
AT5G00580
AT5G00580.1
AT5G00580.2
AT5G00580.3
AT5G00580.4
c
WT
MWM
AT5G00580.2
AT5G00580.4
AT5G00580.1
AT5G00580.3
1000bp
500bb
d
CPC1
score
CYP705A12
CYP71A16
MRN1
AT5G00580.1
AT5G00580.2
AT5G00580.3
AT5G00580.4
AT5G06335
AT5G06325
COLDAIR
APOLO
ASCO
CPC2
score
e
Transcript abundance
Pi starvation
NO3− starvation
NAA stimulus
Heat stress
ABA stimulus
log2(FC)
time (hours)
gene
CYP705A12
CYP71A16
AT5G00580
MRN1

Figure2


a
Transcript abundance
CYP705A12
CYP71A16
AT5G00580/MARS
MRN1
log2(FC)
0
−2
−4
−6
−8
b
b
a
ab
a
a
b
a
a
c
b
a
a
a
a
Col−0
RNAi MARS 1
RNAi MARS 2
SALK_133089

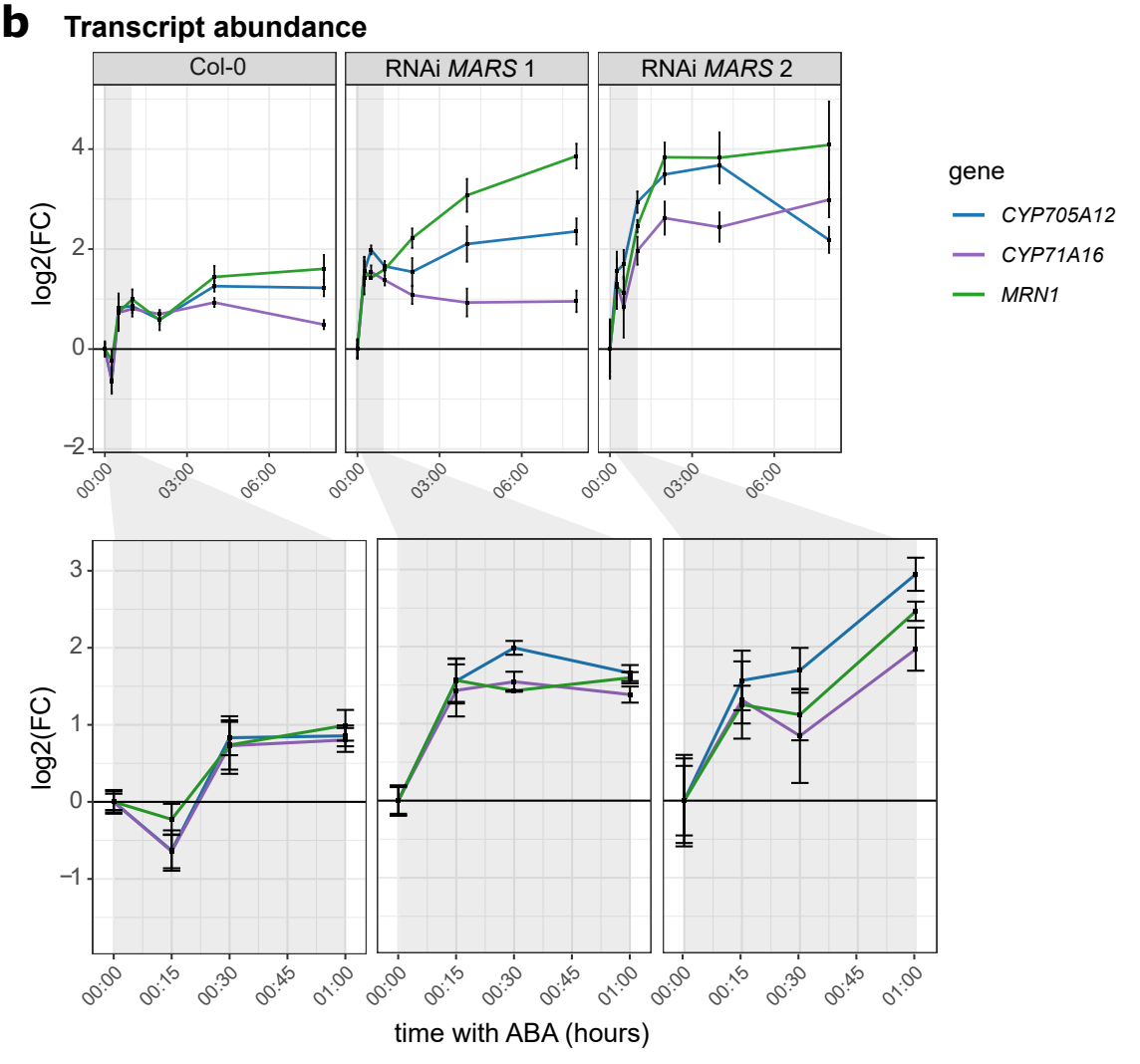
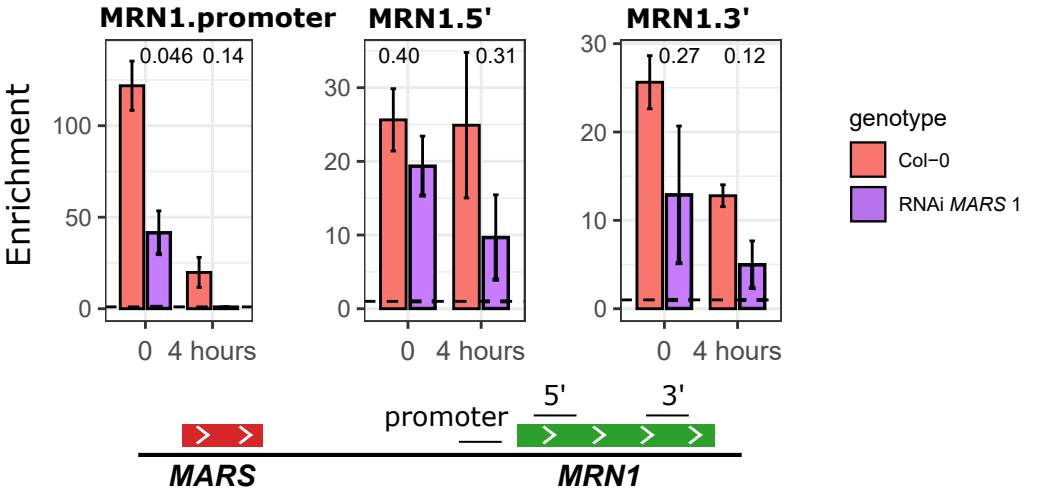

b
Transcript abundance
Col-0
RNAi MARS 1
RNAi MARS 2
log2(FC)
4
2
0
−2
00:00
03:00
06:00
3
2
1
0
−1
00:15
00:30
00:45
01:00
time with ABA (hours)
gene
CYP705A12
CYP71A16
MRN1

Figure3

a **H3K27me3**



b **LHP1**

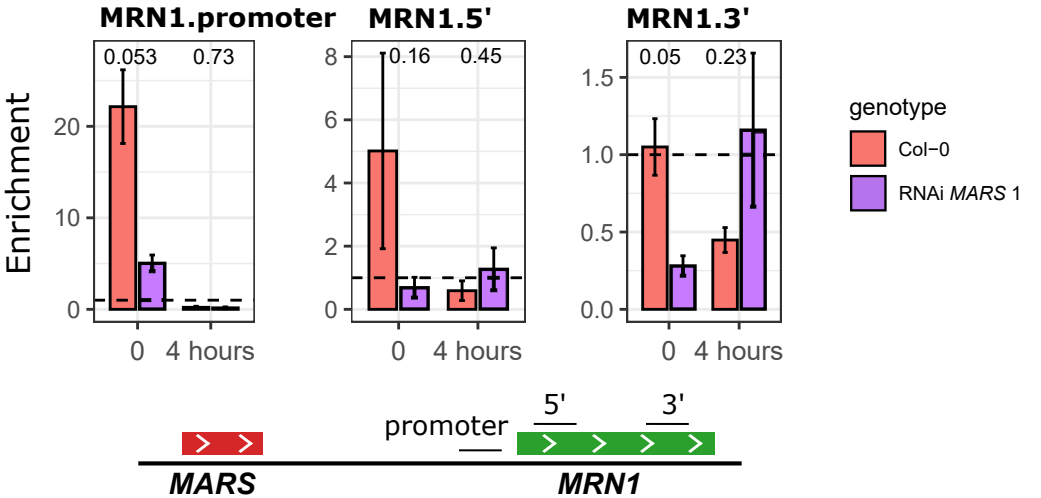
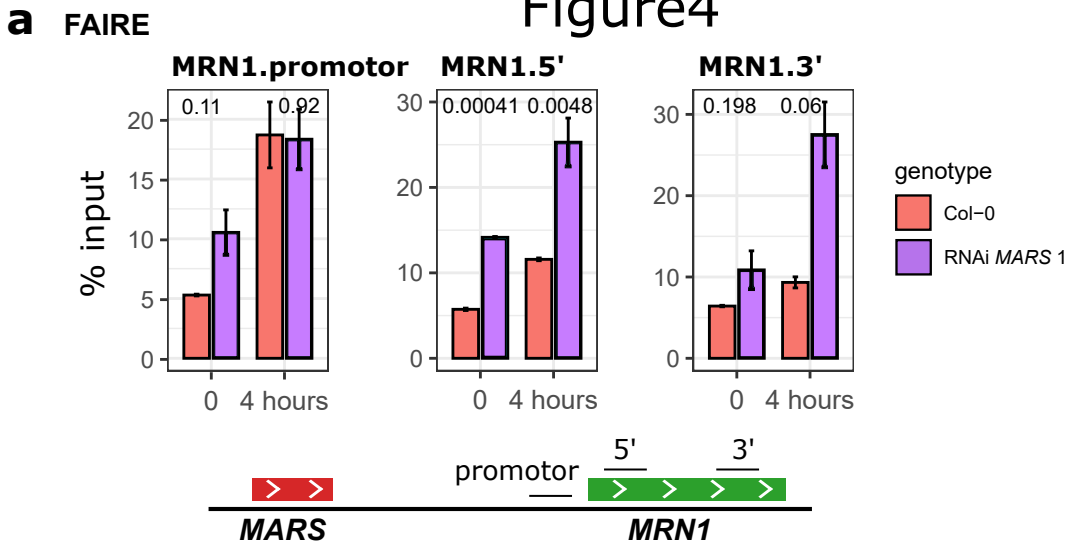
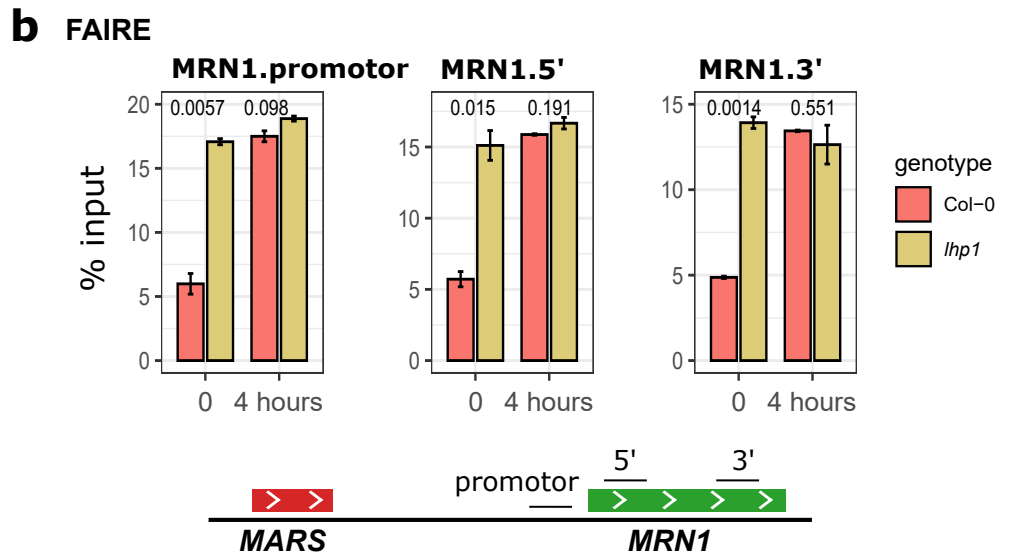
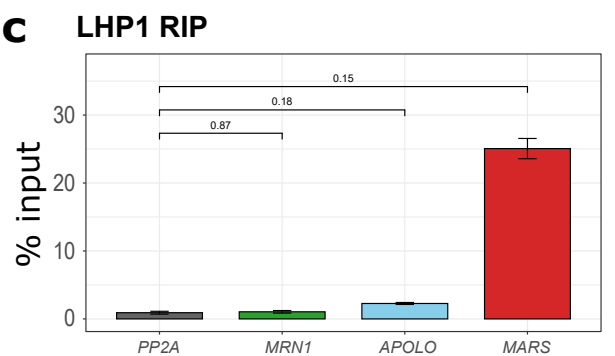


Figure4


a
FAIRE
MRN1.promotor
MRN1.5'
MRN1.3'
0.11
0.92
0.00041
0.0048
0.198
0.06
% input
0
4 hours
genotype
Col−0
RNAi MARS 1
promotor
5'
3'
MARS
MRN1


b
FAIRE
MRN1.promotor
MRN1.5'
MRN1.3'
0.0057
0.098
0.015
0.191
0.0014
0.551
% input
0
4 hours
genotype
Col−0
lhp1
promotor
5'
3'
MARS
MRN1


c
LHP1 RIP
0.15
0.18
0.87
% input
PP2A
MRN1
APOLO
MARS

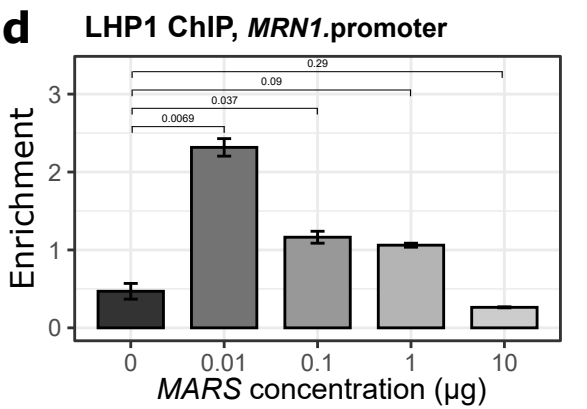
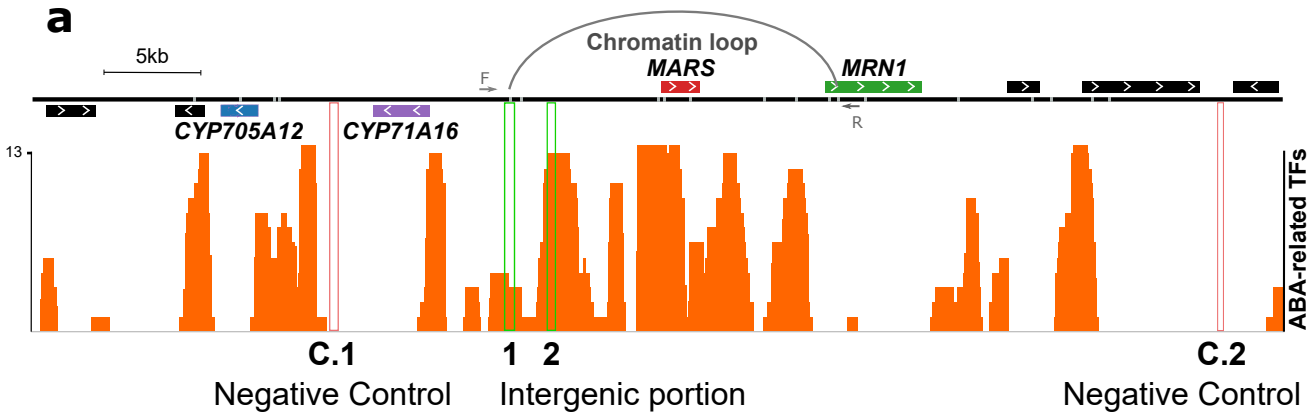
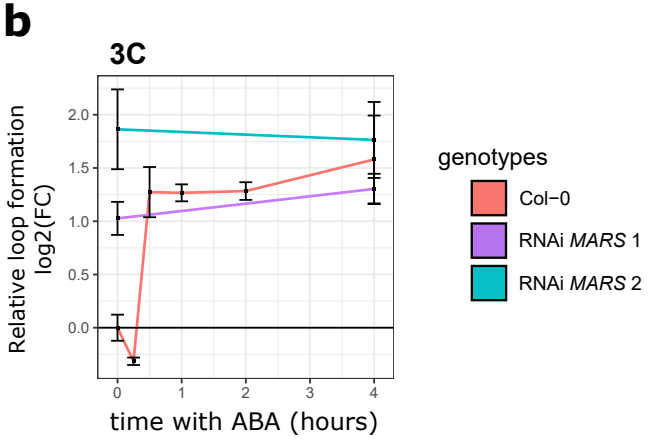
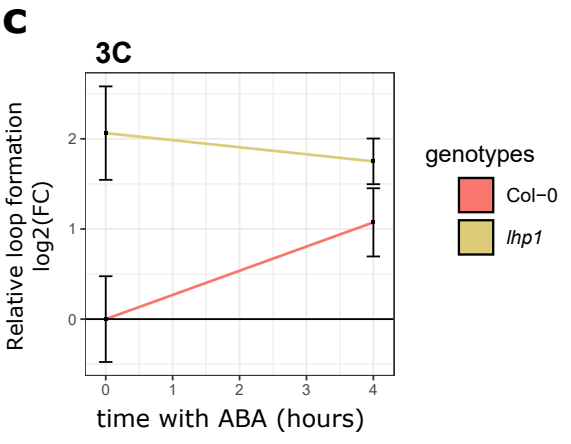

d
LHP1 ChIP, MRN1.promoter
0.29
0.09
0.037
0.0069
Enrichment
0
0.01
0.1
1
10
MARS concentration (µg)

Figure5


a
5kb
Chromatin loop
MARS
MRN1
F
R
CYP705A12
CYP71A16
13
ABA-related TFs
C.1
1
2
C.2
Negative Control
Intergenic portion
Negative Control


b
3C
Relative loop formation
log2(FC)
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
time with ABA (hours)
genotypes
Col−0
RNAi MARS 1
RNAi MARS 2


c
3C
Relative loop formation
log2(FC)
2
1
0
0
1
2
3
4
time with ABA (hours)
genotypes
Col−0
lhp1

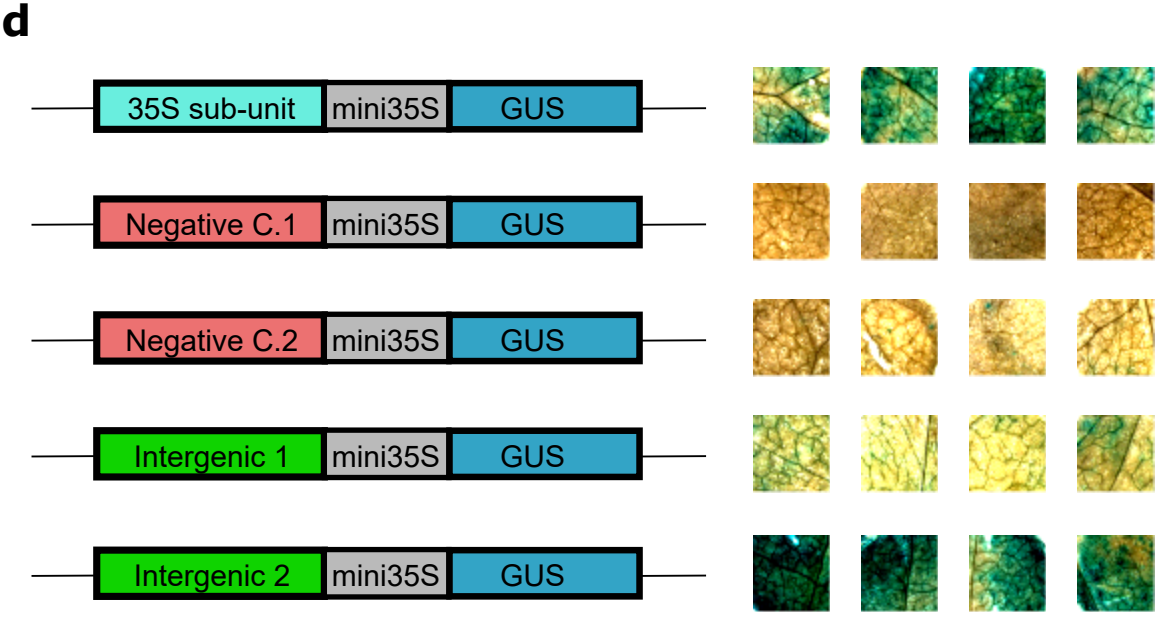
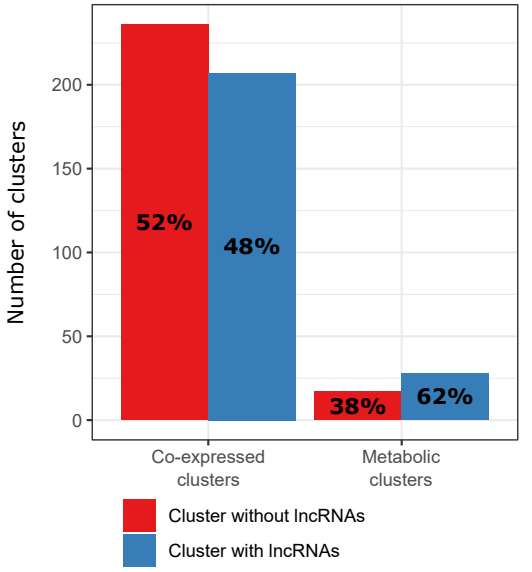
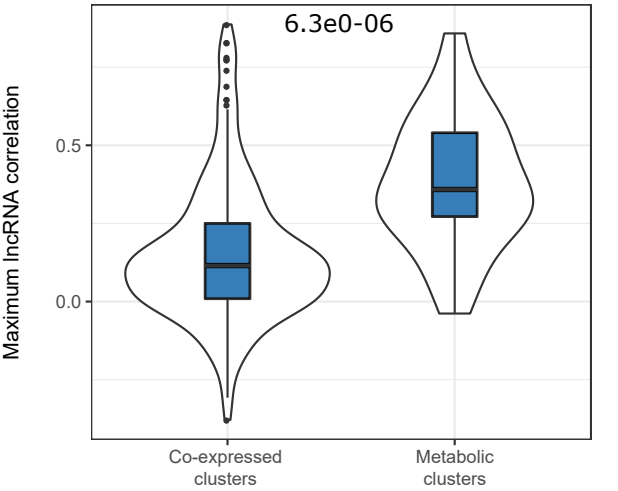

d
35S sub-unit
mini35S
GUS
Negative C.1
mini35S
GUS
Negative C.2
mini35S
GUS
Intergenic 1
mini35S
GUS
Intergenic 2
mini35S
GUS

Figure6

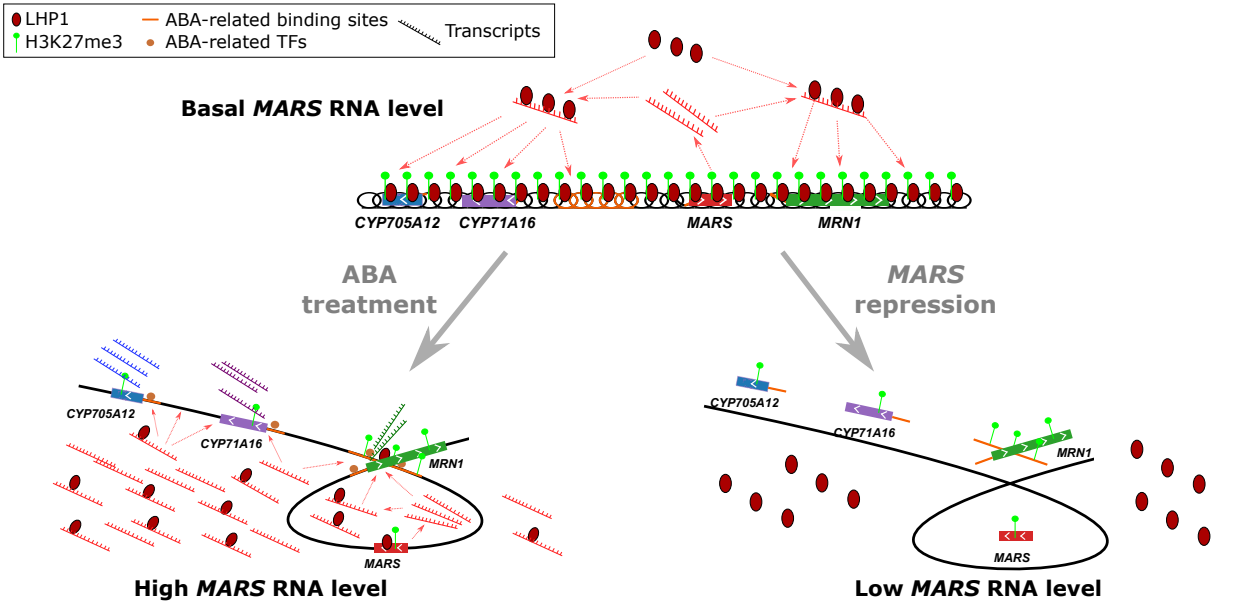
a


Number of clusters
200
150
100
50
0
52%
48%
38%
62%
Co-expressed clusters
Metabolic clusters
Cluster without lncRNAs
Cluster with lncRNAs

b


6.3e0-06
Maximum lncRNA correlation
0.5
0.0
Co-expressed clusters
Metabolic clusters

c


LHP1
H3K27me3
ABA-related binding sites
ABA-related TFs
Transcripts
Basal MARS RNA level
CYP705A12
CYP71A16
MARS
MRN1
ABA treatment
MARS repression
High MARS RNA level
Low MARS RNA level