Plant non-coding RNAs : important regulator of root growth and gene activity

Thomas Roulé Roulé

Thèse de doctorat

NNT : 2021UPASB047

université PARIS-SACLAY

Plant non-coding RNAs: important regulators of root growth and gene activity

ARN-non codants chez les plantes : régulateurs importants de la croissance racinaire et de l'activité des gènes

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 567 : Sciences du Végétal : du gène à l'écosystème (SEVE)
Spécialité de doctorat : Biologie
Unité de recherche : Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France
Référent : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Paris-Saclay, le 05/10/2021, par

Thomas ROULÉ

Composition du Jury

Cécile RAYNAUD Directeur de recherche, Université Paris-Saclay	Présidente & examinatrice
Antoine MARTIN Chargé de recherche, Université Montpellier	Rapporteur
Thierry LAGRANGE Directeur de recherche, Université Perpignan	Rapporteur
Rossana HENRIQUES Assistant professeure, University College Cork	Examinatrice

Direction de la thèse

Martin CRESPI Directeur de recherche, CNRS, IPS2	Directeur
Thomas BLEIN Chargé de recherche, CNRS, IPS2	Co-encadrant

Acknowledgments

Découvrir MARS, vivre un dégât des eaux, une pandémie, jouer au professeur, « quelques » blessures en skate, changer trois fois de logement, me marier ou encore partir aux Etats-Unis… Il s'en est passé des choses ces quatre dernières années… !

My first thank goes to Antoine Martin and Thierry Lagrange for accepting to read and evaluate my manuscript. Also, I would like to thank Rossana Henriques and Cécile Raynaud for being part of my thesis defense.

I want to thank my thesis committee, among which Etienne Delannoy, Nicolas Bouché, Jean-Christophe Palauqui and Benjamin Peret for their advice and helpful comments that help me to construct and verify my hypothesis.

My thanks also goes to the Doctoral School ED 567 which funded my thesis.

I want to thank all my colleagues from the teaching staff at Université Paris-Saclay, among which Sophie Nadot, Céline Charron and Sovanna Tan. More generally I would like to thank all the PhD students I met during this adventure, together with all the personnel and colleagues from the IPS2.

I deeply thank the members of the REGARN team. Aurélie, Christine, Céline S, Céline C, Caroline, Olivier, Hélène, Gautier, Andana, Yu-Ming and Agathe. Jérémie thank you so much for your enthusiasm, sense of humor, joy and scientific advice you provide.

Part of the team for a short but significant moment in my PhD life was Wil Prall, a smart and friendly visiting American PhD student from University of Pennsylvania. Thank you for your critical reading of my US fellowship and all the assistance you provide during my stay in the US. Not part of the team but a strong collaborator, I sincerely thank Federico Ariel for its help and guidance for many experiments. Together with being an amazing scientist you are also a fantastic affectionate person. Also, many thanks to all the undergraduate students I supervised, Nutthalak, Marie-Paule, Arya and Aya, I wish you all the best!

Finally, I want to express all my gratitude to my thesis supervisors Martin Crespi and Thomas Blein. Martin, thank you for your everyday enthusiasm, starting the day with a smiling "comment ça va muchacho?" was super-motivated, profound thanks for your positiveness and so-many smart ideas. You really inspired me during these years, and helped me to grow as a scientist. Thomas, thank you for all the time you gave me, you were always available to discuss my project and provide me relevant insights about how to do science at a high level. From the beginning to the end, you always accompanied me for my professional project and career path, and maintained daily contact even when the Atlantic Ocean was in between us, thank you for everything!

Aux galériens de Nantes

Quelle chance j'ai de vous avoir ! Les moments de détente en votre compagnie ont été, sont et seront toujours un vrai plaisir. Merci d'être là.

A ma famille

Maman, merci pour ton écoute, ton intérêt pour mon sujet, ma carrière professionnelle et personnelle. Tu as toujours été à 100% derrière moi, malgré mes quelques bêtises et étourderies, et je t'en remercie. Mes sœurs, neveux et nièces, merci pour votre bonne humeur, votre soutien et votre présence. J'attends toujours avec un sourire aux lèvres les moments de convivialité où nous nous réunissons tous.

A ma femme (et oui !)

Enfin, je tenais à te remercier Mohena, pour les trop nombreuses fois où je n'ai pas vu l'horloge tourner, pour supporter mon humeur parfois bancale, mais également pour tes « quelques » venus à l'IPS2 et même à l'UCR. Merci de m'accompagner dans cette aventure, jonché d'imprévus. Merci pour ton amour.

A mon père

Je dédie cette thèse à mon père, peut-être l'unique personne sur terre qui aurait été capable de lire l'intégralité de ce manuscrit, références y compris, avec passion et application malgré un niveau d'anglais A1. J'ai grâce à toi acquis une détermination, motivation et passion pour la science qui j'espère me portera le plus loin possible.

Rideau!

Table of contents

Abbreviations

aa	amino acid
ABA	abscisic acid
ABC1K3	ABC1-LIKE KINASE 3
ABCB3	ATP-BINDING CASSETTE B3, MDR3, P-GLYCOPROTEIN 3
ABI	ABA-INSENSITIVE
ActD	Actinomycin D
AG	AGAMOUS
AG-incRNA4	AGAMOUS INTRONIC RNA 4
AGL24	AGAMOUS-LIKE 24
AGO	ARGONAUTE
AMT1	AMMONIUM TRANSPORTER 1
AP1/2/3	APETALA1/2/3
APOLO	AUXIN REGULATED PROMOTER LOOP
APS	ATP sulfurylase
ARF	AUXIN RESPONSE FACTOR
AS	alternative splicing
ASCO	ALTERNATIVE SPLICING COMPETITOR
BGC	biosynthetic gene clusters
BLI	BLISTER
C	Carbon
CASP	CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN 5
CBF1	C-REPEAT/DEHYDRATION-RESPONSIVE ELEMENT BINDING FACTORS
ChIP	chromatin immunoprecipitation
CDS	coding sequence

CEBPB	CCAAT/ENHANCER-BINDING PROTEIN BETA
circRNA	circular RNA
CLF	CURLY LEAF
CNCI	Coding-Non-Coding Index
COME	Coding potential calculator based on multiple evidences
CPC	Coding Potential Calculator
CPL1	C-TERMINAL DOMAIN PHOSPHATASE-LIKE 1
CK	cytokinin
COOLAIR	COLD INDUCED LONG ANTISENSE INTRAGENIC RNA
COLDAIR	COLD ASSISTED INTRONIC NONCODING RNA
CRE	cis-regulatory elements
CSD	COPPER/ZINC SUPEROXIDE DISMUTASE
CUC2	CUP-SHAPED COTYLEDON 2
dCas9	Dead Cas9
dCas9a	Dead Cas9-transcriptional activator
DAS	Day after sown
DCL	DICER-LIKE
DNA	deoxyribonucleic acid
DRIP1/2	DREB2A-INTERACTING PROTEIN1/2
DRIR	DROUGHT-INDUCED LNCRNA
DRM	DOMAINS REARRANGED METHYLASE
ds	double strand
EAR	ERF-ASSOCIATED AMPHIPHILIC REPRESSION
ELENA1	ELF18-INDUCED LONG NONCODING RNA 1
EMF1/2	EMBRYONIC FLOWER 1/2
ENOD40	EARLY NODULIN40
FC	fold-change
FIS	FERTILIZATION-INDEPENDENT SEED
FLC	FLOWERING LOCUS C
FLD	FLOWERING LOCUS D

flg22	flagellin peptide 22
FT	FLOWERING LOCUS T
FUL/AGL8	FRUITFUL/AGAMOUS-LIKE8
GA	gibberellin
GGT2	GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE 2
GOLS3	GALACTINOL SYNTHASE3
GreenNC	Green Non-Coding Database
H3K27me3	histone 3 lysine 27 trimethylation
HEN1	HUA ENHANCER1
HID1	HIDDEN TREASURE1
HST	HASTY
HTT1/2	HEAT-INDUCED TAS1 TARGET1/2
HvE(Z)	ENHANCER OF ZESTE
HYL1	HYPONASTIC LEAVES1
IAA	INDOLE-3-ACETIC ACID
IHI	heterochromatic islands
incRNA	intronic lncRNA
indel	insertion/deletion
IPS1	INDUCED BY PHOSPHATE STARVATION 1
IPT3/7	ISOPENTENYLTRANSFERASE 3/7
JAL22	JACALIN-RELATED LECTIN 2
kb	kilobase
KNAT	KNOTTED 1-LIKE HOMEOBOX
LAIR	LRK ANTISENSE INTERGENIC RNA
LATERALINC	LATERAL LINCRNA
LEC2	LEAFY COTYLEDON 2
LHP1	LIKE HETEROCHROMATIN COMPLEX 1
lincRNA	long intergenic ncRNA
lncRNA	long non-coding RNA
LR	lateral root

LRX2	LEUCINE RICH EXTENSIN2
MARS	MARNERAL SILENCING
MED19a	MEDIATOR SUBUNIT 19A
MEL1	MEIOSIS ARRESTED AT LEPTOTENE 1
miRNA	micro-RNA
mRNA	messenger RNA
MRN1	MARNERAL SYNTHASE 1
MS	mass spectrometry
N	Nitrogen
NAC	NAM, ATAF and CUC
NAT	natural antisense transcript
ncRNA	non-coding RNA
NIR1	NITRITE REDUCTASE 1
NLR	NUCLEOTIDE-BINDING LEUCINE-RICH REPEAT
NRT1.1	NITRATE TRANSPORTER 1.1
NSR	NUCLEAR SPECKLE RNA-BINDING
nt	nucleotide
ORF	open reading frame
PcG	Polycomb group
phasiRNA	Phased secondary siRNA
PHO1/2	PHOSPHATE 1/2
Pi	phosphate
PIC	Pre-Initiation Complex
PID	PINOID
PIF3	PHYTOCHROME- INTERACTING FACTOR 3
PLEK	predictor of long non-coding RNAs and messenger RNAs based on an improved k-mer scheme
PLT1/2	PLETHORA 1/2
Pol	polymerase
PP2C	PHOSPHOENOLPYRUVATE CARBOXYLASE2

PPR	PENTATRICOPEPTIDE REPEAT
PR1	PATHOGENESIS-RELATED GENE1
PRC	Polycomb repressive complex
pre-mRNA	precursor messenger RNA
PRP8	PRE-MRNA PROCESSING 8
PTL	PETALLOSS
QQS	QUA-QUINE STARCH (QQS)
RBP1	RNA-BINDING PROTEIN 1
RdDM	RNA-directed DNA methylation
RDM	RNA-DIRECTED DNA METHYLATION
RDR	RNA DEPENDENT RNA POLYMERASE
RHD6	ROOT HAIR DEFECTIVE 6
RISC	RNA-induced silencing complex
RGF7	ROOT MERISTEM GROWTH FACTOR 7
RKN	root-knot nematode
RNA	ribonucleic acid
RNP	ribonucleoprotein
ROS	reactive oxygen species
RRP	RIBOSOMAL RNA PROCESSING
rRNA	ribosomal RNA
SAFA	SCAFFOLD-ATTACHMENT-FACTOR A
SCF-TIR1	SKP1-CULLIN-F-BOX/TRANSPORT INHIBITOR RESPONSE 1
SCR	SCARECROW
SCL4	SCARECROW- LIKE PROTEIN4
SE	Super Enhancer
SEP3/4	SEPALLATA 3/4
SGS3	SUPPRESSOR OF GENE SILENCING
siRNA	small interfering RNA
slr	solitary
SM	secondary metabolites

SmD1B	SMALL NUCLEAR RIBONUCLEOPROTEIN
SNP	single-nucleotide polymorphism
snRNA	small nuclear RNA
SPL3	SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3
sRNA	small RNA
ss	single strand
STK	SEEDSTICK
STM	SHOOT MERISTEMLESS
SVP	SHORT VEGETATIVE PHASE
SWI/SNF	SWItch/Sucrose Non-Fermentable
TAD	Topologically Associated Domains
TAS3	TRANS-ACTING SIRNA3
TE	transposable element
TES	transcription end site
TF	transcription factor
TFL1	TERMINAL FLOWER 1
TYLCV	tomato yellow leaf curl virus
TriFC	trimolecular fluorescence complementation
tRNA	transfer RNA
TSS	transcription start site
UTR	untranslated region
VDD	VERDANDI
VRN2	VERNALIZATION2
WDR5	WD REPEAT DOMAIN5
WOX1	WUSCHEL RELATED HOMEOBOX 1
WSGAR	WHEAT SEED GERMINATION ASSOCIATED RNA
WUS	WUSCHEL
ZAT6	ZINC FINGER OF ARABIDOPSIS THALIANA 6

List of Figures and Tables

Introduction

Results

I Introduction

1. Non coding RNAs: from junk to utility

The availability of genome sequences of numerous living organisms markedly changed our vision of the genome. Until the 1940s, it was proposed that genome length could positively correlate with the organism's complexity (Elliott and Gregory, 2015). Nevertheless, the strong differences of genome length observed between closely related organisms and the existence of very long genomes within simple organisms, noted in the early 1950s, give rise to a convolutional quandary, the so-called "C-value paradox" (Thomas, 1971). Later, the uncovering of large non-protein-coding DNA regions within the genome resolved the lack of correlation between genome length and species complexity. Anyhow, this finding raises multiple questions concerning the relevance of these non-protein-coding sequences: How these sequences appear? Are they functional? Why don't they exist in all organisms? Together, these inquiries spawn the "C-value enigma" (Gregory, 2001). The extent advances of RNA sequencing methods revealed that more than 90% of eukaryotic genomes are transcribed to RNA with only 2-10% of individual transcripts translated into protein (Gil and Ulitsky, 2020). As expected, this tremendous quantity of transcripts is significantly different from each other, some are constitutively activated and participate in the basal function of cell life, such as gene transcription or traduction; others are specifically activated upon particular stimuli or tissues, presenting regulatory function.

In this first part, I will describe the different types of transcripts found within a cell, their characteristics and specificity, and focus on a specific class of transcripts: the long non-coding RNAs (lncRNAs).

1.1 A large diversity of transcripts

1.1.1 Cooperation of non-coding and coding transcripts for protein synthesis

The protein composition within a cell directly influences its fate. For coding genes, the final step of gene expression is the mRNA translation, in which the genetic information contained within the transcript is decoded into a protein (Cramer, 2019).

First, the gene is transcribed into an RNA molecule which is getting mature through the addition of a poly(A) tail and a N7-methylated guanosine linked to the first nucleotide in 5′. These modifications increased the RNA stability and translation efficiency (Shatkin, 1976). The linear RNA sequence can then be modified, or spliced, by the spliceosome, leading to the eviction of some portion of the transcript (Filichkin et al., 2015). In humans, 73 mutations impairing the splicing process has been linked to cancer predisposition (Tate et al., 2019), strengthening the importance of this mechanism for cell life. Most of the time, during splicing all intronic regions are removed from the transcript. However, certain introns can be retained and certain exons can be skipped, leading to different mature RNA from one locus through Alternative Splicing (AS). Interestingly, in plants the majority of AS relates to intron retention, whereas in humans it consists in exon skipping (Sammeth et al., 2008; Wang et al., 2008). The spliceosome is constituted of small nuclear RNAs (snRNAs) that bind to proteins which together form a ribonucleoprotein complex and catalyze the splicing (Wahl et al., 2009). Notably, more than 90 percent of pre-mRNA transcripts are spliced in humans, many in a tissue-dependent or developmental-related manner (Wang et al., 2008). Similarly, in higher plants, more than 60 % of transcripts containing intron undergo AS (Chaudhary et al., 2019). Once mature, the coding transcripts exit the nucleus and are bound and decoded by the ribosomal complex, a large complex form with multiple proteins and ribosomal non-coding RNAs (rRNAs). The ribosomal complex is made of two protein subunits heterogeneous in size, together with four rRNAs implicated in the assembly and functionality of the ribosomes (Merchante et al., 2017). This complex is able to recognize specific sequences from the coding transcripts and produce the corresponding protein with the help of another type of non-coding transcripts: the transfer RNA (tRNA), which harbors the amino-acids (Sanchita et al., 2020). Strikingly, it has been estimated that one amino acid is transferred every 60ms to the forming polypeptides chain (Zaher and Green, 2009). This process is called mRNA translation and it is realized through the cooperation of proteins and non-coding transcripts. In addition, small nucleolar RNAs are another class of structural RNAs that conduct chemical modifications, such as methylation and pseudouridylation, of snRNAs, rRNAs and tRNAs, participating in their maturation (Streit and Schleiff, 2021). Together, mRNA splicing and translation are fascinating examples showing the importance of non-coding transcriptional units cooperating with proteins for the one basal function of the cell: the protein synthesis.

1.1.2 The small non-coding regulatory transcripts

In addition to these house-keeping non-coding RNAs, new regulatory RNAs have been identified in recent years. Small regulatory non-coding RNAs or small RNAs (sRNAs) are defined as transcripts with a length in-between 19-25nt and not translated into protein. They are issued from double strands (ds) of RNA. Depending on their biogenesis and their mechanism of action they can be classified in different major types: small interfering RNAs (siRNAs), microRNAs (miRNAs) and phasiRNAs (Brant and Budak, 2018), all of which, fine-tune the transcriptional or post-transcriptional gene activity (Voinnet, 2009; Budak and Akpinar, 2015).

1.1.2.1 The siRNAs

SiRNAs are 21-24nt length RNAs which inhibit transcription through DNA methylation via RNA dependent DNA methylation (RdDM). This mechanism is notably critical for the silencing of Transposable Element (TE) activity in both animals and plants (Zemach and Zilberman, 2010). Also, RdDM-deficient Arabidopsis mutants present higher TE activity, even though transposition is still pretty rare (Ito et al., 2011). In the TE-rich maize genome, methylated CHH islands are found in-between activated genes and silent-TE, preventing the spreading of active euchromatin to inactive heterochromatin (Li et al., 2015a).

RdDM action is a well described process involving siRNAs that arise from genes transcribed by either RNA Polymerase II (Pol II) or Pol IV (**Figure 1**). Interestingly, Pol IV is a plant-specific Polymerase that shares similar subunits with Pol II (Haag and Pikaard, 2011) and specifically implicated in siRNA biogenesis (Matzke and Mosher, 2014; Matzke et al., 2015).

Upon Pol II transcription, the single strand (ss) siRNA-gene RNA is converted to a dsRNA through transcription of its corresponding strand by the RNA DEPENDENT RNA POLYMERASE6 (RDR6). Dicer-Like2 (DCL2) or DCL4 cut the dsRNA into 21 and 22nt, respectively. One of the strands is finally included in the RNA-Induced Silencing Complex (RISC) through its interaction with Argonaute6 (AGO6) (Wu et al., 2012; Nuthikattu et al., 2013; McCue et al., 2015). Interestingly, Pol II transcripts can also naturally generate hairpin structures through base complementarity leading to a dsRNA-like structure without the need of RDR6. In this case, the

dsRNA is cleaved by DCL3 every 24nt and one of the strands is included in the RISC through its interaction with AGO4 or AGO6 (**Figure 1; Pol II-dependent siRNA production**). On the other hand, Pol IV dependent transcripts are converted into dsRNA with RDR2 (Smith et al., 2007; Law et al., 2013). Then, the dsRNA is processed into 21-24nt through cleavage with DCL2/3 or DCL4. Interestingly, an important proportion of RdDM targets are still methylated within *dcl1/dcl2/dcl3/dcl4* quadruple mutants, indicating a DCL-independent dsRNA cleavage (Yang et al., 2016). Anyhow, one of the strands is loaded into the RISC through interaction with AGO4 or AGO6 (**Figure 1; Pol IV-dependent siRNA production**).

Once the RISC is formed, it associates with the chromatin through base complementarity in-between the ssRNA associated with AGO and the DNA and mediates DNA methylation with the help of other proteins and co-factors. More precisely, for both Pol II or Pol IV dependent siRNA synthesis, the single strand RNA directs the RISC through base complementarity with a nascent Pol V transcript on the region where the DNA methylation occurs. Finally, RISC is able to recruit the DNA methyltransferase DOMAINS REARRANGED METHYLASE2 (DRM2) which methylate the DNA with the help of RNA-DIRECTED DNA METHYLATION1 (RDM1) which bind to both AGO4/AGO6 and DRM2. Interestingly, to maintain Pol V transcript next to the chromatin and facilitate DNA methylation, the yeast and mammalian homolog Ribosomal RNA processing 6 (RRP6), retain the RNA close to the active Pol V. In addition, the Pol V RNA-AGO RNA interaction is strengthened by the INVOLVED IN DE NOVO 2 (IDN2)–IDN2 PARALOGUE (IDP) complex that bind SWIB3 (part of the SWI/SNF complex) participating in chromatin decondensation, facilitating Pol V activity (Zhang et al., 2018a) (**Figure 1**). Methylated DNA positively correlates with a condensed chromatin and a low transcriptional activity. The effect of DNA methylation on gene expression is detailed further in section 2.2.1.

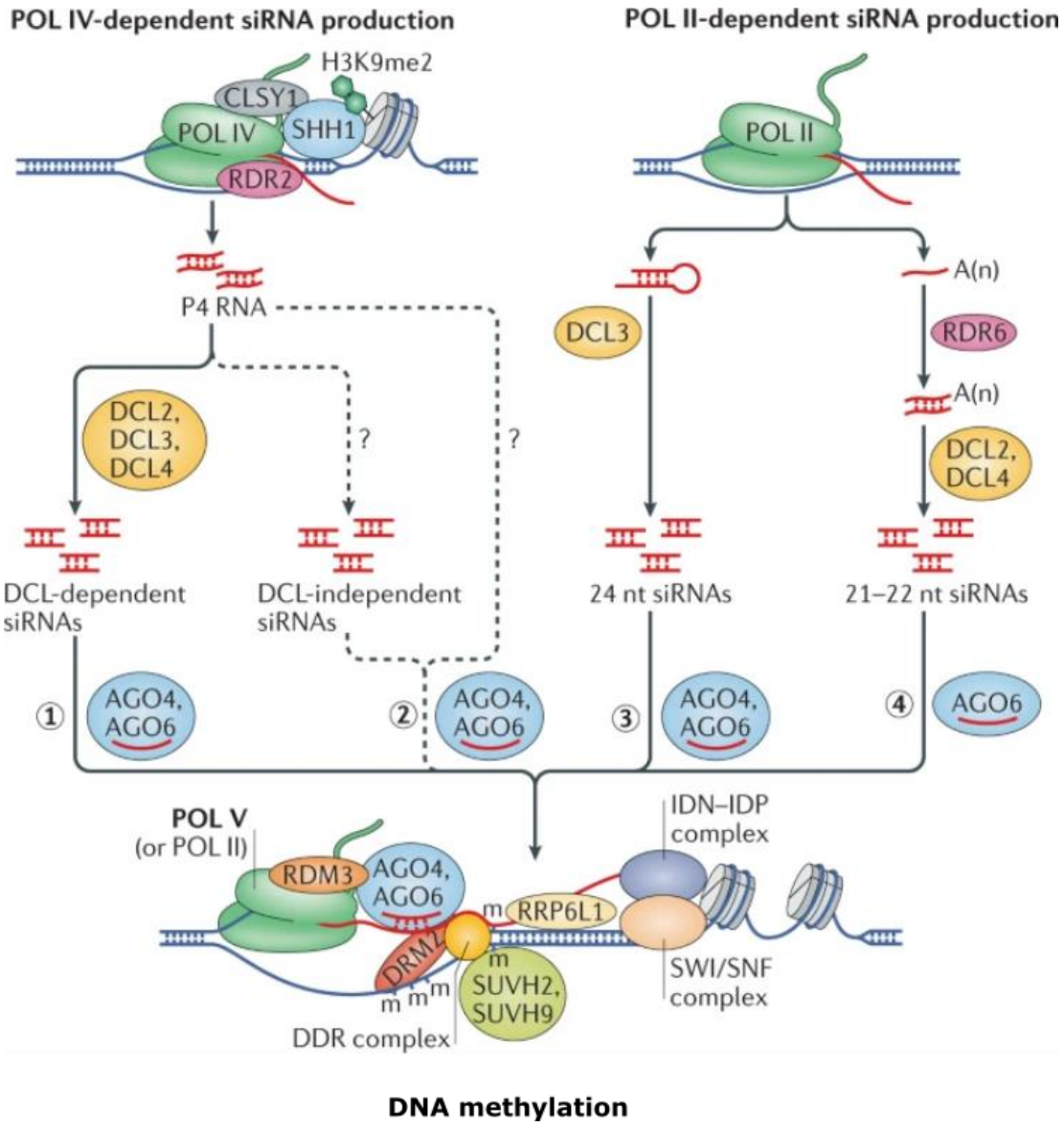


Figure 1: RNA Directed DNA methylation pathways (Adapted from Zhang et al., 2018)**.** For Pol-IV dependent siRNA production, the Pol IV recruitment to the chromatin is facilitated by SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1), which binds dimethylated histone H3 lysine 9 (H3K9me2) (Law et al., 2013; Zhang et al., 2013). Also, Pol IV processing is facilitated by the chromatin remodeler SNF2 DOMAIN-CONTAINING PROTEIN CLASSY 1 (CLSY1) (Zhang et al., 2013; Smith et al., 2007). Pol IV dependent non-coding RNAs (P4 RNAs) are transformed into a dsRNA via the RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). The dsRNA is either cleaved by DICER-LIKE PROTEIN 2 (DCL2), DCL3 and DCL4 (pathway 1), or cleaved by non-DCL proteins (pathway 2), both leading to mainly 24nt siRNAs. These siRNAs interact with ARGONAUTE 4 (AGO4) or AGO6 and pair with a Pol V dependent scaffold transcript which together recruit DOMAINS REARRANGED METHYLASE 2 (DRM2), which methylates the DNA. Interestingly, Pol II can also generate RdDM-related siRNAs. On one hand, Pol II transcripts can harbor a stem loop structure which is cleaved by DCL3 generating 24nt siRNAs (pathway 3). On the other hand, Pol II transcripts can serve as template for RDR6 mediated dsRNA synthesis (Wu et al., 2012; Nuthikattu et al., 2013; McCue et al., 2015), which is then cleaved by DCL2 and DCL4 generating 21-22nt siRNAs (pathway 4). Finally, the Pol II dependent siRNAs transcripts interact with AGO4 or AGO6. The association between the AGOs complex and Pol V is facilitated by RNA-DIRECTED DNA METHYLATION 3 (RDM3) (He et al., 2009; Bies-Etheve et al., 2009). The production of Pol V dependent scaffold RNA needs the DDR complex, comprising the chromatin remodeler DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1, DEFECTIVE IN MERISTEM SILENCING 3 and RDM1. The DDR complex interacts with AGO4, DRM2, SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE PROTEIN 2 (SUVH2) and SUVH9, and bind single-stranded methylated DNA and recruit Pol V (Gao et al., 2010; Kanno et al., 2004; Kanno et al., 2008; Law et al., 2010; Zhong et al., 2012; Liu et al., 2014; Johnson et al., 2014). Finally, the retention of Pol V transcripts is mediated by the RNA-binding proteins RRP6-LIKE 1 (RRP6L1 (Zhang et al., 2014)) and the INVOLVED IN DE NOVO 2 (IDN2)–IDN2 PARALOGUE (IDP) complex, which interacts with the SWITCH/SUCROSE NONFERMENTING (SWI/SNF), a chromatin-remodeling complex (Ausin et al., 2009; Zheng et al., 2010; Ausin et al., 2012; Zhang et al., 2012; Finke et al., 2012; Xie et al., 2012; Zhu et al., 2013).

1.1.2.2 The miRNAs

MicroRNAs (miRNAs) are 19-24nt length RNAs which inhibit gene expression at the post-transcriptional level. Using their base complementarity, they can direct a RISC to other RNAs triggering their cleavage or avoiding their translation into protein (**Figure 2**). Strikingly, it has been estimated that one third of the human genome is regulated by miRNA (Hammond, 2015). Rationally, in plants miRNA are implicated in almost all biological processes such as development and responses to biotic and abiotic stimuli (Budak and Akpinar, 2015). MiRNA genes generate 100-200nt length pri-miRNA transcripts that naturally form a stem-loop structure through base complementarity. In plants, a complex composed notably of DCL1, the dsRNA binding protein HYPONASTIC LEAVES1 (HYL1), nuclear cap binding complex (CBC) and a C2H2-type zinc finger (SERRATE) processed the pri-miRNA by removing 15nt from each extremity of the RNA stem-loop to form the precursor miRNA (pre-miRNA) (Axtell and Meyers, 2018). Then, the pre-miRNA is cleaved into a 20-24nt fragment to form a miRNA/miRNA* duplex. This duplex is methylated at the 3' end by HEN1 which protects them from degradation. Finally, one of the strands is taken by AGO1 to form the RISC, which together mediates mRNA cleavage or translational repression through base complementarity binding of the miRNA complex with the RNA target (**Figure 2**).

Initially, it was thought that plant miRNAs mainly act by transcript cleavage thanks to the high sequence complementarity to their targets. Nevertheless, plant miRNAs are also found enriched on membrane-bound polysomes (Li et al., 2016a) increasing evidence that they also act through translation repression as mammal's miRNA (Millar and Waterhouse, 2005). For example, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3* (*SPL3*), *REVOLUTA, SCARECROW-LIKE PROTEIN 4* (*SCL4*), *APETALA 2* (*AP2*), and *COPPER/ZINC SUPEROXIDE DISMUTASE 1* (*CSD1*) and *CSD2* are all mRNAs targeted by miRNA and subjected to both transcript cleavage and repression of the translation (Yu et al., 2017). Also, as miRNA are relatively small, they may target different mRNA that share similar sequences, likely belonging to the same gene family. For example, in *Arabidopsis thaliana miR156* regulates the transition from vegetative to flowering phase by changing the expression of several *SPL* genes (Wang, 2014), and the *miR167* targets various *ARF* genes implicated in auxin signaling (Xia et al., 2015). Also, miRNA evolution may recruit new targets in one species. For example, the *miR396-GRFs* regulation is conserved

in angiosperms and gymnosperms (Chorostecki et al., 2012; Debernardi et al., 2012) but *Medicago truncatula* miR396 represses *MtGRF* together with two *MtbHLH79* affecting root growth and mycorrhizal colonization (Bazin et al., 2013).

Mutant plants impaired in the miRNA biogenesis pathway present strong developmental defects suggesting miRNA are implicated in a wide range of biological processes (Budak and Akpinar, 2015). For example, *Arabidopsis dcl1* knock-down mutants exhibit pleiotropic developmental defects, and the *dcl1* knock-out is embryo-lethal (Chen, 2005). Similarly, disruption of the *HYL1*, *HEN1*, *DCL1*, *HST* and *AGO1* genes impaired fertility (Cubillos et al., 2012; Oliver et al., 2017). Also in rice, proper function of the *MEIOSIS ARRESTED AT LEPTOTENE 1* (*MEL1*), component of the RISC, is necessary for proper pollen grain development (Nonomura et al., 2007; Komiya et al., 2014). Similarly, the maize *AGO104*, also functioning through RdDM , is directly involved in meiosis (Singh et al., 2011). MiRNAs activity also participates in the plant response to environmental cues. For example, in *Gossypium hirsutum, miRNVL5* expression decreases upon salt stress, positively participating in the plant salt stress response through the increase of expression of *GhCHR*, a positive regulator of salt stress tolerance (Gao et al., 2016). Similarly, the *Arabidopsis miR156* is implicated in heat stress resilience and memory through the regulation of *SPL* genes (Stief et al., 2014), whereas the *miR398* participates in heat tolerance via csd-mRNA cleavage (Guan et al., 2013). Finally, miRNA assists the assimilation of key nutrients. For example, the Arabidopsis *miR395* targets the *ATP sulfurylase (APS)* gene transcript involved in sulfate assimilation, participating in the plant sulfate metabolization (Jones-Rhoades and Bartel, 2004). Also, the *miR399* generated from 6 loci participates in the plant Pi homeostasis through the targeting of *PHO2* mRNA (Pant et al., 2008). *PHO2* encodes for an ubiquitin-conjugating E2 enzyme that mediates the degradation of the PHO1 protein, implicated in Pi loading to the Xylem (Liu et al., 2012a). Logically, overexpression of the *miR399* increased PHO1 protein level leading to Pi hyperaccumulation (Chiou et al., 2006). Given the direct implication of miRNA for plant development and response to environmental stresses, their manipulation has the potential to enhance some agronomic traits.

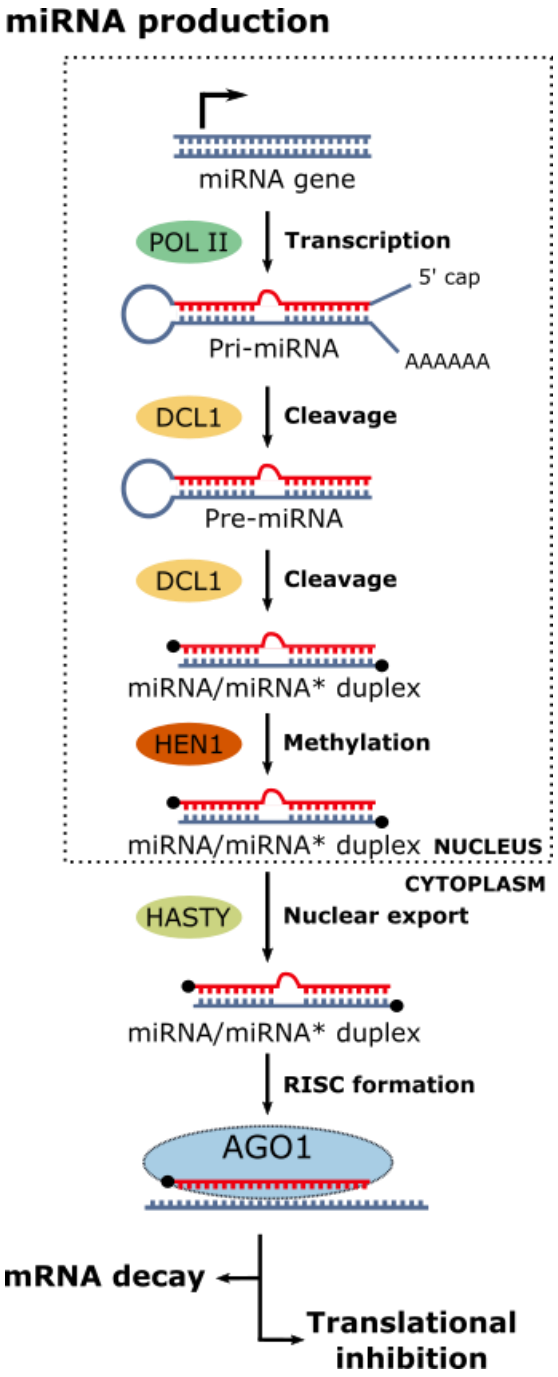


Figure 2: The miRNA pathway (Adapted from Li et al., 2018). Pol II-dependent miRNA precursor transcripts generate a natural hairpin structure through base pair complementarity forming the pri-miRNA. Polyadenylated tail and 5′ cap are removed through DCL1-mediated cleavage generating the pre-miRNA. The pre-miRNA is further cleaved by DCL1 resulting in mature miRNA/miRNA* duplex in which the duplex extremity is methylated by HEN1. HASTY translocate the miRNA/miRNA* duplex from the nucleus to the cytoplasm. Finally, one strand of the miRNA/miRNA* duplex is recruited by AGO1 forming the mature RNA-induced silencing complex (RISC), which interacts with the target mRNA through base complementarity. Binding of the miRNA-RISC complex to the mRNA can generate cleavage of the mRNA or inhibition of the translation.

1.1.2.3 The phasiRNA and tasiRNAs

Phased secondary siRNAs (phasiRNAs) are generated after miRNA-mediated RNA cleavage of specific transcripts and leads to the production of regulatory sRNAs acting likely as miRNAs (Allen et al., 2005). Indeed, the miRNA-mediated RNA cleavage does not always lead to direct degradation of the cleaved transcript. In certain cases, part of the cleaved RNAs, named as phased transcripts, is processed into phasiRNAs through the *RDR/DCL* machinery (Chen et al., 2010). More precisely, SUPPRESSOR OF GENE SILENCING (SGS3) binds to the phased transcripts to avoid their degradation and facilitates their dsRNA transformation through RDR6. Once in double strand form, DCL4 processed them regularly into 21nt long phasiRNA duplex which are methylated by HEN1 and processed into the AGO1-RISC as for miRNA (Brodersen and Voinnet, 2006). Consequently, phasiRNA are likely to function like miRNA to repress gene expression via transcript cleavage or repression of the translation (Chen, 2009; Yu et al., 2017) (**Figure 3**).

PhasiRNA were initially status as tasiRNA for Trans Acting Secondary siRNA since they were initiated from the non-coding *TAS* gene transcripts (Vazquez et al., 2004; Felippes and Weigel, 2009). As for miRNA they are involved in various aspects of plant development and its response to the environment. For example, the *miR390-TAS3-ARF* pathway is implicated in lateral root growth, leaf formation, embryogenesis and is conserved in plants. Notably, in *ago7* or *dcl4* tomato mutants, unable to generate phasiRNAs, the increased level of *ARFs* transcript leads to abnormal leaf shape (Yifhar et al., 2012). Similarly, *Medicago truncatula* and maize *ago7* mutants present cylindrical leaves with irregular polarity (Douglas et al., 2010; Zhou et al., 2013). Also, overexpression of *miR3954* triggers phasiRNAs production from *NAC* genes in citrus, resulting in early flowering (Liu et al., 2017a). Interestingly, a wheat lncRNA *WSGAR* is targeted by *miR9678* and is involved in seed development and germination (Guo et al., 2018). PhasiRNAs are also involved in abiotic stresses. For example, in *Arabidopsis*, *miR173* targets the *HEAT-INDUCED TAS1 TARGET1* (*HTT1*) and *HTT2*, where increasing the *miR173* activity and subsequent phasiRNAs production increase the heat plant sensitivity (Li et al., 2014a). Also, drought-stress resilience of *Populus* plants involved *miR482*, *miR828*, and *miR6445* activity. More precisely, *miR6445* targets *NAC* genes resulting in phasiRNAs productions targeting other *NAC* genes (Xie et al., 2017). Closely, in legume, drought stress triggers the production

of *NAC700*-phasiRNA through the *miR1514a* activity, affecting plant drought tolerance (Sosa-Valencia et al., 2017). Finally, the sweet potato *miR828* accumulates in response to wound and target *IbMYB* and *IbTLD* implicated in lignin and H2O2 content, affecting plant damage recovery (Lin et al., 2012). Wound is a mechanical stress induced under abiotic and biotic stresses. Interestingly, the plant biotic interaction, either beneficial or detrimental, is also governed by phasiRNAs. Among them, the immune receptor *NUCLEOTIDE-BINDING LEUCINE-RICH REPEAT* (*NLR*) and *PENTATRICOPEPTIDE REPEAT* (*PPR*) genes produced phasiRNA after miRNA-mediated cutting and are implicated in plant-microbe interactions, symbiosis and defense (Fei et al., 2013). Other resistance genes are targeted by phasiRNAs, supporting that phasiRNAs are important factors of plant immunity. For example, *miR9863* from barley and wheat targets an *MLA* gene encoding an NLR protein (Liu et al., 2014a). Similarly, *NLR* genes from the norway spruce are targeted by the *miR482/2118* family and generate phasiRNAs (Xia et al., 2015).

Altogether, miRNAs, siRNAs and phasiRNAs, demonstrate the complexity of sRNA-mediated gene regulation, and the importance that they have in all aspects of plant life from organogenesis, to perception of the environment. However non-coding transcripts are not only small or precursors of small RNAs.

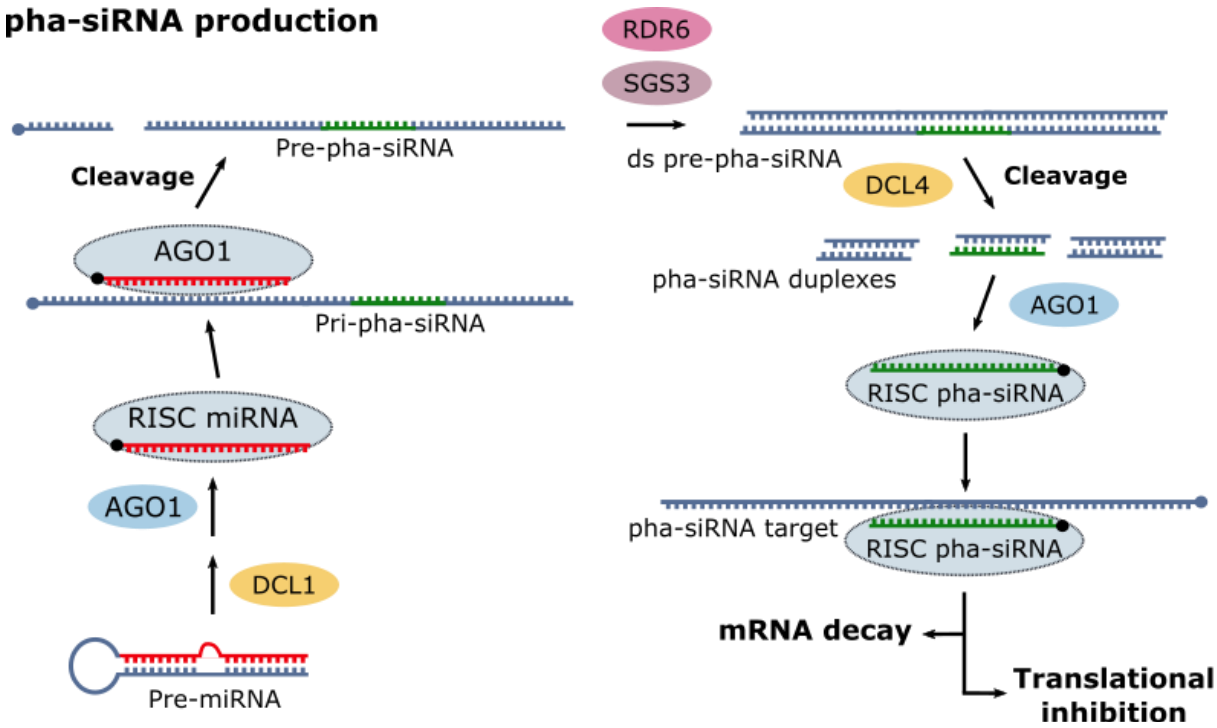


Figure 3: The pha-siRNA pathway (Adapted from Xie et al., 2005). Certain mRNAs cleaved by miRNAs can be processed into phasiRNAs. After miRNA-RISC mediated mRNA cleavage, part of the cleaved transcript, named the pre-pha-siRNA, is converted into a dsRNA through RDR6 and SUPPRESSOR OF GENE SILENCING3 (SGS3). DCL4 cleaves processively in phase the ds-pre-pha-siRNA every 21nt forming phasiRNA duplexes which are methylated by *HEN1*. Finally, one of the strands of the phasiRNA/phasiRNA* duplex is incorporated into the RISC through AGO1 interaction and mediates translational inhibition or transcript degradation of their target transcripts.

1.2 Discovery of the Long Non-Coding RNAs world

Among the plethora of transcripts produced within the cells, the long non-coding RNAs (lncRNAs) emerged recently as important regulators of gene expression. Their low expression level and their cell-type specific transcriptional activity could explain why they were discovered later than other classes of transcripts, with the increased sensitivity of next-generation sequencing methods (Jha et al., 2020). The intensive progress in bioinformatic for genome annotation, mainly through the use of RNA sequencing (RNA-seq) method, participates also in the *in-silico* finding of novel lncRNAs. The increasing number of pipelines able to decipher if a transcript is coding or non-coding based on multiple parameters including, the transcript sequence, secondary structure and the RNA conservation significantly improve the genomic annotations in all studied organisms (Li et al., 2020). For example, the CPC uses both sequence alignment and Open Reading Frame (ORF) length and coverage (Kong et al., 2007), whereas the phyloCSF uses known protein databases together with sequence alignment to classify a transcript as coding or non-coding (Lin et al., 2011). Then, CNCI and PLEK use nucleotide composition such as GC content and *k-mer* occurrence (Sun et al., 2013; Li et al., 2014b), whereas the COME uses structural features of the transcript and epigenetic information to decipher its coding ability (Hu et al., 2017). These tools are based on a machine learning approach, where the software is trained on a set of known non-coding and coding transcripts (Li et al., 2020). More recently, the newly developed tools use a whole set of alignment and non-alignment methods to analyze multiple transcripts features, such as the ORF, RNA secondary structure, peptide isoelectric point and Ficket score. Among these tools, CPC2, emerged as a widely used coding prediction tool for a large range of species (Kang et al., 2017). Thus, even though lncRNAs were initially described as transcriptional noise within the cell, they are now intensively studied and considered as important regulators of cell life from animals to plants (Ariel et al., 2015).

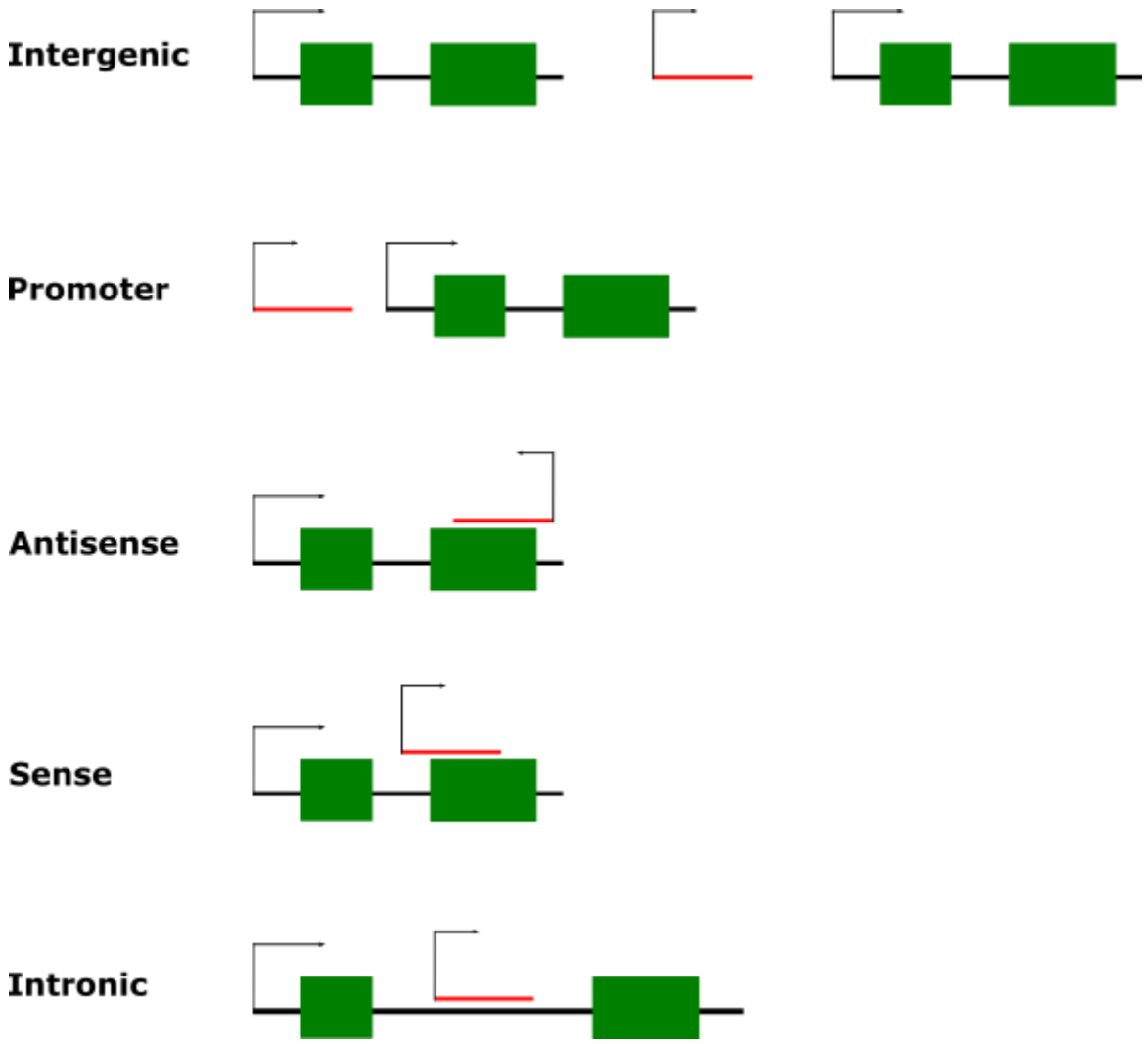


Figure 4: The different classes of lncRNAs. LncRNAs are classified according to their position relative to other genes. Long Intergenic ncRNAs (lincRNAs) are transcribed without overlapping any gene or promoter. Promoter lncRNAs are transcribed within the promoter region of a gene without overlapping the core gene region. LncRNA genes overlapping with an exon of another gene in its opposite direction are named Natural Antisense Transcript lncRNA (NAT-lncRNA), whereas it is named sense lncRNA when it overlaps an exon in the same direction as the transcript. Finally, lncRNA can be transcribed within an intron, generating intronic lncRNAs.

1.2.1 LncRNAs features

LncRNAs are transcripts superior to 200nt in length presenting a low ability to encode for proteins. They can be classified according to their genomic position in relation to other genes (**Figure 4**). First, the intronic lncRNAs are lncRNAs transcribed within the intron of another gene that does not overlap with any exons, independently of its orientation (Ariel et al., 2015), while they are classified as exonic or sense lncRNAs if they overlap with the exon of a gene and on the same strand. Indeed, lncRNAs overlapping with exonic genomic regions but on the other strand of the DNA are classified as natural antisense lncRNA (NAT-lncRNA). LncRNAs transcribed from the promoter of a gene constitute the promoter-lncRNA. Finally, the intergenic lncRNAs (lincRNAs) are lncRNAs transcriptional units localized in-between two genes, thus constituting an individual isolated transcriptional unit (Ariel et al., 2015) (**Figure 4**).

The number of discovered lncRNAs increased together with the development of sequencing and bioinformatic prediction tools. For example, in 2017 the FANTOM5 project identified around 27 919 lncRNAs in humans, whereas in 2019 it has been estimated that the human genome contains around 270 044 lncRNAs (Ma et al., 2018). Similarly in plants, the number of annotated lncRNAs significantly increased over time. For example, the study of Amor (2009) identified only 76 lncRNAs in *Arabidopsis* from detailed expertise of full-length cDNA databases, whereas a more recent study using RNAseq identified 6 510 lncRNAs, among which 4 050 NAT-lncRNAs and 2 460 lincRNAs (Zhao et al., 2018a). Now, more than thousands of lncRNAs have been identified within more than 40 plant species. For example, the *Green Non-Coding Database* (GreeNC) annotated more than 120 000 lncRNAs within 37 plant species and six *algae* (Gallart et al., 2016).

Even though lncRNAs share common features with coding RNAs, such as addition of 5' cap and poly(A) tail, base editing and splicing to give rise to a mature RNA, they also have distinct characteristics (**Figure 5**). Indeed, they tend to be shorter in length, contain less exons and produce a lower number of isoforms as compared to coding RNAs (Golicz et al., 2018; Sarropoulos et al., 2019; Zhao et al., 2020a). Moreover, they are often less abundant than coding RNAs and frequently retained in the nucleus whereas the coding RNA are rapidly translocated to the cytoplasm to be translated. In mammals, increasing studies shed light on subtle differences between mRNA and lncRNAs at the transcriptional level. For example, Pol II

pausing is less frequent in lncRNAs promoter compared to promoters of coding RNAs and could explain the imprecise transcription of some lncRNAs (Schlackow et al., 2017a). Secondly, promoters of lncRNAs tend to have less TF binding sites and fewer histone modifications that could explain their lower abundance as compared to coding transcripts (Mattioli et al., 2019) (**Figure 5**). Indeed, a massively parallel reporter assay, which compared the activity of 2 078 coding and lncRNAs genes promoter, shows that coding gene promoters have higher activity than lncRNA promoters (Mattioli et al., 2019). The low stability of lncRNA transcript could also participate in their poor abundance. Indeed, a study using metabolic labeling (4-thiouracil pulse chase) shows that lncRNAs are 10-fold less stable than coding RNAs (Mukherjee et al., 2017a). Conversely, two studies using ActD to stop transcription shows that their lifespan is similar to coding RNAs (Clark et al., 2012; Melé et al., 2017) hinting that more investigation is needed to understand why lncRNA are less abundant than coding mRNAs (**Figure 5**).

Taking together, lncRNAs constitute an important part of the transcriptome sharing commonality with coding RNAs such as processing and splicing but also unique particularities such as lower abundance, higher nuclear localization, shorter length and higher expression specificity.

1.2.2 Are lncRNAs really non-coding?

As lncRNA molecules are more than 200nt in length, it is likely that they can contain an ORF that could be recognized and decoded by ribosomal units (Li and Liu, 2019), even though classification generally states that lncRNA have no discernable coding potential (Jin et al., 2013; Jégu et al., 2015). More importantly, a long transcript is considered as lncRNA if it is biologically functional in its RNA state, even if it produces some peptides (Ariel et al., 2015). Conversely, some transcripts can have a function as RNA and as protein, making their classification more difficult (Crespi et al., 1994). For example, the *ENOD40* transcript is involved in nucleocytoplasmic trafficking of MtRBP1, but can also generate small peptides participating in root symbiotic nodule organogenesis (Bardou et al., 2011). Similarly, *Steroid receptor RNA activator (SRA)*, regulates steroid receptor-dependent gene expression in its RNA state but can also produce a peptide that modulates the transcriptional activity of *SRA1* gene within peptide

state (Emberley et al., 2003; Hubé et al., 2011). Interestingly, the status as lncRNA or peptide of *SRA* gene depends on the isoform which is produced, strengthening the relevance of AS for the production of bifunctional RNA molecules (Colley and Leedman, 2011). Furthermore, even if the RNA molecule contains an ORF, it does not ensure its ability to produce a protein. In addition to its sequence, the secondary structure harbored by the RNA can modulate its ability to interact with the ribosomes and yield translation products (Xie and Chen, 2017). Indeed, RNA secondary structure can influence the splicing process leading to alternatively spliced mRNA that will have less affinity with ribosomes or lacked a start/stop codon, avoiding their translation into protein (Buratti and Baralle, 2004; Foley et al., 2017).

More generally, increasing evidence from bioinformatic analyses and ribosome profiling, shows that some lncRNA associates with ribosomes strengthening their putative coding abilities (Kageyama et al., 2011; Nam et al., 2016; Yeasmin et al., 2018). Despite the thousands of small ORF discovered in human lncRNAs, very few seem to produce functional peptides such as *MLN* (Anderson et al., 2015) and *HOXB-AS3* (Huang et al., 2017). As the proportion of coding lncRNA varies greatly between ribosomal-profiling studies (Guttman et al., 2013; Ingolia et al., 2014), the Mass spectrometry (MS) emerged as a complementary method to decipher a more precise coding ability of lncRNAs. Nonetheless, MS is more accurate but less sensitive than the ribosomal-profiling studies, leading to a very small proportion of coding-lncRNAs detected through MS approaches only (Verheggen et al., 2017). Thus, the implication of peptidomic approaches together with enrichment protocols may overcome the weak sensitivity of MS (Mustafa et al., 2015).

Taken together, the frontier between coding and non-coding is subtle for the lncRNAs molecule and also dual lncRNA/peptide genes exist which can play different roles.

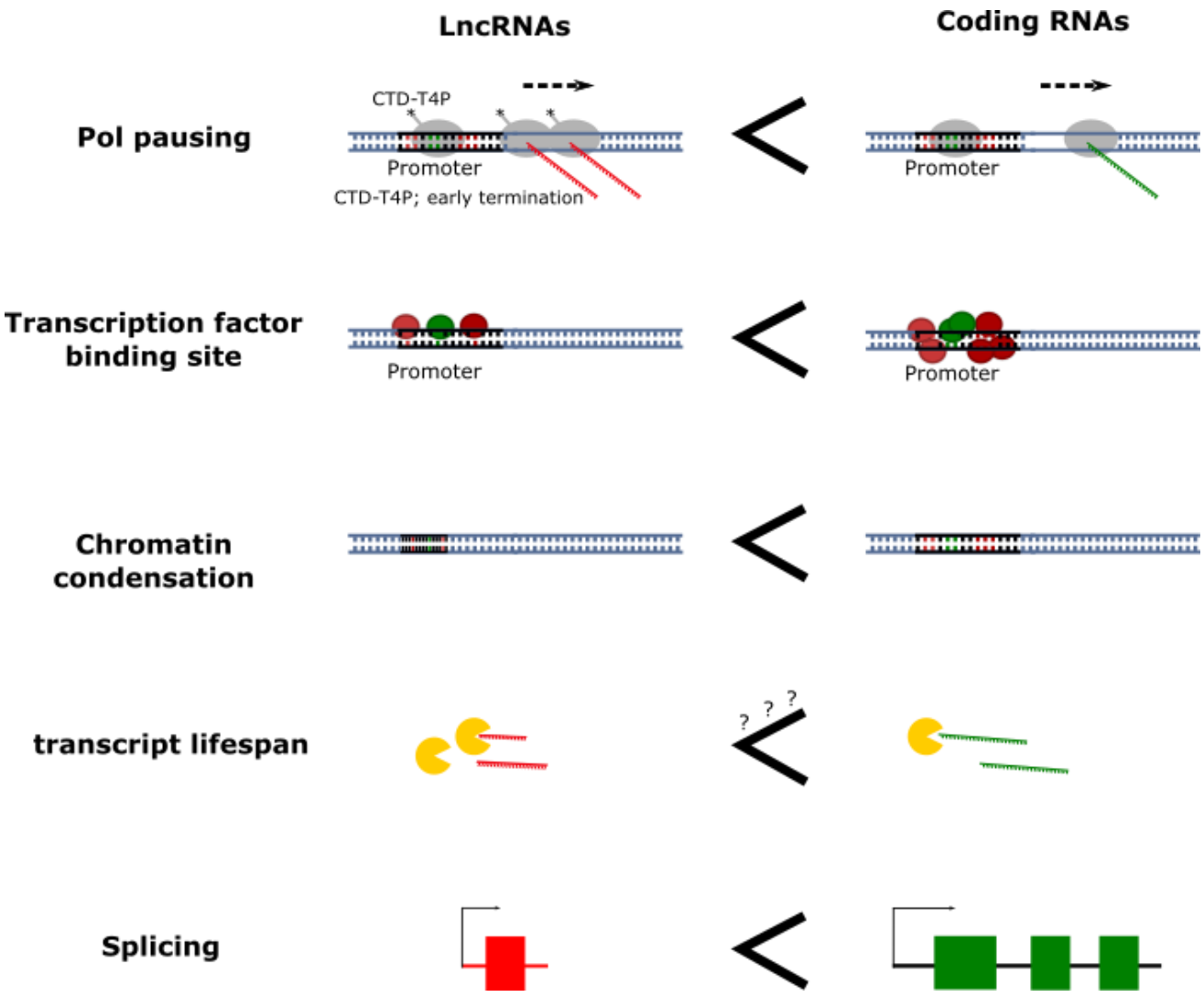


Figure 5. LncRNA specificities compared to coding mRNAs (adapted from Rinn et al., 2020). LncRNAs promoter contains more polymerase C-terminal domain modifications (i.e. threonine 4 phosphorylation (CTD-T4P)) and less TFs binding sites, decreasing the Pol II pausing and transcripts abundance, respectively. Also, the chromatin from lncRNA promoters is more frequently condensed. It is likely that lncRNA half-life is shorter than that of coding mRNAs. Finally, lncRNA are less frequently spliced than coding mRNAs.

1.2.3 Conservation of lncRNAs

The increasing availability of whole genome sequences from different species allows studying the evolutionary relationships between genes (Koonin, 2005). Interestingly, some genomic regions are highly conserved between all living organisms, such as the ribosomal related genes and the homeobox-genes. On the other hand, many genes, mainly non-coding, or intergenic regions are lowly conserved between species and even specific to some clades or individuals (Bürglin and Affolter, 2016). Thanks to the human genome project, it has been observed that structural genetic variants are less frequent in Coding DNA Sequences (CDSs) and introns as compared to other DNA regions such as non-coding genes and intergenic regions (Sudmant et al., 2015). Similarly, in *Arabidopsis thaliana*, the 1001 genome project shows that coding regions accumulate less Single Nucleotide Polymorphism (SNP) than non-coding or intergenic regions (1001 genome consortium, Alonso-Blanco et al., 2016). Likewise, in 66 rice accessions, less than 5% of SNPs and Insertion/deletions (indels) were located in coding regions while representing around 10% of the genome (Zhao et al., 2018b; Sun et al., 2019). Nevertheless, the promoters of non-coding and coding genes are similarly conserved (Ponjavic et al., 2007; Melé et al., 2017). It has been estimated that only less than 2% of the *Arabidopsis thaliana* lncRNAs are conserved at the sequence level across the plant kingdom (Liu et al., 2012a). Concomitantly, it has been observed that highly expressed genes tend to be more conserved than lowly or specifically expressed genes, such as lncRNAs (Contreras-Moreira et al., 2017) (**Figure 5**). Logically, the 18% of sequence-level conserved *Brassicaceae* lncRNA present a higher level of expression as compared to the non-conserved lncRNA (Nelson et al., 2016). Nevertheless, genome analysis of five monocotyledon and five dicotyledon species shows that lncRNA conservation remains high within the same species but strongly decreases at the inter-species level (Deng et al., 2018a). Even though, in all organisms, lncRNAs present similar characteristics, such as a lower size, less exon and number of isoforms per gene, they are less conserved than coding genes at the sequence level. Thus, it is tempting to assume that the non-conserved genome reflects the specific adaptation of an accession or an individual to its environment (Vernikos et al., 2015). It is noteworthy that the sequence of the lncRNA gene may not be conserved although their relative position (synteny) within the genome may be (Mohammadin et al., 2015; Jha et al., 2020), suggesting that the interaction of the lncRNA with its syntenic neighboring coding genes may have biological relevance.

2. Regulation of gene transcription

Gene activity is a dynamic process non-uniformly distributed along the genome and directly dependent on the proteins interacting with the DNA. More precisely, ~147bp of DNA molecule is wrapped along histone octamer, which together form the so-called chromatin critically important for the gene activity (Yamaguchi, 2021). Broadly, the genome can be divided in two different states: heterochromatin, which contains un-transcribed genes and euchromatin, which generally contains actively transcribed genes (Rosa and Shaw, 2013).

The heterochromatin is characterized by a condensed packaged structure due to nucleosome colocalization, limiting gene transcription (Kireeva et al., 2002). It is generally rich in repeated sequences and TE and mainly localized in the centromeric and telomeric regions of the chromosome (Lippman et al., 2004). Whereas, euchromatin is located in chromosomal arms and mostly contains actively transcribed genes (de Nooijer et al., 2009). Euchromatin is characterized by an important plasticity of chromatin condensation, which is controlled by various mechanisms, or "dimensions", including the promoter and cis regulatory element sequences, the epigenetic modifications and the chromatin conformation. Interestingly, part of the heterochromatin, named as facultative, can become euchromatin through particular stimuli, triggering a chromatin decondensation and subsequent gene activity (Grewal and Jia, 2007). In this second part, I will describe the dimension affecting gene activity at the transcriptional level including DNA cis-regulatory sequences (first dimension), epigenetic modifications (second dimension) and genome conformation (third dimension).

2.1 First dimension: Cis-regulatory motifs

In eukaryotes, DNA binding proteins, named Transcription Factors (TFs), bind the 5' outside of the genes called promoters to facilitate the recruitment of the RNA pol complex and the subsequent gene transcription (Allen and Taatjes, 2015). More precisely, they recruit co-transcriptional factors that together change the chromatin condensation facilitating gene transcription and participate in the formation of the Pre-Initiation Complex (PIC), critical for the initiation of gene transcription (Venters and Pugh, 2009) (**Figure 6**). For example, pioneer factors are a particular class of TFs able to bind to condensate chromatin and loosen it, facilitating the interaction of the DNA with other regulatory TFs (Magnani et al., 2011).

TFs bind to the DNA through the recognition of specific DNA sequences named cis-regulatory elements (CREs) (Wilkinson et al., 2017). As compared to prokaryotic TFs that recognize long DNA sequences, eukaryotic TFs recognize short DNA motifs of ~10nt (Stewart and Plotkin, 2012). Thus, genes which participate in the same cellular processes will often share these short similar sequences to be bound together by the same TF (Haberle and Stark, 2018). The specificity of CREs recognition is provided by the DNA binding domain that greatly varies between TF families and can be encoded as a homeodomain, helix-turn-helix and high-mobility group box or other motif, increasing the diversity of recognizable CREs (Mitsis et al., 2020). As gene transcription constitutes the most important and basal mechanism for cell life, it is noteworthy that TFs and CREs evolved in a tightly manner. Notably, bioinformatic evolutionary analysis shows a correlated evolution between TFs and CREs within eukaryotes (Yang et al., 2011). However, SNPs or indels occurring in these CREs could drastically change the responsiveness of a gene to a particular stimulus, by modifying the ability of the TFs to interact with the promoter, likely participating in the plant adaptation to its local environmental stimuli.

Even though TFs are well-known to positively influence gene transcription, through the recruitment of coactivator protein increasing the transcriptional rate or the chromatin decondensation (Cramer, 2019). TFs can also inhibit transcription when recruiting co-transcriptional repressors which increase the condensation of the chromatin, disturbing the ability of RNA Pol to progress along the gene (Cramer, 2019). For example, the ERF-associated amphiphilic repression (EAR) is a plant repressive TFs (Ohta et al., 2001), which when associated with transcriptional activators, convert them into repressors (Hiratsu et al., 2003; Hiratsu et al., 2004). Interestingly, this motif is not specific to a suburb of TFs, as it exists in *AUX-IAA*, *BZR1*, class II *ERF* and C_2H_2 zinc-finger containing proteins, such as *SUPERMAN* (Ohta et al., 2001; Hiratsu et al., 2002; Tiwari et al., 2004; He et al., 2005). More generally, in *Arabidopsis* it has been estimated that 10% of all the TFs act as transcriptional repressors (Ikeda and Ohme-Takagi, 2009).

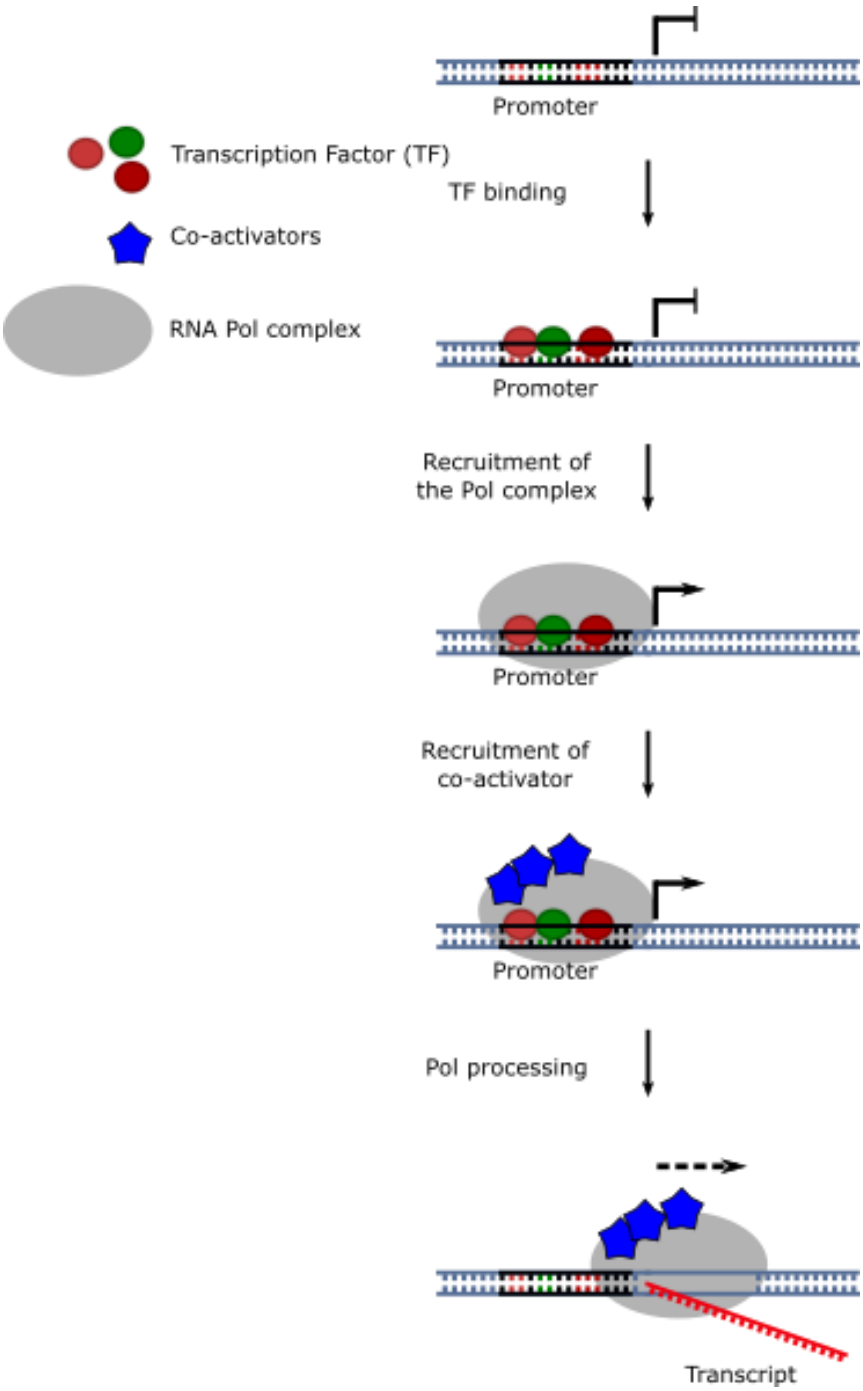


Figure 6. Control of gene expression by promoter' cis-regulatory motif. The promoter region is located upstream of genes and harbor multiple cis-regulatory motifs, recognizable by Transcription Factors (TFs). TFs bind to the promoter region and facilitate the recruitment of the Pol II complex. This TF/Pol II complex associates with co-activators to trigger gene transcription.

As gene proximal promoters estimated at around ~500bp in length (Lenhard et al., 2012), it is likely that multiple TFs cooperate or compete to modulate the transcriptional status of the downstream gene (Morgunova and Taipale, 2017). For example, a large proportion of eukaryotic TFs are functional in a dimer state, interacting either with an identical TF or with another TF from a closely related family (Morgunova and Taipale, 2017). Interestingly, TF dimerization is significantly enriched in complex eukaryotic organisms potentially providing an additional regulatory layer on the control of gene expression in these organisms (Amoutzias et al., 2008). Strikingly, TFs cooperation can also involve indirect interactions, where multiple TFs collaboratively compete with the same histone to access the DNA (Long et al., 2016). In addition to associates within complexes, some TFs are subjected to post-translational modifications affecting their localization, stability or ability to interact or dimerize (Everett et al., 2010). For example, phosphorylation of the TF CCAAT/enhancer-binding protein beta (CEBPB) during mammal pregnancy is necessary for its activation (Lynch et al., 2011). Similarly, plethora of plant bZIP TFs are subjected to phosphorylation upon activation, such as the ABI5–ABF–AREB implicated in the ABA signaling pathway (Schütze et al., 2008). Also, the phosphorylation of the rice bZIP TF TRAB1 at the Ser102 residue is necessary for its activity, ensuring a proper maturation and dormancy of plant embryos (Hobo et al., 1999; Kagaya et al., 2002).

With the development of highly sensitive TF profiling methods, such as Chip-seq, it is now well established that gene expression is not only controlled by interaction between TFs and promoter but also through the binding of TFs to cis-regulatory sequences located far away from the gene they regulate, that were named enhancer (Arnone and Davidson, 1997; Johnson et al., 2007). Enhancers are intergenic regulatory sequences bound by TFs that interact with other promoters to strengthen their activity. Enhancer-mediated regulation of gene expression is detailed further in section 2.3.3.2.

Collectively, the physical interaction between TFs and their corresponding genomic cis-regulatory sequences, promoter and enhancers, constitute a highly complex layer of regulation of gene activity within the cells.

2.2 Second dimension: Epigenetic marks

The DNA molecule is very long and could fit inside the nucleus only thanks to an organized packing that is the result of a tight organization of a complex comprising the DNA molecule, RNA and proteins, altogether forming the chromatin. This DNA packing is controlled by chromatin modifications that alters locally its level of condensation without altering the sequence of the DNA (Prakash and Fournier, 2018). Namely, epigenetic study the changes of gene function that are mitotically or meiotically heritable without involving a shift in DNA sequence (Dupont et al., 2009). Among these epigenetic modifications, DNA methylation is often associated with gene silencing (Matzke and Mosher, 2014), whereas histone modifications can be associated with active or silenced gene activity (Ueda and Seki, 2020).

2.2.1 DNA methylation

DNA methylation occurs on cytosine bases and corresponds to the addition of a methyl group to form 5-methylcytosine. In plants, DNA methylation can occur on different context: CG, CHG and CHH (where H represents A, T or C) (Lister et al., 2008), even though the majority of DNA methylation is observed in a non-CG context (Kenchanmane Raju et al., 2019). In agreement, loss of CG methylation in *met1* mutant triggers strong developmental defects (Kankel et al., 2003; Saze et al., 2003; Reinders et al., 2009), whereas disruption of the non-CG methylation does not result in strong physiological phenotype (Chan et al., 2006). In *Arabidopsis*, all methylation contexts are significantly enriched within repeated sequences, such as TE. Nevertheless, some genes are methylated through CG or CHH contexts (Zhang et al., 2006). DNA methylation strongly correlates with low transcriptional activity. Concomitantly, in the majority of eukaryotic cells methylation is highly concentrated in the heterochromatin (Lister et al., 2008). Strikingly, whereas DNA methylation within gene promoter correlates with a low transcriptional activity, methylation within the gene body may be involved in the transcriptional activation (Elhamamsy, 2016). For example, the active X chromosome within female mammals is significantly richer in methylation within the gene body than the inactive X chromosome (Hellman and Chess, 2007). Nonetheless in plants DNA methylation within the gene body may disturb Pol II elongation inhibiting gene transcription (Zilberman et al., 2007).

In plants siRNA can direct de-novo methylation through RdDM, described in section 1.1.2.1. Interestingly, phasiRNA, described in 1.1.2.3, can also be recruited into the RdDM pathway, through *AGO4* or *AGO6*, to mediate de-novo methylation (Matzke and Mosher, 2014). Despite that RdDM is often associated with CHH context it can methylate C in all contexts (Kenchanmane Raju et al., 2019). Although RdDM predominantly occurs in TE rich regions, mutants impaired in the RdDM pathway doesn't present a strong increase of TEs activity suggesting RdDM is not the only factor silencing TE (Wendte and Schmitz, 2018). Nevertheless, under heat stress in maize the RdDM prevents the TE activity, resulting in heritable transposition (Guo et al., 2021).

RdDM and more generally, DNA methylation are implicated in a wide-range of biological processes in plants, from development to stress signaling pathway. For example, RdDM demonstrated its importance for proper plant reproduction in the model plant *Arabidopsis*, tomato, rice and *Brassica rapa* (Chow et al., 2020). Heat stress decreases the level of DNA methylation in multiple plant species, such as *Gossypium hirsutum* (Ma et al., 2018), *Glycine max L.* (Hossain et al., 2017)*, Brassica napus cv. Topas* (Li et al., 2016b)*, Oryza sativa* (Folsom et al., 2014) and *Arabidopsis* (Naydenov et al., 2015). Concomitantly, *Arabidopsis* RdDM defective mutants are hypersensitive to heat stress (Popova et al., 2013). Similarly, DNA methylation is reduced under cold stress within *Arabidopsis*, *Castanea sativa Mill.*, *Populus tremula*, and *Cucumis sativus* L. (Conde et al., 2017; Lai et al., 2017; Xie et al., 2019). Interestingly, chemical mediated inhibition of DNA methylation increased freezing tolerance in *Arabidopsis* plants (Xie et al., 2019). Finally, various mutants impaired in the RdDM pathway (*nrpd2a-1, dcl3–1, ago4–3, clsy1–1* and *drd1–6*) present a constitutive expression of defense related genes, among them the *PATHOGENESIS-RELATED 1 (PR1),* important for the systemic acquired resistance, illustrating the importance of RdDM for the plant biotic stress response (McCue et al., 2012).

Altogether, DNA methylation constitutes an important additional layer for the control of gene expression within eukaryotic cells. In humans more than 70% of CG-rich regions are methylated (Strichman-Almashanu et al., 2002) and are directly implicated in many diseases such as cancer development (Wang and Lei, 2018). Similarly in *Arabidopsis*, 14% of all the cytosines are methylated (Zhang et al., 2006) and mutant plants impaired in the DNA methylation pathway harbor strong developmental defects (Zhang et al., 2018a).

2.2.2 Chemical modifications of histones

DNA folding around histone proteins is made by two turns of DNA around an octamer of two subunits of each of the core subunits of histones H2A, H2B, H3, and H4 forming the so-called nucleosomes, also corresponding to the unit of the chromatin (Tatchell and Van Holde, 1977). Post-translational modifications of histones influence their co-localization and the way DNA is wrapped around them (Tatchell and Van Holde, 1977) and directly influence the condensation level of the chromatin. Rationally, histone co-localization positively correlates with high chromatin condensation and low transcriptional activity (Xu et al., 2017a). Several protein types are involved in these histone modifications. Writer enzymes mediate chemical modifications of histones, such as acetylation, methylation, phosphorylation or ubiquitination; whereas eraser enzymes remove them (Xu et al., 2017a). Reader proteins physically interact with histone epigenetic marks to strengthen their stability (Liu et al., 2010a). Among them, the Polycomb group (PcG) mediates gene repression mainly through trimethylation of the Lysine 27 on Histone 3 (H3K27me3). PcG proteins encompass two complexes: Polycomb Repressive Complex1 (PRC1) and PRC2, containing both writers, readers and erasers proteins, conserved in plants and animals **(Figure 7)**. PRC2 is involved in the spreading and de-novo trimethylation of H3K27 (Kim, 2020). Whereas, the PRC1 cannot perform de-novo trimethylation but can facilitate the spreading of H3K27me3 (Veluchamy et al., 2016) and maintain its occupancy during cell replication (Derkacheva et al., 2013). Additionally, PRC1 is involved in ubiquitination of H2AK119 and H2AK121, also participating in the chromatin condensation (Kim, 2020). Regarding the vast diversity of chromatin modifying protein, other than PcG, affecting the chemical state of histone, I will focus on the PcG and the H3K27me3 epigenetic mark, a complex widely studied in my thesis.

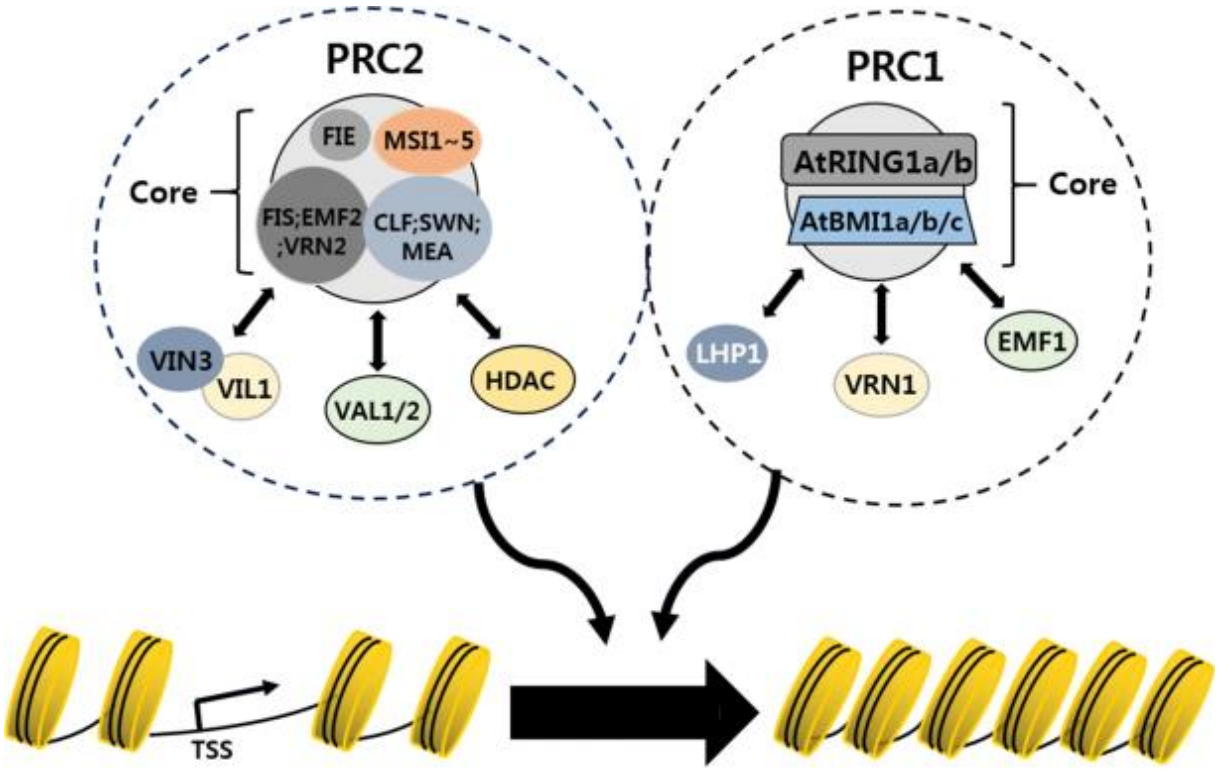


Figure 7. Polycomb-mediated chromatin condensation (Adapted from Kim, 2020). The polycomb complexes, PRC2 and PRC1, harbor several core components, highly conserved within eukaryotes. Also, both sub-complexes are associated with several other regulators. On one hand, PRC2 core associates with two PHD-finger domain proteins (VIN3 and VIL1), two B3 domain proteins (VAL1 and VAL2), and some HDAC proteins, among which HDA19. On the other hand, PRC1 core associates with LHP1, VRN1 and EMF1. PRCs complexes interact with histone proteins and mediates trimethylation of H3K27 and ubiquitination of H2AK119/121 maintaining the chromatin in a condensed and transcriptionally silenced state.

The histone methyltransferase activity of PRC2 mediates H3K27me3 deposition within thousands of genomic loci encompassing many developmental and stress-related genes (Turck et al., 2007; Zhang et al., 2007; Lafos et al., 2011). Specifically, the FIE subunit binds to the H3 tail (Kim and Sung, 2014), and the MSI1-5 subunits directly binds to the nucleosome, both facilitating the PRC2-chromatin interaction. Finally, the PRC2-core subunits FIS, EMF2, VRN2, CLF, SWN and MEA, with the help of additional cofactors, mediates trimethylation of H3K27 (**Figure 7**). Interestingly, the trimethylation of H3K27 solely is not sufficient to maintain the chromatin in a condensed structure. Indeed, H3K27me3 serves as a signal to recruit the PRC1, which prevent the recruitment of remodeling factors, such as SWI/SNF (Shao et al., 1999; Francis and Kingston, 2001), through ubiquitination of H2A (Cao et al., 2002; Eskeland et al., 2010; Wani et al., 2016; Yin et al., 2021), maintaining the chromatin in a silenced state (Cao et al., 2002). In agreement, co-localization of PRC2 and PRC1 in many H3K27me3-rich sites is observed from animals to plants (Tolhuis et al., 2006; Turck et al., 2007; Ku et al., 2008; Kim et al., 2012; Zhou et al., 2017; Yin et al., 2021). Nevertheless, some H2AUb are detected in non-H3K27me3 loci, suggesting PRC1-H2Aub may act independently of H3K27me3 deposition (Kim et al., 2012). Adversely, the non-PcG proteins eukaryotic-conserved Jumonji C (JmjC) (Crevillén, 2020), removed methylation on H3K27 through its histone demethylase activity (Klose et al., 2006; Black et al., 2012), thus participating in the H3K27me3 equilibrium (Shen et al., 2021).

Arabidopsis mutants affected in *PRC2* show important developmental defects including, loss of vegetative growth, early flowering, leaf serration, abnormal flower architecture and sterility. Concomitantly, genes involved in developmental transition are strongly affected in PRC2 mutants. For example, during germination, *EMF1* and *FIE*, part of PRC2, represses *ABSCISIC ACID INSENSITIVE 3* (*ABI3*) and *LEAFY COTYLEDON 2* (*LEC2*) which are involved in the maturation of embryo (Ueda and Seki, 2020). In agreement, PRC2 mutants exhibit lipids accumulation in root tips, characteristic of an abnormal seed to seedling transition (Suzuki et al., 2007). Consistently, PRCs mediates the transition from seed to seedling through the targeting of genes positively regulated by ABA signaling and negatively regulated by GA, two hormones implicated in seed development and germination (Bouyer et al., 2011; Kim et al., 2012), respectively (Finkelstein et al., 2008). Later in the plant life, vegetative to flowering transition is also tightly regulated by PRC2. Indeed, key regulator of floral transition such as: *floral MADS box genes AGAMOUS* (*AG*), *APETALA 1* (*AP1*), *AP3*, *SEPATALLA 1* (*SEP1*), *PETALLOSS*

(*PTL*), and *FRUITFUL/AGAMOUS-LIKE8* (*FUL/AGL8*) are bound by PRC2 (Ueda and Seki, 2020). These floral-related genes have been found highly up-regulated in PRC1 and PRC2 mutants, in agreement with the rapid transition from vegetative to flowering in these mutants (Pu and Sung, 2015). In addition, the flowering genes *FLC* and *LFY* are also targeted by the PRCs (Kim et al., 2012). Indeed, when temperature decreases, the *FLC* locus is subjected to a PRCs-mediated dynamic epigenetic switch from H3K4me3 to H3K27me3 and H3K9me3 which decreases the transcription of *FLC*. Moreover, major genes involved in cell differentiation, including *Shoot Meristemless* (*STM*), *Wuschel* (*WUS*), *Knotted 1-Like Homeobox* (*KNATs*), *Cup-Shaped Cotyledon 2* (*CUC2*), *Wuschel Re- lated Homeobox 1* (*WOX1*), *Plethora 1* (*PLT1*), and *PLT2*, are targeted by PRCs and could explain the abnormal leaf and flower shapes (Kim et al., 2012; Deng et al., 2013) observed in PRCs mutants.

For the plant sessile organism, the epigenetic-mediated regulation of gene expression is a very convenient mechanism to quickly adapt their transcriptome to the current environmental condition. For example, the MSI1 PRC2 subunits fine-tune ABA signaling, modulating the plant sensitivity to drought and salt stresses (Alexandre et al., 2009; Mehdi et al., 2016). Concomitantly, the level of expression of the barley *CLF* homolog *HvE(Z)* (*Enhancer of zeste*) significantly increased in response to ABA application (Kapazoglou et al., 2010). In response to drought stress, the PRC1 subunits BMI1a/DRIP2 and BMI1b/DRIP1, targets the TF DREB2A to proteasome-mediated degradation, negatively influencing the plant drought-stress response (Qin et al., 2008). Also, loss of *CLF* function significantly reduced the plant tolerance to drought stress (Liu et al., 2014b). Interestingly, CLF also interacts with the plant-specific coiled-coil protein BLISTER (BLI) promoting the plant cold stress resistance (Schatlowski et al., 2010; Purdy et al., 2011). Following, EMF1 and EMF2 represses the cold stress-induced gene *COR15A* (Kim et al., 2010). More precisely, the decrease in temperature reduces H3K27me3 occupancy at *COR15A* and *GALACTINOL SYNTHASE3* (*GOLS3*) genes, logically increasing their transcriptional activity (Kwon et al., 2009). Interestingly, the return to normal temperature does not recover the initial level of H3K27me3 even though *COR15A* and *GOLS3* gene expression decreases (Kwon et al., 2009).

Together, PcG proteins constitute key regulators of gene expression, through modulation of H3K27me3 and H2AUb occupancy, influencing the plant stress signaling pathways and critical developmental processes such as seed development and flowering (Kleinmanns and Schubert, 2014).

2.3 Third dimension: chromatin conformation

The DNA packing within the nucleus is not randomly organized (Tiang et al., 2012). Even more, spatial organization of the chromatin participates in the regulation of gene expression (Diesinger et al., 2010). Microscopic studies first addressed chromosome architecture within the nucleus: electron microscopy was used to assess the basic profile of nuclear architecture, whereas fluorescence microscopy shed light on the nuclear compartmentalization and co-localization of chromosomes (Solovei et al., 2002; Dehghani et al., 2005; Belton et al., 2012). Later, 3C-related techniques, based on digestion and ligation of cross-linked DNA, showed the importance of chromosome architecture for the regulation of gene activity (Dekker et al., 2002; Belton et al., 2012).

2.3.1 General configuration of genome within the nucleus

Eukaryotic genomes, from animals to plants, occupy well defined territories within the nucleus (Schubert et al., 2012). As the general genome characteristics, such as genome length, TE abundance and epigenetic marks, the genome spatial organization also vary greatly between organisms and can be classified into three major conformations: Rabl, Rosette, and Bouquet structure (**Figure 8**).

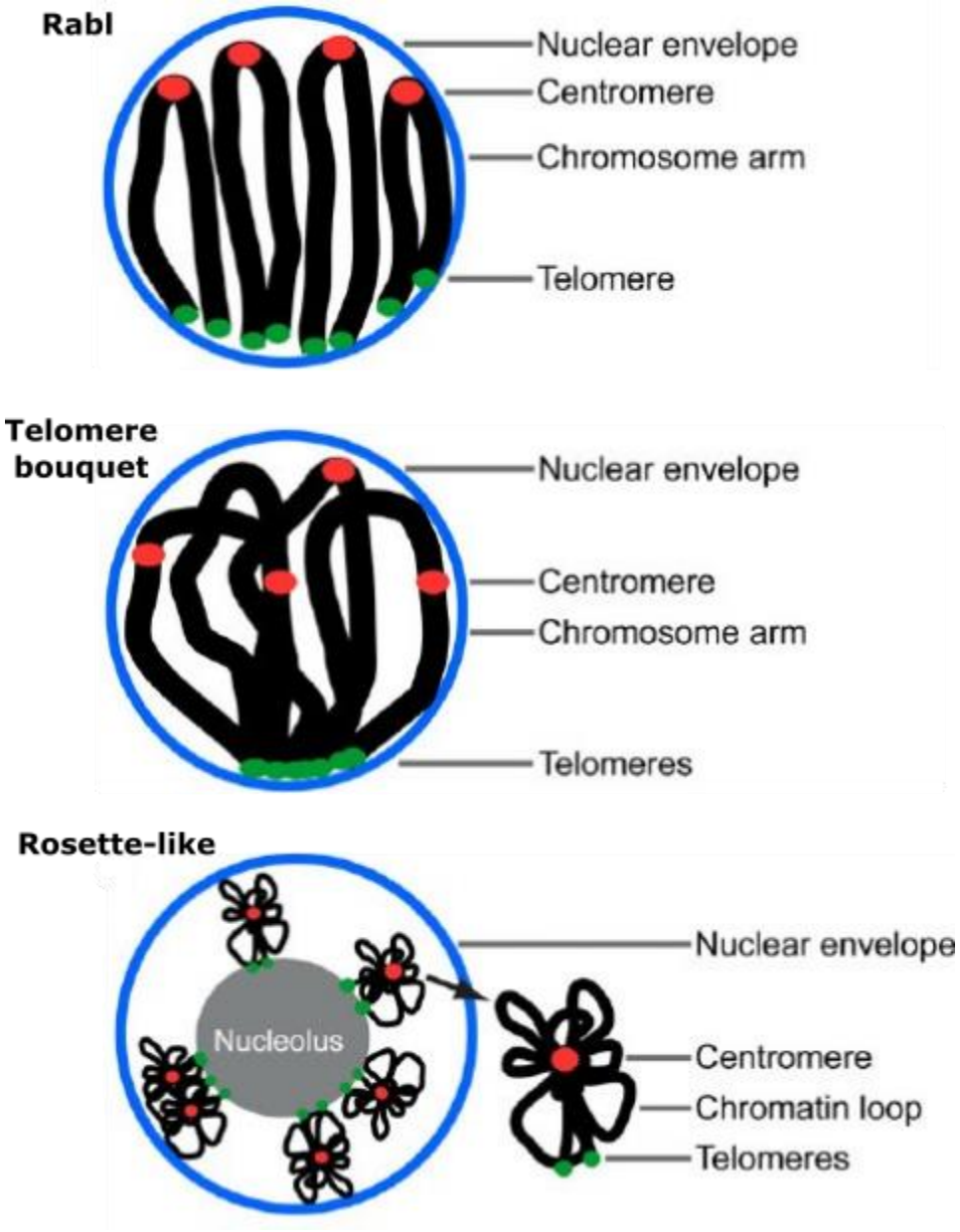


Figure 8. Chromosome arrangements within the nucleus. (Adapted from Tiang et al., 2012) The Rabl genome configuration is found in many large plant genome species and is characterized by a spatial separation of telomeres and centromeres. Interestingly, the Bouquet genome conformation is also found in large plant genomes, such as maize. In this configuration, the telomeres interact with each other close to the nuclear periphery whereas the centromeres and chromosome arms are located within the nucleoplasm region. Finally, small plant genomes, such as *Arabidopsis*, showed a Rosette-like configuration, in which the heterochromatin (centromeres and telomeres) tends to interact with itself whereas the euchromatin (chromosome arms) does not.

For example, plants with large genomes, such as wheat (Abranches et al., 1998) and barley (Noguchi and Fukui, 1995), preferentially harbored the Rabl configuration (**Figure 8; Rabl**). In this configuration the chromosomes bend in their centromere, mainly during interphase and meiotic prophase, leading to spatial separation of centromere and telomere (Rosa and Shaw, 2013). Consequently, telomeres clustered together in the nuclear periphery. This configuration is also found in the animal kingdom such as mice (Van Driel and Fransz, 2004) and yeast (Mizuguchi et al., 2015). Maize genome presents a Bouquet conformation where telomeres are close to each other's and localized in the nuclear periphery in opposition to chromosome arms and centromeres located in the nucleoplasm region (Bass et al., 2000; Cowan et al., 2001) (**Figure 8; Telomere bouquet**). The strong similarities between the Rabl and Bouquet configurations make this classification controversial among the scientific community (Rodriguez-Granados et al., 2016). Finally, plants with small genomes, such as *Arabidopsis*, present a Rosette-like configuration (**Figure 8; Rosette-like**). In this state, heterochromatin is clustered together and forms chromocenters from which euchromatin evades (Fransz et al., 2002). The important packing of the genome within the nucleus ultimately leads to intra and inter chromosomal interactions. In *Arabidopsis*, intra-chromosomal interactions have been detected in centromeric and pericentromeric segments, in agreement with the Rosette-like conformation, but also to a lesser extent within the chromosomal arms. Whereas, inter-chromosomal interactions are mainly detected within the pericentromeric and telomeric regions (Moissiard et al., 2012; Grob et al., 2013; Feng et al., 2014; Grob et al., 2014; Liu et al., 2016). Interestingly regions close to the centromere are less subjected to inter-chromosomal interactions as compared to segments far away from it, i.e. telomere. This indicates that chromosome arms are potentially more flexible, in terms of spatial localization, than the anchor-like centromere (Grob et al., 2013).

2.3.2 Long range interactions

In animals and large plants genome long-range interactions relate to the AB compartments, where A corresponds to active transcription and B corresponds to gene silencing (Lieberman-Aiden et al., 2009; Dong et al., 2017). These regions are dynamically regulated under stress or tissue specificity in agreement with gene expression changes (Roy et al., 2015).

Curiously, such genome compartmentalization is absent from the small genome of *Arabidopsis*. Instead, within chromosome arms, some heterochromatic regions interact with each other forming heterochromatic islands (IHIs). These regions are marked with the silenced H3K9me2 epigenetic mark and enriched in TE. IHIs includes few transcribed genes, most of them implicated in transposon silencing (Grob et al., 2014). In harmony, IHIs also interact with telomeric regions. These interactions are altered in RdDM-related mutants such as *morc6*, *met1*, and *ddm1* (Feng et al., 2014; Liu and Weigel, 2015), strengthening the importance of DNA methylation in IHI-telomeres interactions and more generally the maintenance of a proper nuclear architecture. Also, the LHP1 protein from the PRC1, in addition to its role as recruiter of PRC2 and spreader of H3K27me3 marks, is involved in shaping the conformation of the chromatin. Indeed, genome-wide conformation analysis of the *lhp1* mutant shed light on important chromatin conformation changes both at the long and local-range scale. This modification of chromatin conformation together with the impairment of H3K27me3 deposition, can explain changes of expression of multiple genes in the *lhp1* mutant (Veluchamy et al., 2016).

2.3.3 Short-range interaction

The development of high-resolution chromatin conformation methods in both animals and plants, allows the identification of multiple chromatin loops at the local or gene scale. These interactions can regulate gene expression in a wide variety of way involving looping within a gene, two genes or between a gene and a cis-regulatory intergenic region (Miele and Dekker, 2008; Cavalli and Misteli, 2013; Gibcus and Dekker, 2013). Interestingly, in animals and large plant genomes, these short-range interactions are not randomly positioned within the genomes. Indeed, some genomic regions from the same chromosome present high interaction frequency. These regions are named Topologically Associated Domains (TADs) in which the DNA sequence localized within the TAD more frequently interacts with each other than with the genomic region outside of the TAD. Thus, it has been proposed that these topological units restrict the chromatin looping interactions between promoter and enhancers with genes (Rodriguez-Granados et al., 2016). Nevertheless, the small *Arabidopsis* genome does not contain such well-defined independently regulated regions (Gibcus and Dekker, 2013; Liu et al., 2016).

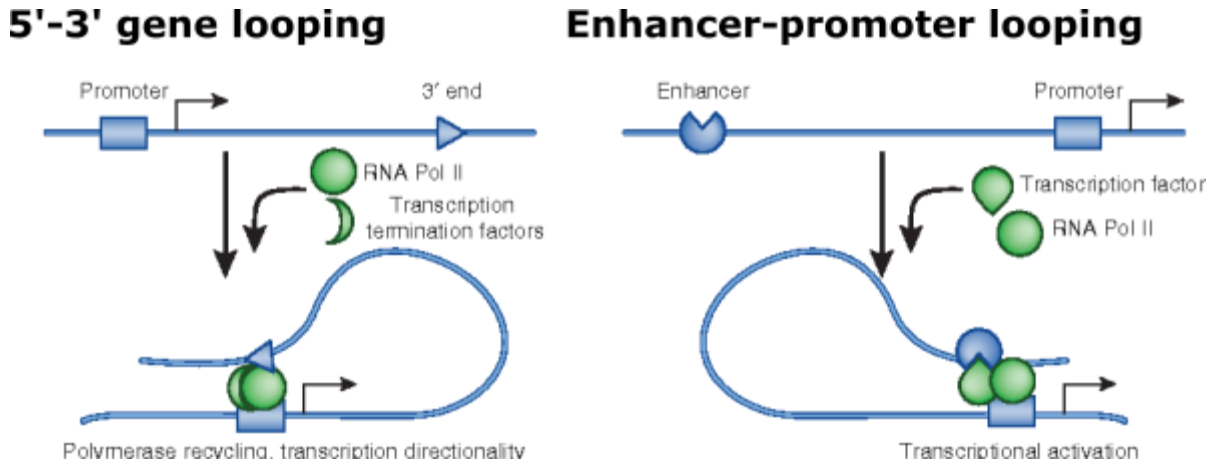


Figure 9. Short range interactions. (Adapted from Cavalli and Misteli, 2013). The intragenic chromatin looping (5'-3- gene looping) puts in close proximity the TSS and the TES of a gene, generally facilitating the transcriptional activity through Pol II recycling. The enhancer-promoter looping joins enhancer regions with target promoters, positively affecting gene transcription at distant loci.

2.3.3.1 Looping within a gene locus

Gene looping is when DNA-DNA interaction occurs between the Transcription Start Site (TSS) and the Transcription End Site (TES) of a gene. In yeast and mammal cells these types of interaction have been linked to Pol II recycling, in which the Pol II processing is renewed several times to extend the production of RNA from the concerning gene (**Figure 9; 5'-3' gene looping**). For example, the transcription initiation factor TFIIB participates in the collocation of promoter and TES at actively transcribed genes enabling quick and efficient Pol II recycling (O'Sullivan et al., 2004; Singh and Hampsey, 2007). In *Arabidopsis*, this type of chromatin looping configuration has also been described in *IPT3* and *IPT7* genes, both involved in cytokinin biosynthesis. Interestingly, the chromatin remodeler BAF60 can disrupt the chromatin loop under certain stimuli decreasing the expression of *IPT3* and *IPT7* (Jégu et al., 2015). Likewise, the epigenetically regulated *FLC* gene also presents an additional layer of regulation involving changes of chromatin conformation. Strikingly, the *FLC* TSS-TES interactions negatively correlate with gene expression. Indeed, in this case, the disruption of the gene looping configuration facilitates the epigenetic switch which is occurring during vernalization from inactive to active transcriptional state (Crevillén et al., 2013). Similarly, a gene looping within the Arabidopsis *WUS* locus is negatively associated with gene transcription. Excluding the promoter region, this gene loop blocks the recruitment of Pol II to the core gene, limiting gene transcription (Guo et al., 2018). Lastly, when actively transcribed the flowering-repressing gene *TFL1* harbored a chromatin loop putting in close proximity its promoter with its TSS. In floral meristem the binding of AP1 together with MADS-box proteins (SOC1, AGL24, SVP, and SEP4) to the *TFL1* promoter disrupts the chromatin loop, repressing *TFL1* expression, allowing flower development (Liu et al., 2013)

2.3.3.2 Enhancer loop

The development of technologies assessing the chromatin-condensation at the genome scale allow the discovery of multiple intergenic CREs, named enhancer (Kyrchanova and Georgiev, 2021). For example, it has been estimated that the human genome contains 300 000 enhancers (Li et al., 2018a). Similarly, more than thousands of enhancers have been discovered in *Arabidopsis* (Zhu et al., 2015; Yan et al., 2019), maize (Oka et al., 2017; Ricci et al., 2019) and rice (Sun et al., 2019). Enhancers are generally 100-1000bp regions rich in TFs binding

sites. Their chromatin feature is quite similar to promoter region, where both can bind TFs which correlates with chromatin decondensation (Kyrchanova and Georgiev, 2021). Nevertheless, enhancers are not able to recruit the Pol machinery which initiates the gene transcription. Interestingly, enhancers can generate non-poly adenylated transcripts, referred as enhancers RNA, most of them unstable and staying in the nucleus (Andersson, 2015). Constitutively, or under certain stimuli, enhancer regions physically interact with gene promoters. The physical interaction between enhancer and gene promoter together with their activation (binding of TFs) positively affects gene transcription (**Figure 9; Enhancer-promoter looping**). Multiple enhancers can be clustered within the same genomic loci forming "super-enhancers" which can reach more than thousands of bp (Whyte et al., 2013). Interestingly, sub-domain of these super-enhancers can act independently or together to activate a large number of gene promoters (Hay et al., 2016; Shin et al., 2016; Huang et al., 2018).

Many enhancers are closely related to their promoter target. For example, in *Drosophila* it has been estimated that 20% of the enhancer' promoter targets are located within 100kb from the enhancer (Ghavi-Helm et al., 2014; Kvon et al., 2014). Similarly, during *Arabidopsis* floral development, enhancer activity significantly positively correlates with the expression of the closely located genes (Yan et al., 2019). Illustratively, a chromatin loop positively regulates the *VERDANDI* (*VDD*) gene involved in female gametophyte identity (Matias-Hernandez et al., 2010). Also, the floral organ identity is controlled by a dimer formed with SEEDSTICK (STK) and SEPALLATA3 (SEP3), which directly mediate the formation of two enhancer-chromatin loops directly implicated in the transcriptional activity of the female gametophyte identity gene *VERDANDI* (*VDD*) (Mendes et al., 2013). In maize, many enhancer-promoter interactions occur within long-distance (Salvi et al., 2007; Studer et al., 2011; Huang et al., 2018; Ricci et al., 2019). For example, the two maize *b1* epialleles involved in flavonoid pigmentation present a significantly different transcription rate. Notably, the promoter of the most expressed one is interacting with an enhancer located 100kb upstream from *B1* gene (Louwers et al., 2009). Taking together, the 3D nuclear organization of the chromatin constitutes, in addition to cis-regulatory sequences and epigenetic modifications, a third complex regulatory layers for the regulation of gene transcriptional activity. Notably, the epigenetic landscape and chromatin condensation is directly implicated in the 3D genome organization from the chromosome to the gene scale.

3 Regulation of gene expression by lncRNAs

In the last decade, next generation sequencing methods unveiled that the large majority of the eukaryotic genome is transcribed, whereas only a few transcripts are translated into proteins. For example, 95% of the human genome is transcribed and only 2% encodes for protein. Similarly in *Arabidopsis* 71% of the genome is transcribed, whereas less than 2% of the transcripts are coding (Lucero et al., 2020). This large diversity of non-coding transcripts includes cell housekeeping transcripts (rRNA, tRNA, snRNA and snorRNA), small regulatory transcripts and long regulatory transcripts, all of which are described in section 1.1. NcRNAs recently emerged as important regulators of gene expression acting both at the transcriptional and post-transcriptional level in multiple biological processes. This third part will present more in detail, the miscellaneous mechanisms that lncRNAs molecules use to fine-tune gene expression, and includes a revision about the importance of lncRNAs for plant root growth and development.

3.1 Modulation of the transcriptional activity by lncRNAs

Transcription is realized by a remarkable cooperation between the Pol machinery, TFs and other complexes (Woychik and Hampsey, 2002). Among them, the MEDIATOR complex mediates the communication between TFs and Pol II, influencing the transcriptional rate (Allen and Taatjes, 2015). Interestingly, in *Arabidopsis*, the lncRNA *ELF18-INDUCED LONG-NONCODING RNA1* (*ELENA1*) is able to bind to *MED19a*, a subunit of the MEDIATOR complex, and modulate its recruitment to specific promoters (Seo et al., 2017) (**Figure 10A**). Indeed, upon pathogen attack *ELENA1* transcript abundance increases and binds to the promoter region of *PATHOGENESIS- RELATED1* (*PR1*) gene facilitating the recruitment of MED19a in this region, enhancing the expression of *PR1* (Seo et al., 2017). Likewise, in humans *PANDA* and *DHFR* lncRNAs are able to directly interact with TFs modulating their binding to gene promoters, directly influencing their target gene activity. LncRNAs can also act negatively on gene transcription. For example, in *Arabidopsis*, *HIDDEN TREASURE1* (*HID1*) lncRNA decreases *PHYTOCHROME- INTERACTING FACTOR 3* (*PIF3*) transcriptional activity by binding to its first intron. Downregulation of *HID1* expression increases *PIF3* gene activity and the subsequent elongation of the hypocotyl. The mechanism by which *HID3* decreases gene expression is still

unclear, but one explanation is that *HID1* forms a ribonucleoprotein complex interacting with the *PIF3* genomic region, thus impairing Pol II progression along the *PIF3* gene (Wang et al., 2014).

When two genes oriented in the opposite direction overlap, they cannot be transcribed at the same moment, otherwise the two Pol II complexes will collapse and their respective progression on the DNA will be blocked and then degraded. In *Arabidopsis*, the lncRNA *SVALKA* takes advantage of this phenomenon to regulate its neighboring reverse-oriented gene *CBF1* involved in cold tolerance. Prolonged cold exposure triggers a transcriptional read-through of the *SVALKA* genomic region leading to Pol II collision between *SVALKA* and *CBF1* transcriptions, decreasing *CBF1* transcription rate (Kindgren et al., 2018) (**Figure 10B**). In agreement, in yeast, it has been shown that after the Pol collision, both Pol II polymerases are removed from the DNA through proteolysis (Hobson et al., 2012). In yeast, transcriptional read through of *CUT60* lncRNA reached the promoter of its neighboring similarly orientated gene *ATP16*, repressing its expression (Du Mee et al., 2018). Surprisingly, when two genes are in the same orientation, the transcriptional readthrough can lead to transcripts fusion (Kaessmann, 2010), increasing the transcripts diversity or leading to abnormal transcripts. Indeed, it has been estimated that 65% of human genes can form chimeric RNAs through association of the RNA sequence with another gene (Birney et al., 2007; Gingeras, 2009). Taken together, lncRNAs can change the transcriptional efficiency of a gene through interaction with transcription-related proteins or through transcriptional readthrough both as sense or antisense leading to different transcriptional outputs.

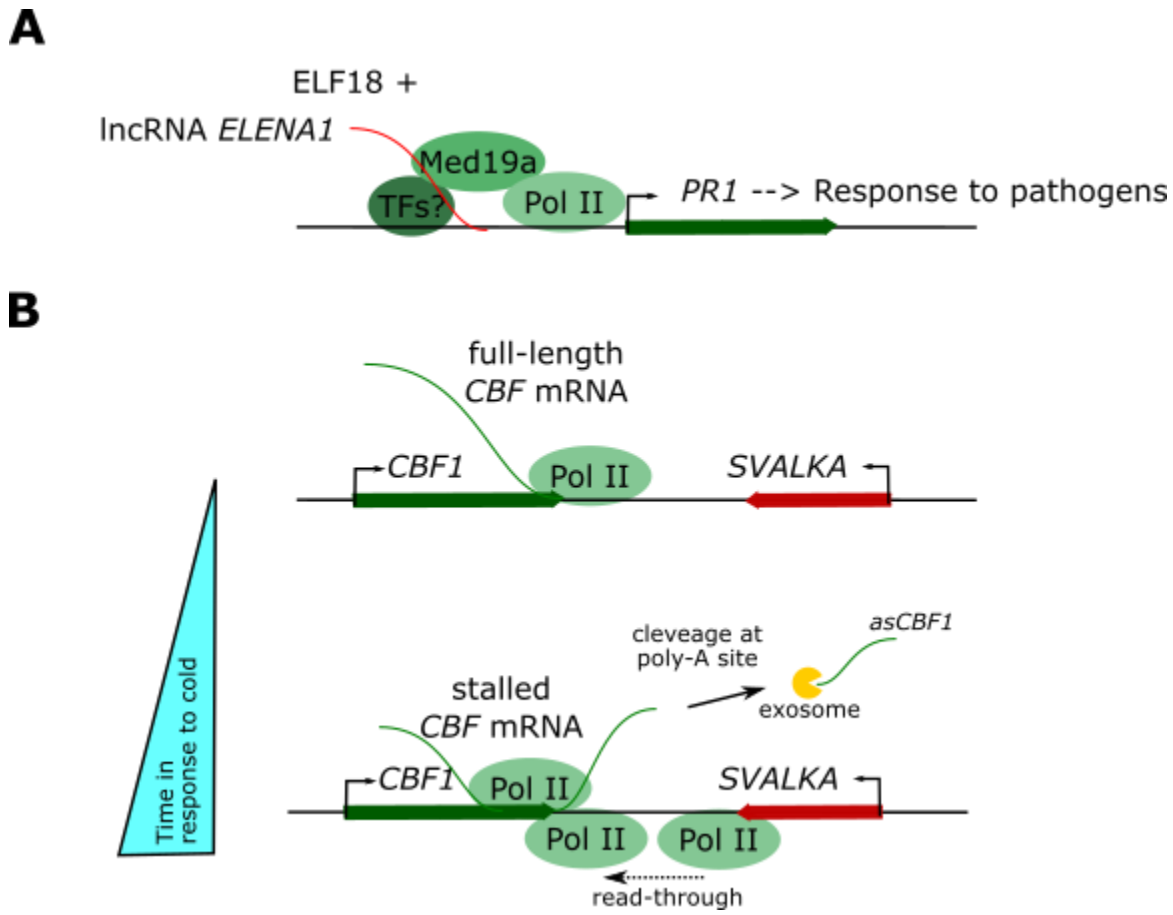


Figure 10. Pol-mediated regulation of gene transcription by lncRNAs (Adapted from Lucero et al., 2020). **A.** The *Arabidopsis ELENA1* lncRNA physically interacts with the Med19a subunit of the MEDIATOR complex and facilitates its recruitment to the *PR1* gene promoter, activating its transcription in response to ELF18. **B.** A transcriptional Pol II read-through within the *Arabidopsis* lncRNA *SVALKA* locus is occurring during a prolonged cold exposure, triggering the transcription of the exosome-sensitive lncRNA *asCBF1* which is cleaved at its polyadenylated sites. The transcription of *asCF1* causes Pol II collision with the *CBF1* locus, directly stopping *CBF1* transcription.

3.2 LncRNAs-mediated modification of the epigenetic landscape

LncRNAs are able to physically interact with histone modifier proteins. For example, the human *PANDA* lncRNA, in addition to interact with the TF *NF-YA*, can binds to the *scaffold-attachment-factor A* (*SAFA*) and the PRC1-2 to modulate cell senescence (Hung et al., 2011). Similarly, in *Arabidopsis*, *AGAMOUS INTRONIC RNA 4* (*AG-incRNA4*) binds to the CLF sub-unit from the PRC2 complex (**Figure 11A**). *AGAMOUS* (*AG*) gene encodes a MADS TF involved in flower development (Schubert et al., 2006). In the absence of *AG-incRNA4*, there is a reduction of H3K27me3 on *AG* chromatin and an induction of *AG*, suggesting a role of *AG-incRNA4* in the recruitment of PRC2 to this region (**Figure 11A**). Interestingly, the intronic *AG-incRNA4* lncRNA also mediate the PRC2 recruitment to its promoter mediating its own silencing (Wu et al., 2018). Also, *FLOWERING LOCUS C* (*FLC*) gene is a MADS TF involved in flowering transition through a complex chromatin state switch involving the action of three lncRNAs. First, *COOLAIR*, an antisense transcript of the *FLC* gene, physically interacts with the 5′ region of *FLC* and mediates, with *FLOWERING LOCUS D* (*FLD*) the demethylation of H3K4me2 resulting in *FLC* transcriptional repression (Liu et al., 2010b). Additionally, the *COLD ASSISTED INTRONIC NONCODING RNA* (*COLDAIR*) *FLC* intronic lncRNA interacts with the CLF PRC2 sub-unit to further regulate epigenetic silencing of *FLC* gene through H3K27me3 deposition (Heo and Sung, 2011). Similarly, the *COLDWRAP* lncRNA arising from the promoter of *FLC* is also able to interact with the PRC2 complex and could modulate *FLC* silencing in different ecotypes (Kim and Sung, 2017) (**Figure 12A**). In rice, the expression of the *LRKs* gene cluster involved in grain yield is modulated by an antisense lncRNA, named *LRK ANTISENSE INTERGENIC RNA* (*LAIR*), transcribed from a region inside the *LRK1* gene. This lncRNA is able to recruit the *OsWDR5* (*WD REPEAT DOMAIN5*), involved in the H4K16 acetylation, of the *LRK* genes region, increasing their expression. Concomitantly, the overexpression of *LAIR* lncRNA drastically increased grain yields in rice (Wang et al., 2018) (**Figure 11B**).

LncRNAs can also modulate the chromatin state by being processed into siRNA to trigger gene silencing through DNA methylation via the RdDM mechanism. In this case, the lncRNA can be transcribed by Pol IV and Pol V, two Pol II homologs which are specific to the plant kingdom. For example, in *Arabidopsis*, an auxin stimulus triggers a strong upregulation of both *PID*, a key regulator of auxin transport, and its neighboring lncRNA AUXIN-REGULATED

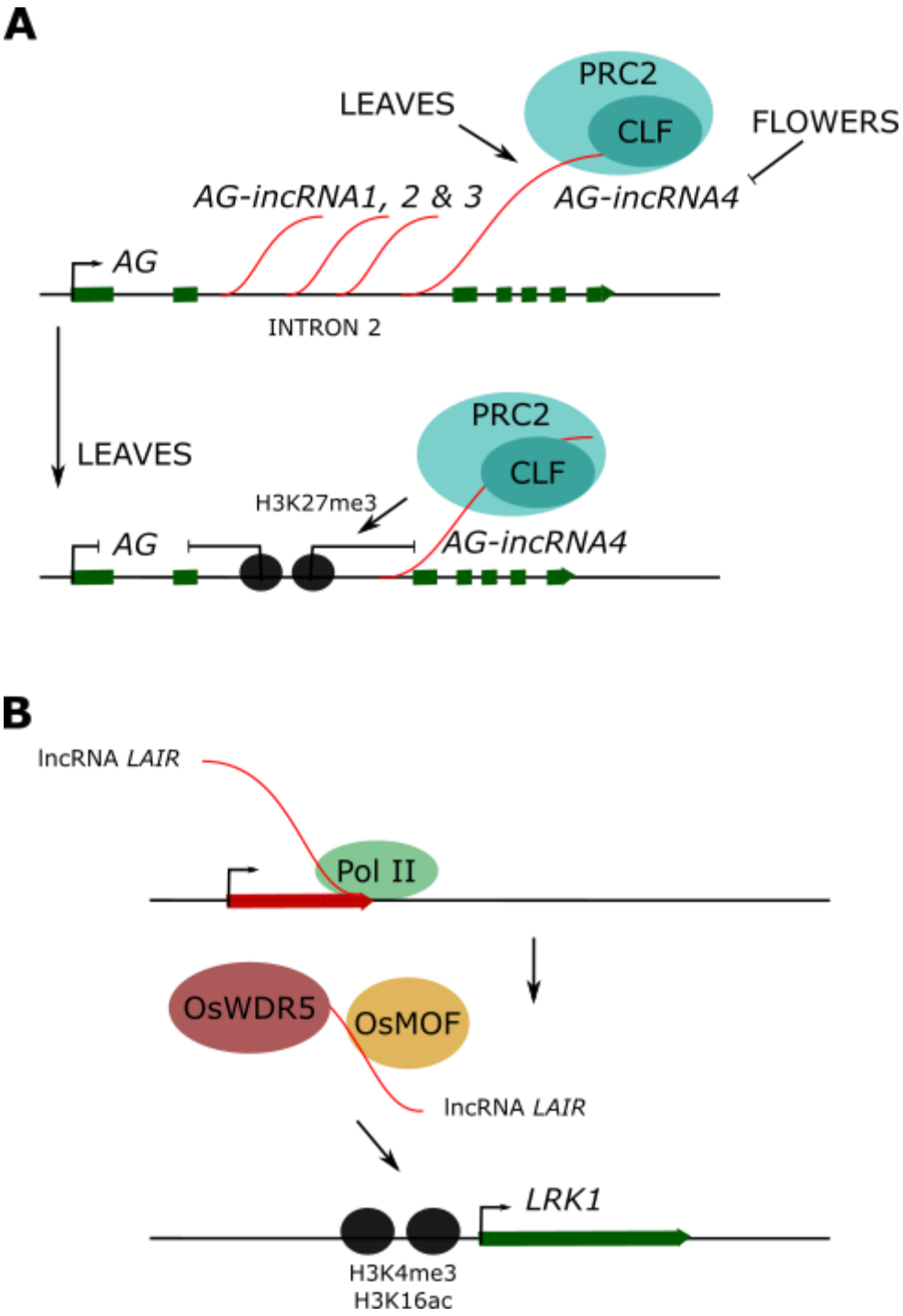


Figure 11. Epigenetic-mediated regulation of gene transcription by lncRNAs. **A.** The *Arabidopsis* intronic lncRNA *AG-incRNA4* recruits PRC2 to the second intron of the *AGAMOUS* gene through interaction with the CLF subunit, directly influencing H3K27me3 deposition. Interestingly, the *AG-incRNA4* is repressed in flowers but not in leaves, impairing *AGAMOUS* transcription in a tissue-specific manner (Adapted from Lucero et al., 2020). **B.** The rice lncRNA *LAIR* recruits the epigenetic modifier *OsMOF* and *OsWDR5a*, increasing H3K4me3 and H3K16ac deposition within the *LRK1* gene, enhancing its transcription (Adapted from Wang et al., 2018).

PROMOTER LOOP (*APOLO*) through the disruption of a chromatin loop encompassing the *APOLO* region and the PID promoter likely via an active DNA demethylation. In this linear chromatin configuration, both *APOLO* and *PID* are expressed. Gradually, Pol II-dependent *APOLO* transcripts recruit LHP1 protein to reform the chromatin loop while PRC2 complex deposits H3K27me3 to strengthen the compaction of the chromatin loop (see details in 3.3). Interestingly, *APOLO* is also transcribed by Pol IV-V and generates siRNA that triggers DNA methylation through the RdDM pathway, maintaining the silencing of the *APOLO-PID* regions (**Figure 12B**). As the epigenetic status and chromatin condensation sometimes directly influences the genome topology, it was proposed that lncRNA-mediated modification of the epigenetic landscape can influence the 3D configuration of the genome.

3.3 Chromatin architecture changes through lncRNAs activity

In the last decade increasing examples show the ability of lncRNAs to modulate the chromatin conformation (Quinodoz and Guttman, 2014). The addition of RNase-A into the nucleus has proven that euchromatin is maintained in a decondensed configuration through nuclear-located RNAs (Caudron-Herger et al., 2011; Caudron-Herger and Rippe, 2012). In addition, genome-wide conformation analyses in mammalian cells show that TAD boundaries [MC5] [tr6] are regulated through RNA transcription (Barutcu et al., 2019). Indeed, introduction of transcriptional inhibitors within chromatin extracts significantly weakens the TAD frontiers (Barutcu et al., 2019). More specifically, the *COLDWRAP* lncRNA transcribed within the FLC-epigenetically regulated flowering gene mediates the formation of a repressive intragenic chromatin loop which blocks Pol II transcription, inhibiting *FLC* expression allowing plant to flower (Kim and Sung, 2017) (**Figure 12A**). Strikingly, it has also been shown that the *APOLO lncRNA* also acts in trans, through the formation of RNA:DNA duplexes named R-loops. R-loops containing the *APOLO RNA* are able to remove LHP1 away from the chromatin in order to modulate the local chromatin conformation and influence gene transcription (Ariel et al., 2020) (**Figure 12B**). Interestingly, a significant proportion of genomic loci recognized by the *APOLO* RNA contain auxin-responsive genes. Among them, the LEUCINE RICH EXTENSIN2 (LRX2), involved in cell wall remodeling upon lateral root emergence (Lewis et al., 2013), and ROOT HAIR DEFECTIVE 6 (RHD6), a key regulator of root hair initiation (Lin et al., 2015; Moison et al., 2021).

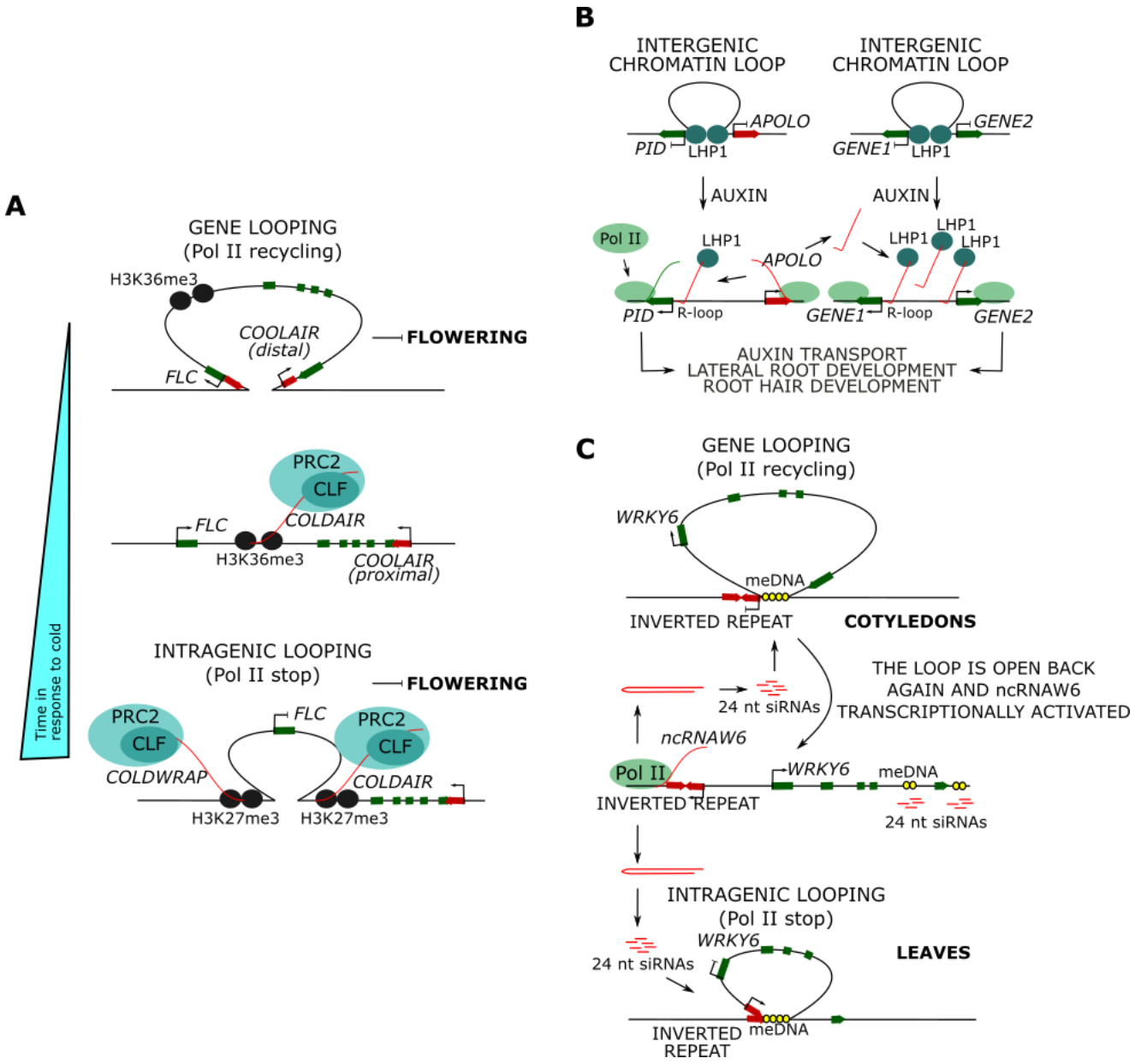


Figure 12. Chromatin loop-mediated regulation of gene expression by lncRNAs (Adapted from Lucero et al., 2020). **A.** The floral repressing gene *FLOWERING LOCUS C* (*FLC*) is regulated by three lncRNAs. *COOLAIR* distal isoform lncRNA transcripts accumulation correlates with a gene looping joining the promoter and TES of the *FLC* locus, likely participating in Pol II recycling. Cold exposure triggers the disruption of the chromatin looping, allowing the production of a *COOLAIR* proximal isoform and the intronic lncRNA *COLDAIR*. *COLDAIR* lncRNA recruits PRC2 to the *FLC* locus, increasing H3K27me3 deposition. The *COLDWRAP* lncRNA promoter starts to be transcribed later after longer cold exposure and also recruit the PRC2 complex within the FLC upstream region. Interestingly, the *COLDWRAP*-mediated PRC2 recruitment within *FLC* region mediates the formation of an intragenic chromatin looping encompassing the first intron of *FLC*. Consequently, the Pol II processing is disturbed and the *FLC* gene is silent. **B.** In *Arabidopsis*, an intergenic chromatin loop encompassing the *APOLO* and *PID* promoters, is disrupted in response to auxin stimulus, triggering *APOLO* and *PID* gene transcription. Interestingly, *APOLO* directly interacts with the DNA forming R-loop in the *PID* locus and other auxin-responsive genes. *APOLO* also interacts with the PRC1 component LHP1 chromatin protein, modulating its binding to the chromatin where *APOLO* forms R-loops. Consequently, *APOLO* regulates neighboring and distant genes within the *Arabidopsis* genome, through fine-tuning LHP1 binding to these loci, thus influencing the auxin transport (Ariel et al., 2014), lateral root (Ariel et al., 2020) and root hair development (Moison et al., 2021). **C.** The sunflower promoter lncRNA *ncRNAW6* fine-tunes *WRKY6* gene expression in a tissue-specific manner. In the cotyledons, *ncRNAW6* mediates DNA methylation within *WRKY6* gene through RdDM, together with the formation of a chromatin loop encompassing the whole *WRKY6* region. The gene looping allows Pol II recycling, positively correlating with *WRKY6* gene transcription but negatively with *ncRNAW6*. Gradually, the gene looping is disrupted allowing *ncRNAW6* expression and the re-formation of the loop. An alternative chromatin loop is formed within the leaves putting in close proximity the *ncRNAW6* and the *WRKY6* core gene. Here, the *ncRNAW6* is still expressed and the chromatin loop is not disrupted, thus repressing *WRK6* gene transcription.

In sunflower, a gene loop putting in close proximity the promoter and TES of *HaWRKY6* allows an efficient Pol II recycling, increasing *HaWRKY6* expression. Interestingly, this active chromatin loop is only found in cotyledons tissues (**Figure 12C**). Whereas, in the leaves a chromatin loop encompassing the promoter and the middle of the gene (4th intron) of *HaWRKY6* block the elongation of Pol II, decreasing *HaWRKY6* expression. Strikingly, a RdDM-functioning lncRNA is transcribed from the *HaWRKY6* promoter. In cotyledons, the active chromatin loop avoids its transcription which in term maintains a high level of expression of *HaWRK6*, while its absence in leaves allows the repression of *HaWRK6* through RdDM (Gagliardi et al., 2019) (**Figure 12C**).

3.4 LncRNAs mediating post-transcriptional regulation of gene expression

We saw previously that housekeeping ncRNAs participate in the maturation of mRNA through splicing (see section 1.1.1). Interestingly, some lncRNA can also impact splicing. This is the case of the Arabidopsis lncRNA *Alternative Splicing COmpetitor* (*ASCO*) which can physically interact with the nuclear speckle RNA-binding proteins (NSRs), involved in AS regulation during developmental processes in *Arabidopsis* (Bardou et al., 2014). *ASCO*-NSRs interactions impaired the splicing of auxin responsive genes influencing auxin-driven lateral root formation in *Arabidopsis* (**Figure 13A**). More recently the *ASCO* lncRNA have been found to interact with other components of the spliceosome (SmD1b and PRP8a), shedding light on the relationship between *ASCO* action on AS and plant resistance to pathogens (Rigo et al., 2020) (**Figure 13B**). More precisely, increased levels of *ASCO* impaired the full splicing of flagellin-related transcripts. Consequently, RNAi-mediated *ASCO* silencing increases root growth sensitivity to flagellin 22 (Rigo et al., 2020). Other lncRNAs regulate post-transcriptional processes by perturbing miRNA regulation. For example, the gene *PHOSPHATE 2* (*PHO2*), a key regulator of phosphate (Pi) homeostasis, is targeted by *miR399* (see section 1.1.2.2.). The *AtIPS1* lncRNA shares a highly similar 23nt sequence with *PHO2*, corresponding to the recognition sequence of *miR399*. Interestingly, in *IPS1*, there is a 3 nt sequence mismatches in the middle of the *miR399* binding site leading to the formation of a bulge avoiding *IPS1* cleavage. Consequently, *IPS1* is acting like a substrate of the *miR399* which remains stuck on *IPS1* as it cannot cleave this RNA. Therefore, miR399 action on *PHO2* is inhibited by the target mimicry

lncRNA (Franco-Zorrilla et al., 2007). These miRNA "sponges" can block miRNA-mediated post-transcriptional regulation of specific mRNA targets. Interestingly, the IPS1-mediated *miR399* inhibition is also conserved in maize, tomato and *Medicago truncatula* ((Wang et al., 2017a; Du et al., 2018) (**Figure 13C**). Also, in *Arabidopsis*, 407 putative lncRNA miRNA-sponge has been identified in response to blue-light; one of which directly implicated in the blue-light-mediated photomorphogenesis and mannitol stress response through the sequestration of *miR167*, involved in hypocotyl elongation (Sun et al., 2018; Sun et al., 2020a).

As two thirds of the human transcriptome is regulated by miRNA, it is likely that many lncRNAs are involved in this type of post-transcriptional regulation (Borah et al., 2018). Interestingly, circular RNAs (circRNAs) are a recent class of non-coding RNAs formed through reverse splicing of an exon or cyclization of an intron within a gene, generally coding (Tian et al., 2020). They present a covalent 5'-3' ends linkage (Santer et al., 2019), are non-polyadenylated and significantly more stable than linear lncRNA (Memczak et al., 2013). Abundant in the human genome, and expressed in a tissue-specific manner (Shao et al., 2017), many have been shown to act as miRNA sponges (Salzman et al., 2012) indicating that these circRNAs may also act as *IPS1*. Thousands of circRNAs have been identified in humans (Glažar et al., 2014), many of which seem directly implicated in cancer (Su et al., 2019). Similarly, thousands of circRNAs have been detected in nearly 30 plant species, among which *Arabidopsis* and many crops (Wang et al., 2015; Ye et al., 2015; Wang et al., 2017b; Dong et al., 2019). Despite their responsiveness to abiotic and biotic stresses, their biological functions in plants remain to be elucidated (Zhang et al., 2020). Now, the only example of a plant functional circRNA in the regulation of gene expression is the circRNAs emerging from the *SEP3* gene transcript, a positive regulator of female organ development (Favaro et al., 2003). Here, the circRNAs formation modulates flower development by impacting the splicing of its cognate mRNA *SEP3* (Conn et al., 2017). Altogether, lncRNA molecules constitute important regulators of gene expression. Physically interacting with the epigenetic modifying complexes, TFs or the DNA itself, they fine-tune transcriptional activity of target genes. Additionally interacting with key modulators of the mRNA splicing and stability, they also influence mRNA abundance at the post-transcriptional level, thus globally constituting new elements of the regulatory networks controlling cell gene expression.

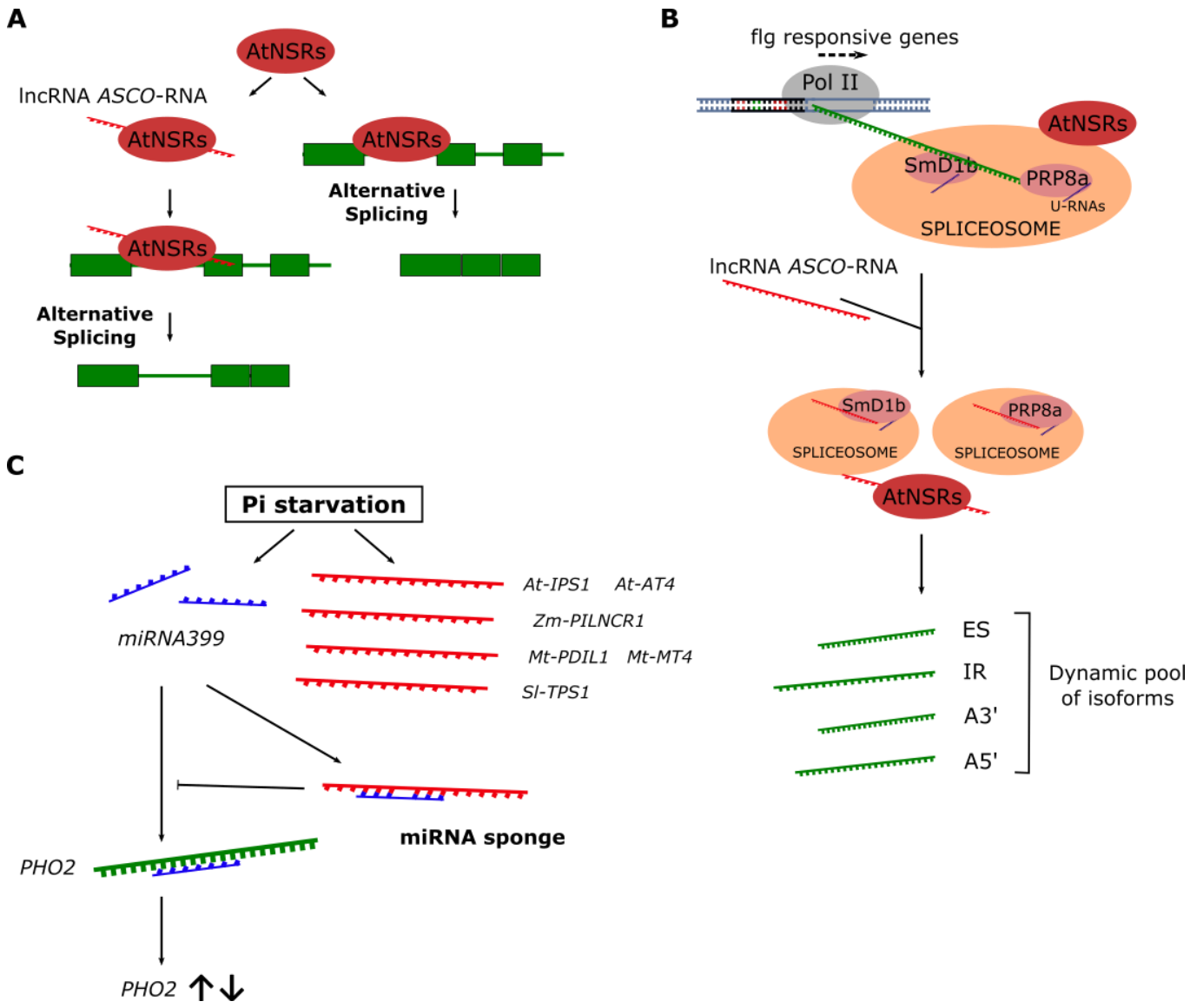


Figure 13. Post-transcriptional regulation of gene expression by lncRNAs. A. The Arabidopsis lncRNA *ASCO* physically interacts with the NSRs protein modulating the NSR-dependent alternative splicing of auxin-responsive gene transcripts (Adapted from Zhang et al 2019). **B.** In addition to hijack NSR proteins, the *Arabidopsis ASCO* lncRNA interacts with PRP8a and SmD1b, core subunits of the spliceosome, modulating the alternative splicing of flagellin-responsive genes (Adapted from Rigo et al., 2020). **C.** Phosphate starvation triggers the expression of the Arabidopsis *IPS1* lncRNA and *miR399* and their orthologs in multiple plant species. The *IPS1* lncRNA sequester *miR399*, preventing it to cleave the miR399-target *PHO2* transcript, thus participating in modulating *PHO2* transcript abundance upon Pi deprivation (Adapted from Zhang et al., 2019).

3.5 Review: Regulatory long non-coding RNAs in root growth and development

The root is a dynamically regulated organ that constantly adapts its architecture according to the plant surrounding and its stage of development. Notably, some lncRNAs are involved in the control of root gene expression, influencing the root growth and development. The following part will be presented under the form of a review, under revision in *Biochemical Society Transactions*, in which I participate as first author. It presents the current involvement of lncRNAs in the quantitative control of the root architecture under physiological and environmentally adverse conditions. This review also discusses how lncRNAs participate in the initiation of new root organs such as lateral roots or nodules.

Regulatory long non-coding RNAs in root growth and development

Thomas Roulé[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*

Abstract

Fixed to the soil through their roots, plants have evolved sophisticated mechanisms of gene regulation to defend against environmental constraints. Among them, long non-coding RNAs (lncRNAs) are an emerging class of RNAs, regulating gene expression at both the transcriptional and post-transcriptional level. Highly responsive to environmental cues or developmental processes they are involved in fine-tuning the plant response to these signals. Roots, in addition to anchoring the plant to the soil, allow the plant to absorb the major part of its mineral nutrient. Also, roots directly sense environmental constraints such as mineral nutrient availability and various stresses (abiotic or biotic) and dynamically adapts its growth according to it. Here, we review the role of lncRNAs for the control of root growth and development. In particular, we highlight their role in fine-tuning the main root growth and the development of root lateral organs, such as lateral roots and nodules. In addition, we explore their involvement in plant response to stresses and the regulation of nutrient assimilation and homeostasis, resulting in modification of root architecture. Given their crucial role in many developmental and stress-related biological mechanisms it is likely that lncRNAs will be targeted in plant breeding programs to subtly acclimate the emerging crops to the coming environmental changes.

KEYWORDS: lncRNA, root growth, root development, plant nutrition, stresses

Introduction

The development of next generation sequencing methods to analyze transcription in multiple cell types and conditions revealed that more than 90% of the eukaryotic genome is transcribed (Chen et al., 2021). Strikingly, the large majority of these transcripts remain in their RNA state and generally do not produce peptides constituting the so-called "dark matter" of the genome. From mammals to plants, this tremendous quantity of enigmatic non-coding transcripts is intensively studied, showing remarkable regulatory ways to fine-tune gene expression. Regulatory non-coding transcripts are divided into two main groups based on their size. First, the small RNAs (sRNAs), which are inferior to 200nt, counting the small-interfering RNAs (siRNAs), the micro-RNAs (miRNAs) and the phasing-siRNAs (phasiRNAs), regulating gene expression at the transcriptional or post-transcriptional levels through RNA-dependent DNA methylation or transcript cleavage/inhibition of translation, respectively. Second, the long non-coding RNAs (lncRNAs) are non-coding transcriptional units generally superior to 200nt. Similarly, to coding RNAs, the majority are transcribed by Pol II, poly-adenylated and subjected to splicing events (Chen et al., 2021). Other classes of lncRNAs involved in RNA dependent DNA methylation (RdDM) can also be transcribed by Pol IV and V leading to the production of 21-24 nt siRNAs. When compared to coding transcripts, lncRNAs are more specifically expressed: at certain state of developmental processes, only in certain tissues or in response to environmental stimuli (Rinn and Chang, 2020). In addition, they are generally shorter in length, contain less introns, thus generating a lower number of isoforms compared to coding transcripts (Golicz et al., 2018; Sarropoulos et al., 2019; Zhao et al., 2020a). Those lncRNA influencing gene expression at the transcriptional level are often enriched in the nuclear fraction of the cells. However, many lncRNAs can also be translocated to the cytoplasm, similar to the coding transcripts, to post-transcriptionally regulate gene expression (Fok et al., 2017).
More than thousands of lncRNAs are transcribed within eukaryotic cells, manifesting the likely participation in a wide range of biological processes (Statello et al., 2021). For example in mammals, the *Xist* lncRNA mediates the X chromosome epigenetic inactivation, critical for normal embryonic development (Zhao et al., 2008), whereas *COOLAIR*, *COLDAIR* and *COLDWRAP* lncRNAs mediates the epigenetic silencing of the *Flowering Locus C (FLC)* gene ensuring the end of vernalization and flowering in *Arabidopsis* (Swiezewski et al., 2007; Liu et

al., 2010c). It has been shown that plant lncRNAs are involved also in stress signaling, photomorphogenesis, fertility and plant growth and development (Chen et al., 2021). This review will focus on the influence of lncRNAs on root growth and development, during normal development as well as in response to stresses and their relevance for key root-mediated assimilation of soil nutrients.

Influence of lncRNAs on root growth and development

Phytohormones constitute key regulators of plant growth and development. Among them, auxin, regulates cell division and elongation and is described as a positive regulator of root growth (Hu et al., 2021). Nuclear Speckle RNA-binding Proteins (NSR) proteins are a class of protein involved in Alternative Splicing (AS) and auxin-related developmental processes, among which the development of lateral roots (LR) (Bardou et al., 2014) (**Figure 1A)**. Indeed, *nsra/ nsrb* double mutant is less sensitive to auxin and present a lower number of LR in response to auxin application (Bardou et al., 2014). Interestingly, the lncRNA *Alternative Splicing COmpetitor* (*ASCO*) is able to physically interact with the NSR proteins regulating AS during auxin signaling, directly modifying auxin-related root growth and development (Bardou et al., 2014) (**Figure 1A**). As a result, plants overexpressing *ASCO* exhibit an increased sensitivity to auxin. More recently, immunoprecipitation of the NSRa protein followed by a transcriptomic analysis has served to identify multiple NSR-dependent spliced mRNAs. Strikingly, many lncRNAs, other than *ASCO*, were shown to also interact with NSRa, suggesting that they could also direct NSR-dependent AS and subsequently impact the root architecture (Bazin et al., 2018).

Auxin transport within the root also designs the root architecture by modifying the distribution of auxin inside root tissues. For example, mutants impaired in auxin transport exhibit a significant reduction of root length and meristem size (Billou et al., 2005). The *PINOID* (*PID*) kinase is a key regulator of auxin transport as it seems to phosphorylate PIN proteins directly involved in this process, and its repression leads to a reduced sensitivity to root gravitropism (Sukumar et al., 2009). In normal conditions, the *PID* gene is lowly expressed. Auxin signaling enables the disruption of a chromatin loop encompassing the PID promoter and its neighboring gene, the lncRNA *AUXIN-REGULATED PROMOTER LOOP* (*APOLO*), and

induces the transcription of both *PID* and *APOLO* genes through ARF TFs (Ariel et al., 2014). Later on, *APOLO* physically interacts with the LHP1 chromatin looping protein, reforming the chromatin loop, subsequently repressing *PID* gene activity. In addition, *APOLO* is able to direct DNA methylation to this region and the deposition of the H3K27me3 repressive histone mark, which strengthen the repressive state of the *PID-APOLO* region (**Figure 1B**). As expected, *APOLO* downregulated lines exhibit a reduced sensitivity to root gravitropism as the *pid* mutant (Ariel et al., 2014). More recently, it has been observed that *APOLO* promoter is active during LR development, a process controlled by auxin (Ariel et al., 2020). In the same study, authors showed that the *APOLO* RNA is able to recognize multiple individual loci in *trans* through the formation of R-loops, directly influencing the local chromatin conformation and gene activity of several auxin-responsive genes (Ariel et al., 2020). Interestingly, a significant proportion of *APOLO*-target genes, in addition to being responsive to auxin, were also implicated in LR formation (**Figure 1B**). For example, the *LEUCINE-RICH REPEAT/EXTENSIN2* gene which is involved in cell wall remodeling during LR emergence (Lewis et al., 2013), is targeted by *APOLO*. In agreement, seedlings overexpressing *APOLO* present a significant increase in LR density in response to auxin (Ariel et al., 2020)

Cytokinins (CK) are another crucial class of plant hormone described as an inhibitor of root growth (Amasino, 2005), widely used for plant regeneration to inhibit adventitious root formation (Bishopp et al., 2009). Essential for shoot meristem activity, leaf growth and senescence (Mok et al., 2000), CK are mainly produced in the root tip, but are also synthesized locally in the plant aerial parts (Nordström et al., 2004). Interestingly, the *Sho* gene from *Petunia hybrida,* implicated in CK biogenesis contains an antisense non-coding transcript (Zubko and Meyer, 2007). The expression of *Sho* and of its antisense seems to be correlated and leads to the production of 24nt small RNA in most tissues. This suggests a role of *Sho* antisense transcript to fine tune the accumulation of its sense transcript (**Figure 1C**). Strikingly, roots were the only tissue, where the 24nt were absent suggesting that *Sho* is not degraded in this compartment. This would allow the maintenance of a high level of CK synthesis in the root as compared to other organs. Thus, the *Sho* antisense lncRNA could be implicated in the organ-dependent control of CK biogenesis within Petunia *hybrida* (Zubko and Meyer, 2007).

The *SCARECROW* (*SCR*) gene is required for the asymmetric cell division occurring during *Arabidopsis* root development (Di Laurenzio et al., 1996). Similarly, the SCR-like transcription factors are a family of proteins found in rice and critical for proper root development (Cho and Paszkowski, 2017). More precisely, in non-root organs, such as shoots and flowers, *OsmiR171* is acting as an inhibitor of *SCR-like* genes. Interestingly, a non-coding transcriptional unit, named *MIKKI* ('decoy' in Korean), derived from a retrotransposable element indirectly controls *SCL-like* gene expression through the inhibition of the *miR171* activity (**Figure 1D**). This inhibition is realized through target mimicry or miRNA sponge (Franco-Zorrilla et al., 2007), as the *MIKKI* non-coding transcripts bind to the *miR171* but showed a loss of base pair complementarity at the cleavage site. This resulted in avoiding both *MIKKI* cleavage and led to the sequestration of miR171. Notably, the root length was significantly reduced in plants overexpressing the *MIKKI* transcripts (Cho and Paszkowski, 2017) demonstrating a functional role of this *miR171-MIKKI-SCR* regulatory node. Altogether, lncRNAs constitute emerging regulators of root growth and development through their participation in plant hormone signaling pathway through the regulation of key root-related genes.

LncRNAs as root-stress signaling molecules

The increasing global population and concomitant climate change has resulted in an unsustainable path for global food security. In order to address these challenges new strategies are required that may allow us to prepare our global crops against environmental problems linked to climate change. Interestingly, many lncRNAs are activated upon biotic and abiotic stresses underpinning that certain lncRNAs may be implicated in plant stress resilience. Although stresses act at various levels of the plant, the root compartment is particularly sensitive to adverse environmental conditions such as osmotic-related stresses, cold or pathogen attacks. Impacting root growth directly affects the proper development of agriculturally valuable plants. Large variation in lncRNA expression has been linked to environmental stresses in multiple model and crop plants. For example, transcriptomic analyses revealed that more than hundred lncRNAs are differentially expressed upon drought stress in *Arabidopsis* (Chen et al., 2021), poplar (Shuai et al., 2014), millet (Qi et al., 2013), cassava (Li et al., 2017a) and rice (Chung et al., 2016). Notably, two rice Natural Antisense Transcripts (NATs), *Os02g0250700* and *Os02g0180800,* may be implicated in the transcriptional regulation of a

late embryogenesis abundant (*LEA*) and *cinnamoyl-CoA reductase* (*CCR*) genes, both participating in plant drought tolerance (Chung et al., 2016) (**Figure 1E**). In Arabidopsis, the *DRought-Induced lncRNA* (*DRIR*) regulates a subset of genes implicated in ABA signaling and water translocation, of particular interest for plant drought stress resilience. Interestingly, *DRIR* also regulates the transcriptional activity of *FUCOSYLTRANSFERASE4* and the transcription factor *NAM/ATAF/CUC3* potentially modulating the plant tolerance to drought and salt stresses through these target pathways (Qin et al., 2017) (**Figure 1F**). Interestingly, salt stress also disturbs transcript accumulation of miscellaneous lncRNAs in many plant species, including *Arabidopsis* (Di et al., 2014), soybean (Chen et al., 2019), *Medicago* (Wang et al., 2015), tea bush (Wan et al., 2020), cotton (Zhang et al., 2019), sorghum (Sun et al., 2020b) and poplar (Ma et al., 2019). The cotton *lncRNA973* regulates a subset of genes of the salt-stress signaling pathway, mainly participating in scavenging ROS, directly fine-tuning the plant salt stress tolerance (Zhang et al., 2019) (**Figure 1G**). Remarkably, ectopic overexpression of the cotton *lncRNA973* within *Arabidopsis* significantly increased the plant salt tolerance from germination to seedling establishment, whereas reducing the *lncRNA973* through virus-induced gene silencing (VIGS) in cotton significantly reduced the plant tolerance to salt stress. Additionally, the *lncRNA973* may act as target mimicry for the *miR399* (Deng et al., 2018b), regulating the root phosphate (Pi) assimilation and homeostasis (Kim et al 2011; Bari et al 2006; (Bari et al., 2006; Chiou et al., 2006; Pant et al., 2008; Kim et al., 2011a). Similarly, the root crop cassava regulates drought and cold stress responses through the *lncRNA340* acting as a target mimicry for *miR169* (Li et al., 2017a) (**Figure 1H**).

Low temperature also affects plant growth and development. For example, it has been noticed that prolonged cold exposure increased root hair growth, presumably participating in the plant cold acclimation. Strikingly, plantlets with reduced or increased levels of *APOLO* lncRNA transcripts constitutively produced more root hairs compared to WT (Moison et al., 2021). In agreement, plants with a modified level of *APOLO* present an increased level of *RHD6* transcripts, a key gene critically involved in the regulation of root hair fomation. Notably, the authors showed that *APOLO* controls *RHD6* transcriptional activity during cold stress through the recruitment of WRKY42 to its promoter region (Moison et al., 2021) (**Figure 1B**).

Roots can also be subjected to deleterious biotic stresses which induce significant changes of the plant coding and non-coding transcriptome. For example, the Root-knot nematodes (RKNs) is one of the most destructive pests for agriculture, due to its large spectrum of plant hosts and the serious decrease in plant yield provoked upon infection (Giannakou and Panopoulou, 2019)Interestingly, more than 500 lncRNAs are differentially expressed upon RKNs infection of roots and several may be involved in the nematode stress response (Li et al., 2018b). Furthermore, deregulation of the *ASCO* lncRNA significantly deregulates activity of genes implicated in flagellin response in addition to the auxin-related genes (Rigo et al., 2020) (**Figure 1A**). Curiously, most of these "biotic stress" genes affected by *ASCO* knock-down in root tissues do not overlap with the deregulated genes detected in *nsra*/*nsrb* mutants suggesting an alternative role of *ASCO* for plant biotic stress response. Indeed, recently *ASCO* has been shown to also interact with PRP8a and SmD1b, in addition to NSRs. These proteins are central spliceosome components implicated in the splicing of several flagellin-responsive genes. Accordingly, plants with a reduced level of *ASCO* transcripts are hypersensitive to flagellin treatments as indicated by a significant reduction of primary root growth and root apical meristem size in flagellin-treated plants (Rigo et al., 2020). Collectively these results support that lncRNA activity is highly sensitive to environmental cues, and they are emerging as novel components in the plant response to biotic and abiotic environmental stresses.

LncRNAs-mediating nutrient homeostasis

With the increasing worldwide population, it has been estimated that the global agriculture production needs to be increased by 70% to meet the food demand of 2050, making critical the finding of new alternatives to increase crop yield. Substantial application of nitrogen (N)-containing fertilizers significantly participates in increasing cereal yield and has a strong impact on root architecture, the tissues involved in nutrient uptake. Notably, N-dependent changes in root architecture are found in many plant species (Forde and Walch-Liu, 2009; Forde, 2014) and this has consequences on other root-dependent traits such as water acquisition. Legume plants are able to develop a unique organ on their roots: nodules, in which symbiotic N^2-fixing *rhizobia* capture the atmospheric N (Kistner and Parniske, 2002). Hence, they can have a major impact on soil nitrogen content and are key components of environmentally friendly agricultural practices (Zhao et al., 2020b). Interestingly, the *ENOD40* lncRNA participates in legume nodulation (**Figure 1I**) and certain sequences of this gene are

also conserved in non-legume plants (Gultyaev and Roussis, 2007). RNAi-mediated silencing of *ENOD40* in *Lotus japonicus* suppresses nodule formation (Kumagai et al., 2006). Strikingly, it has been proposed that the legume *ENOD40* lncRNA presents a dual function either as a lncRNA involved in nucleocytoplasmic trafficking of *MtRBP1,* a key regulator of gene expression (Campalans et al., 2004), or through the production of two small peptides interacting with the sucrose synthase protein and participating in sucrose mobilization during nodulation (Bardou et al., 2011). In *Medicago*, another lncRNA consisting of a short variant of the *TAS3* lncRNA participates in nodule formation through target mimicry of *miR390*, preventing the fabrication of trans-acting small interfering RNA targeting TF from the *ARFs* family (Traubenik et al., 2020) (**Figure 1J**). Interestingly, the *TAS3-miR390-ARFs* module is also involved in the lateral root initiation. Notably, the *miR390* is specifically expressed at the sites of lateral root initiation, triggering the processing of *TAS3*-tasiRNAs that mediates inhibition of *ARF2/3* and *ARF4* promoting the lateral root expansion (Marin et al., 2010) (**Figure 1J**).

Apart from legumes, it has been observed that N starvation changes the expression of multiple lncRNAs in *Arabidopsis* (Fukuda et al., 2020), poplar (Chen et al., 2016), barley (Chen et al., 2020), maize (Lv et al., 2016) and rice (Shin et al., 2018). For example, two antisense lncRNAs are significantly induced in rice roots upon N starvation (Shin et al., 2018), likely regulating the *AMT1* gene participating in N assimilation (Ishiyama et al., 2004). Interestingly, the *T5120* lncRNA is directly regulated by *NLP7*, a master regulator of the N signaling pathway, prompting that the *T5120* may participate in *NRT1.1* gene activation (Liu et al., 2019) (**Figure 1K**). As the root constitutes the main compartment for N uptake in land, plants elaborate many strategies to enhance N acquisition efficiency, mainly through changes of root architecture (Walch-Liu et al., 2006). Thus, the implication of lncRNA for the regulation of N signaling pathway, assimilation and nitrogen-fixing symbiosis strengthen the relevance of the non-coding transcriptome for a proper plant N homeostasis in different plants.

In addition to N, phosphate (Pi), is another mineral essential for plant growth and development. Interestingly, Pi starvation also triggers strong changes of root architecture together with changes of gene expression within the root to increase P-nutrient acquisition from the soil environment. Notably, the primary root growth of the *Arabidopsis* Col-0 ecotype is immediately interrupted upon Pi starvation through a Fe-dependent mechanism (Gutiérrez-Alanís et al., 2017). Interestingly, this response is not conserved across all *Arabidopsis*

accessions, such as the L*er* ecotype (Reymond et al., 2006). Notably, it has been observed that the non-coding transcriptomes between Col-0 and L*er* ecotypes among a Pi starvation is strongly different when compared to the coding transcriptomes (Blein et al., 2020). Strikingly, in only one hour of Pi starvation, many lncRNAs are differentially regulated within *Arabidopsis* root tips (Blein et al., 2020). In the same study, authors showed that two lncRNAs enriched in a particular *Arabidopsis* ecotype, *NPC48* and *NPC72*, significantly reduced the main root growth when overexpressed. Furthermore, one of this lncRNAs disturbs the transcriptional activity of key root growth regulator such as *RGF7*, *BIG*, *RPK2* and *CASP5* (Blein et al., 2020) (**Figure 1L**) further supporting that lncRNA are emerging regulators of root growth.

Plants have evolved many adaptive mechanisms to maintain a homeostatic level of Pi within their cells in soil environments with variable levels of combined Pi. For example, upon Pi starvation the *ZAT6* TF is induced and regulates a subset of Pi-responsive genes in *Arabidopsis thaliana*. Interestingly, RNAi-mediated silencing of *ZAT6* is lethal, whereas overexpression limits the expression of Pi-related genes, impairing Pi acquisition and root growth upon seedling development. Indeed, the ZAT6-mediated repression of primary root growth directly regulates the Pi homeostasis through changes of root architecture (Devaiah et al., 2007). In addition, *PHO2* is another gene involved in the Pi-starvation response, directly regulating the plant Pi homeostasis by controlling the degradation of *PHO1*, implicated in Pi loading onto the xylem (Poirier and Bucher, 2002). Notably, *pho2* mutant accumulates toxic amounts of Pi onto the shoot, because of an uncontrolled Pi uptake and translocation into this tissue (Delhaize and Randall, 1993; Dong et al., 1998). Interestingly, *miR399* is induced upon Pi starvation and target *PHO2* transcript limiting Pi assimilation (Liu et al., 2012b). Strikingly, two Arabidopsis lncRNAs, *At4* and *IPS1*, conserved in Medicago, maize and tomato, are also induced within low Pi condition (Burleigh and Harrison, 1999; Shin et al., 2006; Wang et al., 2017a) and share a partial complementary sequence with *miR399*. Concomitantly, the Pi-induced lncRNAs act as miRNA sponge for *miR399* directly influencing *PHO2* transcript level, thus the Pi homeostasis (Burleigh and Harrison, 1999; Shin et al., 2006; Franco-Zorrilla et al., 2007; Wang et al., 2017a) (**Figure 1M**). In agreement, *at4* mutants present a Pi homeostasis defect between the root and shoot compartment under Pi starvation (Burleigh and Harrison, 1999; Shin et al., 2006). Other regulations may appear in other plant species that impinge on this regulatory node. For example, in rice, an antisense lncRNA of *OsPHO1;2* (*cis-NATPHO1;2*) participate in Pi

homeostasis promoting the translation of *OsPHO2* transcript (Jabnoune et al., 2013) (**Figure 1N**). Notably, RNAi-mediated silencing of *cis-NATPHO1;2* decrease the quantity of OsPHO2 protein, reducing Pi content and grain yield. Whereas, *cis-NATPHO1;2* ectopic overexpression increased the PHO1;2 protein level (Jabnoune et al., 2013). Hence antisense lncRNA regulation regulates *PHO2* but this type of regulation is not conserved in *Arabidopsis* or other species.

Altogether, there is growing evidence that lncRNAs are emerging regulators of various biological processes and stress responses. Many Genome Wide Association Studies (GWAS) for diverse agriculturally relevant biological traits mapped to intergenic regions where lncRNAs may exist hinting at a potential function of lncRNA in QTLs for root growth, development and responses to adverse constraints (Liu et al., 2017b; Stagnati et al., 2019; Alseekh et al., 2021; Ma et al., 2021); Liu et al 2017; Stagnati et al 2019; Ma et al 2021). Also, RNA molecules might be more attractive than proteins to modulate gene regulation as their effects may be more transient (Song et al., 2019). Finally, as root growth and development is highly described at the protein coding genes, co-regulatory networks involving lncRNAs may offer new mechanisms for root growth regulation. With less than 1% of the annotated plant lncRNAs functionally characterized, they constitute an immense reservoir for discovering new gene regulatory mechanisms and may offer attractive targets for future plant breeding programs.

Perspectives

- Climate change-related environmental constraints directly or indirectly disturbs the root compartment, impacting water uptake, nutrient acquisition and plant anchoring, ultimately decreasing crop yield. In this context, lncRNAs emerged as interesting molecules for the quantitative control of root-related traits, notably in response to the environment, and could be of interest for plant breeding programs.
- In all transcriptomically investigated crop species lncRNAs have been detected within the root compartment. However, few lncRNAs have been functionally characterized. Nevertheless, current studies show they are majoritarily involved in the quantitative regulation of plant root architecture allowing notably to adapt to root-related environmental stresses, strengthening their importance as key regulators of agronomically relevant traits.
- Despite the important quantity of lncRNAs detected in root, few have been functionally characterized, especially in crop species. Nevertheless, we anticipate that the intensive emerging characterization of these molecules will be used in plant breeding programs to subtly acclimate the crops to the coming environmental changes.

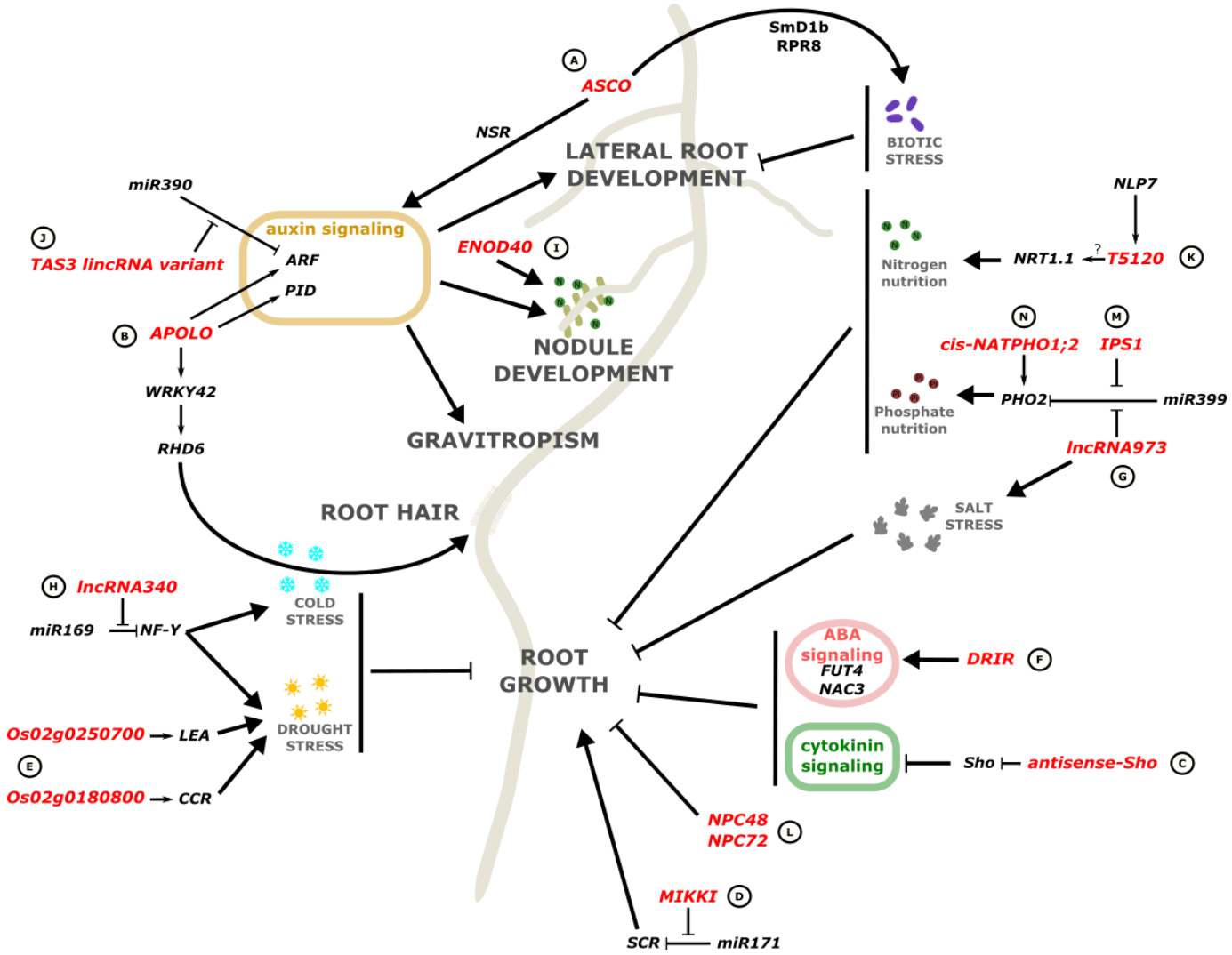


Figure 1. LncRNAs involved root growth and development and its response to environmental stresses. A. The Arabidopsis *ASCO* lncRNA modulates lateral root growth and development modulated by auxin and in response to biotic stress through its interaction with NSR (Bardou et al., 2014) and SmD1b-RPR8 (Rigo et al., 2020), respectively. **B.** The lncRNA *APOLO* fine-tunes the expression of several auxin-responsive genes (Ariel et al., 2020), regulating lateral root development and gravitropism (Ariel et al., 2014). In addition, it modulates the binding of the WRKY42 TF to the *RHD6* promoter, promoting root hair initiation under cold stress (Moison et al., 2020). **C.** In *Petunia hybrida*, the cytokinin-mediated root growth inhibition is modulated by an antisense lncRNA of *Sho* gene, a repressor of cytokinin signaling (Zubko and Meyer, 2007). **D.** The rice *MIKKI* lncRNA shapes the root architecture through target mimicry of the *miR171-SCR* module (Jungnam Cho and Jerzy Paszkowski, 2017). **E.** *Os02g0250700* and *Os02g0180800* are two antisense rice lncRNAs regulating the plant drought tolerance through the regulation of *LEA* and *CCR* genes, respectively (Chung et al., 2016). **F.** *DRIR* is an *Arabidopsis* lncRNA regulating a subset of ABA-related genes, including *FUT4* and *NAC3* (Qin et al., 2017). **G.** *LncRNA973* is a cotton lncRNA participating in the plant salt stress responses (Zhang et al 2019). Also, it acts as target mimicry of the *miR399-PHO2* node for phosphate (Pi) homeostasis (Kim et al., 2011; Bari et al., 2006; Chiou et al., 2006; Pant et al., 2008). **H.** The *lncRNA340* from the crop cassava roots acts as target mimicry of the *miR169-NF-Y* regulatory interaction, participating in the plant cold and drought stress tolerance (Li et al., 2017; Fang et al., 2014). **I.** The *Medicago ENOD40* lncRNA promotes nodule development (Bardou et al., 2011). **J.** A short variant from the *Medicago TAS3* lncRNA acts as target mimicry with the *miR390-ARFs* module, positively influencing nodule formation (Traubenik et al., 2020) and lateral root growth (Marin et al., 2010). **K.** The master regulator of nitrogen signaling *NLP7* in *Arabidopsis* promote the *T5120* lncRNA transcriptional activity, likely participating in *NRT1.1* gene activation known to be implicated in the nitrogen assimilation (Liu et al., 2019). **L.** *NPC48* and *NPC72* are two *Arabidopsis* lncRNAs whose overexpression inhibits primary root growth (Blein et al., 2020). **M.** The *IPS1* lncRNA act as target mimicry for the *miR399-PHO2* module in *Arabidopsis*, *Medicago*, maize and tomato, participating in the plant Pi homeostasis (Burleigh and Harrison, 1999; Shin et al., 2006; Wang et al., 2017; Franco Zorrilla et al., 2007; Wang et al., 2017). **N.** In rice, an antisense lncRNA of *PHO1;2 (cis-NATPHO1;2)*, promote *PHO2* transcript translation participating in the plant phosphate homeostasis (Burleigh and Harrison, 1999; Shin et al., 2006; Wang et al., 2017).

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Author Contributions

The manuscript was drafted by T.R. and revised by M.C. and T.B..

Acknowledgement

IPS2 benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). We thank Olivier Martin for critical reading of the manuscript and apologize to colleagues whose work could not be cited due to space limitations.

Abbreviations

ABA, abscisic acid; AMT1, AMMONIUM TRANSPORTER 1; APOLO, AUXIN REGULATED PROMOTER LOOP; ARF, AUXIN RESPONSE FACTOR; AS, Alternative Splicing; ASCO, ALTERNATIVE SPLICING COMPETITOR; CASP5, CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN 5 ; CCR, CINNAMOYL-COA REDUCTASE; COOLAIR, COLD INDUCED LONG ANTISENSE INTRAGENIC RNA; COLDAIR, COLD ASSISTED INTRONIC NONCODING RNA; CK, Cytokinin; DRIR, DROUGHT-INDUCED LNCRNA*;* FLC, FLOWERING LOCUS C; IPS1, INDUCED BY PHOSPHATE STARVATION 1; LEA, LATE EMBRYOGENESIS ABUNDANT; lncRNA, LONG NON-CODING RNA; LR, Lateral Root; NRT1.1, miRNA, micro-RNA; N, Nitrogen; NITRATE TRANSPORTER 1.1; NSR, NUCLEAR SPECKLE RNA-BINDING PROTEINS; Pi, Phosphate; PID, PINOID; phasiRNAs, phasing-sirRNAs; PHO1/2, PHOSPHATE1/2; QTL, Quantitative Trait Locus; RGF7, ROOT MERISTEM GROWTH FACTOR 7; RHD6, ROOT HAIR DEFECTIVE 6; RPK2, RECEPTOR-LIKE PROTEIN KINASE 2; RKN, Root-knot nematodes; RdDM, RNA dependent DNA methylation; ROS, Reactive Oxygen Species; SCR, SCARECROW, siRNA, small-interfering RNA; sRNA, small RNA; VIGS, Virus-Induced Gene Silencing, Xist, X-inactive specific transcript; ZAT6, ZINC FINGER OF ARABIDOPSIS THALIANA 6.

References

Alseekh, S., Kostova, D., Bulut, M., and Fernie, A. R. (2021). Genome-wide association studies: assessing trait characteristics in model and crop plants. Cell. Mol. Life Sci. 78:5743–5754.

Amasino, R. (2005). 1955: Kinetin arrives. The 50th anniversary of a new plant hormone. Plant Physiol. 138:1177–1184.

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 77:1055-1065.e4.

Bardou, F., Merchan, F., Ariel, F., and Crespi, M. (2011). Dual RNAs in plants. Biochimie 93:1950–1954.

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.

Bari, R., Pant, B. D., and Stitt, M. (2006). Phosphate-Signaling Pathway in Plants 1 [W][OA]. Society 141:988–999.

Bazin, J., Romero, N., Rigo, R., Charon, C., Blein, T., Ariel, F., and Crespi, M. (2018). Nuclear speckle rna binding proteins remodel alternative splicing and the non-coding arabidopsis transcriptome to regulate a cross-talk between auxin and immune responses. Front. Plant Sci. 9:1–13.

Billou, I., Xu, J., Wildwater, M., Willemsen, V., Paponov, I., Frimi, J., Heldstra, R., Aida, M., Palme, K., and Scheres, B. (2005). The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. Nature 433:39–44.

Bishopp, A., Help, H., and Helariutta, Y. (2009). Chapter 1 Cytokinin Signaling During Root Development. 1st ed. Elsevier Inc.

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.

Burleigh, S. H., and Harrison, M. J. (1999). The down-regulation of Mt4-like genes by phosphate fertilization occurs systemically and involves phosphate translocation to the shoots. Plant Physiol. 119:241–248.

Campalans, A., Kondorosi, A., and Crespi, M. (2004). Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in Medicago truncatula. Plant Cell 16:1047–1059.

Chen, M., Wang, C., Bao, H., Chen, H., and Wang, Y. (2016). Genome-wide identification and characterization of novel lncRNAs in Populus under nitrogen deficiency. Mol. Genet. Genomics 291:1663–1680.

Chen, R., Li, M., Zhang, H., Duan, L., Sun, X., Jiang, Q., Zhang, H., and Hu, Z. (2019). Continuous salt stress-induced long non-coding RNAs and DNA methylation patterns in soybean roots. BMC Genomics 20:1–12.

Chen, Z., Jiang, Q., Jiang, P., Zhang, W., Huang, J., Liu, C., Halford, N. G., and Lu, R. (2020). Novel low-nitrogen stress-responsive long non-coding RNAs (lncRNA) in barley landrace B968 (Liuzhutouzidamai) at seedling stage. BMC Plant Biol. 20:1–11.

Chen, Q., Liu, K., Yu, R., Zhou, B., Huang, P., Cao, Z., Zhou, Y., and Wang, J. (2021). From "Dark Matter" to "Star": Insight Into the Regulation Mechanisms of Plant Functional Long Non-Coding RNAs. Front. Plant Sci. 12:1–13.

Chiou, T. J., Aung, K., Lin, S. I., Wu, C. C., Chiang, S. F., and Su, C. L. (2006). Regulation of phosphate homeostasis by MicroRNA in Arabidopsis. Plant Cell 18:412–421.

Cho, J., and Paszkowski, J. (2017). Regulation of rice root development by a retrotransposon acting as a microRNA sponge. Elife 79:1–21.

Chung, P. J., Jung, H., Jeong, D. H., Ha, S. H., Choi, Y. Do, and Kim, J. K. (2016). Transcriptome profiling of drought responsive noncoding RNAs and their target genes in rice. BMC Genomics 17.

Delhaize, E., and Randall, P. (1993). Characterization of a Phosphate-Accumulator Mutant of A rabidopsis thaliana Advance Access published 1993.

Deng, F., Zhang, X., Wang, W., Yuan, R., and Shen, F. (2018). Identification of Gossypium hirsutum long non-coding RNAs (lncRNAs) under salt stress. BMC Plant Biol. 18:1–14.

Devaiah, B. N., Nagarajan, V. K., and Raghothama, K. G. (2007). Phosphate homeostasis and root development in arabidopsis are synchronized by the zinc finger transcription factor ZAT6. Plant Physiol. 145:147–159.

Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., Zhang, T., Qi, Y., Gerstein, M. B., Guo, Y., et al. (2014). Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. Plant J. 80:848–861.

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J. E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M. G., Feldmann, K. A., and Benfey, P. N. (1996). The SCARECROW gene regulates an asymmetric

cell division that is essential for generating the radial organization of the Arabidopsis root. Cell 86:423–433.

Dong, B., Rengel, Z., and Delhaize, E. (1998). Uptake and translocation of phosphate by pho2 mutant and wild-type seedlings of Arabidopsis thaliana. Planta 205:251–256.

Fok, E. T., Scholefield, J., Fanucchi, S., and Mhlanga, M. M. (2017). The emerging molecular biology toolbox for the study of long noncoding RNA biology. Epigenomics 9:1317–1327.

Forde, B. G. (2014). Nitrogen signalling pathways shaping root system architecture: An update. Curr. Opin. Plant Biol. 21:30–36.

Forde, B. G., and Walch-Liu, P. (2009). Nitrate and glutamate as environmental cues for behavioural responses in plant roots. Plant, Cell Environ. 32:682–693.

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J. A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat. Genet. 39:1033–1037.

Fukuda, M., Fujiwara, T., and Nishida, S. (2020). Roles of non-coding rnas in response to nitrogen availability in plants. Int. J. Mol. Sci. 21:1–15.

Giannakou, I. O., and Panopoulou, S. (2019). The use of fluensulfone for the control of root-knot nematodes in greenhouse cultivated crops: Efficacy and phytotoxicity effects. Cogent Food Agric. 5:1643819.

Golicz, A. A., Singh, M. B., and Bhalla, P. L. (2018). The long intergenic noncoding RNA (LincRNA) Landscape of the soybean genome. Plant Physiol. 176:2133–2147.

Gultyaev, A. P., and Roussis, A. (2007). Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. Nucleic Acids Res. 35:3144–3152.

Gutiérrez-Alanís, D., Yong-Villalobos, L., Jiménez-Sandoval, P., Alatorre-Cobos, F., Oropeza-Aburto, A., Mora-Macías, J., Sánchez-Rodríguez, F., Cruz-Ramírez, A., and Herrera-Estrella, L. (2017). Phosphate Starvation-Dependent Iron Mobilization Induces CLE14 Expression to Trigger Root Meristem Differentiation through CLV2/PEPR2 Signaling. Dev. Cell 41:555-570.e3.

Hu, Y., Omary, M., Hu, Y., Doron, O., Hoermayer, L., Chen, Q., Megides, O., Chekli, O., Ding, Z., Friml, J., et al. (2021). Cell kinetics of auxin transport and activity in Arabidopsis root growth and skewing. Nat. Commun. 12:1–13.

Ishiyama, K., Inoue, E., Tabuchi, M., Yamaya, T., and Takahashi, H. (2004). Biochemical background and compartmentalized functions of cytosolic glutamine synthetase for active ammonium assimilation in rice roots. Plant Cell Physiol. 45:1640–1647.

Jabnoune, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., and Poirier, Y. (2013). A Rice cis-Natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. Plant Cell 25:4166–4182.

Kim, W., Ahn, H. J., Chiou, T. J., and Ahn, J. H. (2011). The role of the miR399-PHO2 module in the regulation of flowering time in response to different ambient temperatures in Arabidopsis thaliana. Mol. Cells 32:83–88.

Kistner, C., and Parniske, M. (2002). Evolution of signal transduction in intracellular symbiosis. Trends Plant Sci. 7:511–518.

Kumagai, H., Kinoshita, E., Ridge, R. W., and Kouchi, H. (2006). RNAi knock-down of ENOD40s leads to significant suppression of nodule formation in Lotus japonicus. Plant Cell Physiol. 47:1102–1111.

Li, S., Yu, X., Lei, N., Cheng, Z., Zhao, P., He, Y., Wang, W., and Peng, M. (2017). Genome-wide identification and functional prediction of cold and/or drought-responsive lncRNAs in cassava. Sci. Rep. 7.

Li, X., Xing, X., Xu, S., Zhang, M., Wang, Y., Wu, H., Sun, Z., Huo, Z., Chen, F., and Yang, T. (2018). Genome-wide identification and functional prediction of tobacco lncRNAs responsive to root-knot nematode stress. PLoS One 13.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3′Processing of AntisenseTranscripts TriggersArabidopsis FLC Chromatin Silencing. Science (80-.). 327:94–97.

Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C., and Chua, N. H. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in arabidopsisC W. Plant Cell 24:4333–4345.

Liu, H., Luo, X., Niu, L., Xiao, Y., Chen, L., Liu, J., Wang, X., Jin, M., Li, W., Zhang, Q., et al. (2017). Distant eQTLs and Non-coding Sequences Play Critical Roles in Regulating Gene Expression and Quantitative Trait Variation in Maize. Mol. Plant 10:414–426.

Liu, F., Xu, Y., Chang, K., Li, S., Liu, Z., Qi, S., Jia, J., Zhang, M., Crawford, N. M., and Wang, Y. (2019). The long noncoding RNA T5120 regulates nitrate response and assimilation in Arabidopsis. New Phytol. 224:117–131.

Lv, Y., Liang, Z., Ge, M., Qi, W., Zhang, T., Lin, F., Peng, Z., and Zhao, H. (2016). Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (Zea mays L.). BMC Genomics 17:1–15.

Ma, J., Bai, X., Luo, W., Feng, Y., Shao, X., Bai, Q., Sun, S., Long, Q., and Wan, D. (2019). Genome-Wide Identification of Long Noncoding RNAs and Their Responses to Salt Stress in Two Closely Related Poplars. Front. Genet. 10:1–13.

Ma, P., Zhang, X., Luo, B., Chen, Z., He, X., Zhang, H., Li, B., Liu, D., and Wu, L. (2021). Transcriptomic and genome-wide association study reveal long noncoding RNAs responding to nitrogen deficiency in maize Advance Access published 2021.

Marin, E., Jouannet, V., Herz, A., Lokerse, A. S., Weijers, D., Vaucheret, H., Nussaume, L., Crespi, M. D., and Maizel, A. (2010). mir390, Arabidopsis TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. Plant Cell 22:1104–1117.

Moison, M., Pacheco, J. M., Lucero, L., Fonouni-Farde, C., Rodríguez-Melo, J., Mansilla, N., Christ, A., Bazin, J., Benhamed, M., Ibañez, F., et al. (2021). The lncRNA APOLO interacts with the transcription factor WRKY42 to trigger root hair cell expansion in response to cold. Mol. Plant 14:937–948.

Mok, M. C., Martin, R. C., and Mok, D. W. S. (2000). Cytokinins: Biosynthesis, metabolism and perception. Vitr. Cell. Dev. Biol. - Plant 36:102–107.

Nordström, A., Tarkowski, P., Tarkowska, D., Norbaek, R., Åstot, C., Dolezal, K., and Sandberg, G. (2004). Auxin regulation of cytokinin biosynthesis in Arabidopsis thaliana: A factor of potential importance for auxin-cytokinin-regulated development. Proc. Natl. Acad. Sci. U. S. A. 101:8039–8044.

Pant, B. D., Buhtz, A., Kehr, J., and Scheible, W. R. (2008). MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. Plant J. 53:731–738.

Poirier, Y., and Bucher, M. (2002). Phosphate Transport and Homeostasis in Arabidopsis. Arab. B. 1:e0024.

Qi, X., Xie, S., Liu, Y., Yi, F., and Yu, J. (2013). Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. Plant Mol. Biol. 83:459–473.

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.

Reymond, M., Svistoonoff, S., Loudet, O., Nussaume, L., and Desnos, T. (2006). Identification of QTL controlling root growth response to phosphate starvation in Arabidopsis thaliana. Plant. Cell Environ. 29:115–125.

Rigo, R., Bazin, J., Romero-Barrios, N., Moison, M., Lucero, L., Christ, A., Benhamed, M., Blein, T., Huguet, S., Charon, C., et al. (2020). The Arabidopsis lnc RNA ASCO modulates the transcriptome through interaction with splicing factors . EMBO Rep. 21:1–19.

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

Sarropoulos, I., Marin, R., Cardoso-Moreira, M., and Kaessmann, H. (2019). Developmental dynamics of lncRNAs across mammalian organs and species. Nature 571:510–514.

Shin, H., Shin, H. S., Chen, R., and Harrison, M. J. (2006). Loss of At4 function impacts phosphate distribution between the roots and the shoots during phosphate starvation. Plant J. 45:712–726.

Shin, S. Y., Jeong, J. S., Lim, J. Y., Kim, T., Park, J. H., Kim, J. K., and Shin, C. (2018). Transcriptomic analyses of rice (Oryza sativa) genes and non-coding RNAs under nitrogen starvation using multiple omics technologies. BMC Genomics 19:1–20.

Shuai, P., Liang, D., Tang, S., Zhang, Z., Ye, C. Y., Su, Y., Xia, X., and Yin, W. (2014). Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in Populus trichocarpa. J. Exp. Bot. 65:4975–4983.

Song, Y., Xuan, A., Bu, C., Ci, D., Tian, M., and Zhang, D. (2019). Osmotic stress-responsive promoter upstream transcripts (PROMPTs) act as carriers of MYB transcription factors to induce the expression of target genes in Populus simonii. Plant Biotechnol. J. 17:164–177.

Stagnati, L., Lanubile, A., Samayoa, L. F., Bragalanti, M., Giorni, P., Busconi, M., Holland, J. B., and Marocco, A. (2019). A genome wide association study reveals markers and genes associated with resistance to fusarium verticillioides infection of seedlings in a maize diversity panel. G3 Genes, Genomes, Genet. 9:571–579.

Statello, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol. 22:96–118.

Sukumar, P., Edwards, K. S., Rahman, A., DeLong, A., and Muday, G. K. (2009). PINOID kinase regulates root gravitropism through modulation of PIN2-dependent basipetal auxin transport in arabidopsis. Plant Physiol. 150:722–735.

Sun, X., Zheng, H., Li, J., Liu, L., Zhang, X., and Sui, N. (2020). Comparative Transcriptome Analysis Reveals New lncRNAs Responding to Salt Stress in Sweet Sorghum. Front. Bioeng. Biotechnol. 8:1–14.

Swiezewski, S., Crevillen, P., Liu, F., Ecker, J. R., Jerzmanowski, A., and Dean, C. (2007). Small RNA-mediated chromatin silencing directed to the 3′ region of the Arabidopsis gene encoding the developmental regulator, FLC. Proc. Natl. Acad. Sci. U. S. A. 104:3633–3638.

Traubenik, S., Reynoso, M. A., Hobecker, K., Lancia, M., Hummel, M., Rosen, B., Town, C., Bailey-Serres, J., Blanco, F., and Zanetti, M. E. (2020). Reprogramming of root cells during nitrogen-fixing symbiosis involves dynamic polysome association of coding and noncoding RNAs. Plant Cell 32:352–373.

Walch-Liu, P., Ivanov, I. I., Filleur, S., Gan, Y., Remans, T., and Forde, B. G. (2006). Nitrogen regulation of root branching. Ann. Bot. 97:875–881.

Wan, S., Zhang, Y., Duan, M., Huang, L., Wang, W., Xu, Q., Yang, Y., and Yu, Y. (2020). Integrated Analysis of Long Non-coding RNAs (lncRNAs) and mRNAs Reveals the Regulatory Role of lncRNAs Associated With Salt Resistance in Camellia sinensis. Front. Plant Sci. 11:1–14.

Wang, T. Z., Liu, M., Zhao, M. G., Chen, R., and Zhang, W. H. (2015). Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in Medicago truncatula using genome-wide high-throughput sequencing. BMC Plant Biol. 15:1–13.

Wang, T., Zhao, M., Zhang, X., Liu, M., Yang, C., Chen, Y., Chen, R., Wen, J., Mysore, K. S., and Zhang, W. H. (2017). Novel phosphate deficiency-responsive long non-coding RNAs in the legume model plant Medicago truncatula. J. Exp. Bot. 68:5937–5948.

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.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science (80-.). 322:750–756.

Zhao, J., Ajadi, A. A., Wang, Y., Tong, X., Wang, H., Tang, L., Li, Z., Shu, Y., Liu, X., Li, S., et al. (2020a). Genome-Wide Identification of lncRNAs During Rice Seed Development. Genes (Basel). 11.

Zhao, Y., Liu, X., Tong, C., and Wu, Y. (2020b). Effect of root interaction on nodulation and nitrogen fixation ability of alfalfa in the simulated alfalfa/triticale intercropping in pots. Sci. Rep. 10:1–11.

Zubko, E., and Meyer, P. (2007). A natural antisense transcript of the Petunia hybrida Sho gene suggests a role for an antisense mechanism in cytokinin regulation. Plant J. 52:1131–1139.

4. Aim of the thesis

Long non-coding RNAs are implicated in a wide-range of plant-developmental and stress-related processes, rapidly emerging as novel regulators of gene expression (Chen et al., 2021). At the start of my PhD, the host team was interested in exploring the diversity of non-coding transcripts for the plant adaptation to environmental constraints.

In this context, I started by studying the conservation of the non-coding transcriptome using two Arabidopsis accessions, under a Pi starvation time-course experiment. These two accessions present a striking difference in root growth architecture in response to Pi-deprivation. Whereas the Col-0 accession immediately stops its primary root growth upon Pi starvation, the L*er* accession does not. The first chapter describes our efforts to compare the features of the non-coding genes as compared to coding genes along these differential responses, shedding light on the unique specificity of the non-coding transcriptome among accessions. I also contributed to the identification of two lncRNAs able to modulate root growth. In the second chapter, I investigate the transcriptomics data emerging from our accession Pi starvation analyses to identify new lncRNAs implicated in root growth and development. Finally, the third chapter constitutes the major part of my thesis work and describes the molecular characterization of an unknown lncRNA implicated in the epigenetic-mediated regulation of a group of genes organized in clusters.

Overall, my thesis work aimed to uncover the implication of the non-coding transcriptomes in the local plant adaptation to its environment, together with finding and characterizing new lncRNAs implicated in the regulation of plant-root-related genes.

II Results

5. The non-coding transcriptome from two *Arabidopsis* ecotypes

Genome conservation is not uniformly distributed along eukaryotic genomes. In humans and plants, it has been shown that coding genes are more frequently conserved than non-coding genes at the sequence level (Sudmant et al., 2015; Alonso-Blanco et al., 2016). Even within the same species, the conservation of non-coding genes, such as lncRNAs, can vary drastically. For example, among the *Arabidopsis* accessions, very few Single Nucleotide Polymorphisms (SNPs) are found in coding mRNAs compared to non-coding genes (Alonso-Blanco et al., 2016). Similarly, in rice less than 4% of all the SNPs and insertion/deletions (indels) were found in coding regions (Zhao et al., 2018b). Even though non-coding genes are poorly conserved at sequence level, their position in relation to neighboring genes within the genome (synteny) is more conserved suggesting that their position more than their sequence could participate in their regulatory function (Mohammadin et al., 2015). Indeed, the act of transcription could be involved in their regulatory activity (Kopp and Mendell, 2018). The specificity of the non-coding genome between organisms of the same species could also reflect the presence of regulatory mechanisms involved in the local adaptation of an individual to its environment. For example, plant response to the lack of phosphate (Pi) drastically changed between ecotypes. Indeed, Col-0 limits rapidly its main root growth through a Fe-dependent pathway involving *LOW PHOSPHATE ROOT* (*LPR1*), *PHOSPHATE DEFICIENCY RESPONSE2 (PDR2), SENSITIVE TO PROTON RHIZOTOXICITY1* (*STOP1*), *and ALUMINUM-ACTIVATED MALATE TRANS- PORTER1* (*ALMT1*); whereas L*er* root growth continues under low Pi condition and is similar to the length observed in a phosphate-containing medium (Reymond et al., 2006). In this context, we thought of interest to characterize the noncoding transcriptomes of Col-0 and L*er* ecotypes under early Pi starvation. Using RNAseq and bioinformatics analysis, we identified many new non-coding genes conserved, enriched, or even specific of one ecotype, in addition to being responsive, or not, to the lack of Pi. We also analyze their links with different classes of small RNAs. Finally, we functionally characterized several of the new intergenic lncRNAs and found that two ecotype-enriched lncRNAs were

involved in primary root growth regulation. These results strengthened that lncRNAs participate in the transcriptomic response to environmental cues and suggest that they may play roles in ecotype adaptation to the soil environment.

In this first chapter, I will present the publication emerging from this study where I contributed by performing qPCR to validate the RNA-seq results, investigate *trans* regulatory effects of lncRNA expression between Col-0 and L*er* genomes and better understand the function of some ecotype-enriched lncRNAs for the root growth and Pi starvation response. Finally, I participate in the writing process and exchange ideas with the reviewers.

5.1 Publication: Landscape of the Noncoding Transcriptome Response of Two Arabidopsis Ecotypes to Phosphate Starvation

Landscape of the Noncoding Transcriptome Response of Two Arabidopsis Ecotypes to Phosphate Starvation[1]

Thomas Blein,[a,b] Coline Balzergue,[c,2] Thomas Roulé,[a,b,2] Marc Gabriel,[e] Laetitia Scalisi,[a,b] Tracy François,[a,b] Céline Sorin,[a,b] Aurélie Christ,[a,b] Christian Godon,[c] Etienne Delannoy,[a,b] Marie-Laure Martin-Magniette,[a,b,d] Laurent Nussaume,[c] Caroline Hartmann,[a,b] Daniel Gautheret,[e] Thierry Desnos,[c] and Martin Crespi[a,b,3,4]

[a]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
[b]Institute of Plant Sciences Paris-Saclay, Université de Paris, 91405 Orsay, France
[c]Aix Marseille University, Commisariat à l'Énergie Atomique, Centre Nationale de la Recherche, Bioscience and Biotechnology Institute of Aix-Marseilles, Unité Mixte de Recherche 7265 Signalisation pour l'Adaptation des Végétaux à leur Environnement (UMR7265 SAVE), 13108 Saint Paul-Lez-Durance, France
[d]Unité Mixte de Recherche MIA-Paris (UMR MIA-Paris), AgroParisTech, Institut National de la Recherche Agronomique, Université Paris-Saclay, 75005 Paris, France
[e]Institute for Integrative Biology of the Cell, Commisariat à l'Énergie Atomique, Centre Nationale de la Recherche, Université Paris Sud, 91198 Gif sur Yvette, France

ORCID IDs: 0000-0001-9788-5201 (T.B.); 0000-0001-6661-9357 (T.R.); 0000-0002-6294-2365 (T.F.); 0000-0002-0866-2063 (E.D.); 0000-0003-4000-9600 (M.-L.M.-M.); 0000-0002-9445-2563 (L.N.); 0000-0003-1071-7868 (C.H.); 0000-0002-1508-8469 (D.G.); 0000-0002-6585-1362 (T.D.); 0000-0002-5698-9482 (M.C.).

Root architecture varies widely between species; it even varies between ecotypes of the same species, despite strong conservation of the coding portion of their genomes. By contrast, noncoding RNAs evolve rapidly between ecotypes and may control their differential responses to the environment, since several long noncoding RNAs (lncRNAs) are known to quantitatively regulate gene expression. Roots from ecotypes Columbia and Landsberg *erecta* of Arabidopsis (*Arabidopsis thaliana*) respond differently to phosphate starvation. Here, we compared transcriptomes (mRNAs, lncRNAs, and small RNAs) of root tips from these two ecotypes during early phosphate starvation. We identified thousands of lncRNAs that were largely conserved at the DNA level in these ecotypes. In contrast to coding genes, many lncRNAs were specifically transcribed in one ecotype and/or differentially expressed between ecotypes independent of phosphate availability. We further characterized these ecotype-related lncRNAs and studied their link with small interfering RNAs. Our analysis identified 675 lncRNAs differentially expressed between the two ecotypes, including antisense RNAs targeting key regulators of root-growth responses. Misregulation of several lincRNAs showed that at least two ecotype-related lncRNAs regulate primary root growth in ecotype Columbia. RNA-sequencing analysis following deregulation of lncRNA NPC48 revealed a potential link with root growth and transport functions. This exploration of the noncoding transcriptome identified ecotype-specific lncRNA-mediated regulation in root apexes. The noncoding genome may harbor further mechanisms involved in ecotype adaptation of roots to different soil environments.

Over the last decade, genome-wide transcriptomics has revealed that a large intergenic part of eukaryotic genomes is transcribed. These transcripts, globally known as noncoding RNAs (Ariel et al., 2015), can regulate genome expression at transcriptional, post-transcriptional, and epigenetic levels, and are generally classified as small (21–24 nucleotides [nt]), long (>200 nt, <100 kb), and circular noncoding RNAs. Plant small RNAs (sRNAs) are produced by processing longer noncoding transcripts that generally contain a hairpin structure or lead to double-strand RNA formation. Plant sRNAs include microRNAs (miRNAs), endogenous small interfering RNAs (siRNAs; generally 21–22 nt long), and, most abundantly, heterochromatin siRNA (24 nt long; Borges and Martienssen, 2015). On the other hand, long noncoding RNAs (lncRNAs) are a heterogeneous group of RNA molecules with a coding capacity <50 amino acids (Chekanova, 2015). lncRNA transcripts are generally polyadenylated and can be intergenic (lincRNAs), intronic, or natural antisense (NATs) with respect to protein-coding genes (Ariel et al., 2015). When compared to mRNAs, lncRNAs are expressed at low levels in a tissue-specific manner or in response to environmental stresses (Liu et al., 2012) and are more frequently accumulated in the nucleus (Derrien et al., 2012), where they can regulate nuclear organization or function (Ariel et al., 2015).

lncRNAs utilize both cis- and trans-modalities of action to regulate gene expression through interactions with ribonucleoproteins and can form scaffolds and/or sequester proteins or RNA molecules as decoys or sponges. However, molecular functions have only been identified for a few lncRNAs in plants. As lncRNA genes lack regions with high primary sequence constraints

(Derrien et al., 2012), it is difficult to use sequence conservation to identify potential functions. Even though the sequences of lncRNAs are not particularly conserved between plant species, they may show relative positional conservation in genomes (Mohammadin et al., 2015).

Finally, lncRNAs could be simply transcriptional byproducts; in this framework, the sole act of their transcription rather than their sequence per se would be the source of the regulatory activity (Kopp and Mendell, 2018).

Resequencing approaches in model species have allowed the determination of whole-genome variations and evolution, from which it has been possible to provide the characterization of pan-genomes composed of "core" genomes (present in all accessions) and "dispensable" genomes (those specific to two or more accessions or even unique sequences specific to only one accession). Core genes are frequently highly expressed whereas dispensable genes are variably expressed, and generally in a tissue-specific manner (Contreras-Moreira et al., 2017). The dispensable genomes may play important roles in the capacity of individual organisms to cope with environmental conditions (Vernikos et al., 2015). Indeed, identification of natural variations in large worldwide populations (accessions) of Arabidopsis (*Arabidopsis thaliana*) showed an average of one SNP per 10 bp more frequently located in intergenic regions than in coding mRNAs (The 1001 Genomes Consortium, 2016). This has also been observed recently in rice (*Oryza sativa*), for which only 3.5% of SNPs and 2.5% of small insertion-deletions (InDels) were located in coding regions (Zhao et al., 2018). This latter observation would explain why lncRNAs differ even between closely related plant species (Nelson et al., 2017). In plants, three mechanisms have been proposed for the origin of lncRNAs: evolution from transposable element (TE) sequences, pseudogenization of protein-coding gene sequences, or duplication of preexisting lncRNAs (Kapusta and Feschotte, 2014)

The inorganic phosphate (Pi) accumulated in the upper soil layer is perceived by plants at the root apex (Svistoonoff et al., 2007). Accessions of ecotypes Columbia (Col) and Landsberg *erecta* (L*er*) of Arabidopsis display different primary root growth and architecture in response to Pi starvation (Reymond et al., 2006). The identification of LOW PHOSPHATE ROOT1 (LPR1), a major quantitative trait locus, has been done in recombinant inbred lines obtained by crosses of accessions presenting this opposite root response to low Pi and it has been linked to differential expression of *LPR1* in root apexes (Reymond et al., 2006; Svistoonoff et al., 2007). When the primary root tip of a Col seedling encounters a low-Pi medium, cell elongation in the transition zone rapidly decreases and cell proliferation in the root apical meristem (RAM) progressively ceases as callose deposition occurs in RAM plasmodesmata (Müller et al., 2015; Abel, 2017; Gutiérrez-Alanís et al., 2018). Root growth inhibition in low Pi depends on iron (Fe) availability in soil or media (Svistoonoff et al., 2007; Ward et al., 2008), as Fe concentrations clearly increase in Col plants during Pi starvation (Misson et al., 2005; Hirsch et al., 2006; Baxter et al., 2008). Indeed, inhibition of cell elongation and the RAM arrest are Fe-dependent (Svistoonoff et al., 2007; Ward et al., 2008; Müller et al., 2015; Abel, 2017; Balzergue et al., 2017; Mora-Macías et al., 2017; Gutiérrez-Alanís et al., 2018). Interestingly, in low-Pi conditions, Fe accumulates in the elongation zone, but not in the RAM, and more generally, in Col plants, Fe is redistributed among tissues (Mora-Macías et al., 2017; Gutiérrez-Alanís et al., 2018). By contrast, in L*er* seedlings that are subject to low Pi, elongation and proliferation of root cells in the root apex continue, thereby sustaining root growth (Reymond et al., 2006). The corresponding regulatory system controlling root inhibition involves LPR1, PHOSPHATE DEFICIENCY RESPONSE2 (PDR2), SENSITIVE TO PROTON RHIZOTOXICITY1(STOP1), and ALUMINUM-ACTIVATED MALATE TRANSPORTER1 (ALMT1). The LPR1-PDR2 and STOP1-ALMT1 modules allow Fe accumulation in roots under low Pi (Ticconi et al., 2009; Abel, 2017; Balzergue et al., 2017; Gutiérrez-Alanís et al., 2018). From these results, one concludes that interactions between Pi and Fe determine the differential growth response of Col and L*er* ecotypes. In this work, we identified and characterized the noncoding transcriptomes of Col and L*er* root apexes during early Pi starvation responses. Thousands of Arabidopsis lncRNAs, notably in the L*er* accession, were identified, with only a minor fraction linked to sRNA production. Several "ecotype-specific" or "ecotype-enriched" variants were highly conserved at the DNA level and showed expression variation correlated with changes in the expression of key regulators of the Pi-starvation response. Functional analysis of five lncRNAs in Col revealed two further regulators of primary root growth, allowing us to hypothesize that

[1]This work was supported by the Agence Nationale pour la Recherche (grant nos. ANR–12–ADAP–0019 [RNAdapt], ANR–16–CE12–0032 [SPLISIL], and ANR–17–EUR–0007 [Saclay Plant Sciences Graduate School of Research] managed under an Investments for the Future program [grant no. ANR–11–IDEX–0003–02]) and The King Abdulla University of Science and Technology (KAUST) International Program (grant no. OCRF–2014–CRG4).

[2]These authors contributed equally to the article.

[3]Author for contact: martin.crespi@ips2.universite-paris-saclay.fr.

[4]Senior author.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Martin Crespi (martin.crespi@ips2.universite-paris-saclay.fr).

T.B., C.B., T.R., L.S., T.F., and C.G. performed gene expression analysis and root phenotyping; T.B., M.G., D.G., E.D., and M.-L.M.-M. performed statistical analysis and bioinformatics; C.B., C.S., and A.C. were involved in sample preparation and processing; T.B., C.H., T.D., L.N., and M.C. directed experimental work; T.B., C.H., T.D., and M.C. designed experiments and wrote the manuscript; and all authors provided comments on the article and approved it.

www.plantphysiol.org/cgi/doi/10.1104/pp.20.00446

lncRNA expression patterns contribute to the modulation of environmental responses in different ecotypes.

RESULTS

Col and L*er* Root Tip Transcriptome Assemblies

We characterized the root-tip transcriptome of Col and L*er* ecotypes, which present contrasting root phenotypes in response to Pi deficiency (Reymond et al., 2006). We performed comparative whole-genome transcriptomic analyses using paired-end sequencing of three biological replicates of root tips during a short kinetics (0, 1, and 2 h) of low (10 μM) Pi treatment (Supplemental Table S1). To avoid possible differences related to the *erecta* mutation present in the L*er* ecotype, we used the Col^{er105} mutant here. For each ecotype, the reads were independently mapped to their reference genome (Supplemental Fig. S1A): TAIR10 for Col (Lamesch et al., 2012) and L*er* v7 for L*er* (Gan et al., 2011) that shared the same TAIR10 annotation (unlike L*er* v8; Zapata et al. [2016]). We predicted previously unannotated transcripts by comparing our data to TAIR10. The homology of these predicted transcripts in Col and L*er* was determined by mapping them onto the other genome. We retained as transcripts only RNA molecules of at least 200 nt. When these previously uncharacterized transcripts overlapped with pre-existing annotations, fusions were generated. Transcripts identified by this pipeline (Supplemental Fig. S1A) were compared with those from different Arabidopsis databases: Araport 11 (Cheng et al., 2017), RepTas (Liu et al., 2012), CANTATAdb (Szcześniak et al., 2016), miRBase v21 (Kozomara and Griffiths-Jones, 2014), and with those from two previous studies concerning lncRNAs (Ben Amor et al., 2009; Li et al., 2016). Finally, we used COME software to determine the potential coding capacity of identified transcripts (Hu et al., 2017). On the basis of both database information and COME predictions we classified the corresponding genes as coding or noncoding.

In total, we identified 5,313 and 6,408 previously uncharacterized putative genes in Col and L*er* ecotypes,

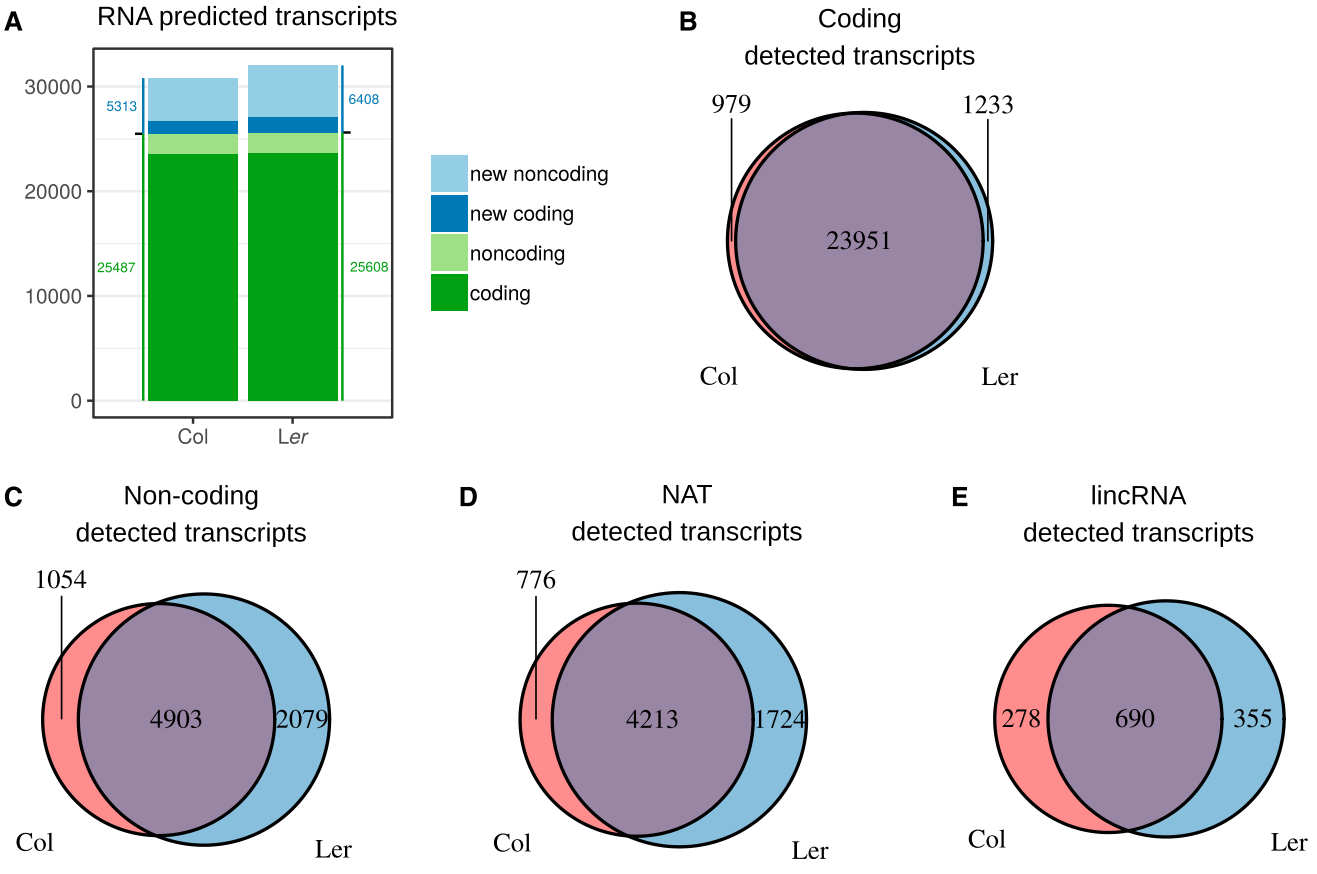


Figure 1. Identification of the transcripts and their occurrence across the two ecotypes. A, Number of predicted coding and noncoding transcripts in the two ecotypes, classified by type. New transcripts refers to genes not characterized in previously published studies. B to E, Predicted transcripts in each ecotype were classified as coding (B) or noncoding (C). For the latter case, two subclasses are defined: antisense of another annotation (NAT; D) and intergenic (lincRNA; E). In contrast to coding genes, many noncoding RNAs, notably lincRNAs, were detected only in one ecotype despite the high DNA sequence similarity in both ecotypes.

respectively (Fig. 1A; Supplemental Table S2; Supplemental Files S1 and S2). In root apexes, these identified genes were predominantly noncoding RNAs: 76% and 77% of the total previously uncharacterized genes in the case of Col and L*er*, respectively (Fig. 1A; Supplemental Fig. S1, B and C). As expected, noncoding genes were globally less expressed than coding genes (Supplemental Fig. S2, A and B). Genes specifically detected in one ecotype belong much more often to the noncoding (>40% of the total noncoding genes) than to the coding class (<8% of the total coding genes; Fig. 1, B and C), notably true for lincRNA genes (52% of the total lncRNAs) as compared to NATs (34% of the total NATs; Fig. 1, D and E). Overall, expression of noncoding genes is more ecotype specific than that of coding genes.

We detected a greater number of previously uncharacterized genes in the L*er* ecotype (Fig. 1, A and C). Such differences do not result from library sequencing saturation. Indeed, in the last 2% of sequencing reads, <10 additional genes of this type were detected (Supplemental Fig. S2C). Thus, sequencing was deep enough to detect expressed genes, and the difference in gene detection between Col and L*er* does not result from a sequencing bias.

We next sought to determine whether these previously uncharacterized detected genes expressed in L*er* (coding or noncoding) could correspond to specific parts of the L*er* genome missed or rearranged in the Col genome. Out of these 7,357 genes, only 41 and 53 genes in Col and L*er*, respectively, coincided with missing DNA sequences in the other ecotype (Fig. 2A), showing that the DNA sequence of the different previously uncharacterized genes is largely conserved apart from a few SNPs. Thus, the ecotype differences in transcript accumulation came from a shift in transcription that could be due to the deregulation of gene regulators, the accumulation of small sequence differences in promoters, or to specific differences in epigenetic status in the lncRNA-producing region due to TE insertions or other rearrangements possibly at large distance from the differentially expressed loci.

Evolutionary Analysis of lncRNA Genes Expressed in Root Tips

Mechanisms modifying root architecture result from local signaling that occurs at the root tip (Svistoonoff et al., 2007; Thibaud et al., 2010; Müller et al., 2015; Balzergue et al., 2017). We therefore characterized the Arabidopsis genes expressed in the root apex, taking advantage of the extensive sequence information in Arabidopsis accessions (The 1001 Genomes Consortium, 2016). According to current annotations, genes were considered as non-NAT (no gene on the other strand) or NAT (presence of a gene on the other strand). For Arabidopsis species, we calculated the rate of SNPs accumulated in the different types of genes among all accessions (Fig. 2B). As expected, TEs accumulated many more SNPs than coding genes, whereas non-NAT lncRNAs and structural RNA genes showed an intermediate level of SNPs between TEs and coding genes. By contrast, the amount of SNPs was generally similar for NAT lncRNA and coding genes, as can be justified by the fact that the coding regions are under strong selection pressure.

To investigate sequence evolution at a larger scale, we used the PhastCons score that represents an interspecies

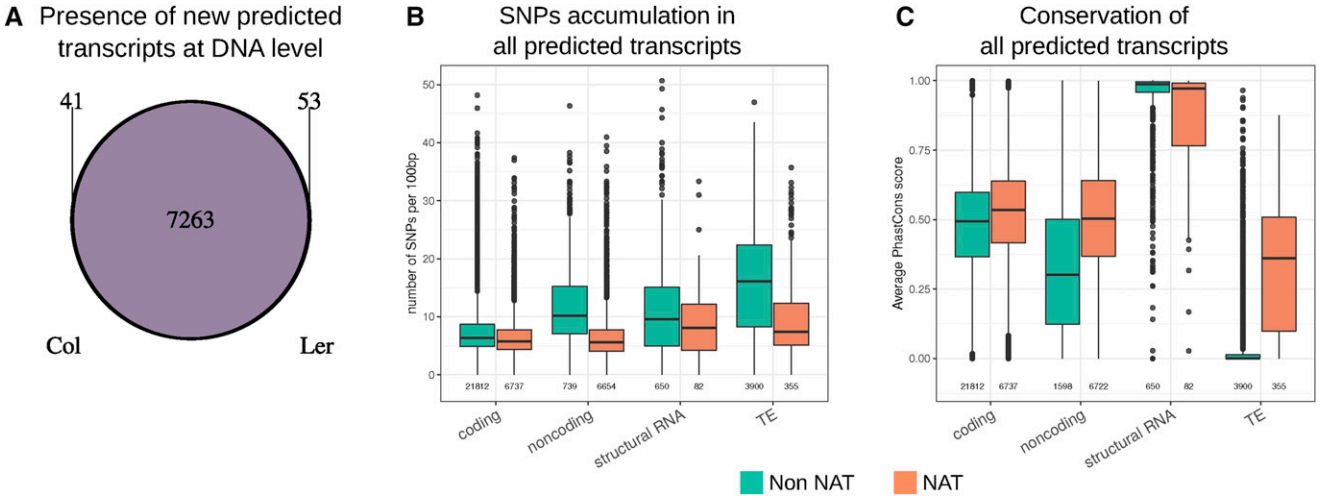


Figure 2. Characterization of transcripts at the DNA level. A, Detection of the DNA sequences of previously uncharacterized predicted transcripts in the two ecotypes (minimum of 90% sequence identity along 90% of the RNA length). The large majority of RNAs come from common DNA regions from both ecotypes. B, SNP accumulation per 100 bp of transcript length for each type of transcript according to data from The 1001 Genomes Project (The 1001 Genomes Consortium, 2016). C, Conservation among plant species (average PhasCons score) of each type of transcript according to genomic position in relation to other annotations. In B and C, Non-NAT refers to transcripts which do not overlap with annotations on the other DNA strands, independent of annotation type (coding, noncoding, and structural RNA or TE).

level of nt conservation (normalized between 0 and 1) according to the alignment of 20 angiosperm genomes (Fig. 2C; Hupalo and Kern, 2013). As expected, structural RNA genes were strongly conserved (median score of 1), whereas TEs were not (median score of 0). Coding genes presented a score between these two extremes (~0.5). Interestingly, non-NAT lncRNA genes showed an intermediate score between those of coding and transposable genes (median score of ~0.3), whereas NAT lncRNA genes again showed the same degree of conservation as coding genes. These observations further suggest that NAT lncRNA genes are strongly constrained, whereas intergenic noncoding genes allow more variability even though they are more constrained than TEs.

Few lncRNA Transcripts Colocalize with Small RNA-Generating Loci

In animals and plants, some lncRNA loci colocalize with regions producing sRNA molecules (Matzke and Mosher, 2014). Therefore, we asked whether the lncRNA loci identified could generate sRNAs. Using similar samples previously prepared for the lncRNA studies, we prepared sRNA libraries for each ecotype and sequenced sRNAs to obtain a full description of the sRNAome mapped on each ecotype genome. Only a minority of the lncRNAs accumulate sRNAs, but of those, most contained sequences capable of generating nonphased RNA molecules of 21/22 or 24 nt (Supplemental Fig. S3, A and B) and only a small fraction of lncRNAs overlapped with phased siRNAs or were miRNA precursors.

We then analyzed the potential link between siRNAs and lncRNAs in each ecotype. The majority of lncRNAs (6,452 genes of 7,850 detected) did not lead to accumulation of siRNAs. This is also true for the lncRNAs specifically detected only in one ecotype (2,688 genes out of 3,110 ecotype-specific genes), since many of them did not generate any siRNA in either ecotype (long in Col and not detected [ND] in L*er*, or vice versa; Fig. 3A). Thus, the differential detection of lncRNA between ecotypes could not be linked to a change in the processing of siRNA by the encoding lncRNA loci.

We then wondered whether different sRNA processing by lncRNAs could occur between the two ecotypes. We first looked at lncRNAs that accumulated sRNAs in only one accession: (1) lncRNAs that could

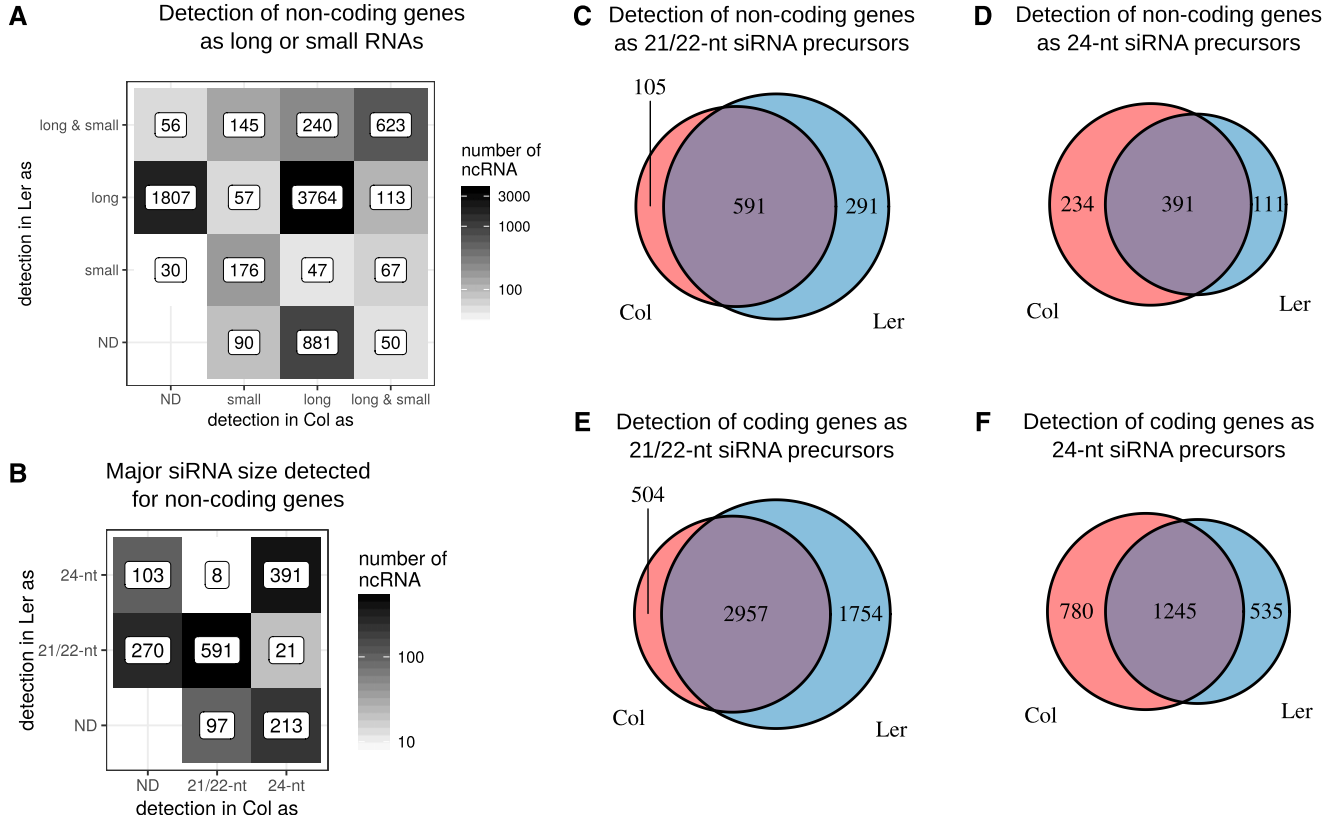


Figure 3. LncRNAs as sRNA precursors. The major specificity difference between Col and L*er* is the lncRNA component of the transcriptome. A, Identification of noncoding transcripts as siRNAs or long RNAs. B, Distribution of the major siRNA sizes for noncoding transcripts detected as long in the two ecotypes. There is no major change of siRNA size between the two accessions. C, Detection of noncoding RNAs as 21-nt and 22-nt siRNA precursors. D, Detection of noncoding RNA as 24-nt siRNA precursors. E, Detection of coding RNA as 21-nt and 22-nt siRNAs precursors. F, Detection of coding RNA as 24-nt siRNAs precursors. Detection threshold for small RNA set at 1 read per million. ND, Not detected; ncRNA, noncoding RNA.

generate siRNAs in only one ecotype while being detected as lncRNAs in both ecotypes (long in L*er* and long and small in Col [113 genes], and vice versa [240 genes]); and (2) loci that produce siRNAs only in one ecotype yet are detected as lncRNAs only in the other (long in L*er* and small in Col [57 genes], or vice versa [47 genes]). It is known that 21/22-nt siRNAs act on gene transcripts, whereas 24-nt siRNAs mediate chromatin modifications (Matzke and Mosher, 2014). Thus, a difference in the size of accumulated siRNAs for a given gene in one ecotype could indicate a modification of posttranscriptional (21/22 nt) or epigenetic (24 nt) regulation in the other ecotype. Among the lncRNA genes accumulating siRNA, a large portion accumulated the same size in both ecotypes (591 for 21/22-nt siRNA and 391 for 24-nt siRNA) or produced siRNA in just one ecotype (367 for 21/22-nt and 316 for 24-nt; Fig. 3B). Among the 1,694 lncRNA genes accumulating siRNAs, only 29 accumulated a different size of siRNAs between the two ecotypes. Therefore, no major change of reciprocal posttranscriptional or transcriptional regulation of lncRNA by sRNAs could be established between ecotypes.

Finally, we investigated the specificity of detection of sRNAs between the two ecotypes. First, we studied lncRNA genes predicted to produce phased 21/22-nt siRNAs. Among the seven predicted lncRNAs, only two were specific to L*er* (Supplemental Fig. S3C). Second, searching for miRNA loci, we found that 23 and 12 of the 191 detected miRNAs were specifically detected in Col and L*er*, respectively (Supplemental Fig. S3D), a proportion related to the variation detected for protein-coding genes. Third, we analyzed the proportion of specific expression for the vast majority of 21/22-nt and 24-nt siRNAs located in coding or noncoding genes. Altogether, the L*er* ecotype produces a larger number of 21/22-nt siRNAs specifically linked to this ecotype (Fig. 3, C and E), whereas Col is more enriched in ecotype-specific 24-nt siRNAs (Fig. 3, D and F), suggesting that in these loci, links with differential posttranscriptional and epigenetic regulations among ecotypes occurred.

Overall, the major difference in the noncoding transcriptome of the two ecotypes was linked to lncRNAs and not associated with small RNAs, even though in certain cases sRNAs may be involved in ecotype-specific regulation.

Differential Accumulation of Transcripts between Ecotypes in Early Response to Pi Deficiency

Root growth arrest in the Col ecotype occurs in the first hours of low-Pi sensing by the root tip (Balzergue et al., 2017), whereas root growth continues in L*er*. To determine the effects of short kinetics in Pi deficiency, we examined gene expression patterns in the two ecotypes in response to this stress. Principal component analysis (PCA) showed a data dispersion that allowed a clear distinction between effects of the ecotype (first axis; Supplemental Fig. S4A) and of the kinetics (second axis; Supplemental Fig. S4A). Thus, we used a multifactor analysis that takes into account the ecotype, the kinetics, and their interaction to investigate differential gene expression independent of coding classification, as coding and noncoding genes had comparable dispersion in our experiments. For each comparison, we confirmed the distribution of *P*-values as a criterion of statistical robustness (Rigaill et al., 2018). After processing the differential analyses, we interpreted the results by separating the genes as "coding" or "noncoding" as defined above.

For coding genes, we observed 3,315 genes differentially expressed between the two ecotypes over the kinetics, with 2,504 differentially expressed between at least two kinetics points over the two ecotypes (Fig. 4A; Supplemental Fig. S4B; Supplemental Table S3). The number of differentially expressed coding genes between ecotypes or along the stress kinetics was similar. However, the response to phosphate starvation was significantly impacted by ecotype in only 55 genes ("interaction" of the two factors; Fig. 4A). Upregulation was observed in 1,566 and 1,749 coding genes along the

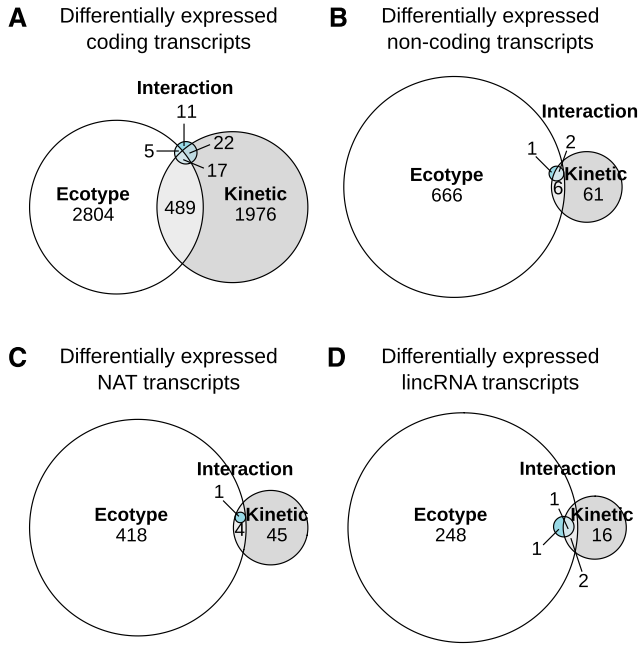


Figure 4. Differentially expressed genes according to ecotype and kinetic effects. Statistical analysis revealed differentially expressed genes between ecotypes and kinetics during phosphate starvation treatments for coding and noncoding genes. The differentially expressed genes can be grouped according to their significant link with genotype effect (different level between the two ecotypes), kinetic effect (differential between any pair of time points in the phosphate starvation kinetics), and the interaction of the two effects (showing differential expression in response to phosphate stress according to genotype). After determining the global distribution, genes were partitioned between coding (A) and noncoding (B) transcripts. Among noncoding genes, we sorted transcripts according to their being antisense to another annotation (C) or intergenic (D).

kinetics in Col and L*er*, respectively (Supplemental Fig. S4B). Interestingly, a clear bias of expression between ecotypes could be observed for noncoding genes (Fig. 4B). Indeed, 675 (666 + 6 + 2 +1) noncoding genes were differentially expressed between the two ecotypes, whereas only 70 (61 + 6 + 2) were differentially expressed along at least one point of the kinetics. Comparable biases were observed for both classes of noncoding genes, lincRNAs and NATs (Fig. 4, C and D). Globally, 146 lincRNAs and 236 NATs were significantly upregulated in Col compared to L*er* and 106 lincRNAs and 187 NATs in L*er* compared to Col (Supplemental Fig. S4C).

We used reverse transcription quantitative PCR (RT-qPCR) analysis of independent replicates of Col, Col^{er105}, and L*er* to confirm the differential expression of 14 lincRNA genes (seven in Col and seven in L*er*) previously identified in the RNA-sequencing (RNA-seq) analysis. We were thereby able to confirm the differential expression of 12 lncRNAs (Fig. 5). Globally, Col and Col^{er105} showed similar expression levels despite minor differences. To investigate any dominant expression effect from one ecotype, we investigated the level of expression of these lncRNAs in the F1 offspring of Col and L*er* crosses. Among the 12 differentially expressed genes, an intermediate level of expression was always detected (of which eight were statistically significant; Fig. 5). This suggests independent regulation of lncRNAs between the parental genomes and discards major dominant "trans" regulatory effects of lncRNA expression between genomes.

One interesting possibility is that specific lncRNAs may be expressed in the Col and L*er* genomes in relation to known regulators of the Pi-starvation response. As a first such case, we were able to identify two specific *Ler* antisense lncRNAs to the Pie transporter *AT5G43370/PHT1.2* gene (Mudge et al., 2002), which is expressed at a higher level in L*er* compared to Col (Supplemental Fig. S5A). The increase of Pi transporter expression in L*er* might impart an increased Pi uptake. As a second case, we found that a Col-expressed NAT RNA is complementary to *SPX4*, a critical regulator of phosphate responses (Duan et al., 2008); in our analysis, it shows reduced expression in Col compared to L*er* (Supplemental Fig. S5B). In other cases, we observed that two consecutive coding transcripts showing differential levels of expression among ecotypes flank a lincRNA with an ecotype-specific expression pattern (Supplemental Fig. S5C), suggesting that various cis effects may be involved in these differential ecotype-linked expression patterns.

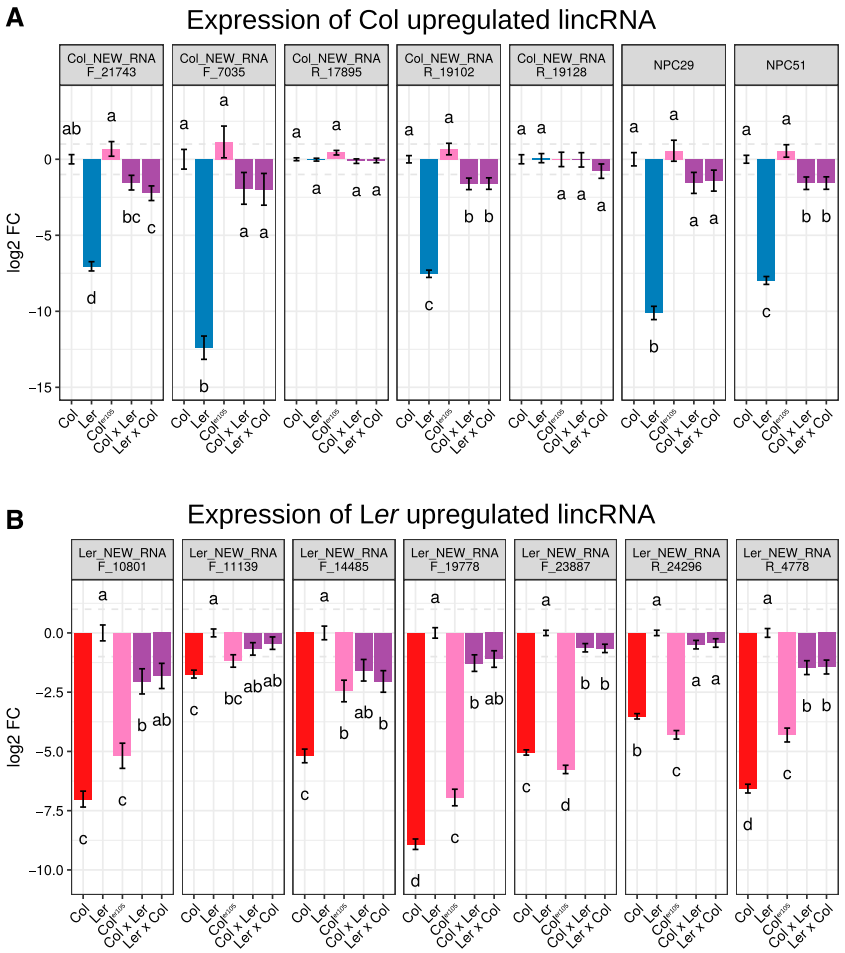


Figure 5. Expression of strongly deregulated lncRNAs between the two ecotypes. The level of expression of strongly deregulated lncRNAs between ecotypes was investigated by RT-qPCR in roots of 11-d-old plants of Col, L*er*, Col^{er105}, and hybrids between Col and L*er* grown under control conditions. A, LincRNAs upregulated in Col relative to L*er*. B, LincRNAs upregulated in L*er* relative to Col. Measurements represent the log 2-fold change (FC) compared to Col (A) or L*er* (B) grown in the same high-phosphate conditions. Error bars represent the SE ($n = 8$; for details, see Supplemental Table S5). Results were analyzed by one-way ANOVA followed by Tukey's honestly significant difference (HSD) mean-separation test. Lowercase letters indicate statistical difference among groups ($P \leq 0.05$).

Differential Accumulation of sRNAs between Ecotypes

The differential accumulation of sRNAs of 21/22and 24 nt was also examined in each ecotype and during the Pi starvation response. PCA of these sequencing data again clearly separated sRNA abundance between ecotypes but not at the level of the kinetics response (Supplemental Fig. S6, A and B; Supplemental Table S3). We identified 416 coding and 211 noncoding genes that accumulated 21/22-nt siRNAs differentially between ecotypes, with generally more siRNAs in L*er* (298 coding genes, 83 lincRNAs, and 40 NATs) than in Col (118 coding genes, 49 lincRNAs, and 39 NATs; Supplemental Fig. S6D; Supplemental Table S3). A greater number of genes accumulated 24-nt siRNAs between ecotypes differentially, showing upregulated siRNAs in L*er* (758 coding genes, 189 lincRNAs, and 67 NATs) compared to Col (391 coding genes, 104 lincRNAs, and 69 NATs; Supplemental Fig. S6E).

Concerning miRNA, as for other small RNAs, the PCA analysis showed only differences between ecotypes (Supplemental Fig. S6C). Indeed, 38 miRNAs were differentially expressed between the two ecotypes (15 and 23 for Col and L*er*, respectively, Supplemental Fig. S6F; Supplemental Table S3). Interestingly, the families of miR399 and miR397 specifically accumulated in the L*er* ecotype. These miRNAs target the *PHOSPHATE2* (*PHO2*) and *NITROGEN LIMITATION ADAPTATION* (*NLA*) transcripts, the encoded proteins of which are known to act together to allow degradation of the Pi transporter PHT1;4 (Park et al., 2014). In L*er*, the higher amount of miR399 and miR397 might be expected to lead to a lower level of PHO2 and NLA and therefore a higher level of PHT1;4 protein; upregulation could increase Pi uptake if there were no counteracting posttranslational regulations affecting Pi transporters (Bayle et al., 2011; Nussaume et al., 2011; Ayadi et al., 2015). However, we observed no difference between ecotypes in accumulation of transcripts of those targets, namely *PHO2* and *NLA*, nor of *PHT1;4* (as also previously reported; Shin et al., 2004; Ayadi et al., 2015). This result suggests that the promoter activity of *PHO2* and *NLA* may compensate for the increased accumulation of these miRNAs in *Ler*.

Misregulation of lncRNA Expression Affects Primary Root Growth in Col

The different patterns of lncRNA expression between ecotypes may induce regulation of root-growth responses. We selected five lncRNA genes, *NPC15*, *NPC34*, *NPC43*, *NPC48*, and *NPC72*, that showed differential expression among ecotypes to study the impact of their expression on Col primary root growth. In RNA-seq data, three of these lncRNA genes were more highly expressed in Col (*NPC15*, *NPC43*, and *NPC72*) and two in L*er* (*NPC34* and *NPC48*; Fig. 6A). The potential dominant expression patterns of these lncRNA genes were evaluated in two F1s of Col × L*er* reciprocal crosses. Expression analysis by RT-qPCR confirmed the RNA-seq results for *NPC15*, *NPC34*, and *NPC72* genes (Fig. 6B), whereas for *NPC48*, the differential expression was only detected in the Col^{er105} mutant, suggesting that the *erecta* mutation affects the expression of this gene. No differential expression could be detected for the *NPC43* gene. This latter result might be due to the concomitant accumulation of antisense transcripts (*NPC504*) in this locus. The differential expression of lncRNA genes between ecotypes could be linked to

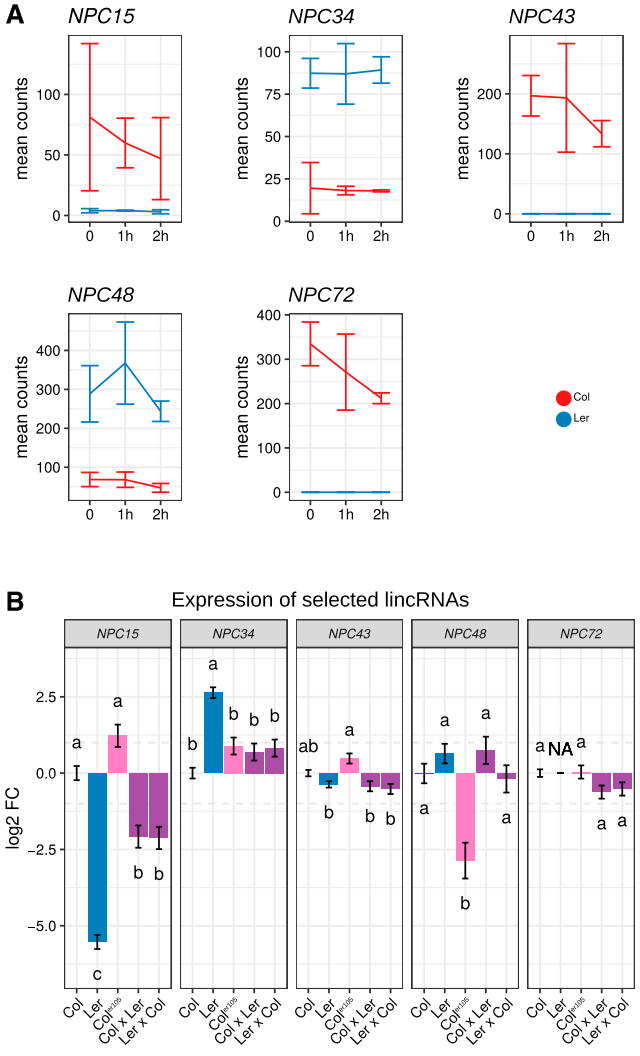


Figure 6. Expression of selected lncRNAs in Col and L*er*. A, Expression profiles of selected lincRNAs in Col and L*er* for early Pi starvation kinetics (RNA-seq data, average expression ± SD, and three replicates). Five selected lincRNAs showed differential expression between Col and L*er* at each time point. B, Level of expression of selected lincRNAs in roots of 11-d-old plants grown under high-Pi conditions in Col, L*er*, Col^{er105}, and hybrids between Col and L*er*. Measurements represent corrected means of log 2-fold changes (FC) compared to Col measured by RT-qPCR. Error bars represent the SE ($n \geq 7$; for details, see Supplemental Table S5). Results were analyzed by one-way ANOVA followed by Tukey's HSD mean-separation test. Lowercase letters indicate statistical difference among groups ($P \leq 0.05$).

genetic changes at their loci. No significant modifications (except a few SNPs) were detected between Col and L*er* for the *NPC34*, *NPC43*, and *NPC48* loci (Supplemental Fig. S7). By contrast, the *NPC15* locus contains an insertion of 2,417 nt in L*er* v8 and *NPC72* is completely missing in L*er* v7 and v8 genomes (Supplemental Fig. S7, A–E). Therefore, genome modifications at the locus level could explain the specific expression pattern of *NPC72* and *NPC15* genes in L*er*.

To support the potential actions of lncRNAs at a phenotypic level, we used overexpressing (35S

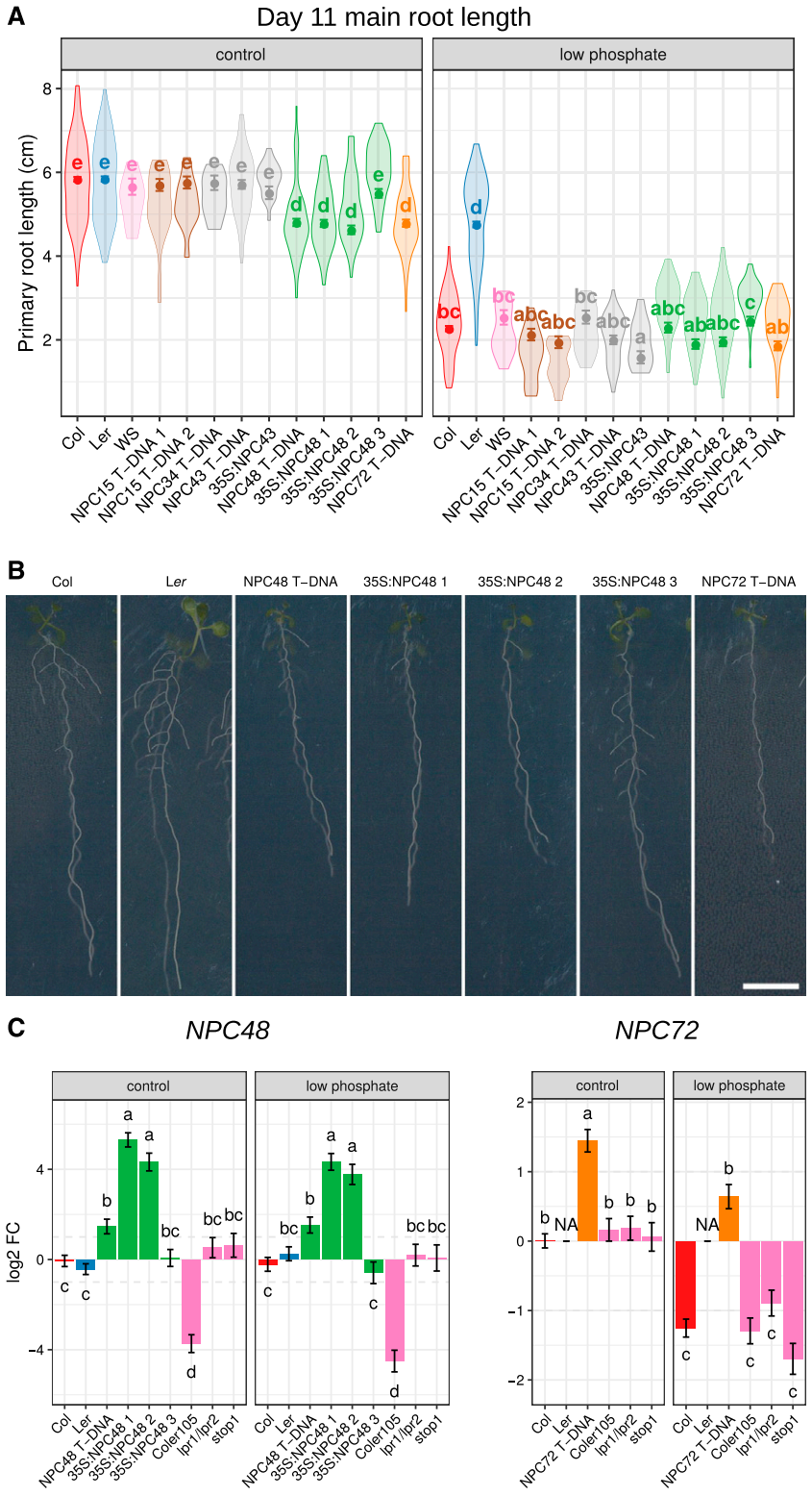


Figure 7. Overexpression of the lincRNAs NPC48 and NPC72 affects primary root growth. A, Mean primary root length according to genotype and Pi condition at the age of 11 d after sowing ($n \geq 23$; for details, see Supplemental Table S6). B, Representative pictures of roots of each genotype 11 d after sowing under high-Pi conditions. Scale bar = 1 cm. C, Expression levels of *NPC48* and *NPC72* in roots of 11-d-old plants grown under high-Pi conditions in lines deregulated in *NPC48* or *NPC72* and mutants affected in Pi-related root arrest. Measurements represent the log 2-fold changes (FC) compared to Col ($n \geq 4$; for details, see Supplemental Table S5). Measurements represent corrected means of primary root growth (A) or of the fold change compared to Col (C). Error bars represent the SE. Results were analyzed by two-way ANOVA (A) or one-way ANOVA (C) followed by Tukey's HSD mean-separation test. Lowercase letters indicate statistical difference among groups ($P \leq 0.05$).

promoter) and T-DNA insertion lines for the same five genes (*NPC15*, *NPC34*, *NPC43*, *NPC48*, and *NPC72*) in Col to monitor the effects on root growth in control and low-Pi conditions. In control conditions, only *NPC48* and *NPC72* overexpression lines led to significant root growth reduction compared to Col (Fig. 7, A and B; Supplemental Fig. S8, A–E). The T-DNA insertions have been mapped in the 5′ region of the *NPC48* and *NPC72* loci. In these lines, the lncRNAs were overexpressed (Fig. 7C). Furthermore the *npc48* T-DNA line strongly supported the phenotype observed with the *35S:NPC48* lines. Repeating the analysis in low-Pi conditions known to inhibit root growth in Col but not in L*er*, we observed minor differences in root length across the different lines. A priori, the ratio of root growth in control and low-Pi conditions should highlight potential differences in Pi sensitivity of transgenic lines. This ratio was significantly increased for *NPC48* and *NPC72* lines compared to Col (Supplemental Fig. S8F).

NPC48 and *NPC72* deregulated lines presented a significant decrease in root length in control conditions, but not in low-Pi conditions. Perhaps even in control conditions the mutants act as if they are partially limited in Pi. Hence, we asked whether these phenotypes could be linked to a root growth arrest due to oversensitive perception of Pi starvation under control conditions or an alteration of Pi systemic sensing (which would affect Pi uptake). This does not seem to be the case, since (1) known Pi-starvation markers were not deregulated in roots under control conditions; and (2) these markers were induced in these lines to the same extent as in Col (Supplemental Fig. S9, A and B). Then, we investigated the local Pi signaling response, exploiting the genes *LOW PHOSPHATE ROOT1* (*LPR1*) and *LPR2* and the transcription factor *STOP1*, which are known to be locally involved in primary root growth arrest under low Pi (Svistoonoff et al., 2007; Ticconi et al., 2009; Müller et al., 2015; Balzergue et al., 2017; Mora-Macías et al., 2017). Expression analysis of *NPC48* and *NPC72* genes in *lpr1/lpr2* and *stop1* mutant lines revealed no significant variation of these lncRNA expression patterns (Fig. 7C). Reciprocally, no significant expression variation was detected for the LPR1/LPR2 pathway in *NPC48* or *NPC72* lines (*LPR1*, *LPR2*, *STOP1*, *ALMT1*, and *MATE* genes; Supplemental Fig. S9, C to G).

To gain further insight into the function of *NPC48*, we performed an RNA-seq analysis of Col ecotype overexpressing *NPC48* (*35S:NPC48-1*) or not under control conditions (where the phenotype was observed) in order to assess the impact of *NPC48* deregulation on the genome-wide transcriptome (Fig. 8). Among the 158 differentially expressed genes, 140 were coding genes, and the majority of these were upregulated in correlation with increased *NPC48* expression (Fig. 8, A and B). In contrast, the great majority of noncoding genes were downregulated in *35S:NPC48-1*, including two lincRNAs and 15 NATs, but none of their antisense coding genes. Since *NPC48* was upregulated in the L*er* ecotype,

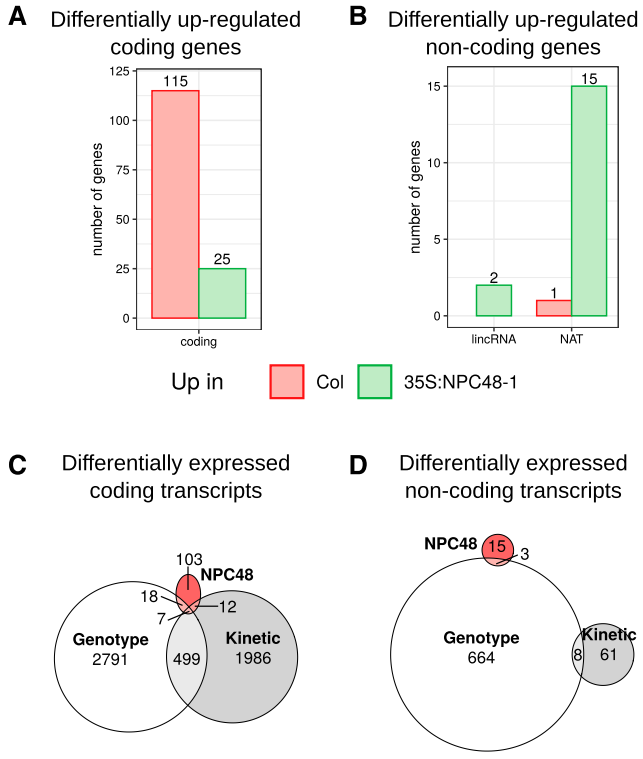


Figure 8. Differentially expressed genes in plants overexpressing NPC48. A and B, Numbers of coding (A) and noncoding (B) genes showing statistically different expression levels in RNA-seq data between Col and *35S:NPC48-1*. C and D, Venn diagrams showing how differentially expressed coding (C) and noncoding (D) genes in *35S:NPC48-1* correlate with those between Col and L*er* or those found during the phosphate starvation kinetics.

we asked whether, among the deregulated genes, several could be linked to ecotype- or phosphate kinetics-related variations (Fig. 8, C and D). However no such direct link could be made, as the large majority of deregulated genes are not linked to any of these categories. To further confirm the deregulation of specific targets, we also tested another *35S:NPC48* line showing an intermediate level of expression compared to *35S:NPC48-1* (Supplemental Fig. S10). Several genes could not be confirmed in this second line, notably the strongly downregulated gene in *35S:NPC48-1* encoding the iron-regulated gene AT3G01260 (Rodríguez-Celma et al., 2013). Among the genes differentially expressed, we could identify several linked to nutrient transport and root hair growth, such as *ABCB3* (Shibata et al., 2018), *CPL1* (Zhang et al., 2016), and *JAL22* (Diet et al., 2006). *ABCB3* and *CPL1* were repressed by overexpression of *NPC48*, in contrast to *JAL22* (Supplemental Fig. S10). There were also several genes encoding known growth regulators of primary root growth that might be linked to the *NPC48* overexpression phenotype (Supplemental Table S8). These include *RGF7*, *BIG*, *RPK2*, and *CASP5*, which were upregulated. Differentially expressed genes are significantly enriched in genes regulated during iron starvation

(25 of 138 genes; $P = 1.62 \times 10^{-3}$; Rodríguez-Celma et al., 2013). Hence, we tested, in the two different *NPC48* overexpressing lines, whether overexpression of *NPC48* might lead to iron-related phenotypes such as (1) modification of the expression of genes related to iron responses (Supplemental Fig. S11, A and B), (2) significant root growth in response to changes in the iron concentration of the growth medium (Supplemental Fig. S11, C and D), or (3) modification of iron accumulation in root tips (Supplemental Fig. S11, E and F). However, no clear link could be made with any alteration of Fe-related responses dependent on *NPC48* misregulation.

Altogether, the identification of ecotype-related lncRNAs allowed us to characterize further regulators of primary root growth. However, overexpression of *NPC48* did not affect the Col transcriptome in a manner that could be linked to L*er* expression patterns. Nevertheless, several genes related to transport, growth regulation, and root hair function are deregulated by misregulation of this ecotype-specific lncRNA.

DISCUSSION

Until recently, transcriptome studies were mainly focused on protein-coding gene transcripts and ignored lncRNAs. Variation in the nt sequences or expression patterns of the noncoding genome can have less pleiotropic effects than changes in the protein sequence of critical regulators. However, it is now commonly accepted that lncRNAs can play central roles in development and response to environmental conditions by their expression in a particular cell-type (Ariel et al., 2015). In the current study, using strand-specific RNA-seq analysis in root tips, we identified thousands of previously uncharacterized lncRNAs (lincRNAs and NATs) expressed at low levels from two Arabidopsis accessions. We focused on root growth, as it is a complex trait that is responsive to the soil environment (Petricka et al., 2012) and impacts a large number of loci spread across the genome. Interestingly, in our study, we observed many lncRNA genes that were differentially or specifically expressed in Col or L*er*, in contrast to the number of protein-coding genes. It had already been shown in other systems that intraspecies variation is strongly linked to the noncoding part of the genome. For example, ~45% of disease-associated human SNPs mapped to noncoding regions of the genome (Ning et al., 2017). In chicken (*Gallus gallus*), domestication traits governing body morphology or behavior are under selection and often associated with lncRNA genes (Wang et al., 2017). Similarly, in plants, the comparison of SNPs associated with fruit phenotypes in two tomato (*Solanum lycopersicum*) cultivars also corresponded to noncoding genomic regions (Scarano et al., 2017). The SNPs could act directly at the level of lncRNA expression or affect the expression of lncRNA-neighboring genes (Kopp and Mendell, 2018). lncRNAs are thus elements to be considered in genetic association studies.

Surprisingly, very few ecotype-specific lncRNAs coincided with the deletion of specific DNA sequences in one particular ecotype (Fig. 3). Hence, we propose that the lncRNA differential expression of a relatively similar DNA molecule results from shifts in transcription rate or stability of lncRNAs that could be connected to SNP or InDel polymorphisms in promoters and/or rearrangements distant from lncRNA loci in the two ecotypes (e.g. transposon insertions). As lncRNAs can repress or activate the transcription of other genes, the expression polymorphisms observed between the two ecotypes could also result in a cascade of cis-local or trans-distal action on target genes (Ariel et al., 2015; Marchese et al., 2017). It is noteworthy that the majority of ecotype-specific lncRNAs identified did not colocalize with siRNAs and thus could not reflect putative gene silencing differences between ecotypes (either transcriptional or posttranscriptional processes; Matzke and Mosher, 2014). This points to the lncRNA itself, or its transcription, being linked to the quantitative regulation of target gene expression (Marchese et al., 2017).

We were able to confirm, by RT-qPCR, the expression levels of 12 lincRNAs among the 14 chosen for validation, supporting the expression variation identified by RNA-seq (Fig. 5). Allele-specific expression is known to affect productivity in plants (Springer and Stupar, 2007). Moreover, in Arabidopsis, heterosis has been reported for different traits, such as flowering time (Seymour et al., 2016) or phosphate acquisition (Narang and Altmann, 2001). As lncRNAs are able to modify chromatin and thus alter gene expression, we added, in our expression analysis, the study of the F1 resulting from reciprocal crosses between Col and L*er*. In the F1 hybrid, the 12 confirmed differentially expressed lncRNA genes chosen for validation globally exhibited an additive expression pattern compared to their parents. This is consistent with results obtained in maize (*Zea mays*) F1 hybrids, where additivity is frequently observed for lncRNAs (Li et al., 2014).

Analysis of the pan-genome (restricted to coding genes) of 19 Arabidopsis ecotypes showed that at least 70 accessory genes could be identified in each ecotype (Contreras-Moreira et al., 2017). In response to stress, accessory genes can explain at least part of the phenotypic difference of behavior observed among ecotypes (Gan et al., 2011). Since we find that lncRNAs have a lower selection pressure (Fig. 2), our results give further evidence that they might play a similar role as accessory coding genes in response to environmental changes.

Root apexes play an important role in sensing external stimuli. We examined the gene expression profile soon after stress application (at 1 and 2 h) in two ecotypes that present differences in response to Pi starvation. In the two ecotypes, the number of differentially expressed coding genes during stress kinetics was similar to that between ecotypes. By contrast, a clear bias of specific expression of lncRNAs and siRNAs was identified. However, among the coding genes that are differentially expressed along the Pi kinetics, one-quarter

are also differentially expressed between the two ecotypes, whereas for the noncoding genes, this fraction is one-eighth (Fig. 4). This suggests that the main part of the ecotype-specific phosphate response comes from the coding portion of the genome. However, despite their lack of response to early stress application, lncRNAs may still influence the plant response to stress by priming the chromatin conformation for a fast response of the coding part of the genome.

The general adaptation of root architecture in response to low Pi consists of an arrest of primary root growth after the perception of Pi limitation. In Arabidopsis, studies concerning plant Pi homeostasis during Pi deficiency characterized the *IPS1/AT4* lncRNA controlling the distribution of Pi from root to shoot. It acts as a target mimic for miR399, which regulates *PHO2* mRNAs (Lin et al., 2008). Moreover, Yuan et al. (2016) identified lncRNAs differentially expressed in roots and shoots of plants grown in the presence or absence of Pi for 10 d. Those authors suggested that a coexpression between lncRNAs and adjacent coding genes may be linked to cis-regulation by lncRNAs of target genes involved in Pi-starvation processes. Interestingly, in fission yeast (*Schizosaccharomyces pombe*), two of the three genes of the Pi regulon are repressed in Pi-rich medium by the transcription of lncRNA genes (Shah et al., 2014) that are present in the 5′ region (cis-regulation). The molecular mechanisms that govern root growth modification by Pi have been mostly elucidated in Col plants. For the local impact of Pi (restricted to root architecture), Pi deficiency is sensed by the root tips and primary root growth inhibition is induced by both the reduction of cell elongation (*STOP1* and *LPR1/LPR2* pathways) and the progressive arrest of meristem division (*LPR1/LPR2* pathway), notably linked to the presence of iron in the medium. Expression analysis in response to Pi deficiency in our mutant lines *NPC48* and *NPC72* did not link these two lncRNAs to Pi starvation root growth arrest mediated by *LPR1/LPR2* and *STOP1* pathways.

The overexpression of *NPC48* leads to reduction of the main root growth. Transcriptome analysis of the strongest deregulated line (*35S:NPC48-1*) shows enrichment in genes that are deregulated during iron starvation (Rodríguez-Celma et al., 2013). However, no visible relation to iron homeostasis could be confirmed in the lines overexpressing *NPC48*. Apart from the link to iron starvation, no enrichment of any pathway could be demonstrated. Few identified deregulated genes could be confirmed in the second line, possibly because of the lower level of expression of *NPC48* in this line or because of the specificity of the T-DNA insertions of both lines. Notably, the iron-regulated gene *AT3G01260* was only deregulated in the *35S:NPC48-1* line. Few potential regulators of root growth or genes related to root hair growth and nutrient transport are modified. Nevertheless, we confirmed the presence of these latter genes in the second overexpressing line, supporting their link with *NPC48* overexpression. Overexpression of *NPC48* may decrease the absorption of essential nutrients, leading to a restriction of root growth by global nutrition deficiency. This lncRNA is a quantitative regulator of primary root growth, but its overexpression did not show any major alteration in the transcriptome. However, the overexpression did modify expression of several genes dealing with root growth or nutrient assimilation, and this is likely linked to its quantitative phenotype. Only a few core regulators of root growth rate have been identified up to now (Satbhai et al., 2015; Motte et al., 2019). Clearly, one can expect more subtle regulators to exist, and *NPC48* might be one of them.

In Arabidopsis, using grafts between ecotypes presenting a high frequency of SNPs, Thieme et al. (2015) showed that about 2,000 mRNAs, among which 9,300 contain SNPs, could move in plants that were subjected to Pi deficiency for 2 weeks. These mRNAs were transported from root to shoot or shoot to root. The authors suggested that these mobile mRNAs might function widely as specific signaling molecules coordinating growth, cell differentiation, and stress adaptation of distant plant organs. As the lncRNAs described here have 3′ polyadenyylated tails and are probably 5′ capped, it is tempting to assume that at least some of them can be transported through the xylem and/or the phloem and may contribute to systemic signaling responses.

Globally, the in-depth exploration of the noncoding transcriptome of two ecotypes presented in this work identified thousands of previously uncharacterized lncRNAs with ecotype-specific expression. Statistical analysis among ecotypes identified several coregulations between coding and noncoding genes (including sRNAs). These coregulations are likely linked to the evolution of different regulatory mechanisms among ecotypes grown in diverse soil environments, and our detailed study of specific cases has provided two ecotype-related lncRNAs that are potentially involved in regulating primary root growth.

MATERIALS AND METHODS

Plant Growth

Seeds were surface-sterilized and sown on a horizontal line in plates vertically disposed in a growing chamber (16-h photoperiod; intensity 90 μE; 21°C). The growth medium was previously described in Balzergue et al. (2017). The −Pi and +Pi agar medium contained 10 and 500 μM Pi, respectively.

For the root apex isolation, seeds were sown on 1-cm bands of nylon membrane (Nitex 100 μm). After 1 week on +Pi agar medium, the membranes were transferred to −Pi agar medium. Plants were sampled at time points 0, 1, and 2 h after transfer. Each biological replicate is a pool of >100 root apexes cut at 0.5 cm from the root tip.

Arabidopsis Lines

The *stop1* (SALK_114108, Nottingham Arabidopsis Stock Center [NASC] reference N666684), *lpr1;lpr2* (Svistoonoff et al., 2007), *npc15 T-DNA1* (SALK_027817; NASC reference N527817), *npc15 T-DNA2* (SALK_090867; NASC reference N590867), *npc43 T-DNA* (SALK_007967; NASC reference N507967), *npc48 T-DNA* (SAIL_1165_H01; NASC reference N843057), and *npc72 T-DNA* (SAIL_571_C12; NASC reference N824316) lines are in the

ecotype Columbia (Col) of Arabidopsis (*Arabidopsis thaliana*) background. Overexpressing lines *35S:NPC43*, *35S:NPC48-1*, *35S:NPC48-2*, and *35S:NPC48-3* were retrieved from Ben Amor et al. (2009) and are in the Col-0 background. *npc34 T-DNA* (FLAG_223D08 or FLAG_228A07) is in the Wassilewskija background. Col^{er105} is in the Columbia background (Col-0) with the null allele erecta-105.

Library Construction and Sequencing

For each time point (0, 1, and 2 h), total RNA of three biological replicates of the Col^{er105} and L*er* pool of root apexes were extracted following the RNeasy micro kit (Qiagen) protocol. One microgram of total RNA of each sample was used for mRNA library preparation using the Illumina TruSeq Stranded mRNA library preparation kit according to the manufacturer's instructions. Libraries were sequenced on a HiSeq 2000 Sequencing System (Illumina) using 100-nt paired-end reads.

For *35S:NPC48-1* RNA-seq analysis, total RNA of three biological replicates of whole roots from Col and *35S:NPC48-1* were extracted using a *Quick*-RNA Miniprep kit (Zymo Research). One microgram of total RNA of each sample was used for mRNA library preparation using an Illumina Truseq Stranded mRNA library preparation kit according to the manufacturer's instructions. Libraries were sequenced on a NextSeq 500 Sequencing System (Illumina) using 75-nt single-end reads.

sRNAs of root apexes were extracted using the *mir*Vana miRNA Isolation Kit (Ambion). sRNA libraries were constructed using the Ion Total RNA-Seq Kit v2 (Ion Torrent, Life Technologies) according to the manufacturer's instructions. Libraries were then sequenced using IonProton and the adapters.

Previously Uncharacterized Transcript Identification

According to their ecotype of origin, transcript reads were aligned to the TAIR10 (Lamesch et al., 2012) or L*er* v7 (Gan et al., 2011) genome. The previously uncharacterized transcripts were predicted in each ecotype independently and then cross-positioned on the other genome to analyze homology. Using the information available on the Col genome from databases, the different transcripts were classified as coding and noncoding (lncRNAs), lncRNAs included lincRNAs and NATs. See the detailed protocol in Supplemental Materials and Methods.

sRNA Analysis

The cleaned sRNA reads were aligned on the TAIR10 or L*er* v7 genome. For Araport11 annotations and previously uncharacterized genes predicted here, the accumulation of sRNA was analyzed to classify them as miRNA, phased siRNA cluster, or siRNA cluster and determine the main siRNA size. See detailed protocol in Supplemental Materials and Methods.

Expression Analysis

For each annotation, mRNA reads were counted to estimate the level of expression of each gene. These counts were used for differential gene expression analysis. Using siRNA accumulation on each annotation, differential accumulation of 21/22-nt and 24-nt sRNAs on coding and noncoding genes was computed independently. Bonferroni correction of the *P*-value was used for each analysis. Differentially expressed genes or siRNA accumulations were defined as having an adjusted *P*-value inferior to 0.01. See the detailed protocol in Supplemental Materials and Methods.

Measurement of the Primary Root Length

Images were taken with a flat scanner and root lengths were measured using RootNav software (Pound et al., 2013).

RT-qPCR

Total RNA was extracted from whole roots using the Quick-RNA MiniPrep kit (Zymo Research, USA) according to the manufacturer's instructions. Reverse transcription was performed on 500 ng total RNA using the Maxima Reverse Transcriptase (Thermo Scientific). RT-qPCR was performed on a 480 LightCycler thermocycler (Roche) according to the manufacturer's instructions with Light cycler 480 SYBR Green I Master (Roche) and the primers listed in Supplemental Table S4. We used PP2A subunit PD (AT1G13320) as a reference gene for normalization.

Statistics and Reproducibility of Experiments

For each measure (root length, qPCR expression level), the least-squares means were computed. This approach makes it possible to correct for inter-repetition variation. Data are presented as least-squares means ± SE. The statistical significance tests are included in the legend of each figure. See detailed protocol in Supplemental Materials and Methods.

Accession Numbers

Sequence files generated during this study have been deposited in the NCBI GEO database under the accessions GSE128250, GSE128256 and GSE151005. Names and accession numbers of genes mentioned are listed in Supplemental Table S9.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Characteristics of identified transcripts.

Supplemental Figure S2. Expression level and detection of coding and noncoding genes.

Supplemental Figure S3. Ecotype-specific classification of lncRNAs as siRNA precursors.

Supplemental Figure S4. Ecotype effect on gene expression.

Supplemental Figure S5. Genome organization and correlation of expression at selected loci.

Supplemental Figure S6. Ecotype effect on siRNA accumulation.

Supplemental Figure S7. Genome homology at selected ncRNA loci

Supplemental Figure S8. Deregulation of selected noncoding RNAs.

Supplemental Figure S9. Deregulation of NPC48 and NPC72 does not change the expression of phosphate starvation-related genes.

Supplemental Figure S10. Expression analysis of genes regulated in the *35S:NPC48-1* line in a different independent transgenic line.

Supplemental Figure S11. Overexpression of *NPC48* does not affect iron homeostasis and/or phenotypic responses.

Supplemental Table S1. Mapping efficiency for each sequence sample.

Supplemental Table S2. Genomic information of previously uncharacterized transcripts compared to TAIR10.

Supplemental Table S3. Differential gene expression analysis.

Supplemental Table S4. Sequence of primers used in this study.

Supplemental Table S5. Number of samples used for each genotype and condition in the RT-qPCR experiments.

Supplemental Table S6. Number of samples used for each genotype and condition in root-length measurements.

Supplemental Table S7. Software used for the bioinformatics analyses.

Supplemental Table S8. Differentially expressed genes in *35S:NPC48-1*.

Supplemental Table S9. Gene name abbreviations.

Supplemental File S1. Identified Col transcripts.

Supplemental File S2. Identified L*er* transcripts.

Supplemental Materials and Methods. Detailed bioinformatic and extra materials and methods.

ACKNOWLEDGMENTS

We thank Ambre Miassod and Janina Lüders (Institute of Plant Sciences Paris-Saclay, Centre Nationale de la Recherche Scientifique, Institut National

de la Recherche Agronomique, Université Evry, Université Paris-Saclay) for technical assistance with gene expression experiments. We thank Olivier Martin (Institute of Plant Sciences Paris-Saclay, Centre Nationale de la Recherche Scientifique, Institut National de la Recherche Agronomique, Université Evry, Université Paris-Saclay) for critical reading of the manuscript.

Received April 15, 2020; accepted April 30, 2020; published May 13, 2020.

LITERATURE CITED

Abel S (2017) Phosphate scouting by root tips. Curr Opin Plant Biol **39:** 168–177

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

Ayadi A, David P, Arrighi J-F, Chiarenza S, Thibaud M-C, Nussaume L, Marin E (2015) Reducing the genetic redundancy of Arabidopsis PHOSPHATE TRANSPORTER1 transporters to study phosphate uptake and signaling. Plant Physiol **167:** 1511–1526

Balzergue C, Dartevelle T, Godon C, Laugier E, Meisrimler C, Teulon J-M, Creff A, Bissler M, Brouchoud C, Hagège A, et al (2017) Low phosphate activates STOP1-ALMT1 to rapidly inhibit root cell elongation. Nat Commun **8:** 15300

Baxter IR, Vitek O, Lahner B, Muthukumar B, Borghi M, Morrissey J, Guerinot ML, Salt DE (2008) The leaf ionome as a multivariable system to detect a plant's physiological status. Proc Natl Acad Sci USA **105:** 12081–12086

Bayle V, Arrighi J-FJ-F, Creff A, Nespoulous C, Vialaret J, Rossignol M, Gonzalez E, Paz-Ares J, Nussaume L (2011) *Arabidopsis thaliana* high-affinity phosphate transporters exhibit multiple levels of posttranslational regulation. Plant Cell **23:** 1523–1535

Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, et al (2009) Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. Genome Res **19:** 57–69

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol **16:** 727–741

Chekanova JA (2015) Long non-coding RNAs and their functions in plants. Curr Opin Plant Biol **27:** 207–216

Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. Plant J **89:** 789–804

Contreras-Moreira B, Cantalapiedra CP, García-Pereira MJ, Gordon SP, Vogel JP, Igartua E, Casas AM, Vinuesa P (2017) Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species. Front Plant Sci **8:** 184

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res **22:** 1775–1789

Diet A, Link B, Seifert GJ, Schellenberg B, Wagner U, Pauly M, Reiter WD, Ringli C (2006) The *Arabidopsis* root hair cell wall formation mutant *lrx1* is suppressed by mutations in the *RHM1* gene encoding a UDP-L-rhamnose synthase. Plant Cell **18:** 1630–1641

Duan K, Yi K, Dang L, Huang H, Wu W, Wu P (2008) Characterization of a sub-family of Arabidopsis genes with the SPX domain reveals their diverse functions in plant tolerance to phosphorus starvation. Plant J **54:** 965–975

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, et al (2011) Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature **477:** 419–423

Gutiérrez-Alanís D, Ojeda-Rivera JO, Yong-Villalobos L, Cárdenas-Torres L, Herrera-Estrella L (2018) Adaptation to phosphate scarcity: Tips from *Arabidopsis* roots. Trends Plant Sci **23:** 721–730

Hirsch J, Marin E, Floriani M, Chiarenza S, Richaud P, Nussaume L, Thibaud MC (2006) Phosphate deficiency promotes modification of iron distribution in Arabidopsis plants. Biochimie **88:** 1767–1771

Hu L, Xu Z, Hu B, Lu ZJ (2017) COME: A robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. Nucleic Acids Res **45:** e2

Hupalo D, Kern AD (2013) Conservation and functional element discovery in 20 angiosperm plant genomes. Mol Biol Evol **30:** 1729–1744

Kapusta A, Feschotte C (2014) Volatile evolution of long noncoding RNA repertoires: Mechanisms and biological implications. Trends Genet **30:** 439–452

Kopp F, Mendell JT (2018) Functional classification and experimental dissection of long noncoding RNAs. Cell **172:** 393–407

Kozomara A, Griffiths-Jones S (2014) miRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res **42:** D68–D73

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al (2012) The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. Nucleic Acids Res **40:** D1202–D1210

Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE, et al (2014) Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biol **15:** R40

Li S, Yamada M, Han X, Ohler U, Benfey PN (2016) High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. Dev Cell **39:** 508–522

Lin SI, Chiang SF, Lin WY, Chen JW, Tseng CY, Wu PC, Chiou TJ (2008) Regulatory network of microRNA399 and *PHO2* by systemic signaling. Plant Physiol **147:** 732–746

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H (2012) Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. Plant Cell **24:** 4333–4345

Marchese FP, Raimondi I, Huarte M (2017) The multidimensional mechanisms of long noncoding RNA function. Genome Biol **18:** 206

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. Nat Rev Genet **15:** 394–408

Misson J, Raghothama KG, Jain A, Jouhet J, Block MA, Bligny R, Ortet P, Creff A, Somerville S, Rolland N, et al (2005) A genome-wide transcriptional analysis using *Arabidopsis thaliana* Affymetrix gene chips determined plant responses to phosphate deprivation. Proc Natl Acad Sci USA **102:** 11934–11939

Mohammadin S, Edger PP, Pires JC, Schranz ME (2015) Positionally-conserved but sequence-diverged: Identification of long non-coding RNAs in the Brassicaceae and Cleomaceae. BMC Plant Biol **15:** 217

Mora-Macías J, Ojeda-Rivera JO, Gutiérrez-Alanís D, Yong-Villalobos L, Oropeza-Aburto A, Raya-González J, Jiménez-Domínguez G, Chávez-Calvillo G, Rellán-Álvarez R, Herrera-Estrella L (2017) Malate-dependent Fe accumulation is a critical checkpoint in the root developmental response to low phosphate. Proc Natl Acad Sci USA **114:** E3563–E3572

Motte H, Vanneste S, Beeckman T (2019) Molecular and environmental regulation of root development. Annu Rev Plant Biol **70:** 465–488

Mudge SR, Rae AL, Diatloff E, Smith FW (2002) Expression analysis suggests novel roles for members of the Pht1 family of phosphate transporters in *Arabidopsis*. Plant J **31:** 341–353

Müller J, Toev T, Heisters M, Teller J, Moore KL, Hause G, Dinesh DC, Bürstenbinder K, Abel S (2015) Iron-dependent callose deposition adjusts root meristem maintenance to phosphate availability. Dev Cell **33:** 216–230

Narang RA, Altmann T (2001) Phosphate acquisition heterosis in *Arabidopsis thaliana*: A morphological and physiological analysis. Plant Soil **234:** 91–97

Nelson ADL, Devisetty UK, Palos K, Haug-Baltzell AK, Lyons E, Beilstein MA (2017) Evolinc: A tool for the identification and evolutionary comparison of long intergenic non-coding RNAs. Front Genet **8:** 52

Ning S, Yue M, Wang P, Liu Y, Zhi H, Zhang Y, Zhang J, Gao Y, Guo M, Zhou D, et al (2017) LincSNP 2.0: An updated database for linking disease-associated SNPs to human long non-coding RNAs and their TFBSs. Nucleic Acids Res **45**(D1): D74–D78

Nussaume L, Kanno S, Javot H, Marin E, Pochon N, Ayadi A, Nakanishi TM, Thibaud MC (2011) Phosphate import in plants: Focus on the PHT1 transporters. Front Plant Sci **2:** 83

Park BS, Seo JS, Chua N-H (2014) NITROGEN LIMITATION ADAPTATION recruits PHOSPHATE2 to target the phosphate transporter PT2 for degradation during the regulation of *Arabidopsis* phosphate homeostasis. Plant Cell **26:** 454–464

Petricka JJ, Winter CM, Benfey PN (2012) Control of *Arabidopsis* root development. Annu Rev Plant Biol **63:** 563–590

Pound MP, French AP, Atkinson JA, Wells DM, Bennett MJ, Pridmore T (2013) RootNav: Navigating images of complex root architectures. Plant Physiol **162:** 1802–1814

Reymond M, Svistoonoff S, Loudet O, Nussaume L, Desnos T (2006) Identification of QTL controlling root growth response to phosphate starvation in *Arabidopsis thaliana*. Plant Cell Environ **29:** 115–125

Rigaill G, Balzergue S, Brunaud V, Blondet E, Rau A, Rogier O, Caius J, Maugis-Rabusseau C, Soubigou-Taconnat L, Aubourg S, et al (2018) Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. Brief Bioinform **19:** 65–76

Rodríguez-Celma J, Lin WD, Fu GM, Abadía J, López-Millán AF, Schmidt W (2013) Mutually exclusive alterations in secondary metabolism are critical for the uptake of insoluble iron compounds by Arabidopsis and *Medicago truncatula*. Plant Physiol **162:** 1473–1485

Satbhai SB, Ristova D, Busch W (2015) Underground tuning: Quantitative regulation of root growth. J Exp Bot **66:** 1099–1112

Scarano D, Rao R, Corrado G (2017) In silico identification and annotation of non-coding RNAs by RNA-seq and de novo assembly of the transcriptome of tomato fruits. PLoS One **12:** e0171504

Seymour DK, Chae E, Grimm DG, Martín Pizarro C, Habring-Müller A, Vasseur F, Rakitsch B, Borgwardt KM, Koenig D, Weigel D (2016) Genetic architecture of nonadditive inheritance in *Arabidopsis thaliana* hybrids. Proc Natl Acad Sci USA **113:** E7317–E7326

Shah S, Wittmann S, Kilchert C, Vasiljeva L (2014) lncRNA recruits RNAi and the exosome to dynamically regulate pho1 expression in response to phosphate levels in fission yeast. Genes Dev **28:** 231–244

Shibata M, Breuer C, Kawamura A, Clark NM, Rymen B, Braidwood L, Morohashi K, Busch W, Benfey PN, Sozzani R, et al (2018) GTL1 and DF1 regulate root hair growth through transcriptional repression of *ROOT HAIR DEFECTIVE 6-LIKE 4* in *Arabidopsis*. Development **145:** dev159707

Shin H, Shin HS, Dewbre GR, Harrison MJ (2004) Phosphate transport in Arabidopsis: Pht1;1 and Pht1;4 play a major role in phosphate acquisition from both low- and high-phosphate environments. Plant J **39:** 629–642

Springer NM, Stupar RM (2007) Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. Plant Cell **19:** 2391–2402

Svistoonoff S, Creff A, Reymond M, Sigoillot-Claude C, Ricaud L, Blanchet A, Nussaume L, Desnos T (2007) Root tip contact with low-phosphate media reprograms plant root architecture. Nat Genet **39:** 792–796

Szcześniak MW, Rosikiewicz W, Makałowska I (2016) CANTATAdb: A collection of plant long non-coding RNAs. Plant Cell Physiol **57:** e8

The 1001 Genomes Consortium (2016) 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell **166:** 481–491

Thibaud MC, Arrighi JF, Bayle V, Chiarenza S, Creff A, Bustos R, Paz-Ares J, Poirier Y, Nussaume L (2010) Dissection of local and systemic transcriptional responses to phosphate starvation in Arabidopsis. Plant J **64:** 775–789

Thieme CJ, Rojas-Triana M, Stecyk E, Schudoma C, Zhang W, Yang L, Miñambres M, Walther D, Schulze WX, Paz-Ares J, et al (2015) Endogenous *Arabidopsis* messenger RNAs transported to distant tissues. Nat Plants **1:** 15025

Ticconi CA, Lucero RD, Sakhonwasee S, Adamson AW, Creff A, Nussaume L, Desnos T, Abel S (2009) ER-resident proteins PDR2 and LPR1 mediate the developmental response of root meristems to phosphate availability. Proc Natl Acad Sci USA **106:** 14174–14179

Vernikos G, Medini D, Riley DR, Tettelin H (2015) Ten years of pan-genome analyses. Curr Opin Microbiol **23:** 148–154

Wang YM, Xu HB, Wang MS, Otecko NO, Ye LQ, Wu DD, Zhang YP (2017) Annotating long intergenic non-coding RNAs under artificial selection during chicken domestication. BMC Evol Biol **17:** 192

Ward JT, Lahner B, Yakubova E, Salt DE, Raghothama KG (2008) The effect of iron on the primary root elongation of Arabidopsis during phosphate deficiency. Plant Physiol **147:** 1181–1191

Yuan J, Zhang Y, Dong J, Sun Y, Lim BL, Liu D, Lu ZJ (2016) Systematic characterization of novel lncRNAs responding to phosphate starvation in *Arabidopsis thaliana*. BMC Genomics **17:** 655

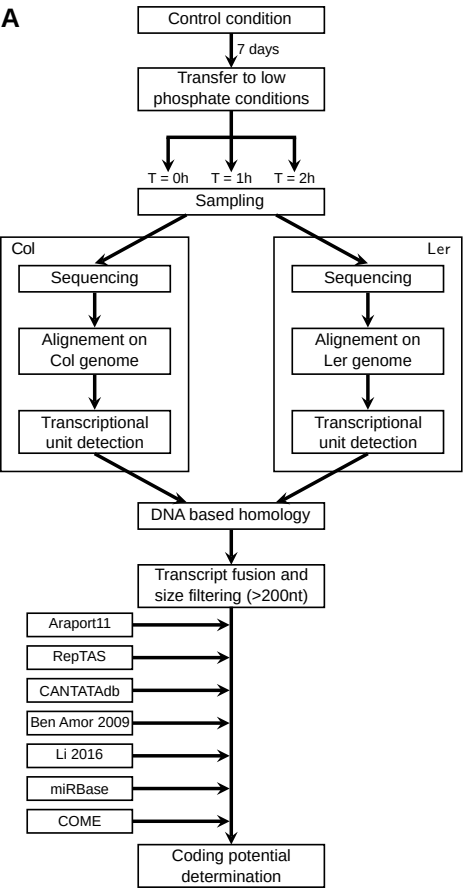
Zapata L, Ding J, Willing E-M, Hartwig B, Bezdan D, Jiao W-B, Patel V, Velikkakam James G, Koornneef M, Ossowski S, et al (2016) Chromosome-level assembly of *Arabidopsis thaliana* L*er* reveals the extent of translocation and inversion polymorphisms. Proc Natl Acad Sci USA **113:** E4052–E4060

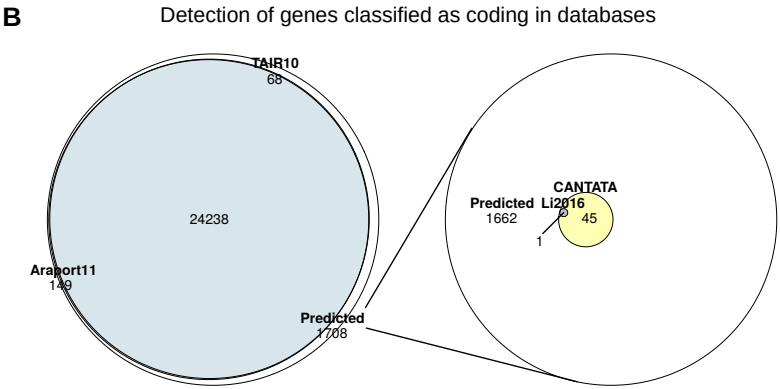
Zhang B, Yang G, Chen Y, Zhao Y, Gao P, Liu B, Wang H, Zheng ZL (2016) C-terminal domain (CTD) phosphatase links Rho GTPase signaling to Pol II CTD phosphorylation in *Arabidopsis* and yeast. Proc Natl Acad Sci USA **113:** E8197–E8206

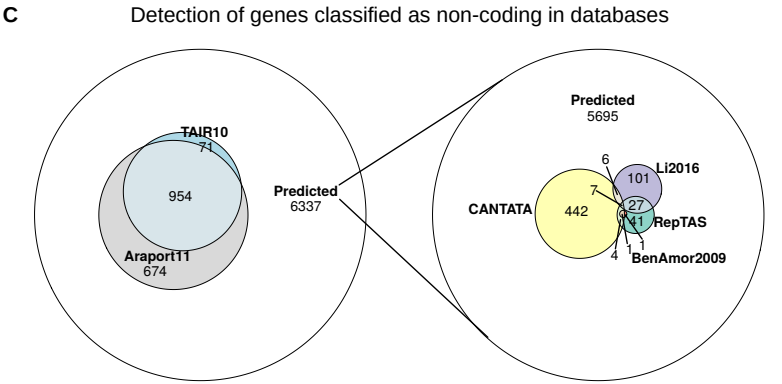
Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T, et al (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat Genet **50:** 278–284

Supplemental Figure S1 - Characteristic of identified transcripts

A, Flowchart of identification of lncRNAs responsive to Pi starvation in two Arabidopsis ecotypes. Plants of the two ecotypes were grown in control condition for seven days before transfer to low phosphate condition. Root tips were then sampled at time point 0 h, 1 h and 2 h after transfer. After RNA extraction, PolyA transcripts were sequenced. The retrieved reads were then mapped independently for the two ecotypes onto their respective genomes (TAIR10 for Col and Ler v7 for Ler). Based on this mapping we predicted new transcriptional units on each genome compared to TAIR10 annotation available on both genomes. We then aligned the resulting transcriptional data onto the opposite genome to compute a homology and fused the overlapping transcripts. Only transcripts with a length superior to 200 nt were kept for further analysis. Each transcript was then categorized in one of the 4 following classes: coding, structural RNA (rRNA, tRNA, snRNA, snoRNA, ...), transposable element and non-coding RNA. The classification was estimated based first the overlap with already annotated transcripts (by order of importance in Araport11, RepTAS database, CANTATA database, BenAmor et al. 2009, Li et al. 2016 and miRBase v21). For the transcripts that were not found in any of these databases, their coding potential was predicted using COME. B and C, occurrence of the different detected genes in the different Arabidopsis databases for genes predicted as coding (B) or as non-coding (C).

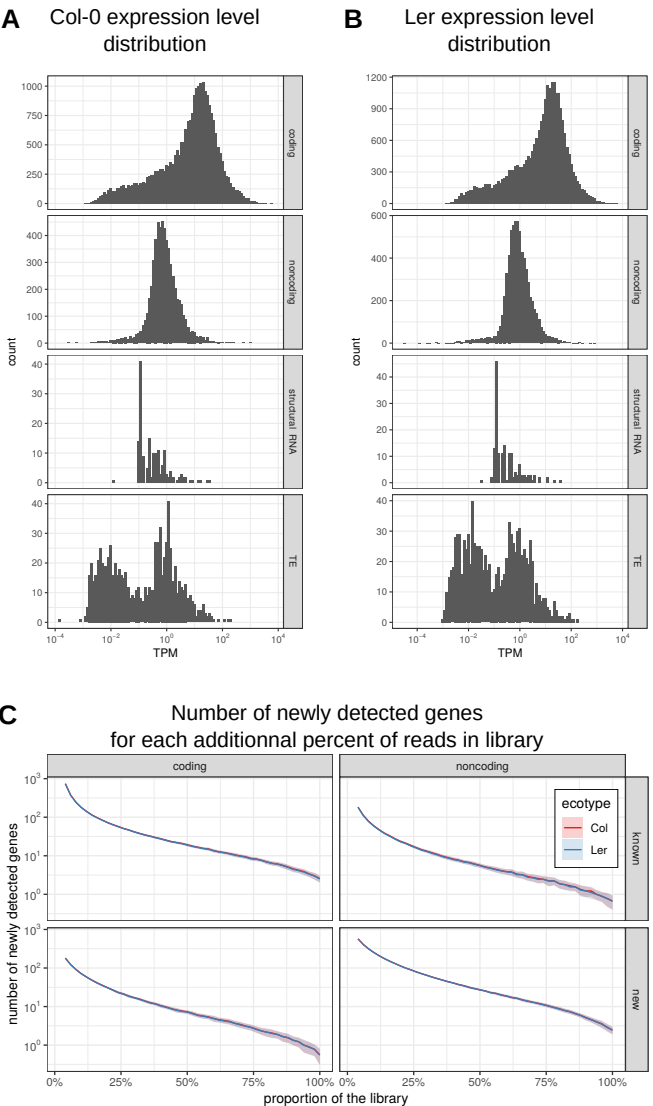






Supplemental Figure S2 - Expression level and detection of coding and non-coding genes

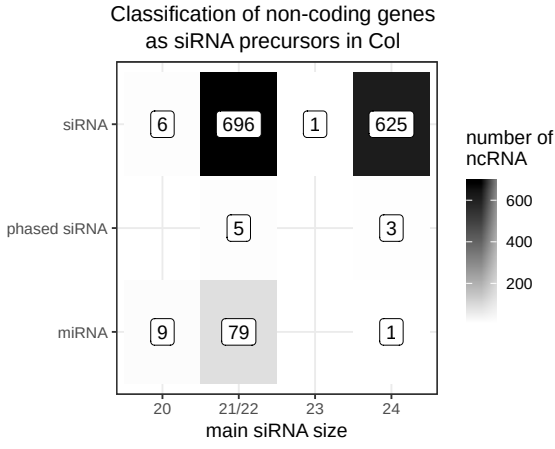
A and B, distribution of coding genes, non-coding genes, structural RNA and transposable elements according to their level of expression in Col (A) or Ler (B). C, Number of newly detected genes per percent of additional sequencing reads of the library. Line corresponds to the median of 100 bootstraps, grey shadow corresponds to the standard deviation.



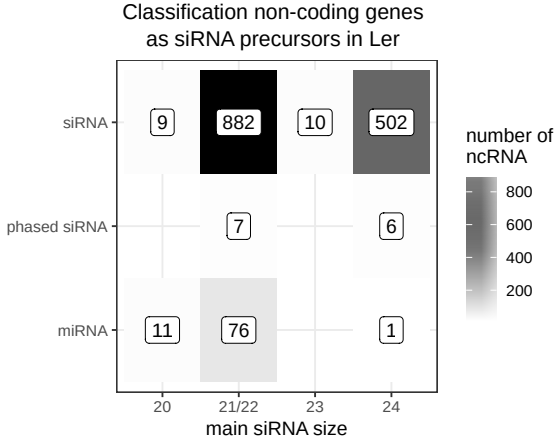
Supplemental Figure S3 - Ecotype specific classification of lncRNAs as siRNA precursors

A and B, ShortStack classification of all non-coding genes as siRNA precursors expressed at more than 1 RPM of major siRNA in Col (A) or Ler (B). C, Detection specificity of non-coding RNA as phased 21 nt and 22 nt siRNAs precursors at the level of 1 RPM in each ecotype. D, Detection specificity of miRBase miRNAs at the level of 1 RPM in each ecotype. The majority of the miRNAs are detected in both ecotypes.

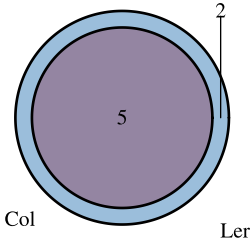
A



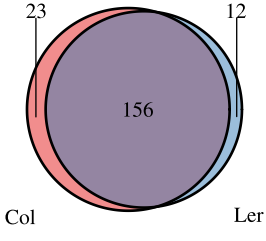
B



C Detection of non-coding genes as 21/22nt phased siRNA precursors

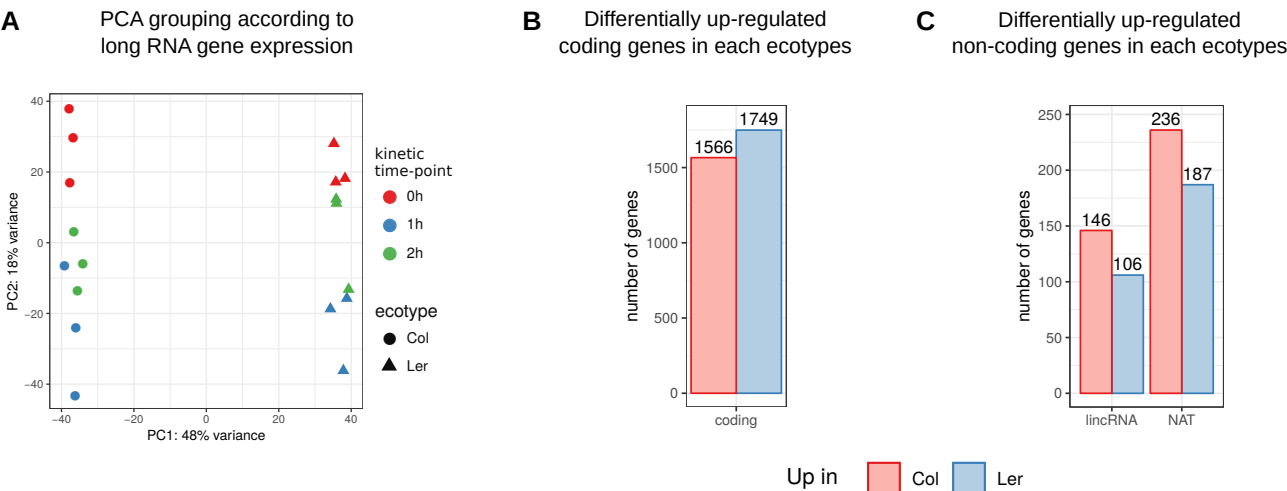


D Detection of miRBase miRNAs



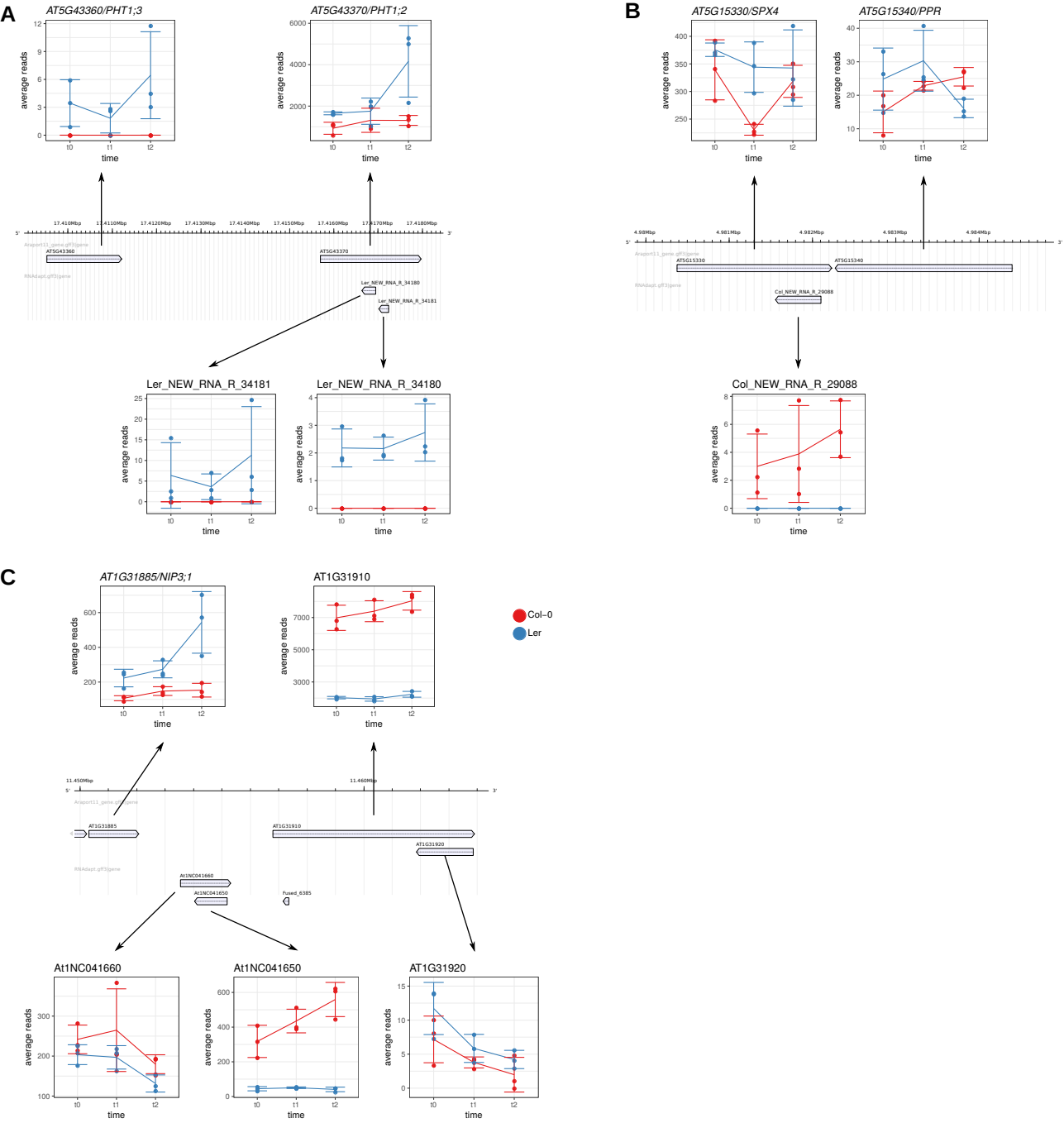
Supplemental Figure S4 - The ecotype effect on gene expression

A, PCA analysis showing the effect of ecotype and phosphate kinetics on the variance of gene expression. For the 3315 coding genes B, and 675 non-coding genes C, that are ecotype differentially expressed, we give the number of genes that accumulated more in Col and in Ler.



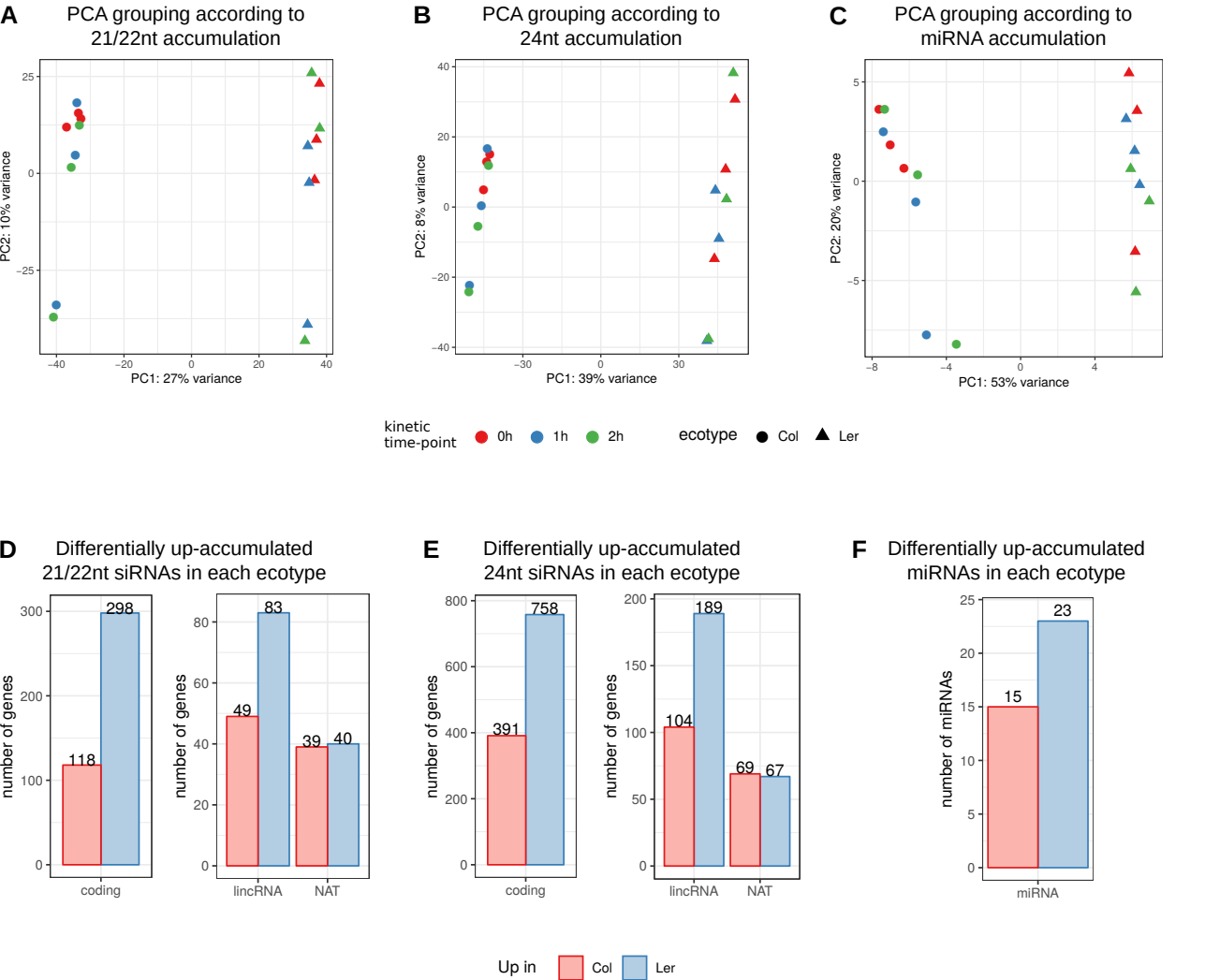
Supplemental Figure S5 - Genome organization and correlation of expression at selected loci

A, Increased expression of *PHT1;2* in Ler correlated with the specific expression of the two NATs *Ler_NEW_R_34181* and *Ler_NEW_R_34180*. B, Decreased expression of *SPX4* gene in Col correlated to the expression of the NAT *Col_NEW_RNA_R_29088*. C, Differential expression of *NIP3;1* and *AT1G1910* between Col and Ler correlated with the expression of lincRNA *At1NC041650*. Error bars represent standard deviation, (n = 3).



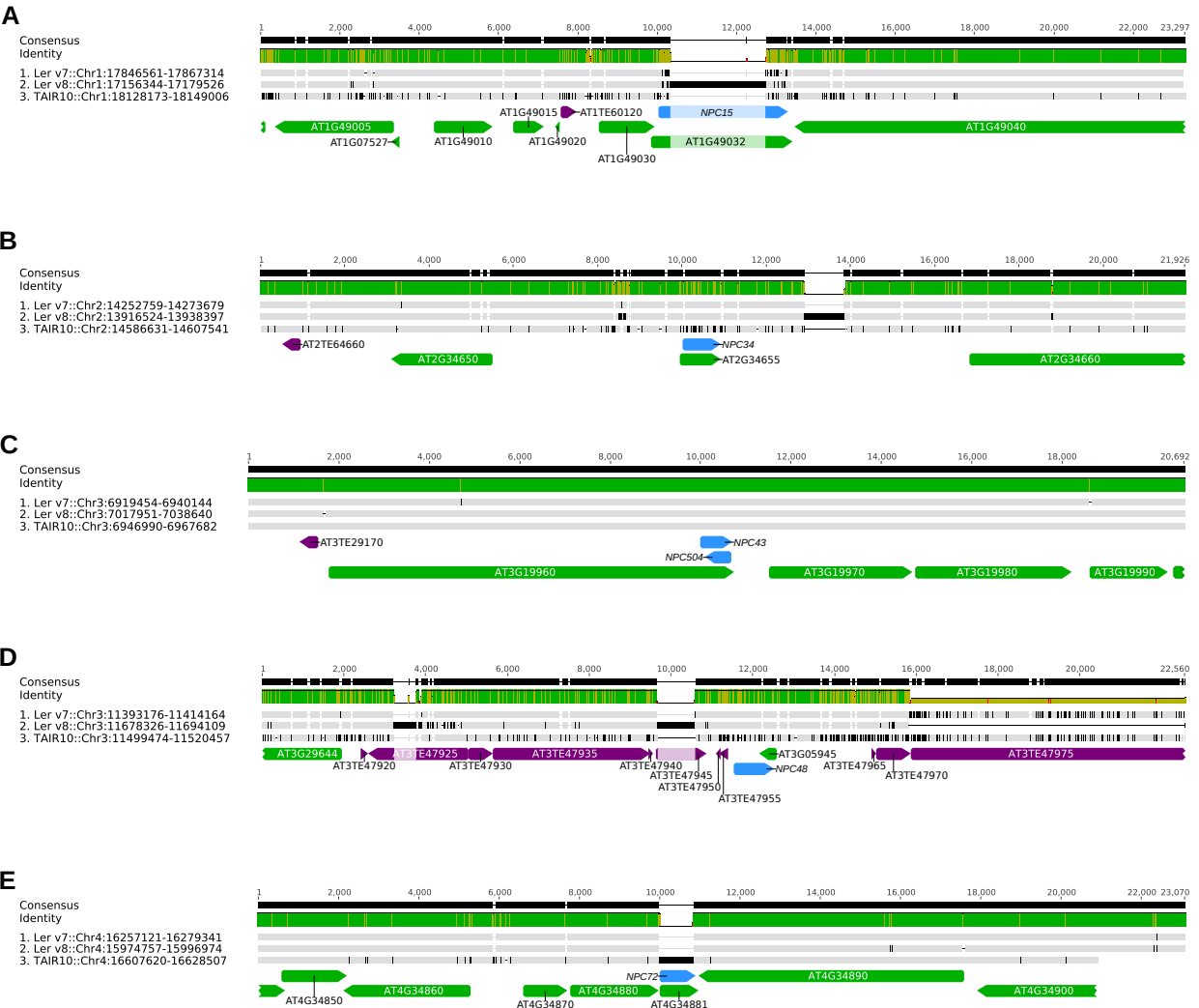
Supplemental Figure S6 - The ecotype effect on siRNA accumulation

A to C, PCA analysis showing the effect of genotype and phosphate kinetics on the variance between samples for 21/22nt (A), 24nt (B) and miRNAs (C). The samples can be well separated according to genotype (Col or Ler) but not according to the phosphate kinetics time point. D to F, For ecotype differentially accumulated siRNAs, the number of differentially accumulated siRNAs in either ecotype for 21/22nt siRNA precursors (D), 24nt siRNA precursors (E) and miRNAs (F).



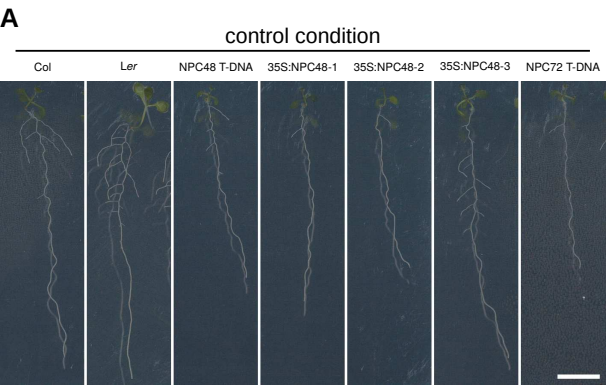
Supplemental Figure S7 - Genome homology at selected ncRNA loci

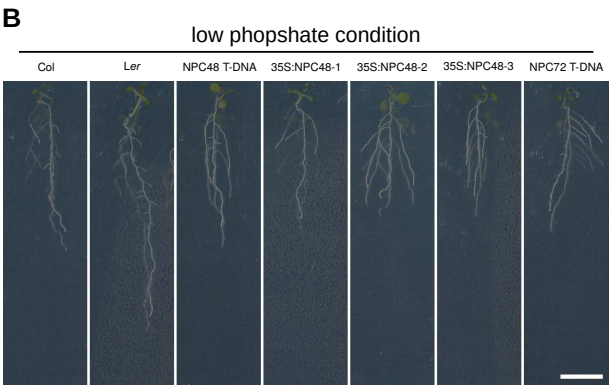
A to E, genome alignment around NPC15 (A), NPC34 (B), NPC43 (C), NPC48 (D) and NPC72 (E) between the reference sequence of Arabidopsis Col and two sequences from Ler (Ler v7 and Ler v8). The identity line is color coded according to conservation: green is for full conservation, yellow is for mismatches. For each sequence, dashes represent missing sequences, plain grey represents identical sequences and plain black represents specific insertion or mismatches.

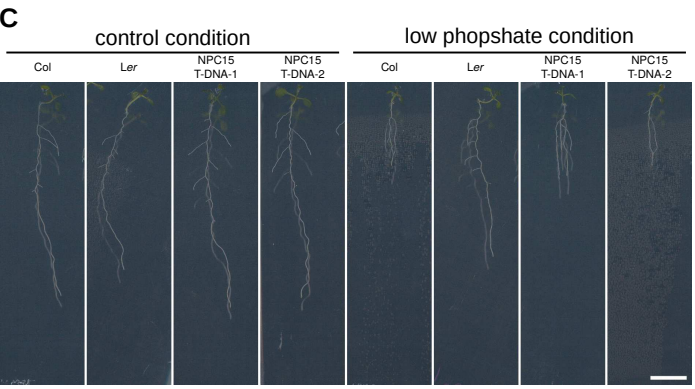


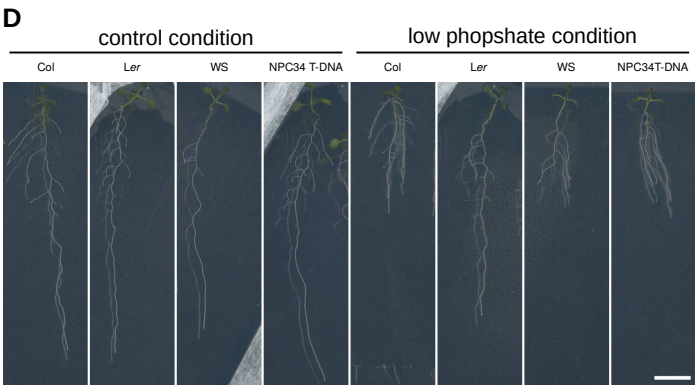
Supplemental Figure S8 - Deregulation of selected non-coding RNAs

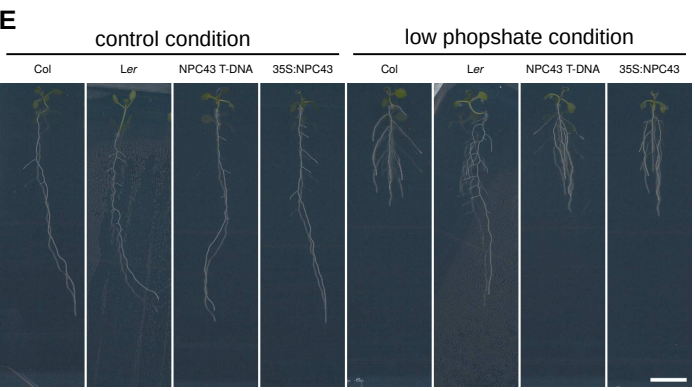
A to E, Representative pictures of a root of each genotype 11 days after sowing on control or low phosphate condition. The scale represents 1 cm. F, Main root length ratio between root of 11-day-old plants grown under low phosphate and that grown under control condition. Measures represent corrected means of all possible ratios (n ≥ 14, details in Supplemental Table S6). G to I, Expression levels of *NPC15*, *NPC34* and *NPC43* in root of 11-day-old plants grown under control or low phosphate conditions of lines deregulated in the corresponding non-coding gene. Measures represent corrected means of log 2 fold changes compared to Col (B and F) or WS (D). Error bars represent standard error (n ≥ 2, details in Supplemental Table S5). F to I, results were analyzed by two-way analysis of variance (ANOVA) followed by Tukey's HSD posthoc test: groups with different letters are statistically different ($p \leq 0.05$).

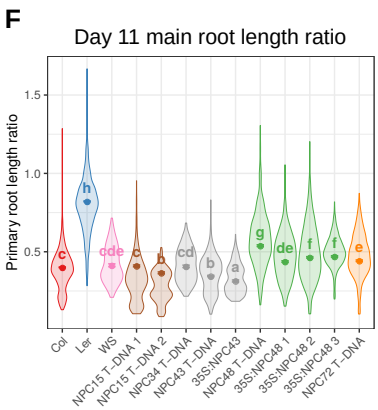


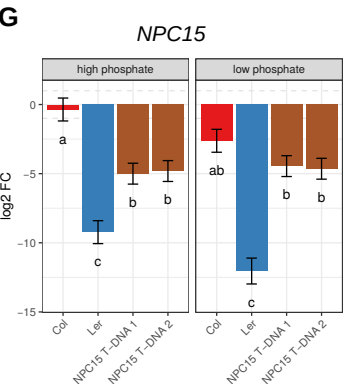


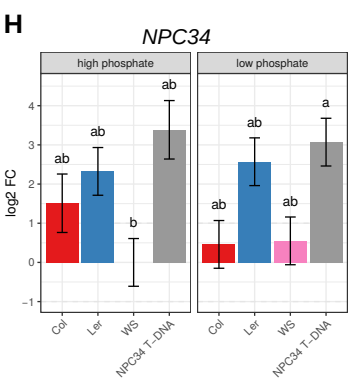


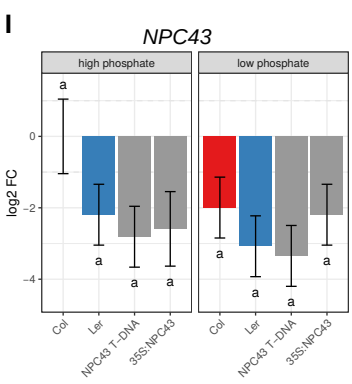






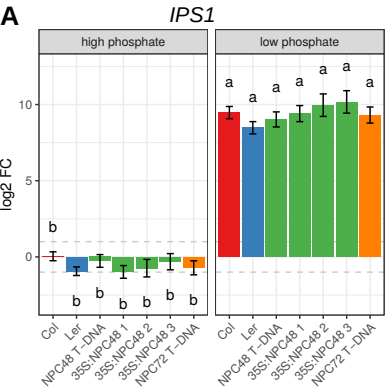


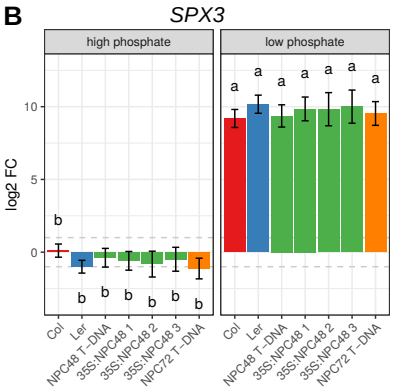


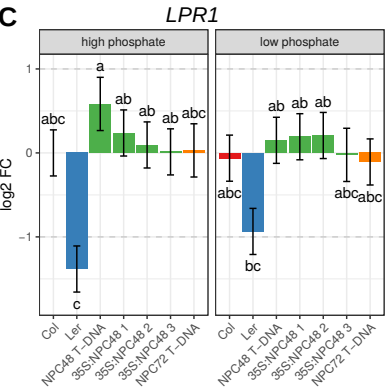


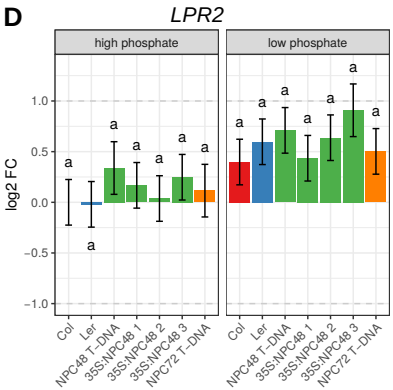
Supplemental Figure S9 - Deregulation of *NPC48* and *NPC72* do not change the expression of phosphate starvation related genes

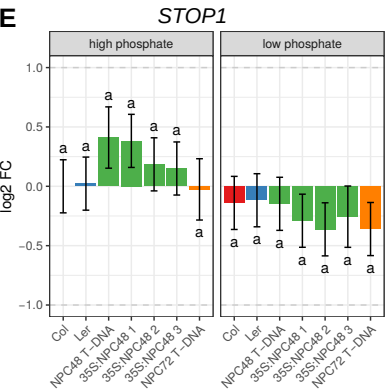
A to G, level of expression in root of 11-day-old plants grown under high phosphate condition or low phosphate condition of genes involved in phosphate sensing, *IPS1* (A) and *SPX3* (B), or in the response to phosphate related growth arrest, *LPR1* (C), *LPR2* (D), *STOP1* (D), *AMLT1* (F) and *MATE* (G). Measures represent corrected means of log 2 fold changes compared to Col measured by RT-qPCR. Error bars represent standard error (n ≥ 3, details in Table S5). Results were analyzed by two-way ANOVA followed by Tukey's HSD posthoc test: groups with different letters are statistically different ($p \leq 0.05$).

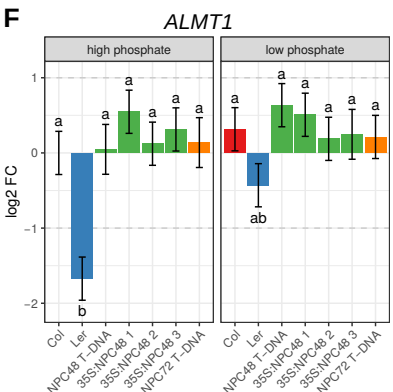


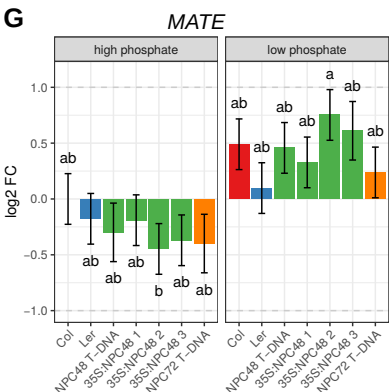






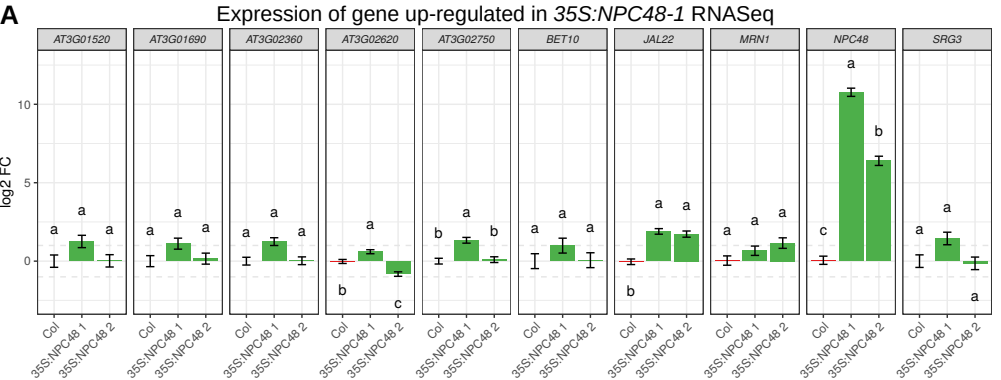


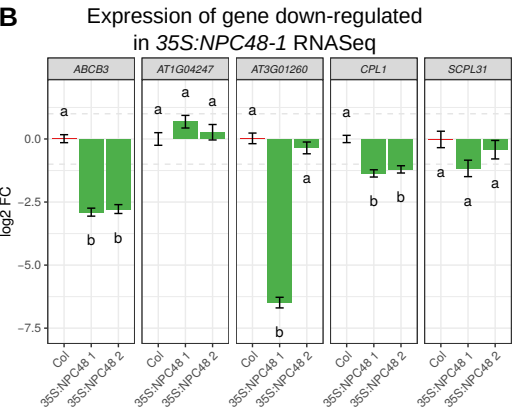




Supplemental Figure S10 - Expression analysis of genes regulated in 35S:NPC48-1 line in a different independent transgenic line

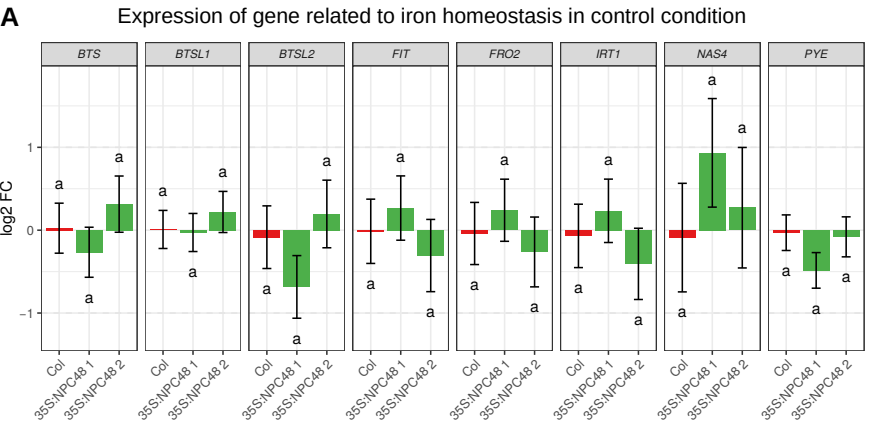
A and B, Expression levels in root of 11-day-old plants grown under control condition of genes differentially expressed in the RNA-seq between Col and *35S:NPC48-1*. Note that *AT1G2160* is only de-regulated in one transgenic line. Measures represent corrected mean log 2 fold changes compared to Col measured by RT-qPCR. Error bars represent standard error (n ≥ 3, details in Table S5). Results were analyzed by one-way ANOVA followed by Tukey's HSD posthoc test: groups with different letters are statistically different ($p \leq 0.05$).

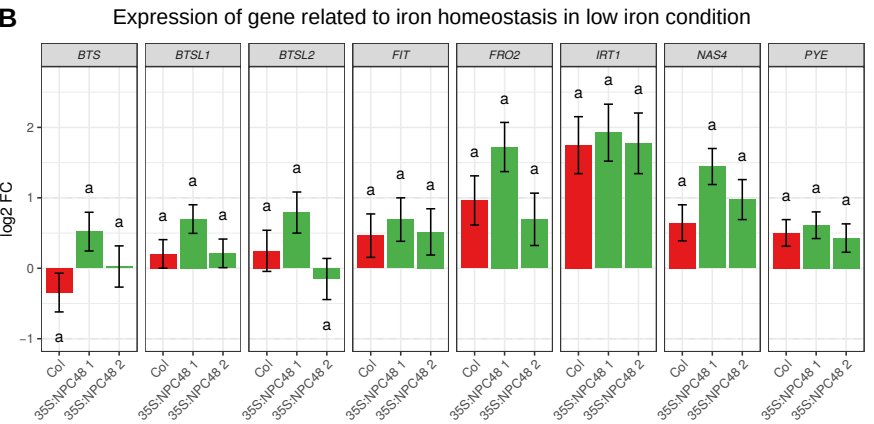


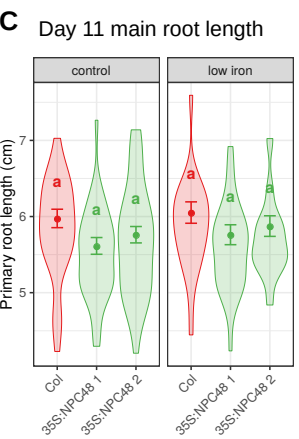


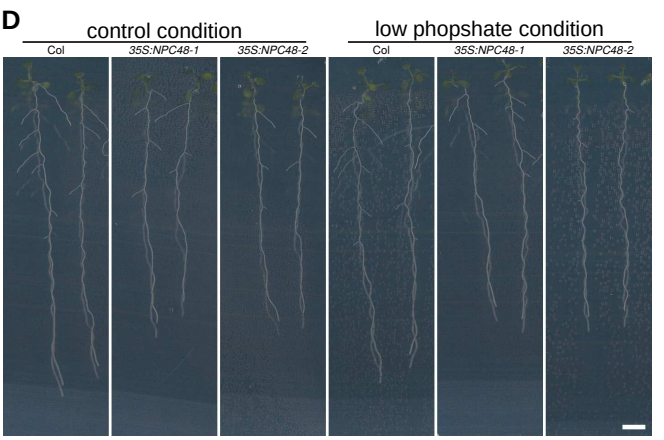
Supplemental Figure S11 - The over-expression of *NPC48* does not affect iron homeostasis and/or phenotypic responses

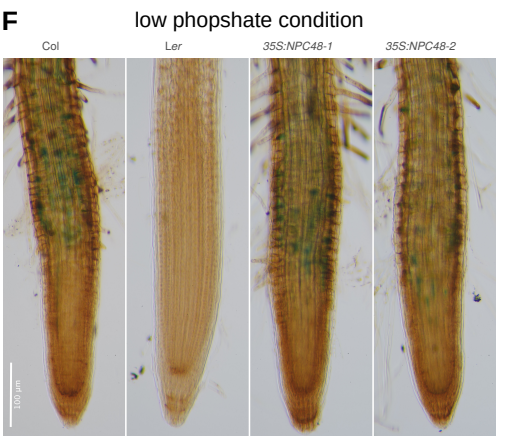
Expression levels in root of 11-day-old plants grown under control conditions (A) or low phosphate conditions (B) of genes related to iron homeostasis. Description of gene names is shown in Supplementary Table S9. Measures represent log 2 fold changes compared to Col measured by RT-qPCR. No statistically significant differences in the expression of these genes were found (n ≥ 2, details in Table S5). C, mean primary root length according to genotype and iron condition 11 days after sowing (n ≥ 26, details in Table S6). D, Representative pictures of roots for each genotype 11 days after sowing in control or low phosphate conditions, the scale represents 0.5cm. E and F, Fe distribution in primary root tips. Four-days old seedlings of the indicated genotypes, grown on high phosphate, were transferred to high phosphate (E) or low phosphate (F) plates for 48 h prior to Perls/DAB staining. Images are representative of two independent experiments. the scale represent 100µm. Measures represent corrected means of the FC (A, B) or length of the primary root (C). Error bars represent standard error. Results were analyzed by one-way ANOVA followed by Tukey's HSD posthoc test: groups with different letters are statistically different ($p \leq 0.05$).











Supplemental Materials and Methods

New transcript identification

All reads were quality trimmed using Trimmomatic and remaining ribosomal sequences were removed using sortMeRNA (Kopylova et al., 2012). According to their ecotype of origin, mRNA cleaned reads were aligned on TAIR10 (Lamesch et al., 2012) or Ler v7 (Gan et al., 2011) genome using TopHat2 (version 2.0.13, (Kim et al., 2013)) with the following arguments: `--max-multihits 1 --num-threads 8 -i 20 --min-segment-intron 20 --min-coverage-intron 20 --library-type fr-firststrand --microexon-search -I 1000 --max-segment-intron 1000 --max-coverage-intron 1000 --b2-very-sensitive`. Independently of ecotype, new transcripts were predicted using GFFprof included in RNAprof (Tran et al., 2016)

The transcripts predicted on each genome, TAIR10 and L*er* v7, were positioned on the other genome, respectively L*er* v7 and TAIR10, and on L*er* v8 (Zapata et al., 2016) using blastn (from BLAST suite 2.2.29+) using a maximum e-value of 10^{-4}. For each transcript, the different blast hits fragments were fused together if the distance between two fragments was less than 5000 nucleotides and placed on the same strand of the chromosome. Only hits with at least 90 % of sequence identity and where the length was conserved (at least 90 % and less than 110 % of length outside of insertion) were kept. For each transcript only the best hit was conserved according first to the conservation of the sequence length and then identity. In case of hits of the same strength, a higher priority was given when the chromosome and then the strand were conserved. Each transcript was therefore placed on each of the three genomes.

For L*er*-predicted transcripts only those positioned on L*er* v8 were kept. On each genome independently, transcripts coming from the same ecotype (GFFprof prediction) or the other one (blastn positioning) were fused using cuffmerge (version 1.0.0) with default parameters. Only transcripts longer than 200nt in either ecotype were kept for further processing.

Based on the position of the transcripts on the TAIR10 genome, new transcripts were annotated according to already known transcripts in the following databases: Araport 11 (Cheng et al., 2017), RepTas (Liu et al., 2012), CANTATAdb (Szcześniak et al., 2016), miRBase v21 (Kozomara and Griffiths-Jones, 2014) lncRNAs predicted from Ben Amor et al. (2009) and root predicted lncRNAs from Li et al. (2016). GffCompare (version 0.10.4, https://ccb.jhu.edu/software/stringtie/gffcompare.shtml) was used for the comparison. In case of overlap with a known transcript (=, c, k, j, e, o codes of GffCompare), the closest transcript was used to determine the identification and the coding potential of the transcript. For previously non-discovered transcripts we used the COME software (Hu et al., 2017) to predict their coding potential.

Library saturation analysis

The saturation of libraries was computed using the `RPKM_saturation.py` script included in RSeQC v3.0.0 (Wang et al., 2012; Wang et al., 2016). Raw counts were quantified at gene level from resampled bam for each additional 2% of the reads from 2% to 100% of the reads (`-l 2 -u 100 -s 2`). The saturation was computed 100 times. For each library and each sampling the number of detected gene (known or new and coding or non-coding) was defined as the number of genes having at least one read.

Small RNA analysis

The cleaned small RNA reads were aligned on TAIR10 or Ler v7 genome using ShortStack (version 3.8.5, Johnson et al. (2016)) without mismatch (`--mismatches 0`), keeping all primary multi-mapping (`--bowtie_m all`) and correcting for multi-mapped reads according to the uniquely mapped reads (`--mmap u`).

For each annotation in Araport11 (mRNA coding and non-coding and TE) and each new annotation predicted in this study according to mRNA sequencing the accumulation of small RNA was analyzed using ShortStack with default parameters. The counts for 21nt and 22nt were summed for each sample. Using these new counts, the DicerCall, defined as the size of the majority of the reads of a cluster, was recomputed. The other description and counting are according to ShortStack prediction.

Expression analysis

For each annotation, coding and non-coding, mRNA reads were counted with htseq-count (Anders et al., 2015) using strand specific and intersection strict mode (`--stranded=reverse -t gene --mode=intersection-strict`). These counts were used for differential gene expression analysis with DEseq2 (v1.16.1(Love et al., 2014)) using a linear model and as factors the ecotype (two levels), the kinetic time (3 levels), the interaction between the two and the replicate (3 levels). Differential comparisons between the two ecotypes were computed in average over the 3 time points of the kinetic. Differential comparisons between each point of the kinetic were computed on average over the 2 ecotypes. Low counts were discarded using DESeq2 independent filtering with default parameters and raw p-values were adjusted with the Bonferroni method. Differentially expressed genes were defined as having an adjusted p-value lower than 0.01.

Differential siRNA accumulation was computed using DESeq2 with a model taking into account only the genotype (two levels) as factor, and using the counts of ShortStack. Differential accumulation was computed independently for the 21/22nt on one side and the 24nt on the other side and limited to coding and non-coding genes. Bonferroni correction of the p-value was used and differential siRNAs were defined as having an adjusted p-value inferior to 0.01.

Statistics and reproducibility of experiments

Statistical analyses were performed using R (v3.4.2 (R Core Team, 2017)) with the help of the `tidyverse` (v1.2.1 (Wickham, 2017)) and `emmeans` packages (Lenth, 2019). For each measure (root length, qPCR expression level, experimental repetition), the least-squares means were computed taking into account all the factors (genotype and condition) in a linear model. This allows correcting for inter-repetition variation. Data are presented as least-squares means±SEM.

The results for statistical significance tests are included in the legend of each figure. 'n' values represent the number of independent samples in a repetition, i.e. the number of roots or pools of root per condition. The number of independent experiments is denoted as "repetition". For each analysis, the detail number of repetition and number

of sample per repetition are available in Supplemental Table S5 and Supplemental Table S6.

PERLs-DAB staining

PERLs-DAB-staining was adapted from Roschzttardtz et al. (2009). Plants were incubated 15 minutes in 0.5% (v/v) HCl, 2% (w/v) K-ferrocyanide, washed twice with dH2O and incubated 15 minutes in methanol containing 10 mM Na-azide and 0.3% (v/v) H2O2. After washing with 100 mM Na-phosphate buffer (pH 7.4), plants were incubated 5 min in the same buffer containing 0.0125% (w/v) DAB (Sigma-Aldrich) and 0.005% (v/v) H2O2. The reaction was stopped by washing twice with 100 mM Na-phosphate buffer (pH 7.4) and plants were optically cleared with 8:2:1 hydro-chlorate:water:glycerol (w/v). Roots were observed and documented with the Axiozoom-V16 macroscope Zeiss (microscopy equipment from the ZoOM, CEA Cadarache) using the 2.3x objective.

Supplemental Tables

Supplemental Table S1 - Mapping efficiency for each sequence sample

Ecotype	Time	replicates	original reads (mate1 + mate2)	mapped reads (mate1 + mate2)	
				number	from original
Col	0	col_t0_A	48886542	47745820	97.67%
		col_t0_B	49805886	48532469	97.44%
		col_t0_C	49359434	48386295	98.03%
	1h	col_1_A	53743064	52363370	97.43%
		col_1_B	58599074	57292554	97.77%
		col_1_C	61443288	60383962	98.28%
	2h	col_2_A	57988404	56707066	97.79%
		col_2_B	51030526	49897582	97.78%
		col_2_C	54236002	53278567	98.23%
Ler	0	ler_t0_A	54211518	52284390	96.4%
		ler_t0_B	58151120	56716239	97.5%
		ler_t0_C	55034548	53704417	97.6%
	1h	ler_t15_A	56236294	54483784	96.9%
		ler_t15_B	64959674	62942186	96.9%
		ler_t15_C	51056408	49324788	96.6%
	2h	ler_t1_A	57013228	55131860	96.7%
		ler_t1_B	51457646	50199625	97.6%
		ler_t1_C	58824614	57416162	97.6%

Supplemental Table S4 - Sequence of primers used in this study

Gene	Forward primer	Reverse primer
Col_NEW_RNA_F_21743	CATGAAATCGGGAGAGCAACCACA	ACCACAAGATCCGGTTTGAAGTTGAT
Col_NEW_RNA_F_7035	TAGCGGTCATGGCCGGTTTATGT	CGGCTTAAGGAATCGTGATAAACGACG
Col_NEW_RNA_R_17895	AAGTCCTCCGCCGACGATGGG	CGACGTCGCAGCAATTGGGCT
Col_NEW_RNA_R_19102	AGGGATAAGCTCTTCTTTTTCCCCTCA	GTGAATCTAGCCGTGAGTTTGGTGAT
Col_NEW_RNA_R_19128	AAACGAAACTTCTTGTCGGAGAACAGC	GGGATATTGTTGGAGCAAAACAAGCGA
NPC29	TCGAGCCAGCCATCAAACACAGG	GCGTTTGCGCCGTTCACCAT
NPC51	CTCCGTCGACGATACCAAAGGCG	CGTCGATCATCATTCCGGCGGT
Ler_NEW_RNA_F_10801	TCCTAGATCCCGTCGCTCAAAAATGC	GCGGCTGCGACGATGGAGAAA
Ler_NEW_RNA_F_11139	ATCGCCACAAACACCACCACCA	CGGTTCTCGCGACGATGGCT
Ler_NEW_RNA_F_14485	ACATGGTACACGACCCAACCCATCT	TGATCGCCACTGATTTTGAGAGCCG
Ler_NEW_RNA_F_19778	ACCCGGAAAAAGACAGAACAAACATGC	CTTCCCTCCCATCTCCTTCCTCTGT
Ler_NEW_RNA_F_23887	TGAGTCACCGCCTCGGGAGC	CGTCCTAGGTCGGGTCGGGG
Ler_NEW_RNA_R_24296	GGAGCTTGCCTTCAAGTACCGC	AACGAGATCTTTCCGAGCCGTGA
Ler_NEW_RNA_R_4778	TGCATACGTCTTTTCAACTCATCCACG	ATAGCTGCGGAGGCACAAGAGAA
NPC15	TCCAATGCTCAAATAATCTCTTG	AACTCGTCAAAAGACCCAATTT
NPC34	CTAGCAACAGAGACCAACCC	GTGGCTTCCATAGCGCCGGA
NPC43	GAACAGTCCGACTCCAAGCC	CAAAAGTCACGATTTGGCGAGT
NPC72	ACCGGAATCAGCTCAACG	CGACAACCGGGATATACCAC
NPC48	TGCCTGTTCTTCAAATCAACAC	ACCAACAATTGGACGAAGAATC
IPS1	ATGGCGAAATGGTTCTGCTA	CCTCCTCTTGTTCGCTTGTC
SPX3	TGAAGACTGCAGAAGGCTGA	CGAAGCTTGCCAAAGGATAG
LPR1	ACTGGCGGCAGGATTGGAGGA	CCAACCCCTCTCGTGTGCCG
LPR2	GCCTGGCTCTTGAGTTGCGTCA	CGGTTACGGTACACGAGCGGG
STOP1	ACCCACGAGAAGCACTGCGG	GCAGGCGTGTGTCCCTGGAA
ALMT1	GGCAGTGTGCCTACAGGATT	TGAGTTTCCCGATTCCGAGC
MATE	GCATAGGACTTCCGTTTGTGGCA	CGAACACAAACGCTAAGGCA
AT3G01520	GCGGTGAACGCTTCGACGAT	TCGACGACTTGTACGTGGAGCA
AT3G01690	CCCTTTCCCCCACCGCGAAA	GCGGCGTTGCCATGGGAGTA
AT3G02360	ACCGGCATGAAAGGCACAGGA	AGCCCGCTAAGGAACCTCGC
AT3G02620	TGCCGCAGAGGATCAGGAGGT	TCTCTCCCGTAAATCCAGCTGAACGA
AT3G02750	TCCGGTTGGGTTTCTGTGTTCTCCTC	TGCTGCTGCTGCTGCTGCTT

BET10	GCTTGACGACGTCCACAGGGG	AGGGTGCTCCCCTTCACAGCA
JAL22	GGTGGTGGTGATGGGCACGA	CGCTTGGAGAGGAAGGCCACA
MRN1	ACTTTTTCACTGGTCGGTACCGAAGAC	TGCATAAAGTGTCGCACTCATTCGACTA
SRG3	CTGCATGTAAACCGCCTCGGCT	CCGCAACCCCGCTCTTGTCA
ABCB3	GCCTCGGATGCTTCATCACCGC	GGGGCTTGTAGGGCAAGAGCCT
AT1G04247	ACCTACAAAATAAGCCTGGCACAACC	GGTGGGGAGGTTGGGACTTTTCT
AT3G01260	CGGAGTTGGCGAAAATCGAAGGGT	TCGGTTCCGGTTATACTCATCTCGACG
CPL1	TGTAGTGGCGAATACCATGCGCTC	GGCAAGGCGTTGAGGGTCCA
SCPL31	TCACTCGCATGAAGAGACTTCTGAACA	CACCGAGGCTCATGGGTGGA
BTS	ACCATGTCGATCTCCGGCTG	CAAGAGAATATGTTTGCGCTACATT
BTSL1	TGTCCCTTGGCTTCCAATGCTGG	TCGCGGAGTTTCAGCAACGGT
BTSL2	CCAGCATGTTCAATTCTGCCAACTCA	TCCAGCTTTGGAGGCAAAGGGT
FIT	TCGGTCTAGGACTTTGATCTCTG	TCTTGAACATACAACACTGCATCT
FRO2	TTCACCGTTCATGGTCTTTGTT	GAGCTATCTCTCCGGCCAAATT
IRT1	CTCTTTGCTTCCATCAAATGTTC	CCTAACGCTATTCCGAATGG
NAS4	TGTTCTTGGCTGCTCTTGTAGG	CAAGGCTCAACGATTGGATAGA
PYE	CAGGACTTCCCATTTTCCAA	CTTGTGTCTGGGGATCAGGT

Supplemental Table S7 – Software used for the bioinformatic analyses

Name	Version	Reference	Non default arguments
TopHat2	2.0.13	Kim et al. (2013)	`--max-multihits 1 --num-threads 8 -i 20 --min-segment-intron 20 --min-coverage-intron 20 --library-type fr-firststrand --microexon-search -I 1000 --max-segment-intron 1000 --max-coverage-intron 1000 --b2-very-sensitive`
GFFprof	1.2.6	Tran et al. (2016)	
blastn	2.2.29+	https://www.ncbi.nlm.nih.gov/books/NBK279690/	
cuffmerge	1.0.0	http://cole-trapnell-lab.github.io/cufflinks/cuffmerge/	
GffCompare	0.10.4	https://ccb.jhu.edu/software/stringtie/gffcompare.shtml	
COME	1.0	Hu et al. (2017)	
RSeQC	3.0.0	Wang et al. (2012) Wang et al. (2016)	`-l 2 –u 100 –s 2`
ShortStack	3.8.5	Johnson et al. (2016)	`--mismatches 0 --bowtie_m all --mmap u`
htseq-count	0.6.1	Anders et al. (2015)	`--stranded=reverse -t gene --mode=intersection-strict`
R	3.4.2	R Core Team (2017)	
DESeq2	1.16.1	Love et al. (2014)	
tidyverse	1.2.1	Wickham (2017)	
emmeans	1.3.3	Lenth (2019)	

2 ***Supplemental Table S9 – Gene name abbreviation***

Name	Short Name	AGI
INDUCED BY PHOSPHATE STARVATION1	*IPS1*	*AT3G09922*
SPX DOMAIN GENE 3	*SPX3*	*AT2G45130*
LOW PHOSPHATE ROOT1	*LPR1*	*AT1G23010*
LOW PHOSPHATE ROOT2	*LPR2*	*AT1G71040*
SENSITIVE TO PROTON RHIZOTOXICITY 1	*STOP1*	*AT1G34370*
ALUMINUM-ACTIVATED MALATE TRANSPORTER 1	*ALMT1*	*AT1G08430*
MATE	*MATE*	*MATE*
BROMODOMAIN AND EXTRATERMINAL DOMAIN PROTEIN 10	*BET10*	*AT3G01770*
JACALIN-RELATED LECTIN 22	*JAL22*	*AT2G39310*
MARNERAL SYNTHASE 1	*MRN1*	*AT5G42600*
SENESCENCE-RELATED GENE 3	*SRG3*	*AT3G02040*
ATP-BINDING CASSETTE B3	*ABCB3*	*AT4G01820*
C-TERMINAL DOMAIN PHOSPHATASE-LIKE 1	*CPL1*	*AT4G21670*
SERINE CARBOXYPEPTIDASE-LIKE 31	*SCPL31*	*AT1G11080*
BRUTUS	*BTS*	*AT3G18290*
BTS LIKE1	*BTSL1*	*AT1G74770*
BTS LIKE2	*BTSL2*	*AT1G18910*
FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR	*FIT*	*AT2G28160*
FERRIC REDUCTION OXIDASE 2	*FRO2*	*AT1G01580*
IRON-REGULATED TRANSPORTER 1	*IRT1*	*AT4G19690*
NICOTIANAMINE SYNTHASE 4	*NAS4*	*AT1G56430*
POPEYE	*PYE*	*AT3G47640*

Supplemental references

Amor B Ben, Wirth S, Merchan F, Laporte P, D'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, et al (2009) Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. Genome Res **19**: 57–69

Anders S, Pyl PT, Huber W (2015) HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics **31**: 166–169

Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J **89**: 789–804

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, et al (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature 477: 419–423

Hu L, Xu Z, Hu B, Lu ZJ (2017) COME: A robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. Nucleic Acids Res **45**: e2

Johnson NR, Yeoh JM, Coruh C, Axtell MJ (2016) Improved Placement of Multi-mapping Small RNAs. G3: Genes|Genomes|Genetics **6**: 2103–2111

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14**: R36

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

Kozomara A, Griffiths-Jones S (2014) MiRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. doi: 10.1093/nar/gkt1181

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40: D1202–D1210

Lenth R (2019) emmeans: Estimated Marginal Means, aka Least-Squares Means.

Li S, Yamada M, Han X, Ohler U, Benfey PN (2016) High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Dev Cell **39**: 508–522

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H (2012) Genome-Wide Analysis Uncovers Regulation of Long Intergenic Noncoding RNAs in Arabidopsis. Plant Cell **24**: 4333–4345

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. doi: 10.1186/s13059-014-0550-8

R Core Team (2017) R: A language and environment for statistical computing. http://www.R-project.org/. R Found. Stat. Comput. Vienna, Austria,

Roschzttardtz H, Conéjéro G, Curie C, Mari S (2009) Identification of the endodermal vacuole as the iron storage compartment in the Arabidopsis embryo. Plant Physiol **151**: 1329–1338

Szcześniak MW, Rosikiewicz W, Makałowska I (2016) CANTATAdb: A collection of plant long non-coding RNAs. Plant Cell Physiol **57**: e8

Tran VDT, Souiai O, Romero-Barrios N, Crespi M, Gautheret D (2016) Detection of generic differential RNA processing events from RNA-seq data. RNA Biol **13**: 59–67

Wang L, Nie J, Sicotte H, Li Y, Eckel-Passow JE, Dasari S, Vedell PT, Barman P, Wang L, Weinshiboum R, et al (2016) Measure transcript integrity using RNA-seq data. BMC Bioinformatics 17: 1–16

Wang L, Wang S, Li W (2012) RSeQC: quality control of RNA-seq experiments. Bioinforma Oxford Engl **28**: 2184–2185

Wickham H (2017) tidyverse: Easily Install and Load the "Tidyverse."

Zapata L, Ding J, Willing E-M, Hartwig B, Bezdan D, Jiao W-B, Patel V, James GV, Koornneef M, Ossowski S, et al (2016) Chromosome-level assembly of Arabidopsis thaliana L er reveals the extent of translocation and inversion polymorphisms. Proc Natl Acad Sci **113**: E4052–E4060

6. Exploring transcriptomes to find new cis-regulatory root-related lncRNAs

6.1 LATERALINC, new regulator of lateral root growth in Arabidopsis

6.1.1 Introduction

Long non-coding RNAs (lncRNAs) with their low abundance and poor level of conservation were initially thought to be non-functional or "junk" DNA. Nowadays, it is clear that many lncRNAs are functional, and constitute important regulators of cell life. For example, the *Xist* lincRNA is the major actor of the X-inactivation process in mammal, disturbing its functionality leads to a plethora of medical significance such as cancer predisposition (Brown, 1999), Rett syndrome (Huppke et al., 2006), autoimmunity (Simmonds et al., 2014), autism (Talebizadeh et al., 2005) and recurrent miscarriage (Sullivan et al., 2003). In plants, several lncRNAs rising from the *FLC* locus are necessary for a normal flowering, ensuring the transition to the next generation (Jarroux et al., 2017). Outside of *FLC* regulation, bioinformatic analyses revealed nearly fifty lncRNA with increased abundance in roots (Amor et al., 2009). Interestingly, ectopic overexpression of two of these lncRNAs affects root growth in *Arabidopsis* (Amor et al., 2009). One of them, namely *APOLO*, has been intensively studied and is now known as an important regulator of the plant auxin and cold stress response (Ariel et al., 2014; Ariel et al., 2020; Moison et al., 2021). Also, two ecotype-enriched lincRNAs, namely *NPC48* and *NPC72* in L*er* and Col-0, respectively, disturbs primary root growth when overexpressed in Col-0 (Blein et al., 2020). LncRNAs can regulate gene activity in *trans* or in *cis* by modifying the expression of a distant or a neighboring gene, respectively. In plants, several *cis*-regulatory lncRNAs have been described and are involved in flowering control (*FLC* related lncRNAs), auxin transport (*APOLO*), grain yield (*LAIR*) and freezing tolerance (*SVALKA*) (Chen et al., 2021). To date, apart from *APOLO* (Ariel et al., 2014), no cis-RdDM-acting lncRNA has been described as root growth regulator. Thus, we decided to look for novel root-related lncRNAs using the transcriptomics data from Blein et al (2020) as template for the analysis. Taking advantage of the data on long and small transcriptomes, we focused our search on lincRNAs able to generate 24nt siRNA, thus, likely functioning through RdDM regulation.

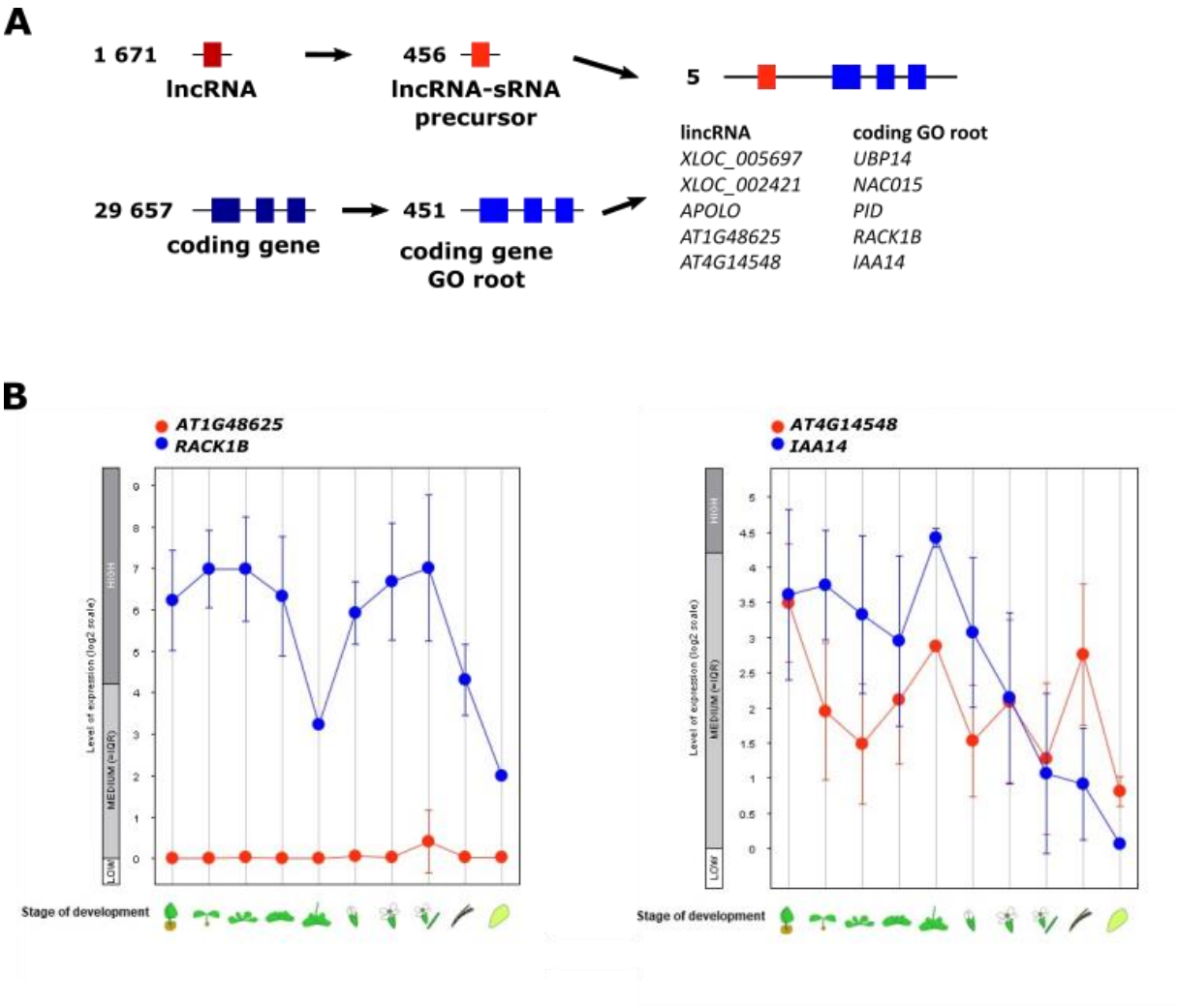


Figure 1. Identification of the *LATERALINC* lincRNA. A. Flowchart of identification of lncRNA-sRNA neighbors to root-related genes. LncRNAs co-localizing with 24nt length sRNA and the neighbor coding gene with a Gene Ontology related to root growth and development were collected using the Blein et al (2020) transcriptomic analyses and the GO consortium (Berardini et al 2004), respectively. The lncRNAs-sRNA potential precursor neighbors to a root-related gene were isolated. The resulting five lncRNAs and the name of the neighboring gene are indicated on the right. **B.** Transcript abundance of lncRNA-sRNA and neighbors' genes throughout plant life for two lncRNA/coding gene couples. Genevestigator snapshot from the Development condition search tool (Hruz et al., 2008). **C.** Schematic illustration of the different isoforms of *AT4G14548/LATERALINC* genomic region. First line corresponds to the genomic region whereas the other lines present the various isoforms of *AT4G14548/LATERALINC* and *AT4G14550/IAA14*. For each isoform, exons are indicated with rectangles and introns with solid lines. *LATERALINC* has a single exon.

In this second chapter, I investigated the transcriptomic analyses from the first chapter (Blein et al., 2020) to search for an RdDM-cis-acting lincRNA neighbor to a critical root growth modulator gene on a list derived from the literature (Berardini et al., 2004). We expect to identify in this way, lncRNAs modulating the root growth through an eventual regulation of the neighboring gene. Among many others, I found that the deregulation of an intergenic lncRNA (lincRNA) neighbor to the gene *IAA14*, a main regulator of the auxin-mediated Lateral Root (LR) development, disturbs the root system architecture. We named it *LATERAL ROOT LINCRNA* (*LATERALINC*). Despite this genomic position, the *LATERALINC* is not responsive to auxin and its silencing does not change *IAA14* transcript abundance. Altogether, the *LATERALINC* may regulate LR growth or development through a *IAA14*-independent pathway.

6.1.2 Results and discussion

6.1.2.1 *LATERALINC* is positively correlated with IAA14 during plant development

Among the 1 671 lincRNAs detected in *Arabidopsis thaliana* root tips (Blein et al., 2020), 456 co-localize with 24 nt siRNAs, potentially acting through RdDM. Interestingly, five lincRNAs/siRNA precursors are neighbors to root-related coding genes: *XLOC_005697, XLOC_002421, AT1G48625, AT2G34655* (*APOLO* lncRNA), and *AT4G14548* (**Figure 1A**). Using the Genevestigator database (Hruz et al., 2008), and to strengthen the relationship between the lincRNA and its putative neighbor target gene in the control of root architecture, we look for uncharacterized lincRNAs and target genes expressed in the root compartment and positively correlated during plant development. *XLOC_005697* and *XLOC_002421* do not have public identifiers and thus cannot be searched in this data. The *AT1G48625* lincRNA was only detected in inflorescences and poorly correlated with its putative target *AT1G48630/RACK1B* during plant development (**Figure 1B**). Interestingly, both *AT4G14548* lincRNA and its putative target *AT5G14550/IAA14* are detected in root and positively correlated during plant development (**Figure 1B**). Notably, their expression seems to decrease throughout plant life, reaching their lower level of expression in flower-related organs (**Figure 1B**). Intriguingly, *AT4G14548* lincRNA is separated from the *IAA14* gene 3′ ends by only 59bp (**Figure 1C**).

Twenty-nine *Aux/IAA* genes are annotated in *Arabidopsis* and encode short-nuclear proteins involved in the inactivation of the *Auxin Response Factors* (*ARF*) proteins, consequently repressing the auxin transcriptional responses. Auxin stimulus increases the interaction between *Aux/IAA* proteins and the *Skp1-Cullin-F-box/Transport Inhibitor Response 1* (*SCF-TIR1*) complex, promoting the proteasome mediated degradation of the *Aux/IAA* proteins. Notably, gain-of-function mutations of 10 *IAA* proteins affect plant development. Among them, *iaa14/solitary root* (*sli*) gain of function mutation stops the auxin-mediated cell division in the pericycle, blocking the development of LR (Fukaki et al., 2005; Vanneste et al., 2005). Hence, *AT4G14548* lincRNA could influence the LR development through the regulation of *IAA14* gene expression. Thus, we decided to name it *LATERALINC* for *LATERAL root LINCrna*.

6.1.2.2 Lateral root growth is impaired in *LATERALINC* downregulated lines

To assess the physiological role of *LATERALINC*, we decided to test the physiological impact of *LATERALINC* downregulation on root architecture. To this end, we modified *LATERALINC* expression without affecting the encoding DNA region using an RNAi construct triggering post-transcriptional lincRNA degradation, and isolated three independent lines (RNAi *LATERALINC* 1, RNA *LATERALINC* 2 and RNAi *LATERALINC* 3). As expected, the downregulation of *LATERALINC* does not affect the primary root length (**Figure 2A**) but triggers a significant reduction of LR length compared to Col (**Figure 2B**). Notably, the LR inhibition was stronger in the RNAi line 2 and line 3 as compared to the line 1 (**Figure 2B**). Taken together, the deregulation of *LATERALINC* through RNAi inhibition influences the LR length, hinting the possibility that *LATERALINC* modulates the transcriptional activity of *IAA14* or other LR-development related genes.

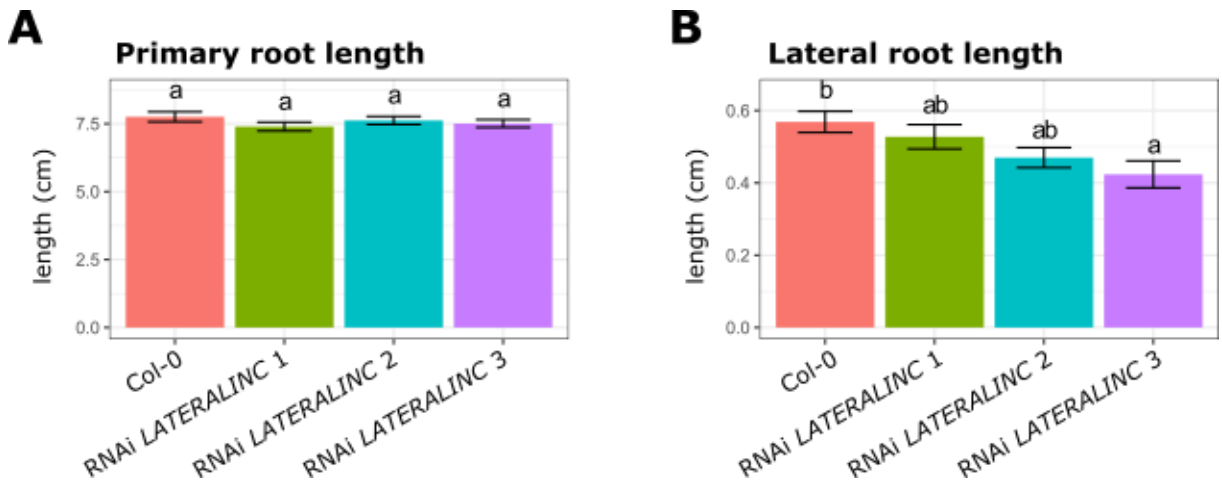


Figure 2. *LATERALINC* shaped the root architecture. For Col and RNAi *LATERALINC* lines; **A.** mean primary root length and **B.** mean lateral root length. Letters indicate a statistical group determined by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test. For each condition, letters indicate statistical differences between genotypes ($p \leq 0.05$).

6.1.2.3 *LATERALINC* expression is correlated with the one of *IAA1* but its downregulation does not affect *IAA14* and *IAA1* genes expression

Thus, we assessed the transcriptional activity of *IAA14*, key regulator of LR development (Fukaki et al., 2005; Vanneste et al., 2005) within the *LATERALINC* RNAi lines. Transcript abundance of *LATERALINC* was significantly decreased in the RNAi lines, as expected. Remarkably, the level of downregulation of *LATERALINC* correlates with the LR phenotype (**Figure 3A and 2B**), strengthening the relationship between LR length and *LATERALINC* expression. However, *LATERALINC* downregulation does not affect the *IAA14* gene expression (**Figure 3A**), pointing that the *LATERALINC*-mediated LR growth may occur independently of this gene. Nevertheless, as *IAA14* gene activity increased in response to auxin (Fukaki et al., 2005; Vanneste et al., 2005), we wondered whether the *IAA14* auxin-mediated transcript accumulation is disturbed upon *LATERALINC* downregulation. Nonetheless, the increased transcript abundance of *IAA14* in response to auxin were comparable between the RNAi lines and Col strengthening that *LATERALINC* acted independently of *IAA14* to modulate the LR length (**Figure 3B**). Thus, we look at other closely located genes (using available transcriptomics data from Araport11) in case *LATERALINC* act in *cis* on other neighbors (**Figure 3C**). Strikingly, we found that *LATERALINC* transcriptional accumulation positively correlates with *AT4G14560*/*IAA1* located at 10.5 kp from the 5' end of the *IAA14* gene. *IAA1* is also a substrate of *SCF-TIR1* like *IAA14* but it is involved in cell elongation and differentiation in the aerial parts of *Arabidopsis* plants (Yang et al., 2004; Ku et al., 2009) and not in roots. Despite this positive correlation observed in databases (**Figure 3C**), the RNAi-mediated *LATERALINC* downregulation does not affect the *IAA1* transcript abundance in control condition or in response to auxin (**Figure 3A and 3B**). Surprisingly, the *AT4G14540* auxin-mediated transcript accumulation is disturbed upon *LATERALINC* downregulation. More precisely, the *AT4G14540* gene transcripts start to accumulate 48 hours after the auxin application in Col whereas it accumulates sooner, as early at 8 hours for the RNAi lines 1-2 and 24 hours for the RNAi line 3, and return to a basal level at 24 and 48 hours for RNAi lines 1-2 and RNAi line 3, respectively (**Figure 3B**). Interestingly, *AT4G14540* encodes for NF-YB3 TF, implicated in heat stress tolerance (Sato et al., 2019). Notably, plants overexpressing *NF-YB3* showed an enhanced tolerance to a lethal heat stress (Sato et al., 2019). Taken together, *LATERALINC* does not regulate *IAA14 and IAA1* gene expression but may be involved in the transcriptional regulation of *NF-YB3* in response to auxin.

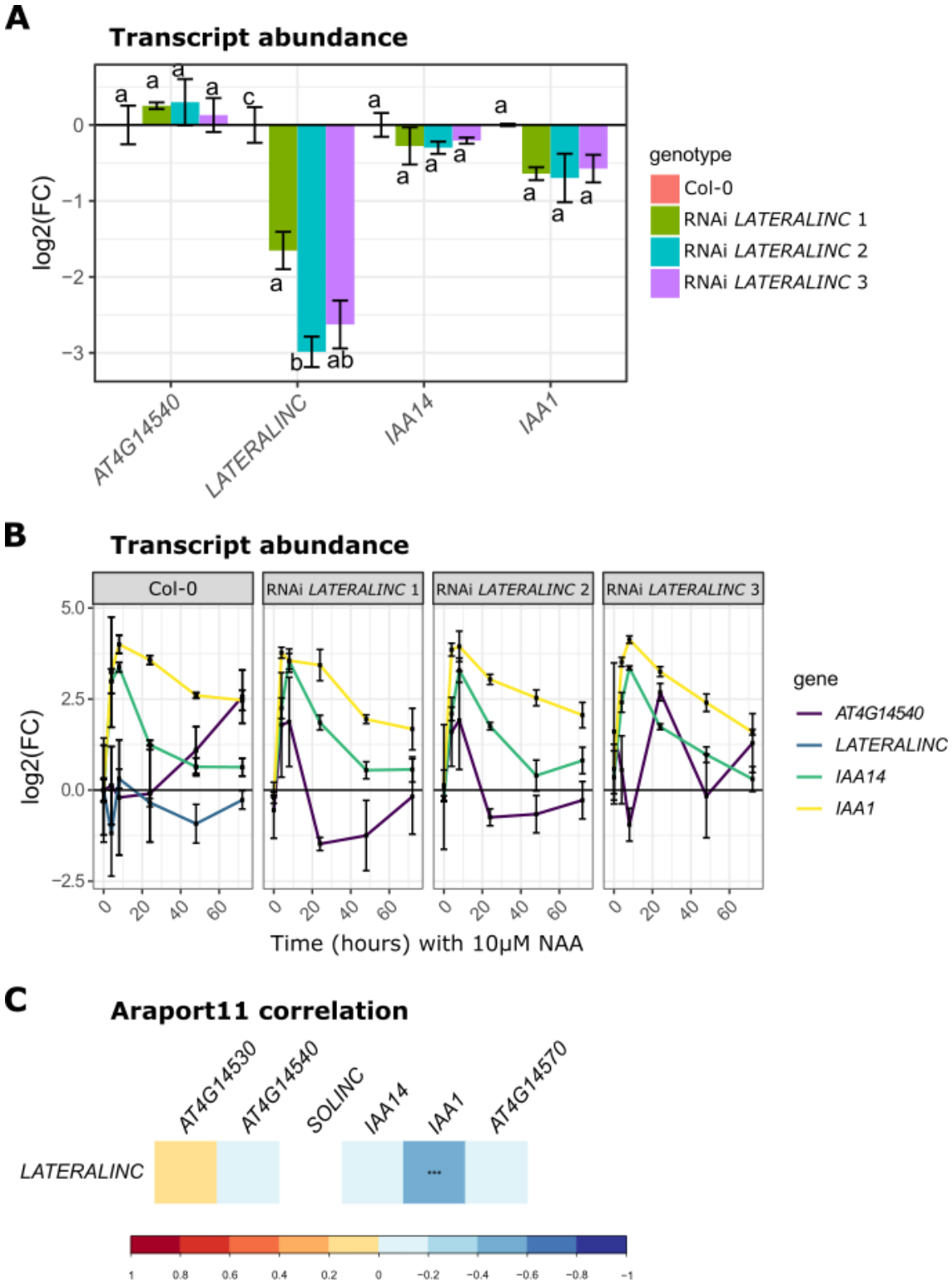


Figure 3. *LATERALINC* does not regulate *IAA14* but its expression correlates with the one of *IAA1*. Transcript levels of *LATERALINC* and its neighboring genes for Col and RNAi *LATERLINC* lines; **A.** in control condition and **B.** in response to auxin stimulus. Gene expression data are expressed as the mean ± standard error (n = 3) of the log2 fold change compared to the Col or Col at time 0h for A or B, respectively. For each gene, letters indicate statistical differences between genotypes ($p \leq 0.05$). **C.** 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 $p \leq 0.001$).

6.1.3 Methods

6.1.3.1 Data processing and statistical analysis

Except if stated otherwise, all data processing and statistical analyses were realized in R v3.4.2 (R Core Team, 2017) with the help of the tidyverse (v1.2.1; (Wickham et al., 2019)).

6.1.3.2 Identification of *LATERALINC*

The coding gene, lncRNA and siRNA annotations were collected using the transcriptomics data from Blein et al (2020). The root-related Gene Ontology (GO) was collected from the GO consortium (Berardini et al., 2004) and included all the GO terms related to root growth and development (see GO term and identifier in **Table 1**). The lncRNA-sRNA precursor neighboring genes were collected with bedtools closest using the default parameter (Quinlan and Hall, 2010).

6.1.3.3 Lines generation

All plants used in this study are in Columbia-0 background. RNAi-LATERALINC were obtained using the pFRN binary vector (Ariel et al., 2012) bearing 250bp of the 5' exon of *LATERALINC* gene (see primers in **Table 2**), previously sub-cloned into the pENTR/D-TOPO vector. Arabidopsis plants were transformed using *Agrobacterium tumefaciens* Agl-0 (Clough and Bent, 1998a).

6.1.3.4 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). Plants were grown on solid half-strength MS medium (MS/2) supplemented with 0.7% sucrose and supplemented with 0.8g/L agar (Sigma-Aldrich, A1296 #BCBL6182V), buffered at pH 5.6 with 3.4mM 2-(N-morpholino) ethane sulfonic acid. For root phenotype characterization, the root length was measured at 12 day after sawing (DAS) using RootNav software (Pound et al., 2013) from images taken with a flat scanner. For the treatment with auxin, seedlings were sprayed with 10µM 1-Naphthaleneacetic acid (NAA) at 12 DAS.

Table 1. List of Gene Ontology related to the root growth and development used for *LATERALINC* identification.

GO term	GO Identifier
root hair cell development	GO:0080147
root development	GO:0048364
primary root development	GO:0080022
regulation of root development	GO:2000280
post-embryonic root morphogenesis	GO:0010101
maintenance of root meristem identity	GO:0010078
root hair elongation	GO:0048767
root epidermal cell differentiation	GO:0010053
lateral root formation	GO:0010311
regulation of root meristem growth	GO:0010082
lateral root development	GO:0048527
embryonic root morphogenesis	GO:0010086
root meristem specification	GO:0010071
root meristem growth	GO:0010449
root hair cell differentiation	GO:0048765
regulation of root morphogenesis	GO:2000067
root radial pattern formation	GO:0090057
root cap development	GO:0048829
root hair initiation	GO:0048766
lateral root morphogenesis	GO:0010102
adventitious root development	GO:0048830
root hair cell tip growth	GO:0048768
root morphogenesis	GO:0010015
post-embryonic root development	GO:0048528
negative regulation of lateral root development	GO:1901332
regulation of lateral root development	GO:2000023
root system development	GO:0022622
lateral root branching	GO:0080181
root hair tip	GO:0035619
regulation of post-embryonic root development	GO:2000069
root hair	GO:0035618

6.1.3.5 RT-qPCR

Total RNA was extracted from roots 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 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 2**. Data were analyzed using the ΔΔCt method using *PROTEIN PHOSPHATASE 2A SUBUNIT A3* (*AT1G13320*) for gene normalization (Czechowski et al., 2005) and time 0 for time-course experiments.

6.1.4 Conclusion and perspectives

IAA14 is a key auxin-dependent regulator of LR development (Fukaki et al., 2005; Vanneste et al., 2005). Even though *LATERALINC* is a lincRNA involved in the quantitative control of LR length, *LATERALINC* downregulation does not seem to affect *IAA14* transcripts' abundance, hinting that the *LATERALINC*-mediated LR development is occurring independently of the *IAA14* neighboring gene. In addition, *LATERALINC* is not responsive to exogenous auxin application whereas *IAA14* gene expression increased (**Figure 3B**). Interestingly, the auxin kinetics pinpoints that the RNAi-mediated *LATERALINC* downregulation influences the expression of its neighboring gene *AT4G14540*/*NF-YB3*. The *NF-Ys* are a family of TF widely conserved within eukaryotes (Dolfini et al., 2012). In *Arabidopsis*, there are 3 main families of *NF-Ys (NF-YA, NF-YB* and *NF-YC)* which include around 10 genes in each class (Petroni et al., 2013). *NF-Ys* are involved in a wide-range of biological processes, among which, root nodule formation (Combier et al., 2006; Zanetti et al., 2010), flower development (Nakashima et al., 2009), seed maturation (Nambara et al., 1998) and abiotic stress responses (Yoshida et al., 2011). The rapid and transient upregulation of *NF-YB3* in response to auxin observed in the RNAi *LATERALINC* lines as compared to Col hints at a potential relationship between *LATERALINC* and *NF-YB3*. *NF-YB3* is induced under a mild heat stress and its ectopic overexpression in-plants increased the plant resistance to lethal heat stress (Sato et al., 2019) making noteworthy to investigate the influence of *LATERALINC* downregulation on the *NF-YB3* transcriptional responsiveness and plant resistance to heat stress, even though it is not directly linked to LR development. Furthermore, the *NF-YA* have been shown to regulate root architecture (Sorin et al., 2014), making of interest to test whether *NF-YB3* deregulated lines display subtle root phenotypes and whether it regulates *LATERALINC* expression.

Table 2. List of primer used in *LATERALINC* study.

Primers used to generate the RNAi lines	**Forward primer**	**Reverse primer**
LATERALINC	CACCTCCTTTCATCTACGGGATTTGCTGC	CCGTAAACGGCGGTGAGACA
Primers used for transcripts abundance analysis from cDNA	**Forward primer**	**Reverse primer**
AT4G14540	ACCGGTGAGGCTTCTGACAAGTG	GGCTCCACGTAGTCCTCAAACCC
LATERALINC	TGGTCGTTCCAACCATGCCAGAG	ACGGCGGTGAGACAAACCAACA
IAA14	AAGGCGACGGTTCCTCCACCA	CCACCGGTGAGGAACTACCGGAAA
IAA1	CGTTTGGGATTACCCGGAGCACA	TGTTGAGTCGTTGTTCTTGCGCTTG
AT1G13320 (housekeeping gene; Czechowski et al., 2005)	GAGCTGAAGTGGCTTCCATGAC	GGTCCGACATACCCATGATCC

The *LATERALINC* deregulation significantly reduces LR length suggesting that *LATERALINC* is a positive regulator of LR extension. Highly resolutive temporal phenotypic analyses of LR emergence and growth are needed to understand the reason behind the LATERALINC-mediated LR length decrease (i.e. Delay in the LR initiation? Lower LR growth rate? Early termination of LR growth?). LR emergence and growth are tightly regulated by a plethora of genes, involving auxin biosynthesis, transport, signaling and degradation, but also auxin unrelated genes such as cell wall remodeler, water transporter, other hormones such as ABA and/or ROS-related genes (Banda et al., 2019). Transcriptomic analyses of the *LATERALINC* RNAi lines may be informative to find the putative target genes of *LATERALINC*, helping to understand their LR length phenotype. Also, phenotypic and transcriptomic analyses of plants overexpressing *LATERALINC* may help to strengthen the relationship between *LATERALINC* and root architecture (i.e. do the *LATERALINC*-overexpressing plants present a reverted phenotype, a higher LR length? Does *LATERALINC* overexpression disturbs the transcriptional activity of key LR regulators?). Finally, as *LATERALINC* produces 24nt-length siRNA, its potential function through the RdDM pathway to regulate its target gene activity merits to be explored. Thus, the finding of differentially expressed genes within *LATERALINC* deregulated lines sharing a 24nt length sequence complementarity with *LATERALINC* and involved in root growth could constitute another interesting bioinformatics analysis to search for putative target genes linked to RdDM dependent-*LATERALINC* regulation. Indeed, all DNA sequences complementary to siRNAs are subjected to be targeted for methylation, as shown through genome wide methylation and transcript profiling of RdDM mutants (Herr et al., 2005; Onodera et al., 2005; Pontier et al., 2005).

Hence, *LATERALINC* opens new perspectives to understand how this root growth regulator impinges on the regulatory networks involved in LR length determination.

6.2 *MARS*, a lncRNA implicated in the transcriptional regulation of an embedded gene cluster

6.2.1 Introduction

Bacteria, fungi and plants can synthesize numerous bioactive secondary metabolic (SM) compounds principally to cope with their environments or to interact with other organisms. For many of these SMs, the metabolic pathways that govern their synthesis are currently unknown. Curiously, some of these organic molecules are generated by enzymes whose corresponding genes are co-localized at a specific chromosome locus. These latter genes are referred to as biosynthetic gene clusters (BGC) or metabolic gene clusters (Yu et al., 2016).

In plants, BGCs have been identified in both monocots and dicots; the size of these clusters varies from approximately 35 kb to several hundred kb and include three to ten genes (Nützmann and Osbourn, 2015). These clusters, arising by the recruitment of duplicated genes and neofunctionalization (Nützmann and Osbourn, 2015), are generally present in chromosome regions under strong selection pressure (Boutanaev et al., 2015). It has been proposed that gene clustering could be a strategy to allow co-inheritance of genes involved in the pathway, conferring a selective advantage to plants. Indeed, plant BGCs are either species specific or limited to a taxonomic subgroup (Field et al., 2011; Castillo et al., 2013). Additionally, the genes of plant BGCs are generally co-expressed in particular organs or under specific conditions (Nützmann et al., 2018). For example, in *Arabidopsis*, the better characterized BGCs are the ones involved in the synthesis of triterpenes (Nützmann et al., 2018). These pathways are specifically expressed in roots where they can produce more than 50 metabolites. Recently, the important role of these SMs in the establishment of root microbiota has been clearly demonstrated (Huang et al., 2019). Intriguingly, as BCGs, lncRNA are also mainly expressed in a tissue specific manner and rarely conserved between and within the same plant species. For example, the comparative transcriptomic *Arabidopsis* ecotypes study from the chapter one of my thesis highlighted that the non-coding transcriptome shows a more ecotype-specific expression than the coding transcriptome, hinting at the importance of the non-coding genomes for the plant local adaptation (Blein et al., 2020). Interestingly, a *LRK* (*leucine-rich repeat receptor kinase*) genes cluster is important for grain yield in rice and the coregulation of the genes cluster is mediated by an antisense lncRNA transcribed from *LRK1* genomic region, supporting a link between non-coding transcriptional units and the transcriptional activity of gene clusters (Wang et al., 2018).

Table 1. List of the putative cis-acting-Col-0-enriched lincRNA and their putative target.

ID_lincRNA	ID	pearson_corr
AT5G38005	AT5G38010	0.950200034
AT5G38005	AT5G38000	0.88107496
AT5G24105	AT5G24100	0.820041589
AT1G06135	AT1G06130	0.813064507
AT1G32172	AT1G32170	0.781090041
AT3G61198	AT3G61190	0.754414744
AT1G18745	AT1G18750	0.745933303
AT3G26612	AT3G26610	0.739124547
AT5G43725	AT5G43740	0.735286136
AT5G00580	AT5G42590	0.709489331
AT2G18735	AT2G18740	0.70369652
AT2G42485	AT2G42490	0.684668185
AT5G40275	AT5G40280	0.666250635
AT1G66173	AT1G66170	0.64411278
AT5G03285	AT5G03290	0.642228588
AT3G12965	AT3G12960	0.624938217
AT3G60972	AT3G60970	0.617147251
AT2G23040	AT2G23050	0.60973819
AT2G15128	AT2G15130	0.593576616
AT3G04485	AT3G04490	0.585352432
AT5G24735	AT5G24710	0.583849213
AT1G61226	AT1G61230	0.580564818
AT1G31935	AT1G31940	0.579607346
AT1G16635	AT1G16630	0.565890198
AT5G15022	AT5G15020	0.556896227
AT2G42485	AT2G42480	0.554159457
AT2G05995	AT2G06000	0.548526096
AT3G22886	AT3G22890	0.544771084
AT4G02005	AT4G02010	0.528030352
AT3G14185	AT3G14190	0.524807596
AT3G60238	AT3G60240	0.523816256
AT1G52855	AT1G52857	0.517909401
AT3G25795	AT3G25790	0.506450705
AT1G19968	AT1G19970	0.500899032

In this chapter, I will investigate the function of a lncRNA embedded in the marneral genes cluster. After briefly introducing how this cluster-related lncRNA was identified, I will present the preprint emerging from this study where I contributed by designing and performing the large majority of the experiments and participating in the writing process. This preprint will be followed by additional results and perspectives to thoroughly characterize the lncRNA function.

6.2.2 Identification of the *MARS* lncRNA

Using the transcriptomics data from Blein et al (2020), I wanted to identify lincRNAs that regulate their neighboring genes in *cis*. In the screen presented in the 2nd chapter, I used the Blein et al. (2020) lincRNA-siRNA precursor annotation, together with the genomic positioning and public gene expression database (Genevestigator (Hruz et al., 2008)) to find putative RdDM-cis-acting lincRNAs implicated in the regulation of root growth and/or development. Differently, I decided this time to explore the transcriptomic analysis of both Col-0 and L*er*, together with the genomic positioning to find putative lncRNAs that may be enriched in a specific ecotype and have potential action as cis-acting lincRNAs. Between Col-0 and L*er* ecotypes, the variability of the transcriptional activity within the non-coding genome may contribute to the modification of expression observed for the coding genome. Thus, after filtering the genes significantly enriched in Col-0 ecotypes compared to L*er*, I isolated the Col-enriched lincRNA their neighboring Col-enriched coding. To strengthen the functional relationship between the lincRNA and the neighboring coding gene, I performed a Pearson correlation analysis among the 128 lincRNA-coding gene pairs using the normalized counts from the Pi starvation kinetics from Blein et al (2020) (**Figure 1**).

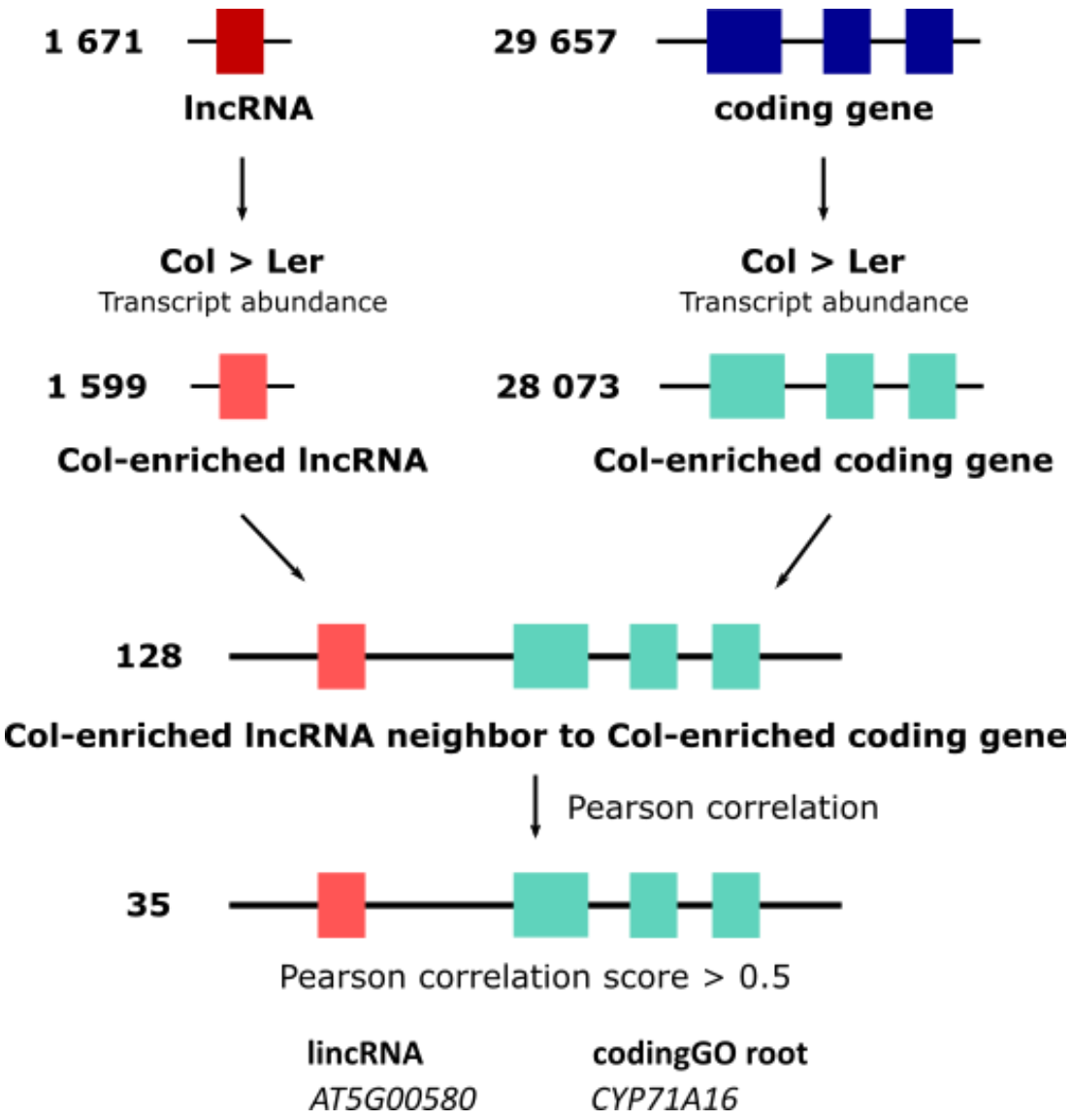


Figure 1. Identification of the *MARS* lincRNA. Flowchart of identification of the *MARS* lincRNA. LincRNAs and coding genes were collected using the Blein et al (2020) transcriptomic analysis and those significantly enriched in Col-0 ecotype as compared to L*er* were retrieved. Pearson correlation scores were calculated for a lncRNA and its upstream and downstream neighboring genes.

Focusing on the lincRNA-coding gene pairs presenting a Pearson correlation score superior to 0.5, I isolated 35 putative cis-acting lincRNA candidates (see **Table1**). From them, I isolated the *AT5G00580* lincRNA localized in-between the genes *CYP71A16 (AT5G42590)* and *MARNERAL SYNTHASE 1* (*MRN1*, *AT5G42600)*, both belonging to the marneral cluster together with *CYP705A12* (*AT4G42580*) (**Figure 1**). Indeed, taking into account that the repression or activation of this cluster of genes seems to be in part governed by the repressive histone mark H3K27me3 (Nützmann et al., 2018) and that lncRNA can modulate the deposition of this mark (Chen et al., 2021), we decided to focus on this intriguing candidate and wondered whether a link between the epigenetic landscape of the marneral cluster and the lncRNA that is transcribed from it could exist. Therefore, I investigated a potential function of this lincRNA in the epigenetic regulation of gene expression of the marneral cluster.

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

Under favorable revision in *Molecular Plant*, our characterization of the MARS lncRNA for its implication in marneral cluster genes regulation is available in BioRxiv as a preprint (10.1101/2020.08.10.236562).

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], Nosheen Hussain[4], Moussa Benhamed[1,2], Jose Gutierrez-Marcos[4], 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

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

*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, 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 of four triterpenes, 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 *CYP71A16*, 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.

MATERIAL AND 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 **Supplementary Table 1**), previously sub-cloned into the pENTR/D-TOPO vector. *UBQ:MARS* were obtained using the GreenGate Cloning system (Lampropoulos et al., 2013) including the *UBQ10* promoter (pGGA006). Arabidopsis plants were transformed using *Agrobacterium tumefaciens* AgI-0 (Clough and Bent, 1998b). The T-DNA inserted line *SALK_133089* was ordered to NASC (N633089). Homozygous mutants were identified by PCR (see primers in **Supplementary Table 1**).

Seeds of *rnm1* (Go et al., 2012), *35S:MRN1* (Field et al., 2011), and *mro1-2* (*cyp71a16*, (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) 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 500µM Pi 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 treatment with water, exogenous ABA or auxin, seedlings were sprayed with 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 lightning 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 phenotype characterization, seedlings were sown in control media and transferred at day 6 in control medium or 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. Finally, for seed germination assay, 0.5µM ABA was supplemented or not to the medium. Germination rate was evaluated twice a day. Seeds were considered germinated when the seed coat was 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 (Sigma-Aldrich) and treated with DNase (Fermentas) as indicated by the manufacturers. Reverse transcription was performed using 1µg total RNA and 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 **Supplementary Table 1**. Data were analyzed using the ΔΔCt method using *PROTEIN PHOSPHATASE 2A SUBUNIT A3* (*AT1G13320*) for gene normalization (Czechowski et al., 2005) and time 0 for time-course experiments.

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; 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 Protein A Dynabeads (Invitrogen) were incubated 12 hours at 4 °C with the samples. Immunoprecipitated DNA was recovered using

Phenol:Chloroform:Isoamilic Acid (25:24:1, Sigma) followed by ethanol precipitation and quantified by qPCR. For input samples, non-immunoprecipitated sonicated chromatin was processed in parallel.

In-vitro transcribed *MARS and GFP* RNA were obtained from a PCR product amplified from wild-type genomic cDNA and pB7FWG2 plasmid harbouring the *GFP* gene, respectively, using the T7 promoter included in the forward primer used for amplification (**Supplementary Table 1**). PCR products were purified by agarose electrophoresis and 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 resuspended in 1 mL of nuclei lysis buffer and split into five tubes. An increasing amount 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 described above.

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)

FAIRE was performed as described by (Simon et al., 2012). After chromatin purification following the ChIP protocol, only 50 µl from the 500 µl of purified chromatin were used (diluted to 500 µl in 10 mM Tris-HCl pH 8). Quantification was performed by qPCR using the same set of primers as for ChIP.

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 600uL 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.

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.

RNA immunoprecipitation (RIP)

For RIP, the *lhp1* mutants complemented with the *ProLHP1:LHP1:GFP* (Nakahigashi et al., 2005) were treated for 4h with ABA. After crosslinking and chromatin extraction as for ChIP, ten percent of resuspended chromatin was kept 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 (Ariel et al., 2014). The enrichment was determined as the percentage of cDNA detected after IP taking the input value as 100%.

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 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 determined by qPCR using a DNA region uncut by Hind III to normalize the amount of DNA across samples.

Transcriptional activation assay in tobacco leaves

The *GUS* reporter system for validating the activity of the putative enhancer element was adapted from (Yan et al., 2019). Different DNA fragments were cloned in the GreenGate system (Lampropoulos et al., 2013) fused to a minimal 35S promoter element from CAMV (synthesized by Eurofins Genomics). The sub-unit B3 from 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 **Supplementary Table 1**.

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 (Jefferson et al., 1987). In addition, a single leaf was also stained with the different constructs to compare their relative activity. 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 (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 on Arabidopsis (Kautsar et al., 2017). 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 analyses

To compute the correlation of expression in different organ of Arabidopsis 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 removing the adaptors with Trim Galore with default parameters, the 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 correlation of expression 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 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 (Wei et al., 2017).

Inside each co-expressed and metabolic clusters of genes, 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 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. 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).

RESULTS

The marneral gene cluster contains 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 (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 *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**). 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 (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 (**Supplementary 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**). 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 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, i.e. constitutively expressing *AT5G00580* under the control of the *UBI10* promoter or knocking down by RNAi. Over-expression lines did not affect the accumulation of protein-coding transcripts in the marneral gene cluster, regardless if the plants were treated or not with ABA (**Figure S2**). On the other hand, *AT5G00580* knock-down by RNAi displayed only in a slight basal induction of *MRN1* (**Figure 2A**). 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 *S3A***) 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 S3B *and* S3C**). 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 S3A**). The same behavior was observed using a higher concentration of ABA (**Figure S4**). 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 200 bp upstream the transcription start site (TSS) of *MARS* gene that we named *mrs1-1*. We found that *mrs1-1* partially impairs the transcriptional accumulation of *MARS* and protein-coding genes in the marneral cluster, mainly *CYP71A16* (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,

CYP71A16, 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 *CYP71A16 gene* (*mro1-2;* (Field et al., 2011)), which did not influence the expression of the other marneral cluster genes (**Figure S6A,B**) or known ABA-responsive genes (**Figure S6C,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 *S7***), but resulted in a decrease in *MARS* transcripts abundance under control condition (**Figure S7A,B**). 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* down-regulated 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 S8A,B,C,D**). 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 S8A,B**). 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 S8E**). 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 S8E;** 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. Collectively, our results indicate that *MARS* can modulate various ABA-related physiological responses, 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 S9)**. ATAC-Seq data (Sijacic et al., 2018) also revealed that the marneral cluster exhibits a high chromatin condensation in shoots (**Figure S9**). These data suggest that the marneral cluster is in a epigenetically silent state in aerial organs, thus correlating with its low expression level in leaves (Yu et al., 2016).

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 S10**). Interestingly, exogenous ABA triggered a strong reduction of H3K27me3 deposition throughout the marneral cluster (**Figure 3A and S10**). 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 3A, S10 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 *CYP71A16* and *MARS* (**Figure 3B and S11**). Remarkably, LHP1 recognition was strongly impaired in response to ABA as well as in RNAi-*MARS* seedlings (**Figure 3B and S11**). 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 S12A**). In agreement, RNAi-*MARS* plants did not exhibit altered levels of DNA methylation across the *MARS* gene body, indicating that *MARS* silencing occurs at post-transcriptional level, without affecting the epigenetic state of the endogenous locus (**Figure S12B).** Notably, the global chromatin status of the cluster was less condensed in RNAi-*MARS* seedlings in response to ABA (**Figure 4A and S12A**), in agreement with a decrease of both H3K27me3 deposition and LHP1 binding (**Figure 4A,B, S10 and S11**) 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 4B and S13**), 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 S14A**), as observed in the RNAi-*MARS* seedlings (**Figure 2B**) with decondensed chromatin (**Figure 4A and S12A**).

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 *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 nuclear structural ncRNA *U6* (**Figure S14B**), involved in the spliceosome. Then, we confirmed by RNA immunoprecipitation (RIP) that LHP1 can interact with *MARS in vivo,* in contrast to *MRN1* or a randomly selected housekeeping gene (*PP2A*) taken as negative controls (**Figure 4C**).

LHP1 binding to the marneral cluster was impaired both in response to exogenous ABA (concomitantly inducing *MARS*, **Figure 3B, S11 and 2B**) and in RNAi-*MARS* seedlings, hinting at a stoichiometry-dependent action of *MARS* on LHP1 recognition of the marneral cluster. Therefore, we used chromatin extracts from RNAi-*MARS* line 1 seedling, that contains 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 S14C**) successfully promoted LHP1 binding to the cluster, in contrast to higher concentrations (between 1 and 10 µg of RNA). Moreover, the *in vitro*-transcribed *GFP* RNA was not able to promote LHP1 binding, supporting the relevance of the specific *MARS*-LHP1 stoichiometric interaction for LHP1-target recognition (**Figure 4D and S14C**). Altogether, our results suggest that the physical interaction of the nuclear-enriched lncRNA *MARS* to LHP1 modulates its binding to proximal chromatin in a dual manner 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 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 interaction linking the intergenic region between *CYP71A16* and *MARS* and the *MRN1* locus (indicated as "Chromatin loop" in **Figure 5A**). 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 5B**). 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**).

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 (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 5C**). 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 13 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 *CYP71A16* 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**

5A). We thus assessed the capacity of these genomic regions to enhance the transcriptional activity of MRN1. To this end, we generated transcriptional reporter line 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 *CYP71A16* and the other at the 3′ end of *AT5G42620* locus (**Figure 5A**). Among the two putative distal enhancers tested, one was able to activate *GUS* expression (Intergenic region 2, **Figure 5D and Figure S15),** 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 and that 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 (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 of gene clusters for co-expressed neighboring genes (Yu et al., 2016) and one for metabolic gene clusters (PlantiSMASH (Kautsar et al., 2017)). Among the 390 clusters of co-expressed neighboring genes, 189 (48%) contained at least one lncRNA embedded within the cluster. Most importantly, among the 45 metabolic clusters, 28 (62%) include lncRNAs inside the cluster (**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 (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). 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, chromatin loop conformation is related to LHP1 binding and is modulated by *MARS* in a dual manner. 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 in response to ABA directly modulates LHP1 binding to the marneral cluster (**Figure 6C**; in response to ABA). *In vitro*, high level of *MARS* RNA can titrate LHP1 binding even though a minimal amount of *MARS* RNA is required to recruit LHP1 to the marneral locus (**Figure 4D and S14C**). *In vivo*, the overexpression of *MARS* does not affect the transcriptional activity of the marneral cluster, suggesting that additional ABA-related factors are required to trigger the activation of the cluster

upon LHP1 titration (**Figure S2**). Also, when *MARS* levels are too low compared to basal levels (*in vitro* (**Figure 4D and S14C**) or *in vivo* with the RNAi lines (**Figure 3B and S11**)), recruitment of LHP1 to the cluster is also impaired (**Figure 6C;** *MARS* repression). Thus, the concentration-dependent dual effect of *MARS* on LHP1 binding (**Figure 4D and S14C**) appears as a key factor determining the dynamics of the marneral cluster epigenetic landscape.

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. 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.

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.

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.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

ACKNOWLEDGEMENT

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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, 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.

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 the different isoforms of *AT5G00580* transcripts. First line corresponds to the *AT5G00580* genomic region whereas the other lines present the various isoforms. For each isoform, exons are indicated with rectangle 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 of noncoding transcripts.

(D) 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 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* (*SALK_133089)*. Transcriptional abundance is shown as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype. Letters indicate a statistical 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* and *mrs1-1* (*SALK_133089*). Gene expression data are shown as the mean ± standard error (n = 3) of the log2 fold change compared to the Col-0 genotype at time 0h.

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

(A) H3K27me3 deposition over the *MRN1* promoter 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* promoter 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 shown 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 shown 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. *MRN1* 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 percentage of input (signal measured before immunoprecipitation).

(D) LHP1 binding to the *MRN1* promoter region in chromatin from RNAi-*MARS* 1 seedlings upon increasing amounts of *in-vitro* transcribed *MARS* or *GFP* RNAs. After incubation (see Methods), the samples were crosslinked for LHP1 ChIP-qPCR. Higher values indicate more LHP1-DNA interaction. Results are shown 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 approaches the *MRN1* locus with 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 on the marneral cluster (Song et al., 2016). 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 shown 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 shown 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 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 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 (Yu et al., 2016) 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 (Kautsar et al., 2017).

(B) Maximum level of correlation between a lncRNA and any clustered gene determined 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 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 approaches the enhancer element and *MRN1* proximal promoter, leading to a 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.

Figure1

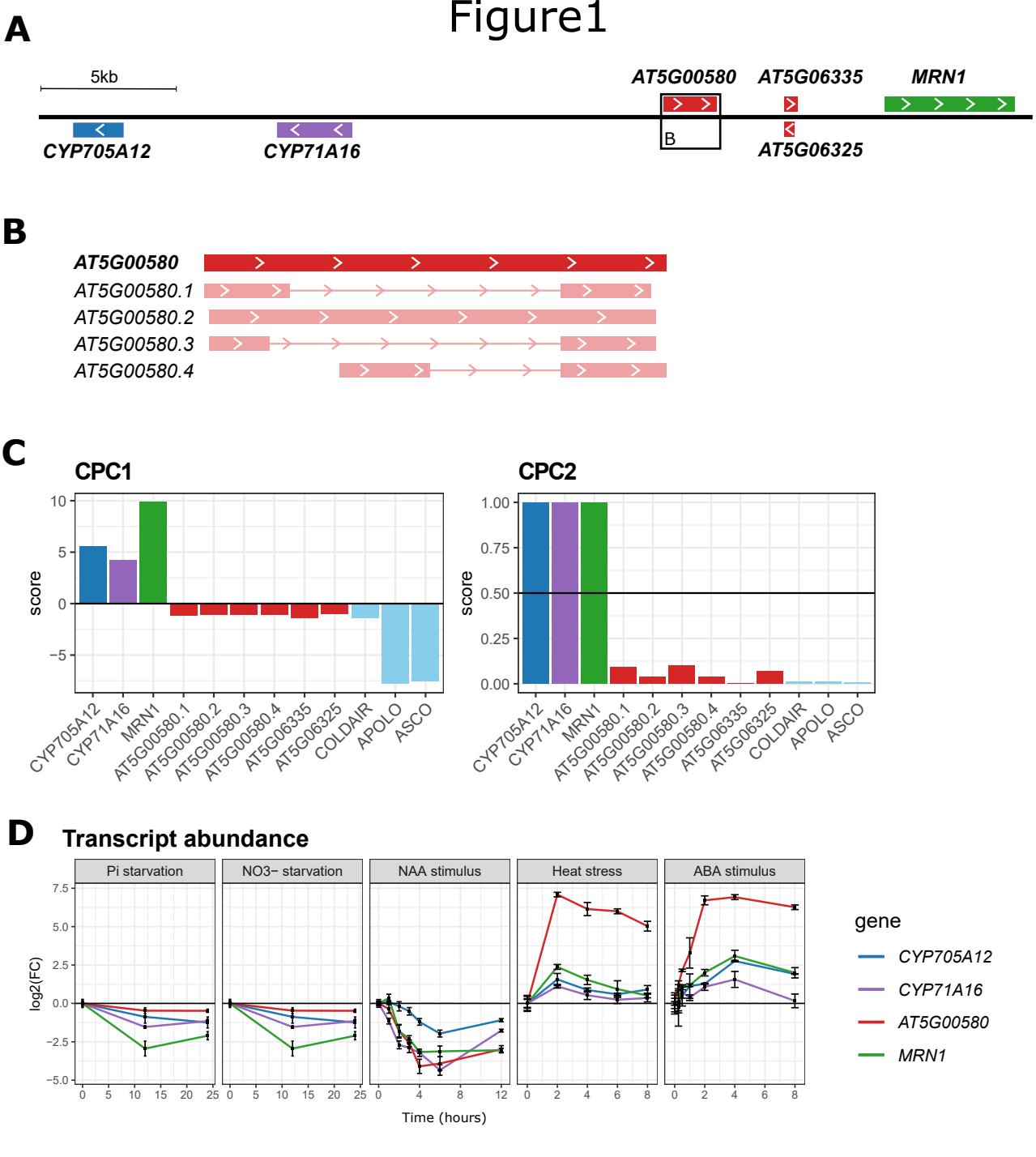
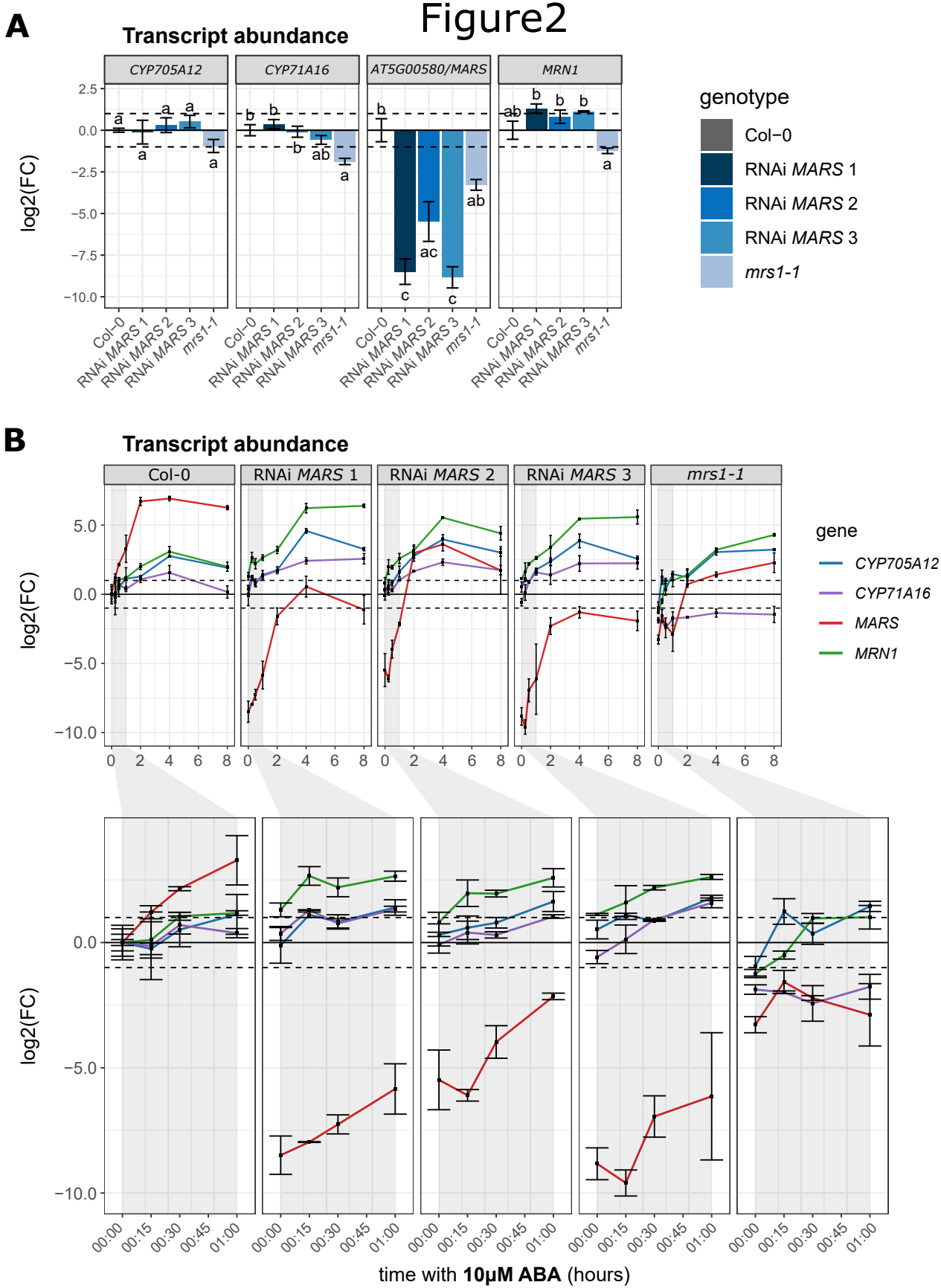
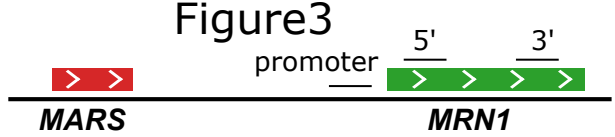
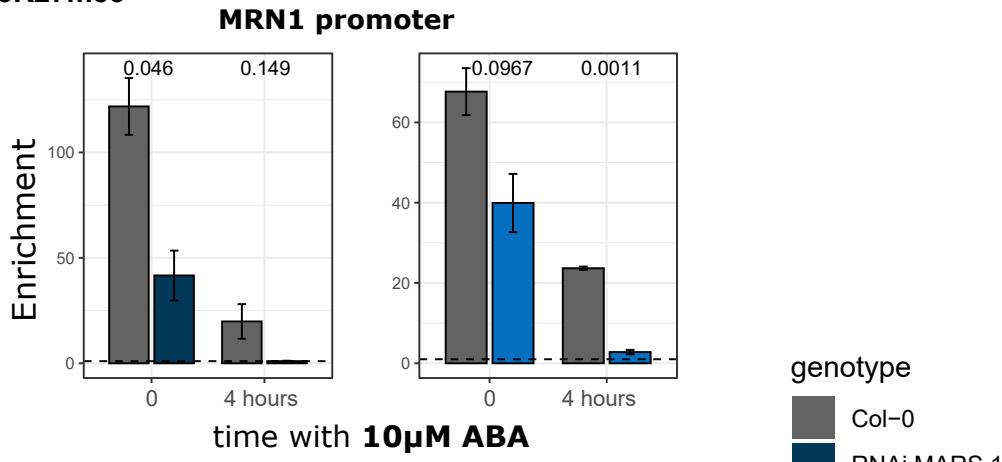

A
5kb
AT5G00580
AT5G06335
MRN1
CYP705A12
CYP71A16
B
AT5G06325
B
AT5G00580
AT5G00580.1
AT5G00580.2
AT5G00580.3
AT5G00580.4
C
CPC1
CPC2
score
10
5
0
−5
1.00
0.75
0.50
0.25
0.00
CYP705A12
CYP71A16
MRN1
AT5G00580.1
AT5G00580.2
AT5G00580.3
AT5G00580.4
AT5G06335
AT5G06325
COLDAIR
APOLO
ASCO
D
Transcript abundance
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
CYP71A16
AT5G00580
MRN1

Figure2


A
Transcript abundance
CYP705A12
CYP71A16
AT5G00580/MARS
MRN1
log2(FC)
2.5
0.0
−2.5
−5.0
−7.5
−10.0
a
a
a
a
b
b
b
ab
a
b
ac
c
c
ab
ab
b
b
b
a
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)
5.0
0.0
−5.0
−10.0
0
2
4
6
8
gene
CYP705A12
CYP71A16
MARS
MRN1
00:00
00:15
00:30
00:45
01:00
time with 10µM ABA (hours)



A **H3K27me3**



B **LHP1**

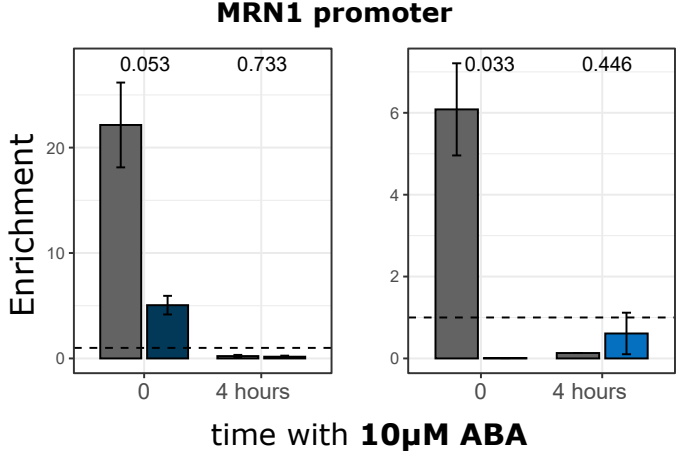
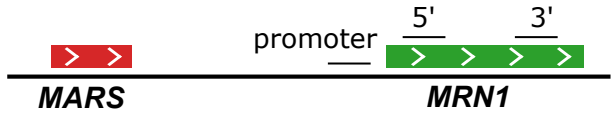
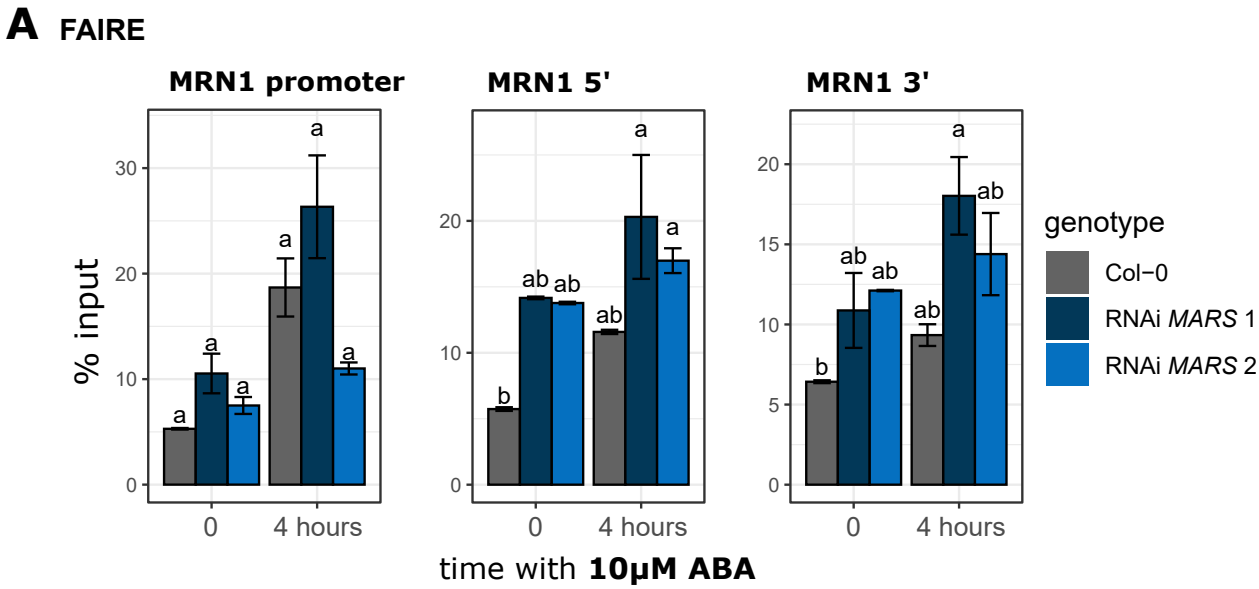
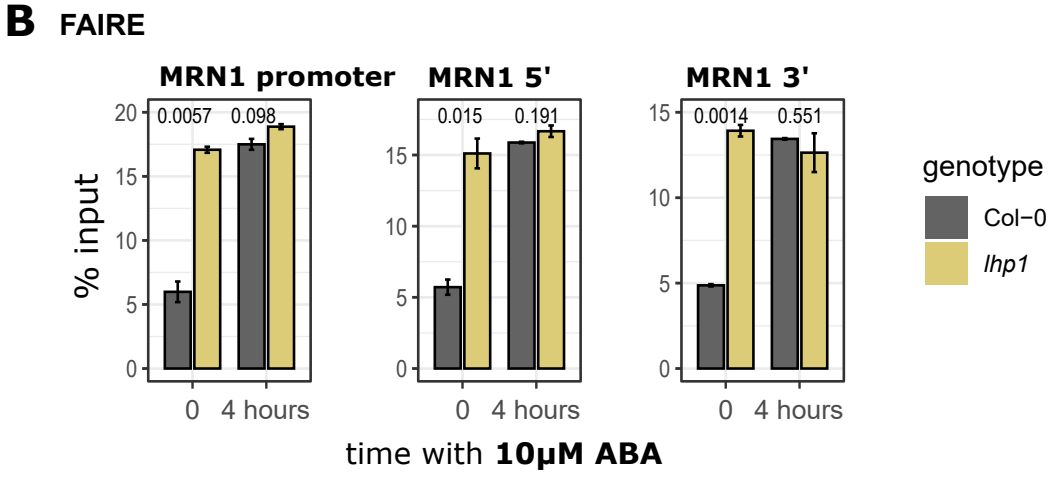
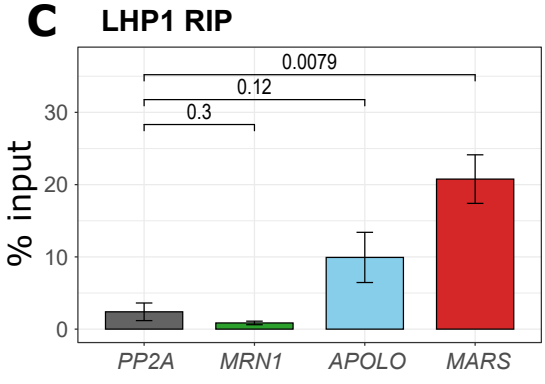


Figure4


promoter
5'
3'
MARS
MRN1


A FAIRE
MRN1 promoter
MRN1 5'
MRN1 3'
% input
time with 10μM ABA
0
4 hours
genotype
Col−0
RNAi MARS 1
RNAi MARS 2


B FAIRE
MRN1 promoter
MRN1 5'
MRN1 3'
0.0057
0.098
0.015
0.191
0.0014
0.551
% input
time with 10μM ABA
genotype
Col−0
lhp1


C LHP1 RIP
0.0079
0.12
0.3
% input
PP2A
MRN1
APOLO
MARS

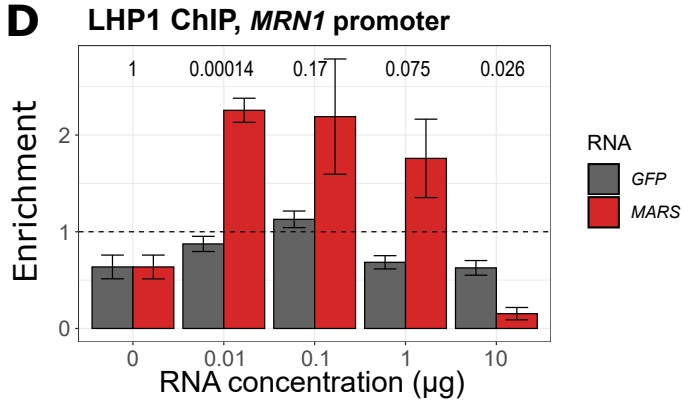
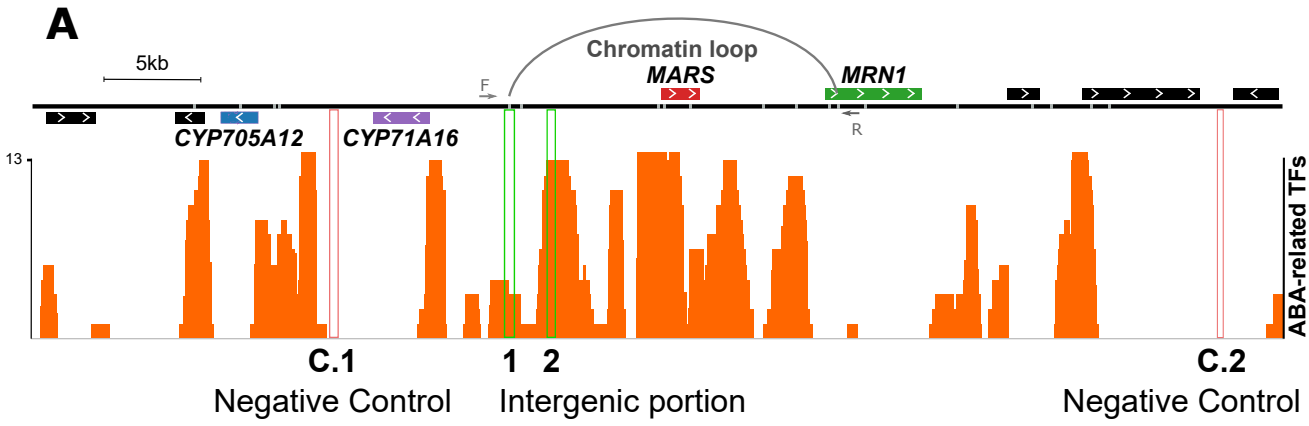
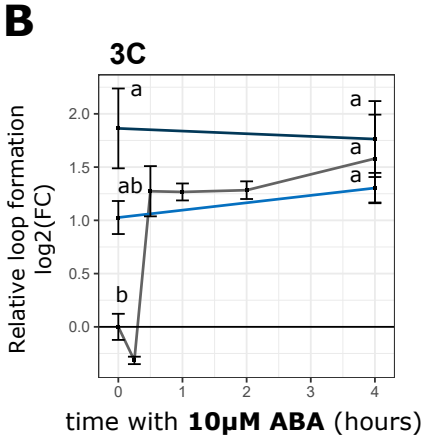
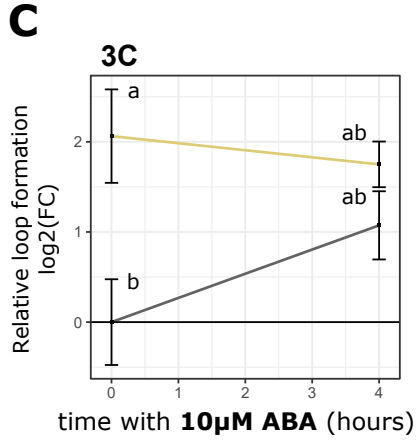
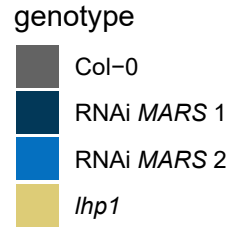

D LHP1 ChIP, MRN1 promoter
1
0.00014
0.17
0.075
0.026
Enrichment
RNA concentration (μg)
RNA
GFP
MARS

Figure5


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

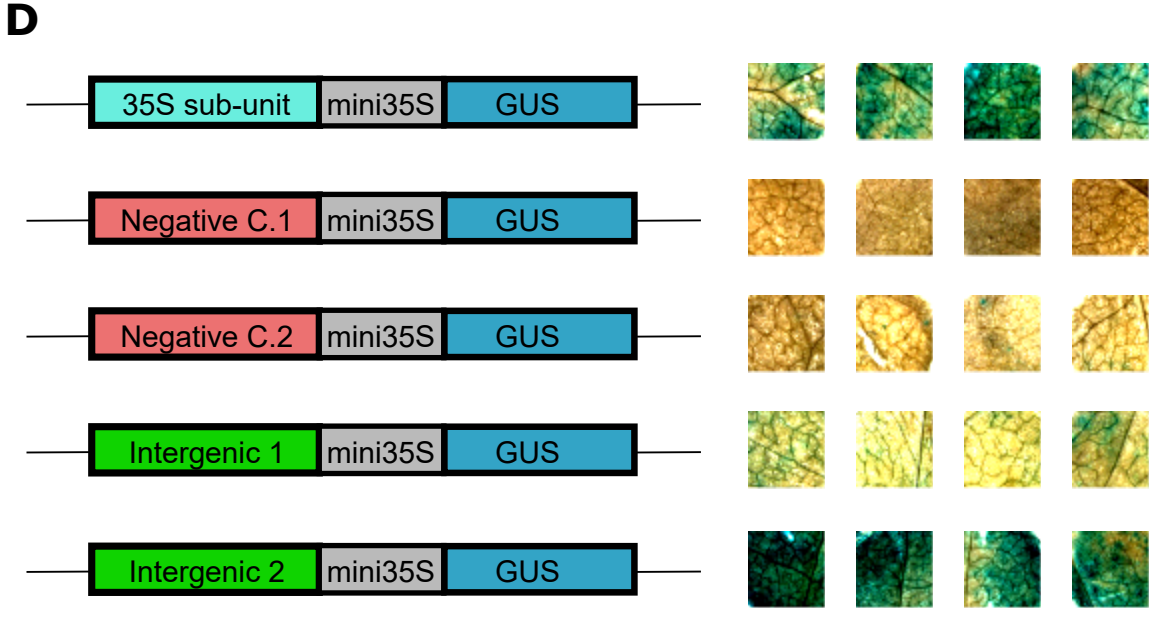
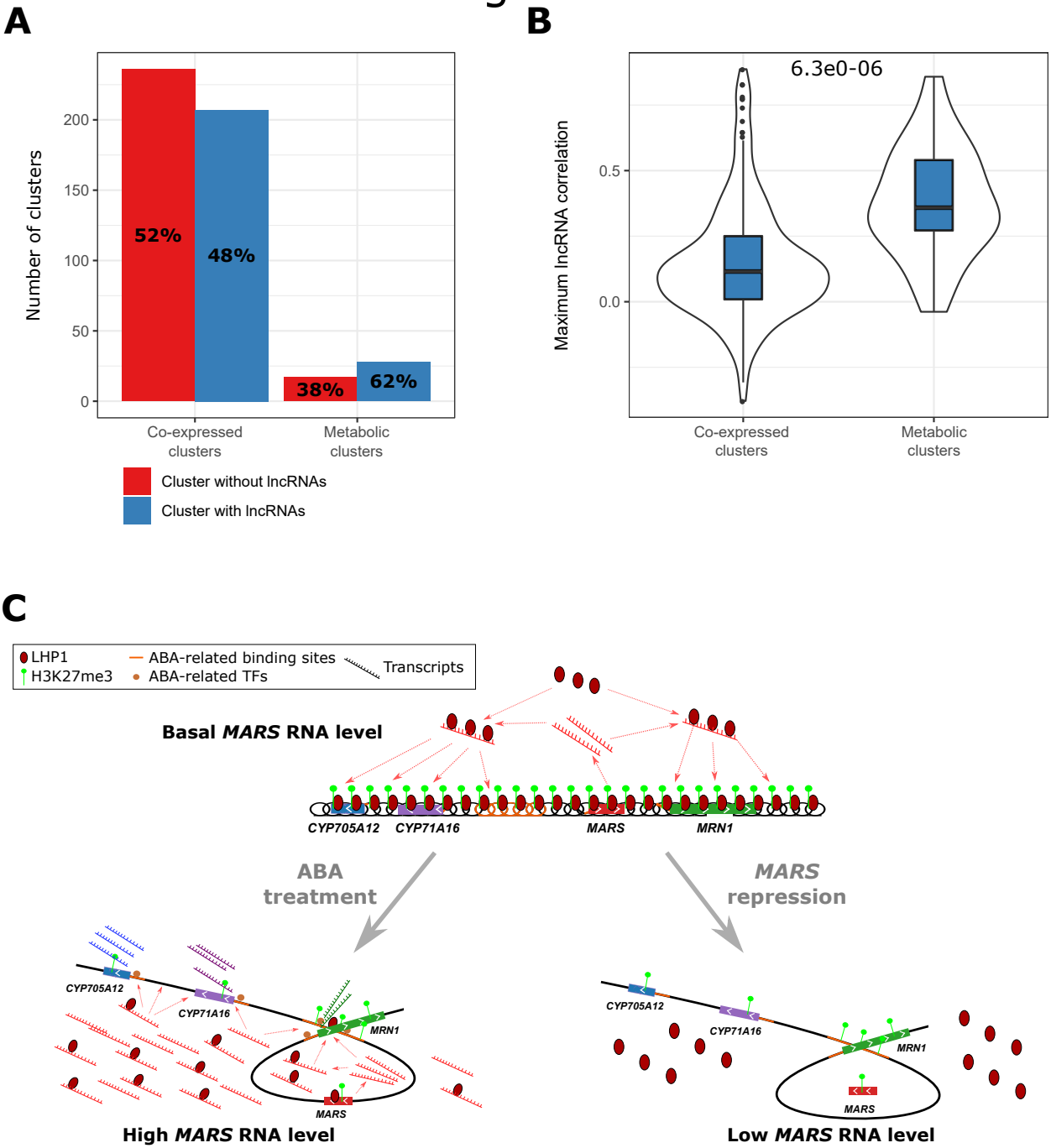
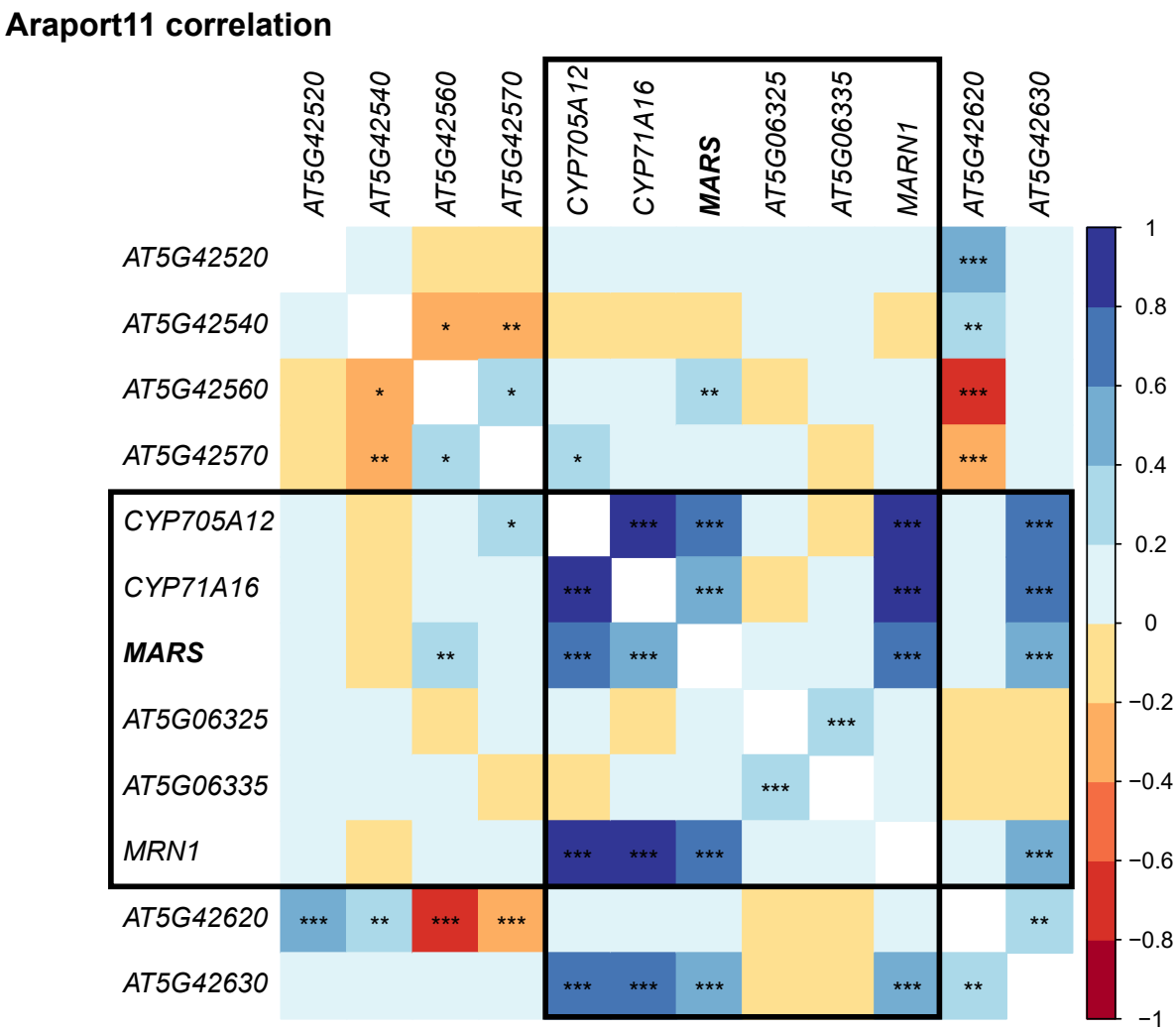

D
35S sub-unit
mini35S
GUS
Negative C.1
Negative C.2
Intergenic 1
Intergenic 2

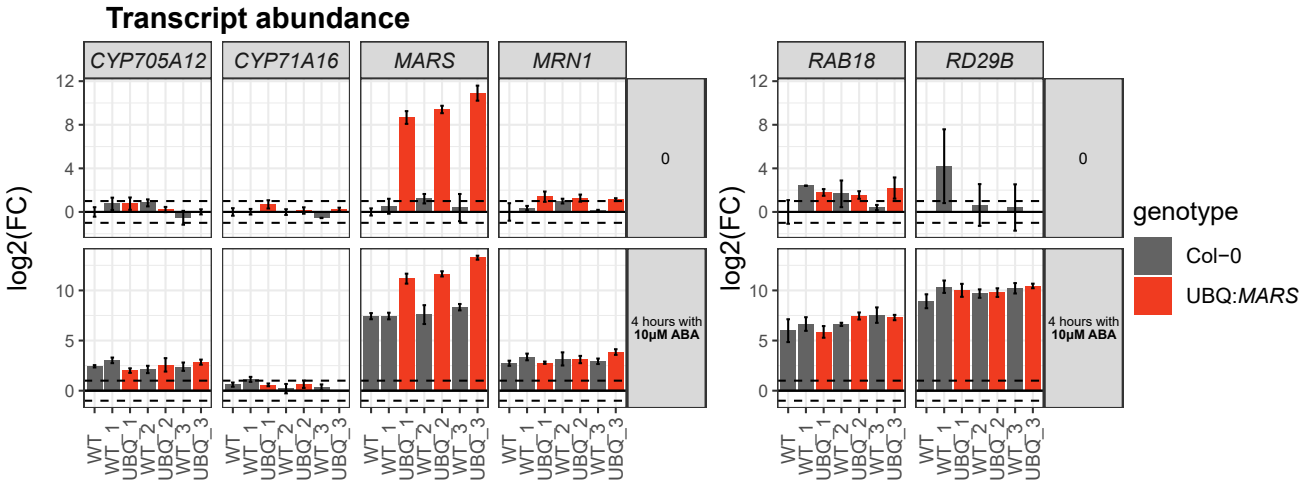
Figure6


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
Maximum lncRNA correlation
6.3e0-06
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

FigureS1

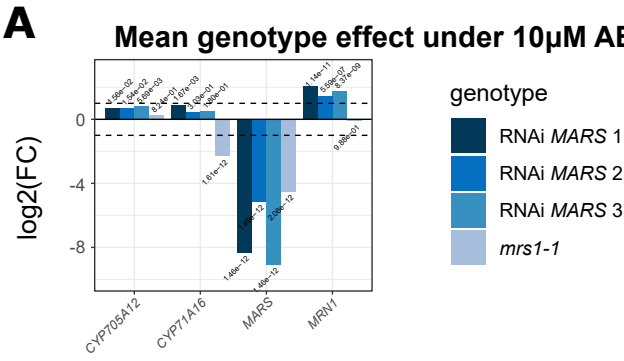

Araport11 correlation
AT5G42520
AT5G42540
AT5G42560
AT5G42570
CYP705A12
CYP71A16
MARS
AT5G06325
AT5G06335
MRN1
AT5G42620
AT5G42630
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

FigureS2

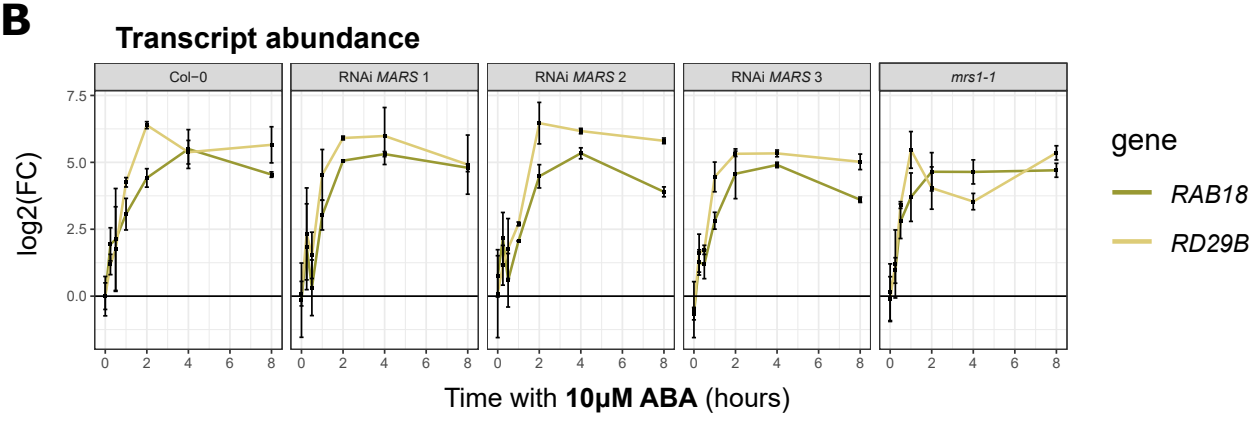

Transcript abundance
CYP705A12
CYP71A16
MARS
MRN1
RAB18
RD29B
log2(FC)
0
4 hours with 10µM ABA
WT_1
WT_1
UBQ_1
WT_2
UBQ_2
WT_3
UBQ_3
genotype
Col−0
UBQ:MARS

FigureS3

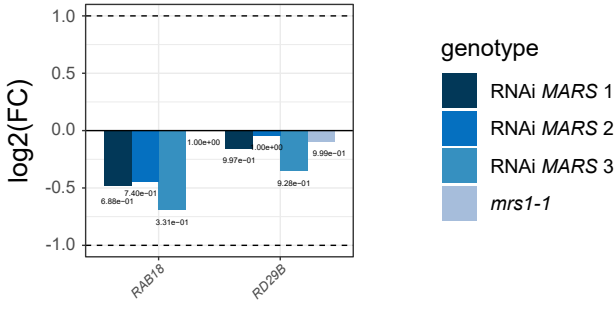
A



B

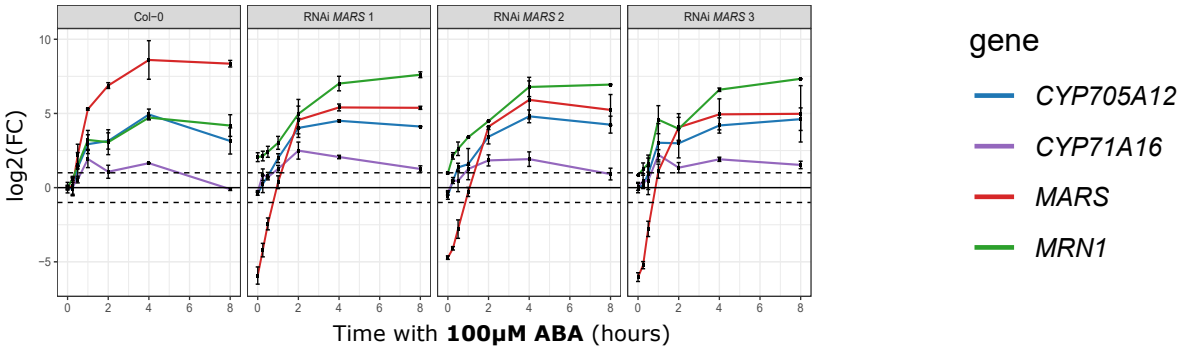


C

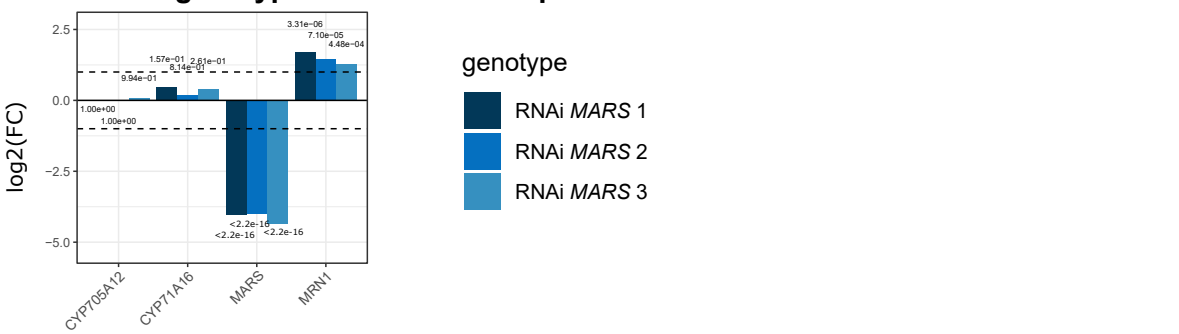


FigureS4

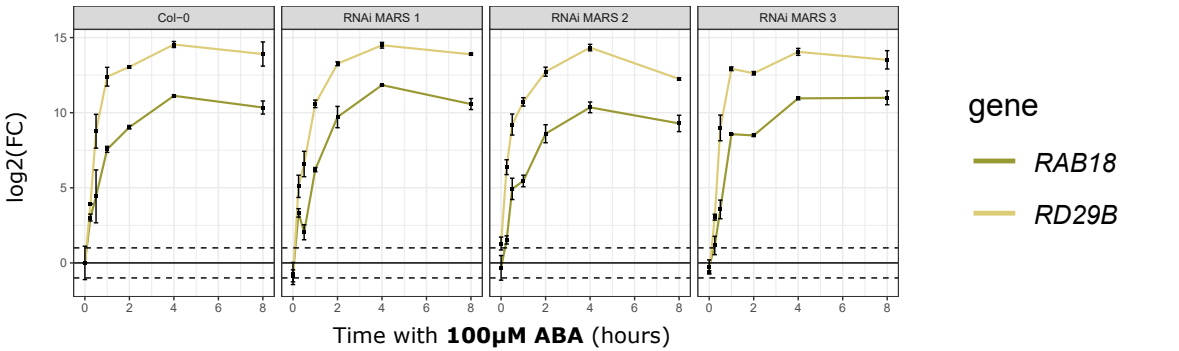
A



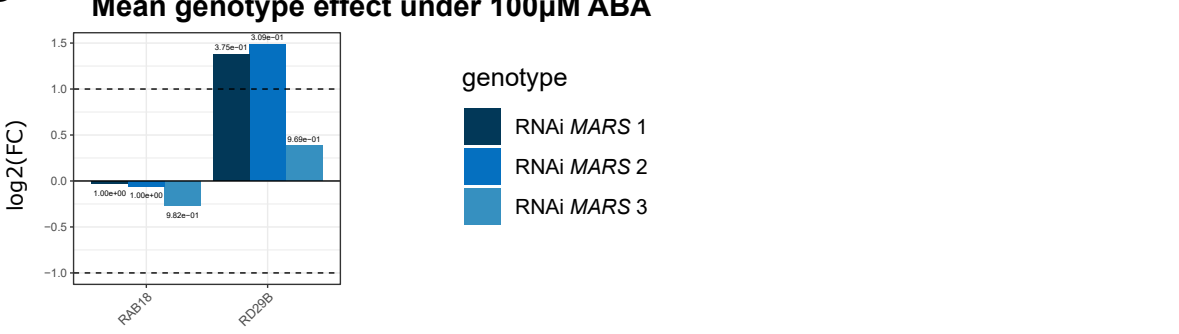
B



C



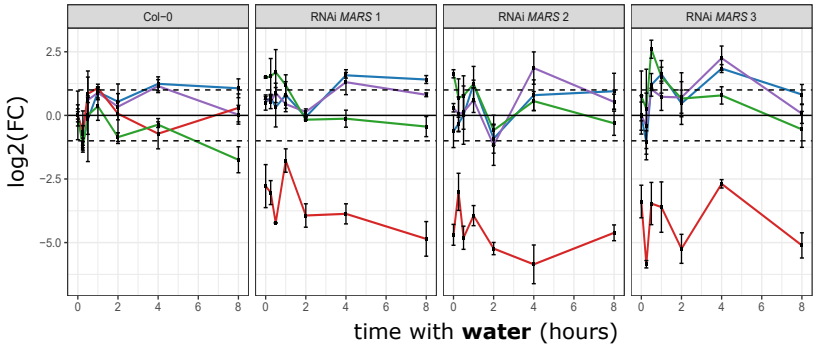
D

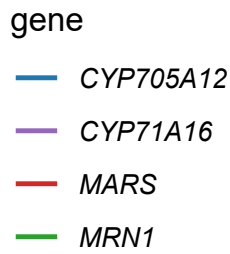


FigureS5

A

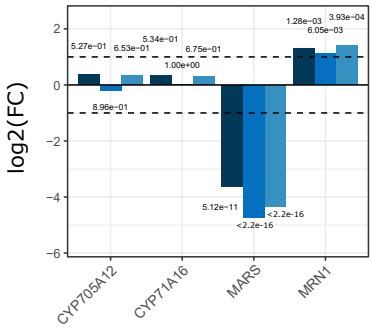
Transcript abundance

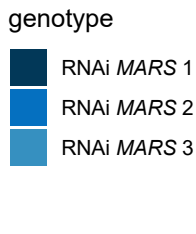




B

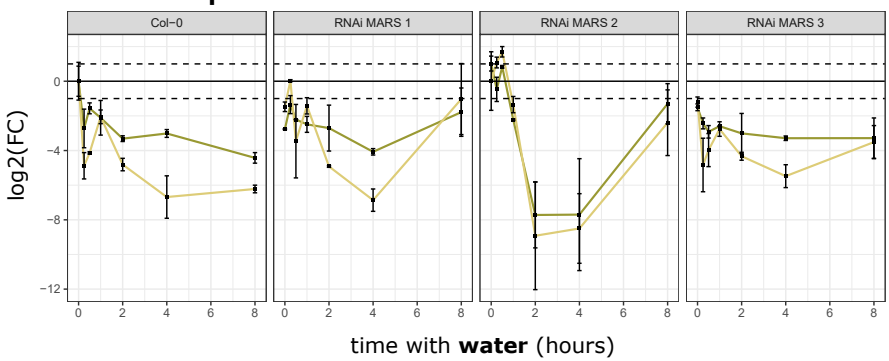
Mean genotype effect under water





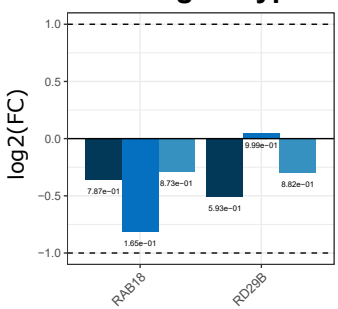
C

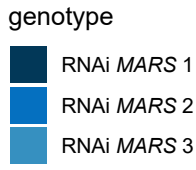
Transcript abundance



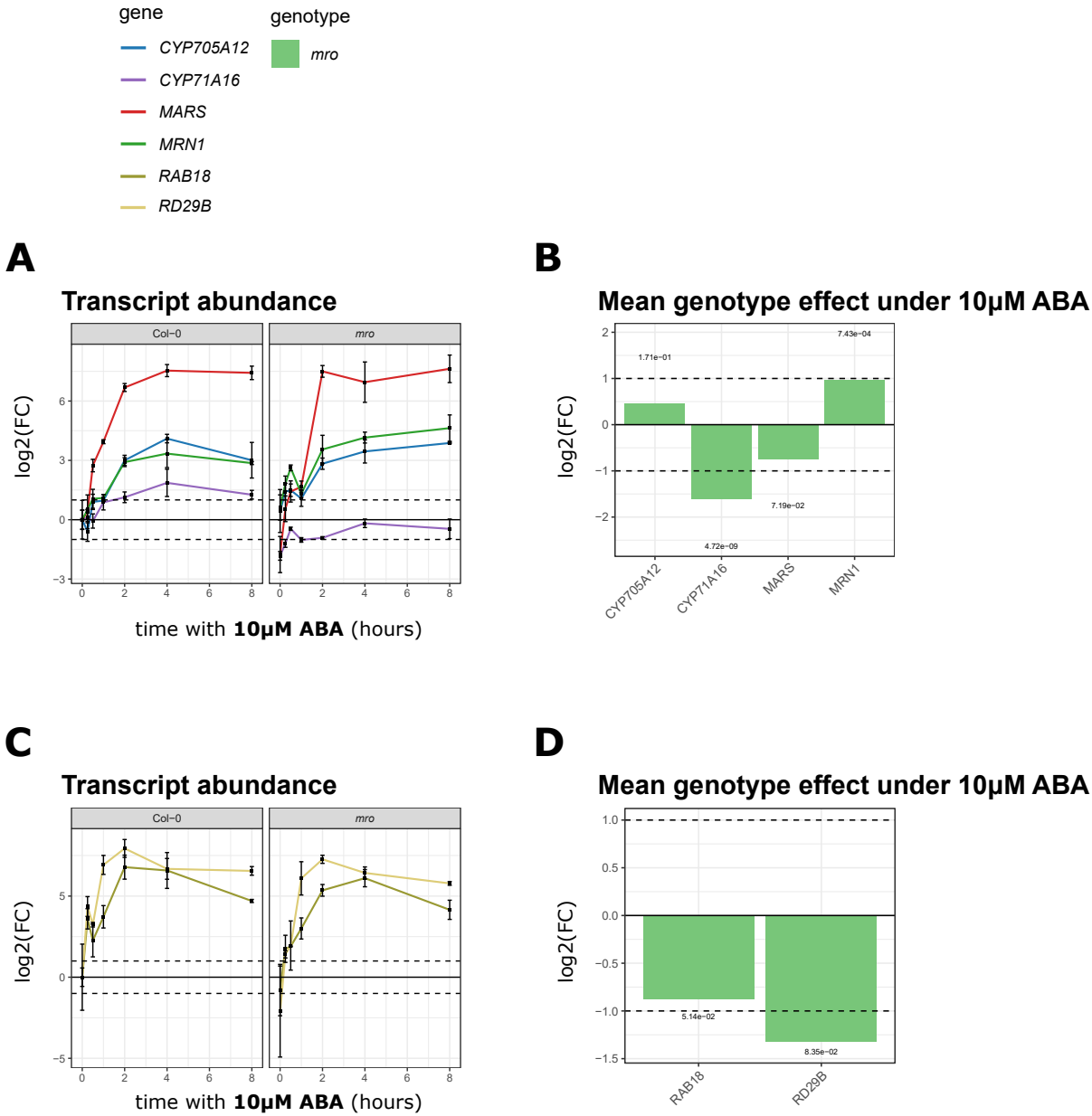
D

Mean genotype effect under water

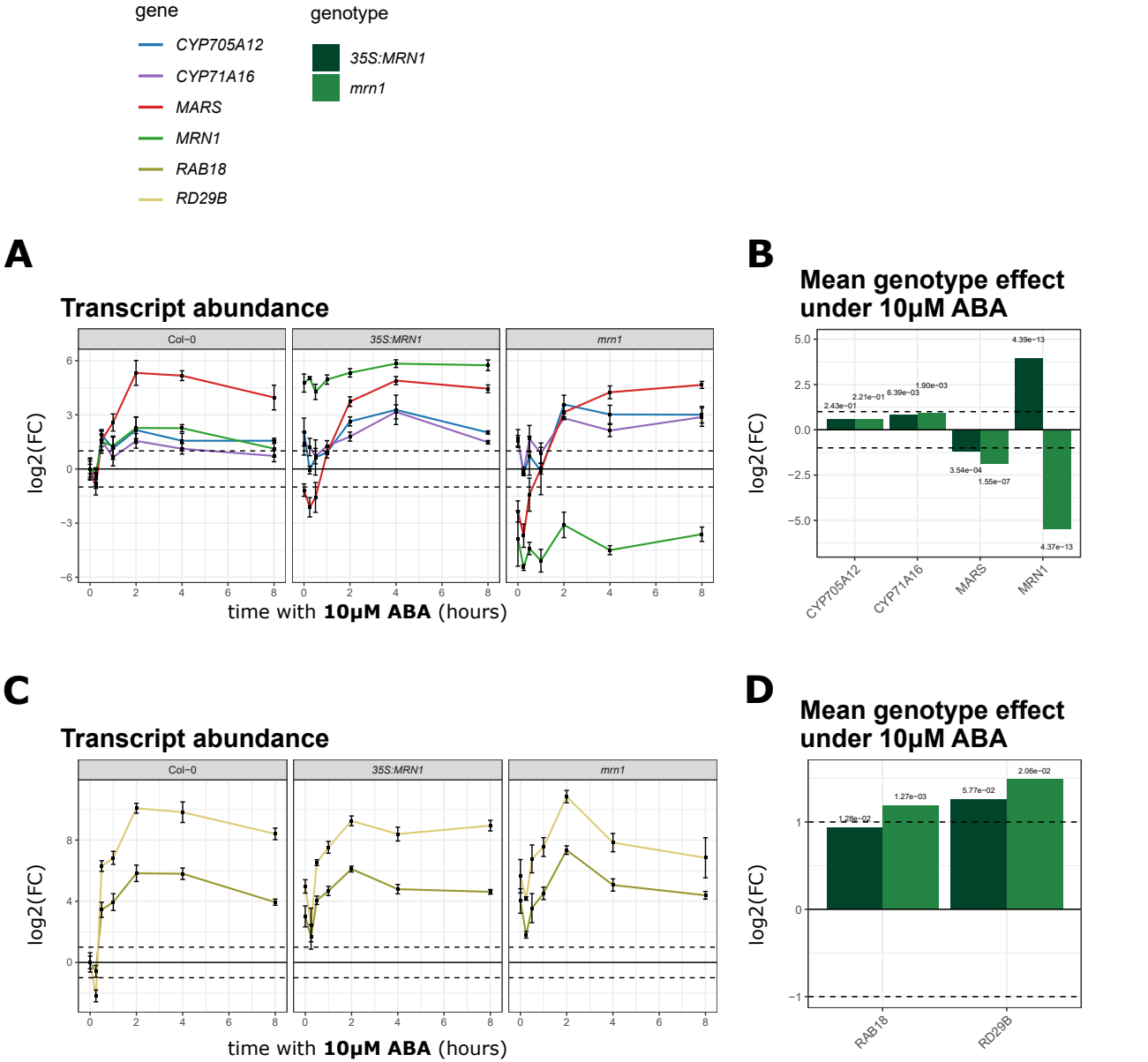




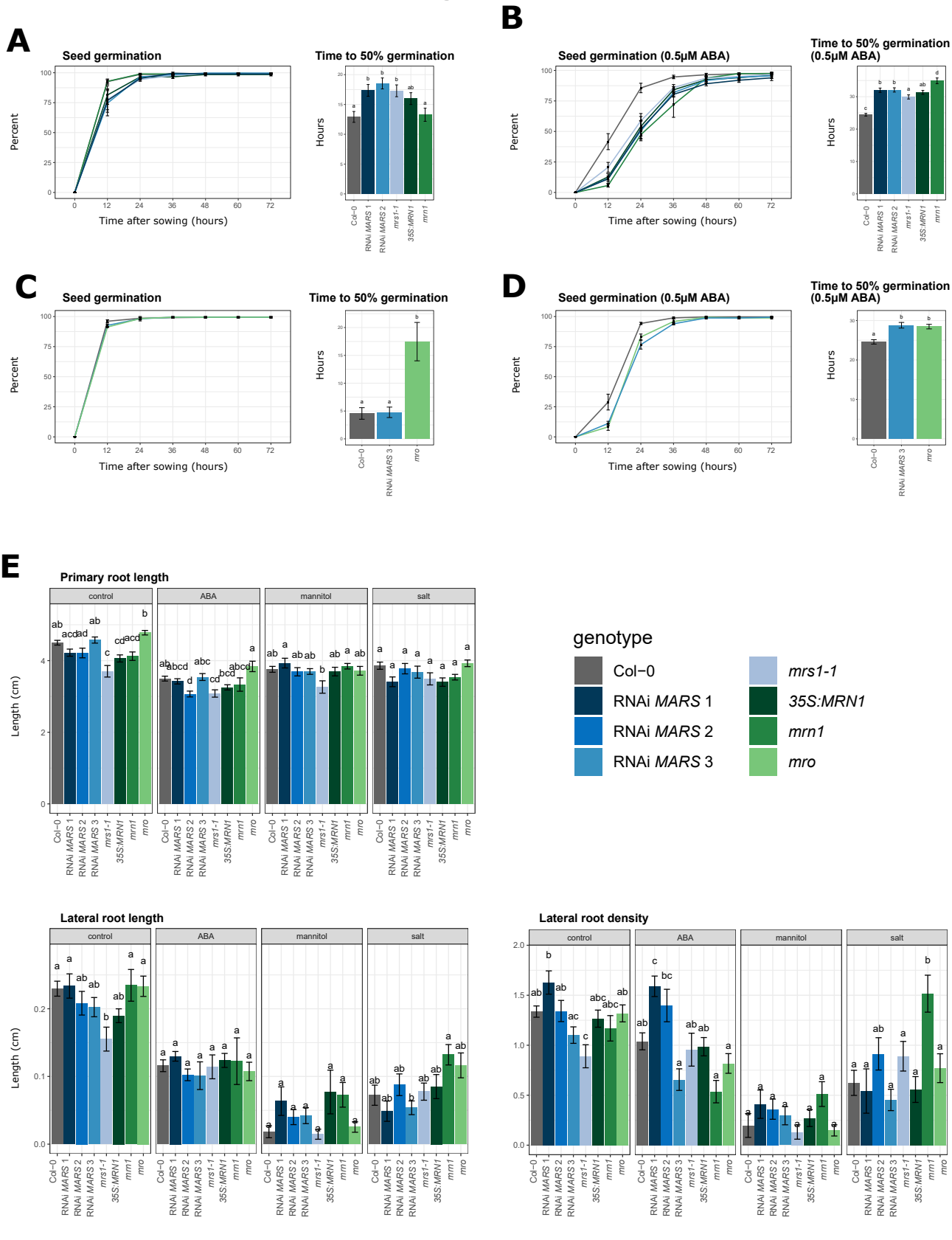
FigureS6


gene
CYP705A12
CYP71A16
MARS
MRN1
RAB18
RD29B
genotype
mro
A
Transcript abundance
Col−0
mro
log2(FC)
time with 10μM ABA (hours)
B
Mean genotype effect under 10μM ABA
1.71e−01
7.43e−04
7.19e−02
4.72e−09
CYP705A12
CYP71A16
MARS
MRN1
C
Transcript abundance
Col−0
mro
log2(FC)
time with 10μM ABA (hours)
D
Mean genotype effect under 10μM ABA
5.14e−02
8.35e−02
RAB18
RD29B

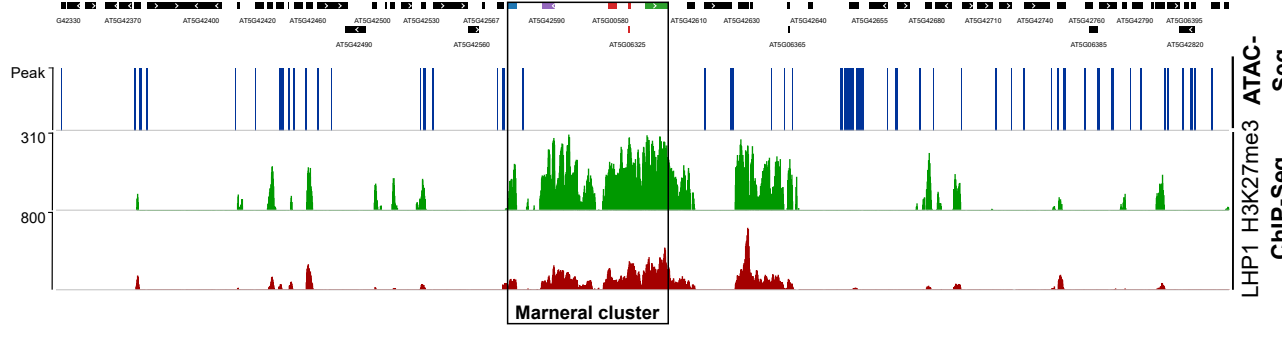
FigureS7


gene
CYP705A12
CYP71A16
MARS
MRN1
RAB18
RD29B
genotype
35S:MRN1
mrn1
A
Transcript abundance
Col−0
35S:MRN1
mrn1
log2(FC)
time with 10μM ABA (hours)
B
Mean genotype effect under 10μM ABA
log2(FC)
2.43e−01
2.21e−01
6.39e−03
1.90e−03
3.54e−04
1.55e−07
4.39e−13
4.37e−13
CYP705A12
CYP71A16
MARS
MRN1
C
Transcript abundance
Col−0
35S:MRN1
mrn1
log2(FC)
time with 10μM ABA (hours)
D
Mean genotype effect under 10μM ABA
log2(FC)
1.28e−02
1.27e−03
5.77e−02
2.06e−02
RAB18
RD29B

FigureS8

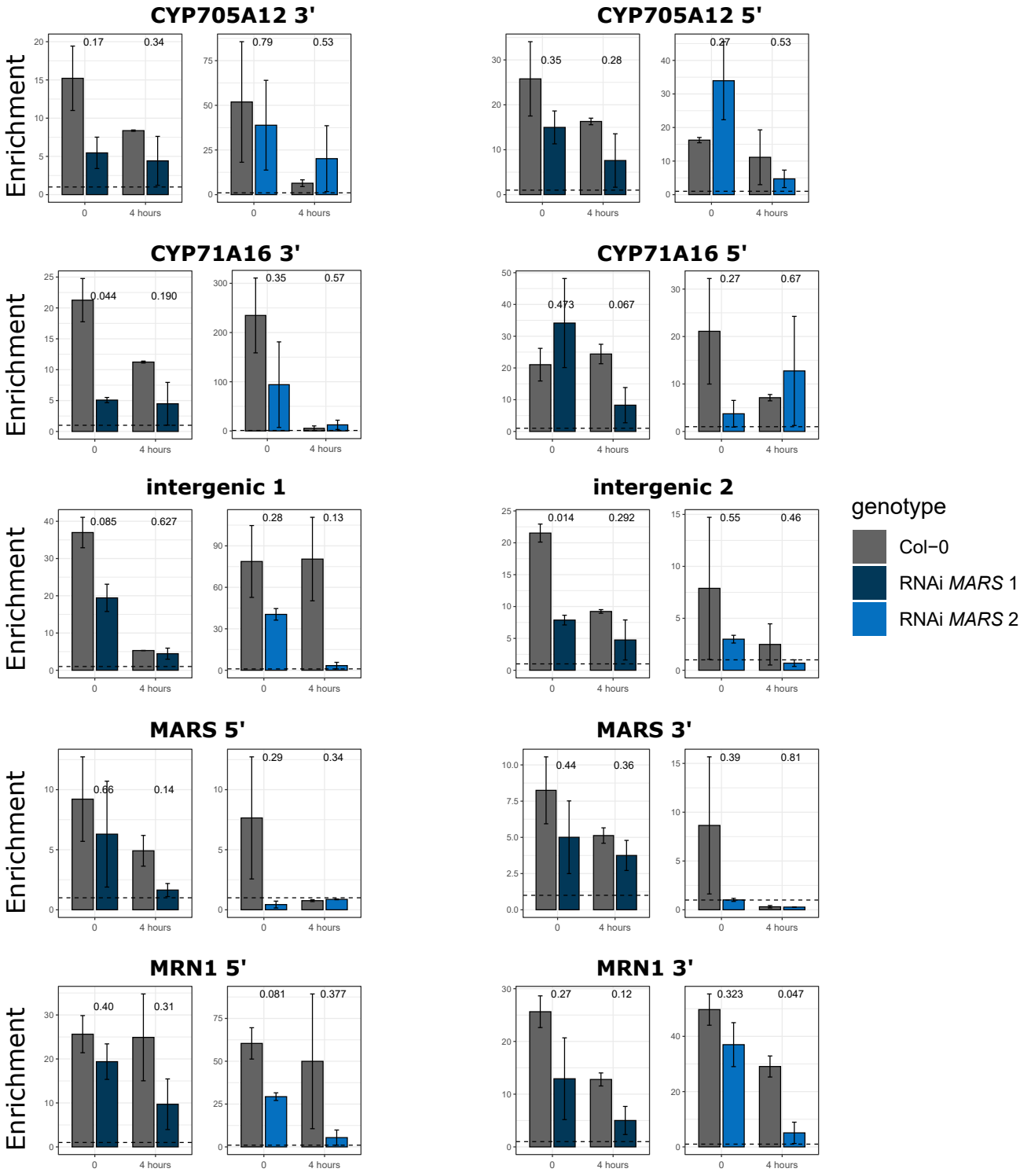

A
Seed germination
Percent
Time after sowing (hours)
Time to 50% germination
Hours
B
Seed germination (0.5μM ABA)
Time to 50% germination (0.5μM ABA)
C
D
E
Primary root length
control
ABA
mannitol
salt
Length (cm)
genotype
Col−0
RNAi MARS 1
RNAi MARS 2
RNAi MARS 3
mrs1-1
35S:MRN1
mrn1
mro
Lateral root length
Lateral root density

FigureS9

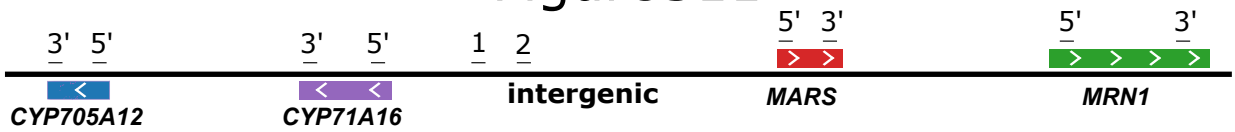

G42330
AT5G42370
AT5G42400
AT5G42420
AT5G42460
AT5G42490
AT5G42500
AT5G42530
AT5G42560
AT5G42567
AT5G42590
AT5G00580
AT5G06325
AT5G42610
AT5G42630
AT5G42640
AT5G06365
AT5G42655
AT5G42680
AT5G42710
AT5G42740
AT5G42760
AT5G06385
AT5G42790
AT5G06395
AT5G42820
Peak
310
800
ATAC-Seq
LHP1 H3K27me3 ChIP-Seq
Marneral cluster

FigureS10

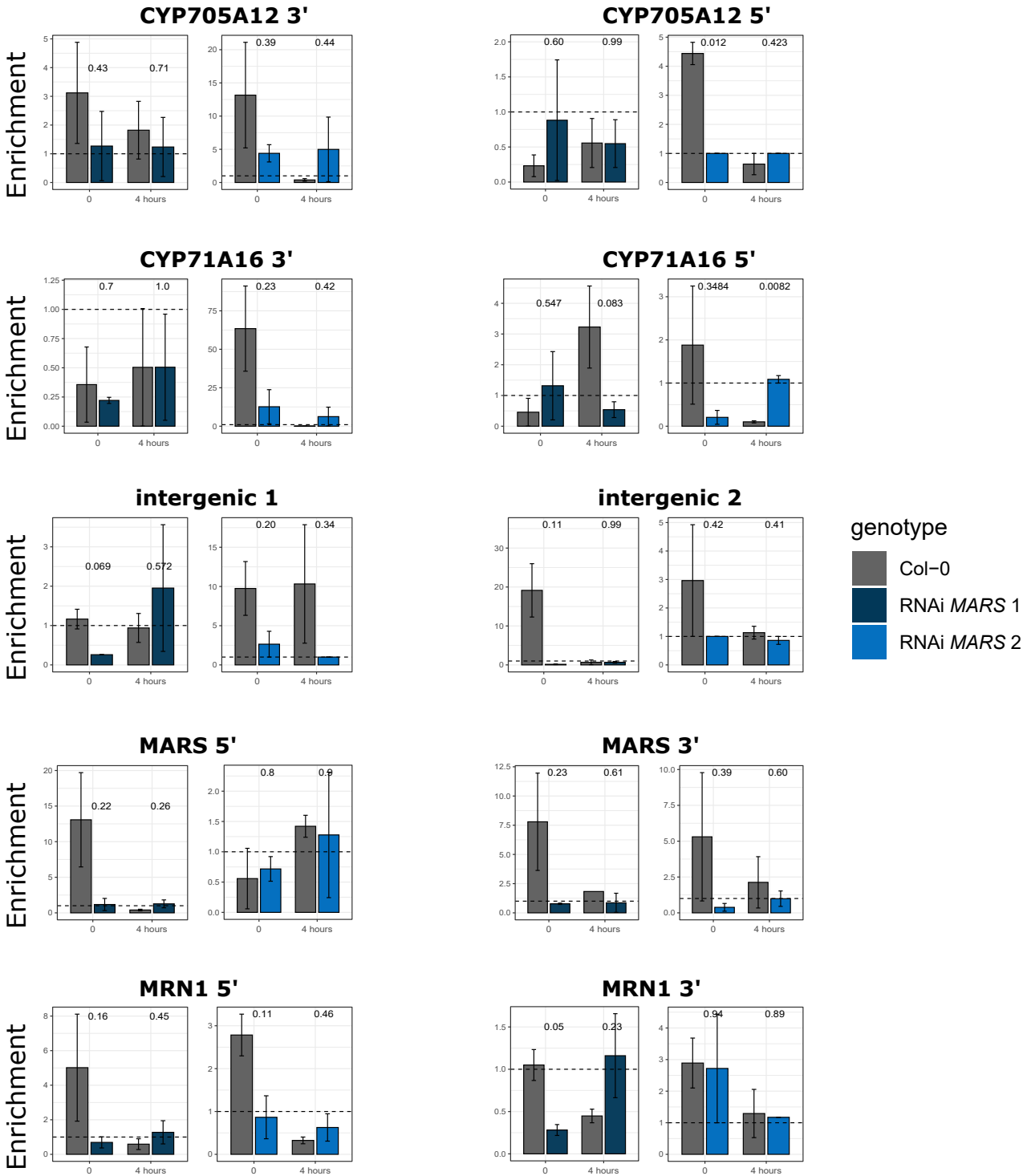
H3K27me3



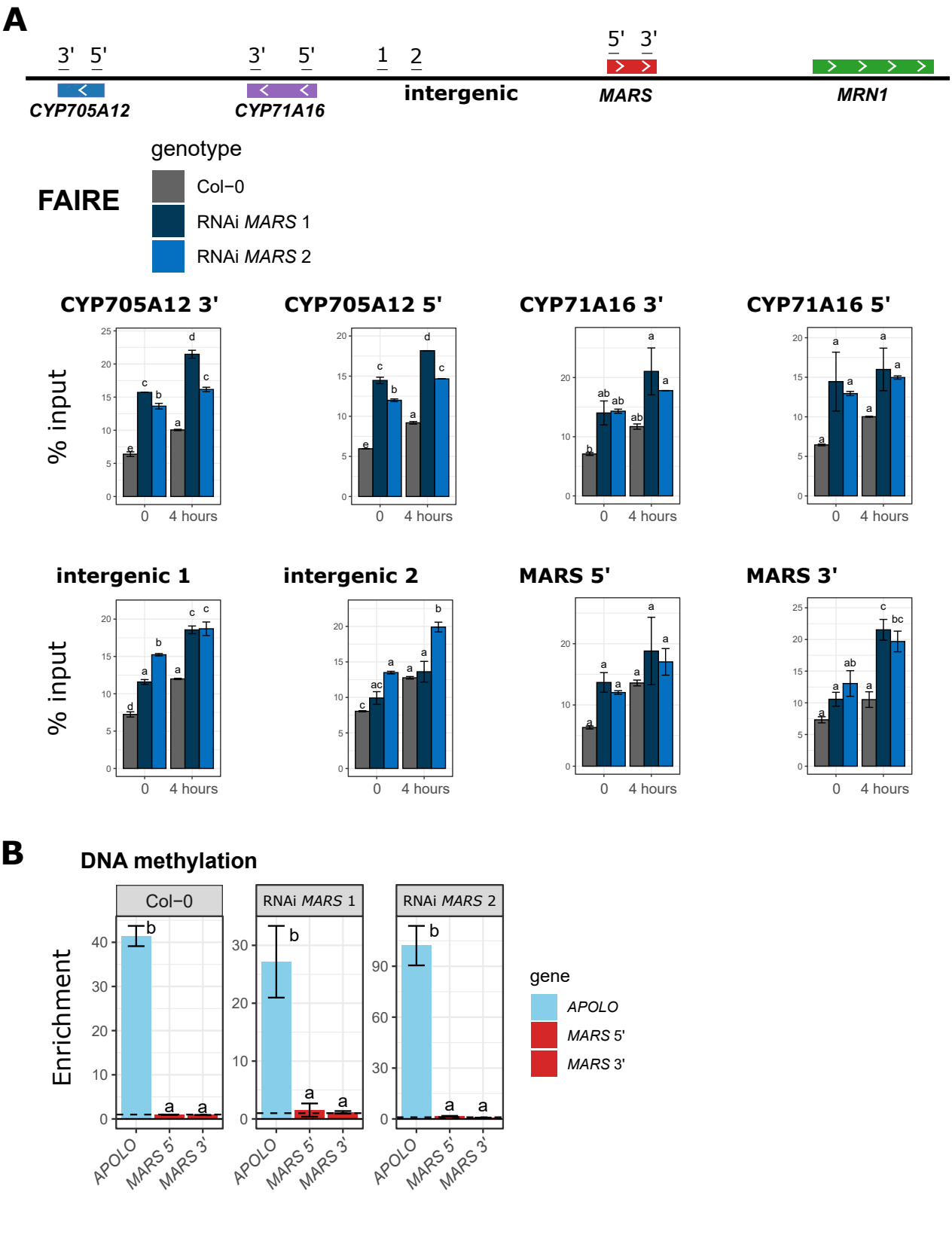
FigureS11



LHP1

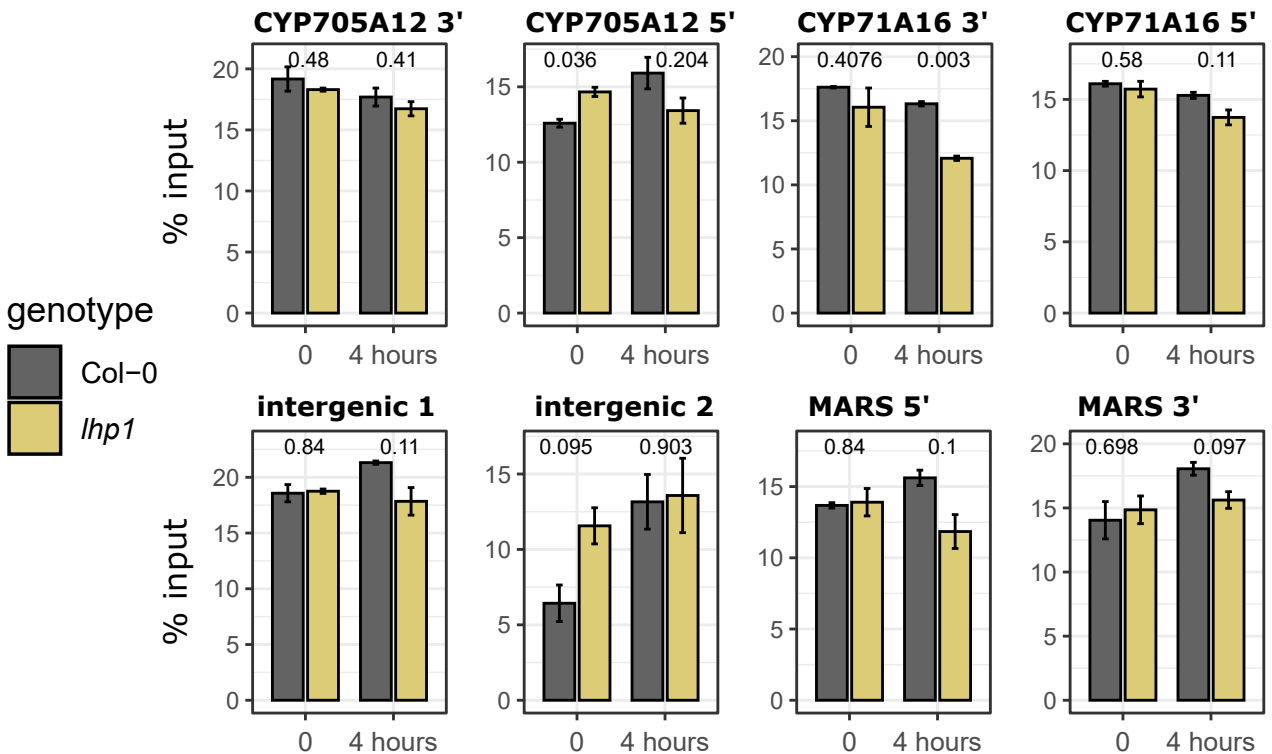


FigureS12

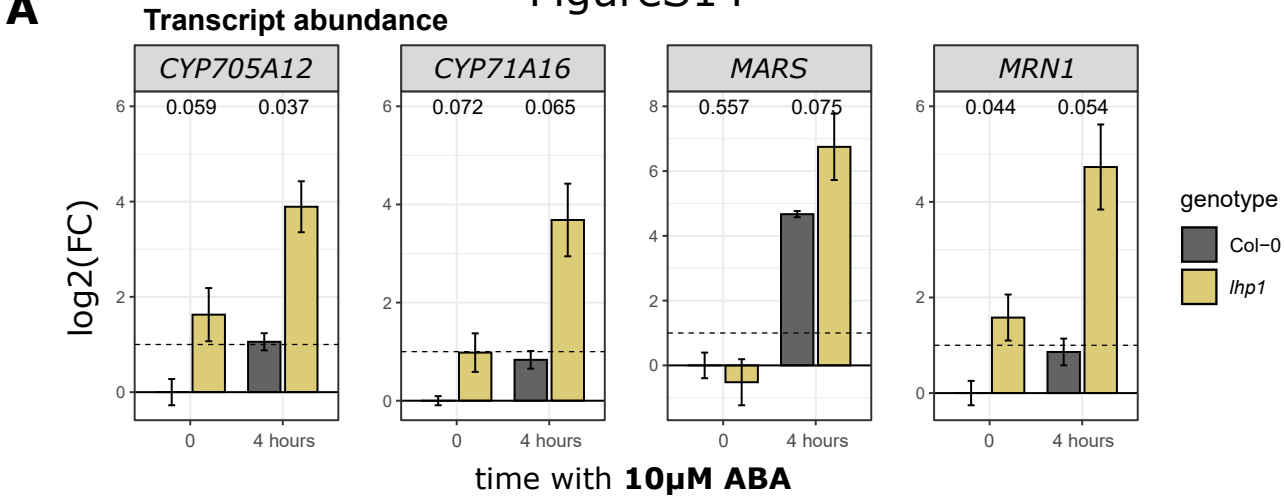

A
3' 5'
3' 5'
1 2
5' 3'
CYP705A12
CYP71A16
intergenic
MARS
MRN1
genotype
FAIRE
Col−0
RNAi MARS 1
RNAi MARS 2
CYP705A12 3'
CYP705A12 5'
CYP71A16 3'
CYP71A16 5'
% input
0
4 hours
intergenic 1
intergenic 2
MARS 5'
MARS 3'
B
DNA methylation
Col−0
RNAi MARS 1
RNAi MARS 2
Enrichment
APOLO
MARS 5'
MARS 3'
gene

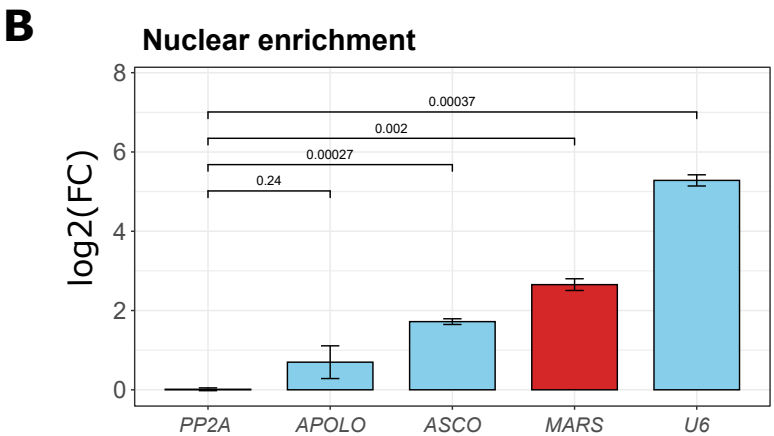
FigureS13

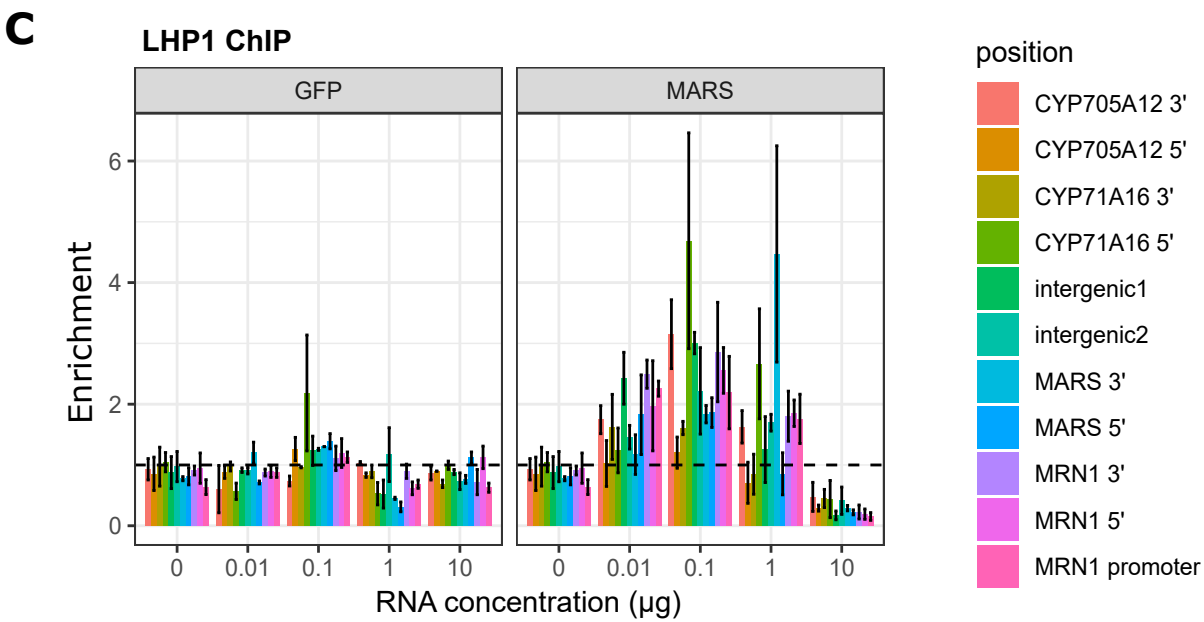
FAIRE



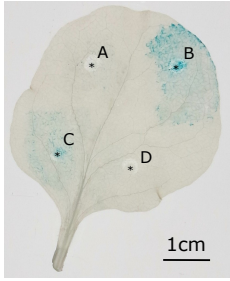
FigureS14

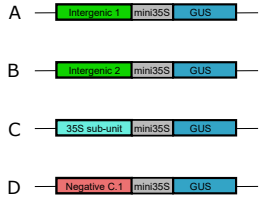

A
Transcript abundance
CYP705A12
CYP71A16
MARS
MRN1
0.059
0.037
0.072
0.065
0.557
0.075
0.044
0.054
log2(FC)
0
4 hours
time with 10µM ABA
genotype
Col−0
lhp1


B
Nuclear enrichment
log2(FC)
0.00037
0.002
0.00027
0.24
PP2A
APOLO
ASCO
MARS
U6


C
LHP1 ChIP
GFP
MARS
Enrichment
0
0.01
0.1
1
10
RNA concentration (µg)
position
CYP705A12 3'
CYP705A12 5'
CYP71A16 3'
CYP71A16 5'
intergenic1
intergenic2
MARS 3'
MARS 5'
MRN1 3'
MRN1 5'
MRN1 promoter

FigureS15


A
B
C
D
1cm


A
Intergenic 1
mini35S
GUS
B
Intergenic 2
mini35S
GUS
C
35S sub-unit
mini35S
GUS
D
Negative C.1
mini35S
GUS

TableS1	**List of primers used in the study**	
Primers used for SALK_133089 homozygous selection	***Forward primer***	***Reverse primer***
	AGTCCAGGTTTGGTTTGGTTC	ACATGTTCTTTTGGCAAGCAG
Primers used to generate the RNAi lines	***Forward primer***	***Reverse primer***
MARS	TCAATGATAGGACTACTACTCATGGCCCAGG	GATAGGACTACTACTCATGGCCCAGGTAAAC
Primers used for transcripts abundance analysis from cDNA	***Forward primer***	***Reverse primer***
CYP705A12	CAACAAGTGTTGTTTTCTCCGGTACG	GGGCTAAGATTGCATCCCTCGG
CYP71A16	CTGGTGTTTCTTTGTCGGCAGTTGT	CGGTCGGGGAATACATTCCGAGTT
MARS	ACTTTTTCACTGGTCGGTACCGAAGAC	TGCATAAAGTGTCGCACTCATTCGACTA
MRN1	GGGAGAAAGTGCTCTCTCTTGCCCTAA	GCGGCGCGATGAACAGGAGA
AT1G13320 (housekeeping gene; Czechowski et al., 2005)	GAGCTGAAGTGGCTTCCATGAC	GGTCCGACATACCCATGATCC
Primers used for ChIP, meDIP and FAIRE experiments	***Forward primer***	***Reverse primer***
CYP705A12.3'	GGAGTGCACTTAAGCGGATGAGCC	GCGATTGGAGCGATGGTGCAGT
CYP705A212.5'	GATGCGGAGATGGAGAAGAGGTCCA	CCTGCCTCCGAGCCCTCCTT
CYP71A16.3'	TGTGAGGTCGTTAAGCTCATGGAGAGG	GGAGCTGGAATCGATGTTTGCCGA
CYP71A16.5'	ATTGTGCCTTCCTCGCCGCT	TGGAGACGCTGGAAGAAGCAAGT
intergenic1	GGCTTTTGTTGACTACTATTACAGGCGGG	GGGTTTAGGGTTTATGGTTTAGGGTTTAGGG
intergenic2	ACACCTATTAGTGACATCCACAAAGCGT	ACATTTTTGTGGTGTTAGGTGTGTGAAGC

MARS.5'	TTCCTTCAGTGAGACCAGACGCTTTCA	ACGGTTTCGCGTCCCGACC
MARS.3'	ACTTTTTCACTGGTCGGTACCGAAGAC	TGCATAAAGTGTCGCACTCATTCGACTA
MRN1.5'	TCGAGCAGAAGATCCCGAGAGTGAGA	CGTTAGCCTGACATGCCGCGT
MRN1.3'	GGGAGAAAGTGCTCTCTCTTGCCCTAA	GCGGCGCGATGAACAGGAGA
APOLO	GTGGCTTCCATAGCGCCGGA	CTAGCAACAGAGACCAACCC
Primers used for *in-vitro* RNA transcription	***Forward primer***	***Reverse primer***
MARS RNA	TAATACGACTCACTATAGGGGGCGACATTCACAAAACGTCGTTAAATA	GTTGGCAGCACCCATGTTTAGATGTCC
GFP RNA	GTAATACGACTCACTATAGGATGGTGAGCAAGGGCGAG	GGCTTGTACAGCTCGTCCATGC
Primers used for transcript abundance analysis from sub-cellular localization and RIP experiments	***Forward primer***	***Reverse primer***
AT1G13320 (housekeeping gene; Czechowski et al., 2005)	GAGCTGAAGTGGCTTCCATGAC	GGTCCGACATACCCATGATCC
APOLO	GTGGCTTCCATAGCGCCGGA	CTAGCAACAGAGACCAACCC
ASCO	CCCATCGCACTGATCGGCGG	TCGAGCGCTGCCGTCTTCAC
MARS	ACTTTTTCACTGGTCGGTACCGAAGAC	TGCATAAAGTGTCGCACTCATTCGACTA
U6	CGGGGACATCCGATAAAATTGG	CGATTTGTGCGTGTCATCCTTG
MRN1	GGGAGAAAGTGCTCTCTCTTGCCCTAA	GCGGCGCGATGAACAGGAGA
Primers used for 3C experiment	***Forward primer***	***Reverse primer***
Chromatin loop	GGCTTTTGTTGACTACTATTACAGGCGGG	CCAGACCAGTCATACACTCCTAGAACCTG
Control	GTCCGAATCTTACGGACCGGATTGTC	ACTGATAAACCCATCACCGGTGTTTCC

Primers used for the GUS transient assay	***Forward primer***	***Reverse primer***
Negative control 1	AACAGGTCTCAACCTAACTAAGTGTTACTAAATCATCTCACC	AACAGGTCTCTTGTTCATTTGTAAATCTTTTAAAGCCCTTG
Negative control 2	AACAGGTCTCAACCTTGCTCAGAATGCCCCTACCT	AACAGGTCTCTTGTTCCTTGGTAACCCAAGCAACCA
Intergenic 1	AACAGGTCTCAACCTGATTTTGGAGTTCCTGGTAAATGT	AACAGGTCTCTTGTTATGTCGCACTCATTCGACTA
Intergenic 2	AACAGGTCTCAACCTGCATGTTGCCTATGATTAGAAGGA	AACAGGTCTCTTGTTACTCGATTTGGAGAAGTGTTTCAA
minimal 35S promoter element	AACAGGTCTCAAACAGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTG	AACAGGTCTCTAGCCCAGCGTGTCCTCTCCAAATGAAATGAACTTCCTTATATAGAGGAAGGGTCTTGC
sub-unit B3 from 35S promoter element	AACAGGTCTCAACCTCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCA	AACAGGTCTCTTGTTTGGGTGGGGGTCCATCTTTGGGACCACTGTCGGCAGAGGCATCTTCAACGATG

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. *MARS* cannot regulate the marneral cluster in *trans*

Transcript abundance of the marneral cluster genes in response to 10µM ABA treatment in lines overexpressing *MARS* from the *UBIQUITIN 10* promoter (UBQ). RNA was extracted from individual T2 plants from 3 independent lines. The genotypes have been classified as wild-type or overexpressor (UBQ) according to the level of expression of *MARS* taking a Col-0 line as the *MARS* wild-type background. Results are expressed as the mean ± standard error ($n \geq 4$) of the log2 fold change compared to the Col-0 genotype at time 0h.

Figure S3. *MARS* modulates the response of marneral genes to ABA without altering the plant sensitivity to an exogenous treatment

(A) Average genotype effect on transcript levels of each marneral cluster gene in three independent RNAi-*MARS* lines compared to Col-0 in response to 10µM ABA according to two-way analysis of variance (ANOVA) including genotype and time as additive factors. Data are presented in Fig 2B. For each effect, numbers indicate the p-value of the difference between the RNAi lines and Col-0 by Tukey's post-hoc test.

(B) Transcript levels of two ABA marker genes, *RAB18* and *RD29B,* in response to 10µM 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.

(C) Average genotype effect on the transcript levels of two ABA marker genes, *RAB18* and *RD29B,* in RNAi-*MARS* lines compared to Col-0 in response to 10µM ABA according to two-way analysis of variance (ANOVA) including genotype and time as additive factors. Data are presented in Fig S3B. 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 S4. *MARS* modulates the response of marneral genes to high concentrations of ABA

(A) Transcript abundance of the genes of the marneral cluster in response to 100 µM ABA in RNAi-*MARS*. 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* 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 S4A. 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 levels 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 S4C. 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 levels of the marneral cluster genes in response to water 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 levels of two ABA marker genes, *RAB18* and *RD29B*, in response to water 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 S3A. 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. Deregulation of *CYP71A16* does not modulate marneral cluster genes expression nor plant sensitivity to ABA

(A) Transcript levels of the genes of the marneral cluster in response to 10µM ABA in *mro1-2 (CYP71A16* knock-out*)* mutant. 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 the *mro1-2 (CYP71A16* knock-out*)* mutant compared to Col-0 in response to 10µM ABA according to two-way analysis of variance (ANOVA) including genotype and time as additive factors. Data are presented in Fig S6A. For each effect, numbers indicate the p-value of the difference between the *cyp71a16* mutant and Col-0 by Tukey's post-hoc test.

(C) Transcript levels of two ABA marker genes in response to 10µM ABA in *the mro1-2 (CYP71A16* knock-out*)* mutant. 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 *mro1-2 (CYP71A16* knock-out*)* mutant compared to Col-0 in response to 10µM ABA according to two-way analysis of variance (ANOVA) including genotype and time as additive factors. Data are presented in Fig S6C. For each effect, numbers indicate the p-value of the difference between the *cyp71a16* mutant and Col-0 by Tukey's post-hoc test.

Figure S7. Deregulation of *MRN1* does not modulate marneral cluster genes expression nor plant sensitivity to ABA

(A) Transcript levels of the genes of the marneral cluster in response to 10µM ABA in *35S:MRN1* and *mrn1* mutants. 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 genes of the marneral cluster in *35S:MRN1* and *mrn1* mutants compared to Col-0 in response to 10µM 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 *MRN1* mutants and Col-0 by Tukey's post-hoc test.

(C) Transcript levels of the genes of two ABA marker genes, *RAB18* and *RD29B*, in response to 10µM ABA in *35S:MRN1* and *mrn1* mutants. 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 *35S:MRN1* and *mrn1* mutants compared to Col-0 in response to 10µM 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 *MRN1* mutants and Col-0 by Tukey's post-hoc test.

Figure S8. *MARS* modulates seed germination and mannitol-dependent root growth 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 collected from plants grown separately. Time for 50% germination (T50) is indicated on the right. 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 ($p \leq 0.05$).

(B) Percentage of germinated seeds in a medium containing 0.5µM ABA. Results are expressed as the mean ±standard error (n = 7) from a batch of ≈50 seeds collected from plants grown separately. Time for 50% germination (T50) is indicated on the right. 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 ($p \leq 0.05$).

(C) Percentage of germinated seeds in a control medium. Results are expressed as the mean ±standard error (n = 4) from a batch of ≈50 seeds collected from plants grown separately. Time for 50% germination (T50) is indicated on the right. 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) Percentage of germinated seeds in a medium containing 0.5µM ABA. Results are expressed as the mean ±standard error (n = 4) from a batch of ≈50 seeds collected from plants grown separately. Time for 50% germination (T50) is indicated on the right. 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 ($p \leq 0.05$).

(E) Mean primary root length, lateral root length and lateral root density according to the genotype and the condition of 11-day-old seedlings. Letters indicate a statistical group determined by one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test. For each condition, letters indicate statistical difference between genotypes ($p \leq 0.05$).

Figure S9. The epigenetic landscape of the marneral cluster and surrounding genomic region

First track represents DNA accessibility determined by the ATAC-seq (44). ATAC-peaks are indicated with a blue rectangle and correspond to relaxed chromatin. Second and third tracks show H3K27me3 and LHP1 ChIP-Seq, respectively (43). The three experiments shown here have been performed using Arabidopsis shoot. Gene annotation is shown at the top.

Figure S10. *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 expressed 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 S11. *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 expressed 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 S12. *MARS* modulates chromatin condensation of the marneral cluster genomic region

(A) 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.

(B) DNA methylation of the *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 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 ($p \leq 0.05$). The RNAi-transgene does not affect DNA methylation at the *MARS* locus.

Figure S13. 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 S14. Nuclear-enriched *MARS* RNA modulates LHP1 binding to the marneral cluster and influences the subsequent ABA response of the marneral cluster genes

(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. Numbers are p-value of the difference between the two genotypes determined by Student t-test.

(B) 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.

(C) LHP1 binding in RNAi-*MARS*-derived chromatin to different sites across the marneral cluster upon increasing amounts of *in-vitro* transcribed *MARS* or GFP 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.

Figure S15. The intergenic region between *CYP71A16* and *MARS* is able to activate gene transcription

Agroinfiltration of the different constructs from Figure 5D in the same tobacco leaf. Points of agroinfiltration for each construct are indicated with a star (*), and each letter indicates the corresponding constructs. Same result has been observed in two other independent tobacco leaves.

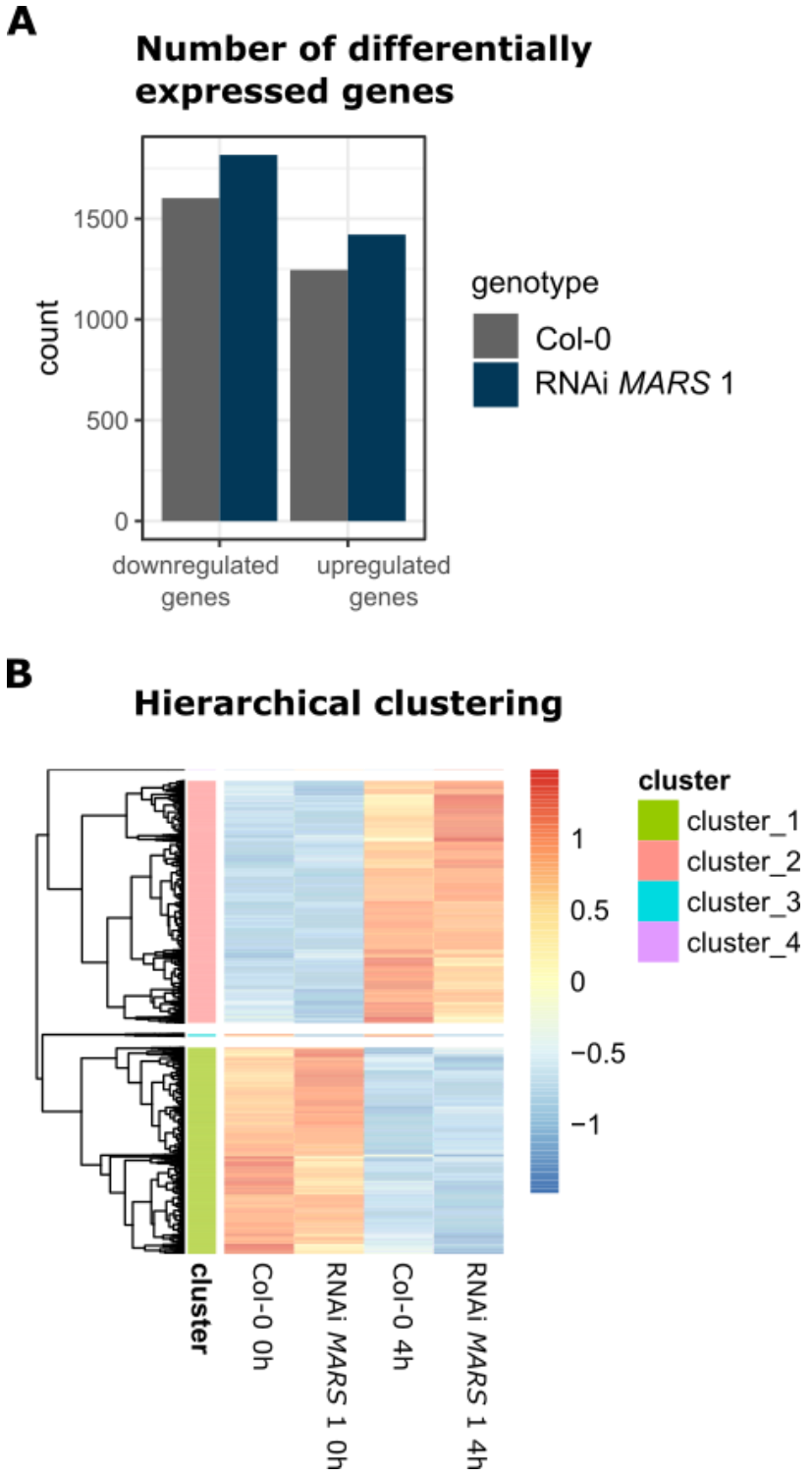


Figure 2. Differentially expressed genes in RNAi *MARS* 1. A. Number of differentially expressed (DE) genes after an ABA treatment to Col and RNAi *MARS* 1. **B.** Hierarchical clustering of all DE genes between Col and RNAi *MARS* 1 based on the DEseq2 normcounts values. Square colors correspond to the standardized normalized counts values by row as shown in the legends on the left.

6.2.4 Additional results and discussion

6.2.4.1 Deregulation of genes involved in the Carbon/Nitrogen equilibrium and cell oxidation status in the RNAi *MARS* line

MARS deregulation significantly affects seed germination and root growth under osmotic stress (Roulé et al. preprint). Thus, we wondered whether other genes outside from the marneral cluster, are also mis-regulated in the RNAi lines further supporting the phenotypes observed. To this end, we analyzed the complete transcriptome of *MARS* RNAi line 1 in response to an ABA stimulus by RNA-seq. While ABA treatment induces a strong gene deregulation in wild type, we found many more differentially expressed genes in response to ABA in the RNAi line compared to Col (3 238 versus 2 849, respectively), without any bias toward induced or repressed genes by ABA treatment (**Figure 2A**). However, under control condition, only 29 genes were differentially regulated between the two genotypes, none of them implicated in seed germination, ABA or stress-related mechanism. To further investigate the impact of the *MARS*-deregulation on the plant ABA response I conducted a hierarchical clustering of all the differentially expressed genes. This allowed us to determine two main clusters (cluster 1 and 2) that grouped the great majority of the differentially expressed genes (5 313 among 5 343 genes) that behave the same way in the two genotypes. Two smaller clusters can be isolated that present opposite expression profiles between the genotypes (**Figure 2C and 3A**). The cluster 4 contains three genes always significantly more expressed in the RNAi line, whereas the cluster 3 contains 27 genes always significantly less expressed in the RNAi line (**Figure 3A**). Notably, *QUA-QUINE STARCH* (*QQS*) from cluster 4 is a gene that when overexpressed, decreases the seed starch content (Li et al., 2009; Li and Wurtele, 2015; Li et al., 2015b). As high starch content positively correlates with better germination rate (Zhao et al., 2018c), it supports the delay in germination we observed in the *MARS* deregulated lines (**Figure 3B**). Additionally, two genes, namely *GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE 2* (*GGT2*) and *ABC1-LIKE KINASE 3* (*ABC1K3*) from cluster 3, are involved in the maintenance of a normal cell oxidation status (Ohkama-Ohtsu et al., 2007; Martinis et al., 2013) and are significantly less abundant in the RNAi line as compared to the Col (**Figure 3B**). As germination and osmotic stress are known to generate oxidative stresses (Li et al., 2017b) it is likely that the reduced activity of these genes participates in the reduced seed germination and root growth under osmotic stresses.

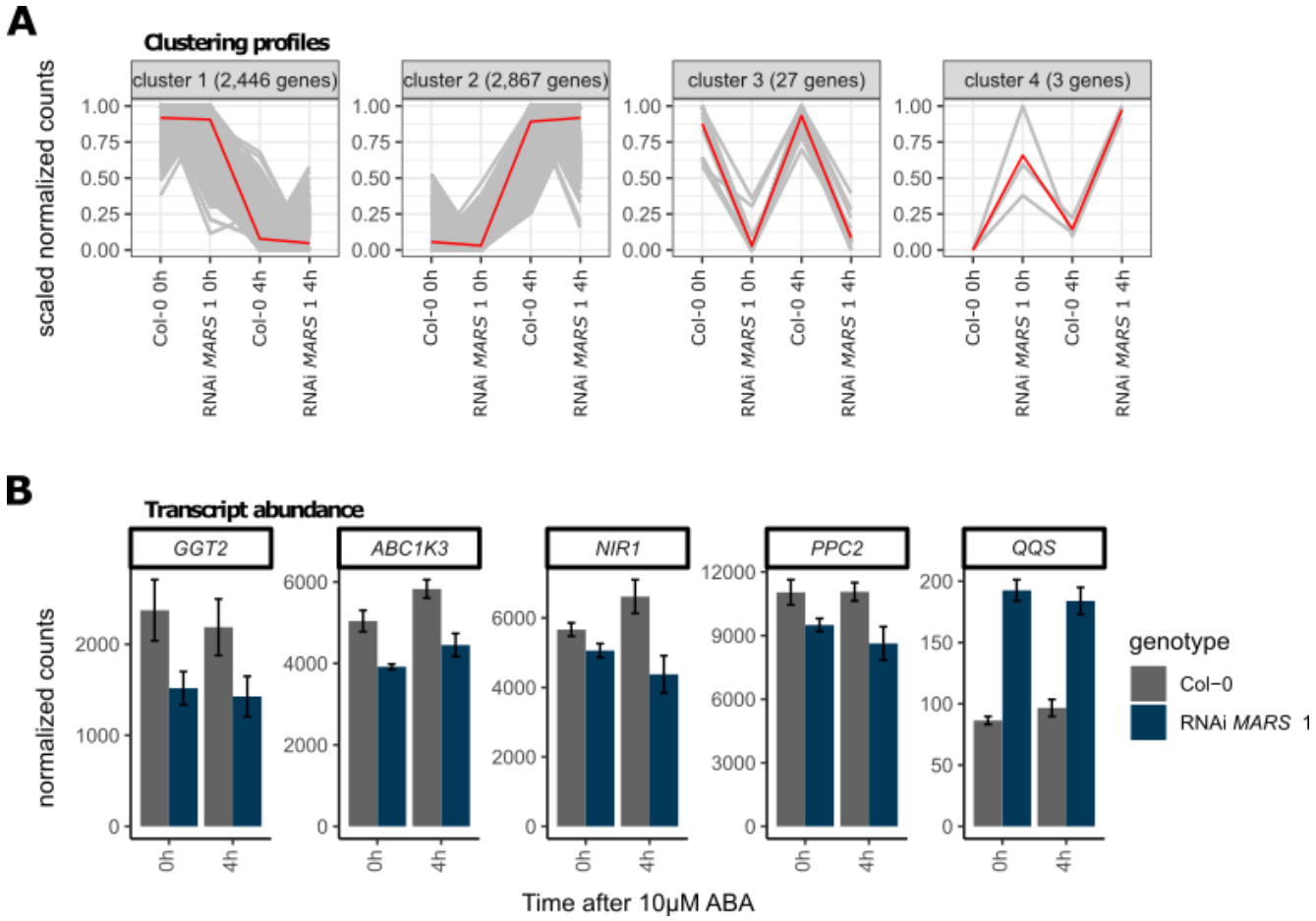


Figure 3. *MARS* downregulation impaired gene implicated in carbon/nitrogen equilibrium and cell oxidation status. A. Cluster profiles of k-means clusters based on the DEseq2 normcounts values from Col and RNAi *MARS* 1. Each grey line corresponds to a gene. The red line corresponds to the mean of all the lines in the cluster. **B.** Transcript abundance of selected indicated 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.

Finally, two genes involved in nitrate assimilation and metabolization, *NITRITE REDUCTASE1* (*NIR1*) and *PHOSPHOENOLPYRUVATE CARBOXYLASE2* (*PPC2*), respectively, are also significantly less abundant in the RNAi line which could participate in the lower germination efficiency observed in the *MARS* downregulated lines as nitrate promote seed germination at low concentration in many plant species (Shi et al., 2015; Duermeyer et al., 2018) (**Figure 3B**). Taken together, the disturbance of genes involved in the C/N balance and cell oxidation status may participate in the lower seed germination and root growth under osmotic stress in the *MARS* deregulated lines.

6.2.4.2 *MARS* physically interact with the marneral cluster genomic region to titrate LHP1 binding

In Roulé et al (preprint) we showed that *MARS* physically interacts with LHP1 and that it modulates its binding to the chromatin in a dose-dependent manner. To increase our understanding on how *MARS* recruit LHP1 to the marneral cluster we decided to look if the *MARS* RNA itself is able to physically interact with the marneral genomic region. Indeed, the *APOLO* lncRNA, which also physically interacts with LHP1, modulates LHP1 binding throughout the genome via the formation of R-loops. More precisely, the RNA-DNA *APOLO* duplex decoys LHP1 away from the chromatin, affecting chromatin configuration and gene expression (Ariel et al., 2020). Hence, we decided to use a ChIRP-qPCR approach (Ariel et al., 2020) to characterize the interaction of the *MARS* RNA with the chromatin DNA. To this end, we designed biotinylated probes complementary to the *MARS* RNA. After confirming the specificity of our probes through biotin immunoprecipitation followed by RNA extraction (**Figure 4A**), we extracted and quantified the chromatin genomic DNA copurified together with the *MARS* RNA. Strikingly, we found that *MARS* is significantly enriched in the *MRN1* genomic region and the intergenic region 2 of the DNA (**Figure 4B**) as compared to control oligos. Interestingly, the *MARS* enrichment corresponds to the region where LHP1 binding was detected (**Figure 4C**) with higher LHP1 binding and *MARS RNA*-DNA interaction in the intergenic region 2 and the *MRN1* 5' and promoter region. Taking together, *MARS* physical interaction with the marneral cluster genomic region may facilitate the recruitment and binding of LHP1 to the chromatin within this genomic region, as it has already been shown with the *APOLO* lincRNA and LHP1 binding to auxin-responsive genes (Ariel et al., 2020).

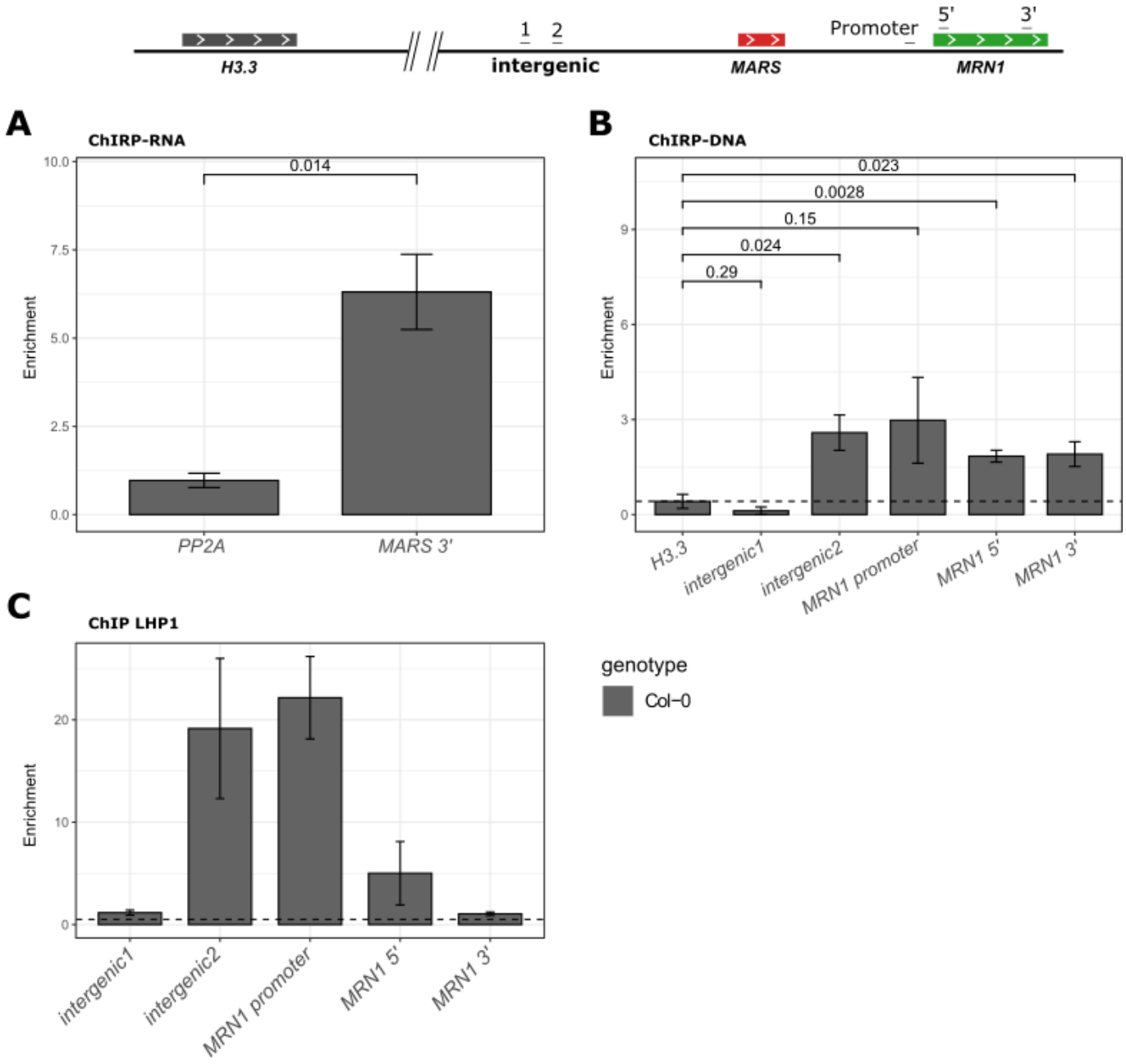


Figure 4. The *MARS* lncRNA physically interacts with the marneral cluster genomic region. A. The biotinylated probes against the *MARS* sequence significantly capture the MARS RNA (ChIRP RNA) Results are shown as the mean ± standard error (n = 3) of the *MARS/LacZ* probes ratio, used as non-specific oligos. The *PP2A* gene transcripts are used as a negative control. **B.** *MARS* RNA associates with the marneral genomic region (ChIRP-DNA). Enrichment of MARS RNA on marneral cluster genomic region is shown as the mean ± standard error (n = 3) of the *MARS/LacZ* probes ratio. Numbers are the p-value of the difference between the *H3.3* (*AT4G00430*) taken as negative control and the different marneral genomic position determined by Student t-test. **C.** LHP1 binding to the marneral genomic region in Col-0 (ChIP). Higher values of ChIP-qPCR indicate more LHP1 deposition. Deposition of LHP1 is found on chromatin DNA using qPCR amplification and oligos from the *MRN1* 5′ region, intergenic 2 and *MRN1* promoter. Results are shown as the mean ± standard error (n = 2) of the LHP1/IgG ratio, used as IP control.

6.2.5 Methods

6.2.5.1 Data processing and statistical analysis

Except if stated otherwise, all data processing and statistical analyses were realized in R v3.4.2 (R Core Team, 2017) with the help of the tidyverse (v1.2.1; Wickham, 2019).

6.2.5.2 Identification of *MARS*

The coding gene and lncRNA annotations were collected using the transcriptomics data from Blein et al (2020). Similarly, the Col-enriched genes, together with their respective normalized counts all along the Pi shortage were retrieved from the Blein et al (2020) study. The lncRNA neighboring genes were picked up with bedtools closest using the default parameter (Quinlan et al., 2010). Pearson correlation analyses were realized with the help of corrplot (v0.9; (Wei et al., 2021)) R packages.

6.2.5.3 Plant materials and growth condition

All plants used in this study are in the Columbia-0 (Col-0) background. The RNAi-*MARS* were obtained as described (Roulé et al preprint). Plants were grown in plates containing solid half-strength MS medium (MS/2) supplemented with 0.7% sucrose, vertically positioned within a growing chamber. Plants were grown for 11 days after sowing in long day conditions (16 h in light 150uE; 8 h in dark; 21°C). ABA treatments were conducted by spraying plants with water containing 10µM ABA for four hours.

6.2.5.4 Library construction and sequencing

Total RNA of whole seedlings of three biological replicates from control and ABA-treated plants were extracted using TRI Reagent (Sigma-Aldrich) and treated with DNase (Fermentas) as indicated by the manufacturers. For each sample, libraries were processed with one microgram of total RNA using an Illumina Truseq Stranded mRNA library preparation kit following the manufacturer's instructions. Libraries were sequenced on a NextSeq 500 Sequencing System (Illumina) using 75-nt single-end reads.

6.2.5.5 Differential expression analysis and clustering

All reads 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 2012) 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 genotypes and the treatment (two levels each). Low counts were discarded using DESeq2 independent filtering with default parameters and raw p-values were adjusted with the 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 et al 2019) R packages. The optimal number of clusters were defined through the Elbow method with pkgs (v1.8.0; (Zhang et al., 2018b)). Clustering profiles were determined from the clustered heatmap.

6.2.5.6 ChIRP followed by qPCR

The ChIRP protocol has been adapted from Chu et al (Chu et al., 2012) and already described in Ariel et al (2014; 2020). Briefly, two grams of ABA-treated seedlings of 11 DAS were cross-linked 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). After a brief centrifugation, the supernatants were collected and incubated with 100 pmol of complementary biotinylated probes for the *MARS* transcript. Biotinylated probes targeting the LacZ RNA were used as negative control. Magnetic streptavidin beads were used to recover the probes and RNA/chromatin bind to it. For ChIRP-RNA, the co-purified ribonucleoprotein complexes were eluted and RNA were extracted with 1mL of TRI Reagent (Sigma-Aldrich)

following the instructions of the manufacturer. Reverse transcription was performed using the total RNA and the Maxima Reverse Transcriptase (Thermo Scientific). For ChIRP-DNA, the DNA from the co-purified ribonucleoprotein complexes were extracted using Phenol:Chloroform:Isoamyl Acid (25:24:1, Sigma) followed by ethanol precipitation. For both ChIRP assays, qPCR was performed on a Light Cycler 480 with SYBR Green master I (Roche) in standard protocol (40 cycles, 60°C annealing). For ChIRP-RNA data were analyzed using the ΔΔCt method using *PROTEIN PHOSPHATASE 2A SUBUNIT A3* (*AT1G13320*) for gene normalization. For ChIRP-DNA the enrichment was calculated as the ratio of *MARS*/LacZ probes taking the *H3.3* (*AT4G40030*) genomic region as a negative control for statistical enrichment analysis.

6.2.5.7 ChIP followed by qPCR

ChIP was performed using anti-IgG (Millipore,Cat#12-370) 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. Briefly, 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 Protein A Dynabeads (Invitrogen) were incubated 12 hours at 4 °C with the samples. Immunoprecipitated DNA was recovered using Phenol:Chloroform:Isoamilic Acid (25:24:1, Sigma) followed by ethanol precipitation and quantified by qPCR. For input samples, non-immunoprecipitated sonicated chromatin was processed in parallel.

6.2.6 Conclusion and perspectives of complementary results

The marneral is a key secondary metabolite for *Arabidopsis thaliana* growth and development and it is produced by the enzymes encoded in three genes organized in one cluster (Field et al., 2011; Go et al., 2012). Interestingly, we noticed that a lincRNA embedded in this cluster is positively co-regulated with the coding genes of the marneral cluster and it is involved in their ABA-mediated gene expression. The downregulation of *MARS* transcript abundance results in ABA-dependent transcriptomic changes. Notably, thirty genes showed a contrasted expression pattern between Col and the RNAi *MARS* 1 line. Among them, genes involved in the C/N balance and the antioxydant regulatory system (**Figure 3B**), that could participate in the reduced seed germination efficiency and enhance osmotic stress sensitivity observed within the *MARS* and marneral cluster genes deregulated lines (Li et al., 2017; Zhao et al., 2018; Roulé et al preprint).

As an additional analysis, we showed that the *MARS* transcript physically interacts with the genomic region of the marneral cluster (**Figure 4B**). Remarkably, the *MARS*-DNA interaction corresponds to the chromatin regions showing binding to LHP1 (**Figure 4C**), strengthening the dependency between *MARS* and LHP1 binding. Interestingly, the *APOLO* lncRNA fine-tunes the chromatin topology and epigenetic landscape of independent genomic regions all along the *Arabidopsis* genome through the formation of R-loop (Ariel et al., 2020). *APOLO*-R-loop formation mediates the binding of LHP1 onto the chromatin, resulting in transcriptional activity changes of more than 150 genes, involved in a wide-range of physiological processes, including root gravitropism (Ariel et al., 2014), root hair formation (Moison et al., 2020), LR development and more generally auxin signaling pathway (Ariel et al., 2020). Thus, the transcriptional activity changes observed within the *MARS* downregulated line could be directly controlled by the *MARS* RNA itself through the formation of R-loop, or be a consequence of the marneral genes' increased transcriptional activity. To confirm this assumption, it is necessary to profile genome-wide the DNA-RNA hybrid R-loops as well as the *MARS* transcript DNA localization under the ABA stimulus inducing the accumulation of *MARS*. Indeed, even though the marneral epigenetic landscape is changing upon *MARS*

downregulation (decrease of H3K27me3 mark and LHP1 deposition), we could not correlate this with changes in marneral genes transcription, indicating that an additional ABA-requiring factor is needed for the *MARS*-mediated transcriptional activity changes of marneral cluster genes. Likely, the chromatin is affected without affecting transcription but predisposing this locus to a modified response against the ABA stimulus. This resembles a priming response as sown for many environmental stresses (Hilker et al., 2016).

Hence, *MARS* is a lncRNA able to regulate the marneral cluster epigenetic landscape and chromatin conformation, which changes the marneral cluster genes transcriptional responsiveness to ABA stimulus. *MARS* downregulation affects the plant sensitivity to osmotic stress and ability of the seeds to germinate, through the deregulation of marneral cluster genes and/or other ABA-related genes. With nearly five hundred clusters within *Arabidopsis* (Yu et al., 2016; Kautsar et al., 2017) together with more than thousands of lncRNAs (Blein et al., 2020), it is likely that other lncRNAs could be implicated in the co-regulation of clustered genes. Particularly, we found that lncRNAs within clusters are significantly more co-regulated with metabolic clustered genes than non-metabolic clustered genes, hinting at an intriguing putative co-evolutionary relationship between BGCs and lncRNAs (Roulé et al., preprint).

III Conclusions and perspectives

7. The non-coding transcriptome, signature of the plant local environment

7.1 Conservation of non-coding genes

The extensive characterization of the root transcriptomes of two *Arabidopsis* accessions presenting a contrasted response to phosphate (Pi) shortage, likely revealed a co-evolutionary link, in-between non-coding and coding genes, potentially linked to different root growth regulatory mechanisms. The QTL associated with the Col-0 and L*er* differential root architecture under Pi shortage was mapped to the *LOW PHOSPHATE ROOT1* (*LPR1*) quantitative trait locus (Reymond et al., 2006). Notably, the Col-specific root growth arrest under low Pi condition depends on iron (Fe) availability in the soil or media. During a Pi shortage in the Col-0 accession, a Fe-dependent callose deposition is occurring within the root elongation zone, avoiding cell elongation and consequently stopping cell division within the Root Apical Meristem (RAM) (Balzergue et al., 2017; Gutiérrez-Alanís et al., 2017). This mechanism is governed by LPR1-PDR2 and STOP1-ALMT1 modules which control Fe accumulation within the root under low Pi conditions. Interestingly, the *LPR1* gene is absent in the L*er* ecotype, participating to its continuous root growth even in Pi starvation (Reymond et al., 2006). We detected a surprising bias of expression for non-coding genes upon the early Pi starvation plant response. Indeed, as compared to the protein-coding genes, the non-coding genes were significantly more differentially regulated in-between the ecotypes, suggesting they could also participate in the divergent Pi-starvation-dependent root growth response observed in Col-0 and L*er* accessions. Notably, we could identify ecotype-enriched expression of the root growth regulators *PHT1;2*, *SPX4* and *NIP3;1* correlating with the expression of closely located ecotype-enriched lncRNAs (Blein et al., 2020).

We observed a lower level of conservation for non-coding genes as compared to coding genes (Blein et al., 2020). In agreement, the 1000 human genome (Sudmant et al., 2015) and 1001 *Arabidopsis* genome (2016) consortiums show that the gene sequence conservation between individuals is higher in coding than in non-coding portions of the genome. Nevertheless, this does not explain the significant gene expression differences observed between the two ecotypes. Clearly, it will be also relevant to compare the level of conservation of gene promoters instead of the gene core sequences. Interestingly, two studies in mammalian organisms revealed that the level of conservation of the non-coding gene promoter is as comparable as that of coding genes (Ponjavic et al., 2007; Melé et al., 2017). Notably, some lncRNA promoters contain *cis*-regulatory motifs such as transcription factor (TF) binding sites which are conserved within 11 tetrapods species (Necsulea et al., 2014). Thus, the survey of promoter sequence conservation between Col-0 and L*er*, together with the transcriptomic information we produced could increase our understanding of the basis for the non-coding genes' accession-specific expression. In our context, it is conceivable that non-differentially expressed genes will have conserved promoter sequences in-between Col-0 and L*er*, whereas the ecotype-specifically expressed genes will not. This difference may be more marked for non-coding genes as compared to coding genes since the formers are much more expressed in an ecotype-specific way (Blein et al., 2020). Finally, it is likely that a higher proportion of ecotype specific-expressed genes will be involved in the plant environmental cues perception as compared to the generally expressed genes. Indeed, as ecotype specific genes are enriched in lowly conserved non-coding genes (Blein et al., 2020), it is unlikely that they present a constitutive expression. Rather, they should be specifically expressed under specific environmental constraints, constituting a transcriptomic signature of their respective ecotype environment.

Another point to consider is that the positioning of non-coding genes throughout the genome may be conserved independently of their sequences. For example, in zebrafish it has been estimated that only 5% of lncRNAs are conserved at the sequence level within the animal kingdom (Ulitsky et al., 2011; Hezroni et al., 2015). Nevertheless, lncRNAs appear more often in syntenic positions as opposed to protein-coding genes (Ulitsky et al., 2011). The enriched appearance of lncRNAs in conserved genomic positions indicates that there is an evolutionary pressure to preserve lncRNAs in these positions while the rest of the genome is rearranged. In agreement, the SyntDB lncRNAs database has defined positionally conserved lncRNAs across primates (Bryzghalov et al., 2020). Altogether, it could be noteworthy to look for syntenic conservation of non-coding

genes, together with their transcriptional activity, to find syntenic-conserved non-coding genes between plant species, potentially participating in the regulation of their closely located protein-coding genes.

7.2 The lncRNA features: an advantage for a quick adaptation to environmental changes?

The comparison of lncRNA promoter versus the ones of coding genes has been intensively studied in mammals. It has been revealed that lncRNA promoters contain less TF binding site, are more subjected to Pol II pausing, and present a higher level of chromatin condensation as compared to the coding genes promoter (Schlackow et al., 2017b; Golicz et al., 2018; Sarropoulos et al., 2019). Also, lncRNA often demonstrate a strong cell-specific or tissue-specific expression patterns as compare to protein coding mRNAs (Cabili et al., 2011; Wang et al., 2014). All these characteristics may participate in the lower abundance of non-coding transcripts as compared to coding transcripts (Mattioli et al., 2019). Indeed, one of the particularities of non-coding genes is that they are expressed, or detected, in a lower extent compared to coding genes. This assessment has also been confirmed in our studies, where non-coding transcripts in both Col-0 and L*er* were significantly less abundant than the coding transcripts (Blein et al., 2020). This could explain why lncRNA molecules were discovered later as the sensitivity of the transcriptomic tools increased. Another explanation of their lower abundance could be that long non-coding transcripts are less stable than coding transcripts. For this latter hypothesis, several studies in mammals have used metabolic labelling (Mukherjee et al., 2017b) or drugs inhibiting gene transcription (Clark et al., 2012; Melé et al., 2017) to assess global transcript stability. However, results vary according to the method used, making it difficult to conclude whether transcript stability participates in the lower transcript abundance of lncRNAs. Thus, it could be interesting to repeat these experiments within *Arabidopsis* or other plant species, using both methods, to know if lncRNAs are less stable than coding mRNAs in the plant kingdom. Anyhow, the lower stability of non-coding transcripts, together with their high responsiveness and specificity of expression to environmental cues (Di et al., 2014), could participate in the plant's rapid adaptation to its environment. Indeed, with low transcript abundance and high responsiveness, a plant's transcriptome can rapidly adapt to changes in its environment, increasing the rapid adaptability of the cell to reach an equilibrium with all the environmental signals it perceives.

7.3 Two ecotype-associated lncRNAs in the environmental control of plant growth and development

Exploring the Col-0 and L*er* transcriptomic landscapes allowed us to discover new root-expressed lncRNAs involved in the quantitative control of root architecture. Notably, the overexpression of two ecotype-enriched lincRNAs, namely *NPC48* and *NPC72* from Ler and Col-0, respectively, resulted in a significant limitation of primary root growth under normal conditions (Blein et al., 2020). As ectopic overexpression correlates with root-architecture changes, it is likely that these two lincRNAs act in trans to regulate their target genes. Transcriptomic analyses of *NPC48*-overexpressing plants have been conducted. Strikingly, despite the strong phenotypic impact of the *NPC48* overexpression, the transcriptomes were not dramatically different from Col plants. Thus, it is likely that *NPC48* is a subtle root growth regulator. Interestingly, these two root-growth-modulators lincRNAs, *NPC48* and *NPC72*, were initially identified in a genome-wide bioinformatic analyses of cDNA libraries that identified 76 Arabidopsis lncRNAs (Amor et al., 2009). In the same study, plants overexpressing the NPC48 present strong developmental defects, including late flowering and abnormal leaf shape (Amor et al., 2009). Regarding the strong influence of *NPC48* overexpression on plant development it is surprising that the transcriptome of the *NPC48*-overexpressing plants is not dramatically affected (Blein et al., 2020). As the *NPC48* transcriptome analyses have been realized on the root compartment, one explanation for the poor root transcriptomic changes detected, together with the striking aerial phenotypes observed could be that the *NPC48* lncRNA is more implicated in the control of gene expression in late stages of plant development, notably in the aerial part of the plant. Additionally, the root growth limitation observed in the *NPC48*-overexpressing line could simply be an indirect phenotype, triggered by strong plant aerial defects. Hence, *NPC48* could be an important regulator of the leaf and flower establishment. On the other hand, the influence of *NPC72* overexpression on the transcriptome is unknown, but it has been shown that *NPC72* is responsive to salt stress (Amor et al., 2009), hinting at a potential relationship between *NPC72* and the plant salt-stress response. To go further, it could be interesting to grow *NPC72*-overexpressing lines under salt stress conditions and perform phenotypic and transcriptomic analyses. Also, it could be noteworthy to generate and analyze at the transcriptomic and phenotypic level, knock-down (RNAi) or knock-out (CRISPR deletion) mutant lines for both *NPC48* and *NPC72* to better understand their function as plant, direct or indirect, root-growth regulators.

7.4 An RdDM-acting lincRNA regulates the root system architecture

The extensive characterization of the small and long non-coding transcriptome of Col-0 and L*er* ecotypes in the early response of Pi shortage constitutes a great resource to look for functional lncRNAs. In addition to *NPC48* and *NPC72*, the *LATERALINC* lincRNA (described in chapter 2) also participates in the quantitative control of the root architecture, likely through RdDM. Notably, *LATERALINC* downregulation significantly reduced the Lateral Root (LR) length in control conditions. Initially, *LATERALINC* was selected as a putative regulator of its downstream neighboring gene *IAA14*. However, the absence of impact on *IAA14* gene expression between Col and *LATERALINC*-downregulated lines reinforces that the LR-related phenotype is occurring independently of *IAA14*. However, its upstream neighboring gene *NF-YB3* may be regulated by *LATERALINC* in an auxin-dependent context. Indeed, low levels of *LATERALINC* transcripts significantly delayed the *NF-YB3* induction in response to auxin, strengthening the potential regulatory relationship between these genes. As *NF-YB3* has been described as a positive regulator of heat stress tolerance (Sato et al., 2019), it could be noteworthy to investigate the *LATERALINC* deregulated line's sensitivity to heat stress. Notably, the soil temperature can influence many root-related developmental processes such as LR initiation, root branching, and more generally root growth (Pregitzer et al., 2000). More than thousands of genes are differentially regulated within root hair upon heat stress in soybean (Valdés-López et al., 2016), all of which hinting at a potential relationship between the root architecture and the ability of plants to face heat stress episodes. Hence, *LATERALINC* may be involved in fine-tuning the plant heat stress tolerance through modulation of the root system architecture via regulation of its neighboring gene *NF-YB3*. Nevertheless, *LATERALINC* may regulate other genes in *trans* involved in LR development or growth. To confirm this hypothesis, we could investigate whether overexpressing *LATERALINC* in *trans* modifies LR architecture. Also, transcriptomic analyses of *LATERALINC*-deregulated lines could be used to identify new direct or indirect *LATERALINC'* targets. Finally, the *LATERALINC*-mediated auxin-dependent regulation of *NF-YB3* may also directly participate in the auxin-dependent LR initiation and growth as this trait has not been studied within the *NF-YB3* overexpressing plants (Sato et al., 2019).

Taken altogether, the intensive characterization of the coding and non-coding transcriptome within two *Arabidopsis* ecotypes in the early Pi starvation response constituted a nice opportunity to discover new modulators of root-growth. Indeed, the deep-investigation of these transcriptomics data allowed us to identify several lincRNAs that modify root architecture: *NPC48* and *NPC72*, subtle regulators of the primary root growth, and *LATERALINC,* a regulator of LR development. In addition, the screening used to find the *MARS* lincRNA also identified more than thirty lincRNAs potentially involved in the regulation of expression of their neighboring genes, lying ahead of deeper investigation. The roots, in addition to anchoring them to the soil, allow water uptake and mineral nutrition required for plant growth and development. Climate change, together with the unsustainable use of mineral-enriched fertilizer, makes critical the deep understanding of the regulatory mechanisms behind the establishment of the root system architecture. Since lncRNAs regulate gene expression in a quantitatively-manner and are mainly expressed only under certain constraints (specific development stage or environment condition), they could be good candidates for targeting in plant breeding programs to subtly acclimate the emerging crops to the coming environmental changes without strongly affecting plant growth. This may also serve to improve the plant nutrition under adverse environmental conditions and in nutrient-poor soils, interesting agricultural traits for the near future.

8. The *MARS* lncRNA, a novel actor in the plant response to environment

8.1 *MARS*-mediated marneral genes expression changes is involved in the plant response to its environment

The marneral cluster is a 30kb region within *Arabidopsis thaliana*, containing three genes: *MARNERAL SYNTHASE1* (*MRN1*), *CYP705A12* and *CYP71A16*, directly involved in the biosynthesis and metabolism of the marneral (Field et al., 2011). The marneral is a triterpene critical for proper plant development (Go et al., 2012). Indeed, knock-out of *MRN1* triggers general plant developmental defects such as a delay in flowering, abnormal leaf shape and less grain per siliques (Go et al., 2012). Interestingly, three additional transcriptional units are embedded in the marneral cluster. According to their features and the use of bioinformatic tools, we classified these transcripts as lncRNAs. Notably one of these transcriptional units positively correlates with the expression of the marneral genes cluster. Strikingly, RNAi-mediated decrease of this transcriptional unit, significantly enhance the marneral genes transcriptional response to an ABA stimulus. Therefore, we named this lncRNA *MARS* for *MARneral Silencing*. High responsiveness to environmental cues is characteristic of lncRNAs. Many lncRNAs are differentially regulated under abiotic or biotic stresses such as drought, heat, osmotic, hypoxic, cold and nutrient deficiency stresses or *Sclerotinia sclerotiorum* and *Fusarium oxysporum* infection, tomato yellow leaf curl virus (TYLCV) and stripe rust (Zhang et al., 2013). Furthermore, a systematic comparison of lncRNAs in *Miscanthus lutarioriparius* revealed that, when these plants face a new environment, the expression of their lncRNAs is modified with much larger amplitude than for coding mRNAs (Xu et al., 2017b). Similarly in *Arabidopsis*, lncRNAs show an enhanced transcriptional stress-response specificity to cold, heat, drought, salt and high light stresses as compared to coding mRNAs (Di et al., 2014). In agreement, *MARS* was significantly more upregulated in response to heat stress and ABA stimulus as compared to the protein-coding genes of the marneral cluster.

The phytohormone ABA is involved in a wide range of biological processes, such as seed development and germination, but also in the perception and signal transduction of many environmental cues (Vishwakarma et al., 2017). Even though we showed that the marneral cluster misregulation is associated with an increased sensitivity to osmotic stress and a seed germination delay, we have not looked at other ABA-related physiological responses such as seed maturation. Indeed, in rice almost five hundred lncRNAs exhibit different expression patterns during seed development, highlighting that lncRNAs might be important in this process (Zhao et al., 2020). In *Arabidopsis*, the non-coding transcriptome upon silique development has not been explored yet, even though time-course transcriptomic approaches during *Arabidopsis* fruit development and maturation have been realized (Mizzotti et al., 2018). Similarly, the transcriptome of *Arabidopsis* dormant seeds have been investigated without pointing out the non-coding part of it (Dekkers et al., 2016; Nelson et al., 2017). Thus, it could be noteworthy to re-investigate all these transcriptomic analyses to test if the marneral cluster genes present a peculiar expression pattern upon seed establishment. This may guide us for the associated phenotypic analysis, such as the number of grain per silique, weight of grains and seed dormancy maintenance in *MARS* and the marneral protein-coding gene mutants.

Finally, in addition to ABA, the marneral cluster responds to auxin stimulus, N and Pi starvation indicating that the marneral could be involved in a wide-range of non-ABA related physiological processes. As the marneral was initially described as a critical triterpene for the normal development of plants (Go et al., 2012), including the aerial organs, it is likely that *MARS* may regulate such developmental processes in a marneral-dependent manner. In agreement, the *APOLO* lncRNA is involved in both developmental and stress related mechanisms, such as LR development (Ariel et al., 2020) and plant cold-stress response (Moison et al., 2020). Thus, it could be noteworthy to look at other plant developmental phenotypes in *MARS* deregulated lines compared to Col.

8.2 ABA as precursor or signaling molecule for the marneral biosynthesis and metabolization

The strong developmental defects of *mrn1* knock-out mutant are associated with cell elongation defects in both the root and the shoot apical meristem (Go et al., 2012). In agreement, the *MARS*-mediated deregulation of marneral cluster genes significantly affects many ABA-related plant physiological processes including, root growth under osmotic stress or seed germination. It is tempting to assume that the marneral triterpene may participate in the ABA signaling pathway or that ABA is responsible for the biosynthesis of this secondary metabolite. Indeed, the squalene synthase (*BfSS1*) gene, involved in the biosynthesis of phytosterol and triterpene in *Bupleurum falcatum*, is induced after ABA treatment indicating that this hormone may participate in the regulation of biosynthesis of the plant phytosterol and triterpene (Kim et al., 2011b). Furthermore, ABA has been reported to regulate some of the secondary metabolites biosynthesis within plant cell cultures (Zhao et al 2000; Luo et al 2001). In addition, many ABA-related genes are differentially expressed upon the apple russet development, a process characterized by a significant triterpenoid composition shift (Falginella et al., 2021). Finally in grapevine, ABA and GA increased the content of mono and sesqui-terpenes in leaves and berries (Murcia et al., 2017). Marneral triterpene quantification through Mass-Spectrometry (MS), together with transcriptional activity measurements of marneral gene within *Arabidopsis* mutant impaired in ABA biosynthesis or signaling may provide additional evidence about the importance or direct relationship between the ABA signaling and the marneral synthesis.

8.3 *MARS*: an enhancer lncRNA located within a Super Enhancer region?

MARS is a lncRNA able to modulate the binding of LHP1, a protein linked to epigenetic modifications on chromatin (Chen et al., 2021) and a member of PRC1 who also impacts chromatin looping (Kim, 2020). Notably, knock-out of *LHP1* results in strong modifications of genome topology together with a decrease in H3K27me3 mark (Veluchamy et al., 2016). In agreement, we observed that LHP1 binding on the marneral cluster genomic region positively correlates with a condensed "linear" chromatin organization and poor gene transcription. Using RNA immunoprecipitation (RIP) assays we confirmed *in-vivo* that *MARS* interacts with LHP1. Nevertheless, RIP is a cross-linked based method that, cannot decipher whether *MARS* directly interacts with LHP1, or interacts with another member of PRC1 or PRC2. To test this, the use of direct RNA-protein interaction methods such as Tri-FC (Seo and Chua, 2017), electrophoretic mobility shift assay (Rio, 2014) or co-transfection of LHP1 and *MARS*-RNA binary vectors in a heterologous transient system will be required. Anyhow, we demonstrated that the *MARS*-mediated LHP1 binding fine-tunes the activity of an *MRN1*-enhancer located 18kb from *MARS*. In agreement, Nützman et al (2020) demonstrated that the marneral cluster harbors a compact topology in roots where the cluster is expressed as compared to the shoot where it is less active (Field et al., 2011; Nutzman et al., 2018). As the *MRN1*-enhancer activity (i.e. leading to chromatin decondensation and relaxation for transcription to occur) increases upon *MRN1* and *CYPs* gene activation, it is conceivable that the *MRN1*-enhancer can also target other genes within the cluster. Furthermore, the over-represented TF binding sites upstream and downstream to the *MRN1*-enhancer points that this intergenic region could be a Super Enhancer (SE). Indeed, SEs are genomic regions containing multiple closely located enhancers (Huang et al., 2018). Finally, the *MARS* genomic region itself could be an enhancer region for the marneral cluster even though the large majority of transcribed enhancers generates small short, bi-directional, unspliced and non-polyadenylated transcripts. Recently, certain enhancers were shown to also generate long, unidirectional, spliced and polyadenylated transcripts such as the *MARS* RNA (Sartorelli and Lauberth, 2020). An integrative analysis of chromatin configuration and condensation upon an ABA stimulus could help to gain a complete view of the activity of the marneral cluster enhancers. For example, the use of capture Hi-C (i.e. probes

enriching the marneral cluster genomic region followed by topological analyzes) together with ATAC-seq (method allowing the identification of decondensed regions of the chromatin susceptible to *in vitro* transposition) upon ABA stimulus would help to precisely localized the marneral enhancers, and the spatial arrangements in relation to their relative targets in chromatin linked to ABA action.

8.4 *MARS* act in *cis* or in *trans* for the control of marneral cluster gene expression?

RNAi-mediated approach, insertional mutant and ectopic overexpression were used to change *MARS* transcripts abundance. Unfortunately, none of these methods could clearly conclude whether the *MARS* lncRNA act in *cis* (*MARS* transcription) or *trans* (*MARS* RNA molecule acting at a distance) to regulate the marneral cluster expression. Interestingly, we showed that addition of *in-vitro* transcribed *MARS* RNA to the *MRN1* chromatin region is able to titrate LHP1 to the chromatin in a dose-dependent manner suggesting a *trans*-regulatory function. However, as the ectopic *in vivo* overexpression did not result in gene expression changes, it is likely that other ABA-related factors are needed for this *trans* activity to titer LHP1 in the frame of the *MARS*-dependent marneral cluster regulation. CRISPR-based deletion approaches may be used to remove the *MARS* locus from the marneral cluster to further propose that the *MARS*-dependent regulation occurs in *cis* on its embedded cluster (i.e. that *MARS* transcription is linked to its regulatory function). Nevertheless, deletion of a 2kb region within a three-dimensionally regulated locus may result in unrelated side-effects on the topological status of the locus. An alternative will be the use of a dead-Cas9 (dCas9) based approach, together with a transcriptional inhibitor or activator (dCas9a) of the targeted loci. By activating and inactivating MARS *in situ* without changing the DNA sequence, this may help to conclude whether the *MARS* lncRNA transcription is required for its function. For example, a comparable molecular phenotype between plants overexpressing *MARS* in *trans* (ectopic overexpression) or in *cis* (dCas9a), will reinforce that *MARS* act in *trans* (i.e. *MARS* transcription not involved in its regulatory function) for the marneral cluster regulation. DCas9 systems start to be widely used in plants (Lowder et al., 2018; Lee et al., 2019; Moradpour and Abdulah, 2020). Notably, *Flowering Locus T* (*FT*) gene expression and epigenetic status have been modified in

Arabidopsis remarkably resulting in flowering time changes (Lee et al., 2019). Nevertheless, as the addition of *in-vitro* transcribed *MARS* RNA is sufficient to change the epigenetic landscape of the marneral cluster purified chromatin, it is likely that *MARS* act, at least partially, in *trans*.

While prokaryotic organisms use operon configuration to ensure gene co-expression (Lawrence, 2002), the localization of eukaryotic genes within the genome is quite random, even for genes involved in the same biological function (Nutzman et al., 2018). However, Biosynthetic Gene Clusters (BGCs) are a eukaryotic exception in which genes involved in the same biological pathway are embedded in a local hotspot (Hurst et al., 2004; Michalak, 2008). In *Arabidopsis thaliana*, the local chromatin conformation of three secondary metabolite clusters, including the marneral cluster, exhibit a compact configuration when active, suggesting a conserved topologically-based regulatory mechanism (Nutzman et al., 2020). As we noticed that a significant proportion of gene cluster contain a co-regulated lncRNA, notably for the metabolic clusters (Roulé et al., preprint), it is conceivable that lncRNA-mediated local epigenetic remodeling may exists in other cluster than the marneral and constitute a conserved mechanism to ensure cluster genes co-regulation.

Also, regarding the high quantity of lncRNAs and enhancers, yet uncharacterized it is likely that other lncRNAs sharing similar molecular mechanisms as *MARS* lncRNA will be found elsewhere in *Arabidopsis thaliana* genome, or more broadly in other plant species. Taken all together the discovery of the *MARS* lincRNA, together with its mechanistic characterization, led us to think that lncRNAs could constitute general regulators of the activity of enhancer in plants.

References

Abranches, R., Beven, A. F., Aragón-Alcaide, L., and Shaw, P. J. (1998). Transcription sites are not correlated with chromosome territories in wheat nuclei. *J. Cell Biol.* **143**:5–12.

Alexandre, C., Möller-Steinbach, Y., Schönrock, N., Gruissem, W., and Hennig, L. (2009). Arabidopsis MSI1 is required for negative regulation of the response to drought stress. *Mol. Plant* **2**:675–687.

Allen, B. L., and Taatjes, D. J. (2015). The Mediator complex: A central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**:155–166.

Allen, E., Xie, Z., Gustafson, A. M., and Carrington, J. C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**:207–221.

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M. M., Cao, J., Chae, E., Dezwaan, T. M. M., Ding, W., et al. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* **166**:481–491.

Alseekh, S., Kostova, D., Bulut, M., and Fernie, A. R. (2021). Genome-wide association studies: assessing trait characteristics in model and crop plants. *Cell. Mol. Life Sci.* **78**:5743–5754.

Amasino, R. (2005). 1955: Kinetin arrives. The 50th anniversary of a new plant hormone. *Plant Physiol.* **138**:1177–1184.

Amor, B. Ben, Wirth, S., Merchan, F., Laporte, P., D'Aubenton-Carafa, Y., Hirsch, J., Maizel, A., Mallory, A., Lucas, A., Deragon, J. M., et al. (2009). Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res.* **19**:57–69.

Amoutzias, G. D., Robertson, D. L., Van de Peer, Y., and Oliver, S. G. (2008). Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* **33**:220–229.

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**:595–606.

Andersson, R. (2015). Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays* **37**:314–323.

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

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**:1055-1065.e4.

Arnone, M. I., and Davidson, E. H. (1997). The hardwiring of development: Organization and function of genomic regulatory systems. *Development* **124**:1851–1864.

Axtell, M. J., and Meyers, B. C. (2018). Revisiting criteria for plant microRNA annotation in the Era of big data. *Plant Cell* **30**:272–284.

Balzergue, C., Dartevelle, T., Godon, C., Laugier, E., Meisrimler, C., Teulon, J. M., Creff, A., Bissler, M., Brouchoud, C., Hagège, A., et al. (2017). Low phosphate activates STOP1-ALMT1 to rapidly inhibit root cell elongation. *Nat. Commun.* **8**.

Banda, J., Bellande, K., von Wangenheim, D., Goh, T., Guyomarc'h, S., Laplaze, L., and Bennett, M. J. (2019). Lateral Root Formation in Arabidopsis: A Well-Ordered LRexit. *Trends Plant Sci.* **24**:826–839.

Bardou, F., Merchan, F., Ariel, F., and Crespi, M. (2011). Dual RNAs in plants. *Biochimie* **93**:1950–1954.

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.

Bari, R., Pant, B. D., and Stitt, M. (2006). Phosphate-Signaling Pathway in Plants 1 [W][OA]. *Society* **141**:988–999.

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.

Bass, H. W., Riera-Lizarazu, O., Ananiev, E. V., Bordoli, S. J., Rines, H. W., Phillips, R. L., Sedat, J. W., Agard, D. A., and Cande, W. Z. (2000). Evidence for the coincident initiation of homolog pairing and synapsis during the telomere-clustering (bouquet) stage of meiotic prophase. *J. Cell Sci.* **113**:1033–1042.

Bazin, J., Khan, G. A., Combier, J. P., Bustos-Sanmamed, P., Debernardi, J. M., Rodriguez, R., Sorin, C., Palatnik, J., Hartmann, C., Crespi, M., et al. (2013). MiR396 affects mycorrhization and root meristem activity in the legume Medicago truncatula. *Plant J.* **74**:920–934.

Bazin, J., Romero, N., Rigo, R., Charon, C., Blein, T., Ariel, F., and Crespi, M. (2018). Nuclear speckle rna binding proteins remodel alternative splicing and the non-coding arabidopsis transcriptome to regulate a cross-talk between auxin and immune responses. *Front. Plant Sci.* **9**:1–13.

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**:268–276.

Berardini, T. Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L. A., Yoon, J., Doyle, A., Lander, G., et al. (2004). Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiol.* **135**:745–755.

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.

Billou, I., Xu, J., Wildwater, M., Willemsen, V., Paponov, I., Frimi, J., Heldstra, R., Aida, M., Palme, K., and Scheres, B. (2005). The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. *Nature* **433**:39–44.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., et al. (2007). Identification

and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**:799–816.

Bishopp, A., Help, H., and Helariutta, Y. (2009). *Chapter 1 Cytokinin Signaling During Root Development*. 1st ed. Elsevier Inc.

Black, J. C., Van Rechem, C., and Whetstine, J. R. (2012). Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* **48**:491–507.

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.

Borah, P., Das, A., Milner, M. J., Ali, A., Bentley, A. R., and Pandey, R. (2018). Long non-coding rnas as endogenous target mimics and exploration of their role in low nutrient stress tolerance in plants. *Genes (Basel).* **9**:1. – 17.

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.

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., Goodrich, J., Renou, J. P., Grini, P. E., Colot, V., et al. (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet.* **7**.

Brant, E. J., and Budak, H. (2018). Plant small non-coding RNAs and their roles in biotic stresses. *Front. Plant Sci.* **9**:1–9.

Brodersen, P., and Voinnet, O. (2006). The diversity of RNA silencing pathways in plants. *Trends Genet.* **22**:268–280.

Brown, C. J. (1999). Skewed X-Chromosome Inactivation : Cause or Consequence ? **91**:3–4.

Bryzghalov, O., Szcześniak, M. W., and Makałowska, I. (2020). SyntDB: Defining orthologues of human long noncoding RNAs across primates. *Nucleic Acids Res.* **48**:D238–D245.

Budak, H., and Akpinar, B. A. (2015). Plant miRNAs: biogenesis, organization and origins. *Funct. Integr. Genomics* **15**:523–531.

Buratti, E., and Baralle, F. E. (2004). Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Mol. Cell. Biol.* **24**:10505–10514.

Bürglin, T. R., and Affolter, M. (2016). Homeodomain proteins: an update. *Chromosoma* **125**:497–521.

Burleigh, S. H., and Harrison, M. J. (1999). The down-regulation of Mt4-like genes by phosphate fertilization occurs systemically and involves phosphate translocation to the shoots. *Plant Physiol.* **119**:241–248.

Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**:1915–1927.

Campalans, A., Kondorosi, A., and Crespi, M. (2004). Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in Medicago truncatula. *Plant Cell* **16**:1047–1059.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in polycomb-group silencing. *Science (80-.).* **298**:1039–1043.

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.
Chan, S. W. L., Henderson, I. R., Zhang, X., Shah, G., Chien, J. S. C., and Jacobsen, S. E. (2006). RNAi, DRD1, and histone methylation actively target developmentally important Non-CG DNA methylation in Arabidopsis. *PLoS Genet.* **2**:0791–0797.
Chaudhary, S., Khokhar, W., Jabre, I., Reddy, A. S. N., Byrne, L. J., Wilson, C. M., and Syed, N. H. (2019). Alternative splicing and protein diversity: Plants versus animals. *Front. Plant Sci.* **10**:1–14.
Chen, X. (2005). microRNA biogenesis and function in plants. *FEBS Lett.* **579**:5923–5931.
Chen, X. (2009). Small RNAs and their roles in plant development. *Annu. Rev. Cell Dev. Biol.* **25**:21–44.
Chen, H. M., Chen, L. T., Patel, K., Li, Y. H., Baulcombe, D. C., and Wu, S. H. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc. Natl. Acad. Sci. U. S. A.* **107**:15269–15274.
Chen, M., Wang, C., Bao, H., Chen, H., and Wang, Y. (2016). Genome-wide identification and characterization of novel lncRNAs in Populus under nitrogen deficiency. *Mol. Genet. Genomics* **291**:1663–1680.
Chen, R., Li, M., Zhang, H., Duan, L., Sun, X., Jiang, Q., Zhang, H., and Hu, Z. (2019). Continuous salt stress-induced long non-coding RNAs and DNA methylation patterns in soybean roots. *BMC Genomics* **20**:1–12.
Chen, Z., Jiang, Q., Jiang, P., Zhang, W., Huang, J., Liu, C., Halford, N. G., and Lu, R. (2020). Novel low-nitrogen stress-responsive long non-coding RNAs (lncRNA) in barley landrace B968 (Liuzhutouzidamai) at seedling stage. *BMC Plant Biol.* **20**:1–11.
Chen, Q., Liu, K., Yu, R., Zhou, B., Huang, P., Cao, Z., Zhou, Y., and Wang, J. (2021). From "Dark Matter" to "Star": Insight Into the Regulation Mechanisms of Plant Functional Long Non-Coding RNAs. *Front. Plant Sci.* **12**:1–13.
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.
Chiou, T. J., Aung, K., Lin, S. I., Wu, C. C., Chiang, S. F., and Su, C. L. (2006). Regulation of phosphate homeostasis by MicroRNA in Arabidopsis. *Plant Cell* **18**:412–421.
Cho, J., and Paszkowski, J. (2017). Regulation of rice root development by a retrotransposon acting as a microRNA sponge. *Elife* **79**:1–21.
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.
Chorostecki, U., Crosa, V. A., Lodeyro, A. F., Bologna, N. G., Martin, A. P., Carrillo, N., Schommer, C., and Palatnik, J. F. (2012). Identification of new microRNA-regulated genes by conserved targeting in plant species. *Nucleic Acids Res.* **40**:8893–8904.
Chow, H. T., Chakraborty, T., and Mosher, R. A. (2020). RNA-directed DNA Methylation and

sexual reproduction: expanding beyond the seed. *Curr. Opin. Plant Biol.* **54**:11–17.

Chu, C., Quinn, J., and Chang, H. Y. (2012). Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* Advance Access published 2012, doi:10.3791/3912.

Chung, P. J., Jung, H., Jeong, D. H., Ha, S. H., Choi, Y. Do, and Kim, J. K. (2016). Transcriptome profiling of drought responsive noncoding RNAs and their target genes in rice. *BMC Genomics* **17**.

Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., Dinger, M. E., and Mattick, J. S. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**:885–898.

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

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

Colley, S. M., and Leedman, P. J. (2011). Steroid Receptor RNA Activator - A nuclear receptor coregulator with multiple partners: Insights and challenges. *Biochimie* **93**:1966–1972.

Combier, J. P., Frugier, F., De Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernié, T., Ott, T., Gamas, P., Crespi, M., et al. (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes Dev.* **20**:3084–3088.

Conde, D., Le Gac, A. L., Perales, M., Dervinis, C., Kirst, M., Maury, S., González-Melendi, P., and Allona, I. (2017). Chilling-responsive DEMETER-LIKE DNA demethylase mediates in poplar bud break. *Plant Cell Environ.* **40**:2236–2249.

Conn, V. M., Hugouvieux, V., Nayak, A., Conos, S. A., Capovilla, G., Cildir, G., Jourdain, A., Tergaonkar, V., Schmid, M., Zubieta, C., et al. (2017). A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nat. Plants* **3**:4–8.

Contreras-Moreira, B., Cantalapiedra, C. P., García-Pereira, M. J., Gordon, S. P., Vogel, J. P., Igartua, E., Casas, A. M., and Vinuesa, P. (2017). Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species. *Front. Plant Sci.* **8**:1–16.

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**.

Cowan, C. R., Carlton, P. M., and Cande, W. Z. (2001). The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiol.* **125**:532–538.

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature* **573**:45–54.

Crespi, M. D., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y., Petrovics, G., Kondorosi, E., and Kondorosi, a (1994). enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J.* **13**:5099–5112.

Crevillén, P. (2020). Histone Demethylases as Counterbalance to H3K27me3 Silencing in Plants. *iScience* **23**:1–8.

Crevillén, P., Sonmez, C., Wu, Z., and Dean, C. (2013). A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J.* **32**:140–148.

Cubillos, F. A., Coustham, V., and Loudet, O. (2012). Lessons from eQTL mapping studies: Non-coding regions and their role behind natural phenotypic variation in plants. *Curr. Opin. Plant Biol.* **15**:192–198.

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.

de Nooijer, S., Wellink, J., Mulder, B., and Bisseling, T. (2009). Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. *Nucleic Acids Res.* **37**:3558–3568.

Debernardi, J. M., Rodriguez, R. E., Mecchia, M. A., and Palatnik, J. F. (2012). Functional specialization of the plant miR396 regulatory network through distinct microRNA-target interactions. *PLoS Genet.* **8**.

Dehghani, H., Dellaire, G., and Bazett-Jones, D. P. (2005). Organization of chromatin in the interphase mammalian cell. *Micron* **36**:95–108.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science (80-.).* **295**:1306–1311.

Dekkers, B. J. W., Pearce, S. P., Van Bolderen-Veldkamp, R. P. M., Holdsworth, M. J., and Bentsink, L. (2016). Dormant and after-ripened Arabidopsis thaliana seeds are distinguished by early transcriptional differences in the imbibed state. *Front. Plant Sci.* **7**:1–15.

Delhaize, E., and Randall, P. (1993). Characterization of a Phosphate-Accumulator Mutant of A rabidopsis thaliana Advance Access published 1993.

Deng, W., Buzas, D. M., Ying, H., Robertson, M., Taylor, J., Peacock, W. J., Dennis, E. S., and Helliwell, C. (2013). Arabidopsis Polycomb Repressive Complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC Genomics* **14**:1–12.

Deng, P., Liu, S., Nie, X., Weining, S., and Wu, L. (2018a). Conservation analysis of long non-coding RNAs in plants. *Sci. China Life Sci.* **61**:190–198.

Deng, F., Zhang, X., Wang, W., Yuan, R., and Shen, F. (2018b). Identification of Gossypium hirsutum long non-coding RNAs (lncRNAs) under salt stress. *BMC Plant Biol.* **18**:1–14.

Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgová, I., Mahrez, W., Nanni, P., Bischof, S., Gruissem, W., and Hennig, L. (2013). Arabidopsis MSI1 connects LHP1 to PRC2 complexes. *EMBO J.* **32**:2073–2085.

Devaiah, B. N., Nagarajan, V. K., and Raghothama, K. G. (2007). Phosphate homeostasis and root development in arabidopsis are synchronized by the zinc finger transcription factor ZAT6. *Plant Physiol.* **145**:147–159.

Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., Zhang, T., Qi, Y., Gerstein, M. B., Guo, Y., et al. (2014). Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. *Plant J.* **80**:848–861.

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J. E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M. G., Feldmann, K. A., and Benfey, P. N. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. *Cell* **86**:423–433.

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.

Dolfini, D., Gatta, R., and Mantovani, R. (2012). NF-Y and the transcriptional activation of CCAAT promoters. *Crit. Rev. Biochem. Mol. Biol.* **47**:29–49.

Dong, B., Rengel, Z., and Delhaize, E. (1998). Uptake and translocation of phosphate by pho2 mutant and wild-type seedlings of Arabidopsis thaliana. *Planta* **205**:251–256.

Dong, P., Tu, X., Chu, P. Y., Lü, P., Zhu, N., Grierson, D., Du, B., Li, P., and Zhong, S. (2017). 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Mol. Plant* **10**:1497–1509.

Dong, Y., Chen, H., Gao, J., Liu, Y., Li, J., and Wang, J. (2019). Bioactive ingredients in Chinese herbal medicines that target non-coding RNAs: Promising new choices for disease treatment. *Front. Pharmacol.* **10**:1–30.

Douglas, R. N., Wiley, D., Sarkar, A., Springer, N., Timmermans, M. C. P., and Scanlon, M. J. (2010). Ragged seedling2 encodes an ARGONAUTE7-like protein required for mediolateral expansion, but not dorsiventrality, of maize leaves. *Plant Cell* **22**:1441–1451.

Du, Q., Wang, K., Zou, C., Xu, C., and Li, W. X. (2018). The PILNCR1-miR399 regulatory module is important for low phosphate tolerance in maize. *Plant Physiol.* **177**:1743–1753.

Du Mee, D. J. M., Ivanov, M., Parker, J. P., Buratowski, S., and Marquardt, S. (2018). Efficient termination of nuclear lncrna transcription promotes mitochondrial genome maintenance. *Elife* **7**:1–24.

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.

Dupont, C., Armant, D. R., and Brenner, C. A. (2009). Epigenetics: Definition, mechanisms and clinical perspective. *Semin. Reprod. Med.* **27**:351–357.

Elhamamsy, A. R. (2016). DNA methylation dynamics in plants and mammals: overview of regulation and dysregulation. *Cell Biochem. Funct.* **34**:289–298.

Elliott, T. A., and Gregory, T. R. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos. Trans. R. Soc. B Biol. Sci.* **370**.

Emberley, E., Huang, G. J., Hamedani, M. K., Czosnek, A., Ali, D., Grolla, A., Lu, B., Watson, P. H., Murphy, L. C., and Leygue, E. (2003). Identification of new human coding steroid receptor RNA activator isoforms. *Biochem. Biophys. Res. Commun.* **301**:509–515.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A., et al. (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Mol. Cell* **38**:452–464.

Everett, L., Hansen, M., and Hannenhalli, S. (2010). *Regulating the Regulators: Modulators of Transcription Factor Activity*. Methods Mo. (ed. Methods Mol Biol) Methods Mol Biol.

Falginella, L., Andre, C. M., Legay, S., Lin-Wang, K., Dare, A. P., Deng, C., Rebstock, R., Plunkett, B. J., Guo, L., Cipriani, G., et al. (2021). Differential regulation of triterpene biosynthesis induced by an early failure in cuticle formation in apple. *Hortic. Res.* **8**.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M., and Colombo, L. (2003). MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. *Plant Cell* **15**:2603–2611.

Fei, Q., Xia, R., and Meyers, B. C. (2013). Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* **25**:2400–2415.

Felippes, F. F., and Weigel, D. (2009). Triggering the formation of tasiRNAs in Arabidopsis thaliana: The role of microRNA miR173. *EMBO Rep.* **10**:264–270.

Feng, S., Cokus, S. J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S. E. (2014). Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Mol. Cell* **55**:694–707.

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.

Filichkin, S., Priest, H. D., Megraw, M., and Mockler, T. C. (2015). Alternative splicing in plants: Directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. Plant Biol.* **24**:125–135.

Finkelstein, R., Reeves, W., Ariizumi, T., and Steber, C. (2008). Molecular aspects of seed dormancy. *Annu. Rev. Plant Biol.* **59**:387–415.

Fok, E. T., Scholefield, J., Fanucchi, S., and Mhlanga, M. M. (2017). The emerging molecular biology toolbox for the study of long noncoding RNA biology. *Epigenomics* **9**:1317–1327.

Foley, S. W., Kramer, M. C., and Gregory, B. D. (2017). RNA structure, binding, and coordination in Arabidopsis. *Wiley Interdiscip. Rev. RNA* **8**.

Folsom, J. J., Begcy, K., Hao, X., Wang, D., and Walia, H. (2014). Rice fertilization-independent Endosperm1 regulates seed size under heat stress by controlling early endosperm development. *Plant Physiol.* **165**:238–248.

Forde, B. G. (2014). Nitrogen signalling pathways shaping root system architecture: An update. *Curr. Opin. Plant Biol.* **21**:30–36.

Forde, B. G., and Walch-Liu, P. (2009). Nitrate and glutamate as environmental cues for behavioural responses in plant roots. *Plant, Cell Environ.* **32**:682–693.

Francis, N. J., and Kingston, R. E. (2001). Francis and Kingston 2001 **2**.

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J. A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**:1033–1037.

Fransz, P., De Jong, J. H., Lysak, M., Castiglione, M. R., and Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. U. S. A.* **99**:14584–14589.

Fukaki, H., Nakao, Y., Okushima, Y., Theologis, A., and Tasaka, M. (2005). Tissue-specific expression of stabilized SOLITARY-ROOT/IAA14 alters lateral root development in Arabidopsis. *Plant J.* **44**:382–395.

Fukuda, M., Fujiwara, T., and Nishida, S. (2020). Roles of non-coding rnas in response to nitrogen availability in plants. *Int. J. Mol. Sci.* **21**:1–15.

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.

Gallart, A. P., Pulido, A. H., De Lagrán, I. A. M., Sanseverino, W., and Cigliano, R. A. (2016). GREENC: A Wiki-based database of plant IncRNAs. *Nucleic Acids Res.* **44**:D1161–D1166.

Gao, S., Yang, L., Zeng, H. Q., Zhou, Z. S., Yang, Z. M., Li, H., Sun, D., Xie, F., and Zhang, B. (2016). A cotton miRNA is involved in regulation of plant response to salt stress. *Sci. Rep.* **6**:1–14.

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E. E. M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**:96–100.

Giannakou, I. O., and Panopoulou, S. (2019). The use of fluensulfone for the control of root-knot nematodes in greenhouse cultivated crops: Efficacy and phytotoxicity effects. *Cogent Food Agric.* **5**:1643819.

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

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

Gingeras, T. R. (2009). Implications of chimaeric non-co-linear transcripts. *Nature* **461**:206–211.

Glažar, P., Papavasileiou, P., and Rajewsky, N. (2014). CircBase: A database for circular RNAs. *Rna* **20**:1666–1670.

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.

Golicz, A. A., Singh, M. B., and Bhalla, P. L. (2018). The long intergenic noncoding RNA (LincRNA) Landscape of the soybean genome. *Plant Physiol.* **176**:2133–2147.

Gregory, T. R. (2001). Coincidence, coevolution, or causation? DNA content, cellsize, and the C-value enigma. *Biol. Rev.* **76**:65–101.

Grewal, S. I. S., and Jia, S. (2007). Heterochromatin revisited. *Nat. Rev. Genet.* **8**:35–46.

Grob, S., Schmid, M. W., Luedtke, N. W., Wicker, T., and Grossniklaus, U. (2013). Characterization of chromosomal architecture in Arabidopsis by chromosome conformation capture. *Genome Biol.* **14**:1–19.

Grob, S., Schmid, M. W., and Grossniklaus, U. (2014). Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Mol. Cell* **55**:678–693.

Guan, Q., Lu, X., Zeng, H., Zhang, Y., and Zhu, J. (2013). Heat stress induction of miR398 triggers a regulatory loop that is critical for thermotolerance in Arabidopsis. *Plant J.* **74**:840–851.

Gultyaev, A. P., and Roussis, A. (2007). Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. *Nucleic Acids Res.* **35**:3144–3152.

Guo, G., Liu, X., Sun, F., Cao, J., Huo, N., Wuda, B., Xin, M., Hu, Z., Du, J., Xia, R., et al. (2018). Wheat miR9678 affects seed germination by generating phased siRNAs and modulating abscisic acid/gibberellin signaling. *Plant Cell* **30**:796–814.

Guo, W., Wang, D., and Lisch, D. (2021). RNA-directed DNA methylation prevents rapid and heritable reversal of transposon silencing under heat stress in Zea mays. *PLoS Genet.* **17**:1–28.

Gutiérrez-Alanís, D., Yong-Villalobos, L., Jiménez-Sandoval, P., Alatorre-Cobos, F., Oropeza-Aburto, A., Mora-Macías, J., Sánchez-Rodríguez, F., Cruz-Ramírez, A., and Herrera-Estrella, L. (2017). Phosphate Starvation-Dependent Iron Mobilization Induces CLE14 Expression to Trigger Root Meristem Differentiation through CLV2/PEPR2 Signaling. *Dev. Cell* **41**:555-570.e3.

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., and Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**:240–251.

Haag, J. R., and Pikaard, C. S. (2011). Multisubunit RNA polymerases IV and V: Purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* **12**:483–492.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**:621–637.

Hammond, S. M. (2015). An overview of microRNAs. *Adv. Drug Deliv. Rev.* **87**:3–14.

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Sharpe, J. A., Suciu, M. C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.* **48**:895–903.

He, J. X., Gendron, J. M., Sun, Y., Gampala, S. S. L., Gendron, N., Sun, C. Q., and Wang, Z. Y. (2005). BZR1 is a transcriptional repressor with dual roles in brassinosteroid homeostasis and growth responses. *Science (80-.).* **307**:1634–1638.

Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science (80-.).* **315**:1141–1143.

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

Herr, A. J., Jensen, M. B., Dalmay, T., and Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science (80-.).* **308**:118–120.

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., and Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep.* **11**:1110–1122.

Hilker, M., Schwachtje, J., Baier, M., Balazadeh, S., Bäurle, I., Geiselhardt, S., Hincha, D. K., Kunze, R., Mueller-Roeber, B., Rillig, M. C., et al. (2016). Priming and memory of stress responses in organisms lacking a nervous system. *Biol. Rev.* **91**:1118–1133.

Hiratsu, K., Ohta, M., Matsui, K., and Ohme-Takagi, M. (2002). The SUPERMAN protein is an active repressor whose carboxy-terminal repression domain is required for the development of normal flowers. *FEBS Lett.* **514**:351–354.

Hiratsu, K., Matsui, K., Koyama, T., and Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. *Plant J.* **34**:733–739.

Hiratsu, K., Mitsuda, N., Matsui, K., and Ohme-Takagi, M. (2004). Identification of the minimal repression domain of SUPERMAN shows that the DLELRL hexapeptide is both necessary and sufficient for repression of transcription in Arabidopsis. *Biochem. Biophys. Res. Commun.* **321**:172–178.

Hobo, T., Kowyama, Y., and Hattori, T. (1999). A bZIP factor, TRAB1, interacts with VP1 and mediates abscisic acid- induced transcription. *Proc. Natl. Acad. Sci. U. S. A.* **96**:15348–15353.

Hobson, D. J., Wei, W., Steinmetz, L. M., and Svejstrup, J. Q. (2012). RNA Polymerase II Collision Interrupts Convergent Transcription. *Mol. Cell* **48**:365–374.

Hossain, M. S., Kawakatsu, T., Kim, K. Do, Zhang, N., Nguyen, C. T., Khan, S. M., Batek, J. M., Joshi, T., Schmutz, J., Grimwood, J., et al. (2017). Divergent cytosine DNA methylation patterns in single-cell, soybean root hairs. *New Phytol.* **214**:808–819.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes. *Adv. Bioinformatics* **2008**:1–5.

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

Hu, L., Xu, Z., Hu, B., and Lu, Z. J. (2017). COME: A robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids*

Res. **45**:1–13.

Hu, Y., Omary, M., Hu, Y., Doron, O., Hoermayer, L., Chen, Q., Megides, O., Chekli, O., Ding, Z., Friml, J., et al. (2021). Cell kinetics of auxin transport and activity in Arabidopsis root growth and skewing. *Nat. Commun.* **12**:1–13.

Huang, J. Z., Chen, M., Chen, D., Gao, X. C., Zhu, S., Huang, H., Hu, M., Zhu, H., and Yan, G. R. (2017). A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cell* **68**:171-184.e6.

Huang, J., Li, K., Cai, W., Liu, X., Zhang, Y., Orkin, S. H., Xu, J., and Yuan, G. C. (2018). Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat. Commun.* **9**.

Huang, A. C., Jiang, T., Liu, Y. X., Bai, Y. C., Reed, J., Qu, B., Goossens, A., Nützmann, H. W., Bai, Y., and Osbourn, A. (2019). A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science (80-.).* **364**.

Hubé, F., Velasco, G., Rollin, J., Furling, D., and Francastel, C. (2011). Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. *Nucleic Acids Res.* **39**:513–525.

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.

Huppke, P., Maier, E. M., Warnke, A., Brendel, C., Laccone, F., and Gärtner, J. (2006). Very mild cases of Rett syndrome with skewed X inactivation. *J. Med. Genet.* **43**:814–816.

Hurst, L. D., Pál, C., and Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**:299–310.

Ikeda, M., and Ohme-Takagi, M. (2009). A novel group of transcriptional repressors in arabidopsis. *Plant Cell Physiol.* **50**:970–975.

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J. S., Jackson, S. E., Wills, M. R., and Weissman, J. S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep.* **8**:1365–1379.

Ishiyama, K., Inoue, E., Tabuchi, M., Yamaya, T., and Takahashi, H. (2004). Biochemical background and compartmentalized functions of cytosolic glutamine synthetase for active ammonium assimilation in rice roots. *Plant Cell Physiol.* **45**:1640–1647.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**:115–120.

Jabnoune, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., and Poirier, Y. (2013). A Rice cis-Natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. *Plant Cell* **25**:4166–4182.

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.

Jégu, T., Domenichini, S., Blein, T., Ariel, F., Christ, A., Kim, S. K., Crespi, M., Boutet-Mercey, S., Mouille, G., Bourge, M., et al. (2015). A SWI/SNF chromatin remodelling protein controls cytokinin production through the regulation of chromatin architecture. *PLoS One* **10**:1–18.

Jha, U. C., Nayyar, H., Jha, R., Khurshid, M., Zhou, M., Mantri, N., and Siddique, K. H. M. (2020). Long non-coding RNAs: Emerging players regulating plant abiotic stress response

and adaptation. *BMC Plant Biol.* **20**:1–20.

Jin, J., Liu, J., Wang, H., Wong, L., and Chua, N. H. (2013). PLncDB: Plant long non-coding RNA database. *Bioinformatics* **29**:1068–1071.

Johnson, D. S., Mortazavi, A., and Myers, R. M. (2007). Protein-DNA Interactions Advance Access published 2007.

Jones-Rhoades, M. W., and Bartel, D. P. (2004). Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* **14**:787–799.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**:1313–1326.

Kagaya, Y., Hobo, T., Murata, M., Ban, A., and Hattori, T. (2002). Abscisic acid-induced transcription is mediated by phosphorylation of an abscisic acid response element binding factor, TRAB1. *Plant Cell* **14**:3177–3189.

Kageyama, Y., Kondo, T., and Hashimoto, Y. (2011). Coding vs non-coding: Translatability of short ORFs found in putative non-coding transcripts. *Biochimie* **93**:1981–1986.

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.

Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., Riddle, N. C., Verbsky, M. L., and Richards, E. J. (2003). Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* **163**:1109–1122.

Kapazoglou, A., Tondelli, A., Papaefthimiou, D., Ampatzidou, H., Francia, E., Stanca, M. A., Bladenopoulos, K., and Tsaftaris, A. S. (2010). Epigenetic chromatin modifiers in barley: IV. The study of barley Polycomb group (PcG) genes during seed development and in response to external ABA. *BMC Plant Biol.* **10**.

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.

Kenchanmane Raju, S. K., Ritter, E. J., and Niederhuth, C. E. (2019). Establishment, maintenance, and biological roles of non-CG methylation in plants. *Essays Biochem.* **63**:743–755.

Kim, D. H. (2020). Current understanding of flowering pathways in plants: focusing on the vernalization pathway in Arabidopsis and several vegetable crop plants. *Hortic. Environ. Biotechnol.* **61**:209–227.

Kim, D. H., and Sung, S. (2014). Polycomb-mediated gene silencing in Arabidopsis thaliana. *Mol. Cells* **37**:841–850.

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

Kim, S. Y., Zhu, T., and Renee Sung, Z. (2010). Epigenetic regulation of gene programs by EMF1 and EMF2 in Arabidopsis. *Plant Physiol.* **152**:516–528.

Kim, W., Ahn, H. J., Chiou, T. J., and Ahn, J. H. (2011a). The role of the miR399-PHO2 module in the regulation of flowering time in response to different ambient temperatures in Arabidopsis thaliana. *Mol. Cells* **32**:83–88.

Kim, Y. S., Cho, J. H., Park, S., Han, J. Y., Back, K., and Choi, Y. E. (2011b). Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in Bupleurum falcatum. *Planta* **233**:343–355.

Kim, S. Y., Lee, J., Eshed-Williams, L., Zilberman, D., and Sung, Z. R. (2012). EMF1 and PRC2

cooperate to repress key regulators of arabidopsis development. *PLoS Genet.* **8**.

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**.

Kireeva, M. L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V. M. (2002). Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription. *Mol. Cell* **9**:541–552.

Kistner, C., and Parniske, M. (2002). Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* **7**:511–518.

Kleinmanns, J. A., and Schubert, D. (2014). Polycomb and Trithorax group protein-mediated control of stress responses in plants. *Biol. Chem.* **395**:1291–1300.

Klose, R. J., Kallin, E. M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7**:715–727.

Komiya, R., Ohyanagi, H., Niihama, M., Watanabe, T., Nakano, M., Kurata, N., and Nonomura, K. I. (2014). Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. *Plant J.* **78**:385–397.

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.

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **39**:309–338.

Kopp, F., and Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**:393–407.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**.

Ku, S. J., Park, J. Y., Ha, S. B., and Kim, J. (2009). Overexpression of IAA1 with domain II mutation impairs cell elongation and cell division in inflorescences and leaves of Arabidopsis. *J. Plant Physiol.* **166**:548–553.

Kumagai, H., Kinoshita, E., Ridge, R. W., and Kouchi, H. (2006). RNAi knock-down of ENOD40s leads to significant suppression of nodule formation in Lotus japonicus. *Plant Cell Physiol.* **47**:1102–1111.

Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**:91–95.

Kwon, C. S., Lee, D., Choi, G., and Chung, W. Il (2009). Histone occupancy-dependent and -independent removal of H3K27 trimethylation at cold-responsive genes in Arabidopsis. *Plant J.* **60**:112–121.

Kyrchanova, O., and Georgiev, P. (2021). Mechanisms of enhancer-promoter interactions in higher eukaryotes. *Int. J. Mol. Sci.* **22**:1–20.

Lafos, M., Kroll, P., Hohenstatt, M. L., Thorpe, F. L., Clarenz, O., and Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during arabidopsis differentiation. *PLoS Genet.* **7**.

Lai, Y. S., Zhang, X., Zhang, W., Shen, D., Wang, H., Xia, Y., Qiu, Y., Song, J., Wang, C., and Li, X. (2017). The association of changes in DNA methylation with temperature-dependent sex determination in cucumber. *J. Exp. Bot.* **68**:2899–2912.

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**.

Law, J. A., Du, J., Hale, C. J., Feng, S., Krajewski, K., Palanca, A. M. S., Strahl, B. D., Patel, D. J., and Jacobsen, S. E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**:385–389.

Lawrence, J. G. (2002). Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell* **110**:407–413.

Lee, J. E., Neumann, M., Duro, D. I., and Schmid, M. (2019). CRISPR-based tools for targeted transcriptional and epigenetic regulation in plants. *PLoS One* **14**:1–17.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**:233–245.

Lewis, D. R., Olex, A. L., Lundy, S. R., Turkett, W. H., Fetrow, J. S., and Muday, G. K. (2013). A kinetic analysis of the auxin transcriptome reveals cell wall remodeling proteins that modulate lateral root development in Arabidopsis. *Plant Cell* **25**:3329–3346.

Li, J., and Liu, C. (2019). Coding or noncoding, the converging concepts of RNAs. *Front. Genet.* **10**:1–10.

Li, L., and Wurtele, E. S. (2015). The QQS orphan gene of Arabidopsis modulates carbon and nitrogen allocation in soybean. *Plant Biotechnol. J.* **13**:177–187.

Li, L., Foster, C. M., Gan, Q., Nettleton, D., James, M. G., Myers, A. M., and Wurtele, E. S. (2009). Identification of the novel protein QQS as a component of the starch metabolic network in Arabidopsis leaves. *Plant J.* **58**:485–498.

Li, S., Liu, J., Liu, Z., Li, X., Wu, F., and He, Y. (2014a). HEAT-INDUCED TAS1 TARGET1 mediates thermotolerance via heat stress transcription factor A1a-directed pathways in arabidopsis. *Plant Cell* **26**:1764–1780.

Li, A., Zhang, J., and Zhou, Z. (2014b). PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* **15**:1–10.

Li, Q., Gent, J. I., Zynda, G., Song, J., Makarevitch, I., Hirsch, C. D., Hirsch, C. N., Dawe, R. K., Madzima, T. F., McGinnis, K. M., et al. (2015a). RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proc. Natl. Acad. Sci. U. S. A.* **112**:14728–14733.

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

Li, S., Le, B., Ma, X., Li, S., You, C., Yu, Y., Zhang, B., Liu, L., Gao, L., Shi, T., et al. (2016a). Biogenesis of phased siRNAs on membrane-bound polysomes in Arabidopsis. *Elife* **5**:1–24.

Li, J., Huang, Q., Sun, M., Zhang, T., Li, H., Chen, B., Xu, K., Gao, G., Li, F., Yan, G., et al. (2016b). Global DNA methylation variations after short-term heat shock treatment in cultured microspores of Brassica napus cv. Topas. *Sci. Rep.* **6**:1–13.

Li, S., Yu, X., Lei, N., Cheng, Z., Zhao, P., He, Y., Wang, W., and Peng, M. (2017a). Genome-wide identification and functional prediction of cold and/or drought-responsive lncRNAs in cassava. *Sci. Rep.* **7**.

Li, Y., Wang, Y., Xue, H., Pritchard, H. W., and Wang, X. (2017b). Changes in the mitochondrial protein profile due to ROS eruption during ageing of elm (Ulmus pumila L.) seeds. *Plant Physiol. Biochem.* **114**:72–87.

Li, Y., Shi, W., and Wasserman, W. W. (2018a). Genome-wide prediction of cis-regulatory

regions using supervised deep learning methods. *BMC Bioinformatics* **19**:1–14.

Li, X., Xing, X., Xu, S., Zhang, M., Wang, Y., Wu, H., Sun, Z., Huo, Z., Chen, F., and Yang, T. (2018b). Genome-wide identification and functional prediction of tobacco lncRNAs responsive to root-knot nematode stress. *PLoS One* **13**.

Li, J., Zhang, X., and Liu, C. (2020). The computational approaches of lncRNA identification based on coding potential: Status quo and challenges. *Comput. Struct. Biotechnol. J.* **18**:3666–3677.

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.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (80-.).* **326**:289–293.

Lin, M. F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**:275–282.

Lin, J. S., Lin, C. C., Lin, H. H., Chen, Y. C., and Jeng, S. T. (2012). MicroR828 regulates lignin and h2O2 accumulation in sweet potato on wounding. *New Phytol.* **196**:427–440.

Lin, Q., Ohashi, Y., Kato, M., Tsuge, T., Gu, H., Qu, L. J., and Aoyama, T. (2015). GLABRA2 directly suppresses basic helix-loop-helix transcription factor genes with diverse functions in root hair development. *Plant Cell* **27**:2894–2906.

Lippman, Z., Gendrel, A. V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., Lavine, K., Mittal, V., May, B., Kasschau, K. B., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**:471–476.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., and Ecker, J. R. (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* **133**:523–536.

Liu, C., and Weigel, D. (2015). Chromatin in 3D: Progress and prospects for plants. *Genome Biol.* **16**:1–6.

Liu, C., Lu, F., Cui, X., and Cao, X. (2010a). Histone methylation in higher plants. *Annu. Rev. Plant Biol.* **61**:395–420.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010b). Targeted 3′Processing of AntisenseTranscripts TriggersArabidopsis FLC Chromatin Silencing. *Science (80-.).* **327**:94–97.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010c). Targeted 3′Processing of AntisenseTranscripts TriggersArabidopsis FLC Chromatin Silencing. *Science (80-.).* **327**:94–97.

Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C., and Chua, N.-H. (2012a). Genome-Wide Analysis Uncovers Regulation of Long Intergenic Noncoding RNAs in Arabidopsis. *Plant Cell* **24**:4333–4345.

Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C., and Chua, N. H. (2012b). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in arabidopsisC W. *Plant Cell* **24**:4333–4345.

Liu, C., Teo, Z. W. N., Bi, Y., Song, S., Xi, W., Yang, X., Yin, Z., and Yu, H. (2013). A Conserved Genetic Pathway Determines Inflorescence Architecture in Arabidopsis and Rice. *Dev. Cell* **24**:612–622.

Liu, J., Cheng, X., Liu, D., Xu, W., Wise, R., and Shen, Q. H. (2014a). The miR9863 Family

Regulates Distinct Mla Alleles in Barley to Attenuate NLR Receptor-Triggered Disease Resistance and Cell-Death Signaling. *PLoS Genet.* **10**.

Liu, N., Fromm, M., and Avramova, Z. (2014b). H3K27me3 and H3K4me3 chromatin environment at super-induced dehydration stress memory genes of arabidopsis thaliana. *Mol. Plant* **7**:502–513.

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.

Liu, Y., Ke, L., Wu, G., Xu, Y., Wu, X., Xia, R., Deng, X., and Xu, Q. (2017a). miR3954 is a trigger of phasiRNAs that affects flowering time in citrus. *Plant J.* **92**:263–275.

Liu, H., Luo, X., Niu, L., Xiao, Y., Chen, L., Liu, J., Wang, X., Jin, M., Li, W., Zhang, Q., et al. (2017b). Distant eQTLs and Non-coding Sequences Play Critical Roles in Regulating Gene Expression and Quantitative Trait Variation in Maize. *Mol. Plant* **10**:414–426.

Liu, F., Xu, Y., Chang, K., Li, S., Liu, Z., Qi, S., Jia, J., Zhang, M., Crawford, N. M., and Wang, Y. (2019). The long noncoding RNA T5120 regulates nitrate response and assimilation in Arabidopsis. *New Phytol.* **224**:117–131.

Long, H. K., Prescott, S. L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**:1170–1187.

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.

Lowder, L. G., Zhou, J., Zhang, Y., Malzahn, A., Zhong, Z., Hsieh, T. F., Voytas, D. F., Zhang, Y., and Qi, Y. (2018). Robust Transcriptional Activation in Plants Using Multiplexed CRISPR-Act2.0 and mTALE-Act Systems. *Mol. Plant* **11**:245–256.

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

Lv, Y., Liang, Z., Ge, M., Qi, W., Zhang, T., Lin, F., Peng, Z., and Zhao, H. (2016). Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (Zea mays L.). *BMC Genomics* **17**:1–15.

Lynch, V. J., May, G., and Wagner, G. P. (2011). Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* **480**:383–386.

Ma, Y., Min, L., Wang, M., Wang, C., Zhao, Y., Li, Y., Fang, Q., Wu, Y., Xie, S., Ding, Y., et al. (2018). Disrupted genome methylation in response to high temperature has distinct affects on microspore abortion and anther indehiscence. *Plant Cell* **30**:1387–1403.

Ma, J., Bai, X., Luo, W., Feng, Y., Shao, X., Bai, Q., Sun, S., Long, Q., and Wan, D. (2019). Genome-Wide Identification of Long Noncoding RNAs and Their Responses to Salt Stress in Two Closely Related Poplars. *Front. Genet.* **10**:1–13.

Ma, P., Zhang, X., Luo, B., Chen, Z., He, X., Zhang, H., Li, B., Liu, D., and Wu, L. (2021). Transcriptomic and genome-wide association study reveal long noncoding RNAs responding to nitrogen deficiency in maize Advance Access published 2021.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: Directing transcriptional regulators within the chromatin environment. *Trends Genet.* **27**:465–474.

Marin, E., Jouannet, V., Herz, A., Lokerse, A. S., Weijers, D., Vaucheret, H., Nussaume, L., Crespi, M. D., and Maizel, A. (2010). mir390, Arabidopsis TAS3 tasiRNAs, and their AUXIN

RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. *Plant Cell* **22**:1104–1117.

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.

Matias-Hernandez, L., Battaglia, R., Galbiati, F., Rubes, M., Eichenberger, C., Grossniklaus, U., Kater, M. M., and Colombo, L. (2010). VERDANDI is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. *Plant Cell* **22**:1702–1715.

Mattioli, K., Volders, P. J., Gerhardinger, C., Lee, J. C., Maass, P. G., Melé, M., and Rinn, J. L. (2019). High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. *Genome Res.* **29**:344–355.

Matzke, M. A., and Mosher, R. A. (2014). RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**:394–408.

Matzke, M. A., Kanno, T., and Matzke, A. J. M. (2015). RNA-directed DNA methylation: The evolution of a complex epigenetic pathway in flowering plants. *Annu. Rev. Plant Biol.* **66**:243–267.

McCue, A. D., Nuthikattu, S., Reeder, S. H., and Slotkin, R. K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet.* **8**.

McCue, A. D., Panda, K., Nuthikattu, S., Choudury, S. G., Thomas, E. N., and Slotkin, R. K. (2015). ARGONAUTE 6 bridges transposable element m RNA -derived si RNA s to the establishment of DNA methylation . *EMBO J.* **34**:20–35.

Mehdi, S., Derkacheva, M., Ramström, M., Kralemann, L., Bergquist, J., and Hennig, L. (2016). The WD40 domain protein MSI1 functions in a histone deacetylase complex to fine-tune abscisic acid signaling. *Plant Cell* **28**:42–54.

Melé, M., Mattioli, K., Mallard, W., Shechner, D. M., Gerhardinger, C., and Rinn, J. L. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* **27**:27–37.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**:333–338.

Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finzi, L., Kater, M. M., and Colombo, L. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in Arabidopsis. *Plant Cell* **25**:2560–2572.

Merchante, C., Stepanova, A. N., and Alonso, J. M. (2017). Translation regulation in plants: an interesting past, an exciting present and a promising future. *Plant J.* **90**:628–653.

Michalak, P. (2008). Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* **91**:243–248.

Miele, A., and Dekker, J. (2008). Long-range chromosomal interactions and gene regulation. *Mol. Biosyst.* **4**:1046–1058.

Millar, A. A., and Waterhouse, P. M. (2005). Plant and animal microRNAs: Similarities and differences. *Funct. Integr. Genomics* **5**:129–135.

Mitsis, T., Efthimiadou, A., Bacopoulou, F., Vlachakis, D., Chrousos, G. P., and Eliopoulos, E. (2020). Transcription factors and evolution: An integral part of gene expression (Review). *World Acad. Sci. J.* **2**:3–8.

Mizuguchi, T., Barrowman, J., and Grewal, S. I. S. (2015). Chromosome domain architecture and dynamic organization of the fission yeast genome. *FEBS Lett.* **589**:2975–2986.

Mizzotti, C., Rotasperti, L., Moretto, M., Tadini, L., Resentini, F., Galliani, B. M., Galbiati, M., Engelen, K., Pesaresi, P., and Masiero, S. (2018). Time-course transcriptome analysis of arabidopsis siliques discloses genes essential for fruit development and maturation. *Plant Physiol.* **178**:1249–1268.

Mohammadin, S., Edger, P. P., Pires, J. C., and Schranz, M. E. (2015). Positionally-conserved but sequence-diverged: Identification of long non-coding RNAs in the Brassicaceae and Cleomaceae. *BMC Plant Biol.* **15**:1–12.

Moison, M., Pacheco, J. M., Lucero, L., Fonouni-Farde, C., Rodríguez-Melo, J., Mansilla, N., Christ, A., Bazin, J., Benhamed, M., Ibañez, F., et al. (2021). The lncRNA APOLO interacts with the transcription factor WRKY42 to trigger root hair cell expansion in response to cold. *Mol. Plant* **14**:937–948.

Moissiard, G., Cokus, S. J., Cary, J., Feng, S., Billi, A. C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B. R., McCord, R. P., et al. (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. *Science (80-.).* **336**:1448–1451.

Mok, M. C., Martin, R. C., and Mok, D. W. S. (2000). Cytokinins: Biosynthesis, metabolism and perception. *Vitr. Cell. Dev. Biol. - Plant* **36**:102–107.

Moradpour, M., and Abdulah, S. N. A. (2020). CRISPR/dCas9 platforms in plants: strategies and applications beyond genome editing. *Plant Biotechnol. J.* **18**:32–44.

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.

Morgunova, E., and Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* **47**:1–8.

Mukherjee, N., Calviello, L., Hirsekorn, A., De Pretis, S., Pelizzola, M., and Ohler, U. (2017a). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24**:86–96.

Mukherjee, N., Calviello, L., Hirsekorn, A., De Pretis, S., Pelizzola, M., and Ohler, U. (2017b). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24**:86–96.

Murcia, G., Fontana, A., Pontin, M., Baraldi, R., Bertazza, G., and Piccoli, P. N. (2017). ABA and GA3 regulate the synthesis of primary and secondary metabolites related to alleviation from biotic and abiotic stresses in grapevine. *Phytochemistry* **135**:34–52.

Mustafa, G. M., Larry, D., Petersen, J. R., and Elferink, C. J. (2015). Targeted proteomics for biomarker discovery and validation of hepatocellular carcinoma in hepatitis C infected patients. *World J. Hepatol.* **7**:1312–1324.

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.

Nakashima, K., Fujita, Y., Kanamori, N., Katagiri, T., Umezawa, T., Kidokoro, S., Maruyama, K., Yoshida, T., Ishiyama, K., Kobayashi, M., et al. (2009). Three arabidopsis SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy. *Plant Cell Physiol.* **50**:1345–1363.

Nam, J. W., Choi, S. W., and You, B. H. (2016). Incredible RNA: Dual functions of coding and noncoding. *Mol. Cells* **39**:367–374.

Nambara, E., Kawaide, H., Kamiya, Y., and Naito, S. (1998). Characterization of an Arabidopsis thaliana mutant that has a defect in ABA accumulation: ABA-dependent and ABA-independent accumulation of free amino acids during dehydration. *Plant Cell Physiol.* **39**:853–858.

Naydenov, M., Baev, V., Apostolova, E., Gospodinova, N., Sablok, G., Gozmanova, M., and Yahubyan, G. (2015). High-temperature effect on genes engaged in DNA methylation and affected by DNA methylation in Arabidopsis. *Plant Physiol. Biochem.* **87**:102–108.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J. C., Grützner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**:635–640.

Nelson, A. D. L., Forsythe, E. S., Devisetty, U. K., Clausen, D. S., Haug-Batzell, A. K., Meldrum, A. M. R., Frank, M. R., Lyons, E., and Beilstein, M. A. (2016). A genomic analysis of factors driving lincRNA diversification: Lessons from plants. *G3 Genes, Genomes, Genet.* **6**:2881–2891.

Nelson, S. K., Ariizumi, T., and Steber, C. M. (2017). Biology in the dry seed: Transcriptome changes associated with dry seed dormancy and dormancy loss in the arabidopsis GA-insensitive sleepy1-2 mutant. *Front. Plant Sci.* **8**:1–21.

Noguchi, J., and Fukui, K. (1995). Chromatin arrangements in intact interphase nuclei examined by laser confocal microscopy. *J. Plant Res.* **108**:209–216.

Nonomura, K. I., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., and Kurataa, N. (2007). A germ cell-specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell* **19**:2583–2594.

Nordström, A., Tarkowski, P., Tarkowska, D., Norbaek, R., Åstot, C., Dolezal, K., and Sandberg, G. (2004). Auxin regulation of cytokinin biosynthesis in Arabidopsis thaliana: A factor of potential importance for auxin-cytokinin-regulated development. *Proc. Natl. Acad. Sci. U. S. A.* **101**:8039–8044.

Nuthikattu, S., McCue, A. D., Panda, K., Fultz, D., DeFraia, C., Thomas, E. N., and Keith Slotkin, R. (2013). The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol.* **162**:116–131.

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

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.

O'Sullivan, J. M., Tan-Wong, S. M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N. J. (2004). Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* **36**:1014–1018.

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.

Ohta, M., Matsui, K., Hiratsu, K., Shinshi, H., and Ohme-Takagi, M. (2001). Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell* **13**:1959–1968.

Oka, R., Zicola, J., Weber, B., Anderson, S. N., Hodgman, C., Gent, J. I., Wesselink, J. J., Springer, N. M., Hoefsloot, H. C. J., Turck, F., et al. (2017). Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* **18**:1–24.

Oliver, C., Pradillo, M., Jover-Gil, S., Cuñado, N., Ponce, M. R., and Santos, J. L. (2017). Loss of function of Arabidopsis microRNA-machinery genes impairs fertility, and has effects on homologous recombination and meiotic chromatin dynamics. *Sci. Rep.* **7**:1–14.

Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O., and Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation- dependent heterochromatin formation. *Cell* **120**:613–622.

Pant, B. D., Buhtz, A., Kehr, J., and Scheible, W. R. (2008). MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *Plant J.* **53**:731–738.

Petroni, K., Kumimoto, R. W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B. F., and Mantovani, R. (2013). The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. *Plant Cell* **24**:4777–4792.

Poirier, Y., and Bucher, M. (2002). Phosphate Transport and Homeostasis in Arabidopsis. *Arab. B.* **1**:e0024.

Ponjavic, J., Ponting, C. P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**:556–565.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M. A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev.* **19**:2030–2040.

Popova, O. V., Dinh, H. Q., Aufsatz, W., and Jonak, C. (2013). The RdDM pathway is required for basal heat tolerance in arabidopsis. *Mol. Plant* **6**:396–410.

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.

Prakash, K., and Fournier, D. (2018). Evidence for the implication of the histone code in building the genome structure. *BioSystems* **164**:49–59.

Pregitzer, K. S., King, J. S., Burton, A. J., and Brown, S. E. (2000). Responses of tree fine roots to temperature. *New Phytol.* **147**:105–115.

Pu, L., and Sung, Z. R. (2015). PcG and trxG in plants - friends or foes. *Trends Genet.* **31**:252–262.

Purdy, S. J., Bussell, J. D., Nelson, D. C., Villadsen, D., and Smith, S. M. (2011). A nuclear-localized protein, KOLD SENSITIV-1, affects the expression of cold-responsive genes during prolonged chilling in Arabidopsis. *J. Plant Physiol.* **168**:263–269.

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**.

Qi, X., Xie, S., Liu, Y., Yi, F., and Yu, J. (2013). Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. *Plant Mol. Biol.* **83**:459–473.

Qin, F., Sakuma, Y., Tran, L. S. P., Maruyama, K., Kidokoro, S., Fujita, Y., Fujita, M., Umezawa, T., Sawano, Y., Miyazono, K. I., et al. (2008). Arabidopsis DREB2A-interacting proteins function as Ring E3 ligases and negatively regulate plant drought stress-responsive gene expression. *Plant Cell* **20**:1693–1707.

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.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.

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.

Reinders, J., Wulff, B. B. H., Mirouze, M., Marí-Ordóñez, A., Dapp, M., Rozhon, W., Bucher, E., Theiler, G., and Paszkowski, J. (2009). Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes Dev.* **23**:939–950.

Reymond, M., Svistoonoff, S., Loudet, O., Nussaume, L., and Desnos, T. (2006). Identification of QTL controlling root growth response to phosphate starvation in *Arabidopsis thaliana*. *Plant. Cell Environ.* **29**:115–125.

Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., Noshay, J. M., Galli, M., Mejía-Guerra, M. K., Colomé-Tatché, M., et al. (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **5**:1237–1249.

Rigo, R., Bazin, J., Romero-Barrios, N., Moison, M., Lucero, L., Christ, A., Benhamed, M., Blein, T., Huguet, S., Charon, C., et al. (2020). The Arabidopsis lnc RNA ASCO modulates the transcriptome through interaction with splicing factors . *EMBO Rep.* **21**:1–19.

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

Rio, D. C. (2014). Electrophoretic mobility shift assays for RNA-protein complexes. *Cold Spring Harb. Protoc.* **2014**:435–440.

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.

Rosa, S., and Shaw, P. (2013). Insights into chromatin structure and dynamics in plants. *Biology (Basel).* **2**:1378–1410.

Roy, S., Siahpirani, A. F., Chasman, D., Knaack, S., Ay, F., Stewart, R., Wilson, M., and

Sridharan, R. (2015). A predictive modeling approach for cell line-specific long-range regulatory interactions. *Nucleic Acids Res.* **43**:8694–8712.

Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K. A., Meeley, R., Ananiev, E. V., Svitashev, S., Bruggemann, E., et al. (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci. U. S. A.* **104**:11376–11381.

Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., and Brown, P. O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**.

Sammeth, M., Foissac, S., and Guigó, R. (2008). A general definition and nomenclature for alternative splicing events. *PLoS Comput. Biol.* **4**.

Sanchita, Trivedi, P. K., and Asif, M. H. (2020). Updates on plant long non-coding RNAs (lncRNAs): the regulatory components. *Plant Cell. Tissue Organ Cult.* **140**:259–269.

Santer, L., Bär, C., and Thum, T. (2019). Circular RNAs: A Novel Class of Functional RNA Molecules with a Therapeutic Perspective. *Mol. Ther.* **27**:1350–1363.

Sarropoulos, I., Marin, R., Cardoso-Moreira, M., and Kaessmann, H. (2019). Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* **571**:510–514.

Sartorelli, V., and Lauberth, S. M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. *Nat. Struct. Mol. Biol.* **27**:521–528.

Sato, H., Suzuki, T., Takahashi, F., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2019). NF-YB2 and NF-YB3 have functionally diverged and differentially induce drought and heat stress-specific genes. *Plant Physiol.* **180**:1677–1690.

Saze, H., Scheid, O. M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* **34**:65–69.

Schatlowski, N., Stahl, Y., Hohenstatt, M. L., Goodrich, J., and Schuberta, D. (2010). The CURLY LEAF interacting protein BLISTER controls expression of polycomb-group target genes and cellular differentiation of Arabidopsis thaliana. *Plant Cell* **22**:2291–2305.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N. J. (2017a). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol. Cell* **65**:25–38.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N. J. (2017b). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol. Cell* **65**:25–38.

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J.* **25**:4638–4649.

Schubert, V., Berr, A., and Meister, A. (2012). Interphase chromatin organisation in Arabidopsis nuclei: Constraints versus randomness. *Chromosoma* **121**:369–387.

Schütze, K., Harter, K., and Chaban, C. (2008). Post-translational regulation of plant bZIP factors. *Trends Plant Sci.* **13**:247–255.

Seo, J. S., and Chua, N.-H. (2017). Trimolecular Fluorescence Complementation (TriFC) Assay for Direct Visualization of RNA-Protein Interaction in planta. *Bio-Protocol* **7**:1–6.

Seo, J. S., Sun, H. X., Park, B. S., Huang, C. H., Yeh, S. D., Jung, C., and Chua, N. H. (2017). ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in arabidopsis. *Plant Cell* **29**:1024–1038.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W., and Kingston,

R. E. (1999). Stabilization of chromatin structure by PRC1, a polycomb complex. *Cell* **98**:37–46.

Shao, Y., Li, J., Lu, R., Li, T., Yang, Y., Xiao, B., and Guo, J. (2017). Global circular RNA expression profile of human gastric cancer and its clinical significance. *Cancer Med.* **6**:1173–1180.

Shatkin, A. J. (1976). Capping of eucaryotic mRNAs. *Cell* **9**:645–653.

Shen, Q., Lin, Y., Li, Y., and Wang, G. (2021). Dynamics of h3k27me3 modification on plant adaptation to environmental cues. *Plants* **10**.

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.

Shin, H., Shin, H. S., Chen, R., and Harrison, M. J. (2006). Loss of At4 function impacts phosphate distribution between the roots and the shoots during phosphate starvation. *Plant J.* **45**:712–726.

Shin, H. Y., Willi, M., Yoo, K. H., Zeng, X., Wang, C., Metser, G., and Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* **48**:904–911.

Shin, S. Y., Jeong, J. S., Lim, J. Y., Kim, T., Park, J. H., Kim, J. K., and Shin, C. (2018). Transcriptomic analyses of rice (Oryza sativa) genes and non-coding RNAs under nitrogen starvation using multiple omics technologies. *BMC Genomics* **19**:1–20.

Shuai, P., Liang, D., Tang, S., Zhang, Z., Ye, C. Y., Su, Y., Xia, X., and Yin, W. (2014). Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in Populus trichocarpa. *J. Exp. Bot.* **65**:4975–4983.

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.

Simmonds, M. J., Kavvoura, F. K., Brand, O. J., Newby, P. R., Jackson, L. E., Hargreaves, C. E., Franklyn, J. A., and Gough, S. C. L. (2014). Skewed X chromosome inactivation and female preponderance in autoimmune thyroid disease: An association study and meta-analysis. *J. Clin. Endocrinol. Metab.* **99**:127–131.

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.

Singh, B. N., and Hampsey, M. (2007). A Transcription-Independent Role for TFIIB in Gene Looping. *Mol. Cell* **27**:806–816.

Singh, M., Goel, S., Meeley, R. B., Dantec, C., Parrinello, H., Michaud, C., Leblanc, O., and Grimanelli, D. (2011). Production of viable gametes without meiosis in maize deficient for an ARGONAUTE protein. *Plant Cell* **23**:443–458.

Smith, L. M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F. K., Herr, A. J., Pikaard, C. S., and Baulcombe, D. C. (2007). An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell* **19**:1507–1521.

Solovei, I., Cavallo, A., Schermelleh, L., Jaunin, F., Scasselati, C., Cmarko, D., Cremer, C., Fakan, S., and Cremer, T. (2002). Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH). *Exp. Cell Res.* **276**:10–23.

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.

Song, Y., Xuan, A., Bu, C., Ci, D., Tian, M., and Zhang, D. (2019). Osmotic stress-responsive promoter upstream transcripts (PROMPTs) act as carriers of MYB transcription factors to induce the expression of target genes in Populus simonii. *Plant Biotechnol. J.* **17**:164–177.

Sorin, C., Declerck, M., Christ, A., Blein, T., Ma, L., Lelandais-Brière, C., Njo, M. F., Beeckman, T., Crespi, M., and Hartmann, C. (2014). A miR169 isoform regulates specific NF-YA targets and root architecture in Arabidopsis. *New Phytol.* **202**:1197–1211.

Sosa-Valencia, G., Palomar, M., Covarrubias, A. A., and Reyes, J. L. (2017). The legume miR1514a modulates a NAC transcription factor transcript to trigger phasiRNA formation in response to drought. *J. Exp. Bot.* **68**:2013–2026.

Stagnati, L., Lanubile, A., Samayoa, L. F., Bragalanti, M., Giorni, P., Busconi, M., Holland, J. B., and Marocco, A. (2019). A genome wide association study reveals markers and genes associated with resistance to fusarium verticillioides infection of seedlings in a maize diversity panel. *G3 Genes, Genomes, Genet.* **9**:571–579.

Statello, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**:96–118.

Stewart, A. J., and Plotkin, J. B. (2012). Why transcription factor binding sites are ten nucleotides long. *Genetics* **192**:973–985.

Stief, A., Altmann, S., Hoffmann, K., Pant, B. D., Scheible, W. R., and Bäurle, I. (2014). Arabidopsis miR156 regulates tolerance to recurring environmental stress through SPL transcription factors. *Plant Cell* **26**:1792–1807.

Streit, D., and Schleiff, E. (2021). The Arabidopsis 2′-O-Ribose-Methylation and Pseudouridylation Landscape of rRNA in Comparison to Human and Yeast. *Front. Plant Sci.* **12**:1–19.

Strichman-Almashanu, L. Z., Lee, R. S., Onyango, P. O., Perlman, E., Flam, F., Frieman, M. B., and Feinberg, A. P. (2002). A Genome-Wide Screen for Normally Methylated Human CpG Islands That Can Identify Novel Imprinted Genes. *Genome Res.* **12**:543–554.

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* **43**:1160–1163.

Su, M., Xiao, Y., Ma, J., Tang, Y., Tian, B., Zhang, Y., Li, X., Wu, Z., Yang, D., Zhou, Y., et al. (2019). Circular RNAs in Cancer: Emerging functions in hallmarks, stemness, resistance and roles as potential biomarkers. *Mol. Cancer* **18**:1–17.

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M. H. Y., et al. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature* **526**:75–81.

Sukumar, P., Edwards, K. S., Rahman, A., DeLong, A., and Muday, G. K. (2009). PINOID kinase regulates root gravitropism through modulation of PIN2-dependent basipetal auxin transport in arabidopsis. *Plant Physiol.* **150**:722–735.

Sullivan, A. E., Lewis, T., Stephenson, M., Odem, R., Schreiber, J., Ober, C., and Branch, D. W. (2003). Pregnancy outcome in recurrent miscarriage patients with skewed X chromosome inactivation. *Obstet. Gynecol.* **101**:1236–1242.

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., Liu, Y., Chen, R., and Zhao, Y. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **41**.

Sun, Z., Li, M., Zhou, Y., Guo, T., Liu, Y., Zhang, H., and Fang, Y. (2018). Coordinated

regulation of Arabidopsis microRNA biogenesis and red light signaling through Dicer-like 1 and phytochrome-interacting factor 4. *PLoS Genet.* **14**:1–21.

Sun, J., He, N., Niu, L., Huang, Y., Shen, W., Zhang, Y., Li, L., and Hou, C. (2019). Global Quantitative Mapping of Enhancers in Rice by STARR-seq. *Genomics, Proteomics Bioinforma.* **17**:140–153.

Sun, Z., Huang, K., Han, Z., Wang, P., and Fang, Y. (2020a). Genome-wide identification of Arabidopsis long noncoding RNAs in response to the blue light. *Sci. Rep.* **10**:1–10.

Sun, X., Zheng, H., Li, J., Liu, L., Zhang, X., and Sui, N. (2020b). Comparative Transcriptome Analysis Reveals New lncRNAs Responding to Salt Stress in Sweet Sorghum. *Front. Bioeng. Biotechnol.* **8**:1–14.

Suzuki, M., Wang, H. H. Y., and McCarty, D. R. (2007). Repression of the Leafy Cotyledon 1/B3 regulatory network in plant embryo development by VP1/Abscisic Acid Insensitive 3-Like B3 genes. *Plant Physiol.* **143**:902–911.

Swiezewski, S., Crevillen, P., Liu, F., Ecker, J. R., Jerzmanowski, A., and Dean, C. (2007). Small RNA-mediated chromatin silencing directed to the 3′ region of the Arabidopsis gene encoding the developmental regulator, FLC. *Proc. Natl. Acad. Sci. U. S. A.* **104**:3633–3638.

Talebizadeh, Z., Bittel, D. C., Veatch, O. J., Kibiryeva, N., and Butler, M. G. (2005). Brief report: Non-Random X chromosome inactivation in females with autism. *J Autism Dev Disord* **35**:675–681.

Tatchell, K., and Van Holde, K. E. (1977). Reconstitution of chromatin core particles. *Biochemistry* **16**:5295–5303.

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., et al. (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**:D941–D947.

Thomas, C. A. (1971). The genetic organization of chromosomes. *Annu. Rev. Genet.* **5**:237–256.

Tian, Y., Xing, Y., Zhang, Z., Peng, R., Zhang, L., and Sun, Y. (2020). Bioinformatics Analysis of Key Genes and circRNA-miRNA-mRNA Regulatory Network in Gastric Cancer. *Biomed Res. Int.* **2020**.

Tiang, C. L., He, Y., and Pawlowski, W. P. (2012). Chromosome organization and Dynamics during interphase, mitosis, and meiosis in plants. *Plant Physiol.* **158**:26–34.

Tiwari, S. B., Hagen, G., and Guilfoyle, T. J. (2004). Aux/IAA Proteins Contain a Potent Transcriptional Repression Domain. *Plant Cell* **16**:533–543.

Tolhuis, B., Muijrers, I., De Wit, E., Teunissen, H., Talhout, W., Van Steensel, B., and Van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. *Nat. Genet.* **38**:694–699.

Traubenik, S., Reynoso, M. A., Hobecker, K., Lancia, M., Hummel, M., Rosen, B., Town, C., Bailey-Serres, J., Blanco, F., and Zanetti, M. E. (2020). Reprogramming of root cells during nitrogen-fixing symbiosis involves dynamic polysome association of coding and noncoding RNAs. *Plant Cell* **32**:352–373.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M. L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R. A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet.* **3**:0855–0866.

Ueda, M., and Seki, M. (2020). Histone modifications form epigenetic regulatory networks to regulate abiotic stress response1[OPEN]. *Plant Physiol.* **182**:15–26.

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., and Bartel, D. P. (2011). Conserved function

of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**:1537–1550.

Valdés-López, O., Batek, J., Gomez-Hernandez, N., Nguyen, C. T., Isidra-Arellano, M. C., Zhang, N., Joshi, T., Xu, D., Hixson, K. K., Weitz, K. K., et al. (2016). Soybean roots grown under heat stress show global changes in their transcriptional and proteomic profiles. *Front. Plant Sci.* **7**:1–12.

Van Driel, R., and Fransz, P. (2004). Nuclear architecture and genome functioning in plants and animals: What can we learn from both? *Exp. Cell Res.* **296**:86–90.

Vanneste, S., De Rybel, B., Beemster, G. T. S., Ljung, K., De Smet, I., Van Isterdael, G., Naudts, M., Iida, R., Gruissem, W., Tasaka, M., et al. (2005). Cell cycle progression in the pericycle is not sufficient for SOLITARY ROOT/IAA14-mediated lateral root initiation in Arabidopsis thaliana. *Plant Cell* **17**:3035–3050.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., Hilbert, J. L., Bartel, D. P., and Crété, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of arabidopsis mRNAs. *Mol. Cell* **16**:69–79.

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.

Venters, B. J., and Pugh, B. F. (2009). How eukaryotic genes are transcribed regulation of eukaroytic gene transcription B. J. Venters and B. F. Pugh. *Crit. Rev. Biochem. Mol. Biol.* **44**:117–141.

Verheggen, K., Volders, P. J., Mestdagh, P., Menschaert, G., Van Damme, P., Gevaert, K., Martens, L., and Vandesompele, J. (2017). Noncoding after All: Biases in Proteomics Data Do Not Explain Observed Absence of lncRNA Translation Products. *J. Proteome Res.* **16**:2508–2515.

Vernikos, G., Medini, D., Riley, D. R., and Tettelin, H. (2015). Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* **23**:148–154.

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.

Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell* **136**:669–687.

Wahl, M. C., Will, C. L., and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136**:701–718.

Walch-Liu, P., Ivanov, I. I., Filleur, S., Gan, Y., Remans, T., and Forde, B. G. (2006). Nitrogen regulation of root branching. *Ann. Bot.* **97**:875–881.

Wan, S., Zhang, Y., Duan, M., Huang, L., Wang, W., Xu, Q., Yang, Y., and Yu, Y. (2020). Integrated Analysis of Long Non-coding RNAs (lncRNAs) and mRNAs Reveals the Regulatory Role of lncRNAs Associated With Salt Resistance in Camellia sinensis. *Front. Plant Sci.* **11**:1–14.

Wang, J. W. (2014). Regulation of flowering time by the miR156-mediated age pathway. *J. Exp. Bot.* **65**:4723–4730.

Wang, Y. P., and Lei, Q. Y. (2018). Metabolic recoding of epigenetics in cancer. *Cancer Commun. (London, England)* **38**:25.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F.,

Schroth, G. P., and Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**:470–476.

Wang, Y., Fan, X., Lin, F., He, G., Terzaghi, W., Zhu, D., and Deng, X. W. (2014). Arabidopsis noncoding RNA mediates control of photomorphogenesis by red light. *Proc. Natl. Acad. Sci. U. S. A.* **111**:10359–10364.

Wang, T. Z., Liu, M., Zhao, M. G., Chen, R., and Zhang, W. H. (2015). Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in Medicago truncatula using genome-wide high-throughput sequencing. *BMC Plant Biol.* **15**:1–13.

Wang, T., Zhao, M., Zhang, X., Liu, M., Yang, C., Chen, Y., Chen, R., Wen, J., Mysore, K. S., and Zhang, W. H. (2017a). Novel phosphate deficiency-responsive long non-coding RNAs in the legume model plant Medicago truncatula. *J. Exp. Bot.* **68**:5937–5948.

Wang, Y., Yang, M., Wei, S., Qin, F., Zhao, H., and Suo, B. (2017b). Identification of circular RNAs and their targets in leaves of Triticum aestivum L. under dehydration stress. *Front. Plant Sci.* **7**:1–10.

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.

Wani, A. H., Boettiger, A. N., Schorderet, P., Ergun, A., Munger, C., Sadreyev, R. I., Zhuang, X., Kingston, R. E., and Francis, N. J. (2016). Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat. Commun.* **7**.

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

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., Freidank, M., Cai, J., and Protivinsky, T. (2021). Package ' corrplot .' *CRAN* Advance Access published 2021.

Wendte, J. M., and Schmitz, R. J. (2018). Specifications of Targeting Heterochromatin Modifications in Plants. *Mol. Plant* **11**:381–387.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., and Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**:307–319.

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.

Wilkinson, A. C., Nakauchi, H., and Göttgens, B. (2017). Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. *Cell Syst.* **5**:319–331.

Woychik, N. A., and Hampsey, M. (2002). The RNA polymerase II machinery: Structure illuminates function. *Cell* **108**:453–463.

Wu, L., Mao, L., and Qi, Y. (2012). Roles of DICER-LIKE and ARGONAUTE proteins in TAS-derived small interfering RNA-triggered DNA methylation. *Plant Physiol.* **160**:990–999.

Wu, H. W., Deng, S., Xu, H., Mao, H. Z., Liu, J., Niu, Q. W., Wang, H., and Chua, N. H. (2018). A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves. *New Phytol.* **219**:1480–1491.

Xia, R., Xu, J., Arikit, S., and Meyers, B. C. (2015). Extensive families of miRNAs and PHAS loci in Norway spruce demonstrate the origins of complex phasiRNA networks in seed plants. *Mol. Biol. Evol.* **32**:2905–2918.

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.

Xie, P., and Chen, H. (2017). Mechanism of ribosome translation through mRNA secondary structures. *Int. J. Biol. Sci.* **13**:712–722.

Xie, J., Yang, X., Song, Y., Du, Q., Li, Y., Chen, J., and Zhang, D. (2017). Adaptive evolution and functional innovation of Populus-specific recently evolved microRNAs. *New Phytol.* **213**:206–219.

Xie, H., Sun, Y., Cheng, B., Xue, S., Cheng, D., Liu, L., Meng, L., and Qiang, S. (2019). Variation in ICE1 Methylation Primarily Determines Phenotypic Variation in Freezing Tolerance in Arabidopsis thaliana. *Plant Cell Physiol.* **60**:152–165.

Xu, Y., Zhang, S., Lin, S., Guo, Y., Deng, W., Zhang, Y., and Xue, Y. (2017a). WERAM: A database of writers, erasers and readers of histone acetylation and methylation in eukaryotes. *Nucleic Acids Res.* **45**:D264–D270.

Xu, Q., Song, Z., Zhu, C., Tao, C., Kang, L., Liu, W., He, F., Yan, J., and Sang, T. (2017b). Systematic comparison of lncRNAs with protein coding mRNAs in population expression and their response to environmental change. *BMC Plant Biol.* **17**:1–15.

Yamaguchi, N. (2021). Removal of H3K27me3 by JMJ Proteins Controls Plant Development and Environmental Responses in Arabidopsis. *Front. Plant Sci.* **12**:1–8.

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., Lee, S., So, J. H., Dharmasiri, S., Dharmasiri, N., Ge, L., Jensen, C., Hangarter, R., Hobbie, L., and Estelle, M. (2004). The IAA1 protein is encoded by AXR5 and is a substrate of SCF TIR1. *Plant J.* **40**:772–782.

Yang, S., Yalamanchili, H. K., Li, X., Yao, K. M., Sham, P. C., Zhang, M. Q., and Wang, J. (2011). Correlated evolution of transcription factors and their binding sites. *Bioinformatics* **27**:2972–2978.

Yang, D.-L., Zhang, G., Tang, K., Li, J., Yang, L., Huang, H., Zhang, H., and Zhu, J.-K. (2016). Dicer-independent RNA-directed DNA methylation in Arabidopsis. *Cell Res.* **26**:1264–1264.

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.

Ye, C. Y., Chen, L., Liu, C., Zhu, Q. H., and Fan, L. (2015). Widespread noncoding circular RNAs in plants. *New Phytol.* **208**:88–95.

Yeasmin, F., Yada, T., and Akimitsu, N. (2018). Micropeptides encoded in transcripts previously identified as long noncoding RNAs: A new chapter in transcriptomics and proteomics. *Front. Genet.* **9**:1–10.

Yifhar, T., Pekker, I., Peled, D., Friedlander, G., Pistunov, A., Sabban, M., Wachsman, G., Alvarez, J. P., Amsellem, Z., and Eshed, Y. (2012). Failure of the tomato trans-acting short interfering RNA program to regulate AUXIN response FACTOR3 and ARF4 underlies the wiry leaf syndrome. *Plant Cell* **24**:3575–3589.

Yin, X., Romero-Campero, F. J., de Los Reyes, P., Yan, P., Yang, J., Tian, G., Yang, X. Z., Mo, X., Zhao, S., Calonje, M., et al. (2021). H2AK121ub in Arabidopsis associates with a less accessible chromatin state at transcriptional regulation hotspots. *Nat. Commun.* **12**.

Yoshida, T., Ohama, N., Nakajima, J., Kidokoro, S., Mizoi, J., Nakashima, K., Maruyama,

K., Kim, J. M., Seki, M., Todaka, D., et al. (2011). Arabidopsis HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression. *Mol. Genet. Genomics* **286**:321–332.

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.

Yu, Y., Jia, T., and Chen, X. (2017). The 'how' and 'where' of plant microRNAs. *New Phytol.* **216**:1002–1017.

Zaher, H. S., and Green, R. (2009). Fidelity at the Molecular Level: Lessons from Protein Synthesis. *Cell* **136**:746–762.

Zanetti, M. E., Blanco, F. A., Beker, M. P., Battaglia, M., and Aguilar, O. M. (2010). A C subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean-Rhizobium etli Symbiosis. *Plant Cell* **22**:4142–4157.

Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr. Biol.* **20**:R780–R785.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W. L., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E., et al. (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell* **126**:1189–1201.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V., Pellegrini, M., Goodrich, J., and Jacobsen, S. E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLoS Biol.* **5**:1026–1035.

Zhang, H., Chen, X., Wang, C., Xu, Z., Wang, Y., Liu, X., Kang, Z., and Ji, W. (2013). Long non-coding genes implicated in response to stripe rust pathogen stress in wheat (Triticum aestivum L.). *Mol. Biol. Rep.* **40**:6245–6253.

Zhang, H., Lang, Z., and Zhu, J. K. (2018a). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**:489–506.

Zhang, A. X., Bjorklund, N., Alvare, G., Ryzdak, T., Sparling, R., Fristensky, B., Alvare, M. G., and Zhang, X. (2018b). 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, P., Li, S., and Chen, M. (2020). Characterization and Function of Circular RNAs in Plants. *Front. Mol. Biosci.* **7**:1–10.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science (80-.).* **322**:750–756.

Zhao, X., Li, J., Lian, B., Gu, H., Li, Y., and Qi, Y. (2018a). Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat. Commun.* **9**:1–12.

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al. (2018b). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* **50**:278–284.

Zhao, M., Zhang, H., Yan, H., Qiu, L., and Baskin, C. C. (2018c). Mobilization and role of starch, protein, and fat reserves during seed germination of six wild grassland species. *Front. Plant Sci.* **9**:1–11.

Zhao, J., Ajadi, A. A., Wang, Y., Tong, X., Wang, H., Tang, L., Li, Z., Shu, Y., Liu, X., Li, S.,

et al. (2020a). Genome-Wide Identification of lncRNAs During Rice Seed Development. *Genes (Basel).* **11**.

Zhao, Y., Liu, X., Tong, C., and Wu, Y. (2020b). Effect of root interaction on nodulation and nitrogen fixation ability of alfalfa in the simulated alfalfa/triticale intercropping in pots. *Sci. Rep.* **10**:1–11.

Zhou, C., Han, L., Fu, C., Wen, J., Cheng, X., Nakashima, J., Ma, J., Tang, Y., Tan, Y., Tadege, M., et al. (2013). The trans-acting short interfering RNA3 pathway and no apical meristem antagonistically regulate leaf margin development and lateral organ separation, as revealed by analysis of an argonaute7/lobed leaflet1 mutant in Medicago truncatula. *Plant Cell* **25**:4845–4862.

Zhou, Y., Romero-Campero, F. J., Gómez-ZambranoÁngeles, Turck, F., and Calonje, M. (2017). H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. *Genome Biol.* **18**:1–13.

Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-wide prediction and validation of intergenic enhancers in arabidopsis using open chromatin signatures. *Plant Cell* **27**:2415–2426.

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39**:61–69.

Zubko, E., and Meyer, P. (2007). A natural antisense transcript of the Petunia hybrida Sho gene suggests a role for an antisense mechanism in cytokinin regulation. *Plant J.* **52**:1131–1139.

Synthèse en français

La grandissante disponibilité des séquences génomiques de divers organismes vivants à changer notre vision du génome. Jusque dans les années 1940, il était proposé que la longueur du génome soit positivement corrélée avec la complexité de l'organisme (Elliot and Gregory, 2015). Cependant les différences de taille de génome entre des organismes proches, ainsi que l'existence de très long génome chez des organismes simple, noté au début des années 1950, donna naissance au dilemme du « C-value paradox » (Thomas, 1971). Après, la découverte de large région génomique ne contenant pas de gène codant résoudra le manque de corrélation entre longueur du génome et complexité de l'organisme. Cette observation soulèvera de nombreuses questions quant à l'intérêt de ces régions non codantes : comment ces séquences sont-elles apparues ? Sont-elles fonctionnelles ? Pourquoi n'existe-t-elle pas chez tous les organismes ?

Plus tard, au début des années 2000, les méthodes de séquençage à haut-débit ont révélé que la grande majorité du génome des eucaryotes est transcrit avec étonnamment, seule une infime partie de ces transcrits donnant naissance à des protéines. En effet, il a été estimé que 95% du génome humain est transcrit avec seulement 2% de transcrit traduit en protéine. Similairement, chez Arabidopsis, 71% du génome est transcrit, avec seulement 2% de transcrits codants (Lucero et al., 2020). Parmi ces transcrits non-codants, nous trouvons, des transcrits structuraux (rRNA, tRNA, snRNA et snoRNA) assurant les fonctions de base de la cellule, tel que les mécanismes de traduction. Mais aussi des petits et longs transcrits régulateurs, inférieure et supérieure à 200nt, respectivement (Ariel et al., 2015). Ces transcrits non-codants régulateurs sont impliqués dans la régulation de l'activité génique au niveau transcriptionnelle et post-transcriptionnelle.

Curieusement, les gènes produisant ces transcrits non-codants régulateurs, présentent une moindre conservation par rapport aux gènes codants entre, et au sein, des espèces. En effet, les projets « 1000 Génomes » et « 1001 génomes » chez l'homme et Arabidopsis, respectivement, ont montré que les gènes codants sont significativement mieux conservés entre individus que les gènes non codants. Aux vues du rôle régulateur des gènes non-codants, ainsi que leur faible niveau de conservation entre les organismes d'une même espèce, il a été proposé que les changements de séquences entre individus au sein des régions non-codantes reflètent l'adaptation d'un individu à son environnement. Dans ce contexte, ma thèse a pour objectif de mieux comprendre les fonctions et importance des unités de transcriptions non-codantes pour l'adaptation des plantes à leur environnement.

Le phosphate (Pi) constitue un des minéraux les plus importants pour les plantes et est absolument nécessaire à leur bon développement. Curieusement, la réponse des plantes aux manque de Pi diffère grandement entre différents écotypes d'Arabidopsis. En effet, la croissance des racines de l'écotype Col-0 s'arrête dès lors que le milieu est pauvre en Pi alors que l'écotype Ler poursuit sa croissance normalement (Reymond et al., 2006). Dans ce cadre, nous avons caractérisé les transcriptomes des écotypes Col-0 et Ler au cours d'une cinétique de réponse au manque de Pi. Par des analyses bioinformatiques, nous avons identifié de nombreux transcrits non-codants, conservés, enrichis, ou spécifiques d'un écotype, et répondant ou non au manque de Pi. De plus, nous avons observé que contrairement aux transcrits codants, les transcrits non-codants sont significativement plus spécifiques d'un écotype. Enfin, nous avons fonctionnellement étudié plusieurs longs ARN non-codants (lncRNA) et trouvé 2 lncRNAs enrichit chez Col-0 ou Ler, impliqué dans la régulation de la croissance racinaire. Ensemble, ces travaux renforcent l'importance des lncRNAs dans le contrôle quantitatif de l'expression des gènes et de l'architecture racinaire, tout en suggérant qu'ils jouent un rôle dans l'adaptation locale des plantes à leur environnement.

Les lncRNAs peuvent réguler l'activité des gènes en cis ou en trans, en modifiant l'expression d'un gène voisin ou distant, respectivement. Chez les plantes, quelques lncRNAs agissant en cis sont impliqué dans le contrôle de la floraison (lncRNAs dérivé du locus FLC), transport de l'auxine (APOLO), rendement en grain chez le riz (LAIR) et tolérance au froid (SVALKA) (Chen et al., 2021). A ce jour, hormis APOLO (Ariel et al., 2014 ; Ariel et al., 2020), aucun lncRNA module l'architecture de la racine via le contrôle de l'expression de son gène voisin. Ainsi, nous avons exploré l'analyse transcriptomique réalisé chez les écotypes Col-0 et Ler au cours d'une cinétique de manque de Pi, afin de détecter des corrélations d'expression positives entre un lncRNA et son gène voisin, impliqué dans le développement ou la croissance de la racine. Sortie de ces analyses, nous avons isolé LATERALINC, un lncRNA modulant la croissance des racines latérales et potentiellement impliqué dans la régulation de son gène voisin NF-YB3, décrit comme impliqué dans la réponse des plantes aux fortes températures (Sato et al., 2019). De plus, nous avons identifié MARS, un lncRNA régulant l'expression d'un groupe de gènes organisé en cluster et impliqué dans la biosynthèse et la métabolisation du marneral, un métabolite secondaire important pour le développement de la plante (Field et al., 2011 ; Go et al., 2012). La majeure partie de mon travail de thèse à consister à une caractérisation moléculaire approfondie du lncRNA MARS, et des mécanismes épigénétiques qu'il utilise pour moduler l'environnement chromatinien du cluster marneral. Brièvement, nous avons démontré que le lncRNA MARS module la compaction et la conformation tridimensionnelle de la chromatine du cluster marneral, afin de moduler la sensibilité des gènes à un stimulus ABA. Nous avons montré que la dérégulation d'expression de MARS perturbe la germination de la graine et la croissance des racines en réponse à des stress osmotiques. Enfin, nous avons observé que de nombreux clusters de gènes, autre que le marneral, contiennent des unités de transcription non-codantes, également corrélé avec les gènes du cluster, suggérant que le contrôle de la co-régulation des gènes en cluster par des lncRNA pourrait être un mécanisme conservé chez les plantes.

Ensemble, ma thèse souligne l'importance des lncRNAs pour l'adaptation des plantes à leur environnement, et illustre un complexe mécanisme de régulation génique chez les plantes impliquant lncRNA, statut épigénétique et conformation tridimensionnelle de la chromatine.

ÉCOLE DOCTORALE
Sciences du végétal : du gène à l'écosystème (SEVE)
université PARIS-SACLAY

Titre : ARN-non codants chez les plantes : régulateur important de la croissance racinaire et de l'activité des gènes
Mots clés : ARN non-codant, croissance racinaire, chromatine

Résumé : Le progrès des méthodes de séquençage à haut-débit au début des années 2000 a révélé que la grande majorité du génome des eucaryotes est transcrit. Étonnamment, seule une infime partie de ces transcrits donnent naissance à des protéines. En effet, il a été estimé que 95% du génome humain est transcrit avec seulement 2% de transcrit traduit en protéine. Cette observation soutient l'hypothèse que les transcrits non-codants constituent d'importants régulateurs de la vie de la cellule. Curieusement, les gènes non-codants présentent une moindre conservation par rapport aux gènes codants, entre et au sein des espèces. Par exemple, chez les plantes, le séquençage d'accessions d'*Arabidopsis* durant le projet 1001 génomes montre des variations de séquences plus fréquentes dans les régions non codantes par rapport aux régions codantes. Au vu du rôle régulateur des gènes non-codants, il a été proposé que ces changements reflètent l'adaptation d'un individu à son environnement. Dans ce contexte, ma thèse a pour objectif de mieux comprendre les fonctions et l'importance des unités de transcriptions non-codantes pour l'adaptation des plantes à leur environnement. Le phosphate (Pi) constitue un des minéraux les plus importants pour les plantes et est absolument nécessaire à leur bon développement. Pour étudier les fonctions des transcrits non-codants durant une pénurie de Pi, nous avons analysé les écotypes Col-0 et L*er* présentant une réponse surprenamment contrastée au manque de Pi (dans un milieu pauvre en Pi, la croissance de la racine principale s'arrête immédiatement chez Col-0, alors qu'elle se poursuit chez l'écotype L*er*). Par rapport à l'activité comparable des gènes codants aux cours de la cinétique de carence en Pi, les gènes non-codants présentent une dynamique très différente en fonction de l'écotype. Élément intéressant, la dérégulation d'expression de deux longs ARN non-codants (lncRNAs) altère l'architecture racinaire, soulignant que les lncRNAs constituent d'importants régulateurs de la croissance racinaire. Notre comparaison du transcriptome de ces deux accessions d'*Arabidopsis* renforce l'importance des transcrits non-codants pour l'adaptation locale des plantes à leur environnement. Profitant de notre étude transcriptomique, nous avons cherché d'autres lncRNAs impliqués dans le contrôle de l'architecture racinaire. Prenant en compte que de nombreux lncRNAs régulent l'activité de leur gène voisin, nous avons recherché des corrélations d'expression positives entre des lncRNAs et leurs gènes voisins, en se focalisant sur les gènes impliqués dans la croissance et le développement de la racine. Parmi, les candidats sélectionnés, nous avons trouvé un lncRNA, que nous avons surnommé *MARneral Silencing* (*MARS*), régulant l'expression d'un groupe de gènes organisés en cluster et impliqué dans la biosynthèse du marneral. Le marneral est un triterpène directement impliqué dans la croissance et le développement de la plante. Nous avons montré que ce lncRNA module la reprogrammation épigénétique du cluster marneral en réponse à l'ABA, influençant directement la réponse des gènes du cluster à ce stimulus. Ainsi, ma thèse souligne l'importance des lncRNAs pour l'adaptation des plantes à leur environnement, et illustre un complexe mécanisme de régulation génique chez les plantes impliquant lncRNA, statut épigénétique et conformation tridimensionnelle de la chromatine.

Title: Plant non-coding RNAs: important regulator of root growth and gene activity
Keywords: non-coding RNA, root growth, chromatin

Abstract: The development of high throughput sequencing methods in the early 2000's revealed that the eukaryotic genomes are tremendously transcribed even though a small part of transcripts give rise to proteins. For example, it has been estimated that 95% of the human genome is transcribed whereas only 2% of the transcripts will produce protein. This surprising observation strengthens the idea that non-coding transcripts are an important regulator of cell life. Interestingly, the non-coding genes show lower conservation than the coding genes between species and even within a species. For example, in plants, the sequencing of *Arabidopsis* accessions during the 1001 genome project demonstrated more frequent sequence variation in non-coding regions as compared to coding regions. Non-coding genes play regulatory roles and it has been proposed that their modification reflect the adaptation of an individual to its environment. In this context, my thesis aims to better understand the function and importance of non-coding transcriptional units for plant adaptation to their environment. Phosphate (Pi) constitutes one of the most important minerals for plants and is absolutely needed for a proper plant development. To investigate the function of non-coding transcript during a shortage of Pi we analyzed Col-0 and L*er* ecotype that present a strikingly contrasted root growth response when subjected to Pi starvation: in medium with low amount of Pi, Col-0 immediately stops its main root growth whereas L*er* ecotype does not. Compared to the similar protein coding genes activity during the Pi shortage, the non-coding genes transcriptional activity shows strong differences between ecotypes. Interestingly, we found that the deregulation of two long non-coding RNAs (lncRNAs) alters root architecture, strengthening that lncRNAs molecules constitute important regulators of root growth. Our comparison of the transcriptomes of these two *Arabidopsis* accessions reinforces the relevance of non-coding transcripts in the plant local adaptation to its environment. Taking advantage of our transcriptomic survey, we look for other functional lncRNAs molecules participating in the control of root architecture. Taking into account that many lncRNAs are involved in the regulation of their neighboring genes, we look for positive correlation of expression between lncRNA and their neighboring coding genes, focusing on coding genes directly involved in root growth and development. Among our putative candidates we found a lncRNA, which we named *MARneral Silencing* (*MARS*), involved in the regulation of marneral cluster genes. The marneral is a triterpene produced by three genes organized in cluster directly involved in plant growth and development. We show that this lncRNA is involved in the epigenetic reprogramming of the marneral cluster in response to an ABA cue, directly influencing marneral genes response to ABA stimulus. Together, my thesis strengthens the importance of lncRNAs on the plant adaptation to its environment, and illustrates a complex chromatin-based mechanism involving lncRNA, epigenetic state and 3D chromatin conformation.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France